Detection and Localisation of Pipe Bursts in a District Metered Area Using an Online Hydraulic Model

Submitted by **Olanrewaju Isaac Okeya**

to the University of Exeter as a thesis for the degree of

Doctor of Engineering in Water Engineering

March 2018

This thesis is available for library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

This page is intentionally left blank

ABSTRACT

This thesis presents a research work on the development of new methodology for near-real-time detection and localisation of pipe bursts in a Water Distribution System (WDS) at the District Meters Area (DMA) level. The methodology makes use of online hydraulic model coupled with a demand forecasting methodology and several statistical techniques to process the hydraulic meters data (i.e., flows and pressures) coming from the field at regular time intervals (i.e. every 15 minutes). Once the detection part of the methodology identifies a potential burst occurrence in a system it raises an alarm. This is followed by the application of the burst localisation methodology to approximately locate the event within the District Metered Area (DMA).

The online hydraulic model is based on data assimilation methodology coupled with a short-term Water Demand Forecasting Model (WDFM) based on Multi-Linear Regression. Three data assimilation methods were tested in the thesis, namely the iterative Kalman Filter method, the Ensemble Kalman Filter method and the Particle Filter method. The iterative Kalman Filter (i-KF) method was eventually chosen for the online hydraulic model based on the best overall trade-off between water system state prediction accuracy and computational efficiency.

The online hydraulic model created this way was coupled with the Statistical Process Control (SPC) technique and a newly developed burst detection metric based on the moving average residuals between the predicted and observed hydraulic states (flows/pressures). Two new SPC-based charts with associated generic set of control rules for analysing burst detection metric values over consecutive time steps were introduced to raise burst alarms in a reliable and timely fashion. The SPC rules and relevant thresholds were determined offline by performing appropriate statistical analysis of residuals.

The above was followed by the development of the new methodology for online burst localisation. The methodology integrates the information on burst detection metric values obtained during the detection stage with the new sensitivity matrix developed offline and hydraulic model runs used to simulate potential bursts to identify the most

likely burst location in the pipe network. A new data algorithm for estimating the 'normal' DMA demand and burst flow during the burst period is developed and used for localisation. A new data algorithm for statistical analysis of flow and pressure data was also developed and used to determine the approximate burst area by producing a list of top ten suspected burst location nodes.

The above novel methodologies for burst detection and localisation were applied to two real-life District Metred Areas in the United Kingdom (UK) with artificially generated flow and pressure observations and assumed bursts. The results obtained this way show that the developed methodology detects pipe bursts in a reliable and timely fashion, provides good estimate of a burst flow and accurately approximately locates the burst within a DMA. In addition, the results obtained show the potential of the methodology described here for online burst detection and localisation in assisting Water Companies (WCs) to conserve water, save energy and money. It can also enhance the UK WCs' profile customer satisfaction, improve operational efficiency and improve the OFWAT's Service Incentive Mechanism (SIM) scores.

This page is intentionally left blank

ACKNOWLEDGEMENTS

I would like to use this opportunity to acknowledge all the people listed below for their contribution to this research.

Firstly, I would like to express the deepest gratitude to my great supervisor, Professor Zoran Kapelan and wonderful Dr. Christopher Hutton. These special thanks are their continuous support and guidance through the research, and also for their patience and cordial relationship. With their expansive knowledge and enthusiasm, they have provided me with invaluable advice.

I would like to acknowledge the financial support received through the STREAM project funded by the Engineering and Physical Sciences Research Council *(*EPSRC*)* and Industrial Collaborator, United Utilities (UU). My further thanks go to the team members of UU Network Modelling team especially Ms. Devina Naga, and Dr. Michele Romano for their generous support, wonderful friendship and providing me intensive information during this research.

I would like to acknowledge all the dedicated and talented members of STREAM-Industrial Doctoral Centre and University of Exeter, Centre for Water Systems, with whom I had the pleasure of working in an inspiring research environment.

In additional, I wish to express my genuine gratitude to my wonderful parents (Chief Isaac and Dr. Elizabeth Okeya), my loving siblings: Temitayo, Olufemi, Olamide and her husband, Dr. Muyiwa Owoniyi for their endless support and constant encouragement they have given me throughout the research period.

My gratitude extends to my wife, Princess Adewunmi who taught that "*I can do all things through Christ who gives me strength*". As a special person in my life, her love, support and encouragement made the journey to complete this research goal easier.

This page is intentionally left blank

TABLE OF CONTENTS

This page is intentionally left blank

LIST OF FIGURES

This page is intentionally left blank

LIST OF TABLES

This page is intentionally left blank

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BDLM	Burst Detection and Localisation Methodology
BDM	Burst Detection Methodology
BLM	Burst Localisation Methodology
CUSUM	Cumulative Sum
DA	Data Assimilation
DMA	District Metered Area
DMF	Demand Multiplication Factor
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EPA	Environmental Protection Agency
EMA	Exponentially Moving Average
ERS	Event Recognition System
EWMA	Exponentially Weighted Moving Average
GIS	Geographic Information System
i-KF	iterative Kalman Filter
KF	Kalman Filter
HM	Hydraulic Model
ISDA	Individual Signal Detection Analysis
MA	Moving Average
MAR	Moving Average of Residual
MAE	Mean Absolute Error
MLE	Maximum Likelihood Estimation
MLR	Multi-Linear Regression
MNF	Minimum Night Flow
NNF	Net Night Flow
OFWAT	Water Services Regulation Authority
PF	Particle Filter
P-MAR	Product of Moving Average of Residuals
PSDA	Products of Signals Detection Analysis
RMSE	Root Mean Squared Error

SA	Statistical Analysis
SCADA	Supervisory Control and Data Acquisition
SIM	Service Incentive Mechanism
SPC	Statistical Process Control
UU	United Utilities
WC	Water Company
WMA	Weighted Moving Average
WDS	Water Distribution Systems
WSS	Water Supply Systems
WSZ	Water Supply Zone
YWS	Yorkshire Water Service

This page is intentionally left blank

1 INTRODUCTION

1.1 MOTIVATION

Accessible potable water is one of the critical essentials to human survival and economic growth of today's society. In most of the developed countries, the water from the environment is treated to meet the drinking water standards and then supplied to the domestic, retail and industrial users. Majority of the potable water is mostly used for toilet flushing, washing, irrigation and industrial uses. However, majority of people in this world who do not have sufficient access to potable water but use poor quality water instead. Therefore, many governments especially in developing countries want their populace to have access to potable water which is causing a rapid expansion of water networks. Such expansion of new water networks will put pressure on existing Water Distribution Systems (WDS) infrastructure and limited water resources. The rapid population growth (increasing water demand) and climate change due to the global warming will put further pressure on the current WDS infrastructure and water resources. It is also evident that water supply is an ongoing critical issue for the future generations in both urban and rural areas. The Water Services Regulation Authority (OFWAT) 2011 report highlights additional challenges that the United Kingdom (UK) Water Companies (WCs) face for the next 25 years and they are as follows (OFWAT, 2011):

1. Consumers' high expectation of water services due to high standards of living.
2. Stringent environmental regulations limiting the amount of raw water that can be abstracted from boreholes and rivers due to climate change and Environmental Agency (EA) aquatic ecosystems protection.
3. High percentage of water losses via unexpected pipe bursts and leaks due to aging water infrastructure.

The WCs all over the world face the choice to either build more resources (i.e. water storages, water networks, water treatments) now or improve the existing management of WDS. The construction cost of building the new resources (i.e. water storages, water network, water treatments) has risen over the last decade. The opportunity to develop sustainable and viable water resources is also limited due to climate change and

budget constraint. Many WCs in the world aim to maximise their current WDS and focus on water conservation efforts. It is widely claimed that the amount of water loss from the WDS ranging between 20% and 40% (Environment Agency, et al., 2012). Thus, the reduction of water loss is a pressing priority for the WCs all over the world. Even though, the WCs in the world are actively rehabilitating and maintaining the WDS infrastructure, but the abnormal events (i.e., pipe bursts/leaks) still persist. These abnormal events occurred due to the ageing WDS infrastructure, pipe crack and loose joints due to poor workmanship. It is pertinent to mention that due to the stochastic nature of the pipe bursts, it is almost impossible to predict the future burst events and their possible locations. The potential solution to reduce water loss is via early detection and location of bursts in near real-time. The development of a real-time tool/system that can perform such detection and location analysis is paramount.

The potential benefits of implementing a near real-time pipe burst detection and localisation methodology are as follows: 1) managing the WDS efficiently under abnormal events to prevent serious damages or long interruptions in service (Bicik, et al., 2009) and 2) it can result in significant financial savings for the UK WCs (Romano, et al., 2013). The financial savings can be the result of following reasons; (i) reducing the associated costs of the water loss' production (i.e., chemical) and distribution (i.e., pumping); (ii) reducing the operational costs of detecting and locating pipe burst; (iii) reducing associated costs of compensating affected local businesses due to pipe bursts; (iv) limiting compensation payments for the damaged infrastructures and properties due to pipe bursts; and (v) avoiding OFWAT's penalties due to the poor OFWAT's serviceability scores.

The UK WCs recognise the aforementioned potential benefits which inspire them to invest in real-time management of WDS research including this research work. The UK WCs have a commitment of reducing carbon footprint, minimising the number of interruptions to water supply, promoting water conservation and reforming their management of Supervisory Control and Data Acquisition (SCADA) system to attain better sustainability of WDS. These UK WCs commitments are appreciated by the public and national water regulatory bodies (i.e., OFWAT). This can further enhance

the UK WCs' profile customer satisfaction and limit the OFWAT's Service Incentive Mechanism (SIM) scores.

1.2 BACKGROUND

There are many pipe burst detection and localisation techniques developed on different principles (Puust, et al., 2010). Majority of the techniques are developed without the consideration of detecting and locating pipe bursts online. The UK WCs are already started to install multiple permanent hydraulic meters (i.e., flow meters and pressure meters) to monitor WDSs. The data gathered from the multiple hydraulic meters at different observation locations can be assessed with hydraulic model to provide WDSs condition status (i.e., normal, abnormal). The changes in WDS sections configuration are due to various factors such as WDS boundary modification, pressure optimisation scheme and addition of new properties layout. These changes can affect the performance of the existing pipe burst detection and localisation techniques. A new and more efficient technique for the online detection and location of pipe burst events in WDSs is required.

Detecting and locating pipe bursts in WDS sections (in real-time) remains a huge challenge for many UK WCs. Many UK WCs have implemented pipe burst detection systems (Mounce et al., 2010; Romano et al., 2013). The developed pipe burst detection techniques are capable of detecting bursts in real-time but they can become redundant if there is a WDS boundary change (i.e., valve closure, DMA configuration change). Currently, vast majority of the pipe bursts detection still depends on customer contacts. Consequently, the UK WCs and some interested water researchers shift their research focus to utilising online hydraulic modelling of WDSs for pipe burst detection and localisation. This represents an opportunity for application of statistical techniques and model-data driven pipe burst detection and localisation methods, which at the same time have to cope with a large number of data and complex WDS configuration. The early detection and location of pipe bursts in real-time can be achieved by using the method proposed in this thesis. Such early detection and localisation of pipe bursts can improve the lead time for burst repair and manage the WDS effectively under abnormal conditions.

The burst detection and localisation techniques that make use of hydraulic model to solve the pipe burst detection and location problem by statistical analysis. These techniques have recently started to appear (e.g., Skworcow and Ulanicki, 2011; Jung and Lansey, 2013; Kang and Lansey, 2014) mainly due to the aspiration of the WCs to make use of a hydraulic model and data (i.e., flows and pressures) from multiple hydraulic meters via SCADA systems.

Hydraulic model is usually used to predict flows and pressures based on the assumptions of a fixed water demand patterns, estimated flow rates, pump scheduling and tank or reservoir elevation (Machell, et al., 2010). A calibrated offline hydraulic model is often used to predict the WDSs hydraulic states or status (i.e., abnormal status due to low pressure, pipe bursts). It is well established that both hydraulic model and observations (i.e., flows, pressures) can be error-prone (Shang, et al., 2006). Therefore, an online hydraulic modelling of WDSs is required to reduce such hydraulic model and observations' error via Data Assimilation (DA) method.

Hutton, et al., (2012), Hatchett et al., (2009) and Shang et al, (2006) highlight many developed DA methods are already applied in hydraulic modelling but they are now being experimented for operational level due to the data scarcity. The operational hydraulic modelling of WDSs is still at the preliminary stages in the UK WCs (Bicik, et al., 2009) because of high computational inefficiency for a large WDS (Shang, et al., 2006; Pauwels & De Lannoy, 2009); hydraulic data quality; complexity of the WDS configuration (Preis, et al., 2011); limited number of hydraulic meters (Romano, et al., 2013) due to budget constraint; hydraulic model inaccuracy due to incomplete and inaccurate data stored in Geographic Information System (GIS) database. Some of the abovementioned limitations can be overcome as the WDS technologies advance or become cheaper over time.

The combination of the following components: hydraulic data (i.e., flows/pressures) and model, a DA method and statistical techniques seems to be a promising combination to observe WDS online. The combination of an online hydraulic and DA method and statistical techniques presents several advantages over other numerical techniques such as: the transient analysis-based (e.g., Kapelan et al., 2003); steady

state analysis-based (e.g., Wu et al., 2010); negative pressure wave-based (e.g., Srirangarajan et al., 2012) and measurement-based statistical/ Artificial Intelligent (AI)-based techniques (e.g., Mounce et al., 2010, 2011; Romano et al., 2012, 2013). For example, model-based burst detection and localisation techniques can still be used if the WDS boundary changes compared to the measurement-based statistical and/or AI-based techniques. Low frequency (e.g., 15 minutes) of hydraulic data is good enough for data analysis compared to the transient analysis-based or negative pressure wave-based method. The model-based burst detection and localisation techniques combined with a DA method and statistical techniques can be improved to detect and locate pipe bursts in a timely and reliable manner. In addition, there is no online model-based burst detection and localisation techniques that utilise a DA method and statistical techniques to detect pipe bursts and pinpoint the burst area within a DMA online. They are mostly tested on limited number of artificial bursts generated via a hydraulic model. This research work discusses the development, implementation and application of a model-based pipe burst detection and localisation technique to detect and locate pipe burst in real-time. The work presented in this thesis is, therefore, relevant for the water industry and has the potential to assist WCs' leakage team.

1.3 THESIS SCOPE AND OBJECTIVES

The overall objective of this thesis is to develop and test and also demonstrate an online hydraulic model-based burst detection and localisation methodology for detecting and locating pipe bursts in WDS sections in real-time.

The following research questions are asked to achieve the aforementioned thesis objective.

1. What are the current capabilities and limitations of (1) developed burst detection and localisation methodologies and (2) data assimilation methods in the literature?
2. Do online hydraulic models (with a data assimilation) improve WDS prediction accuracy when compared to existing offline WDS model?

3. If yes to the above question 2, which online hydraulic models perform reasonably well?
4. Can an AI and/or statistical technique complement the selected online hydraulic model to detect and locate WDS pipe bursts in WDS?
5. How do the proposed burst detection and localisation methods fare on both artificial/real-life case study and flow/pressure data?
6. What are the key findings from the case studies?

The aforementioned research questions are answered in this thesis through the following specific objectives:

1. To review relevant literature on burst detection and localisation methods including their capabilities and limitations and also review relevant data assimilation methods and their practicality in WDS. This is to identify the knowledge gap in the literature;
2. To investigate and develop an online hydraulic model(s) to be used for WDS state estimation in near real-time with the aim to obtain improved prediction accuracy when compared to an offline WDS model. An integral part of this work is the development of a simple water demand forecasting model to forecast water demands every 15 minutes. Online hydraulic model is a combination of a hydraulic model, demand forecasting model and a DA method (i.e., Kalman Filter method, Ensemble Kalman Filter method and Particle Fitter method);
3. To compare performances of alternative online hydraulic models and to decide which online hydraulic model to take forward for the development of burst detection and location methods;
4. To find effective statistical techniques to detect and locate pipe burst using to forecast water demands, correct hydraulic model predictions and detect and locate pipe bursts. This involves utilising online hydraulic model (from step above step 3) and investigating the detection metrics, Statistical Process Control (SPC) Rules to detect pipe bursts and ultimately used to raise detection alarms. Then develop a statistical analysis-based process that make use of offline sensitivity matrix, flow/pressure meters data to approximate the burst flow and area in real-time;

5. To test and demonstrate the robustness the proposed detection and localisation methodology in real time when applied on a number of different real-life UK DMAs with both artificial and real-life flow/pressure data;
6. To review the robustness of the proposed burst detection and localisation methodology in two case studies (two different WDS sections' configuration) including the computational cost to detect and locate pipe bursts.

1.4 THESIS STRUCTURE

This thesis is divided into seven chapters including this introduction.

In Chapter 2, a review of the relevant literature is provided. The review covers key areas of pipe burst detection and localisation techniques for detecting and locating pipe bursts in WDSs. It also highlights their individual advantages and disadvantages. The reviewed techniques are mainly hydraulic techniques which make use of hydraulic data and/or hydraulic model. Data Assimilation (DA) methodologies are also reviewed. A comparison of DA methods is presented.

In Chapter 3, the concept behind the data assimilation method and their variants are described. The integration of a demand forecasting model, a data assimilation method and hydraulic model to make an online hydraulic model is explained.

In Chapter 4, the results of case studies for offline and online hydraulic modelling are presented. The key performances between the selected online hydraulic models are highlighted including computational cost and flow/pressure correction/prediction statistics. The chapter concludes with selecting an online hydraulic model for the development of burst detection and location methods.

In Chapter 5, first the overall methodology for online pipe burst detection and localisation methodology is introduced and its individual constituents are described. The development of water demand forecasting model based on multi-linear regression analysis is described. The formulation of an online hydraulic model is explained. Suitable models to detect pipe bursts and approximate burst area in WDSs are then

presented. This chapter also presents the development of control rules for generating burst alarm and the details of developed technique for estimating burst flow and pinpointing burst nodes in the methodology.

In Chapter 6, a number of case studies to illustrate the proposed burst detection and localisation methodologies are presented. First, the pipe burst detection capabilities of the developed methodology that makes use of artificial hydraulic data from observation locations in one real-life UK DMA are tested and demonstrated. The artificial flow and pressure data are generated via hydraulic model based on the real-life inflow and outflow data. The performances of the developed methodology based on different number of hydraulic meters from different observation locations are reviewed. Then the performance of the detection methodology is tested and reviewed on a real-life DMA and real life hydraulic data (mostly pressure data). Second, the pipe burst localisation capabilities of the developed methodology within a DMA using hydraulic model, statistical techniques and artificial hydraulic data (via hydraulic model) is illustrated on a set of artificial burst locations in one real life DMA. The performance of the burst localisation methodology on a real-life DMA and real life hydraulic data (mostly pressure data) is also reviewed.

In Chapter 7, the key findings of this thesis are summarised, general conclusions are made and directions for the future work

Appendices are included at the end of the thesis.

2 LITERATURE REVIEW

This chapter provides a review of literature relevant to pipe burst detection and localisation techniques with focus on near real-time methodologies. Literature dealing with data assimilation methodologies are also reviewed to establish grounds for development of an online hydraulic modelling of WDS. The aforementioned techniques are reviewed to establish a context for the methodology presented in this thesis.

The chapter is organised as follows:

- Section 2.1 reviews the available burst detection and localisation techniques that have been applied to WDS including their capabilities;
- Section 2.2 reviews DA methods' capabilities and limitations including their implementation in WDS hydraulic modelling;
- Section 2.3 provided a summary of the literature review is including the main conclusions and research gaps.

In the WDS-related literature, the following terminologies Water Supply System (WSS), Water Supply Network (WSN), Water Distribution System (WDS) and Water Distribution Network (WDN) are often used interchangeably. Hence it is pertinent to define the abovementioned terminologies here.

- Water Supply System (WSS) or Water Supply Network (WSN) is an infrastructure for the collection, transmission, treatment, storage, and distribution of water for domestic (homes) use, commercial and industry use and also irrigation use, firefighting use and street flushing use.
- Water Distribution System (WDS) or Water Distribution Network (WDN) is an infrastructure for the distribution of treated water to the point of consumption (potable water).

In this thesis, WDS is used to represent WSS/WSN/WDN. Other terminologies that have been used interchangeably are pipe burst and pipe leak. Pipe burst and pipe leak are defined as follows:

- Pipe burst is an unrecoverable water loss to the environment while interrupting water supply to water customers.

- Pipe leak is an unrecoverable water loss to the environment without interrupting water supply to water customers.

However, in this thesis, pipe burst or pipe leak is defined as unrecoverable and unplanned water loss to the environment from the WDS pipes including pipe fittings.

2.1 BURST DETECTION AND LOCALISATION METHODOLOGIES

2.1.1 Introduction

The detection and location of WDS pipe bursts/leaks are important to WCs around the world due to its opportunity to conserve raw/treated water and save associated cost. The detection and localisation of pipe bursts/leaks in a WDS section (i.e., District Metered Areas (DMAs)) still remains a challenging task due to the stochastic nature of a WDS section. The increasingly frequent installation of plastic pipes nowadays (especially poly-ethylene pipes) in DMAs makes it difficult for some hardware-based techniques (i.e., acoustic equipment-based) to locate pipe bursts (Romano et al., 2013; Bicik et al., 2010). Even though the hardware-based techniques have improved (Puust et al., 2010) but they are costly, labour-intensive, and slow to run. Given this and the fact this thesis aims to develop a numerical model pipe burst detection and location, the literature review on hardware-based techniques is not considered here. However, the information on hardware-based techniques can be found in Li et al. (2015).

Misiunas (2005), Bicik (2010) and Romano (2012) in their PhD theses provided an intensive review of burst detection and location techniques in pipelines and pipe networks. A broad range of methods was considered by the authors, including traditional techniques for burst/leak detection and location, such as flow mass balance (also known as water audits), transient-based method, acoustic logging, ground penetrating radars, tracer gas-based technique and step-tests method.

This section only covers the hydraulic techniques developed to detect and locate bursts/leaks since the contribution of this thesis lies in the combination of a statistical based technique, a data assimilation method and hydraulic model. Hydraulic

techniques make use of hydraulic data (i.e., flow and pressure data) to review the WDS status (i.e., normal – no burst or abnormal - burst). Many of the hydraulic techniques are currently in stage of research and development or applied in a small section of WDS or simple hydraulic model with various successes. The capabilities and limitations of each hydraulic technique are reviewed for real-time burst detection and localisation in a WDS section.

The reviewed hydraulic techniques can be categorised into two groups namely:

- Data-driven (i.e. data analytics type) techniques;
- Hydraulic model-based techniques;

The above techniques are reviewed in Section 2.2.2 and 2.2.3 respectively and then summarised in Section 2.2.4.

2.1.2 Data-Driven Techniques

The data-based techniques analyse the meters signals' values from a SCADA system in near real-time to detect abnormal flows or pressures. These techniques do not make use of hydraulic model, i.e. physically based models to estimate hydraulic states (flow or pressure) of a WDS. Data-driven techniques make use of statistical or similar analysis of hydraulic/other data to extract meaningful information concerning the WDS. Some data-driven techniques make use of sophisticated techniques such as AI (Russell and Norvig, 2009; Holland, 1975) and Statistical Analysis (SA) methods (Atkinson et al., 2007; Nelder, 1990). These AI/SA techniques are employed to improve the reliability of pipe burst detection and localisation.

It is well established that an abnormal event such as burst increases the flow rate and decreases the pressure upstream of the burst/leak. Therefore, the flow meters and pressure meters located at the upstream of the burst/leak or burst are affected. These changes in flow/pressure are normally used to identify the burst/leak occurrence.

The Minimum Night Flow (MNF) (UKWIR, 1994 and 2011) monitoring is the simplest method employed by the UK WCs to detect pipe bursts/leaks. The MNF monitoring

technique involves comparing the current MNF to the acceptable MNF of the DMA at discrete time intervals (e.g., hourly, daily). The disadvantage of such method is the leak is detected only in the following day, after analysing night flow and pressure data. There is no standard mechanism to check if the MNF values are valid. Hence the UK WCs usually compare the current and historical flow/pressure for the same time period and day. This method is often referred as flow/pressure trending to detect existing pipe bursts/leaks based on the difference between the current and historical flow/pressure.

Wang et al. (1993) proposed a method of leak detection based on autoregressive modelling. This method requires pressure measurements from 4 different pressure meters with a sampling frequency of 50 Hz. Two pressure meters were placed at each end of the pipeline. A leak is detected by analysing the time sequences of the pressure gradient at the inlet and outlet of the pipeline. It was shown that a 0.5% leakage in a 120 m long pipe can only be reliably and almost instantly detected by this method. The performance of this method on a simple/complex WDS section is yet to be established. Therefore, its detection capability and limitation remain unknown.

The American Water Works Association (AWWA) and the International Water Association (IWA) published water audit guidelines in 1999 and 2000 respectively. The water audit guidelines (AWWA, 1999; IWA, 2000) assist leak technicians to calculate the amount of lost water based on the estimations of the water produced at the sources, water imported and exported at the WDS section boundary. The water audit provides an overview of water distribution in a WDS section especially an area that is experience high leakage. This approach can provide more accurate leakage estimates if there are continuous high frequency measurements. However, high frequency data are rarely gathered for WDS section monitoring due to high transmission cost and large computer data memory requirement. This technique mostly works well during the night. Therefore, it is not feasible for online burst detection.

Zhang (2001) presented an optimum sequential analysis technique (Sequential Probability Ratio Test) to detect changes in the inlet and outlet flow and pressure measurements. Both flow and pressure measurements were performed with a 30s sampling interval. The system was implemented on a 37 km long propylene pipeline and was shown to detect pipe bursts/leaks (% of average WDS section demand) with

less than 20% error in position. The detection time was less than 20 minutes. The method has been trailed on a simple WDS pipelines hence its capability in a complex WDS section is unknown.

Mounce (2002) proposed the use of a Mixture Density Artificial Neural Network (MD-ANN) (Bishop, 1994) to predict flow value 24 hours ahead based on historical flow data which was then compared to the observed flow data. The predicted and observed flow values were analysed by a classification module. This classification module gives the level of abnormalities in the observed flow value by using a binary leak/no-leak or burst/no-burst indicator. The main strength of the MD-ANN model is the opportunity to use time series data to forecast both flow/pressure reliably. However, the time window used in the experiment is between 12 hours and 24 hours which is deemed too long for burst detection in near real-time.

Mounce and Machell (2006) studied different type of ANNs for classifying flow and pressure data pattern under normal and abnormal conditions. The two ANNs studied were Static ANN and Time-Delay ANN (Haykin, 1994). It was found that the Time-Delayed ANN is capable of learning the simulated leaks' patterns of the engineered events and 75% of the leaks were detected. These engineering events involve opening of a fire hydrant for a specified time period. The classification task of leak is proven to be difficult unless further details of engineered events are known. In reality, the leak-size, causes of the leak and starting time for a leak are likely to be unknown.

Mounce et al. (2006) developed a pipe/leak burst detection methodology that uses the combination of ANN and Fuzzy Inference System (FIS) for an online detection of bursts/leaks at a DMA level. The ANN is based on a mixture density network and trained on a real historical flow data to predict the next 24 hours flow profile and then apply a FIS to detect abnormal flows. The FIS compares the observed flow values with predicted flows over a selected time windows to check if there is abnormal flow within the flow observations. This technique is applied offline to a real WDS flow observations and detects 44% of abnormal flow that were correlated to definite bursts. Based on the results, Mounce et al. (2007) further improved the proposed detection methodology to accurately estimate the average flow of the bursts/leaks. The authors validated the

method using the historical flow data and repaired bursts/leaks data including engineered events. The improved burst/leak detection technique was tested on a real-life WDS section with additional rules in a live environment (Mounce & Boxall, 2010). It detects 36% of raised alarms are correlated to definite bursts and 38% of high abnormal demands (Mounce & Boxall, 2010). Again, the issue with this burst/leak detection methodology is the time taken to generate alarm when abnormal flow or pipe burst occurs. The time window to raise a pipe burst/leak alarm is still between 12 hours and 24 hours which is a significant drawback to the developed burst/leak detection approach.

Mounce et al. (2011) applied a Support Vector Regression (SVR) technique to detect pipe bursts/leaks using the unusual differences between WDS flow/pressure observations and predictions. Prior to applying the SVR to output flow/pressure predictions, historical flow and pressure observations are used to train SVR method. The technique makes use of both flow and pressure observations from a real WDS without using a hydraulic model to detect pipe bursts ranged in size from approximately 10% to 50% of the average daily maximum flow. The technique is applied in a real WDS data to detect burst offline. It is found that SVR methodology raise alarm faster than the hybrid ANN/FIS method when there is a burst. The results show 22% of the raised alarms were false alarms which are higher than the previous studies reported by the same authors (i.e., 18%-false alarms was reported in study by Mounce et al., (2010). However, it is critical to point out that only pressure data was used to perform the burst/leak detection offline.

The pipe burst/leak detection methodologies developed in Mounce (2002, 2006) papers and Mounce et al. (2006, 2007, 2010, and 2011) are effective for pipe burst/leak detection at DMA level in near-real time. The main advantage of their method is the AI-technique (SVR or ANN) relies on time series observed flow/pressure data which is regarded as data driven method. The flow/pressure prediction model is capable of learning flow/pressure profiles in 12/24 hours window. The prediction model may not be adaptable to change the predicted values when there is a change within the DMA section in real-time. However, the authors highlight the challenges to detect pipe

bursts/leaks with pressure data which showed lower success rate of detecting pipe bursts/leaks compared to the flow data only.

Aksela et al. (2009) developed a self-organising Map ANN (Kohonen, 1990) for leak detection. The ANN is trained by using flow data containing the knowledge about reported pipe bursts/leaks. The authors used the daily flow average due to the great variance in behaviour between different days of the week. The results show the leaks in a DMA are detected successfully but the leak assessment frequency is too low to minimise water wastage due to leakages. Hence, the method is unfitting for real-time applications coupled with the heavy reliance on quality training data with little noise (i.e., sufficient record of leaks).

Ye and Fenner (2011) presented a novel burst detection method of using Adaptive Kalman Filtering for an automatic burst/leak detection in WDS based on flow and pressure observations. The Adaptive Kalman Filter modelled the normal flow and pressure data and then calculate the difference between the predicted and observed flow/pressure data. These differences are used as an indication of the abnormal flow/pressure variations relating to the bursts/leaks. The method is applied on a several real DMAs with historical data containing engineering events. The results show the burst events identified by customer contacts are detected successfully. The magnitudes of the burst correspond to the flow residual between the corrected flow and observed flow.

The studies in Ye and Fenner (2011) including Mounce et al. (2006, 2007, 2010 and 2011) show that flow observations are more sensitive to pipe bursts/leaks than pressure observations. Hence, lower success rate of bursts/leaks detection based on pressure data only in their research works. They reported that alarm generated due to pressure observations correspond well to the bursts/leaks or engineered events. The possible reasons why the pressure data are insensitive to small-medium bursts/leaks are as follows:

- The service reservoir can support the pressure in the nearest/immediate downstream DMAs even if a burst/leak has occurred;

- The location of pressure meters especially pressure meter at the highest point of the DMA may be insensitive to small-medium sized bursts/leaks;
- The pressure value at the interval is the average value of previous pressure values between the intervals hence, losing potentially useful information in the data;
- Pressure observations tend to be very noisy;
- The relationship between bursts/leaks location and pressure meters deem complex and not fully understand (i.e., pressure meters close to the bursts/leaks may not be sensitive enough).

Romano et al. (2011, 2012) developed a new methodology that makes use of ANNs to forecast pressure and flow in short term, SPC techniques for short and long-term analysis of the pipe burst based on pressure and flow differences, and Bayesian Inference Systems (BISs) for inferring the probability of a pipe burst and raising corresponding detection alarms. The authors used a wavelet analysis to de-noise the flow and pressure observations before training ANN and BIS. This is to alleviate the problem relating the noisiness of the flow/pressure. The burst methodology is applied on a real flow and pressure data based on engineered events in a rural DMA. The results illustrate that it can successfully identify these events in a fast and reliable manner with a low false alarm rate online. Romano et al. (2012, 2013) presented a methodology based on ordinary kriging to locate the approximate location of the bursts/leaks or engineered events. This method is tested in a rural DMA and its result look promising. However, the developed event detection and localisation methodology is unlikely to work effectively when there is an operational change (i.e., valve closure) with a complex WDS section.

Palau (2012) presented a multivariate statistical technique, called Principal-Component Analysis (PCA) to monitor and control of water inflows into DMAs of urban networks. The PCA technique makes a quick sensitive analysis of the inflows into a DMA without utilising a hydraulic model. The technique defines the control charts for T^2 Hotelling and distance to model that help a leakage technician to identify any anomalous behaviours regarding water use, bursts, or illegal connections of the WDS. The PCA technique simplifies the original set of flow rate data and syntheses the most

significant information into a statistical model that is able to explain most of the behaviour of the WDS. The techniques were applied to a 7 months DMA inflow data and a burst of approximately 5% of the average flow could be detected with a probability between 30% and 95%, depending on the hour of occurrence. Despite its potential to detect burst, the technique hasn't been applied in a real life WDS section or model hence its reliability of the method and detection time are yet to be established.

Arsene et al. (2012) used a combination of fuzzy logic, an ANN technique and a Graph Theory for diagnosis of bursts/leaks and other operational faults in WDS. The two ANNs are trained, the first ANN is trained on Least-Squared loop flows state estimates and the second ANN is trained on Least-Squared loop nodal head estimates. This bursts/leak detection method uses the patterns of state estimates with confidence limits to detect bursts/leaks within the WDS. It is found that there is a high misclassification rate of detecting bursts/leaks with the second ANN. This is because the trained ANN is affected by topological error and operation time periods. The first ANN performs better compared to the second ANN and it has not been applied in a complex WDS section in real-time to assess its validity and the computation time.

Ye and Fenner (2013) developed a model that uses polynomial function relating to historical flow data based on the weighted least squares method for detecting burst unsupervised within the WDS. This approach makes use of Expectation-Maximization (EM) algorithm that uses any historical flow observations to output the normal water flow (non-burst condition). The flow observations from field are then compared to the normal water flow estimated by the EM to detect burst. The burst magnitude is estimated from the difference between the measured and estimated flow. This method is applied to both real and engineered data offline and it detects pipe bursts successfully.

Tao et al. (2013) proposed a burst detection method based on Artificial Immune System (AIS). The AIS is trained on data under normal conditions via clonal immune algorithm. The burst is detected by using the Euclidean distance between the observed data to the predefined data from AIS. The technique is applied offline on a simplified

DMA with synthetic data. It is found that the technique has a good potential of detecting bursts but the application of the technique in a WDS section is yet to be established.

Bakker et al. (2013) presented a burst detection method that utilise the Cumulative Sum (CUSUM) (Misiunas, et al., 2005) and heuristic methods. The method uses the combination of demand forecasting module to forecast 48 hours demand with 15 minutes time steps and CUSUM (Misiunas, 2004) for identification of WDS anomalies (i.e., pipe bursts/leaks). The method was applied on real-time hydraulic model and data from the western part of Netherlands water supply. The results showed all the burst flow that exceeds 20-25% of the average daily flow was detected offline. Therefore, the method is not feasible to detect small to medium sized burst/leak in near real-time.

Ishido and Takahashi (2014) developed a new method for real-time burst/leak detection in WDS using real-time pressure data only. The method uses head loss ratio, a ratio of headloss based on expected pressure to headloss based on observed pressure, a burst/leak indication in a WDS section. This method is applied in a real-life Yokohama WDS section in Japan with its hydraulic data. The results look promising but it is limited to one case study. The reliability of the detection method in various burst scenarios at different time of the day are not established.

Jung et al. (2015) compared three univariate and multivariate Statistical Process Control (SPC) methods with respect to their burst detection effectiveness and efficiency. The three-univariate statistical process control methods used in the paper are: (1) the Western Electric Company rules, (2) the CUSUM method and (3) Exponentially Weighted Moving Average (EWMA). The three multivariate SPC methods are: (1) Hotelling T^2 method, (2) multivariate versions of CUSUM and (3) EWMA. The synthetic data generated via hydraulic model were used to compare the six SPC methods capability. Among the six developed detection methods, the univariate EWMA method was found to be the most effective technique to detect bursts. The authors found difficult to utilise the three univariate SPC methods to detect small-sized burst (below 20% of average daily flow). Therefore, the feasibility of the univariate EWMA method to detect small-sized bursts in a complex WDS section in near real-time is compromised.

Hutton and Kapelan (2015) presented an application of a probabilistic demand forecasting approach to identify pipe bursts. The method produces a probabilistic forecast of future demand under normal conditions. This, in turn, quantifies the probability that a future observation is abnormal. The method, when tested using synthetic bursts applied to a demand time-series for a UK WDS, performed well in detecting bursts, particularly those greater than 5% of mean daily flow at night time. The proposed method is not tested on real-life flow/pressure data (including engineered events) hence the validity of the detection method remains unknown.

Lee et al. (2016) proposed a novel methodology that uses the CUSUM method and a Wavelet Transform (WT) to detect and locate bursts in water pipe networks. The proposed network node matrix represents the candidate locations of bursts for each installed flow/pressure meters. The developed burst detection and location system is validated with real field data obtained from simulated bursts by opening hydrant valves for both simple and complex pipe networks. The result shows the developed algorithm work well. The authors believe the developed algorithm shows a better result compared to those applied to real water-supply systems up to the present (2016). However, the paper didn't reveal its capability of the proposed method in term of detecting or locating pipe bursts at different time periods and burst magnitudes.

In summary, this section reviewed the developed data-based techniques in recent years. The data-driven techniques usually use flow/pressure data and employ SA-based techniques to analyses the flow/pressure signal values for online bursts/leaks detection. These techniques can deal with large amount of noisy and raw data from the hydraulic meters and extract the meaningful information for leakage technicians and operational engineers. However, the data-driven techniques can become redundant when there is an operational change in a WDS section (i.e., valve closure). This limitation can make the data-driven techniques redundant in near real-time. Most of the data-driven detection methods have not been for different burst scenarios at different time of the day. The detection time to raise burst alarm are rarely discussed or mentioned in the literature which indicate most of the data-driven techniques were developed to detect pipe bursts/leaks offline.

The literature on data-driven techniques provided a compelling argument to use both flow meter and pressure meter within a DMA. The information from the flow meter and pressure meter provide further insight of the bursts/leaks occurrence and improve the success rate of pipe burst/leak detections in a reliable and timely manner. It has to be further noted that the observation from the flow meters and pressure meters are not the only source of information for bursts/leaks detection in real-time. A hydraulic model of the DMA or WDS section have been utilised as well to detect bursts/leaks and locate the possible location of bursts/leaks. Hence, the next section looks at the advancement of leak detection via the combination of flow/pressure observations with a hydraulic model.

2.1.3 Hydraulic Model-based Techniques

This section reviews the hydraulic model-based techniques that make use of a hydraulic model of the WDS section and hydraulic data for automated abnormal flow detection. In the UK, each WC divides their WDS section into hundreds of DMAs to assess the abnormal flow especially bursts/leaks associated with the individual DMA. The UK WCs usually install flow meters at the inlet and outlet of a DMA while install a pressure meter at either the highest elevation or critical location of the DMA. The installation of flow/pressure meters are getting cheaper coupled with lower data transmission cost rand faster computing processing unit. This has provided an opportunity for the UK WCs to collect more flow/pressure data in real-time at the low sampling frequencies ranging from every 10 minutes to 24 hours. These flow/pressure data from the WDS section are transmitted to their SCADA system. The real-time collation of flow/pressure data present an opportunity to improve detection of large bursts/leaks in a DMA. It is pertinent to mention that the accuracy and reliably of the flow/pressure meters have improved over the last decade has also improved.

The detection of abnormal flow with hydraulic model are mostly based on either residual analysis between the WDS hydraulic model predictions and the WDS observations (i.e. Bargiela, 1984; Ellul, 1989; Pudar and Liggett, 1992; Powell, 1992; Carpentier and Cohen, 1993; Jung and Lansey, 2013) or evaluation of the pattern of

WDS state estimates (i.e. Gabrys and Bargiela, 1999 and 2000; Caputo and Pelagagge, 2002 and 2003; Izquierdo et al. 2007; Arsene et al. 2012).

Several studies can be found in the literature that propose hydraulic model-based techniques which successfully detect abnormal flow in a real-life WDS section. Many researchers evaluate the current state of the WDS section using a combination of real-time hydraulic data, AI/statistical-based techniques and hydraulic model. The WDS hydraulic model consists of nodes (i.e. reservoir, tank and junction) connected by link (i.e. pipe, pump, valve) which solves the mass and energy conservation equations. Due to limited number of observations, the equations do not fully consider all the WDS variables. Therefore, the state estimation techniques (Piotrowski, 1978; Rao et al. 1974; Sterling and Bargiela, 1984) are widely used to ensure that the WDS hydraulic model predictions closely match the WDS observations. Such techniques process the mass and energy conservation equations and the additional equations relating to the minimisation of errors with limited and/or inaccurate flow/pressure data. These state estimation techniques require high computational overhead due to large amount of equations needed to be solved. However, an optimisation method based on numerically stable factorisation with parallel and distributed computing structure is used to reduce the computational time (Bargiela, 1984; Bargiela and Hainsworth, 1989; Hartley and Bargiela, 1995; Hosseinzaman, 1995). Gabrys and Bargiela (1995) developed a refined ANN to solve the conservation equations of WDS which provide better results at lower computational time due to the parallel and distributed computing structure of the developed ANN.

Ellul (1989) used the residual data between the observed pressure and flow rate and the predicted pressure and flow rate data from WDS hydraulic model. The amount of pipe burst/leak used is proportional to the obtained residual flow. The method is only applied on a simple single-branch of WDS for burst/leak detection. Therefore, its result is not sufficient to validate the capability of burst/leak detection method in a complex DMA with noisy observation data

Pudar and Liggett (1992) used a non-linear derivative-based optimisation method to detect burst/leak with a hydraulic model and “artificial” pressure and flow

measurements. The optimisation method is Levenberg-Marquardt-based algorithm (Levenberg, 1944; Marquardt, 1963). The sum of squared differences between predicted and “artificial” pressure measurements is used as the objective function. Bursts/leaks at nodes were expressed in terms of pressure by an orifice formula. The method is applied on a small pipeline (11 nodes and 7 pipes). The results look promising but the proposed methods have not been applied to real-life case studies yet. This was because the measurement technologies and hydraulic model calibration process used in 1992 were not as advanced as the current technologies and calibration process (year 2016). However, the author concludes that accurate hydraulic model parameters and high number of measured data is required to improve the detection rate and leak magnitude. Therefore, the number of flow/pressure meters and the degree of model calibration required for a WDS section are yet to be established.

Liggett and Chen (1994) used a Transients-based technique which involves studying the nature of unsteady fluid flow caused by rapid flow/pressure fluctuations in the WDS. The rapid flow/pressure fluctuations in WDS can be caused by pipe burst/leaks, valve closures, pump failures, pump start-up and shut-down. The sudden increase/decrease of flow/pressure are used as burst/leak indicator to raise a pipe burst/leak alarm. The result shows successful detection of pipe bursts/leaks offline. However, this technique can’t be applied in real-time because it requires a high frequency of flow/pressure data collection (i.e., every 5 minutes).

A variety of transient-based techniques have been developed to detect abnormalities in pipelines since 1994. Colombo et al. (2009) and Puust et al. (2010) reviewed transients-based techniques and they can be categorised into the 3 groups below:

(1) Leak Reflection Method aims to detect the presence of the leak via data analysis of the transient wave travelling along the pipe reflect at the leak (transient trace). It is regarded as the simple application of transient-based analysis for leakage detection in WDS. The uses of leak reflection method can be found in recent papers such as Ferrante and Brunone (2003a and 2003b); Beck et al. (2005); Lee et al. (2007) and Yang et al. (2013).

(2) Inverse Transient Analysis (ITA) technique makes use of inverse calibration of hydraulic model of water system to the known/measured transient data. Such

techniques involve using pressure measurements to calibrate hydraulic model parameters (i.e., pipes' frictional factor) and determine the size/location of the leak. Authors like Kapelan et al. (2003); Karney et al. (2008); Covas and Ramos (2010); Shamloo and Haghighi (2011) and Stephens et al. (2013) used ITA techniques to solve their respective hydraulic problems.

(3) The Frequency Response Method analyses the transient response in the frequency domain (transformed from time-domain transient response data) for leak identification and localisation (including estimating burst size). Fourier transforms are commonly used to transform time-domain data into the frequency domain. This technique is mostly applied in distribution pipeline systems with few leaks (Mespha et al., 2001; Ferrante and Brunone 2003a; Lee et al., 2006; Duan et al., 2012)

The transient analysis-based techniques require greater understanding of DMA behaviour, unsteady friction, pipe roughness and WDS configurations. The hydraulic model of the studied WDS section is required to be highly calibrated to the highest degree before transient analysis-based techniques can be applied for leakage identification. In addition, most of the transient analysis-based techniques have been tested on field/laboratory pipelines under a control environment. Such transient analysis-based technique is not in position to be used as WDS leaks/bursts detection tool in near real-time.

Gabrys and Bargiela (1999, 2000) presented a Neuro-Fuzzy method to evaluate the WDS state estimates' patterns for bursts/leaks detection and identification. This approach is successfully applied on a water small distributed pipeline with a high number of system observations. In a real-life WDS section, it is unusual to have a high number of observations. Hence, Andersen and Powell (2000) proposed a standard Weight Least Squares (WLS) implicit state-estimation for a WDS section with a lower number of system observations. This method (based on loop equations) is applied on a simple grid water network without considering any associated uncertainties. Since, the hydraulic model uncertainties and WDS observations uncertainties are not taken into consideration. Hence, the Neuro-Fuzzy method is unlikely to be implemented in a real-life WDS section. This is because there is always error within the real-life WDS

observations. Thus, the results can be inaccurate which consequently raise numerous abnormal flow (i.e., pipe burst/leak) alarms.

Caputo and Pelagagge (2002) presented a two-level ANN system where the first-level ANN determines the branch where the bursts/leaks occurs and the second-level ANN estimates bursts/leaks amount and location. The ANN architecture was applied on a simplified WDS section and on a real district heating system (Caputo & Pelagagge, 2003) which showed promising results. The method can identify the location of the bursts/leaks. However, the authors did not consider the uncertainties of WDS observations are not taken into consideration which makes the method unfeasible for online hydraulic modelling of WDS.

Bargiela et al. (2002) integrated Confidence Limit Analysis (CLA) with the Neuro-Fuzzy method developed by Gabrys and Bargiela (2000). The method is applied on a real-life WDS hydraulic model with synthetic data to detect different size of bursts/leaks. The CLA provides the probable range of the WDS state estimates which are compared to the system observations. The Neuro-Fuzzy method integrated with the CLA demonstrated a robust performance with minimum number of misclassification for medium/large bursts/leaks. Furthermore, the application of such method is not reliably established since only synthetic hydraulic data was used.

Poulakis et al. (2003) developed a Bayesian-based methodology to detect anomalies (i.e., pipe bursts/leaks) in flow measurements and deals with the uncertainties in WDS observation and modelling error at the same time. The hydraulic model's pipe roughness, water demands at nodes and artificial flow measurements were perturbed before the method was applied. The results showed that the method is sensitive to the perturbations in the model's parameters and measurements. The burst detection method can deteriorate if a perturbation is more than 5% is added to the WDS observations.

Misiunas et al. (2004, 2005, 2006) used a negative pressure wave-based technique to detect leaks in laboratory pipelines. The technique involved the usage of transient wave arrival time at the measurement station(s) and the knowledge of the wave speed

to detect and locate the leak with hydraulic model. The difference between the flow rate before and after the leak was used to estimate the burst flow.

The negative pressure wave-based technique has been applied to a small real life WDS section and it performed reasonably well to detect bursts in short time (Misiunas et al., 2005 and 2006; Srirangarajan et al., 2010 and 2012). However, these studies show that the performance of the negative pressure wave-based techniques can be affected by (1) changes in demand and the transient waves induced by bursts/leaks or pump start-ups, valve closures can't be easily differentiated; (2) bad measurement due to background noise; (3) complex WDS configuration and (4) location of pressure meters. The implementation of the negative pressure wave-based techniques in real-time can be too complicated due to the complex WDS configurations which affects and weakens the bursts/leaks-induced transient wave. Hence, the existing negative pressure wave-based techniques are not considered for detection of burst/leak in a WDS section.

Shinozuka et al. (2005) presented a method that makes use of ANN and hydraulic model to assess the location and extent of the damage of a burst/leak caused by earthquake. The ANN was trained on a collection of generated flow and pressure data from the hydraulic model. The trained ANN use the Euclidean distance (Deza & Deza, 2009) between a suspected burst/leak to the monitoring stations and then use an indicator for damage assessment. The method was applied on a simple synthetic WDS section which has 1 damaged location and 3 monitoring stations. The WDS section studied was shaped as a rectangular grid with 2 different pipe lengths. The results show the method was sufficient effectively for the studied WDS section. The proposed detection method is not feasible for real-life modelling of WDS section because vast majority of the WDS section are arranged in a rectangular grid.

Puust et al. (2006) proposed a method that uses the Shuffled Complex Evolution Metropolis (Vrugt, et al., 2003) optimisation algorithm to locate pipe burst/leak area. This algorithm estimates the posterior probability density functions of the burst/leak areas. It was successfully applied on two artificial case studies (the first case study includes perturbed measurements while other case study has perfect observations).

The probabilistic methodology provides a final discrete probability value for the burst/leak size which requires high computational time.

Wu and Sage (2006, 2007) proposed a method that utilise Genetic Algorithm (GA) optimisation technique (Holland, 1975) to detect leaks in WDS. The difference between the total leakage rate and the total background leakage rate was treated as the total burst leakage rate. The total leakage rate is the difference between the total DMA demand (derived from inflow and outflow meter) and the estimated DMA consumption. The total leakage rate is distributed across the DMA nodes as a fraction of the total numbers of properties. The background leakage rate for each node is estimated through an empirical formula, captured as the “Internal Condition Factor” which is based on the age of the pipe (i.e., the older a pipe, the greater the background leakage rate). The difference between the total leakage rate and total background leakage rate is known as total burst leakage rate. Then the GA optimisation technique was applying the total burst leakage rate across the hydraulic model’s nodes. The solution fitness was evaluated by comparing the field observed pressures and the simulated pressures. The methods were tested in a complex real-life DMA with engineered events and real field data. The results showed good indications of the leakage “hotspots” (when checked against historical data on leak repairs) in few DMAs. The leakage hotspots are potential location of pipe leaks. This method was incorporated into the WaterGEMS software (Bentley Systems) for offline leak detections.

Izquierdo et al. (2007) also used a Neuro-Fuzzy approach for diagnosing leaks and other faults and anomalies in WDS. The method makes use of the WDS hydraulic model to generate the estimated states with error bounds (fuzzy estimated states) to train the ANN for detecting abnormal flows with WDS field observations and demand predictions. It is found that the method works reasonably well in a small WDS section using synthetic hydraulic data. The disadvantage of this method is a large number of hydraulic meters may be required to provide more information concerning the WDS section and its application in a live environment is unknown.

Mashford et al. (2009, 2012) used Support Vector Machines (SVMs) to analyse a collection of pressure data to obtain information concerning the location and size of

leaks in the WDS. The SVM is trained on perfect pressure data which are generated from EPANET model. The results show that the location and size of leaks are predicted with good degree of accuracy. However, the EPANET model and observation uncertainties are not included in the SVM hence the algorithm is unlikely to be implemented for practical purposes.

Bicik (2010) presented the use of evidential reasoning to estimate the likely location of a burst pipe within a WDS by combining outputs of several models. A novel Dempster-Shafer model is developed, which fuses evidence provided by a pipe burst prediction model, a customer contact model and a hydraulic model to increase confidence in correctly locating a burst pipe. The methods work only well on a number of real life and semi-real case studies.

Skworcow and Ulanicki (2011) proposed an e-FAVOR approach (Borovik, et al., 2009), to detect and locate pipe bursts in a DMA. The method involves carrying out an extended fixed and variable orifice (e-FAVOR) test where the DMA inlet pressure is being stepped up and down, while recording inlet flow, inlet pressure and pressure at selected locations inside the DMA using loggers. The results of the e-FAVOR test are used together with a hydraulic model of the DMA as the inputs to a software tool, which performs series of simulations and facilitates data analysis. The methodology was tested in practice in a manual manner and proved to be effective, but it is time consuming. The reliability of the approach to detect at different period of the day is not yet established.

Arsene et al. (2012) presented a method that analysis the pattern of hydraulic data using the combination of ANN, fuzzy logic and graph theory for the detection of pipe leaks. The method is trained on generated data from the numerical model of WDS based on a Least Squares (LS) loop flows state estimator and a Confidence Limit Analysis (CLA) algorithm for uncertainty quantification. It is found that the method performs well when the method is trained on nodal demand data compared to the combination nodal heads and pipe flow rates data. The major issue with the method is the way the WDS model has to be spanned out as a tree /co-tree model which can be

either labour-intensive or computational expensive to span out a larger WDS section model.

Jung and Lansey (2013) applied the Extended Kalman filter method for burst detection when there is a boundary change within WDS. The detection metric used is standardised Kalman innovation sequence (Mehra & Peschon, 1970). Standardised Kalman innovation is the product of Kalman Gain (refer to equation (4)) and the flow residual between the observed and estimated pipe flow. It is examined to identify a pipe burst. Any standardised Kalman innovation sequence that falls outside the pre-defined threshold is seen as a pipe burst hence, an alarm is raised. The method is applied on a real WDS model with synthetic data generated from the WDS hydraulic model. The flow observations are taken at the interval of every 5 minutes which may not be practical for a large water system due to the computational power and memory. The reliability of the method is unknown since it has not been applied on a real-life data and other burst scenarios (e.g., different burst magnitudes at different time periods, locations and noise levels).

Kang and Lansey (2014) proposed a novel approach for detecting pipe bursts with a WDS specific burst sensitivity tables. The WDS burst sensitivity table is developed by using artificial burst events and analysing the WDS hydraulic responses to the given bursts. This burst detection approach is applied to a simple WDS hydraulic model which has 2 reservoirs, 2 tanks, 2 pumps, 9 nodes and 18 pipes. The authors selected four pressure logger locations and six flow meter locations based on the WDS sensitivity table to monitor the simple network. The results show that the proposed approach is effective to quickly locate bursts and reduce response times. However, the uncertainty of the pressure and flow are not taken into consideration. The application of the method to a real WDS model and real-life flow and pressure observations are yet to be established.

Anjana et al. (2015) proposed the use of a Particle Filter (PF) based technique for the detection of leaks in water pipelines. The developed PF-based detection model use the standard SPC-CUSUM chart was able to detect anomalies in the system. This technique was applied to a real-world network in Mandya (Karnataka, India). It successfully detects a pipe leak in the real network (trunk-main model). Hence the

performance of the technique in a complex DMA in real-time is unknown. This method shows some potential for online detection burst.

Sousa et al. (2015) applied both simulated annealing and graph theory to water distribution networks to locate leaks. This methodology is based on pressure measurements and explores the exchange of information between an optimization model and the hydraulic simulation of the WDS in steady state conditions. The cases considered a single leak and two simultaneous leaks. The result shows the identified pipes are close to the actual leaky pipe. However, the time of leak detection and localisation is unknown. It is difficult to assess the performance of the developed method in different burst scenarios at different time period.

In summary, the hydraulic model-based techniques for pipe bursts/leaks identification are performed by comparing the predicted flow/pressure to the corresponding WDS observed flow/pressure. Most of the hydraulic model-based techniques are tested on a real-life/artificial WDS section for mostly offline pipe burst detection. Many of the hydraulic model-based techniques in the literature assumed the observations are ideal noise-free. The major benefit of hydraulic model -based techniques compare to the data-driven techniques is they can be used to find the optimal location of the hydraulic meters and approximate area of the bursts/leaks. To use the hydraulic model-based techniques effectively for bursts/leaks detection, it is well established that an up-to-date hydraulic model is required. However, it is not necessary to obtain a very accurate hydraulic model because the hydraulic model can never match the reality of the WDS. Most of the hydraulic model-based techniques can detect bursts/leaks and estimate its approximate location in their respective case study. Information relating to the developed burst method's performance in different burst scenarios/time periods or detection time/time taken to approximate bursts/leaks locations are rarely provided in the literature.

The Geography Information System (GIS) of the WDS have improved over the recent years. GIS is even operating live in UK WCs' control room. The information from the GIS is often used to update the hydraulic model parameters especially when there is an operational change (e.g., pump failure or valve closure). Most of the hydraulic

model-based techniques do not consider the effect of the operational change in a WDS section. Therefore, a development of an online hydraulic model-based technique for bursts/leaks detection and localisation methods even in the event of WDS operational change is required. A hydraulic model that reflect the current status of the WDS in near real-time can provide further insight into the affected WDS section when there is a pipe burst/leak.

Several researchers (i.e., Farley et al. 2008 and 2010; Skworcow and Ulanicki, 2011, Kang and Lansey, 2014) have developed a technique to find the optimal location of pressure meters in a DMA. Some of the techniques have been used by UK WCs. Hence it is beneficial to use the suggested optimal location of hydraulic meters from their developed techniques. The meters can assist in detecting bursts/leaks and/or finding the approximate bursts/leaks location reliably. None of the papers in literature have compared the capability of different combinations of hydraulic meters for WDS burst detection.

However, it is also pertinent to mention that the UK WCs are increasing the number of hydraulic meters (particularly pressure sensors) in a DMA. Unfortunately, hydraulic meters' data from the field can be noisy can affect the performances of the developed burst detection/location techniques. Therefore, the next section review data assimilation methods that account for hydraulic model predictions and WDS observation uncertainties to improve hydraulic model state predictions.

2.2.4 Comparison of Burst Detection and Location Techniques

Section 2.2.2 and 2.2.3 described and reviewed several techniques for online bursts/leaks detection and/or location. The techniques analysed have different capabilities, benefits and limitations. The hardware-based techniques for detecting and locating bursts/leaks in WDS were not reviewed given the focus of this thesis on numerical, i.e. software type techniques. These techniques require less man-power to detect/suspect a burst/leak in WDS compared to the hardware-based techniques. The key factors to investigate the feasibility of the reviewed hydraulic techniques for detecting and locating burst/leak in real-time are:

- Ability to detect and locate pipe bursts/leaks in a reliable and timely fashion;
- Ability to process large volumes of raw data (noisy data) from hydraulic meters;
- Ability to be used in changing WDS configurations.

Table 2.1 summarises the main characteristics of the bursts/leaks detection and location techniques that have been reviewed in this section. Table 2.2 summarises their individual advantages and disadvantages.

Table 2.1 and Table 2.2 reveal the potential of the hydraulic model-based techniques have not been fulfilled relating to the bursts/leaks detection and location accuracy in real-time in different burst scenarios. Only the negative pressure wave-based techniques (and possibly transient analysis-based) may have the potential to locate the location of pipe bursts/leaks in a WDS section. However, both techniques require new substation to host sophisticated sensors (acoustic-based hardware), highly calibrated hydraulic model, high frequency of data and higher number of hydraulic meters due to the sensor-to-sensor spacing. These techniques have been successful in a simple distribution pipeline system or WDS section under controlled environment. They are not in position to be implemented in real-time with noisy data from flow meters and pressure meters. However, the reviewed hydraulic data/model-based techniques highlight the usefulness of a hydraulic model in detecting and locating pipe bursts/leaks.

Table 2.1: The Main Characteristics of the Reviewed Hydraulic Techniques for Pipe Bursts/Leaks Detection/Localisation

Technique	**Technique's functionality: bursts/leaks detection/localisation**	**Computational time to detect bursts/leaks**	**Technique's Application: Water network or pipelines**
Data-driven techniques	● Most of the data driven techniques in the literature are developed to detect pipe burst/leaks. ● One data driven technique (Romano et al., 2011, 2012) is developed to locate pipe bursts/leaks in real-time.	Low (every 10 minutes) – High (every 24 hours)	WDS section (i.e., DMA). Pipes' diameter varies between 25mm and 600mm. System Pipelines (i.e., Medium/Large Diameter Trunk Mains). Pipes' diameter varies between 600mm and 1500mm.
Hydraulic model-based techniques	● Majority of the hydraulic model-based techniques in the literature aim to detect pipe bursts/leaks offline. ● Few papers focused on locating pipe bursts/leaks offline. ● Limited number of papers developed model-based technique to detect pipe bursts/leaks in real-time.	Low (every 10 minutes) – High (every 24 hours)	WDS section (i.e., DMA). Pipes' diameter varies between 25mm and 600mm. System Pipelines (i.e., Medium/Large Diameter Trunk Mains). Pipes' diameter varies between 600mm and 1500mm.

Table 2.2: The Main Advantages and Disadvantages of the Reviewed Burst/Leak Detection and Location Techniques

Burst Detection / Localisation Techniques	Advantages	Disadvantages
Data-driven techniques	● Data-driven detection/localisation techniques do not require high frequency flow/pressure measurements. ● Data-driven detection/localisation techniques can perform online monitoring of a WDS section effectively. ● Data-driven detection/localisation techniques work well in DMAs despite the complexity of the DMA because they do not consider DMAs' configuration. ● They also provide an efficient and consistent means for the analysis of large volumes of imperfect data.	● Data-driven detection/localisation techniques can't provide further information on the approximated location of a pipe burst/leak within a DMA. ● Data-driven detection/localisation techniques can't adapt to the changes in the WDS section's operating conditions.
Hydraulic model-based techniques	● Hydraulic model-based techniques do not require high frequency measurements. ● They can perform online monitoring of a WDS section effectively. ● Hydraulic model-based techniques can reduce the size of the approximated location of a pipe burst/leak within a DMA. ● Hydraulic model-based techniques care capable of adapting to the changes in the WDS section's operating conditions. ● Hydraulic model-based techniques can detect and locate leaks in pipeline under controlled environment. ● Hydraulic model-based techniques can reduce the size of the approximated location of a pipe burst/leak within a DMA.	● Hydraulic model-based techniques rely heavily on calibrated hydraulic mode for successful detection and localisation of pipe bursts/leaks. ● Hydraulic model-based techniques require additional AI-Statistical techniques to detect or locate bursts/leaks. They are applicable to a small/medium sized WDS (i.e., DMAs). ● Hydraulic model-based techniques also require accurate measurements at multiple locations. ● Hydraulic model-based techniques can be affected by noisy WDS operations.

The pipe burst/leak detection and location techniques are reviewed using the following key points:

(1) Ability to raise a timely and reliable alarm when there is a fluctuation in flow/pressure data (due to bursts/leaks) in real-time;

(2) Ability to macro/micro locate bursts/leaks in near real-time, i.e. shortly after the burst/leak is detected;

Among the reviewed techniques, the hydraulic model-based techniques make use of hydraulic model and large volume of raw data (noisy data) to detect pipe bursts/leaks in real-time meet the aforementioned criteria. The data-driven techniques do not use the hydraulic model hence offer an efficient and low time-consuming means for the automated on-line pipe bursts/leaks identification in real-life WDS sections. However, these techniques can become redundant when there is an operational change in a WDS section. Also, the hydraulic model-based techniques have the potential to provide additional information about the identified pipe bursts/leaks including the identification of a likely macro/micro burst location and its impact on the water service, all in near real-time.

All of the above speaks in favour of using methods that can combine hydraulic simulation models with online hydraulic meters data for determining the analysed WDS state which, in turn, should enable more accurate pipe burst/leak detection and location. These, data assimilation type methods are reviewed in the next section.

2.2 DATA ASSIMILATION METHODOLOGIES

2.2.1 Introduction

Data Assimilation (DA) methods are used in many scientific disciplines, including meteorology, hydrology and WDS modelling (Evensen 2003; Shang et al., 2006; van Leeuwen 2009; Matgen et al. 2010; Jung and Lansey, 2013). The aim of the review presented here is to highlight the main capabilities and limitations of reviewed DA methods in the context of their potential use for pipe burst/leak detection and location in WDS.

The DA method aims to estimate the hydraulic state(s) of the analysed WDS at a given point in time using observations and state forecasts. It works in a two-step procedure with the steps being the prediction and the correction step. In the prediction step, the DA method predict state estimate (i.e., flow/pressure at a given location) along with their uncertainties. In the correction step, the DA method combines the predicted values with the corresponding WDS observations (along with their uncertainties) to correct the state predictions at the same locations (Bargiela and Hainsworth 1989). The state predictions can include both hydraulic model parameters (i.e., demand coefficients, pipe roughness, and pump/valve status) and/or actual WDS states (boundary inflow/outflow, tank level).

The prediction steps of a DA method for the WDS can be expressed as follows:

$$\boldsymbol{x}_t^f = \boldsymbol{F}\boldsymbol{x}_{t-1}^c + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N[0, \boldsymbol{Q}_t] \tag{1}$$

$$\boldsymbol{y}_t^p = \boldsymbol{H}\boldsymbol{x}_t^f + \boldsymbol{e}_t \qquad \boldsymbol{e}_t \sim N[0, \boldsymbol{R}_t] \tag{2}$$

where subscript *t* is the time step index; superscript *c, f* and *p* are denoted as corrected, forecasted and predicted value respectively; $\boldsymbol{x}_t^f$ and $\boldsymbol{x}_{t-1}^c$ are the forecasted and corrected state vector; $\boldsymbol{F}$ represents the matrix structure that propagate the state vector from t-1 to t; $\boldsymbol{H}$ is the observation operator that maps the state vector to WDS states; $\boldsymbol{y}_t^p$ is the total predicted (modelled) WDS state vector; $\boldsymbol{e}_t$ is the observation error matrix and $\boldsymbol{\omega}_t$ is the state error matrix which is assumed to be zero mean multivariate Gaussian noises with state model error covariance matrix $\boldsymbol{Q}_t$.

The state vector is corrected during the correction step of a DA method when observed data becomes available. The corrected state vector can be used either as the initial conditions or parameters of the hydraulic model for the future WDS state predictions. The accuracy of the state vector forecast depends on the availability of observation data, errors propagated from the initial conditions, model structural and parameter errors.

The DA methods have been developed for real time application to correct state model predictions based on pre-determined understanding of forecast error (Hutton, et al., 2012). DA methods can be categorised into 2 groups; (1) sequential DA method and (2) variational DA method. The variational DA method uses objective function to minimise the error between the predicted WDS state estimates and the studied system observations by adjusting state estimates (i.e., initial boundary conditions or hydraulic model parameters). This method is viewed as a constraint minimisation problem and it is solved iteratively with a gradient based optimisation method. Unlike the variational DA methods, the sequential DA methods take account of the time evolution of WDS observations with the potential to estimate WDS state estimates over time. The sequential DA methodologies are deemed as 'good enough' for online operational purpose although they are not perfect because of the uncertainties (or errors) involved. This section covers a selection of sequential DA methodologies based on empirical (or statistical) approach to solve WDS state estimation problems. However, both Barnes analysis scheme (Barnes, 1964) and Cressman objective analysis scheme (Cressman, 1959) are not reviewed due to their inability to deal with large set of diverse observations and also Cressman objective analysis scheme discard observational error (Schlatter, 1988).

The literature review of sequential DA methods in this section is primarily focused on the correction step. This section also considered the capability and limitation including implementation challenges and computational issues related to each reviewed DA method. Appendix B provides further theoretical details of each reviewed DA method.

The DA methods that are considered in this section include:

- Kalman Filter method
- Extended Kalman Filter method
- Unscented Kalman Filter method
- Ensemble Kalman Filter method
- Particle Filter method.

The correction step of each above-mentioned DA method is reviewed in section 2.3.2 to 2.3.6 and then summarised in section 2.3.7.

2.2.2 Kalman Filter Method

The Kalman Filter (KF) method (Kalman, 1960) aims to estimate the optimal state vector via linear stochastic process (Greg & Gary, 2001). The theoretical background of the KF can found in the literature written by (Welch & Bishop, 2006) and in Appendix B1 of this thesis. Assuming equations (3) and (4) represent a Gaussian linear system, the correction steps of the KF method in WDS context are as follows:

$$\boldsymbol{x_t^c} = \boldsymbol{x_t^f} + \boldsymbol{K_t}(\boldsymbol{z_t} - \boldsymbol{H}(\boldsymbol{y_t^p})) \quad (3)$$

where $\boldsymbol{H}$ is the observation mapping operator that relate WDS observations to the predicted hydraulic states; $\boldsymbol{K}_t$ is the Kalman gain; $\boldsymbol{z}_t$ is the WDS observations vector; $\boldsymbol{P}_t^c$ and $\boldsymbol{P}_t^p$ are the posterior and prior error covariance matrix respectively. The Kalman gain, $\boldsymbol{K}_t$ is viewed as the weight factor based on the prior and observation error covariance matrix:

$$\boldsymbol{K_t} = \boldsymbol{P_t^f H^T}\,(\boldsymbol{H P_t^f H^T} + \boldsymbol{R_t})^{-1} \quad (4)$$

where superscript T indicates the matrix is transposed and $\boldsymbol{R}_t$ is the observation error covariance matrix.

The KF method has been applied in a WDS section to estimate unknown roughness in a linear estimation problem (Todini, 1999) and water quality modelling (Schilling & Martens, 1986). Since the KF method is best suited to linear problems, Walski, et al. (2003) used it in a loop iteration to increase network resilience/guaranteed consumer demand due to nonlinear system of WDS.

Kang and Lansey (2009) compared an iterative-KF method to the Tracking State Estimator (TSE) based on weighted least-squares scheme to estimate WDS states (nodal pressures and chlorine concentrations) in real-time. The iterative-KF method is used to overcome its limitation in a nonlinear system like WDS. The iterative-KF method and TSE quantified the uncertainties in demand estimates and predicted state

variables (pressures and chlorine concentrations). The quantified uncertainties were verified by Monte Carlo simulation. The results indicate flow measurements are better to estimate demands with high level of confidence compared pressure measurements. The results also show the TSE method performs well in a looped WDS network while the iterative-KF method performs well in a series WDS network.

Jung and Lansey (2014) proposed a method that uses the KF method to detect pipe bursts when an operational change occurred. The results show that the proposed approach can be effective in burst detection with the aid of KF method. For further detailed summary of this paper, see page 51.

The main issue with the KF method is the error covariance matrices must be derived at the initial time step. The observation error covariance matrix can be estimated based on the knowledge of the instrumental error variances while prior and process error covariance can be estimated based on the prior information of the state model. Such issue can make the KF method unsuitable for highly non-linear stochastic system like WDS. Hence, the KF method is expanded via Taylor series (Bertino, et al 2003; Evensen, 2003) to accommodate the non-linearity behaviour of the WDS. The expansion of the KF method is known as the Extended Kalman Filter (EKF) method.

2.2.3 Extended Kalman Filter Method

The Extended Kalman Filter (EKF) (Jazwinski, 1970) is the advancement of the KF method (refer to Appendix B2) to accommodate the non-linear system such as WDS.

The state model, ***F*** and observation operator, ***H*** cannot be applied directly on a system with some non-linearity. This operator is approximated with tangent linear operators.

The EKF method has been applied for real time calibration of water demand (Shang et al. 2006). The EKF method involved Autoregressive Integrated Moving average (ARIMA) (Box & Jenkins, 1976) which is used to forecast the water demand coefficients of a small WDS hydraulic model (92 nodes, 2 reservoirs, 3 tanks, 2 pumps and 117 pipes). The EKF method is used to correct the predictions of water demand coefficients

with the aid of usually high number of observed flow rates and pressure heads. The authors show that the EKF method performance depends on the spatial error among the demand coefficients, forecast error, observation error and arbitrary sampling design.

The application of the EKF method is deemed to be unstable in an observed system with large nonlinearities (i.e., WDS) (Hoteit, et al., 2005). Many researchers (i.e., Terejanu, 2008; Pauwels and DeLannoy, 2009) tried to address the instability issues by using higher-order terms of Taylor series to cope with high non-linearity of the observed system (i.e., WDS). Terejanu (2008) also found the state estimates can be biased due to a large variant of observation error. Hence, the Jacobian of the hydraulic network is approximated without other higher-order terms and is limited to the first-order of Taylor series. Terejanu (2008) and Pauwels and DeLannoy (2009) show the EKF method can also neglect observation operator if the higher-order derivatives of Taylor series is used in the correction step. Evensen (2003) highlights the additional computational cost to use higher-order terms of the EKF method. Nevertheless, the EKF method is still capable of providing a good performance in term of improving hydraulic model state predictions despite its high computational cost. Due to the first order approximation of the EKF method, the method can introduce large errors in the corrected state estimates which lead to the divergence of the filter (Wan & van der Merwe, 2000). This flaw can be addressed by a DA method called Unscented Kalman Filter (UKF) in the next section.

2.2.4 Unscented Kalman Filter Method

The Unscented Kalman Filter (UKF) method (Uhlmann, 1995) was developed to address the limitations of the EKF method. The EKF method usually loses information concerning the state estimates when the EKF method tries to linearise an observed system (i.e., WDS). Therefore, the UKF method (refer to Appendix B3) uses an unscented transformation, minimal set points of sigma points (a deterministic sampling technique) to calculate the statistics of a random variable which undergoes a nonlinear transformation. These minimal set points of sigma points are devised via an empirical analysis of the state estimates (Zhang et al., 2009) which is mostly used for Gaussian distribution.

The UKF method is closely related to the EKF method but the differences between these two DA methods are: (1) their implementation/application approach, (2) computational cost and (3) the UKF method requires sigma points and an ensemble of predicted/observed WDS states to estimate the error covariance matrices. The UKF method can be applied to a non-linear system without the knowledge of the Jacobian matrix. It also has the ability to maintain the non-linearity of the system (Kreuzinger, et al., 2008). However, it is noticed that the UKF method does not improve the accuracy of state estimates but it reduces the underestimation of prior error covariance over time (Uhlmann, 1995). Hence, there is a little difference between the corrected state estimates from the UKF and EKF method.

The main disadvantage of the UKF method is the cumulative of underestimated error covariance which can cause the filter to become overconfident over time resulting in underestimating the accuracy of state estimates. The other major disadvantages of the UKF are it is more computationally expensive compared to the EKF method (Kreuzinger, et al., 2008) and it is limited to non-linear systems that have Gaussian noise. Since the corrected state estimates are based on small set of sigma points, the hydraulic state estimates are not truly global approximation of the observed system (i.e., WDS). Due to the complex implementation of the UKF method, the Ensemble Kalman Filter method is usually preferred for hydraulic modelling of the WDS.

2.2.5 Ensemble Kalman Filter Method

The Ensemble Kalman Filter (EnKF) method (Evensen, 1994) is commonly used for spatial-temporal phenomena evaluation like ocean modelling (Evensen, 2009) and weather forecasting (Myrseth, et al., 2009). The EnKF method (refer to Appendix B4) is a suboptimal estimator which is suitable for nonlinear system with a large number of state variables. It was originally developed to overcome some of the problems associated with the EKF method.

Unlike the UKF and EKF method, the EnKF method corrects the ensemble of forecast hydraulic state estimates individually without the need of covariance matrices or

integrating backward in time (Mandel, 2007; Evensen, 2009; Myrseth, et al., 2009). The principle of EnKF method is to approximate the state estimates (vector) and prior error covariance from the ensemble statistics – equations (5) and (6):

$$\boldsymbol{C_Q} = \frac{[\boldsymbol{Y_t} - \boldsymbol{\mu_t^y}]\,[\boldsymbol{Y_t} - \boldsymbol{\mu_t^y}]^{\boldsymbol{T}}}{\boldsymbol{N-1}} \quad (5)$$

$$\boldsymbol{C_R} = \frac{[\boldsymbol{Z_t} - \boldsymbol{\mu_t^z}]\,[\boldsymbol{Z_t} - \boldsymbol{\mu_t^z}]^{\boldsymbol{T}}}{\boldsymbol{N-1}} \quad (6)$$

where μ_t^y and μ_t^z are the ensemble mean of hydraulic model predictions and the corresponding WDS observations respectively; $\boldsymbol{C_Q}$ and $\boldsymbol{C_R}$ are the ensemble WDS hydraulic model predictions and the corresponding WDS observation error covariance respectively.

The ensemble statistics are the nursed to calculate the Kalman gain via $C_Q\,[C_Q + C_R]^{-1}$. The major key issue with the EnKF method is the quantification of the covariance error matrices. The EnKF method relies on sampling design to generate ensemble hydraulic state estimates, hydraulic model predictions and the corresponding WDS observations. The most common method is the perturbed observations method (Burgers, et al., 1998). The perturbed observations method (Burgers, et al., 1998) involves adding random perturbations to both state vectors and WDS observations which introduces sampling errors (Evensen, 2004). The data perturbation is performed to prevent the underestimation of analysis error covariance (Chena, et al., 2013):

The problem with the perturbed observations method is that it affects the prior knowledge between the state estimates which cause the EnKF method to diverge` (Sun, et al., 2009; Sakov and Oke, 2008; Anderson, 2001). Therefore, various deterministic ensemble filters are developed to overcome the limitations of the perturbed observations method. Most of the EnKF method variants combine the ensemble mean of corrected state estimates – equation (7) with an analysis ensemble perturbation:

$$\boldsymbol{X_t^a = \overline{X_t^c} + A_t^c} \quad (7)$$

where $\boldsymbol{X_t^c}$ is the ensemble corrected state vectors; $\boldsymbol{\overline{X_t^c}}$ is the ensemble mean of the corrected state vectors; $\boldsymbol{A_t^c}$ is the analysis ensemble perturbations matrix;

The analysis ensemble perturbations matrix is derived from transformation of predicted ensemble perturbations through a transform matrix. The common variants of transform matrix are as follows (refer to appendix B4):

1. Ensemble Square Root Filter (EnSRF) (Whitaker & Hamill, 2001);
2. Ensemble Adjustment Kalman Filler (EAKF) (Andersons, 2001);
3. Deterministic Ensemble Kalman Filter (DEnKF) (Sakov & Oke, 2008).

The main advantage of the EnKF method and its variants is the speed of correcting the hydraulic state estimates compared to other gradient-based procedures (Tureyen & Onur, 2011) due to exclusion of prior error covariance evolution. The EnKF method does not cause the filter to diverge quickly like the KF method and EKF method in a system with high non-linearity. The perturbed EnKF method is preferred method to the squared-root based the EnKF method due to the easy implementation of the perturbed EnKF method.

The disadvantages of the EnKF method and its variants are: (1) they only account for state forecasts error due to uncertain initial conditions; (2) the state estimates' error due to state model deficiencies are not considered (Tippet, et al., 2003); (3) large number of ensemble member may be required to ensure the stability of the filter hence, increasing the computational cost and time; (4) the EnKF method Kalman gain can give little weigh to the residuals when the number of hydraulic state variables is greater than ensemble number (Myrseth, et al., 2009). An alternative method in dealing with the high nonlinearity of WDS is Particle Filter (refer to section 2.3.6.).

2.2.6 Particle Filter Method

Particle Filter (PF) method is a Sequential Monte Carlo method that is capable of correcting the predicted state estimates (Doucet, et al., 2001). The main principle of the PF method is to use Bayesian formulae to correct predicted state estimates when the WDS observations become available via posterior probability distribution. The posterior probability distribution is calculated as follows:

$$\boldsymbol{P(X_t|Z_t)} = \frac{\boldsymbol{P(Z_t|X_t)\ P(X_t|Z_{t-1})}}{\boldsymbol{P(Z_t|Z_{t-1})}} \quad (8)$$

where $P(\boldsymbol{Z}_t|\boldsymbol{X}_t)$ is the probability of WDS observations given by state forecasts; $P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})$ is the prior probability distribution of state forecasts given by the prior observations and $P(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})$ is the normalisation factor which can be expressed as $\int P(\boldsymbol{Z}_t|\boldsymbol{X}_t)\ P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})\ d\boldsymbol{x}_t$.

Hutton, et al. (2010) highlighted that it is generally difficult to sample directly from the posterior probability distribution itself, hence, Sequential Importance Sampling (SIS) is used instead (Orhan, 2012). The idea of SIS is to use samples drawn from a proposal probability distribution. In other word, the PF method approximates the posterior probability distribution at previous time step (*t-1*) with the weighted set of particles. Unfortunately, there is a discrepancy between the posterior probability distributions at current time step and previous time step. To compensate for this discrepancy, the particle samples (state estimates) and associated weights are reduced to equation (9) and (10) respectively (Arulampalam, et al., 2002):

$$\boldsymbol{P(X_t|Z_t)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{w_t^i \delta}_{\boldsymbol{X_{t-1}^i}} \quad (9)$$

$$\boldsymbol{w_t^i} = \boldsymbol{w_{t-1}^i P\left(Z_t|\ X_t^i\right)} \quad (10)$$

where w_t^i is the updated particle weight; w_{t-1}^i is the particle weight at the previous time step; $(\boldsymbol{X}_t|\boldsymbol{Z}_t)$ is the posterior probability distribution at time step, t; $P\left(\boldsymbol{X}_t^j|\ \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the probabilistic state model (or transition probability distribution) and $Q\left(\boldsymbol{X}_t|\ \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the proposal probability distribution.

The main advantage of the PF method is it does not require specific state estimates distribution form (Pauwels & De Lannoy, 2009) and it can deal with high number of particle samples. Having a large set of particle samples ensures the PF method produce sub-optimal state estimation. The potential advantage of the PF method over the EnKF method is that implementation within any model structure deemed to be easier. This is because the PF method is not required to correct state estimates every time step.

The major disadvantage of the PF method is it estimates the distribution of the limited WDS observations at each time step. Therefore, the particle weights distribution can be skewed when both prior and observation variances are too high (Doucet & Adam, 2008). When the drawn particle samples are not large enough, a resampling algorithm is required (Hutton, et al., 2012; Weerts and El Serafy, 2006). The aim of re-sampling particles is to overcome the PF method degeneracy. Kitagawa (1996) resample particle weights when the samples become highly non-uniform which can neglect the small-particle weights with stratified sampling and duplicate large-particle weights instead. This forces the samplings to focus more on high end probability and becomes dependent which give less information concerning the state estimates (Ching, et al., 2006). The PF method can be re-sampled several times to make the particles completely independent which comes with high computational cost (Ching, et al., 2006).

The PF method has been applied in WDS by both Hutton et al. (2012) and Anjana et al. (2015) to solve their respective WDS problems. Hutton et al. (2012) applied the PF method to estimate the WDS states via demand coefficient correction while Anjana, et al. (2015) combined the PF method and a statistical technique, CUSUM to detect pipe bursts/leaks in a trunk main model respectively. Despite, both authors showed the potential of PF method in hydraulic modelling of WDS. The method is still not widely used to solve WDS problems and its application to account for uncertainty in WDS remains to be evaluated. Full theoretical details of PF method can be found in Appendix B5

2.2.7 Comparison of DA Methods

A number of sequential DA methodologies have been described and reviewed in the previous section. Each DA method has its own capabilities and limitations, computational costs and benefits. Few of the DA methods analysed (i.e., KF method, EKF method) have been applied in water network to estimate WDS parameters or states (i.e., demand coefficients, pipe flows and nodal pressures). Table 2.3 summaries the advantages and disadvantages of DA methodologies and their current application in hydraulic modelling of a WDS section. Table 2.4 provides the key summary of DA methods.

There is no optimal DA method for near real-life hydraulic modelling of WDS. Among the reviewed DA methods, the KF and EKF method work well in a linear system or in a system that have small non-linearity. These DA methods are shown that they work reasonably well in a small-medium WDS section (trunk main or DMA). The EnKF method is developed to work in highly non-linear system and there are 2 variants of the EnKF method. They are as follows: (1) perturbed EnKF method and (2) squared-root-based EnKF method (see Appendix B). The perturbed EnKF method is commonly used due to its simple implementation of the method compared to the squared-root-based EnKF method. However, most of the EnKF methods are still at the research and development stage. The UKF and PF method have the potential to correct state estimates in non-linear system like WDS but at the expense of the higher computational cost. The performance between DA methods in WDS hydraulic modelling remains unknown.

The evaluation of the DA methods should be based on these two points:

1) Difficulty of implementing the DA algorithm.
2) Ability to deal with highly non-linear system - WDS.

It is clear that, among the reviewed DA method, the KF, EnKF and PF method are the possible candidates to correct the hydraulic model states. The state forecasts are corrected via the residuals between the hydraulic model predictions and the corresponding WDS observations along with other uncertainties. The uncertainties are

related to the state forecasts, hydraulic model predictions and the corresponding WDS observations. A comparison study between the KF method, PF method and EnKF method is required to review capability and limitations with a real-life hydraulic model and WDS observations. Based on the results, a DA method would be chosen for burst detection/localisation in WDS.

Table 2.3: Summary of Data Assimilation Methods

Data Assimilation Methods	**Suitability - Linear/Nonlinear system**	**Model forecast**	**Representation of Errors**	**Computational cost**
Kalman Filter (Kalman, 1960)	Linear system	Deterministic model forecast	Variance of instruments to represent observation covariance; guess the initial process and forecast error covariance	Low
Extended Kalman Filter (Jazwinski, 1970)	A system with small non-linearity	Deterministic model forecast	Error covariance is derived via Taylor series.	Medium
Unscented Kalman Filter (Uhlmann, 1995)	Linear and medium non-linear system	Deterministic model forecast	error covariance approximated by *2L+1* sigma points	Medium/High
Ensemble Kalman Filter (Perturbed Observations Methods) (Evensen, 1994)	Linear and high non-linear system	Ensemble estimates	Ensemble Statistics – sample covariance matrix Or Reduced rank approximation of squared root of covariance matrix	High – very high (depending on ensemble size)
Particle Filter (Doucet, et al., 2001)	Linear and high non-linear system	Ensemble estimates	Approximated from proposal probability distribution (Sequential Importance Sampling)	High – very high (depending on number of particles)

Table 2.4: Comparison of Data Assimilation Methodologies and its Application to the WDS

Data Assimilation method	**Advantages**	**Disadvantages**	**Application to WDS**
Kalman Filter (Kalman, 1960)	● Suitable for linear stochastic system models ● Good for tracking targets (i.e. spread of chlorine residual) ● Easy to implement. ● Low computation cost	● Model operator is assumed to be unbiased ● Unsuitable for non-linear stochastic system ● Kalman innovation can be ignored if the filter gets overconfident. ● Difficult to estimate the actual error covariance ● Assume no correlation error between the forecasts and observations	Kang and Lansey (2009) applied KF to a small WDS model to estimate the demand coefficients and chlorine concentration in real-time with synesthetic data. Flow residuals are used to correct demands and chlorine concentration residual are used to correct chlorine concentration.
Extended Kalman Filter (Jazwinski, 1970)	● Linearise system that has small nonlinearity. ● Low computation cost	● Model operator is assumed to be unbiased. ● EKF relies heavily on the addition of Gaussian random variable to stabilise the	Shang, et. al (2006) used EKF to correct demand coefficients in real-time with synesthetic data.

Data Assimilation method	Advantages	Disadvantages	Application to WDS
		corrected forecast error covariance. ● Difficult to estimate the actual error covariance	
Unscented Kalman Filter (Uhlmann, 1995)	● Can maintain the nonlinearity of hydraulic system compared to the Extended Kalman Filter	● Model operator is assumed to be unbiased. ● UKF does not improve the accuracy of hydraulic states estimation. ● Cumulative of underestimated error covariance can cause filter divergence. ● Hydraulic state estimates are not truly global approximation due to small minimal sets of sample points. ● High computational cost because it resembles Monte Carlo sampling	None

Data Assimilation method	Advantages	Disadvantages	Application to WDS
Ensemble Kalman Filter (Evensen, 1994)	● Derivation of tangent linear operator is not required as in EKF. ● Easy to implement. ● Good for highly nonlinear system. ● Posterior error covariance from previous time step is not required.	● Model deficiencies are not considered. ● May require high number of ensemble member to stabilise the EnKF method. ● Kalman gain can give little weigh to Kalman innovation if the ensemble of all forecast hydraulic state estimates are close to its ensemble mean. ● High computation time due to Monte Carlo sampling (ensemble statistics).	None
Particle Filter (Doucet, et al., 2001)	● Good for tracking targets (i.e. spread of chlorine residual) ● Good for highly nonlinear system.	● Particle weights distribution can be skewed when both forecast and observation variance is very high. ● PF diverges quickly, hence, PF requires resampling to stabilise	Anjana et al. (2015) proposed the use of a Particle Filter (PF) based technique for the detection of pipe leaks in water pipelines. The developed PF-based detection model use the standard SPC-

Data Assimilation method	Advantages	Disadvantages	Application to WDS
		PF method at the expense of higher computational cost ● High computation time due to Monte Carlo sampling	CUSUM chart was able to detect anomalies in the system. This technique was applied to a real-world network in Mandya (Karnataka, India). It successfully detects a pipe leak in the real network (trunk-main model). Hence the performance of the technique in a complex DMA in real-time is unknown. This method shows some potential for online detection burst.

2.3 SUMMARY

This chapter provided a review of literature related to the WDS burst detection and localisation techniques. It also provided a review of data assimilation methods that could be or already applied in the WDS. Section 2.1 provides brief introduction and reveal the purpose of the literature review. In Section 2.2 the papers dealing with burst detection and localisation technique in WDS with an emphasis on applications of hydraulic techniques were reviewed. Section 2.3 reviewed WDS data assimilation techniques.

The main conclusions can be drawn from the literature review in Section 2.2 are as follows:

- Two basic types of automated WDS pipe bursts/leaks detection models seem to exist. The data-driven based techniques analyse the signals' values from a SCADA system to detect WDS pipe bursts whilst model-based techniques make use of both WDS hydraulic model and meters/other data for automated WDS pipe burst detection.
- Most of the hydraulic detection and localisation techniques are developed for offline problems including WDS pipe bursts detection. Few burst detection techniques such as statistical/AI-related techniques can be used to detect and locate WDS pipe bursts in the near real-time.
- The application of data-driven techniques performed reasonably well with real-life WDS data for detection purposes but was rarely used for WDS pipe burst localisation. However, these techniques can become redundant / difficult to use when there is a WDS configuration change (i.e. status of valves/pumps, changes or more permanent changes are introduced, such as rezoning) or system load (i.e. demands) changes substantially.
- Some model-based techniques such as negative pressure-based and transient-based techniques can be problematic in a real-life WDS due to a large number of noisy WDS observations and high frequency of observed data required. In addition, accurate WDS transient models still do not seem to exist given the complexity of transient phenomena modelled in the network context. While other model-based techniques in the literature such as Sousa et al. (2015), Anjana et

al, (2015) have not been tested/validated on complex real-life WDS because they were developed based on simple water networks or pipelines studied under laboratory conditions. Noise-free artificial 'observed' data is often used in related case studies.

The above conclusions illustrate that existing burst detection techniques from literature are not perfect. Real-life WDS challenges still present a significant barrier for a number of techniques presented. Despite some advantages of data-driven burst/leak detection techniques (see above), a hydraulic model-based technique still seems to be a viable alternative, for both detecting and locating pipe bursts/leaks in the near real time in a WDS.

It is widely known that model-based techniques make use of the offline calibrated hydraulic models which can deal with changing WDS configuration or load but only up to a point, as they make use of fixed calibration parameter values that were obtained prior to model running in near real-time. Hence, a novel detection and localisation technique, an online hydraulic model is required to overcome the above limitations by using which constantly adapts to changing conditions in the network based on incoming observations. The online hydraulic model would comprise of a combination of techniques including a state estimation technique that considers model and/or data uncertainties.

The novel online hydraulic model technique should be able to perform a near real-time detection and localisation of WDS pipe bursts in a reliable and timely manner within a DMA. Such technique can help WCs to reduce the water losses from a WDS, improve WCs SIM score. It can also help WCs to response quickly to WDS pipe bursts in the near real-time and facilitate appropriate interventions and/or repairs.

From the review of data assimilation methods in Section 2.3, it can be concluded that:

- DA method has become a popular mathematical tool for state estimation in many discipline however, its application in the water industry or research has been so far limited.

- An adaptation of ensemble-based DA method has gained popularity in the WDS-related research to do deal with WDS non-linear hydraulic relationship such as the PF method.
- Application of DA methods in WDS has been so far limited predominantly to simple and small WDS section, often with artificial observation data.

The number of DA methods applied to solve WDS state estimation is scarce compared to other discipline due to lack of good model or poorly calibrated model. The lack of research in the field of online hydraulic model based on a DA method applied for near real-time WDS modelling especially for burst detection and localisation is apparent, which creates the grounds for the work addressed in this thesis.

The most commonly used DA methods in hydraulic engineering-related field are the KF method, EnKF method and PF method. No research papers have so far compared the performance of these methods on a real-life WDS model and its respective observed data. The next chapter overcomes this by analysing the above three DA methods which are integrated with a suitable demand forecasting models to form three respective online WDS hydraulic models. The performances of these three models are compared on a real-life case study in Chapter 4, including comparison to an offline hydraulic model performance. The best performing online hydraulic model is then selected to develop an online hydraulic model for pipe burst detection and localisation in Chapter 5.

3 DATA ASSIMILATION METHODOLOGY

3.1 INTRODUCTION

The application of DA methods in hydraulic modelling of WDS is still at research and development stage. The DA methods are commonly in other disciplines such as meteorology, oceanography, hydrology and climatology (Evensen, 2003; van Leeuwen 2009). The DA method is based on the Bayesian statistical framework including three main components: (1) a state model forecasts (i.e., WDS states/WDS model parameters such as water demand), (2) observation operator, (a numerical model which map state model forecasts to the corresponding WDS observations) and (3) probability density function of the errors of hydraulic model predictions and their corresponding WDS observations.

The potential benefits of utilising online hydraulic modelling of WDS include reducing the operational costs and optimising the operational performance (Rao & Salomons, 2007). Davidson and Bouchart (2006) and Preis et al. (2010) highlighted online hydraulic modelling of WDS provide a greater understanding of system states in the near real-time. Therefore, online hydraulic modelling of WDS can be repeated for a given time-period, to identify optimal control strategies (Broad et al. 2010). To ensure the DA method perform well in any system including WDS, Hutton et al. (2011) revealed the principal sources of uncertainty: model structural uncertainty (hydraulic model), parameter uncertainty (i.e, water demand, pipe roughness), and data or measurement uncertainty (i.e. flow and pressure). Hutton et al. (2011) further explained that the impact of not dealing with the above-mentioned uncertainties can potentially lead to poor model performance or decision. In a WDS hydraulic model, water demand can be argued to be the most significant source of uncertainty that can affect quality of WDS model predictions (Herrera et al., 2010; Preis et al., 2010).

This chapter presents three online hydraulic models to quantify and reduce uncertainty for WDS state estimation. The three selected DA methods are the KF method, EnKF method and PF method. These methods are selected due to their good performance in other disciplines (meteorology and hydrology), their ease of algorithm

implementation and the fact that they represent well a range of DA methods in terms of various characteristics such as computational efficiency and accuracy.

The chapter is organised as follows:

- Section 3.2 explains the difference between the offline and online hydraulic modelling;
- Section 3.3 describes the concept behind the online hydraulic modelling of the WDS;
- Section 3.4 introduces the first online hydraulic model, KF method;
- Section 3.5 introduces the second online hydraulic model called the EnKF method;
- Section 3.6 introduces the third online hydraulic model, PF method;
- Section 3.7 outlines the metrics used to compare the performance of an offline hydraulic model and three online hydraulic models;
- Section 3.8 summaries the chapter findings.

3.2 OFFLINE VS ONLINE HYDRAULIC MODELLING

The offline hydraulic modelling of the WDS make use of historical time series flow/pressure data to calibrate the hydraulic model. Once calibrated this way, the offline model is used to conduct strategic/contingency planning of the WDS and also for pipe bursts/leaks detection. It can also be used to predict future WDS status, flow rate and pressure for network resilience or development planning. The offline calibrated hydraulic model may not represent well the current state of the WDS for operational purposes especially in emergency events (Preis, et al., 2011). The reason for this is that once calibrated, offline model is assumed to represent WDS reality well and is used accordingly, i.e. without being modified for often prolonged periods of time (i.e. until next use or re-calibration which may be months or years away).

Figure 3.1 shows the schematic diagram of an offline and online hydraulic models.

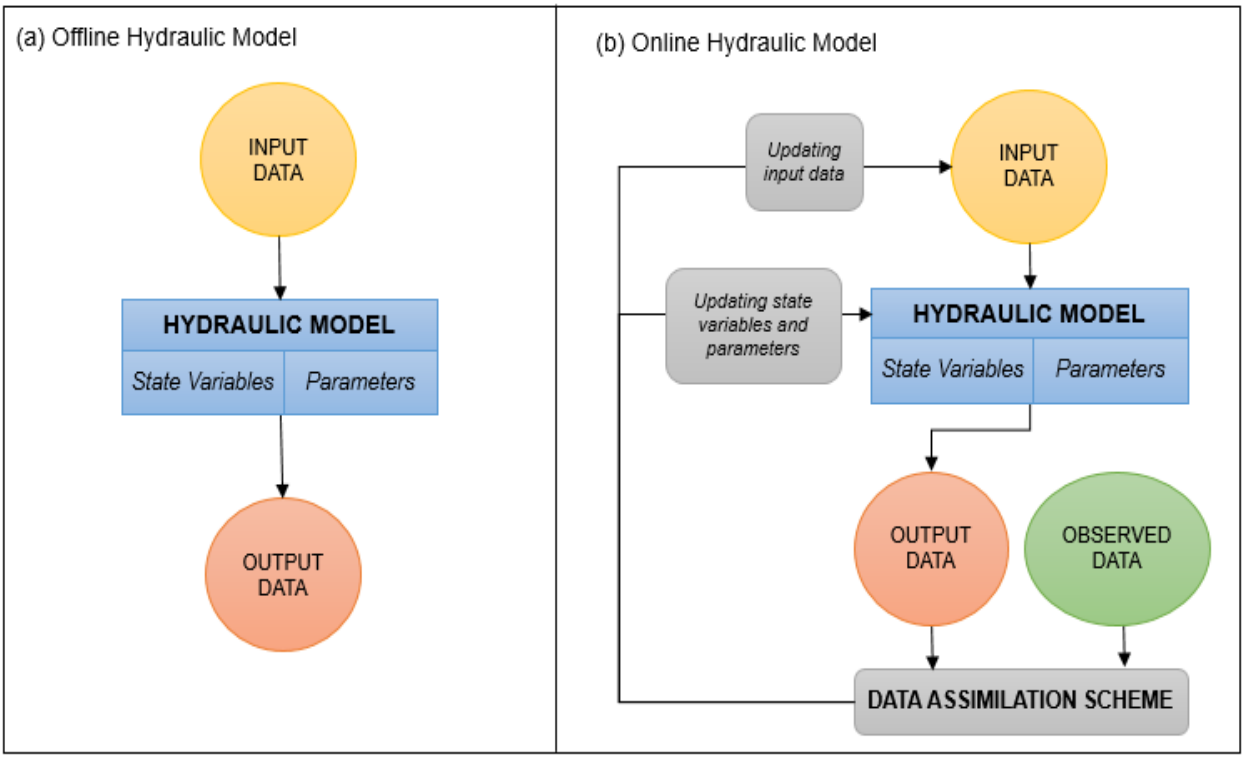


Figure 3.1: The schematic diagram of hydraulic model runs either (a) without a DA method and (b) with a DA method.

The key difference between an offline and online hydraulic model is in the use of DA. Unlike the offline model, the online hydraulic modelling make use of the latest WDS observations along with the uncertainties to correct the predicted WDS parameter and system states at each time step via a DA method (refer to Figure 3.1).

3.3 Online Hydraulic Modelling Concept

3.3.1 Data Assimilation Methods

The online hydraulic model aims to predict the hydraulic states of the WDS in the near real-time and then correct the predictions when the observations become available. This online hydraulic modelling of the WDS involves a combination of the conventional hydraulic model (such as e.g. EPANET 2.0), a Water Demand Forecasting Model (WDFM) and a DA method. The main reason for employing a DA method is its ability to quantify hydraulic model and observations' errors. The DA method corrects hydraulic model states by optimally combining model predictions with the

corresponding WDS observations. This process is regarded as predictor-corrector loop process. The steps of online hydraulic modelling are as follows:

1. **State prediction:** this step is where the WDFM is run to forecast water demands for the next time step. These forecasted water demands are then used to drive the hydraulic model from the known initial system state to the next hydraulic state. The outputs from the hydraulic simulation are pipe flow rates, tank levels and nodal pressures in the network.
2. **State correction:** The DA method (i.e., KF/EnKF/PF method) is used to correct both water demands and WDFM parameters predicted in the previous step. This method is driven by the difference between predicted WDS states and the corresponding WDS observations (coming from flow/pressure meters) at the current time step. The corrected water demands (obtained by correcting the WDFM parameters) are then inserted back into the hydraulic model to obtain the optimal WDS states.

Note that in the case where WDFM is not used, the correction step may involve updating demands (and other) hydraulic model inputs directly. The above prediction and correction steps are repeated at each time step during the online hydraulic modelling.

The starting point for online hydraulic modelling is the offline calibrated hydraulic model whose inputs/parameters are then continuously updated. It is anticipated that, when compared to the offline model, this should result in improved online hydraulic model predictions that are closer to field observations. The online hydraulic model can be run for successive weeks or longer, depending on the purpose of the simulation.

The next section introduced the water demand forecasting model used in this work.

3.3.2 Water Demand Forecasting Model

WDFM is an integral part of the online hydraulic model (refer to Figure 3.1). WDFM is commonly used by WCs around the world to forecast hourly or weekly demand to manage the operation of WDS efficiently. Majority of WDFM researchers such as

Herrera et al. (2010) and Adamowski and Karapataki (2010) identified water demand as the critical parameter to predict the WDS behaviour. Hence, vast majority WDFMs' parameter include at least element of water demands. WDFMs can be found in hydraulic engineering articles such as ARIMA (Box & Jenkins, 1976), M5 Model tress (Quinlan, 1992), Artificial Neural Network (Mounce, 2005), Multi-Linear Regression (MLR) analysis (Adamowski, 2008) and Time series analysis (Adamowski & Karapataki, 2010). Other models can be found in Herrera et al. (2010).

In this thesis, a simple WDFM is chosen after an experimentation of WDFMs based on seasonal Autoregressive Integrated Moving Average (ARIMA) and MLRs. The goal of the WDFM experimentation was to find an appropriate WDFM for this thesis. Hence, the experimentation involved developing several WDFMs using 80% of historic water demands and testing on the remaining 20% of data. The MLR-based model is selected because it predicts water demands more accurately when compared to the ARIMA-based WDFM. This selection is also supported by Adamowski (2008). The MLR-based WDFM is capable of forecasting the water demand every time step (i.e., every 15 minutes) for a fixed lead time forecasting horizon (e.g. 15 minutes). The forecasted demand is a function of demands from previous time steps as follows:

$$\boldsymbol{d_t} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{w_{t-i}} \boldsymbol{d_{t-i}} \qquad (11)$$

where n is the number of demand model parameters; $\boldsymbol{d_{t-i}}$ is the water demand from time step *t-i,* d_t is the forecast demand at current time step and $\boldsymbol{w_{t-i}}$ is the associated weight; *i* is index denoting previous time steps (e.g. *t-1* denotes 15 minutes before current time assuming that time step is 15 minutes).

In the UK, WCs organise their WDS into multiple District Metered Areas (DMAs). The DMA is a defined area of the WDS isolated by valves and monitored by flow meters at all entry and import/export points. The DMA water demand can be estimated from the quantity of water entering and leaving the DMA plus corresponding changes in tank volumes (if any). The recorded DMA inflows and/or outflows are transmitted to the Supervisory Control and Data Acquisition *(*SCADA*)* database. The MLR-based WDFM

uses flow data from the SCADA database to forecast DMA water demands and to calibrate the WDFM parameters.

The steps to develop the MLR-based WDFM is summarised in Figure 3.2.

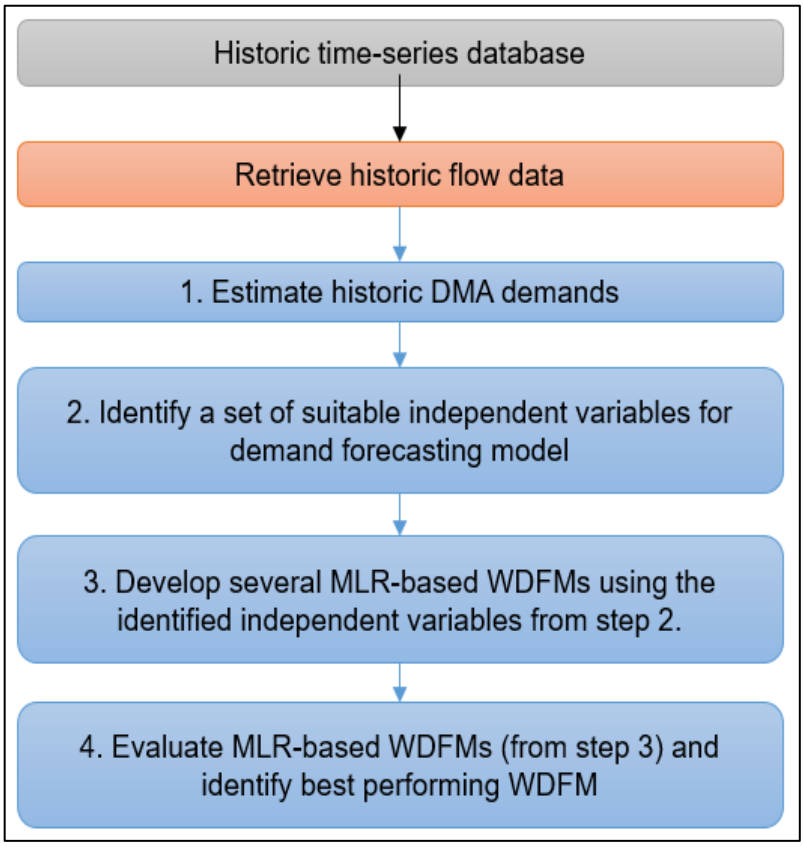


Figure 3.2: The flowchart of developing the water demand forecasting model.

Prior to the first step, the WDFM retrieves '*d*' full days of a DMA historic flow data from the Time Series database. The '*d*' full days of past flow data starts from 00:00 hour to 23:45 hour. Hence, the WDFM collates historic flow data starting from a time step, 23:45 hour of the previous day to the 23:45 hour of the last '*d*' day. The '*d*' days of historic flow dataset is then transferred to first step of the WDFM.

The first step estimates the '*d*' days of DMA water demands using the DMA boundary flow meters (i.e. historic DMA inflows and outflows) and tank level changes (if tank exists in the DMA).

The second step determines the sets of independent variables (demands at different points in the past) for the MLR-based WDFM. In order to find the optimal lag time steps,

the coefficient of correlation, r_d shown in equation (12) is used here. The dependent variable of the MLR-based WDFM is the most recent water demand in the '*d*' days of historic water demand dataset (obtained from step 1). It is denoted here as the '*current*' water demand, d_t. The independent variables are past demands from different lag time steps (*i.e.,* d_{t-1} or d_{t-96}, or d_{t-674}). A r_d value that is close to 1 shows a strong linear relationship between the between the two variables. The correlation coefficient measures the strength of the association between the two water demand variables is as follows:

$$\mathbf{r_d} = \frac{\mathbf{n}\sum \mathbf{d_{t1-i}d_{t2-i}} - \sum \mathbf{d_{t1-i}}\sum \mathbf{d_{t2-i}}}{\sqrt{\mathbf{n}\sum \mathbf{d_{t1-i}^2} - (\sum \mathbf{d_{t1-i}})^2}\sqrt{\mathbf{n}\sum \mathbf{d_{t2-i}^2} - (\sum \mathbf{d_{t2-i}})^2}} \quad (12)$$

where r_d is the water demand correlation coefficient; *n* is the number of points in the observed dataset; *i* is the increment factor; *d* is the water demand; *t1* and *t2* are time lag 1 and 2 respectively; d_{t1} is the water demand at '*current*' time step while d_{t2} is the water demand dataset at *'t2'* time step. *'t2'* can be water demand from the previous time step (i.e, 1 day ago (t-96), day+15 minutes (t-97), week (t-672) assuming the time step interval of 15 minutes).

The third step develops several MLR-based WDFM for the studied WDS using independent variables identified in step 2. After the formation of the MLR-based WDFMs, each forecasting model is calibrated using part of the observed demand data. The Maximum Likelihood Estimation (MLE) method (Andersen, 1970) is used to obtain the values of each WDFM's respective weight coefficients.

The final (fourth) step evaluates several MLR-based WDFMs created in step 3 and identifies the best performing WDFM. In order to review their forecasting performance of each developed MLR-based WDFMs, a different period of the observed demand data (test datasets) is used. This observed demand dataset is not part of the dataset used to develop WDFMs. The best performing WDFM is selected based on the highest correlation coefficient value obtained between the WDFM inputs and outputs.

The WDFM is recalibrated periodically in near real-time. This is done when the moving average of normalized DMA demand error between the estimated and corrected demand is above a predefined threshold value. The moving average value is the mean of normalized DMA demand errors calculated over a sliding window size of a selected length (i.e., 4, 48, 96 time steps). Therefore, the lowest value among the sum of normalized DMA demand errors for different window sizes is used to determine the optimal window size. This is to minimise parameter risk and model risk due to arbitrary small and large window size respectively. The threshold value is defined as the average of normalized DMA demand errors (i.e. differences between the estimated and corrected demands).

3.3.3 Hydraulic Model

In this work, EPANET 2.0 is used as a WDS hydraulic model. This software tool is developed by the United State Environmental Protection Agency (U.S. EPA) to solve nonlinear hydraulics of a WDS. The EPANET uses the Global Gradient Algorithm (GGA), a variant of Newton-Raphson method proposed by Todini and Pilati (1988) to perform hydraulic analysis of any WDS section (refer to Appendix A1 for further details). It is widely used for demand analysis, flushing event or leakage detection in a WDS section, strategic and contingency planning.

The nonlinear hydraulic relationships of WDS are defined by the conservation of mass and energy. They are written in a form of nodal flow continuity and pipe head loss equations. The nodal flow continuity equation is as follows (Kang & Lansey, 2009):

$$\sum_{\boldsymbol{p} \in \boldsymbol{J_{in,n}}} \boldsymbol{Q_p} - \sum_{\boldsymbol{p} \in \boldsymbol{J_{out,n}}} \boldsymbol{Q_p} = \boldsymbol{q_n} \qquad (13)$$

where subscript *n* and *p* are denoting node and pipe respectively; Q_p are the flows of pipes connected to node *n*; q_n is nodal demand; $J_{in,n}$ and $J_{out,n}$ are the set of pipes going into and coming out from a node respectively.

The pipe head loss equation is as follows:

$$\boldsymbol{H_1 - H_2 = h_{L,p}} \quad (14)$$

where H_1 and H_2 are the total energy at upstream (1) and downstream (2) nodes respectively; $h_{L,p}$ is the pipe headloss which calculated via Hazen-Williams formula:

$$\boldsymbol{h_L = K_u \left(\frac{Q}{C_{HW}}\right)^{1.852} \frac{L}{D^{4.87}}} \quad (15)$$

where *D, L* and C_{HW} are diameter, length and Hazen-Williams roughness respectively; and K_u is a dimensionless constant.

Hydraulic model pipe flow and nodal pressure vary at each time step due to changing demands and fixed nodal head (i.e., tank condition) via a quasi-dynamic analysis (also known as Extended Period Simulation) in the EPANET model.

The next three sections provide details of three online hydraulic models considered in this thesis.

3.4 Online Hydraulic Model #1: Kalman Filter Method

The first online hydraulic model considered is the KF method. The KF method is considered in this work as it represents the conventional DA method with widespread use in many areas of engineering. The standard KF method requires linear relationship between system states and parameters being corrected which, obviously, is not the case with WDS. Having said this, the WDS hydraulics are not that much non-linear with key, near quadratic non-linearity arising from the head loss equation (refer to equation (**15**)). Given this, in the work conducted here, the KF method is still used but after being modified to address the above nonlinearity.

The detailed theoretical background of the KF method is described in Appendix B1. The hydraulic state residuals between the predicted and observed states (i.e., flow rates, pressures) are used to correct the model state vector (i.e., forecasted DMA water demands) via the KF method and it can be expressed as:

$$\boldsymbol{x_{t,i}^{c} = x_{t,i-1}^{p} + K_{t,i}\left(z_t - H(x_{t,i-1}^{p})\right)} \quad (16)$$

where subscript *t and i* are the current time step and iteration step respectively, superscript *c* and *p* are the corrected and predicted value; $\boldsymbol{x}_{t,i}$is the corrected state vector (demands and tank levels); $\boldsymbol{x}_{t,i-1}$is the corrected state vector from the previous iteration step); $\boldsymbol{K}_{t,i}$ is the Kalman gain; $\boldsymbol{z}_t$ is the WDS observations (i.e., observed flow rates, tank levels) and ***H*** is the observation operator which converts the model states (i.e., demands, tank level) to the WDS observations (i.e., flow rates, pressure, tank levels).

The standard Kalman gain $\boldsymbol{K}_{t,i}$ (refer to equation (3)) is in literature review is modified to equation (17). This is to consider the nonlinearity of the WDS. The modified Kalman gain is calculated as:

$$\boldsymbol{K_{t,i} = P_{t,i}^{p} H^{T} (H P_{t,i}^{p} H^{T} + R_t)^{-1}} \quad (17)$$

where $\boldsymbol{H}$ is the observation operator that maps the predicted water demands to the corresponding WDS flow observation space; $\boldsymbol{H^T}$ is the transposed observation operator; $\boldsymbol{P}_{t,i}^{p}$ is the prior error covariance matrix and $\boldsymbol{R}_t$ is the observation error covariance matrix;

The prior, observation and process error covariance matrix are estimated from the following equations:

$$\boldsymbol{P_{t,\ i}^{p} = \left[P_{t,i-1}^{c} + P_{t,\ i-1}^{Q}\right]} \quad (18)$$

$$\boldsymbol{R_t = [\bar{r}\, z_t]^2} \quad (19)$$

$$\boldsymbol{P_{t,i}^{Q} = [\bar{p}\, x_{t,i-1}^{p}]^2} \quad (20)$$

where $\boldsymbol{P}_{t,\ i}^{p}$ is the prior error covariance matrix; $\boldsymbol{P}_{t-1,\ i}^{c}$ is the posterior error covariance matrix at the previous time step, *t-1*; $\boldsymbol{P}_{t,\ i-1}^{c}$ is the posterior error covariance matrix at

the current time step, t; $\boldsymbol{P}_t^Q$ is the process error covariance matrix**;** $\boldsymbol{z_t}$ is the WDS observation (i.e. flow rate); $\bar{p}$ is the moving average of normalised residual error between the WDS predictions and the corresponding WDS observations and $\bar{r}$ is the moving average of normalised residual error between the WDS corrections and the corresponding WDS observations.

The main advantage of using moving average of normalised residual data, $\bar{p}$ and $\bar{r}$ is to reduce the impact of anomalous residual errors between observed and predicted/corrected values on state estimation. The anomalous residual data often occur when there is an under/over predicted value or anomalous observation. This either reduces or increases the weight of Kalman gain quicker during iteration hence reducing the number of iteration steps. Whilst the error covariance matrix would make the weight of Kalman gain close to 1 at every iteration step regardless if the observed data is anomalous or predictions are under/over-estimated.

The diagonal element of the observation and process error covariance are calculated using equations (18) and (19) respectively while the off-diagonal terms are zero. The posterior error covariance matrix is calculated as:

$$\boldsymbol{P}_{t,\ i}^{\boldsymbol{c}} = (\boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{t,i}}\,\boldsymbol{H})\boldsymbol{P}_{\boldsymbol{t,\ i}}^{\boldsymbol{p}} \qquad (21)$$

where $\boldsymbol{I}$ is the identity matrix.

Equations (18) to (21) are used iteratively until all state errors (demand errors) between the corrected demands from the current and previous iteration is approaching zero (or reached a defined maximum number of iterations). This is because of the non-linear hydraulic relationship. The final state estimates in equation (16) are taken as the corrected state vector. At the initial iterative step for each time step, $\boldsymbol{P}_{t,i-1}^{c}$ is a zero matrix while the diagonal element of $\boldsymbol{P}_{t,\ i-1}^{Q}$ is the product of average relative state error and state vector.

The above modified KF method can be regarded as an iterative KF method and can be implemented for successive linearisation at each time step.

3.5 ONLINE HYDRAULIC MODEL #2: ENSEMBLE KALMAN FILTER METHOD

The Ensemble Kalman Filter method (EnKF) (Evensen, 1994) is a suboptimal estimator which corrects the ensemble of state vector (demands) without the need of observation operator and covariance matrices. The theoretical background of the EnKF method is described in Appendix B4. The chosen variant of the EnKF method is perturbed EnKF method (Burgers, et al., 1998). The perturbed EnKF method is used in this thesis due to its relatively simple implementation and, at the same time, ability to achieve good prediction accuracy (Sun, et al., 2009). As noted by the latter authors, the perturbed EnKF method require a smaller ensemble size compared to other EnKF method variants such as Singular Evaluative Interpolated KF (SEIK) or Local Ensemble Transform KF (LETKF). Even though both SEIK and LETKF outperform the perturbed EnKF method in terms of forecasting accuracy and state corrections, both SEIK and LETKF methods also require more complex algorithms and larger ensemble sizes resulting in increased computational times. When compared to other inverse methods such as general Least-Square or MLE, the perturbed EnKF corrects the WDS states estimates without the need to do several iterations of model state correction at each time step. Hence, the perturbed EnKF method is preferred variant of the EnKF method in this thesis. Further details of the EnKF variants can be found in Appendix B5.

The correction step of the perturbed EnKF method is as follows:

$$\boldsymbol{X_t^c} = \boldsymbol{X_t^p} + \boldsymbol{K_t^{xz}} \left(\boldsymbol{Z_t} + \boldsymbol{e_t} - \boldsymbol{H}(\boldsymbol{X_t^p}) \right) \tag{22}$$

where $\boldsymbol{e_t}$ contains artificial random noise and the general procedure of calculating Kalman gain is:

$$\boldsymbol{K_t^{xz}} = \boldsymbol{C_t^{xz}} \left[\boldsymbol{C_t^{yy}} + \boldsymbol{C_t^{zz}} \right]^{-1} \tag{23}$$

The ensemble statistics, mean and covariance are estimated as:

$$\boldsymbol{C_t^{xz}} = \frac{\boldsymbol{E_{t,x} E_{t,z}^T}}{\boldsymbol{N-1}} \tag{24}$$

$$\boldsymbol{C_t^{yy}} = \frac{\boldsymbol{E_{t,y}E_{t,y}^T}}{\boldsymbol{N-1}} \quad (25)$$

$$\boldsymbol{C_t^{zz}} = \frac{\boldsymbol{E_{t,z}E_{t,z}^T}}{\boldsymbol{N-1}} \quad (26)$$

$$\begin{cases} \boldsymbol{E_{t,x}} & = \boldsymbol{X_t^p} - \boldsymbol{\mu_{t,x}} \\ \boldsymbol{E_{t,y}} & = \boldsymbol{H}\left(\boldsymbol{X_t^p}\right) - \boldsymbol{\mu_{t,y}} \\ \boldsymbol{E_{t,z}} & = \boldsymbol{Z_t} - \boldsymbol{\mu_{t,z}} \end{cases} \quad (27)$$

where *N* is the ensemble number; $\boldsymbol{X_t^c}$ and $\boldsymbol{X_t^p}$ are the state ensemble matrix; $\boldsymbol{Z}_t$ is the ensemble of WDS observations; $\boldsymbol{K_t^{xz}}$ is the Kalman gain for correcting the state ensemble matrix; $\boldsymbol{C_t^{xz}}$ is the cross error covariance of ensemble state estimates and WDS hydraulic model predictions, $\boldsymbol{C_t^{yy}}$ is the WDS hydraulic model predictions error covariance; $\boldsymbol{C_t^{zz}}$ is the WDS observations error covariance; $\boldsymbol{H}$ *i*s the observation operator; *T* is the transpose of the designated matrix; $\boldsymbol{E}_{t,x}$, $\boldsymbol{E}_{t,y}$ and $\boldsymbol{E}_{t,z}$are the ensemble error of model state estimates, WDS predictions and the corresponding observation errors respectively; $\mu_{t,x}$, $\mu_{t,y}$ and $\mu_{t,z}$ are the ensemble mean of the model state, hydraulic model prediction and the WDS observation respectively.

It is also important to notice the observation operator, $\boldsymbol{H}$ in equation (17) first the model states space to the WDS observation space via a distributed hydraulic model. Then extract the hydraulic model predictions (i.e., flow rates, pressures or tank levels) at the corresponding WDS observations' location. The difference between the extracted hydraulic model prediction and the corresponding WDS observation is used to correct the model states, $\boldsymbol{X_t^p}$. The corrected model states, $\boldsymbol{X_t^c}$ is then inserted into the distributed hydraulic model to be obtained the corrected WDS states (possibly optimal WDS states).

To further reduce the computational time of the EnKF method at each time step, the mean of ensemble corrected model states and corrected WDS states are stored in a database. Therefore, the model state forecasts are based on the historic mean of the corrected model states. The model state forecasts are then perturbed. The ensemble model forecasts are propagated forward to the next observational time step via a distributed hydraulic model, hence equations (22) to (27) are repeated.

3.6 ONLINE HYDRAULIC MODEL #3: PARTICLE FILTER METHOD

The Particle Filter (PF) method (van der Merwe, et al., 2000) makes use of a genetic type mutation selection sampling technique along with a set of particles (samples) to correct model state predictions. The particles are generated from the proposal distribution without requiring assumptions relating to the model state distribution. The particles represent the posterior distribution of WDS process given noisy and/or partial WDS observations. The theoretical background of the PF method is described in Appendix B5.

The PF method is chosen in this thesis due to its ability to approximate the posterior distribution via an empirical distribution (Jardak, et al., 2013) . Secondly, in the WDS-related literature, the performance of PF method has not been compared to both KF and EnKF method on the same case study.

The central concept of the PF method is to approximate the posterior distribution of model state estimates ($\boldsymbol{X_t}$) when WDS observations become available:

$$\boldsymbol{P(X_t|Z_t) = P(Z_t|X_t)P(X_{t-1}|Z_{t-1})} \quad (28)$$

where $P(\boldsymbol{X_t}|\boldsymbol{Z_t})$ represents the likelihood of the current observations, given the model states; $P(\boldsymbol{Z_t}|\boldsymbol{X_t})P(\boldsymbol{X_{t-1}}|\boldsymbol{Z_{t-1}})$ is the prior distribution of model states based on the WDS observations at the previous time step ($\boldsymbol{Z_{t-1}}$).

The posterior distribution is represented by an ensemble of particles (model state estimates) with their associated weight:

$$\boldsymbol{P(X_t|Z_t) = \sum_{i=1}^{Np} w_i \delta(X_t - X_t^i)} \quad (29)$$

where δ is the Dirac delta function, the superscript *i* is the particle index, and the sum of weights is equal to 1.

When the number of particles is large, the PF method approaches the optimal posteriors estimates (Hutton, et al., 2012). The Sequential Importance Sampling (SIS) is used to update the associated particle weights. This procedure is commonly used due to the general difficulty to sample directly from the posterior distribution. Therefore, the particles are then drawn from a proposal distribution $P\left(\boldsymbol{X}_t^i | \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_t\right)$ which leads to the following recursive weight update as each observation is assimilated, derived from Bayes' equation (refer to equation (8)) (Arulampalam, et al., 2002).

$$\boldsymbol{w}_t^i = \boldsymbol{w}_{t-1}^i \frac{\boldsymbol{P}(\boldsymbol{Z}_t | \boldsymbol{X}_t^i)\boldsymbol{P}(\boldsymbol{X}_t^i | \boldsymbol{X}_{t-1}^i)}{\boldsymbol{P}(\boldsymbol{X}_t^i | \boldsymbol{X}_{t-1}^i, \boldsymbol{Y}_t^i)} \tag{30}$$

The proposal probability density (refer to equation (30)) is used as the proposal density, which simplifies the weight update (van Leeuwen, 2009):

$$\boldsymbol{w}_t^i = \boldsymbol{w}_{t-1}^i \frac{\boldsymbol{P}(\boldsymbol{Z}_t | \boldsymbol{X}_t^i)}{\sum_{i=1}^{Np} \boldsymbol{P}(\boldsymbol{Z}_t | \boldsymbol{X}_t^i)} \tag{31}$$

where the superscript *i* is the particle index; subscript t, is the time step index; *w* is the associated weight; *Np* is the number of particles; $\boldsymbol{P}\left(\boldsymbol{Z}_t | \boldsymbol{X}_t^i\right)$ is the likelihood of the current observations, given the model states;

To apply PF method in the WDS context, the conditional probability of the WDS observations given the model states is often assumed to be Gaussian (Salamon & Feyen, 2010). Hence, normal distribution is chosen to reflect the forms of uncertainty present in the model:

$$\boldsymbol{P}\left(\boldsymbol{Z}_t | \boldsymbol{X}_t^i\right) = \frac{1}{\sigma\sqrt{2\pi}} \boldsymbol{exp}^{-\frac{1}{2}\left(\frac{H(X_t^i) - Z_t}{\sigma}\right)^2} \tag{32}$$

where σ is the standard deviation of the observation error.

After the propagation of the ensemble of particles (model state estimates), equations (31) and (32) are applied when a collection of WDS observation become available. The product sum of the particle weights and their associated state predictions is assumed to be the optimal estimate of model states:

$$x_t = \sum_{i=1}^{Np} w_t^i X_t^i \quad (33)$$

The optimal estimate of model states (corrected model states) from equation (33) is then inserted into the distributed hydraulic model to obtain the WDS states. Lack of WDS state modification leads to filter degeneracy, where the models provide a poor approximation of the posterior distribution. van Leeuwen (2009) highlights the larger weights are duplicated at the expense of poorer performing particles and this causes the filter to degenerate from the observations.

To overcome the PF degeneracy, the Stochastic Universal Re-sampling (SUR) is applied to re-sample particles (Salamon and Feye, 2010; van Leeuwen, 2009; Hutton et al. 2011). The SUR is found to be effective in offsetting the PF degeneracy issue and sample impoverishment issue that occur during the particle re-sampling.

The resampling of the particles follows four steps below:

1. Sort the particle weights in ascending order.
2. Choose a random number from a uniform density [0,1/ *Np*].
3. Starting from the random number, *np* segments with length 1/*Np* are laid on the line [0,1].
4. Choose a particle when the end of a line segment falls in the particle's weight bin.

The above SUR method is applied at every time step to ensure the filter do not degenerate.

3.7 Online Hydraulic Model Comparison Metrics

The three online hydraulic models introduced above are compared on a real-life case study in the next section with the aim to decide which online model will be used for real-time burst detection and location. To do this, two performance metrics are defined here.

The first performance metric is the Mean Absolute Error (MAE) which measures the closeness of hydraulic model predictions to their corresponding WDS observations. It is arguably one of the most popular performance metric and is calculated as follows:

$$\boldsymbol{MAE} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} |\boldsymbol{x_t} - \boldsymbol{y_t}| \quad (34)$$

where *n* is the number of data set; *t* represents time step index; x_t is the predicted WDS hydraulic model states and y_t is the system observations.

The second performance metric is the coefficient of determination (R^2) which describes how well a regression line of predicted hydraulic states fits a set of observation data. It ranges between 0 and 1 where 1 indicates that hydraulic model predictions match the corresponding WDS observations perfectly and 0 indicates the opposite. The equation of the coefficient of determination to measure variability of the data set is:

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum_t (\boldsymbol{y_t} - \bar{\boldsymbol{y}}_t)^2}{\sum_t (\boldsymbol{y_t} - \boldsymbol{x_t})^2} \quad (35)$$

where *n* is the number of data set; *t* represents time step index; x_t is the hydraulic model predictions; $\bar{y}_t$ is the mean of the WDS observations and y_t is the WDS observations.

3.8 SUMMARY

After introduction (section 3.1), this chapter describes and contrasts the concepts of offline and online hydraulic models in the WDS context (section 3.2). This is followed by the description of the WDS online hydraulic modelling concept used here (section 3.3) including its three main components: the data assimilation method, the water demand forecasting model and the hydraulic model. Three online hydraulic models are described in sections 3.4-3.6. They all use the same water demand forecasting model (Multi-Linear Regression based, refer to section 3.3.2) and the same hydraulic model (Epanet2 based, refer to section 3.3.3). The three online hydraulic models differ in the

DA method used. The methods analysed are as follows: (1) Iterative KF method, (2) Perturbed EnKF method and (3) PF-SUR method.

The iterative KF method works in a loop iteration to overcome its key limitation to work in a non-linear system like the WDS. The EnKF method employs a perturbation method variant to estimate the optimal WDS states. The PF method is combined with a Stochastic Universal Re-sampling (SUR) method to overcome PF degeneracy and output optimal WDS states. All three DA methods shown are capable of processing data in near real time and are capable of dealing with WDS model non-linearity (Hutton, et al., 2012).

At the end, in section 3.7, two metrics that will be used for comparison analysis between offline and online hydraulic models are defined as follows: MAE and coefficient of determination, R^2.

The results of comparison of offline and three online hydraulic models' performances are shown in the next chapter.

4 CASE STUDY FOR DATA ASSIMILATION METHODOLOGY

4.1 INTRODUCTION

The application of the offline model and online hydraulic models presented in Chapter 3 are illustrated in this chapter. The chapter compares and evaluates the performance of offline and online hydraulic models in a real-life Water Supply Zone (WSZ) model and data. The WSZ in question is divided into several District Metered Areas (DMAs). The historical data used are obtained from the United Utilities (UU) flow and pressure meters deployed in the WSZ. These historic data were stored in the UU database for DMA data record. The availability of real-life flow and pressure data in sufficient quantity was limited in each DMA. This prevented the online hydraulic models being demonstrated on multiple flow and pressure data within a DMA. However, the obtained historic flow and pressure measurements are sufficient to make comparisons between the developed offline/online hydraulic models.

The chapter is organised as follows. After this introduction,

- Section 4.2 describes the case study area used in this thesis;
- Section 4.3 outlines the hydraulic sensor or meter data used for the case study;
- Section 4.4 describes how the offline hydraulic modelling work;
- Section 4.5 explains and defines parameters value for the online hydraulic model;
- Section 4.6 discusses the case study results;
- Section 4.7 summaries the report and provides the concluding remarks.

4.2 CASE STUDY AREA DESCRIPTION

The online and offline hydraulic models are tested on a real network which is renamed for confidentiality reasons as WSZ01. It supplies water to approximately 16,000 customers under gravity from a tank with an average daily demand of 124 l/s. The WSZ01 model consists of 1 tank, 3287 nodes, 2595 pipes and 907 valves. There are 8 DMAs and all the DMAs have 1 inlet flow meter and 1 outlet flow meter except DMA02 which has 2 inlet flow meters. Figure 4.1 shows WSZ01 model without DMA03, DMA05

and DMA08. The reasons for excluding three DMAs will be explained later in this thesis. The hydraulic model displayed in Figure 4.1 will be referred as reduced WSZ01 from henceforth.

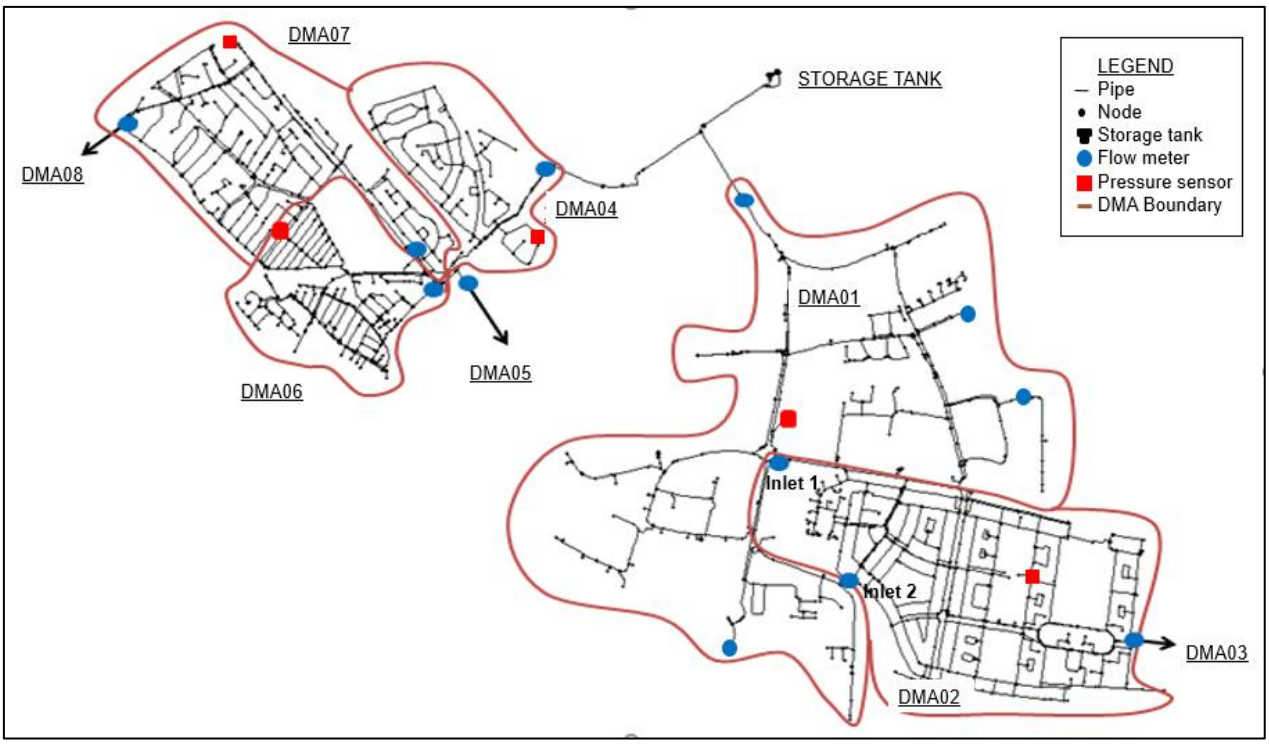


Figure 4.1: Reduced WSZ01 model with flow meter (blue dot) and pressure meter (red square) locations

Table 4.1 shows the proportion of water users in all the 8 DMAs. DMA01 and DMA05 have a large proportion of industrial users compared to other DMAs. This is because DMA01 industrial users include a large retail park, two large pharmaceutical companies and a local airport. The industrial user in DMA05 is a Wastewater Treatment Works (WwTW).

Table 4.1: The percentage of demand in each DMA

Type of User	**DMA01**	**DMA02**	**DMA04**	**DMA06**	**DMA07**
Domestic Users	61%	93%	96%	95%	92%
industrial Users	39%	7%	4%	5%	8%
Average DMA daily demand (l/s)	25.87	18.42	8.62	12.07	8.81

For the purpose of analyses conducted in this Chapter, 5 DMAs (DMA01, DMA02, DMA04, DMA06 and DMA07) are selected and respective hydraulic models are extracted from the WSZ01. This is because both flow and pressure data are sufficient for data analysis and their respective model were last calibrated in 2013. All the 5 DMAs are not a pressured managed DMA. The pressured managed DMAs (DMA03, DMA05 and DMA08) are excluded for further analysis. This is because some of pressure data are missing and other pressure data are nearly flat-lining especially between 00:00 and 05:00 (using the 24-hour clock). The three pressure managed DMAs' model have control rules to fix the pressure data between 23:00 and 06:00. Therefore, the flat-lining DMA pressure data of the three pressured managed DMAs is not sensitive to their respective DMA demand diurnal cycle.

4.3 SENSOR DATA

In this case study, there is a total of 14 flow meters, 1 tank level meter and 5 pressure meters available.

DMA01 has 5 flow meters and 1 pressure meter. One of the four flow meters is located at the inlet of DMA01 to monitor flow entering DMA01. Whilst the other 3 flow meters are used to monitor amount of water going to 3 industrial users within DMA01. The 3 industrial users in DMA01 are a local airport and 2 pharmaceutical companies.

DMA02 has 2 inlet flow meters, 1 outlet flow meter and a pressure meter. The pressure meter is located at the highest point in DMA02. The DMA02 outlet flow meter is the same as DMA03 inlet flow meter.

DMA04 has 1 inlet flow meter and 2 outlet flow meters and a pressure meter. The pressure meter is located at the highest point in DMA04. The 2 outlet flow meters are the DMA05 and DMA06 inlet flow meter.

DMA06 has 1 inlet flow meter and 1 outlet flow meter (to DMA07) along with 1 pressure meter. The pressure meter is located at the highest point in DMA06.

DMA07 has 1 inlet flow meters, 1 outlet flow meter and 1 pressure meter. The pressure meter is located at the highest point in DMA07. The DMA07 outlet flow meter is the same as DMA08 inlet flow meter.

A flow meter is installed at the WSZ01 tank level inlet to observe the tank inflow and a tank level meter is used to observe the water level within the storage tank. The pressure data are gathered at the highest point (elevation) of each DMA respectively (refer to Figure 4.1). The flow and pressure data are gathered between 1st February 2015 and 31st May 2015 (17 weeks).

The next section discusses offline hydraulic modelling and the reduced WSZ01 network model calibration.

4.4 Offline Hydraulic Modelling

The aim of offline calibration of the reduced WSZ01 hydraulic model is to modify or adjust model parameters so that model outputs (tank level, flow and pressure) reflect WDS observations. Figure 4.2 shows the processes of the model calibration process. The model calibration process involves trial and error method where different sets of model parameters are used until the model predictions match the corresponding WDS observations.

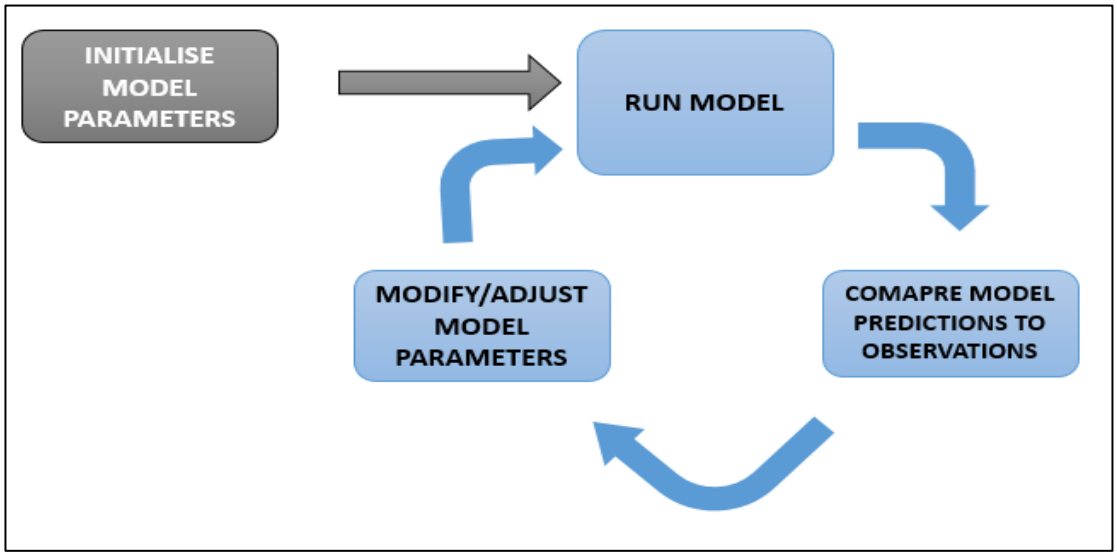


Figure 4.2: Model Calibration Process

The hydraulic data used for the offline calibration are from the DMA boundary flow meters, pressure meters and WSZ01 tank level recorder between 1st and 14th February 2015. The observed data from DMA01 inflow meter, 3 industrial flow meters and DMA01 pressure meter are used to calibrate DMA01 model. DMA02 model is calibrated by using observed data from DMA02 and DMA03 inflow meters and DMA02 pressure meter. DMA04 inflow meter and 2 DMA04 outflow meters (DMA05 and DMA06 flow meter) and DMA04 pressure meter are used to calibrate DMA04 model. DMA06 model is calibrated by using observed flow and pressure data from DMA06 and DMA07 flow meter and DMA06 pressure meter respectively. DMA07 model is calibrated by using observed data from DMA07 and DMA08 flow meter and DMA07 pressure meter.

The hydraulic model calibration (refer to Appendix A3) parameters adjusted using the trial and error technique are as follows:

- the roughness value of the cast iron and ductile iron pipes was increased by 15%;
    - i.e., 2.5mm to 2.8mm, 3mm to 3.4m, 4mm to 4.5mm

The roughness value of iron pipes was increased to increase the headloss across the reduced WSZ01 network.

- the relative opening (i.e. tau value) of the Throttle Control Valve (TCV) located at the DMA inlet was increased;
    - DMA01 TCV setting value from 75 to 110
    - DMA02_2 TCV setting value from 27 to 41
    - DMA06 TCV setting value from 2.5 to 7.5
    - DMA07 TCV setting value from 62 to 80
- Due to the difficulty of finding the roughness value of plastic pipes across the reduced WSZ01 network, TCV setting value of DMA01, 02_2 06 and 07 was increased to reduce the pressure head at their respective DMA inlet. The demand coefficients at 15 minutes intervals were modified so that predicted flow rates match the inlet flow rates. Hence all the DMAs' demand coefficients were changed.

All the above-mentioned modifications were made to ensure that the predicted flow/pressure data from the hydraulic model closely match the observed flow/pressure data.

The DMA demand coefficients is the difference between the DMA inflow and outflow data divided by the total number of properties within the studied DMA. The total number of properties includes small industrial users such as small corner shops and small offices. These small industrial users are unmetered and assumed to have similar domestic demand profile for the model simplification.

The hydraulic model calibration parameters that were remain unchanged were plastic pipes' roughness; DMA04 and DMA02_1 TCV setting value (0); and the base demand. During the reduced WSZ01 models' calibration, the roughness value of plastic pipes was increased whilst the iron pipes' roughness remains unchanged. The results show increment of plastic pipes' roughness value do not have significant impact on the predicted pressure after several trials. Therefore, the roughness of plastic pipes was reversed to their original value. As for the base demand, UU usually use information from billing database and MapInfo or ArcGIS 10.1 to estimate nodal base demand. These databases were not available for this research due to confidential protection of billing information and ArcGIS data.

The flow and pressure observations between 15th and 28th February 2015 are used to validate the hydraulic model calibration. Table 4.2 shows the data statistics for the reduced WSZ01 flow calibration and validation. Table 4.2 reveals a big difference between pre-calibrated flow data and observed flow between 1st and 14th February 2015. The results in Table 4.2 show the importance of calibrating the reduced WSZ01 hydraulic model.

As it can be seen from Table 4.2, the data statistics indicate the modelled (calibrated/validated) flow values of DMA01, DMA04, DMA06 and DMA07 match their respective flow observations. DMA02_1 and DMA02_2 model flow predictions also match the observed flow compared to the pre-calibration flow data.

Table 4.2: Data Statistics for the reduced WSZ01 flow calibration and validation

	Flow Data Statistics						
Data Statistics	**Flow meter**	**DMA 01**	**DMA 02_1**	**DMA 02_2**	**DMA 04**	**DMA 06**	**DMA 07**
	Model index	00172C37	X1920308_	03EA86BD	000D63F7	0016BBF0	02B2B748
MAE (l/s)	**Pre-Calibration***	11.253	3.292	3.960	6.624	5.228	2.807
	Calibration	0.001	0.220	0.234	0.001	0.001	0.001
	Validation	0.001	0.249	0.324	0.001	0.001	0.001
RMSE (l/s)	**Pre-calibration***	15.622	4.306	4.313	8.557	7.166	3.717
	Calibration	0.001	0.466	0.489	0.001	0.001	0.001
	Validation	0.001	0.374	0.433	0.001	0.001	0.001

*Pre-calibration is the previous model calibration before the new calibrated model. In the case of the reduced WDSZ01, it was last calibrated in late 2008.

Table 4.3 shows the data statistics for the reduced WSZ01 pressure calibration/validation. The pressure data statistics shows the reduced WSZ01 model pressure predictions improved after model calibration.

Table 4.3: Data Statistics for the reduced WSZ01 pressure calibration and validation

	Pressure Data Statistics						
Data Statistics	**Pressure meter**	**WSZ 01**	**DMA 01**	**DMA 02**	**DMA 04**	**DMA 06**	**DMA 07**
	Model index	00172CCC_	00172D7A	03E1B5A1	X1920407_	A0020A71	A01E113A
MAE (m)	**Pre-calibration**	0.12	0.85	1.18	0.40	1.54	1.22
	Calibration	0.07	0.35	0.30	0.27	0.93	0.65
	Validation	0.04	0.49	0.53	0.36	1.10	0.58
RMSE (m)	**Pre-calibration**	0.15	1.24	1.51	0.48	1.89	1.75
	Calibration	0.08	0.42	0.38	0.35	1.20	1.04
	Validation	0.05	0.65	0.72	0.42	1.20	0.74

Figure 4.3 to Figure 4.8 show the calibration plot of some DMA flows and pressures between 1st and 7th February 2015. Figure 4.3, Figure 4.5 and Figure 4.7 show the flow comparison between pre-calibration and calibration and observation for DMA01, DMA02 and DMA03 respectively. Figure 4.4, Figure 4.6 and Figure 4.8 show the pressure comparison between pre-calibration and calibration and observation for DMA01, DMA02 and DMA03 respectively.

Figure 4.9 to Figure 4.14 show the validation plot of some DMA flow and pressure between 15th and 21st February 2015. Figure 4.9, Figure 4.11 and Figure 4.13 show the flow comparison between pre-calibration and calibration and observation for DMA01, DMA02 and DMA03 respectively. Figure 4.10, Figure 4.12 and Figure 4.14 show the pressure comparison between pre-calibration and calibration and observation for DMA01, DMA02 and DMA03 respectively.

In Figure 4.3 and Figure 4.9, there are sudden flow spikes (sudden increase) in DMA01 due to water intake at the airport and two pharmaceutical companies. These flow spikes caused DMA01's sudden pressure drops in Figure 4.4 and Figure 4.10. Figure 4.5 show a normal diurnal flow profile for DMA02 but the pressure profile in Figure 4.6 do not reflect the normal conditions of DMA02 .This is because the DMA02 pressure profile is affected by the upstream pressure profile (DMA01) in Figure 4.4.

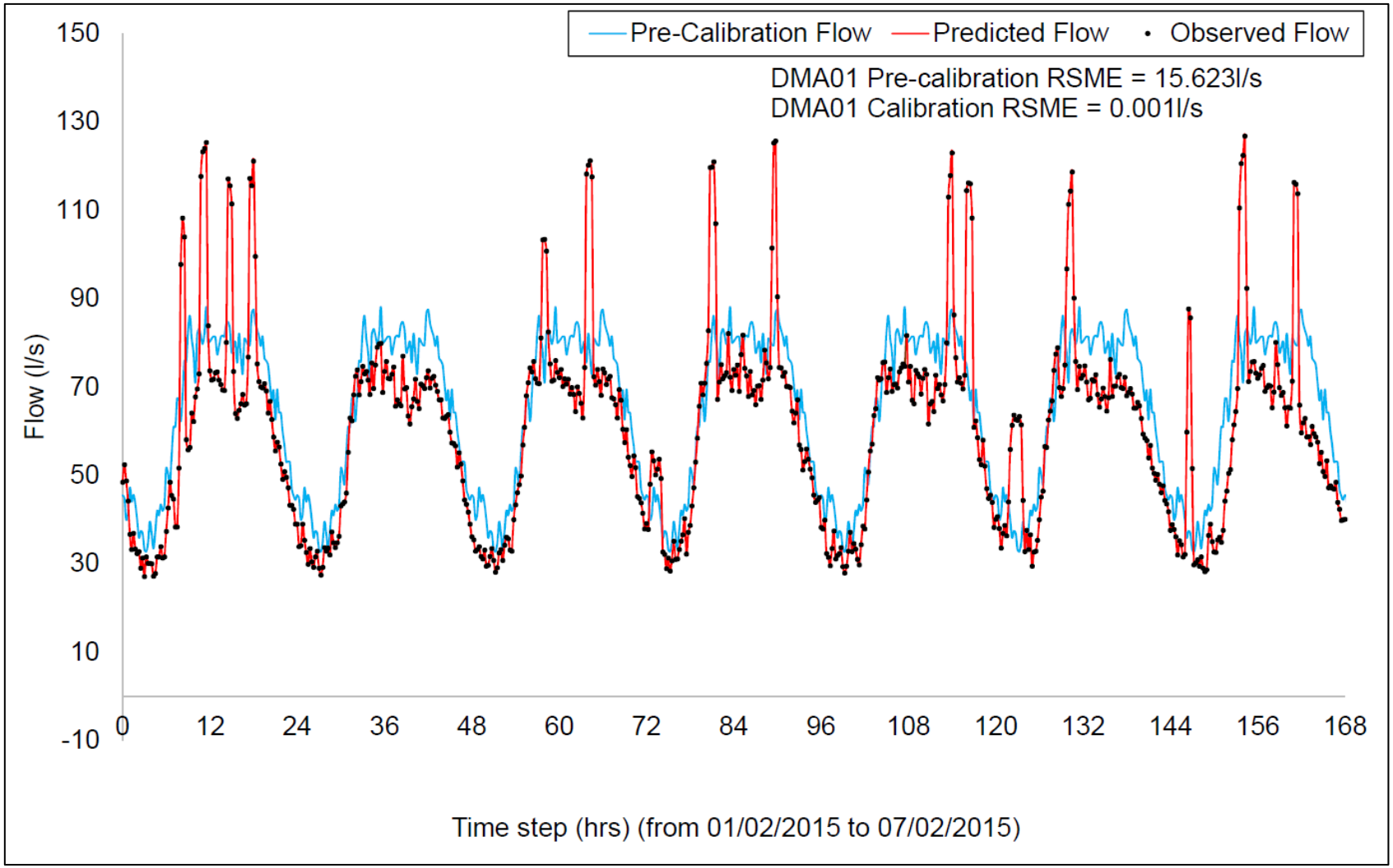


Figure 4.3: Calibration plot of DMA01 hydraulic model for flow between 1st and 7th February 2015

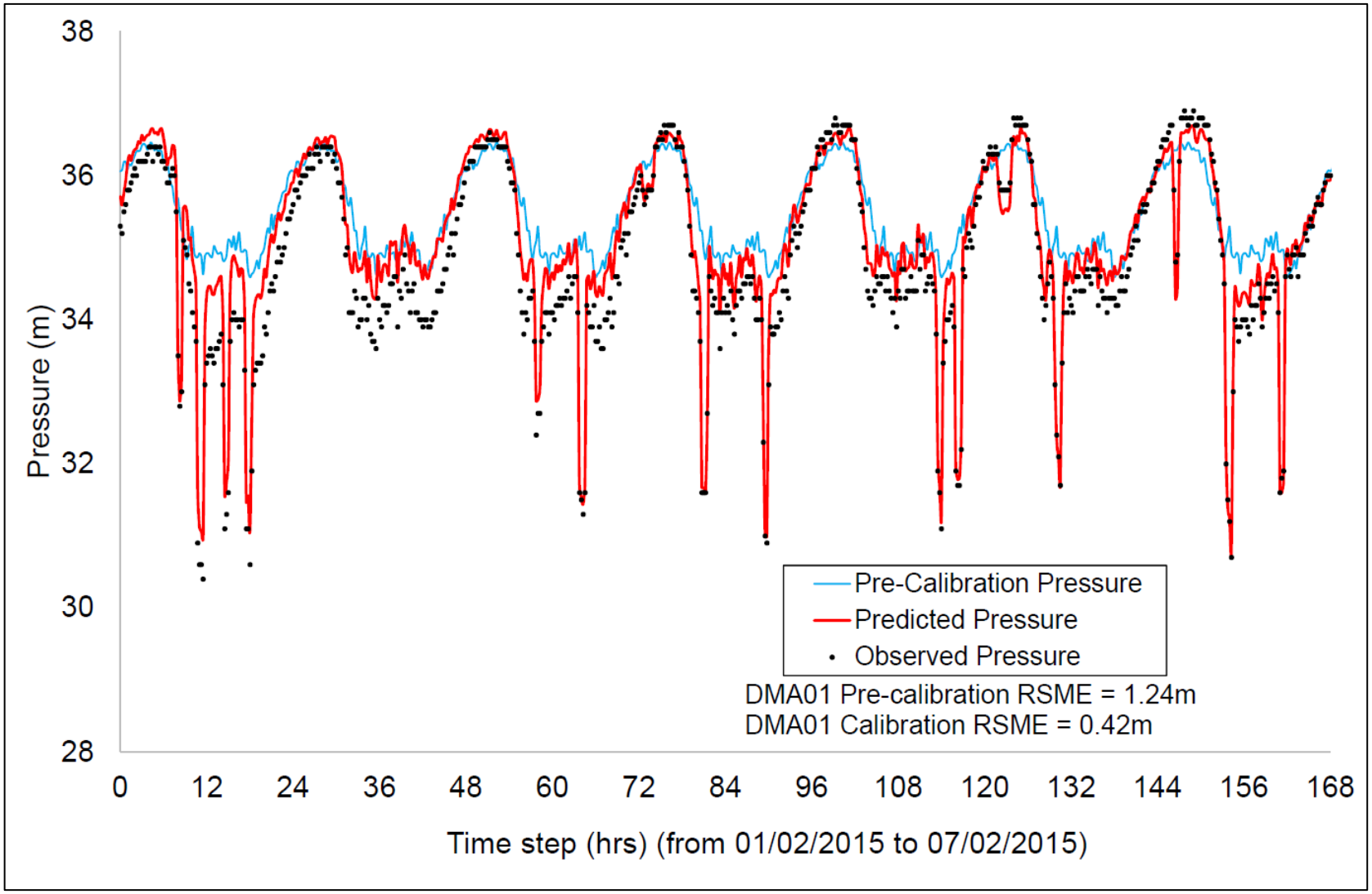


Figure 4.4: Calibration plot of DMA01 hydraulic model for pressure between 1st and 7th February 2015

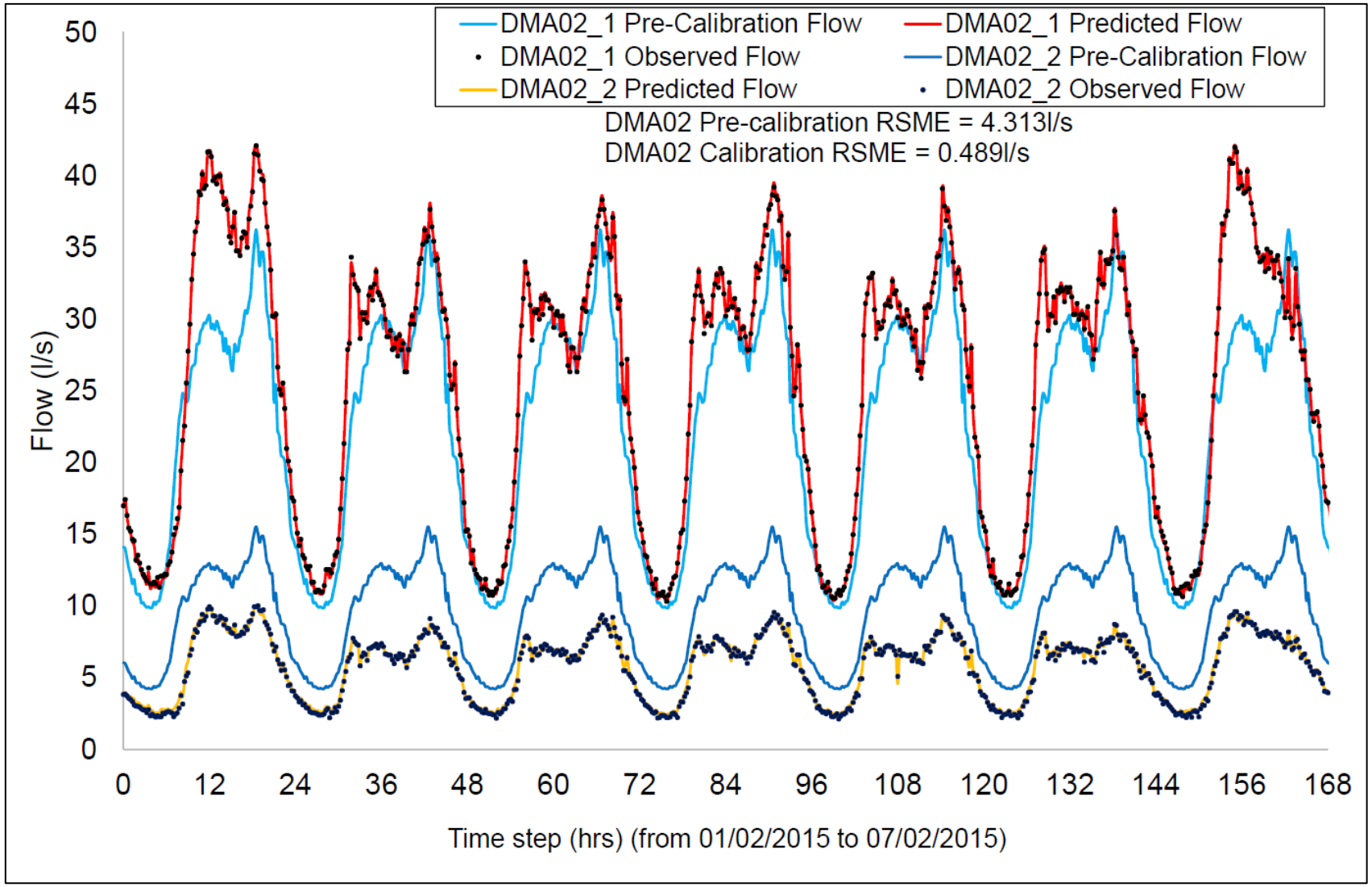


Figure 4.5: Calibration plot of DMA02 hydraulic model for flow between 1st and 7th February 2015

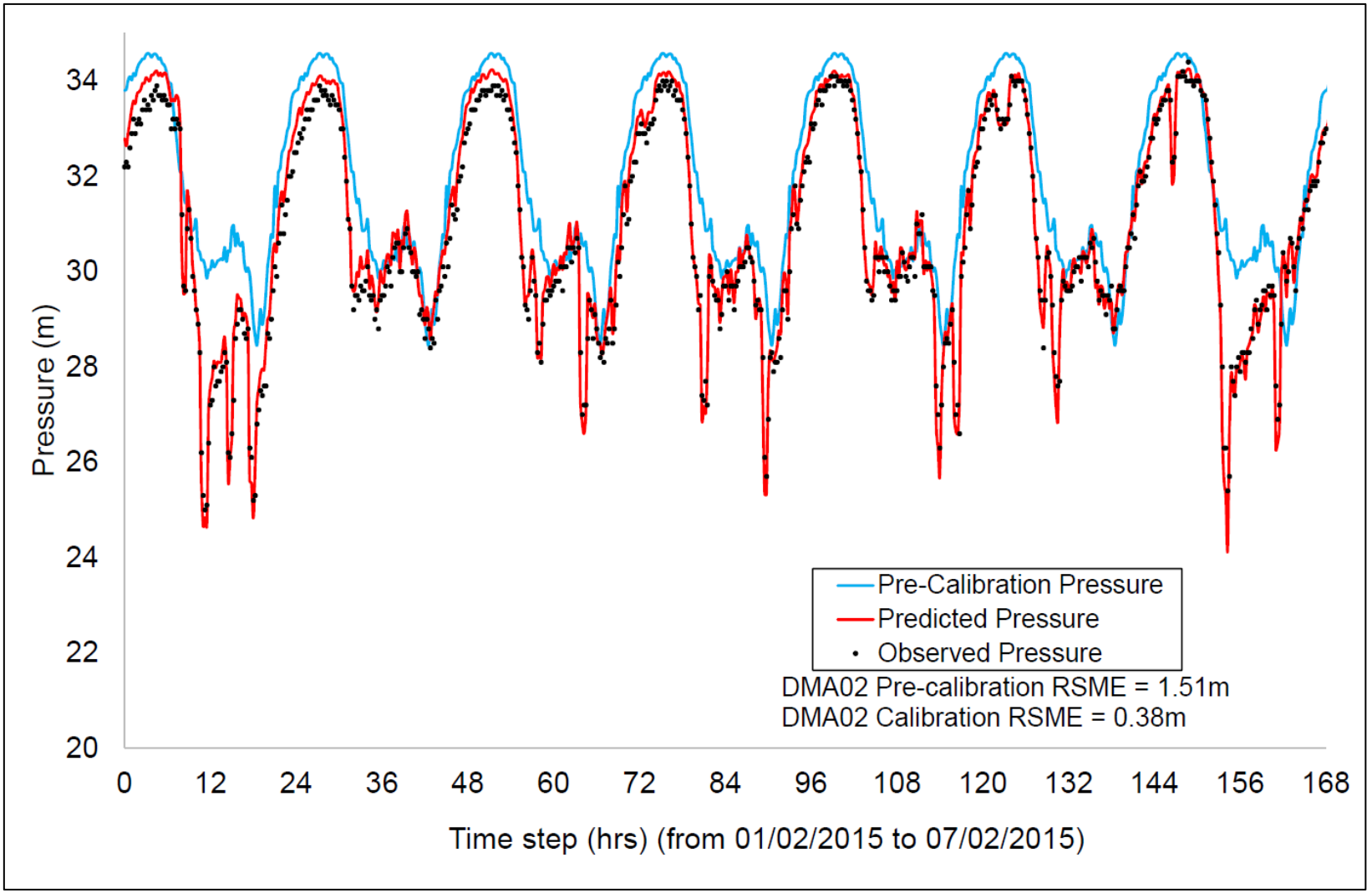


Figure 4.6: Calibration plot of DMA02 hydraulic model for pressure between 1st and 7th February 2015

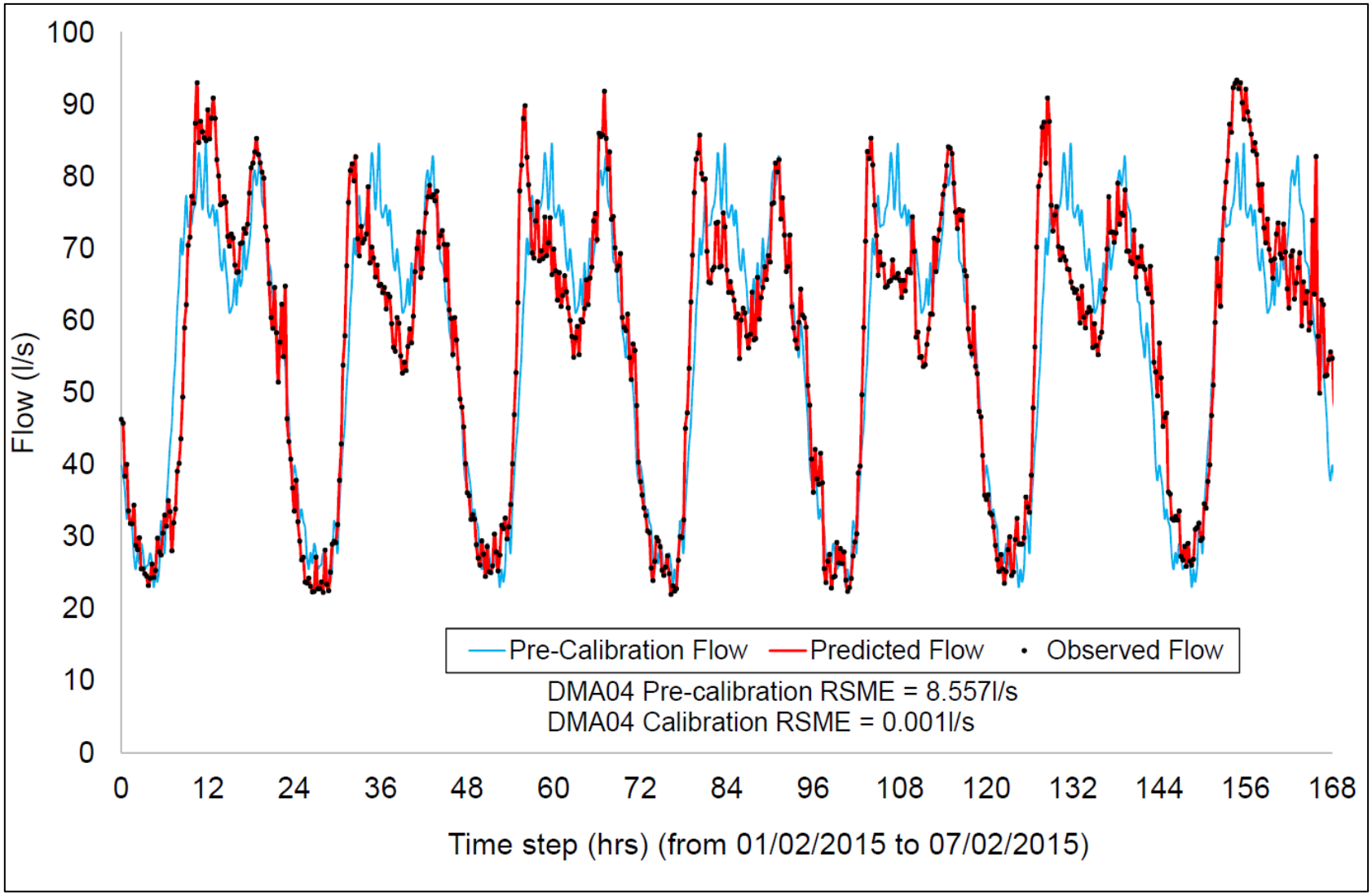


Figure 4.7: Calibration plot of DMA04 hydraulic model for flow between 1st and 7th February 2015

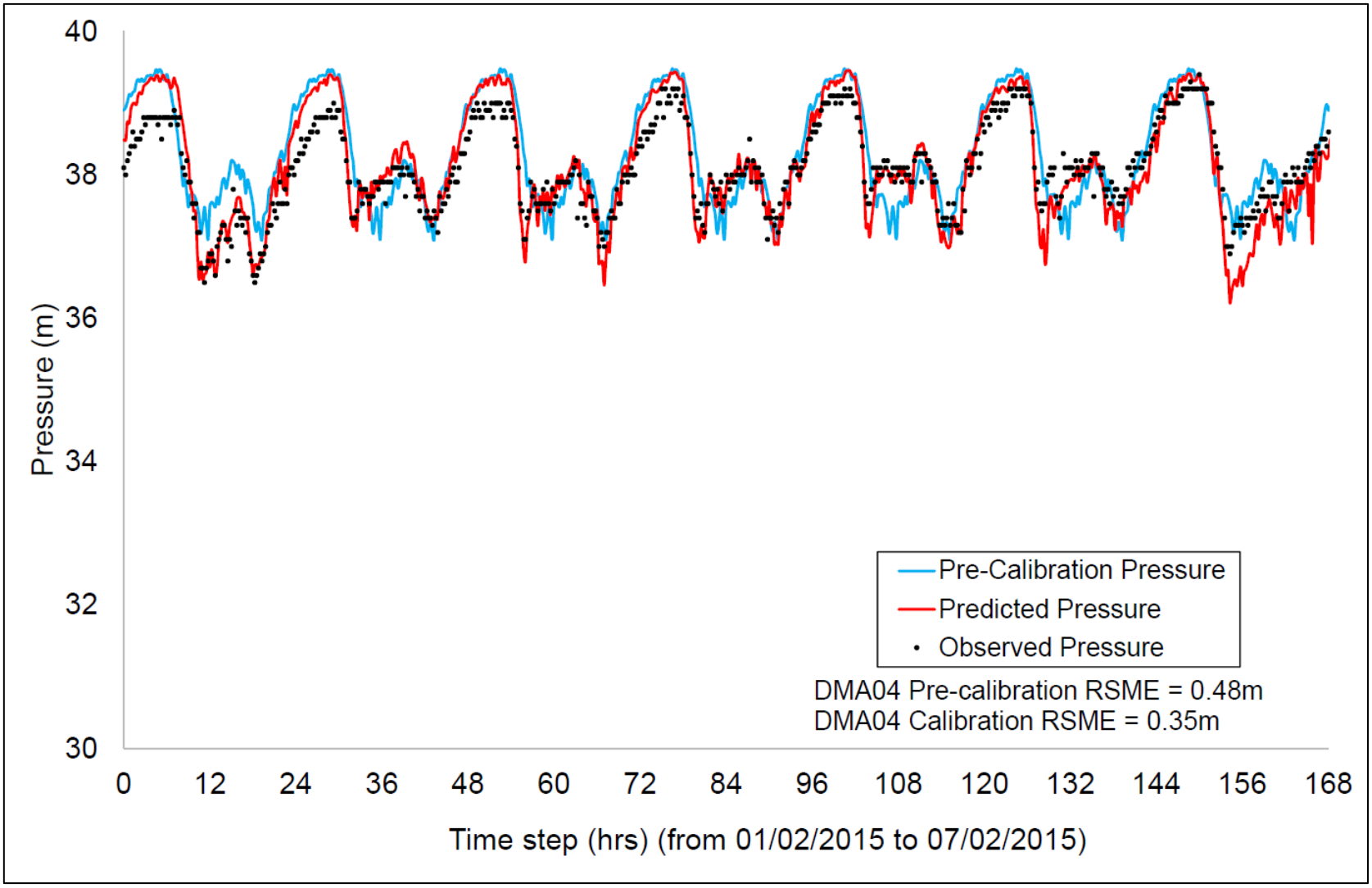


Figure 4.8: Calibration plot of DMA04 hydraulic model for *pressure* between 1st and 7th February 2015

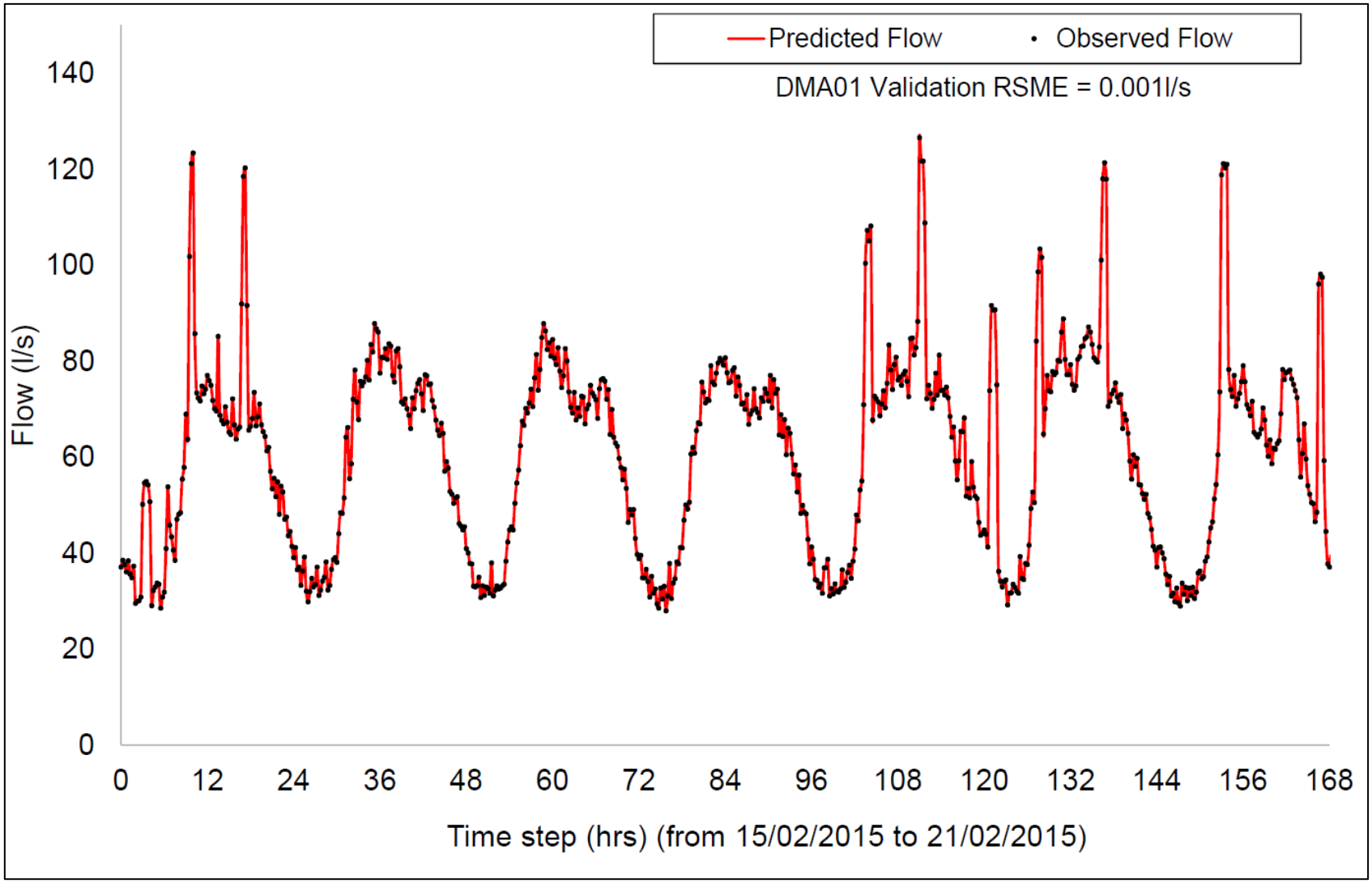


Figure 4.9: Validation plot of DMA01 hydraulic model for flow between 15th and 21st February 2015

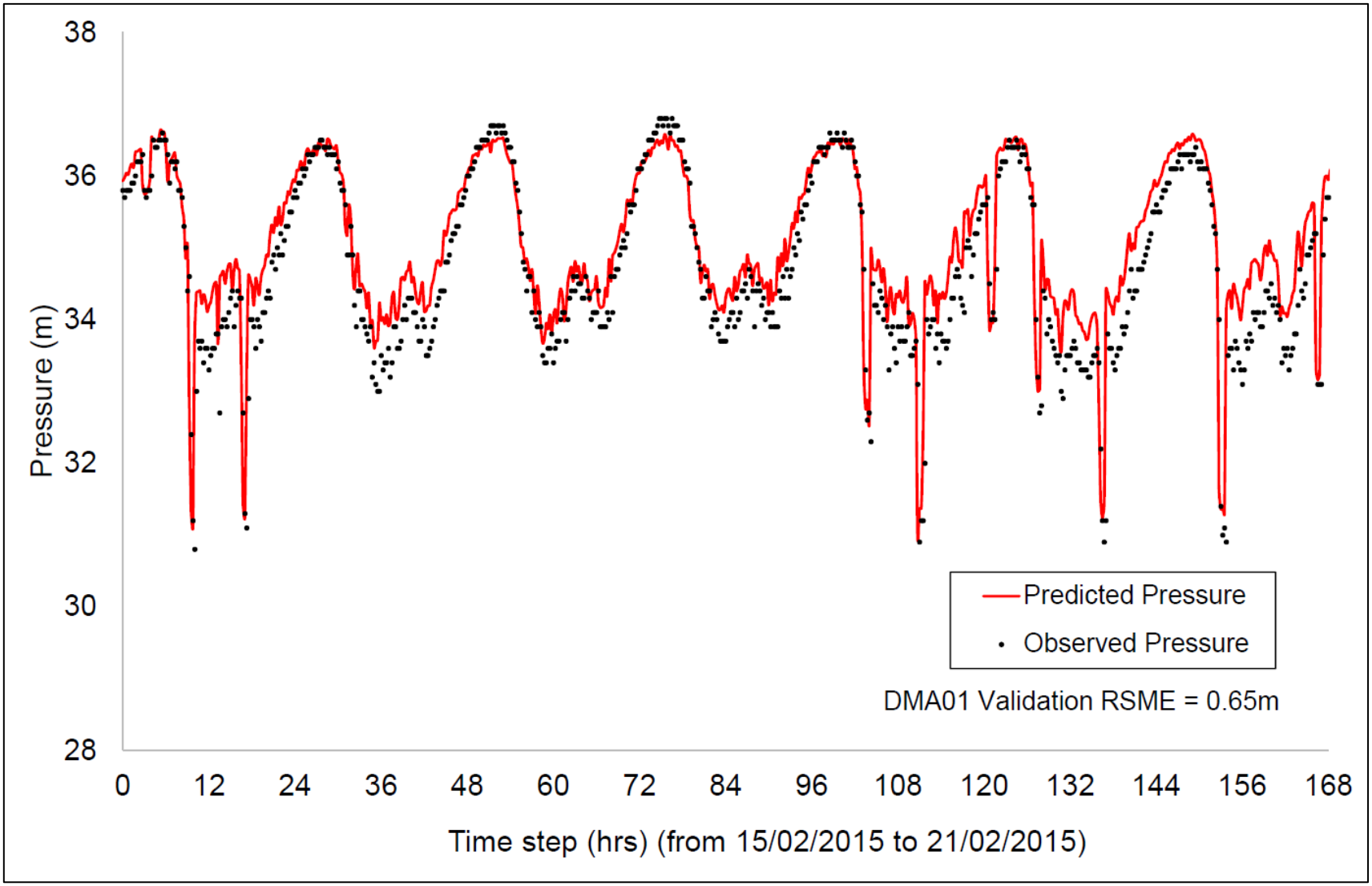


Figure 4.10: Validation plot of DMA01 hydraulic model for pressure between 15th and 21st February 2015

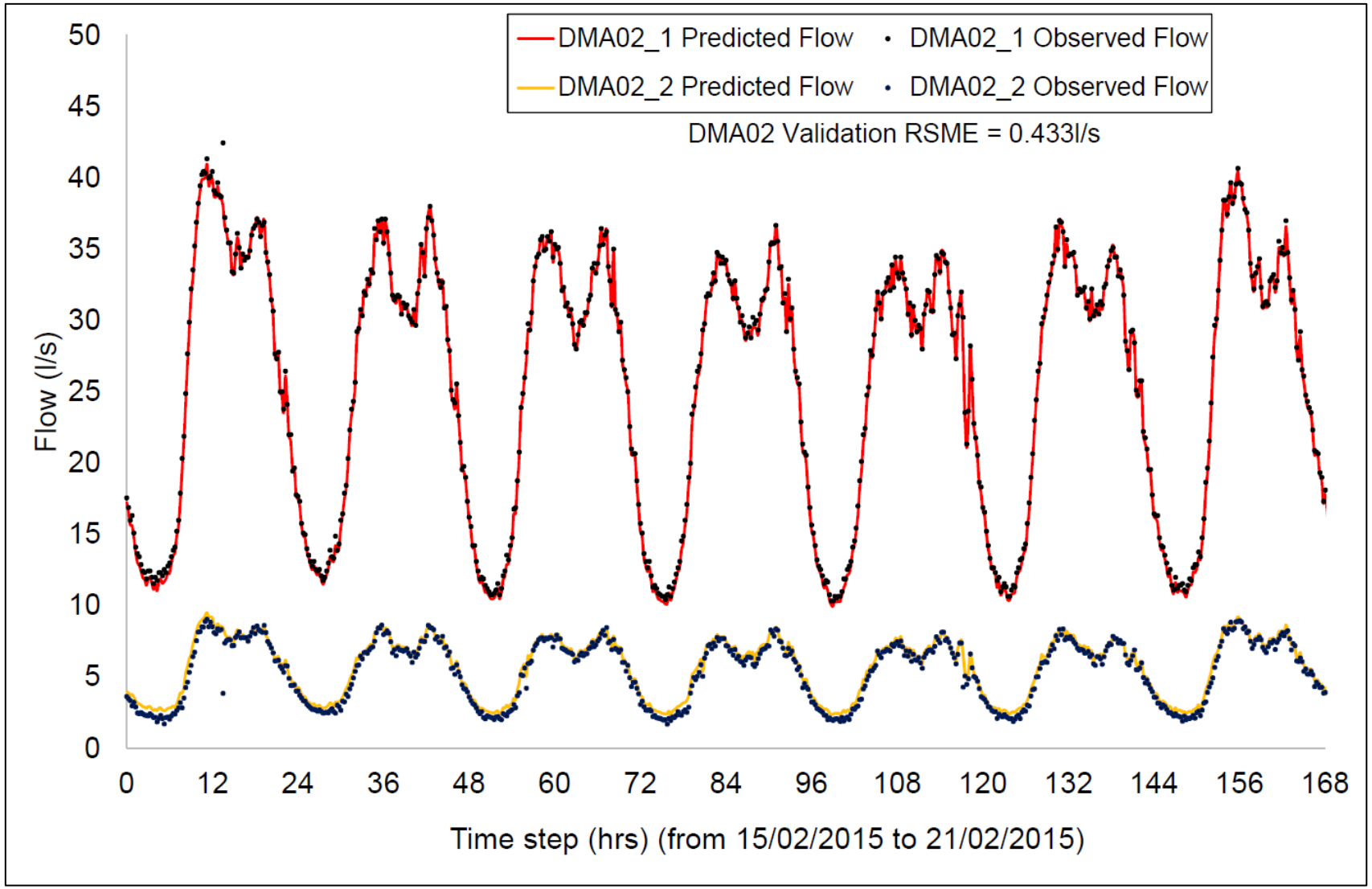


Figure 4.11: Validation plot of DMA02 hydraulic model for flow between 15th and 21st February 2015

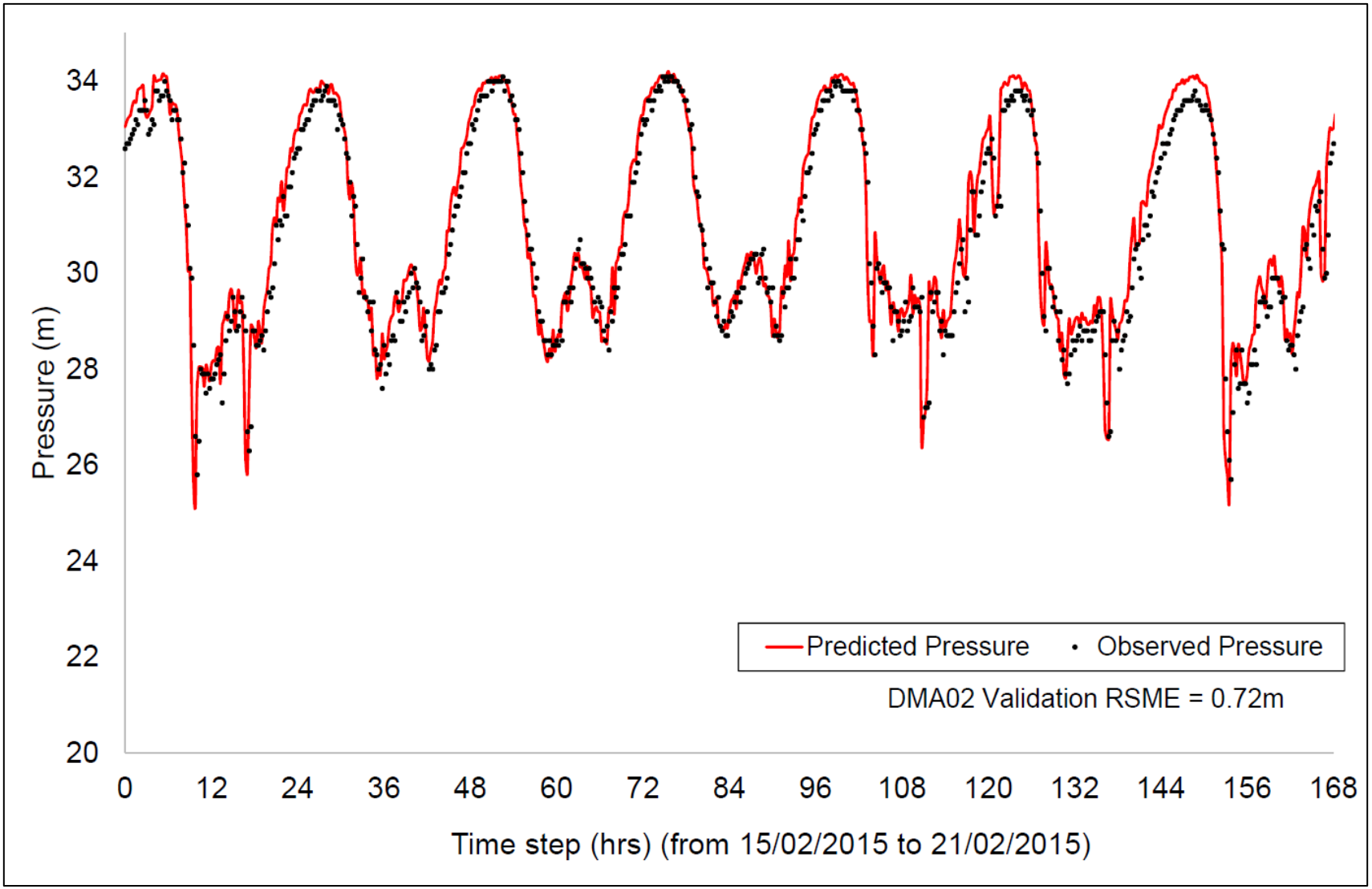


Figure 4.12: Validation plot of DMA02 hydraulic model for pressure between 15th and 21st February 2015

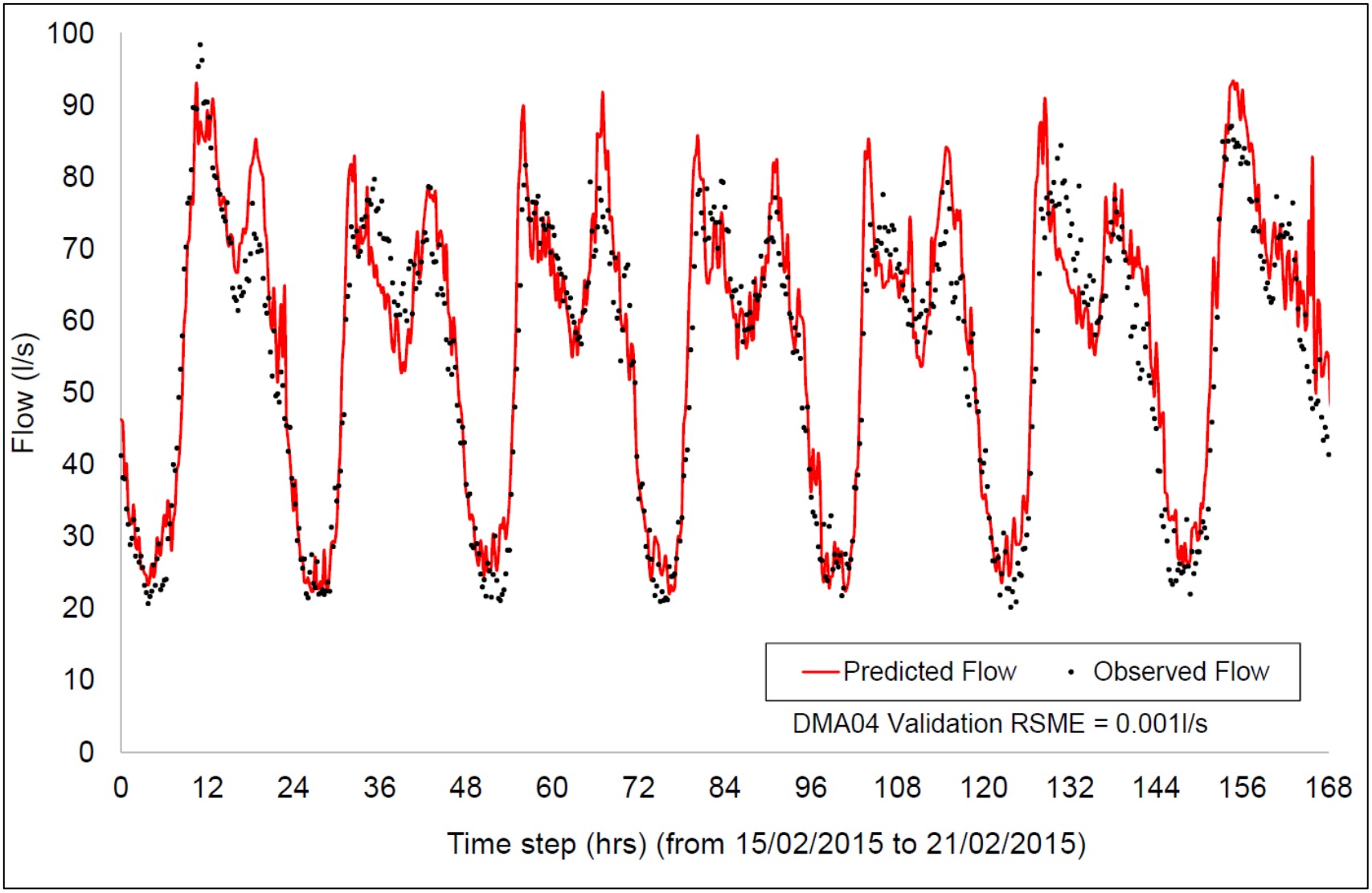


Figure 4.13: Validation plot of DMA04 hydraulic model for flow between 15th and 21st February 2015

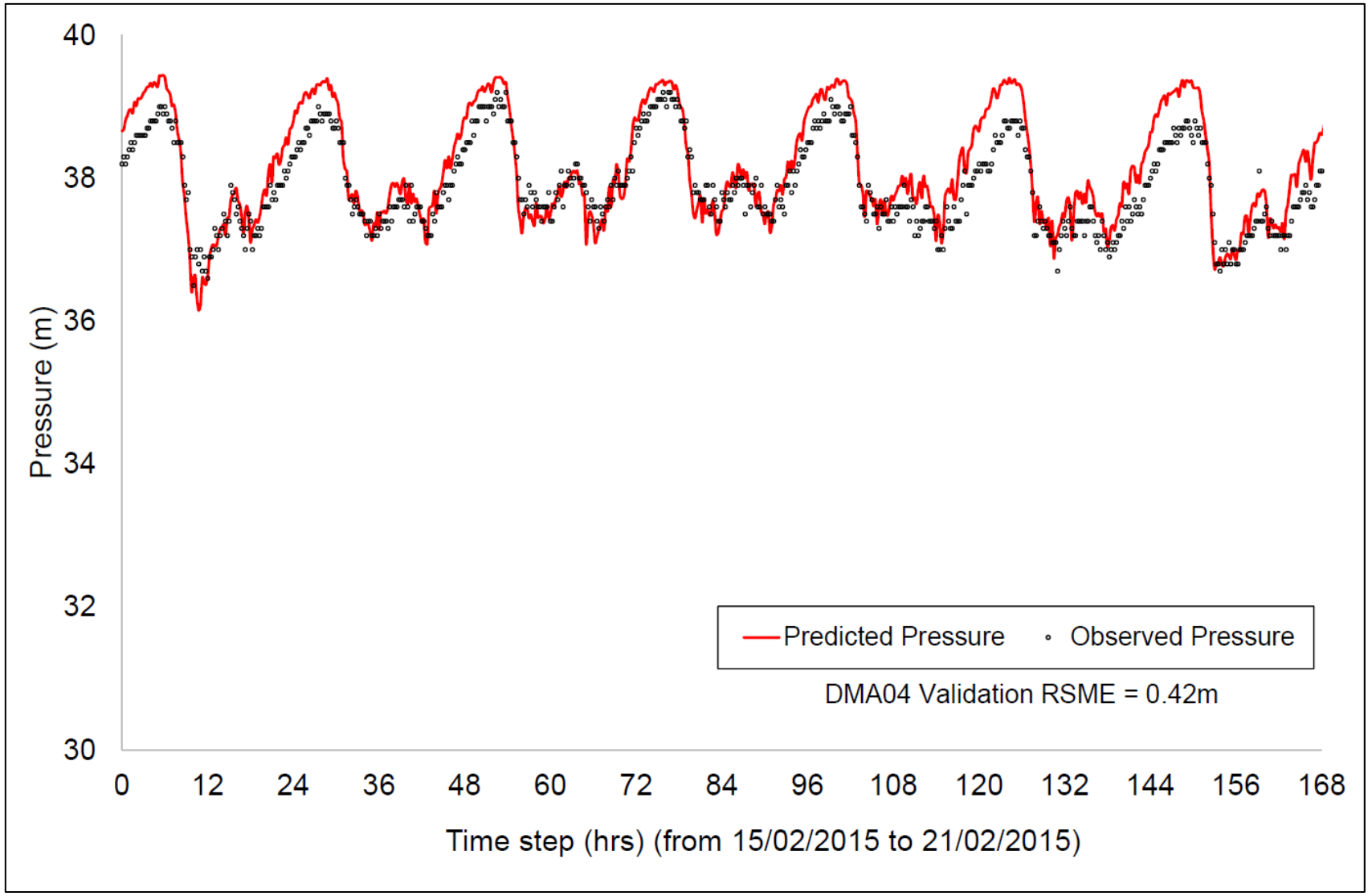


Figure 4.14: Validation plot of DMA04 hydraulic model for pressure between 15th and 21st February 2015

Due to the limited number of flow and pressure observations used for the reduced WSZ01 hydraulic model calibration, it is rational to say that the DMAs' hydraulic model is calibrated offline reasonably well.

The demand profiles between 22nd March and 4th April 2015 are used as 2 weeks' demand coefficients in the calibrated hydraulic model. These demand profiles are the most recent data prior to the offline hydraulic modelling. Water Supply Production Planning Managers at UU usually use recent flow data to derive model demand profiles and then predict the water supply area's future flow and pressure via a hydraulic model. Hence the reason for using the most recent flow data (22nd March and 4th April 2015) for offline hydraulic modelling.

The calibrated model is then run online for 4 weeks starting between 5th April and 2nd May 2015. The WDS state predictions from the offline modelling are obtained as offline hydraulic modelling data of the reduced WSZ01 model network.

The following section explains how online hydraulic model works in the case study and defines key parameters.

4.5 Online Hydraulic Model Implementation

4.5.1 Introduction

The online hydraulic models described in Chapter 3 (refer to Section 3.3) are used in this case study. Each online hydraulic model has 3 key components and they are as follows:

(1) **Water Demand Forecasting model** which forecasts water demands using historic water demands.

(2) **Hydraulic model** (i.e., EPANET 2.0) which models nonlinear relationships described in Section 3.3.3 and generate WDS predictions for the next observational time step using demand forecasts (from step 1).

(3) **Data Assimilation method** which corrects demand forecasts using the differences between WDS predictions and the corresponding WDS observations.

The corrected water demands from the DA method are inserted into hydraulic model to obtain a set of new WDS predictions. These new WDS predictions is referred as WDS corrections. The three components are used at each subsequent time step.

The three selected DA methods, WSZ01 operating conditions and WDFMs are written on Microsoft Visual C++ from scratch since there were no suitable pre-written C++ DA/WDFM templates. The Microsoft Visual C++ is the used to run the offline/online hydraulic models along with the EPANET tool on a HP laptop (with Intel core i5 processor @ 2.30GHz and 6.0 Gb RAM memory).

The next section outlines the model assumptions made for the online hydraulic modelling of reduced WSZ01 network.

4.5.2 Online Hydraulic Model Assumptions

The following assumptions are made in this case study are as follows:

- The EPANET 2.0 offline model is assumed to be sufficiently calibrated. Hence the physical properties such as pipes' roughness, diameter and length are assumed to be correct. They remain constant during the online hydraulic modelling.
- The WDS observations used in this case study are assumed to be accurate with the manufacturer's sensor accuracy.
- The main sources of uncertainties considered are the WDS observations error and water demands error due to model simplification of the WDS realities.
- The EPANET 2.0 offline model configuration is assumed to perfectly represent the real-life WDS topology. Hence other uncertainties such as valve status and inaccurate pipe connectivity are not considered.
- The reduced WSZ01 network is assumed to have no background leakage.

The above assumptions are made in recognition of limited number of hydraulic meters and hydraulic model's limitations.

The next section explains how the MLR-based WDFMs are developed for the case study.

4.5.3 Water Demand Forecasting Model

For each DMA's flow measurements, the methodological steps in section 3.3.2 are used to develop a MLR-based WDFM. The aim of the WDFM is to forecast water demands every 15 minutes with a 15 minute lead time. The forecast water demand is the function of the historical water demands, i.e. demands from the previous time steps. The historic demands are estimated using the DMA boundary flow meters (i.e. historic DMA inflows and outflows). The difference between the inflows and outflows of each DMA at intervals of 15 minutes between 1st February and 7th March 2015 (5 weeks) are assumed to be the water demands for each DMA (given there are no water tanks in any of the five analysed DMAs).

The time series water demand data are used to find several suitable WDFM independent variables for each DMA and large industrial user. In this case study, independent variables are limited to water demand from the previous 15mins ago (*t-1*) until demand from two weeks ago (*t-1344*). This is to limit the time taken to generate the WDFMs.

The number of WDFM independent variables is limited to top six demand points that have strong correlation with the dependent variable after a preliminary test. The preliminary test was performed on year 2013 historic flow measurements (not shown here). The results from the preliminary test shows there is no further benefit if there are more than six independent variables. Hence, the WDFM structure is limited to a maximum of six independent variables. The year 2013 preliminary results also show WDFMs with less than three independent variables do not perform well compared to WDFM with more than three independent variables.

Therefore, six independent variables are chosen based on their correlation coefficient value (that is close to 1) in this case study. The identified independent variables are then used to form several WDFMs for each DMA. The MLE is then used to derive the MLR-based WDFMs' weight coefficients. The WDFMs are tested on two weeks' water demand data between 8th and 21st March 2015 (2 weeks). The best performing MLR-based WDFM for each DMA demand profiles is chosen based on the lowest Root Mean Squared Error (RSME) value. Table 4.4 shows the optimum WDFM based on MLR-analysis for each DMA demand and industrial demand.

Table 4.4: The optimal WDFM based on multi-linear regression analysis for each DMA and industrial demand

Water user	**WDFM**
DMA01 domestic user	$d_{t+1} = 0.466\ d_t + 0.173d_{t-95} + 0.129d_{t-96} + 0.072d_{t-96} + 0.157d_{t-671}$
DMA01 industrial user 1	$d_{t+1} = 0.787d_t + 0.186d_{t-95} + 0.136d_{t-96} - 0.194d_{t-96} + 0.081d_{t-671}$
DMA01 industrial user 2	$d_{t+1} = 0.381\ d_t + 0.247d_{t-670} + 0.165d_{t-671} + 0.082d_{t-672} + 0.140d_{t-1343}$
DMA01 industrial user 3	$d_{t+1} = 0.525\ d_t + 0.159d_{t-96} + 0.114d_{t-671} + 0.066d_{t-672} + 0.106d_{t-1343}$
DMA02 domestic user	$d_{t+1} = 0.708\ d_t + 0.254d_{t-670} - 0.061d_{t-671} + 0.100d_{t-1343}$
DMA04 domestic user	$d_{t+1} = 0.721\ d_t + 0.133d_{t-670} + 0.153d_{t-1343}$
DMA06 domestic user	$d_{t+1} = 0.838\ d_t + 0.006d_{t-96}$
DMA07 domestic user	$d_{t+1} = 0.722\ d_t + 0.153d_{t-671} + 0.123d_{t-1343}$

where d_{t+1} is the forecast demand at the next time step, *t*; d_{t-94} is the corrected demand from 23 hours and 45 minutes ago, d_{t-95} is the corrected demand from a day ago; d_{t-96} is the corrected demand from a day and 15 minutes ago; d_{t-670} is the corrected demand from 167 hours and 45 minutes ago; d_{t-671} is the corrected demand from a week ago; d_{t-672} is the corrected demand from a week and 15 minutes ago and d_{t-1343} is the corrected demand from two weeks ago.

Table 4.4 also shows the four additional WDFMs for industrial demands in DMA01 and DMA05. The DMA01's industrial users 1, 2 and 3 are the Airport and two

pharmaceutical companies, respectively. The DMA05 industrial user is the wastewater treatment works.

The decision to use more than one WDFM for a large highly industrial DMA such as DMA01 and DMA05 was based on the results of WDFMs trail run (not included in this thesis). The WDFM trail runs were performed on 2013 water demand data. The WDFM trail runs involved comparing the performance of one MLR–based WDFM for a highly industrial DMA to multiple MLR-based WDFMs. The multiple WDFMs include one WDFM for the domestic demand and other WDFM(s) for industrial demand(s). It was found that multiple WDFMs outperformed one MLR-based WDFM for a highly industrial DMA.

Even though DMA03, DMA05 and DMA08 model are not part of the reduced WSZ01 model, it is still important to develop their respective WDFM to ensure the upstream flow rates can reflect the reality. Ignoring DMA03, DMA05 and DMA08 demands can consequently affect all the DMAs' pressures and tank level.

In the case study, the historic demand data between 22nd March and 4th April 2015 (2 weeks) are used as the initial estimates of the past demands required for the WDFM. Hence, the online hydraulic modelling of the reduced WSZ01 model started running on the 5th of April 2015. The derived MLR-based WDFMs in Table 4.4 are part of the online hydraulic models.

The forecasted demands from previous time step are used to drive into the EPANET model forward to predict the WDS states. The model predictions are then corrected at the next observational time step via one of the DA methods (refer to section 3.4 to 3.6). The DA methods help in correcting both WDFM parameters and EPANET model states. The next section defines the value for the online hydraulic model #1 parameters.

Prior to the online hydraulic modelling of the reduced WSZ01, each online hydraulic model requires a priori knowledge of the WDS via its historic WDS observed data. Hence, the MLR-based WDFM require historic flow data to develop the WDFM parameters. Whilst each DA scheme requires some level of a prior knowledge of error

statistics to specify prior and observation error covariance. Each WDS state or WDFM parameter error variances account for the model state uncertainty. The observation error variances account for the observation uncertainty. These error variances are pertinent to ensure the model state estimates are corrected consistently and realistically with the scales of the WDS hydraulic model.

4.5.4 Online Hydraulic Model #1: Kalman Filter Method

One of the key elements of an online hydraulic model is the correction step which improves the prediction of the model states. Therefore, the involved parameters in the Kalman gain have to be defined. Specification of the prior error covariance, P_t and observation error covariance, R_t require prediction percent error, $\bar{r}$ and observation percent error $\bar{q}$ to be prescribed respectively (refer to Section 3.1). In this case study, both $\bar{r}$ and $\bar{q}$ values are moving average percent error with a window size of 672 time steps (1 week). The value of $\bar{r}$ and $\bar{q}$ are obtained by online hydraulic model run (iterative KF method) of reduced WSZ01 between 22nd March and 4th April 2015 (2 weeks). The $\bar{p}$*–value* is the average relative error between the flow predictions and the corresponding observations between 29th March and 4th April 2015 (1 week). Whilst the $\bar{r}$*-value* is the average relative error between the flow corrections and the corresponding observations 29th March and 4th April 2015 (1 week). Table 4.5 presents the values of $\bar{r}$ and $\bar{q}$ for each flow meter including the average flow rate. The abbreviation, "IND" in Table 4.5 is abbreviated for industrial users (i.e., manufacturing companies).

Table 4.5: The initial prediction and observation average percent error for each flow meter

Flow meter index	**DMA 01**	**DMA 01_IND1**	**DMA 01_IND2**	**DMA 01_IND3**	**DMA 02_1**	**DMA 02_2**	**DMA 04**	**DMA 06**	**DMA 07**
$\bar{p}$–value (%)	5.5	9.7	35.9	35.0	9.7	8.7	4.7	3.8	4.3
$\bar{r}$–value (%)	1.6	3.0	9.4	9.3	2.3	1.0	1.3	1.0	1.1
Average flow rate (l/s)	65.01	2.27	5.89	0.97	31.20	12.59	59.68	31.01	18.74

Prior to the online hydraulic model run of reduced WSZ01 between 22nd March and 4th April 2015 (1 week), the initial value of $\bar{r}$ is assumed to be Gaussians, $N(0, \sigma_r^2)$ where σ_r^2 = 2.5% of the average flow rate. This observation variance is taken from the flow meter manufacturer datasheet (Siemen MAG flowmeters datasheet, 2010). Whilst the initial value of $\bar{p}$ is zero, assuming the state model is perfect at the initial time step.

Since the online hydraulic modelling starts running from 5th April 2015, the $\bar{p}$-value and $\bar{r}$- value in Table 4.5 are used at the initial time step. In the following time steps, $\bar{p}$-value and $\bar{r}$-value are replaced by their respective moving average of relative error with a window size of one week.

The iteration of the KF method ends when either (1) the errors between the observed and corrected flow rate are below their predefined value or (2) the computations reached a pre-defined maximum iteration number. The predefined value is the product of $\bar{r}$–value and observed flow rate. The pre-defined maximum iteration number is 10. This threshold 10 is chosen due to a test run of the online hydraulic model with 2013 historic data. It took an average of 6-9 iterations for a DMA to reach the above predefined value during the model run in 22nd March and 4th April 2015 (1 week).

The next section defines the parameters' value used for online hydraulic model #2 - the perturbed EnKF method.

4.5.5 Online Hydraulic Model #2: Ensemble Kalman Filter Method

The application of EnKF method is relatively easy to implement. It requires perturbation of WDFM predictions and the WDS observations. The ensemble generation of model predictions is done via Gaussians perturbation of WDFM outputs (forecast demands). The ensemble of forecast demands is perturbed by a normal distribution with mean equal to the forecast demand at each time step, and a variance equal to percent demand error of the forecast demand. The percent water demand error is the average percent error between the predicted demand and observed demand. Table 4.6 displays the percent error for each water demand identified in the reduced WSZ01 network.

Table 4.6: The percent error for each DMA demand

Flow meter index	**DMA 01**	**DMA 01_IND 1**	**DMA 01_IND2**	**DMA 01_IND3**	**DMA02_ 01**	**DMA 02_02**	**DMA 04**	**DMA 06**	**DMA 07**
$\bar{p}$–value (%)	4.3	9.7	35.9	35.0	6.3	6.9	4.0	3.5	3.9

The flow observation ensemble perturbation is drawn from a normal distribution with mean equal to the flow observation at each time step, and a variance equal to the observation average percent error ($\bar{r}$–*value* in Table 4.5) of the observed flow.

After the ensemble generation, the ensemble members are re-adjusted to ensure its ensemble mean prior to perturbation remain unchanged. The ensemble means are same as the WDS flow observations every time step.

The ensemble size of the perturbed EnKF is 40. The ensemble size of 40 is selected due to a sensitivity analysis carried out between the ensemble sizes and model state correction accuracy. The sensitivity analysis involves running the online hydraulic model #2 run between 22nd March and 4th April 2015 at different ensemble sizes from 10 to 100 (at every 5 interval). Then the mode predictions accuracy against ensemble size were reviewed. The sensitivity result (not shown in here) shows the mode predictions accuracy improvement is minor when the ensemble size is above 35.

The next section defines the parameters' value used for the PF-SUR method.

4.5.6 Online Hydraulic Model #3: Particle Filter Method

Prior to the application of the online hydraulic model #3 (PF-SUR), the model parameters' value has to be defined. The number of particles for each model states is equivalent to the ensemble size of the online hydraulic model #2 (perturbed EnKF method) which is 40. This is to compare online hydraulic model #3 against online hydraulic model #2 on the same ensemble size. The perturbation method employed by the online hydraulic model #2 (refer to section 4.5.4) is used to generate the

ensemble of forecast demands and flow observations. Hence, the variance of the prior and observation error is similar as the variance employed in the EnKF method.

The normal distribution (refer to equation (32) in section 3.6) is used to generate the particle weights. The standard deviation for each model state in the normal distribution equation is calculated by using the ensemble WDS state predictions and corresponding observations.

When the particle weights are normalized, each particle weight reveal the importance of their respective particle sample (model state). The normalized particle weight measure from 0 to 1. The normalized particle weights that are close to 1 mean their respective particle sample (predictions) are close to the observed data. Whilst the normalized particle weights that are close to 0 is considered less important. Hence, the SUR method (refer to Section 3.4) is used to resample particle samples. Resampling process involve replacing less important particle samples measured by the SUR method with particle samples that have a greater particle weight. This resampling process is applied at each observational time step to ensure the PF method do not degenerate. After the resampling process, a new set of particle weights is calculated and normalised. Then equation (33) in Section 3.6 is used to obtain the corrected model states.

The model states in this chapter are the forecast water demands and tank levels. Once they are predicted or corrected, the hydraulic model is then run to obtain the WDS pressure and flow values. The forecast water demands and tank levels are corrected at the same time step to prevent filter degeneracy. This part of the correction process is also applied in both online hydraulic model #1 - iterative KF method and #2 - perturbed EnKF method. The aim of this process is to reduce the errors in the initial states at every time step.

The next section discusses the performance of offline and online hydraulic models.

4.6 RESULT AND DISCUSSIONS

The proposed offline and online hydraulic models shown above are applied to the reduced WSZ01 network. The main objective of analyses here is to compare the performance of the offline and online hydraulic model in the case study described in Section 4.2.

Table 4.7 presents a comparison of the WSZ01 tank level prediction statistics. As it can be seen from Table 4.7, the online hydraulic models outperform the offline hydraulic model based on two performance metrics used (Mean Absolute Error and the Coefficient of Determination). Among the three online hydraulic models, the perturbed EnKF method performs better than the other two online hydraulic models. Table 4.7 further shows the iterative KF method is the worst performing online hydraulic model in the term of correcting WDS states (tank levels). However, based on the Coefficient of Determination value, R^2 in Table 4.7, the iterative KF method still performs reasonably well, i.e. the differences in performances of three online hydraulic models are not that large.

Table 4.7: A comparison of the WSZ01 tank level prediction/correction statistics where MAE is Mean Absolute Error between predicted and observed; $R^2$is the Coefficient of Determination between predicted and corresponding observed values

Hydraulic Model	**Predicted Tank level**		**Corrected Tank level**	
	MAE (m)	R^2	MAE (m)	R^2
Offline Model	0.598	0.005	-	-
Iterative KF model	0.193	0.824	0.190	0.917
Perturbed EnKF Model	0.087	0.972	0.018	0.991
PF-SUR Model	0.118	0.862	0.108	0.935

Figure 4.15 shows the plot of observed and predicted tank level over the period of the first two weeks of online hydraulic modelling (between 5th and 18th April 2015).

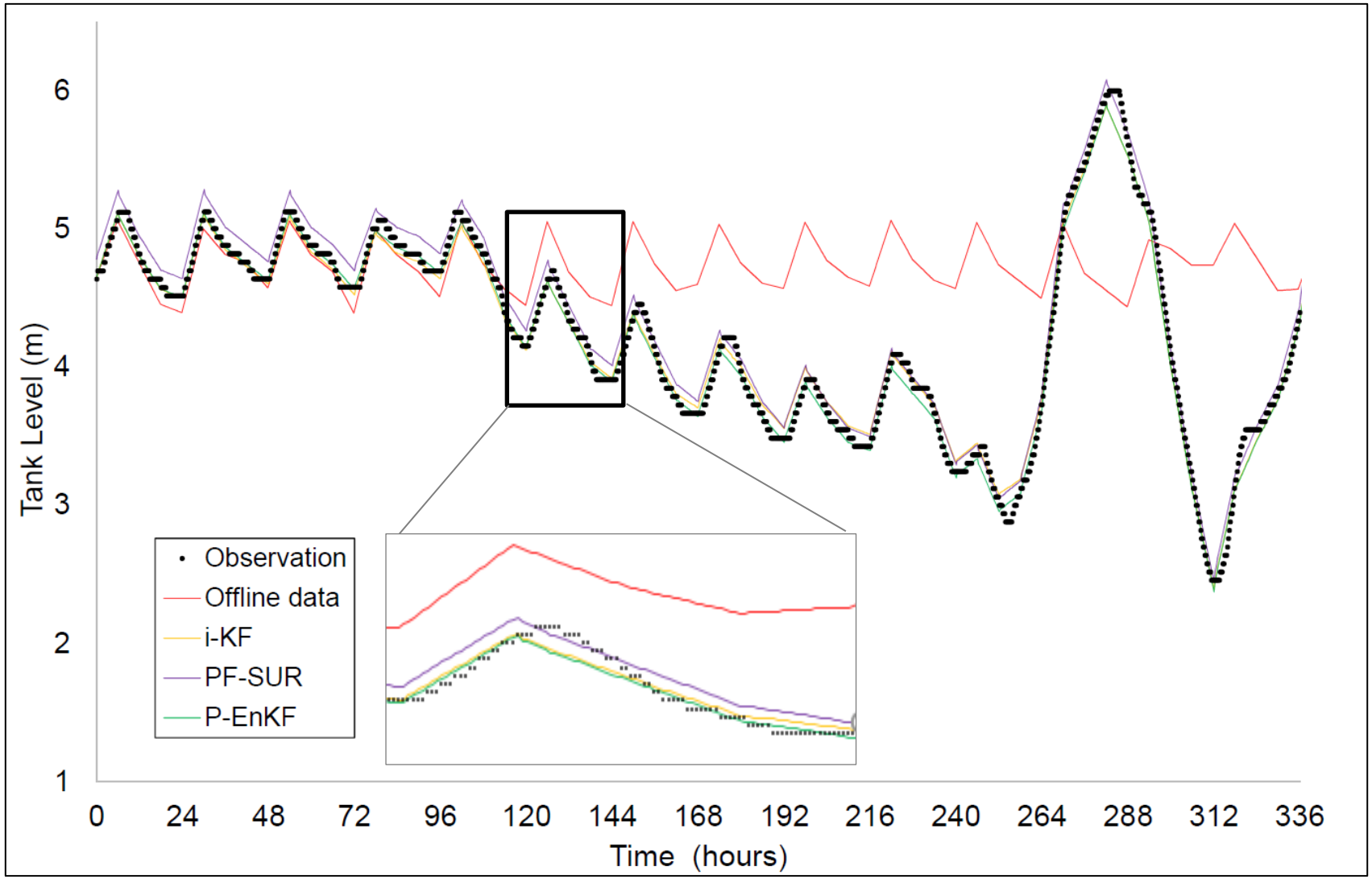


Figure 4.15: A Comparison between observed and predicted tank level at WSZ01 storage tank, node 00172CCC_ every 15 minutes

As it can be seen from Figure 4.15, the tank level predictions of offline hydraulic model do not match the observed tank levels. Hence the offline hydraulic model does not match the WDS realities. Whilst the tank level predictions obtained from the online hydraulic models match the tank level observation. This is because the online hydraulic models correct the predicted WDS state estimates from the previous time step when the WDS observations become available. The offline calibrated model is not performing as well as the three online hydraulic models because it was not able to be adapted to the changing demand in the reduced WSZ01 model.

The tank level in Figure 4.15 shows a sudden increase from 263:00 hour until 281:00 hour until the storage tank is full at 6.033 metres. The tank was deliberately filled up to support another Water Supply Zones (WSZs) demands during an emergency pipeline repair and flushing events was in that WSZs. This explained why there was a sudden decrease in tank level between 287:00 hour and 313:00 hour. The tank level increased again from 313:00 hour because the tank level crossed the minimum threshold level (2.5 metres) which opened the storage tank inlet valve. By opening the inlet valve, the tank was refilled until it reached 6.033m metres level.

Table 4.8 shows a comparison of predicted and corrected water demand for 5 selected DMAs. The comparison of the demand prediction statistics of each hydraulic modelling.

Table 4.8: A comparison of predicted and correct water demand for selected DMAs (R^2 = Coefficient of Determination)

	Method	**DMA01**	**DMA02**	**DMA04**	**DMA06**	**DMA07**
Predicted	**Offline**	0.327	0.902	0.016	0.858	0.902
	Iterative KF	0.595	0.948	0.014	0.840	0.979
	Perturbed EnKF	0.601	0.954	0.014	0.821	0.934
	PF-SUR	0.736	0.947	0.016	0.842	0.923
Corrected	**Iterative KF**	0.635	0.982	0.981	0.981	0.982
	Perturbed EnKF	0.648	0.990	0.938	0.990	0.990
	PF-SUR	0.982	0.985	0.985	0.983	0.985

Table 4.8 demonstrates that three online hydraulic models have lower MAE and higher R^2 values when compared to offline hydraulic model. This is due to the DA methods correcting (i.e. updating) the WDFM parameters at each time step which are used to re-run the online hydraulic models to get the current WDS states. The corrected states of the reduced WSZ01 network are then used as initial conditions at the next time step. As it can be seen from Table 4.8, the perturbed EnKF method generally performs better than the iterative KF and PF-SUR methods.

The Coefficient of Determination values of predicted DMA01 and DMA04 water demand for online hydraulic modelling is significant higher than the offline modelling. This is because both DMA01 and DMA04 have a higher percentage of industrial users which cannot be easily represented by the offline model. The industrial users' daily demand profile varies every week depending on their industrial activities. Hence, the fixed industrial users' demand profile in the offline calibrated hydraulic model cannot reflect the changing demand profile of industrial users.

Table 4.8 shows the offline hydraulic model prediction of water demand seems to perform better than the three online hydraulic models in DMA06 for the predicted values. This is because the DMA06 water demand data variability is the lowest among the DMAs' water demand data. However, the three online hydraulic models improved the demand predictions in DMA06 hence the higher R^2 values for corrected values compared to predicted values.

Table 4.9 illustrates the values of MAE and R^2 for each boundary flow prediction. The result in Table 4.9 shows the flow meter located at upstream DMAs (e.g., DMA01 and DMA04) tend to have higher MAE value (or lower R^2 value) compared to downstream DMAs. This is due to the accumulation of downstream DMAs' predicted water demand errors.

Table 4.10 displays the values of MAE and R^2 for each boundary flow correction. The MAE values of upstream DMAs' corrected flow are much lower than the flow prediction in Table 4.9. The water demand accuracy improves after WDFM parameters correction via the DA scheme. Therefore, the DA scheme helps in reducing the spatial data error associated with upstream flow meters.

Table 4.9: A comparison of the DMA boundary flow prediction statistics

MAE = Mean Absolute Error; R^2 = Coefficient of Determination.

Data Statistics	Method	DMA 01	DMA 02	DMA 03	DMA 04	DMA 05	DMA 06	DMA 07	DMA 08
MAE(l/s)	**Offline**	10.137	0.608	1.041	9.436	2.756	3.941	1.843	0.966
	Iterative KF	5.680	1.192	0.642	7.554	2.669	2.201	1.254	0.630
	Perturbed EnKF	5.613	1.416	0.701	8.398	2.013	2.163	1.185	0.607
	PF-SUR	5.956	0.915	0.600	7.496	2.183	2.266	1.978	0.441
R^2	**Offline**	0.432	0.918	0.903	0.369	0.355	0.901	0.893	0.892
	Iterative KF	0.808	0.946	0.946	0.401	0.725	0.930	0.947	0.944
	Perturbed EnKF	0.831	0.953	0.951	0.395	0.665	0.936	0.952	0.950
	PF-SUR	0.657	0.986	0.948	0.390	0.630	0.901	0.944	0.943

Table 4.10: A comparison of the DMA boundary flow correction statistics

MAE = Mean Absolute Error; R^2 = Coefficient of Determination.

Data Statistics	Online Hydraulic Models	DMA 01	DMA 02	DMA 03	DMA 04	DMA 05	DMA 06	DMA 07	DMA 08
MAE(l/s)	**Iterative KF**	0.459	0.379	0.077	1.816	0.283	0.169	0.113	0.063
	Perturbed EnKF	0.141	0.349	0.036	0.427	0.032	0.055	0.034	0.016
	PF-SUR	1.319	1.090	0.060	1.280	0.526	1.079	0.798	0.407
R^2	**Iterative KF**	0.980	0.969	0.981	0.877	0.965	0.981	0.981	0.981
	Perturbed EnKF	0.982	0.969	0.982	0.970	0.982	0.982	0.982	0.982
	PF-SUR	0.982	0.970	0.982	0.982	0.977	0.981	0.981	0.981

Figure 4.16 and Figure 4.17 show the DMA04 observed and predicted flow rates and pressure between 12th and 18th April 2015 respectively.

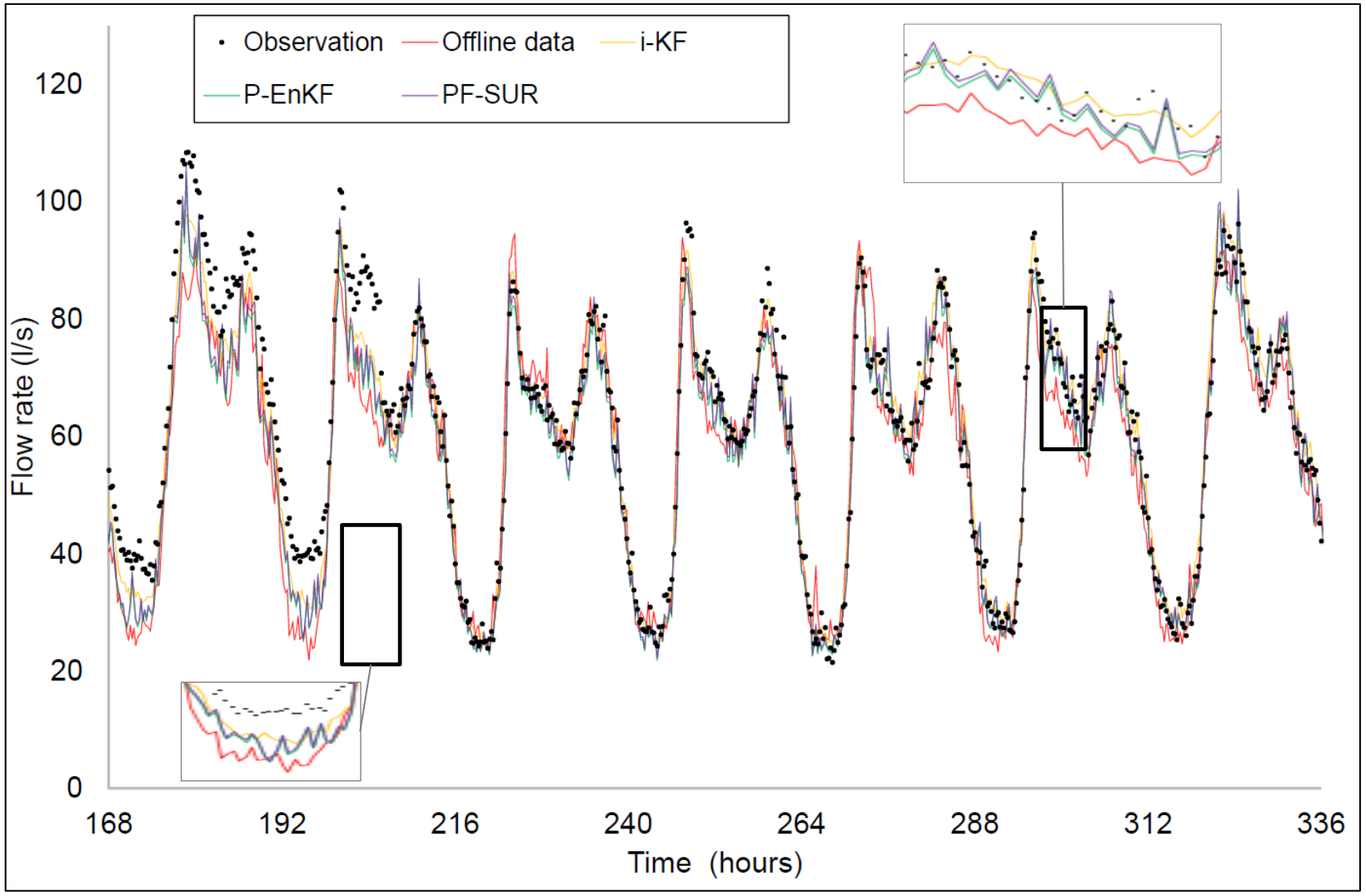


Figure 4.16: Comparison between observed and predicted flow rate at link X32230F7 (DMA04 flow meter) every 15 minutes ahead

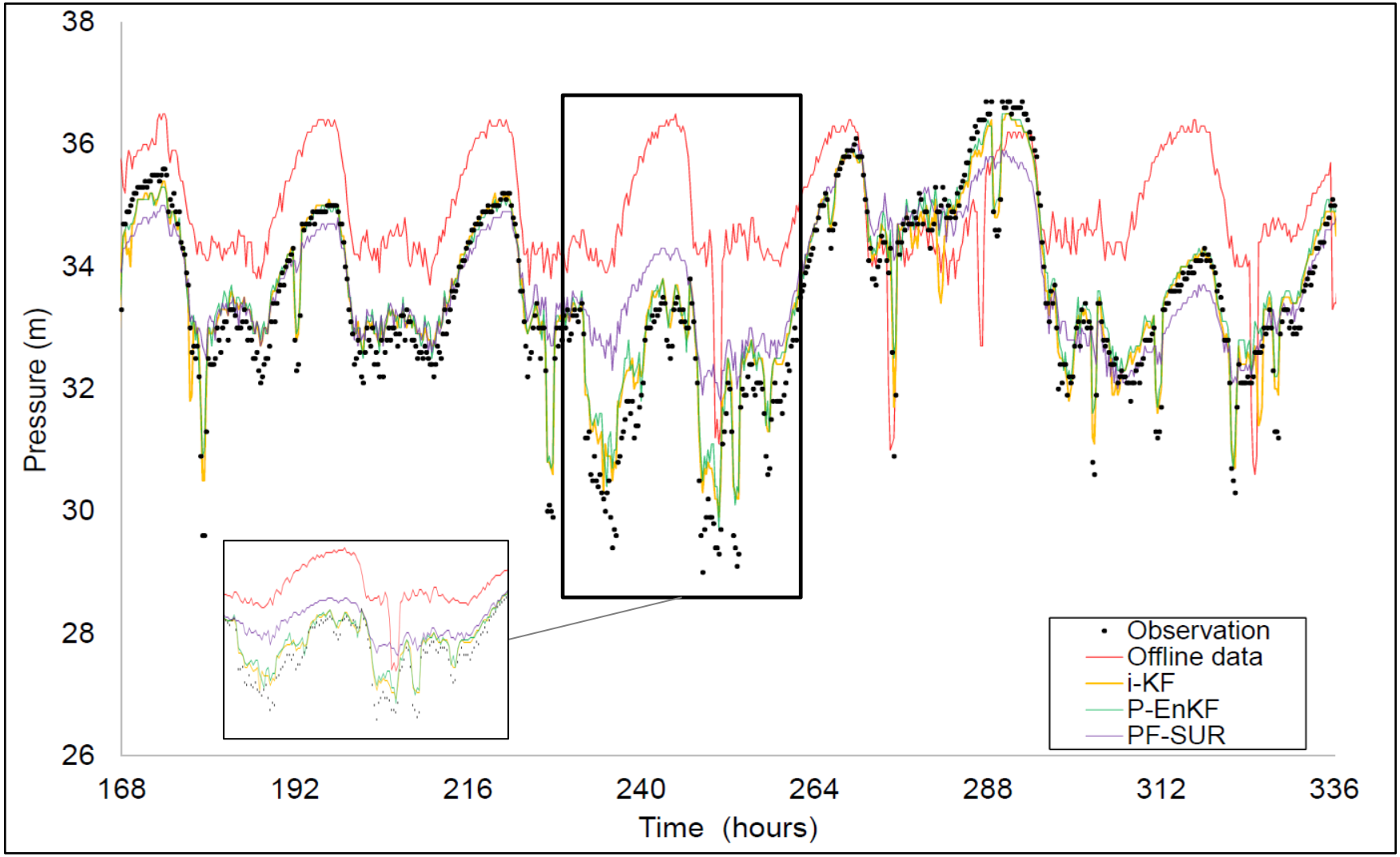


Figure 4.17: Comparison between observed and predicted pressure at node 00172D76 (DMA01 pressure meter) every 15 minutes

As it can be seen from Figure 4.16, the flow predictions made by three online hydraulic models are closer to the flow observations (black dots) when compared to the offline model (red line). The results in Figure 4.17 further justifies the need to use the online hydraulic model to monitor the reduced WSZ in the near real-time.

Table 4.11 and Table 4.12 show the comparison of the reduced WSZ01 network pressure prediction and correction statistics respectively.

Table 4.11: A comparison of the pressure prediction statistics

Data Statistics	**Hydraulic Model type**	**DMA01**	**DMA02**	**DMA04**	**DMA06**	**DMA07**
MAE(m)	**Offline**	2.1	2.7	0.8	1.2	3.0
	Iterative KF	0.5	0.4	0.4	0.8	0.8
	Perturbed EnKF	0.4	0.3	0.3	0.8	0.7
	PF-SUR	1.1	1.3	0.9	1.8	1.4
R^2	**Offline**	0.276	0.657	0.237	0.722	0.742
	Iterative KF	0.801	0.824	0.781	0.863	0.888
	Perturbed EnKF	0.880	0.951	0.882	0.888	0.896
	PF-SUR	0.812	0.878	0.794	0.847	0.881

Table 4.12: A comparison of the pressure correction statistics

Data Statistics	**Hydraulic Model type**	**DMA01**	**DMA02**	**DMA04**	**DMA06**	**DMA07**
MAE (m)	**Iterative KF**	0.3	0.3	0.6	0.8	0.7
	Perturbed EnKF	0.2	0.3	0.7	0.8	0.8
	PF-SUR	0.5	0.5	1.0	1.2	0.9
R^2	**Iterative KF**	0.987	0.990	0.951	0.983	0.991
	Perturbed EnKF	0.987	0.989	0.957	0.985	0.993
	PF-SUR	0.972	0.982	0.952	0.986	0.988

As it can be seen from Table 4.11 and Table 4.12, the MAE and R^2 values shown in all online hydraulic models generally perform better than the offline hydraulic model. Accurate correction of the WDS state (tank level) has helped in reducing the bias introduced at the initial conditions when the WDS hydraulic model propagated forwards in time. This has helped the online hydraulic models to predict pressure compared to offline hydraulic model. Therefore, WDS state (tank level) does not recompense for errors in forecasted demand at each time step. The iterative KF method performed better than the PF-SUR method in term of updating the WDS state (e.g., tank level, flow and pressure).

Table 4.13 shows the comparison of execution time for each data assimilation scheme. The DA methods performed with the aid of EPANET 2.0 and Microsoft Visual C++ on a HP laptop (with Intel core i5 processor @ 2.30GHz and 6.0 Gb RAM memory). It takes iterative KF method 70% less of the perturbed EnKF method time to run online hydraulic modelling of the reduced WSZ01 for four weeks (between 5th April and 2nd May 2015). However, it is still feasible to apply the EnKF method for online hydraulic modelling given the time step in real-time is 15 minutes.

Table 4.13: A comparison of execution time for each data assimilation scheme

Hydraulic Model type	**Execution time for a single time step (s)**	**Execution time for 4 weeks simulation**
Offline	0.82	0h 36m 33s
Iterative KF	2.09	1h 33m 25s
Perturbed EnKF	7.71	5h 45m 32s
PF-SUR	7.91	5h 57m 35s

As it can be seen from Table 4.13, the main advantage of the offline hydraulic model is its computational efficiency. Out of the three online hydraulic models, the i-KF method is by far most computationally efficient.

The results shown in Table 4.8 to Table 4.13 demonstrate that, in terms of prediction accuracy and agreement with observations, the perturbed EnKF method is the best online hydraulic model. Having said this, using this model is computational demanding for a large water network when compared to the iterative KF method. Given that the

iterative KF method seems to be working reasonably well, i.e. that it is not lagging much behind the perturbed EnKF method in terms of prediction accuracy, the iterative KF method will be used in the following chapters to develop, test/validate and demonstrate new methodology to detect and locate pipe bursts in the WDS.

The following section summaries this chapter and findings made in the case study.

4.7 SUMMARY

After introduction (Section 4.1), the chapter describes the case study used to evaluate the performance of an offline model and three online hydraulic models (Section 4.2). This is followed by the outlining the real-time hydraulic meters and other data available for the case study (Section 4.3).

The application of the offline hydraulic model is explained including the hydraulic model calibration process (Section 4.4). The implementation of three online hydraulic models are explained in Sections 4.5.3 to 4.5.5. The aim of all the online hydraulic models is to estimate optimal WDS states by quantifying and reducing model and data uncertainties. These methods are as follows: (1) Iterative KF method, (2) Perturbed EnKF method and (3) PF-SUR method. The model assumptions are also outlined (Section 4.6).

The online hydraulic models outperform the offline hydraulic model based on two performance metrics used (MAE and coefficient of determination). The on-line hydraulic model of a water distribution system is capable of making predictions that can reflect the WDS state more accurately than when using an offline model. This is because the DA methods used in an online hydraulic model updates the system states which minimises the bias in the initial conditions and, in turn, is used to simulate the system state in the next observation time step. On the other hand, the online hydraulic model computational times are larger than the corresponding offline model run times.

Among the three online hydraulic models, the perturbed EnKF method generally performed better when compared to the other two online hydraulic models. However,

the iterative KF method has substantially lower computational time than the other two methods with prediction accuracy that is not lagging much behind the best performing perturbed EnKF model. Hence, the iterative KF method is selected as part of the online hydraulic model that will be used in the following chapters to develop new methodology to detect and locate pipe bursts/leaks in the WDS.

5 BURST DETECTION AND LOCALISATION METHODOLOGY

5.1 INTRODUCTION

The WCs around the UK and rest of the world deal with the WDS pipe burst events due to ageing infrastructure and increasing water demand. Majority of the WDS pipe burst events in the UK were and, in many cases, still are detected via customer's phone calls. Lately, however, UK WCs have become more proactive and have started using different methods for detecting burst/leaks. Most of these methods are based on simple alarming of their flow/pressure meters with flat line thresholds and in some cases on more advanced predictive data analytics (Romano, et al., 2013). These data-driven methods are focusing on detecting pie bursts/leaks and raising corresponding alarms, i.e. not so much on locating and generally diagnosing these events which is still largely done via suitable field inspections where sounding or other equipment is typically used.

An alternative approach to using data-driven methods is to use hydraulic models of DMAs to detect and especially locate/diagnose bursts/leaks. These methods are becoming increasingly popular with the advent of online (i.e. 'live') hydraulic models and the increasing density of flow/pressure meters. The declining cost of flow/pressure meter technologies has reached the point where a larger scale deployment of these devices in a WDS is feasible (Kapelan, et al., 2005) and some of the UK WCs (e.g. Yorkshire Water, United Utilities and Severn Trent Water) are taking advantage of this by installing additional meters in their DMAs.

This chapter presents an online Burst Detection and Localisation Methodology (BDLM) combined with an iterative KF-based online hydraulic model chosen in the previous chapter (refer to Section 3.4). The online BDLM is developed to overcome the limitations of flow or pressure meters data only based techniques and offline hydraulic model-based techniques by constantly adapting to changing conditions in the network based on incoming flow/pressure observations. This way, WDS state is more accurately predicted which, in turn, enables more reliable detection and location of pipe burst and other abnormal events in the system.

After this introduction,

- Section 5.2 describes the overview of online BDLM.
- Section 5.3 provides the theoretical background and the methodological details for detecting WDS pipe burst events in a WDS.
- Section 5.4 describes the proposed model-based localisation method to estimate the likely location of a pipe burst event within a WDS.
- Section 5.5 summaries the report and considers the results of the method in the wider context.

5.2 Burst Detection and Localisation Methodology Overview

The BDLM presented in this chapter aims to detect and locate pipe burst events in near real-time at the DMA level. It also aims to detect and locate such pipe bursts in a reliable and timely fashion.

Figure 5.1 shows a flowchart of the proposed BDLM. The BDLM comprises five major components: (1) A Water Demand Forecasting Model (WDFM) (refer to Section 3.2); (2) a hydraulic model of the water network (i.e., a DMA model); (3) iterative Kalman Filter (i-KF) method (refer to Section 3.4); (4) Burst Detection Methodology (refer to Section 5.3) and (5) Burst Localisation Methodology (refer to Section 5.4).

As it can be observed from Figure 5.1, the developed BDLM proceeds as follows:

1. The WDFM forecasts water demands for the next time step (15 minutes) by using the methodology explained in Section 3.2.2.
2. The forecasted water demands are then used to drive the hydraulic model and derive predicted hydraulic state estimates, i.e. predict flow rates and pressure heads at locations corresponding to observations within the studied water network.

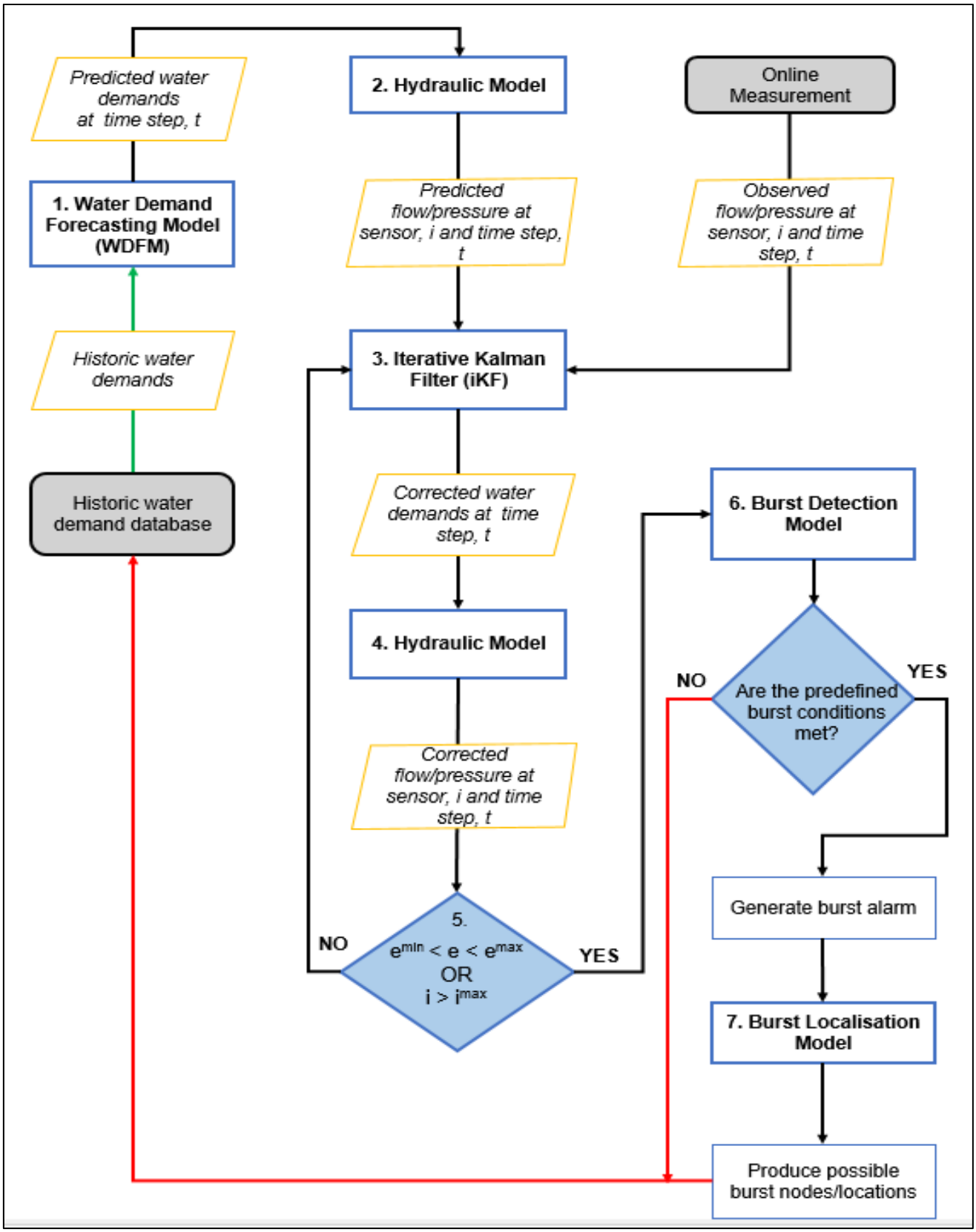


Figure 5.1: The flow chart of the burst detection methodology overview. (The red and green arrow indicate the beginning and end of the BDLM respectively)

3. The chosen DA method, i-KF then utilises the difference between the predicted hydraulic state estimates and the corresponding incoming WDS observations to calculate the corrected demands at the current time step.
4. The hydraulic model is then re-run for the current time-step using the corrected demands from the above step 3, to derive a set of corrected hydraulic state estimates.
5. The next step checks if all the flow errors between the corrected flow values from the current and previous iteration step are within a predefined tolerance value or the maximum number of iteration has been reached. If either of these conditions hold, move to step 6 otherwise repeat steps 3 - 4.
6. The burst detection method (details in Section 5.3) is used to determine if there is a pipe burst event or not in the analysed WDS. This is done by using the flow/pressure observations and the corresponding predictions/corrections obtained at hydraulic meter locations assuming no bursts (or other events) in the system. If the conditions of the burst detection rules are met, then a burst alarm is raised.
7. If and when burst is detected (i.e. burst alarm is raised), the burst localisation method is used to find the approximate area of the pipe burst event (details in Section 5.4).

Steps 1 – 7 are repeated at subsequent time steps in a studied WDS.

5.3 BURST DETECTION METHODOLOGY

5.3.1 Burst Detection Methodology Overview

The Burst Detection Methodology (BDM) is the burst detection component of the BDLM. It initialises once WDS state corrections are obtained (BDLM step 4 in Figure 5.1). The BDM aims to detect pipe burst events in the near real-time. Figure 5.2 shows a flowchart of BDLM's burst detection component.

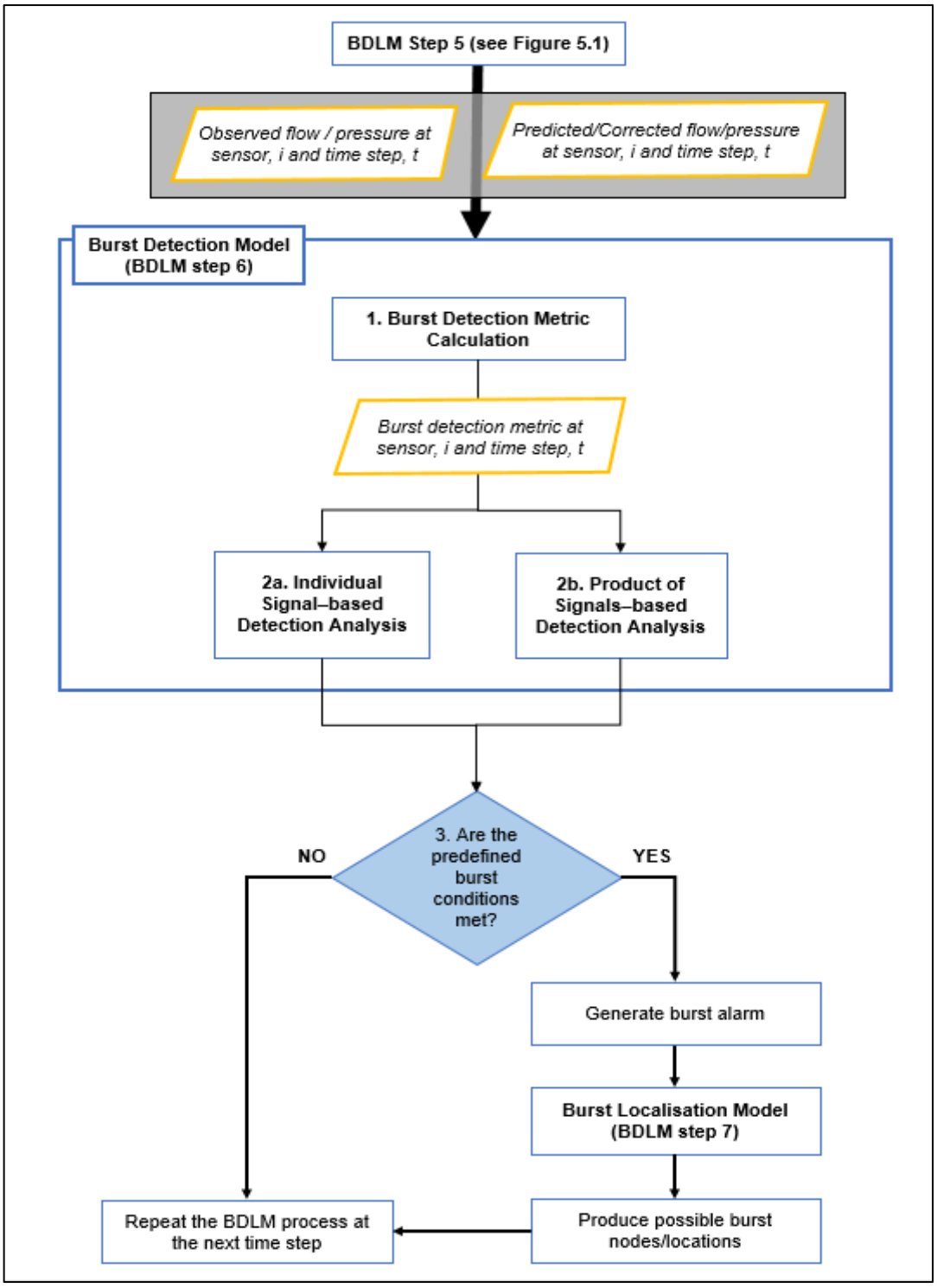


Figure 5.2: A flowchart of BDLM's burst detection method.

As it can be observed from Figure 5.2, it proceeds as follows (after BDLM step 5):

(1) The first step is to retrieve the hydraulic model predictions and their respective WDS observations from the hydraulic meters' location and calculate the burst detection metric values (refer to equation (36)).

(2) Once this is done, two detection analyses are used to detect pipe burst and they are as follows:

a. **Individual Signal-based Detection Analysis (ISDA)**: - reviews the recent MAR values of each flow/pressure meter against the pre-defined Statistical Process Control (SPC) chart limits and rules. Section 5.3.3.1 further describes how the ISDA-SPC detects pipe bursts in real-time along with its predefined statistical control limits and rules.

b. **Product of Signals-based Detection Analysis (PSDA)**: - first calculates the product of MAR value from two flow/pressure meters or more within a DMA and then reviews the recent product values against its pre-defined statistical control limits and rules. This detection analysis is further explained in Section 5.3.3.2 including the pre-defined PSDA-SPC statistical control rules.

If one of the pre-defined generic ISDA/PSDA SPC control rules is met, a burst alarm is raised. When a burst alarm is raised, the next step is to locate the potential area of the pipe bursts. Steps 1 – 2 in are repeated at each subsequent time step.

The BDM steps 2a and 2b work in parallel reviewing the burst detection metric values from different perspectives (e.g. flow/pressure data only and aggregated flow/pressure data). The reason for using two different detection analyses is each analysis focuses on collecting specific evidence that a pipe burst event has occurred. Furthermore, the above BDLM's detection method steps 2b aims to make use of multiple WDS observations from different hydraulic meter (and locations) to improve the discriminatory power to detect pipe burst events. Using two detection analyses also complement each other by independently identifying pipe burst events.

The next section explains how the burst detection metric value is calculated.

5.3.2 Burst Detection Metric

The burst detection metric is defined as follows:

$$\boldsymbol{M_{i,t}} = \frac{\mathbf{1}}{\boldsymbol{n_t}} \sum_{\boldsymbol{t=t-n_t}}^{\boldsymbol{n_t}} (\boldsymbol{X_{i,,t}} - \boldsymbol{Z_{i,t}}) \qquad (36)$$

where $M_{i,t}$ is the burst detection metric value at hydraulic meter *i* and time step *t*; $Z_{i,t}$ is the observed hydraulic state (flow or pressure); $X_{i,t}$ is the predicted/corrected hydraulic state (flow or pressure) obtained by using the online hydraulic model (refer to section 3.4); and n_t is the number of recent observations.

Equation (36) shows the moving average of residuals between predicted and observed flow/pressure data. The burst detection metric is referred as MAR in this thesis. The burst detection metric (i.e. MAR) was chosen after an investigation of alternative approaches to transform predicted/corrected and observed values into burst detection metric. The investigation included comparing the following performance of detection metrics:

- Residuals between hydraulic model predictions and their corresponding WDS observations;
- Normalised Residuals – ratio of residuals to corresponding WDS observations;
- Moving Average of Normalised Residuals – mean of the previous n normalised residuals;
- Residual Autocorrelation - is the Pearson correlation (Stigler, 1989) between hydraulic model predictions and their corresponding WDS observations.
- Squared Residual – square of residuals.
- Moving Mean Squared Residual - mean of the previous n squared residual;
- Exponential**/**Weighted Moving Averages of Residual (EMA/WMA) – is the sum product of weight coefficients and residuals divided by the total weight coefficient. These metrics allocate more weight to recent residuals and less weight to older residuals. However, EMA differs from WMA by taking into account its calculation of all the residual data of the hydraulic meter.
- Ratio – ratio of hydraulic model predictions to their corresponding WDS observations.

The reasons why the above burst detection metrics were discounted are as follows:

- **Normalised Residuals**: During the burst event, flow residuals are normalised by high flow observations which can make the metric values near to the non-burst event values. This undermines the opportunity to detect some pipe burst events (refer to Table 5.1and Figure 5.3). This detection issue also applies to

pressure observations when both flow and pipe burst are low. Therefore, developing generic burst detection control limits and rules to detect pipe bursts via normalised flow/pressure residuals is difficult due to flow/pressure diurnal variation.

- **Squared Residuals** make residual values below 1 much smaller while residual values above 1 become larger. Hence, a unique burst detection control limits and rules would be required for flow and pressure observations individually.
- **Ratio of predicted to observed flow/pressure values** is not sensitive to pipe burst events if observed value is used as the denominator. For example, if flow values are large due to a burst event and predicted flow values is small, the ratio values are close to zero. This makes it difficult to develop burst detection control limits and rules for flow/pressure meters.
- **Residuals** between hydraulic model predictions and WDS observations is discounted due to data noises which increase the likelihood of raising false alarms. Whilst the MAR helps in reducing amount of "noise" and minimising the impact of flow/pressure residual outliers during normal conditions which consequently reducing the number of false alarms.

Table 5.1 shows an example of the flow residuals, normalised flow residuals and flow MAR values assuming there is a pipe burst event in a water network. The non-burst and burst period are shaded in white and yellow respectively.

The MAR values in Table 5.1 are the average of 4 previous flow residual data at each time step. The MAR can be regarded as a short-term moving average. Using 4 recent flow residuals in this example shows that MAR can be used to detect a burst, as seen in substantially increasing MAR values following an occurrence of a burst.

Table 5.1: Example of the flow residuals, normalised flow residuals and flow MAR value

Time (hh.mm)	Predicted flow rates, X (l/s)	Observed flow rates, Z (l/s)	Residual, X - Z (l/s)	Normalised Residual	MAR* (l/s)
13:45	1.423	1.453	-0.030	0.128	-0.070
14:00	1.490	1.577	-0.086	-0.077	-0.073
14:15	1.668	1.613	0.055	-0.125	-0.024
14:30	1.889	1.989	-0.101	-0.026	-0.040
14:45	2.177	1.954	0.223	-0.021	0.023
15:00	2.143	2.142	0.001	-0.055	0.045
15:15	2.268	2.383	-0.115	0.034	0.002
15:30	2.605	2.494	0.112	-0.051	0.055
15:45	2.455	2.930	-0.475	0.114	-0.119
16:00	2.620	3.118	-0.498	0.000	-0.244
16:15	2.975	3.206	-0.231	-0.048	-0.273
16:30	3.065	3.354	-0.288	0.045	-0.373
16:45	3.260	3.566	-0.306	-0.162	-0.331
17:00	3.417	3.681	-0.264	-0.160	-0.272
17:15	3.200	3.689	-0.489	-0.072	-0.337
17:30	3.408	3.738	-0.329	-0.086	-0.347

Figure 5.3 shows a plot of flow residuals, normalised flow residuals and flow MAR values based on the flow data in Table 5.1. The data highlighted in yellow are burst data hence the pipe burst start at 15:45 hour.

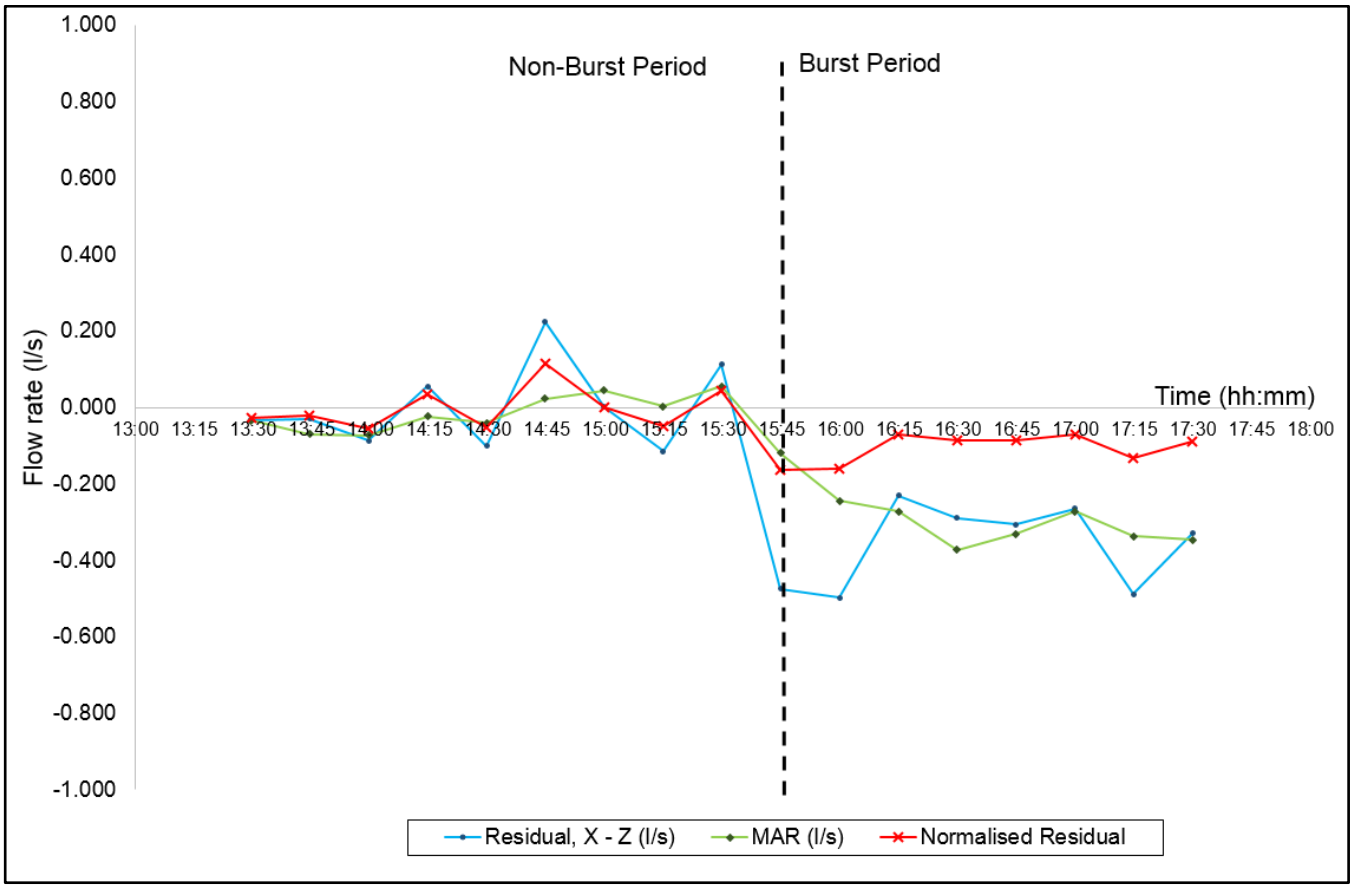


Figure 5.3: Example of flow residuals chart.

As it can be seen from Figure 5.3, normalised flow residual is not sensitive to the pipe burst. At the same time, the ordinary flow residual can be more erratic compared to the MAR data and this erratic nature can undermine the opportunity to detect a pipe burst. The MAR data has the advantage of resist the impact of low/high flow residuals and is more sensitive and stable during the burst period. Therefore, the MAR is selected over residual values as burst detection metric for the above reasons.

It can be argued that Exponentially Moving Average (EMA) and Weighted Moving Average (WMA) can be more sensitive and moves closer to the flow/pressure residual in action. However, the MAR is preferred here due to its resistant to data anomalies or sudden change in residual data under normal conditions.

The next section describes all the burst detection analyses used in this thesis.

5.3.3 Burst Detection Metric Value Analysis

5.3.3.1 *Individual Signal-based Detection Analysis (ISDA)*

The ISDA is the first of the two burst detection analyses with an aim to identify the change in recent MAR values due to abnormal events especially medium/large pipe burst events. The burst detection analysis is based on the SPC–based technique (Montgomery, 2005). The SPC-based Chart is a graphical representation of descriptive statistics that can be used to study how a process variable changes over time. A SPC-based chart uses control limits to detect signals in process data that are not operating under normal conditions. The control limits for a given variable are normally at a distance of the product of a user defined multiplier and the standard deviation of the statistics data from its statistics mean. The theoretical background and application of SPC-based technique can be found in Hardwick (2014) and Qiu (2013).

Due to the complexity of WDS (nonlinear relationship between flow and pressure and distance between the burst and hydraulic meters), some hydraulic meters may be insensitive to some pipe burst events. Therefore, ISDA SPC-based Chart is developed to allow any individual flow/pressure meter to detect any pipe burst events based on their respective MAR values. At each time step, ISDA starts by accumulating the *n*-time-step historical MAR data set of a hydraulic meter. Then estimate a hydraulic meter's MAR mean, $\boldsymbol{\mu_{i,t}}$ and standard deviation, $\boldsymbol{\sigma_{i,t}}$ using the following equations:

$$\boldsymbol{\mu_{i,t}} = \frac{\mathbf{1}}{\boldsymbol{n_t}} \sum_{\mathbf{t}=\ \mathbf{t}-\mathbf{n_t}}^{\mathbf{n_t}} \left(\mathbf{M_{i,\ t}}\right) \tag{37}$$

$$\boldsymbol{\sigma_{i,t}} = \sqrt{\frac{\sum_{\boldsymbol{t}=\mathbf{t}-\mathbf{n_t}}^{\mathbf{n_t}} \left(\mathbf{M_{i,\ t}} - \boldsymbol{\mu_{i,t}}\right)}{\mathbf{n_t}}} \tag{38}$$

where μ_i and $\sigma_i$are the mean and standard deviation of burst detection metric values respectively. *k* is denoted as user defined multiplier of control limits and the subscript, *i* is the hydraulic meter index.

Once the mean and standard deviation are estimated, the ISDA SPC-based Chart primarily checks if the MAR values do exceed pre-defined control limits based on the estimated measures of the MAR's variability. If, at any time step, a burst detection metric falls outside these limits, it may indicate a presence of an abnormal event is taking place in the DMA being analysed. However, an increasing number of successive burst detection metric values above of the high control limit (i.e., $\bar{x} + 4\sigma$ and $\bar{x} + 5\sigma$) indicates a stronger likelihood of a pipe burst in a DMA. Therefore, a Control Chart that reviews the recent history of the burst detection metric data improve the reliability and the effectiveness of pipe burst detection method. If the same observation is made in consecutive time steps, it indicates a high possibility of a pipe burst event occurring in the studied WDS. The ISDA SPC-based Control Rule focuses on the recent history of the MAR hence making this burst detection analysis an effective and reliable detection tool. The ISDA assesses the MAR values in successive time steps using SPC-based Control Rules. A set of Control Rules provide the different criteria for determining if the deviations/outliers are due to a pipe burst event occurring in the studied WDS.

The following modified subset of the Western Electric Control Rules (Western Electric Company, 1956) for detecting 'abnormal' events specifically pipe bursts via IDSA are used here:

- Rule 1: ten consecutive points fall between 2σ and 3σ from the centreline;
- Rule 2: six consecutive points fall between 3σ and 4σ from the centreline;
- Rule 3: four consecutive points fall between 4σ and 5σ from the centreline;
- Rule 4: three consecutive points fall between 5σ and 6σ from the centreline;
- Rule 5: two consecutive points that fall beyond 7σ.

The alarm is risen when either of these rules apply, i.e. as soon as the first of the above conditions is met.

The above Control Rules were identified after relevant sensitivity analysis shown in Chapter 6. The mean and standard deviation used in the control rules are derived from historical trends of flow/pressure P-MAR values for time periods without bursts.

Note that there is no control rule between ±2 standard deviations from the Chart centreline (mean). This is to avoid the possibility to raise false alarm and is based on the sensitivity analysis performed on a real-life case study in Chapter 6. As shown there, majority of '*normal'* MAR values fall within ±2 standard deviations from the Chart centreline (mean). This sensitivity analysis was tested on comparative detection metrics (i.e., residual, normalised residual) in different types of real-life WDS (or DMA). It was concluded that the MAR and rules shown above generated the best event detection results.

The above set of ISDA SPC-based Control Rules can be applied to all flow/pressure meters in a DMA. The Control Rules can be applicable to any water networks (or DMAs) regardless of the chosen combination of hydraulic meters in the near real-time. The next section discusses how two flow/pressure signals are combined and analysed to raise burst alarm if a pipe burst occurs.

5.3.3.2 *Product of Signals-based Detection Analysis (PSDA)*

The PSDA is the second BDLM detection analysis with an aim of detecting pipe bursts that manifest themselves simultaneously on 2 or more flow/pressure meters. This makes the detection more reliable. In PSDA, the product of MAR values from 2 or more flow/pressure meters are evaluated against the pre-defined limits at each time step. The pre-defined limits are based on the estimated measures of the MAR product's variability. Therefore, the PSDA complements the ISDA by detecting pipe burst events missed by the ISDA.

The PSDA starts by multiplying two or more MAR values from the same type of hydraulic meters (flow or pressure). It can be expressed as follows:

$$\boldsymbol{M_{ij,t}} = \prod_{i=m}^{n} \boldsymbol{M_{i,t}} \qquad (39)$$

where $M_{ij,t}$ is the burst detection metric value of two or more flow/pressure meters *i, j*; *m* is hydraulic meter index and *n* is the number of individual hydraulic meters employed in a DMA.

The product value of two or more different MAR values is referred as P-MAR value here. Once P-MAR values are estimated, PSDA SPC-based Chart review recent P-MAR values against the statistical boundaries (mean and standard deviation). Equations (37) and (38) are used to estimate the mean and standard deviation at each time step. The outliers outside the statistical boundaries are reviewed in successive time steps using SPC-based Control Rules. The PSDA is only applicable if there are more than 1 flow/pressure meter within a DMA. For example, if a DMA has 3 flow meters and 3 pressure meters, and labelled as F1, F2, F3 and P1, P2, P3 respectively, there will be 8 PSDA SPC-based charts. This is due to the following hydraulic meter combinations: (1) F1*F2, (2) F2*F3, (3) F1*F3, (4) F1*F2*F3, (5) P1*P2, (6) P2*P3, (7) P1*P3 and (8) P1*P2*P3. Therefore, 8 PDSA-SPC-based charts (comprising 4 flow/pressure charts) will be used analyse P-MAR values for pipe bursts detection.

The following subset of Control Rules for detecting pipe burst events via a PSDA SPC-based chart are used here:

- Rule 1: eight consecutive points fall outside -2σ and 2σ from the centreline;
- Rule 2: five consecutive points fall outside -3σ and 3σ from the centreline;
- Rule 3: four consecutive points fall outside -4σ and 4σ from the centreline;
- Rule 4: three consecutive points that fall outside -5σ and 5σ from the centreline;
- Rule 5: two consecutive points that fall outside -6σ and 6σ from the centreline.

As in the case of ISDA, alarm is risen as soon as either of the above rules apply, i.e. as soon as any of the above conditions is met.

The above Control Rules, motivated by the rules developed for ISDA, were identified after relevant sensitivity analysis shown in Chapter 6. As in the case of ISDA, the mean and standard deviation of the PSDA SPC-based Control Charts are derived from historical trends of flow/pressure P-MAR values for time periods without bursts.

Note that there is no control rule between ±2 standard deviations from the Chart centreline (mean). This is to avoid the possibility to raise false alarm and is based on the sensitivity analysis performed on a real-life case study in Chapter 6.

The next section discusses how all the DMA flow/pressure signals are aggregated and then reviewed for burst detection analysis in real-time.

5.3.4 Final Remarks

The BDM aims to make use of the presence of persistent outliers in flow/pressure meter signals caused by abnormal events (i.e., pipe bursts) to raise burst alarm. The BDM make uses of moving average values of flow/pressure residuals, i.e. differences between online hydraulic model predictions of pressures/flows and the corresponding observations at hydraulic meter locations. The evidence about a potential burst event collected this way is further processed by the ISDA/PSDA SPC-based Control Charts and Rules developed and presented here. This level of sophistication (i.e. as opposed to using flat line thresholds for flow/pressure signal values) is required to maximise the generation of alarms based on genuine burst events and, at the same time, minimise a number of potential false alarms. The false alarms could be related to unusual WDS behaviour, hydraulic meter failure, noisy data or poorly calibrated hydraulic meters.

In addition to the above, it could also be argued that the changing control limits (moving mean and standard deviation) can help in dealing with changing operating conditions of a DMA. Examples of changes in operating conditions can be an increased/decreased flow due to properties development, holiday periods (summer holiday, school half term break) and boundary change in a DMA. The methodology presented here does not account for this presently.

The developed BDM also does not account for one-off events such a big sport events or music festivals. If necessary, these events could be easily accounted for in the detection methodology presented here. However, network modelling engineers and network technicians in the control room are usually aware of these events hence no need for this.

5.4 BURST LOCALISATION METHODOLOGY

5.4.1 Burst Localisation Methodology Overview

The objective of the proposed Burst Localisation Methodology (BLM) is to approximately locate the burst within the DMA in near real-time. The BLM developed and presented here makes the use of the online hydraulic model, detection metric values and Statistical Analyses. The BLM is initialised after a pipe burst event is detected via the developed BDM (in Section 5.3). An overview of principal BLM procedural steps is shown in Figure 5.4.

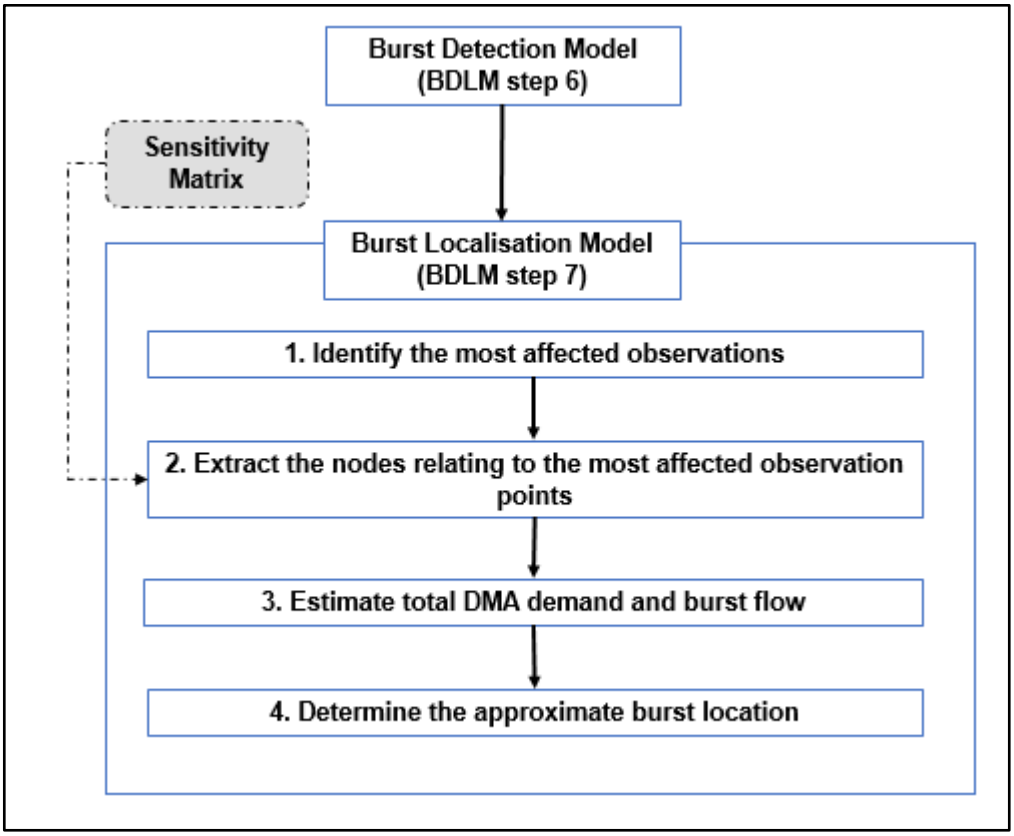


Figure 5.4: The flowchart of Burst Localisation Methodology

As it can be seen from Figure 5.4, a sensitivity matrix is developed offline, i.e. before the BLM methodology is applied online. The sensitivity matrix is a matrix based on the sensitivity of hydraulic model predicted flows and pressures at hydraulic meter locations to bursts simulated at different network nodes. Sensitivity is determined by

using multiple offline hydraulic simulations with the aim to identify the most sensitive hydraulic meters' locations assuming potential bursts occurring at all network nodes.

Once a pipe burst event is detected, the first step is to rank all the flow/pressure meters locations in descending order based on the burst detection metric values obtained. The order (rank) of the observation points indicates the level of individual hydraulic meter signal being affected by the burst (refer to Section 5.4.2 for more details) with the highest rank indicating the hydraulic meter that is most affected by the burst.

The second step extracts the list of likely burst locations from the master sensitivity matrix relating to the most affected flow/pressure meters identified in BLM step 1. Section 5.4.2 describes the process to develop the master sensitivity matrix.

The third step estimates the total DMA demand and burst flow using multiple hydraulic model simulations and several data analysis techniques. The detailed process used to estimate the total DMA demand and burst flow is explained in Section 5.4.3.

The fourth step involves running the hydraulic model with the estimated total DMA demand and the burst flow simulated in turn at each network node shortlisted (as likely burst location) in BLM step 2. The simulation is performed in the near-real time. The total absolute flow/pressure differences between the observed and predicted values at the flow/pressure meter locations are then calculated. The analysed candidate burst locations (network nodes) are ranked in ascending order based on the total flow and pressure residual values. A predefined number of high ranking candidate network nodes are used to determine the burst area. The BLM procedure outlined above is repeated at each time step (e.g., every 15 minutes) following a burst alarm.

The next section explains how the sensitivity matrix is developed.

5.4.2 Development of the Sensitivity Matrix

The sensitivity matrix is a binary matrix. The matrix value of 1 indicates that the observation points are sensitive to burst located at given network node and the

matrix value of 0 means otherwise. The proposed method to build the sensitivity matrix offline follows four steps as summarised in Figure 5.5.

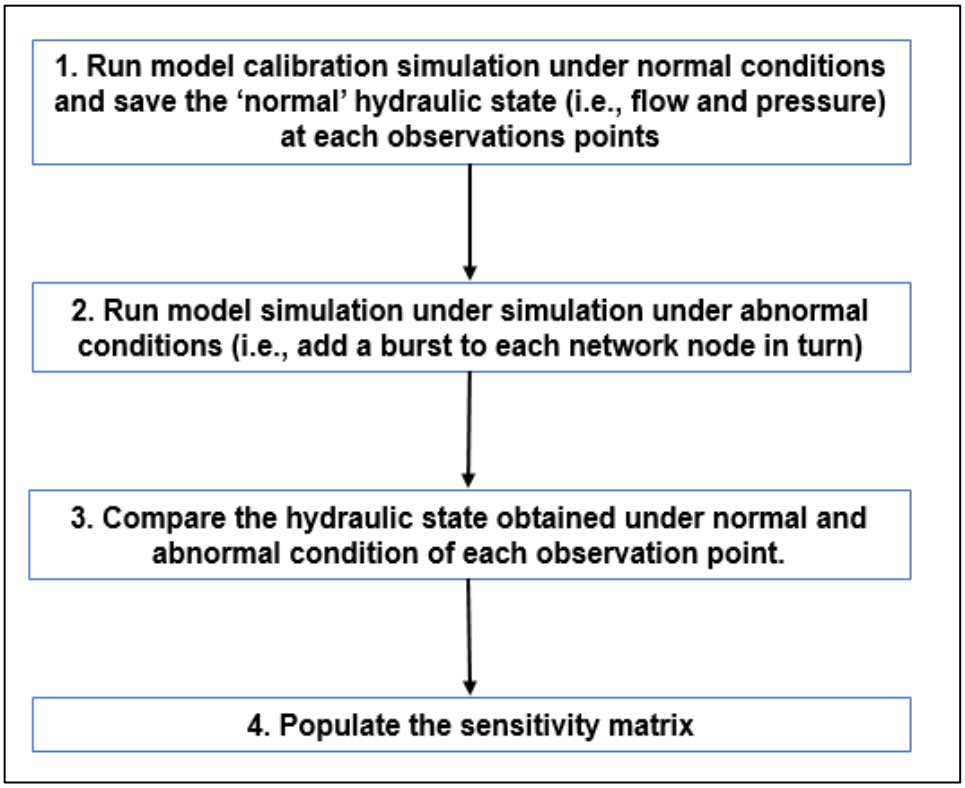


Figure 5.5: The flowchart of hydraulic meters - nodes matrix

A starting point is the well calibrated offline model of the analysed water distribution system. The first step runs hydraulic simulations for 24 hours with fixed time steps (i.e., 15 minutes) under normal conditions (i.e. assuming no bursts and by using demands from the offline calibrated model) to obtain a hydraulic system state (i.e. pressure or flow) at each hydraulic meter's location. The second step repeats the previous step but under abnormal conditions during which bursts are simulated in turn at each network node (i.e. at all possible locations).

The bursts are modelled as pressure-dependent emitter flows (Wu, et al., 2002) located at system nodes, i.e. as follows:

$$\boldsymbol{q_{j,t} = Cp_{j,t}^{e}} \quad (40)$$

where $q_{j,t}$ is the burst flow at node, i at time step, t; C is the emitter discharge coefficient; $p_{j,t}^{e}$ is the nodal pressure at node j, and time step, t and e is the emitter pressure exponent. The emitter exponent of 0.5 in the hydraulic model is assumed as recommended in (Lambert, 2002). This theoretical exponent normally works well for rigid, metal pipes but not so well for the more flexible plastic pipes. The exponent value used here could be changed in practice if necessary.

The third step compares the hydraulic states obtained under normal and burst conditions at each observation point. A threshold value is used to determine whether burst simulated at a given network node affects significantly enough the hydraulic state at each observation point. If the hydraulic state residual error between normal and abnormal conditions is greater than this threshold, then a value of 1 is assigned to the corresponding sensitivity matrix variable (zero otherwise).

This above process is repeated for different burst magnitudes (i.e., between 5% and 40% of the average DMA daily demand) and different time of burst occurrence during the day (i.e., 02:00 to 05.00, 06:30 to 09:30, 18:00 to 21:00) to generate multiple sensitivity matrices. Once multiple sensitivity matrices are generated, the values in all generated matrices are added together to make a master sensitivity matrix.

The fourth step is to convert the master sensitivity matrix (from the above step 3) into a binary matrix by assigning 1 to the corresponding master sensitivity matrix variable values above a threshold and zero otherwise, i.e. as shown in Figure 5.6.

Figure 5.6 shows an example of a master sensitivity matrix generation.

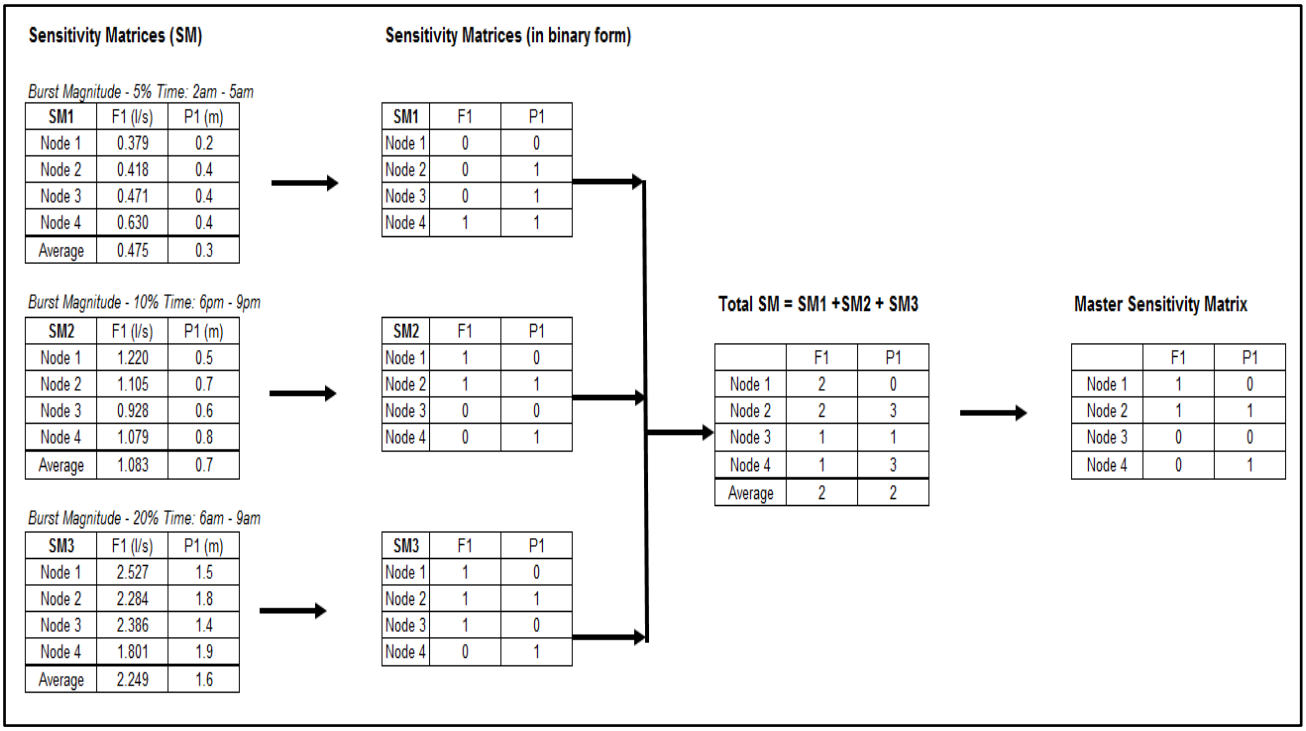


Figure 5.6: Example of a master sensitivity matrix generation

The master sensitivity matrix is then used during online hydraulic modelling when a burst alarm is raised. The next section explains how DMA demand and burst flow are calculated.

5.4.3 Estimation of Total DMA Demand and Burst Flow

The proposed method to determine a total DMA demand and burst flow in the near real-time is shown in Figure 5.7. The total DMA demand is estimated assuming no bursts in the DMA.

The first step estimates the initial total DMA demand and burst flow. The initial total DMA demand is the forecasted DMA demand from a water demand forecasting model (refer to Section 4.2). The burst flow is the difference between the estimated demand based on the inlet/outlet flow observations and the forecasted DMA demand obtained at the point in time when burst is detected (i.e. alarm risen).

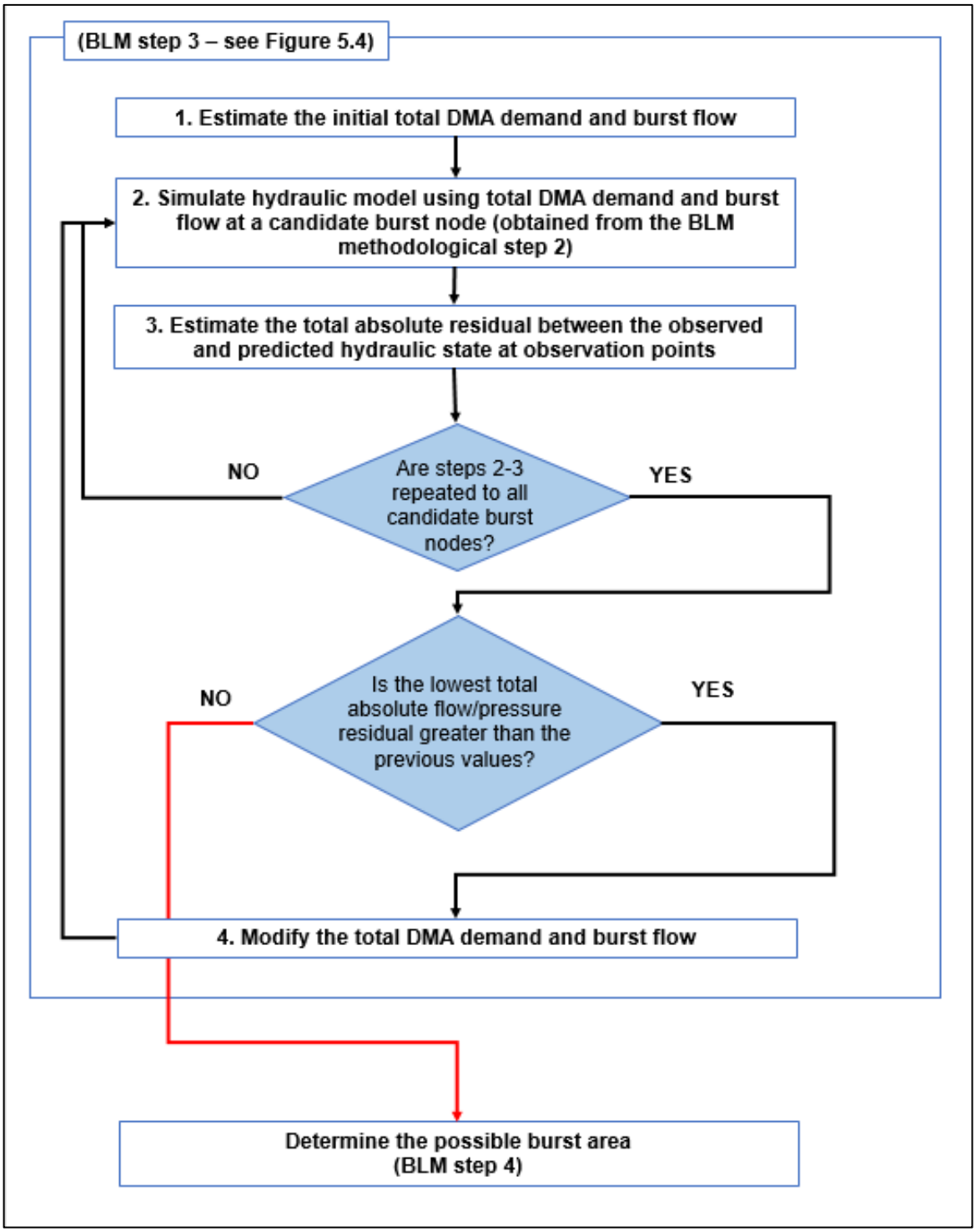


Figure 5.7: The flow chart of total DMA demand and burst flow estimation

The second step simulates the burst at each node obtained from the BLM methodological step 2. Bursts are simulated by using the online hydraulic model with the estimated burst flow and DMA demand. The network pressures and flows predicted this way are then used in the third step to determine the corresponding residuals between the observed and predicted hydraulic states at all hydraulic meter's locations.

Then calculate the total absolute flow/pressure residuals for the candidate burst location. Steps 2 – 3 are repeated for other candidate burst locations to obtain their respective total absolute flow/pressure residuals. Once all this is completed, the candidate burst nodes are ranked in descending order from the lowest to highest total absolute flow/pressure residual to determine the possible burst area (refer to Section 5.4.4 for further details). If the lowest total absolute flow/pressure residuals are greater than lowest total absolute flow/pressure residuals from the previous iteration step. This decision box is only applicable after the second iteration step. The simulation proceeds to step 4.

The fourth step in Figure 5.7 aims to change the burst flow and total DMA demand. This step initially increases the burst flow and decreases the total DMA demand by the same flow increment. When the total DMA demand and burst flow are modified, the above steps 2-3 are repeated to obtain a new lowest total absolute flow/pressure residual. The new lowest total absolute flow/pressure residual value is then compared to its recent values. If the new lowest total absolute flow/pressure residual value is greater than its recent absolute residual value, steps 2-4 are repeated until it becomes less than its recent value. Therefore, the most recent burst flow and total DMA demand are used as the burst flow and total DMA demand at the point in time.

The next section describes how the burst locations are determined.

5.4.4 Determination of Burst Location

Once the total absolute flow/pressure residuals are estimated for all potential burst locations (by using steps 1-4 shown in Figure 5.7), candidate burst nodes are ranked using these values, from the lowest to highest total absolute flow/pressure residual. Table 5.2 shows an example of how this ranking is done at each time step (during the pipe burst period only). As it can be seen from this table, each candidate burst location's ranks obtained for respective total absolute flow and pressure residuals are added together and then used to perform the overall ranking. If the two candidate burst locations have the same 'total rank', the candidate burst location with a lower total absolute flow residual is ranked before the other candidate locations.

Table 5.2: Example ranking of candidate burst nodes

Burst Location (Node index)	Total Absolute Flow Residual (l/s)	Total Absolute Pressure Residual (m)	Node Rank (based on total absolute flow residual)	Node Rank (based on total absolute pressure residual)	Total Rank	Overall Rank
X1	0.458	1.9	4	5	9	5
X2	0.293	1.4	1	2	3	1
X3	0.971	2.3	9	8	17	9
X4	0.450	2.0	3	6	9	4
X5	0.876	1.5	8	3	11	7
X6	0.992	2.1	10	7	17	10
X7	0.539	2.9	5	9	14	8
X8	0.543	1.4	6	2	8	3
X9	0.778	1.7	7	4	11	6
X10	0.305	1.3	2	1	3	2

Table 5.3 reviews historic rankings of potential burst location or nodes to produce a list of likely burst locations. (I.e. node X6 and X10) during the pipe burst period only.

Table 5.3: Example ranking of recurring candidate burst nodes

Burst Location (Node index)	Time step, t-3 (45 minutes ago)	Time step, t-2 (30 minutes ago)	Time step, t-1 (15 minutes ago)	Time step, t	Total Rank	Overall Rank
X1	8	6	4	3	21	5
X2	3	10	5	5	23	6
X3	10	7	9	6	32	9
X4	1	4	7	7	19	4
X5	7	1	2	4	14	3
X6	4	5	1	1	11	1
X7	5	9	10	9	33	10
X8	2	8	8	10	28	8
X9	9	3	6	8	26	7
X10	6	2	3	2	13	2

The highest ranked burst location (i.e. network node) node represents the most likely burst location (e.g. network node X2 in the example shown in Table 5.2) while the lowest ranked node ('X6' in the same example) represents the least likely burst location at that particular time step.

Finally, a predefined number (or percentage) of recurring top ranked network nodes are used to define the likely burst area (refer to Table 5.3).

5.5 SUMMARY

This chapter presented a methodology for an online detection and localisation of WDS pipe burst events. Its key components, the Online Hydraulic Model, Burst Detection Methodology and Burst Localisation Methodology, were discussed. After the introduction in Section 5.1. Section 5.2 provided the high-level overview of the burst and detection and localisation method. Section 5.3 presented two BDM detection analyses to detect WDS pipe burst events in the near real-time. The section explained the background and the methodological details of the model-based detection. Section 5.4 presented a methodology to locate WDS pipe burst events after a burst alarm is raised in the near real-time. The theoretical background and the methodological details of the model-based localisation technique.

The proposed burst detection and location methodology presented here is implemented on Microsoft Visual C++ software platform. It was developed to make use of both flow and pressure data in the near-real time. EPANET 2.0 is used to drive a hydraulic model in time to output hydraulic model predictions. In a scenario installing multiple flow meters continue persist to be challenging in a DMA. The developed methodology can be used with limited number of flow meters with additional pressure meters.

The next chapter test and demonstrate the developed burst detection and localisation methodology developed in this chapter on two case studies. The first case study focused on evaluating the performance of the developed BDLM on a number of artificially created bursts (Section 6.2). The other case study reviews the developed

BDLM’s performance on limited number of real engineered events where bursts where simulated by opening fire hydrants at different times of the day (Section 6.3).

6 CASE STUDIES FOR BURST DETECTION AND LOCALISATION

6.1 INTRODUCTION

This chapter presents the results of burst detection and location analyses carried out on two case studies. The main purpose was to test, evaluate and illustrate the capabilities of the BDLM (developed in Chapter 5) at a DMA level.

The first case study is a real-life DMA with inflow/outflow/pressure data collected from the United Utilities (UU) database in 2015. The second case study is a real-life DMA with real life flow/pressure data in Yorkshire Water Services (YWS). The second case study data were collected as part of the Engineering and Physical Sciences Research Council (EPSRC) NEPTUNE project (Savić et al., 2008).

The chapter is organised as follows. After this introduction, Sections 6.2 and 6.3 report the results obtained on the first and second case study respectively. In each of these sections: -

(i) a general description of the case study including DMA area and hydraulic data details are provided;
(ii) the data used for the BDLM analyses are described;
(iii) BDLM parameters are outlined.

After the general description, the details of the various data analyses and their results are presented and discussed. Once this is done, a summary of the chapter and the main conclusions are given in Section 6.4.

6.2 CASE STUDY #1

6.2.1 Section Overview

This section focuses on testing and evaluating the capabilities of the BDLM that enables performing pipe burst detection and localisation in a DMA. After a general description of the case study area (Section 6.2.2), hydraulic meters data used are outlined (Section 6.2.3) and chosen parameters used for the BDLM are presented

(Section 6.2.5). Section 6.2.4 explains how the artificial flow and pressure data including artificial pipe burst events are generated. Section 6.2.5 and 6.2.6 provides the result of the BDLM detection and localisation capability respectively. A summary of the case study and the main results are summarised in Section 6.2.8.

To analyse the robustness of the BDLM on detecting and locating burst events (simulated by using the case study water network model), nine analyses were conducted to evaluate the capabilities of the methodology that enables performing burst detection and localisation in a DMA.

The main aims of the analyses performed and presented here are as follows:

- The first analysis aims at investigating the performance of the BDLM for different number and location of perfect observed data (using 9 hydraulic meters combinations – refer to Table 6.2).
- The second analysis aims at evaluating the performance of the BDLM for bursts occurring at 3 different time periods of the day (i.e., morning, evening and night).
- The third analysis aims at evaluating the performance of the BDLM by using imperfect observed hydraulic meters data, i.e. meters data with 5 different observation noise levels.
- The fourth analysis aims at evaluating the performance of the BDLM for different burst locations within the DMA (5 different burst locations).
- The fifth analysis aims at evaluating the performance of the BDLM for different burst flows (using 4 different burst magnitudes).
- The sixth analysis aims at evaluating the time taken to raise the burst alarm from the time burst has occurred.
- The seventh analysis aims at investigating why false burst alarms arise during the non-burst period.
- The eight analysis aims at investigating the sensitivity of the closest hydraulic meter to the pipe bursts' location.
- The ninth analysis aims at testing and evaluating the methodology's capabilities on locating various artificial burst events detected by the BDLM.

Details of how these analyses were carried out and of the results obtained are reported in sections 6.2.6 to 6.2.7. The aforementioned analyses are performed on Microsoft Visual C++ along with EPANET 2.0 on a HP laptop (with Intel core i5 processor @ 2.30GHz and 6.0 Gb RAM memory).

6.2.2 Case Study 1 Area Description

The BDLM was tested on a real-life water network called DMA07. The DMA07 water network is part of the WSZ01 water supply zone in United Utilities and it is shown in Figure 6.1 in Section 4.2. Thus DMA, located in the North-West of England, supplies water to approximately 1800 customers under gravity with an average daily demand of 9 l/s. The DMA hydraulic model consists of 527 nodes, 412 pipes and 147 valves. The DMA07 is fed from another, upstream DMA (DMA04). The DMA07 network model is shown in Figure 6.1.

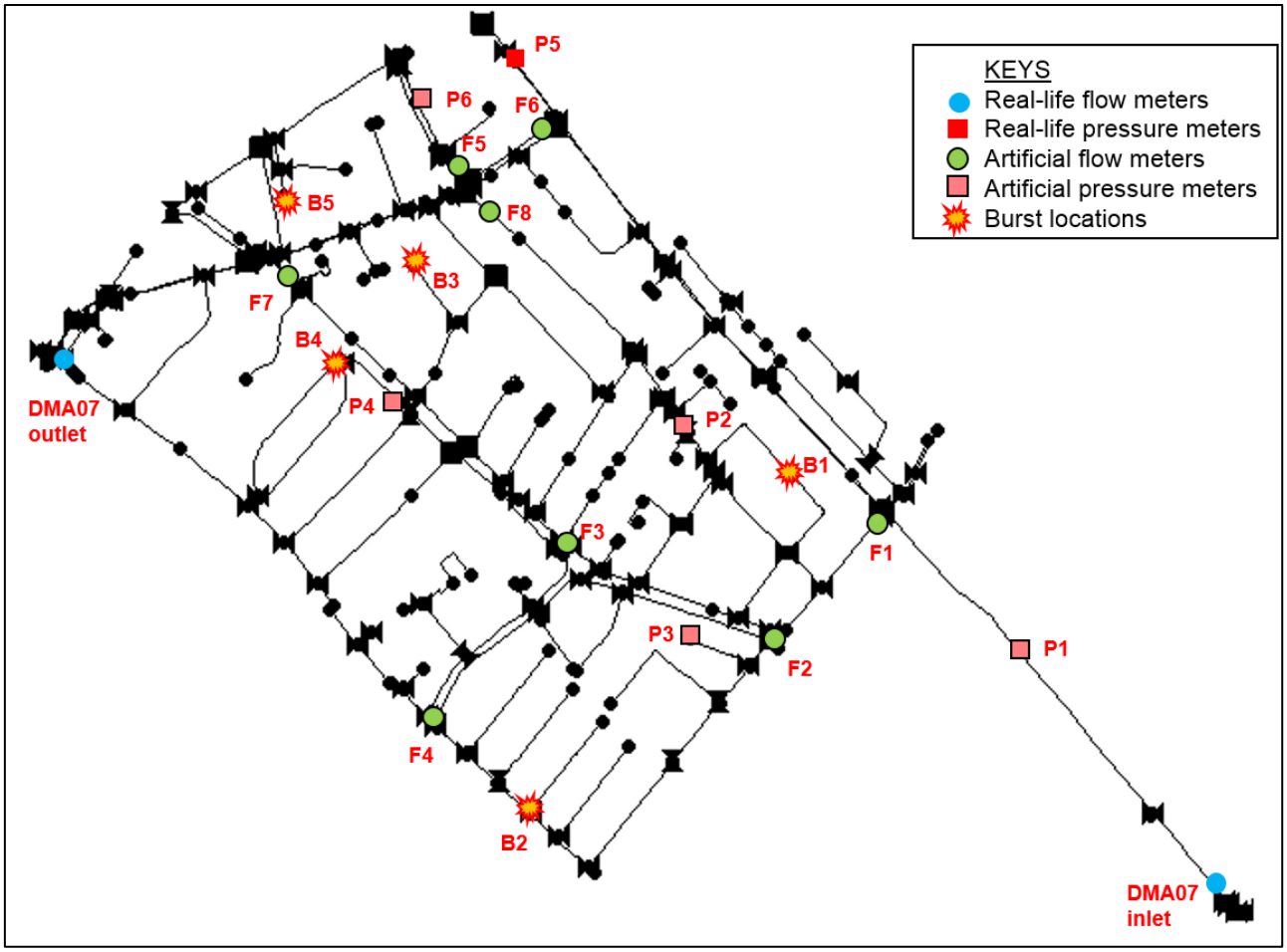


Figure 6.1: The overview of the studied water network - DMA07.

6.2.3 Sensor Data

The available real-life data are the DMA07's inflow and outflow data and pressure data gathered at the highest point of the DMA07 (refer to **Figure 6.1**). The flow and pressure data are gathered between 1st February 2015 and 31st May 2015 (17 weeks). Given limited real hydraulic meters data available, the artificial flow and pressure measurements were generated by using the hydraulic model of the DMA and assuming artificial pipe bursts of different characteristics within the DMA. The 'perfect' flow/pressure observations obtained this way are altered by adding random noise and/or systematic errors, as shown in Table 6.1. The spatial (i.e. nodal) demand allocation is done based on the fraction of DMA07 water demand allocated to each network node and the total DMA demand.

Table 6.1: The type of flow observation noises. $Q_{i,t}$ represents the flow rate for the flow meter, *i* at time step, *t*; and $Q_{i,avg}$ is the average flow rate at flow meter, *i* at the time step, *t.*

Observational noise type	**Type of error**	**Error model**
1	No noise	*N/A*
2	Random noise	$N(0, 0.025Q_{i,t})$
3	Random noise	$N(0, 0.05Q_{i,t})$
4	Systematic errors and random noise	$+0.025Q_{i,avg}$ *and* $N(0, 0.025Q_{i,t})$
5	Systematic errors and random noise	$-0.025Q_{i,avg}$ *and* $N(0, 0.025Q_{i,t})$

The location of the artificial flow and pressure meters (refer to **Figure 6.1**) are selected based on a developed methodology using an optimal meter placement method for DMA monitoring (Farley, et al., 2010). The hydraulic meters placement methodology (Farley, et al., 2010) is used from Bentley Systems' software called "DarwinSampler" (Wu, 2013). An additional 14 artificial hydraulic meter (8 artificial flow meters and 6 artificial pressure meters) are selected to observe the DMA07. This is to analyse the performance of BDLM with a different number and type of artificial flow meters and pressure meters.

Table 6.2: The combination sets of artificial flow meters and artificial pressure meters (refer to **Figure 6.1** for actual network locations)

Hydraulic Sensor Combination index (HSC)	**Artificial flow meter index**	**Artificial pressure** meter **index**
1	F2, F3, F4, F5, F7	P1, P2, P3, P4, P5, P6
2	F2, F3, F4, F5, F8	P1, P2, P3, P4, P5, P6
3	F1, F2, F3, F4, F5	P1, P2, P3, P4, P5, P6
4	F1, F4, F5, F8	P1, P2, P3, P4, P5, P6
5	F2, F3, F5, F6	P1, P2, P3, P4, P5, P6
6	F1, F2, F4, F5	P1, P2, P3, P4, P5, P6
7	F2, F4, F7	P1, P2, P3, P4, P5, P6
8	F2, F3, F6	P1, P2, P3, P4, P5, P6
9	F1, F2, F3	P1, P2, P3, P4, P5, P6

The chosen artificial flow meters in (HSC) 3, 6, and 9 are located on the most critical pipe in the DMA07 while the artificial flow meters in remaining HSCs are chosen randomly. Flow meters F1, F2, F4, F5, F3, F6, F7. F8 and F9 are the top 9 ranking links (pipes) in the ascending order. The number of artificial pressure meters remains the same in all HSCs because the UK WCs can afford to install additional pressure loggers easily compared to flow meters in a DMA.

The information relating to the distance between the artificial flow/pressure meters location and burst location based on the straight-line distance are displayed in Table 6.3.

Table 6.3: The straight-line distance (in metres) between the artificial hydraulic meters and the burst location.

Hydraulic meters		Burst location				
		B1	B2	B3	B4	B5
Flow meter	F1	167	678	783	827	994
	F2	290	475	827	766	986
	F3	352	510	783	440	660
	F4	634	211	704	528	792
	F5	651	1003	167	378	264
	F6	660	1030	264	440	238
	F7	854	924	202	167	123
	F8	616	924	158	334	299
Pressure meter	P1	422	713	1012	1056	1232
	P2	211	651	440	502	660
	P3	299	343	660	634	854
	P4	581	651	255	88	334
	P5	766	1144	343	546	414
	P6	722	1109	264	414	238

6.2.4 Simulated Pipe Burst Events

In the absence of real burst related data, the bursts are generated using equation (40) in Section 5.4.2 as pressure-dependent emitter flows (Wu, et al., 2002) located at network nodes. The emitter exponent of 0.5 is used (Lambert, 2002) while various emitter discharge coefficients have been tested to find the acceptable emitter discharge coefficient to represent the 4 realistic burst flows of varying magnitude. The 4 burst flows used in this case study are 5%, 10%, 20% and 30% of the daily average DMA demand. The bursts are simulated at 3 different time periods for 5 hours: (1) morning peak (06:00 to 11:00), (2) evening peak (16:00 to 21:00) and (3) night (00:30 to 05:30)). Artificial pipe bursts are then simulated at 5 different burst locations (refer to **Figure 6.1**). Therefore, the BDLM's ability to detect and locate bursts was tested and validated on 300 artificial pipe burst events (4 different burst magnitudes, 3

different time periods, 5 different burst locations and 5 different observation noise levels), each under 9 different HSCs totalling 2,700 cases analysed in this case study.

6.2.5 Burst Detection and Localisation Method Parameters

6.2.5.1 Offline Water Demand Forecasting Model Calibration

The difference between the inflows and outflows of the DMA01 at intervals of 15 minutes between 1st February 2015 and 14th March 2015 (5 weeks) is assumed to be the total DMA water demand. The time series of water demand data are used to develop and calibrate the WDFM offline. Based on the cross-correlation results, the WDFM was developed using a combination of water demands from time step, t, t-1, t-670, t-671 and t -1343. The MLE is used to derive the weight coefficients. The WDFM developed this way is as follows:

$$d_{t+\mathbf{1}} = 0.833d_t - 0.134d_{t-1} - 0.032d_{t-67\mathbf{0}} + 0.112d_{t-67\mathbf{1}} + 0.223d_{t-134\mathbf{3}} \quad (41)$$

where d_{t+1} is the forecast demand; d_t is the corrected demand at current time step, t; d_{t-1} is the corrected demand from 15mins ago, t; d_{t-670} is corrected demand from 167.75 hours ago, d_{t-671} is the corrected demand from a week ago and d_{t-1343} is the previous demand from 2 weeks ago.

The WDFM parameters are re-calibrated after each correction of the DMA demands (done as part of online hydraulic modelling – refer to section 3.4) to ensure the outliers do not affect the performance of the WDFM.

6.2.5.2 Offline Hydraulic Model Calibration

The DMA07 hydraulic model is calibrated offline by using the DMA01's inflows, outflows and pressure data (located at the highest elevation) between 1st February 2015 and 14th March 2015 (5 weeks). Further details on how the DMA07 model was calibrated are discussed in Section 4.4.

6.2.5.3 BDLM Parameters

The i-KF method (refer to section 3.4.) is used iteratively until either of the following pre-defined convergence/error conditions is met:

(1) all flow errors estimated as the difference between the corrected flow values from the current and previous iteration step are within a predefined tolerance value. The selected value used here is $\pm$0.25 l/s, to minimise the computational cost of the i-KF method at each time step.

(2) the defined maximum number of iterations has been reached. Using a set of training data, it took between 3 and 8 iteration steps for the i-KF method to correct the forecast demand in one-time step under normal conditions. Hence, a maximum number of 10 iterations was selected.

The flow observations from the DMA boundary (inflow/outflow) are not part of the observed flow data used to correct the forecasted DMA demands. The inclusion of inflow data tends to produce higher corrected DMA demands compared to the corrected DMA demands if inflow data are excluded during the burst period. This is because some flow meters within the DMA are either affected or not affected by the burst depending on the pipe burst's location.

Prior to the burst detection analysis, the value of $\bar{p}$-value and $\bar{r}$-*value* for each flow observation in equation (18) and (19) in section 3.4 respectively are determined by running the BDLM offline between 15th March 2015 and 28th March 2015 (2 weeks) under normal conditions (i.e. without bursts). Table 6.4 shows the initial $\bar{p}$-values and $\bar{r}$-*values* estimated for DMA07. The $\bar{p}$-*value is* the average weekly normalised error between the observed and predicted flow observations and while the $\bar{r}$-*value* is the average weekly normalised error between the observed and corrected flow observations.

Table 6.4: The value of $\bar{p}$-*value* and $\bar{r}$-*value* for each flow observations

Flow meter ID	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**
$\bar{p}$-value	0.051	0.047	0.053	0.050	0.049	0.048	0.013	0.053
$\bar{r}$-value	0.020	0.016	0.023	0.018	0.027	0.024	0.004	0.027

The BDLM makes use of two detection analyses: ISDA and PDSA SPC-based Control Charts/Rules to detect DMA burst pipes. The SPC-based Control Charts review the burst detection metric values against a set of Control Rules (refer to Section 5.3). The burst detection metric is the average of 4 recent flow/pressure residuals. Using 4 recent flow/pressure residuals can be deemed as a good measure of short-term momentum to reflect recent residual variations. As it can be seen from Figure 6.2, MAR with a window size of 4-time steps (MA-4) is more sensitive to small flow residual changes when compared to other MARs (i.e., MA-16, MA-48 and so on).

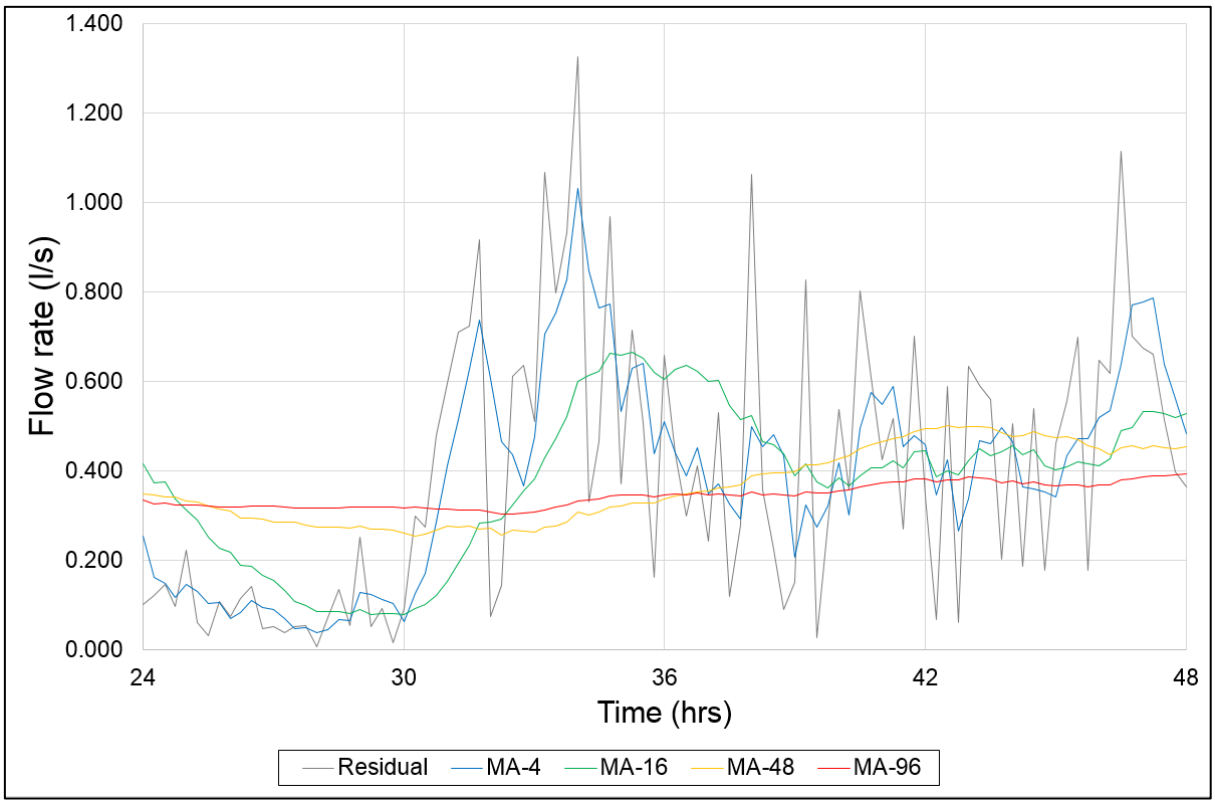


Figure 6.2: Example of flow residuals chart (MA-X is the moving average of the last X flow residuals)

The ISDA/PSDA SPC-based Chart Control Limits (mean and standard deviation (sigma) value) in this case study are estimated using the previous one week of burst detection metric data at each time step. One-week data comprises 672 time-series data (since flow/pressure data are collated every 15 minutes).

The BDLM parameters are determined offline. In the BDLM methodological step 2 (refer to Section 5.4.1), the nodes relating to the top two flow meters affected by the

burst are used for burst localisation analysis. This is because the sensitivity matrix shows that flows at observation points are more sensitive to bursts when compared to pressures at observation points. The flow/pressure threshold value *(in step 2, Section 5.4.2)* to determine if simulated burst at a node has an impact on the flow/pressure meters is the average flow/pressure residual between the hydraulic state under normal and abnormal (burst) conditions. The flow/pressure residual errors above their respective thresholds are considered to be sensitive to the bursts and are used to identify the possible burst area.

The selected value of flow increment (*in step 5, Section 5.4.3*) is 0.2 l/s for the DMA demand and burst flow estimation process (iteration). This increment operator is chosen as small as possible to ensure the burst flow is accurately determined bearing in mind the computational constraints associated with the iterative procedure used. In the BDLM methodology step 4, 10% of the DMA nodes (i.e. a total of 54 nodes here) are used as the maximum number of nodes to formulate the burst area.

The next section discusses the sensitivity data used to set the Control Rules for the ISDA and PSDA SPC-based Control Charts.

6.2.5.4 The development of ISDA and PSDA Control Rules

For the BDLM to reliably the pipe bursts in real-time, the number of Control Rule violations has to be determined. The ISDA/PDSA Control Rules are determined via detailed analysis of MAR/P-MAR data. To obtain the MAR/P-MAR data, each real-life DMA was simulated via BDLM steps 1 to 4 that were run between March 2013 and mid-May 2013. While the January and February 2013 flow/pressure data were used to obtain relevant BDLM parameters.

The following procedure used to determine ISDA/PDSA Control Rules are as follows:

1. Run BDLM under normal conditions to obtain flow/pressure MAR/P-MAR data.
2. Compute the mean and standard deviations of the obtained flow/pressure MAR/P-MAR data.

3. Estimate the acceptable lower and upper control limits using standard deviation of their data and user defined values, *N* (refer to Section 5.3 and Figure 6.3). Figure 6.3 shows the graphical representation of a Control Chart based on the data mean and standard deviations and selected lower and upper control limits. The upper and lower Control Limits are calculated by adding or subtracting the product of a multiplying factor and absolute standard deviation to/from their respective mean.
4. Count the number of selected Control Rule violations.

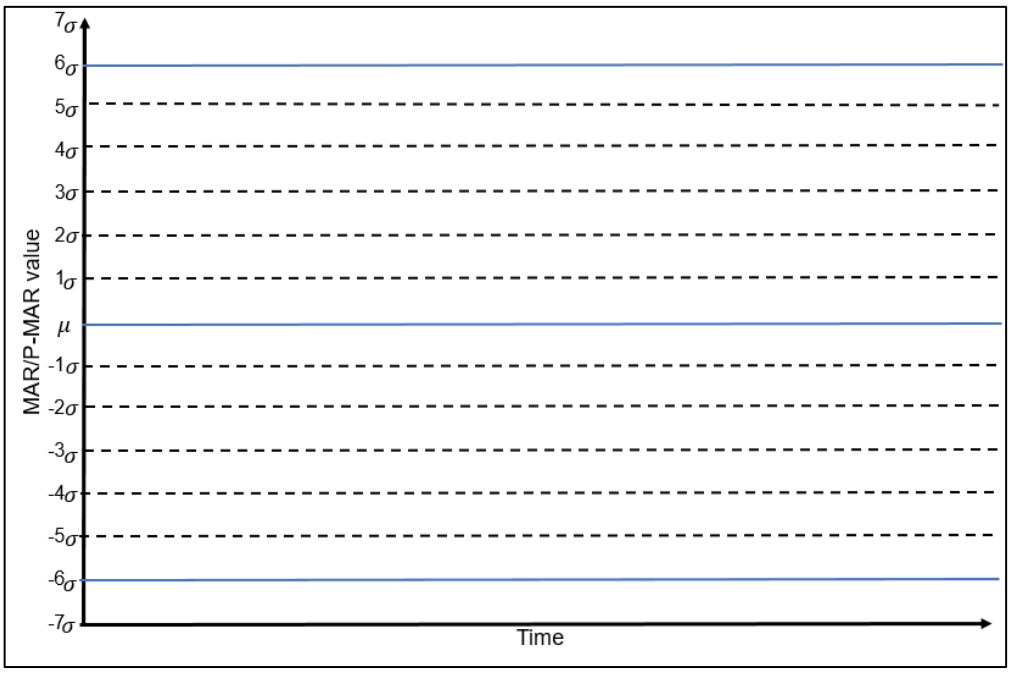


Figure 6.3: Typical Control Chart divided into zones for different control criteria

The generic Control Rules shown in Section 5.3 were develop after an offline sensitivity analysis of 44 real-life flow/pressure data under normal conditions. These flow/pressure data were provided by UU Logger Management personnel as they have no missing data. Among the 44 hydraulic meters data, 16 and 28 are from real life flow meter and pressure meters respectively. These real-life data are obtained from 12 different real-life DMAs and 6 of these real-life DMAs are displayed in the case study in Chapter 4. The 44-real-life flow/pressure data and 12 real-life DMAs are obtained from United Utilities database. The obtained data start from January 2013 to May 2013 (5 months). The results displayed in this case study are limited to the flow/pressure data for reduced WSZ01 water network due to UU's data confidentiality.

Table 6.5 and Table 6.6 show the amount (average percentage) of flow and pressure MAR data that fall within their respective control limits respectively for each DMA. The results in Table 6.5 and Table 6.6 are related to WSZ01 network described in Chapter 4. Table 6.5 and Table 6.6 also show that majority of MAR data points fall within 1σ from the centreline. However, an average of 94% of data points fall within 2σ from the centreline, hence no Control Rules were set to avoid raising many potential false alarms.

Table 6.5: Flow MAR sensitivity results

Control Limits	DMA01	DMA02	DMA04	DMA05	DMA06	DMA07	Overall average*
Within 1σ of the centreline	77.1%	75.8%	75.2%	72.5%	78.1%	76.9%	76.7%
Between 1σ and 2σ from the centreline	16.4%	17.5%	18.8%	21.8%	14.8%	16.9%	17.7%
Between 2σ and 3σ from the centreline	4.4%	5.2%	4.4%	5.1%	5.6%	4.4%	4.1%
Between 3σ and 4σ from the centreline	1.6%	1.4%	1.6%	0.5%	1.3%	1.4%	1.2%
Between 4σ and 5σ from the centreline	0.4%	0.1%	0.0%	0.1%	0.1%	0.3%	0.3%
Between 5σ and 6σ from the centreline	0.1%	0.1%	0.0%	0.1%	0.1%	0.1%	0.1%
Any single MAR point falls outside the 6σ limit from the centreline	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%

*average of 16 real-life flow meters

Table 6.6: Pressure MAR sensitivity results

Control Limits	DMA01	DMA02	DMA04	DMA05	DMA06	DMA07	Overall average*
Within 1σ of the centreline	83.4%	83.1%	68.2%	81.2%	81.8%	76.6%	81.3%
Between 1σ and 2σ from the centreline	11.3%	13.8%	28.2%	13.3%	11.5%	17.5%	15.1%
Between 2σ and 3σ from the centreline	5.3%	2.9%	3.5%	4.5%	5.2%	4.1%	2.6%
Between 3σ and 4σ from the centreline	0.0%	0.4%	0.1%	1.5%	1.4%	1.7%	0.8%
Between 4σ and 5σ from the centreline	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.1%
Between 5σ and 6σ from the centerline	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.1%
Any single MAR point falls outside the 6σ limit from the centreline	0.0%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%

*average of 28 real-life pressure meters

Table 6.7 and Table 6.8 show the amount (average percentage) of P-MAR data that fall within their respective control limits for each DMA. The results in Table 6.7 and Table 6.8 are related to the reduced WSZ01 network described in Chapter 4. During the sensitivity analysis of P-MAR values, 2 artificial flow meters and 2 pressure meters were added to each DMA. The DarwinSampler (Bentley software) (Wu, 2013) was used to determine the optimal locations of the added artificial meters. The purpose of the additional artificial data is to develop to a generic set of PSDA Control Rules for any water networks. Hence 2 artificial flow/pressure data were generated via their respective hydraulic model under normal conditions for each DMA.

Table 6.7: Flow P-MAR sensitivity results

Control Limits	DMA01	DMA02	DMA04	DMA05	DMA06	DMA07	Overall average*
Within 1σ of the centreline	88.6%	99.8%	87.1%	84.2%	82.0%	86.6%	86.2%
Between 1σ and 2σ from the centreline	6.2%	0.2%	7.8%	9.6%	12.6%	8.0%	9.4%
Between 2σ and 3σ from the centerline	3.0%	0.0%	2.8%	3.5%	3.8%	3.3%	3.0%
Between 3σ and 4σ from the centerline	1.4%	0.0%	1.3%	1.9%	0.8%	1.3%	0.8%
Between 4σ and 5σ from the centerline	0.6%	0.0%	0.8%	0.4%	0.5%	0.5%	0.3%
Between 5σ and 6σ from the centerline	0.2%	0.0%	0.3%	0.3%	0.1%	0.2%	0.2%
Any single P-MAR point falls outside the 6σ limit from the centerline	0.2%	0.0%	0.2%	0.3%	0.1%	0.2%	0.2%

*Average from 92 sets of flow P-MAR data

DMA01 has 5 flow meters (3 real-life flow meters and 2 artificial flow meters) hence 10 sets of 2-flow P-MAR values, 10 sets of 3-flow P-MAR values, 3 sets of 4-flow P-MAR values and 1 set of 5-flow P-MAR values. DMA02 has 4 flow meters (2 real life flow meters and 2 artificial flow meters) hence 6sets of 2-flow P-MAR values, 3 sets of 3-flow P-MAR values and 1 set of 4-flow P-MAR values. DMA04, DMA05, DMA06 and DMA07 have 3 sets of 2-flow P-MAR values and 1 set of 3-flow P-MAR values due to (1 real-life flow meters and 2 artificial flow meters).

The overall average in Table 6.7 is based on 92 sets of flow P-MAR data include 56 sets of 2-flow MAR values, 26 sets of 3-flow MAR values, 8 sets of 4-flow MAR values and 2 sets of 5-flow MAR values.

Table 6.8: Pressure P-MAR sensitivity results

Control Limits	DMA01	DMA02	DMA04	DMA05	DMA06	DMA07	Overall average*
Within 1σ of the centreline	82.8%	84.9%	84.0%	81.8%	84.1%	85.0%	80.8%
Between 1σ and 2σ from the centreline	10.2%	7.4%	7.5%	15.2%	8.1%	14.5%	12.1%
Between 2σ and 3σ from the centerline	5.4%	6.1%	7.3%	3.0%	7.0%	0.7%	7.0%
Between 3σ and 4σ from the centerline	1.5%	1.4%	0.8%	0.1%	1.1%	0.1%	0.2%
Between 4σ and 5σ from the centerline	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%
Between 5σ and 6σ from the centerline	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%
Any single P-MAR point falls outside the 6σ limit from the centerline	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

*Average from 201 sets of pressure MAR data

DMA01 has 4 flow meters (2 real-life pressure meters and 2 artificial pressure meters) hence 6sets of 2-pressure P-MAR values, 3sets of 3-pressure P-MAR values and 1 set of 4-pressure P-MAR values. DMA02, DMA04, DMA05, DMA06 and DMA07 have 3 sets of 2-pressure P-MAR values and 1 sets of 3- pressure P-MAR values due to (1 real-life pressure meters and 2 artificial pressure meters).

The overall average in Table 6.8 is based on 201 sets of pressure P-MAR data including 99 sets of 2-pressure MAR values, 58 sets of 3-pressure MAR values, 29 sets of 4-pressure MAR values, 12 sets of 5-pressure MAR values and 3 sets of 6-pressure MAR values.

Table 6.7 and Table 6.8 show that majority of data points fall within 1σ of the centreline. An average of 92% of P-MAR data points fall within 2σ of the centreline, no control rules were developed to avoid raising many potential false alarms.

Table 6.9 shows the frequency of flow/pressure MAR consecutive data points for each control limit.

Table 6.9: The frequency of flow/pressure MAR consecutive data points for each control limit

Number of consecutive data points within the control limits	**2**	**3**	**4**	**5**	**6**	**8**	**10**
Between 2σ and 3σ from the centerline	670	282	103	30	10	6	2
Between 3σ and 4σ from the centerline	203	67	31	11	3	1	0
Between 4σ and 5σ from the centerline	47	16	5	1	0	1	0
Between 5σ and 6σ from the centerline	9	3	1	0	0	0	0
Any single MAR point falls outside the 6σ limit from the centerline	3	1	1	0	0	0	0

As it can be seen from Table 6.9, the frequency decrease when the number of flow/pressure MAR consecutive data points decrease. Table 6.9 is used to set the generic Control Rules for ISDA SPC- based charts for both flow and pressure meters.

Table 6.10 shows the frequency of flow/pressure P-MAR consecutive data points for each control limit.

Table 6.10: The frequency of flow/pressure P-MAR consecutive data points for each control limit

Number of consecutive data points within the control limits	**2**	**3**	**4**	**5**	**6**	**8**	**10**
Between 2σ and 3σ from the centreline	378	117	44	16	8	4	3
Between 3σ and 4σ from the centreline	125	29	24	11	3	1	0
Between 4σ and 5σ from the centreline	45	9	4	1	2	0	0
Between 5σ and 6σ from the centreline	20	3	1	1	0	0	0
Any single MAR point falls outside the 6σ limit from the centreline	9	1	1	2	0	0	0

As it can be seen from Table 6.10, the frequency increases when the number of flow/pressure P-MAR consecutive data points increases. The number of flow/pressure P-MAR consecutive data points for each control limit are determined when the frequency first reach less than 2. For example, for control limits between 5σ and 6σ from the centreline, the number of flow/pressure P-MAR consecutive data points would be 3. This selection process is applied to the ISDA SPC-based Control Rules.

Table 6.10 is used to suggest the generic Control Rules for PSDA SPC-based charts for both flow meters and pressure meters. Even though, artificial flow/pressure data have been used to develop PSDA Control Rules under normal condition. However, the above results and PSDA Control Rules are very similar to the results based on 8 real-life DMAs that have 5/6 real-life pressure meters. These results are not presented in this thesis due to UU's data confidentiality reason.

6.2.6 Testing of Burst Detection Capabilities in the DMA07 Water Network

The BDLM was run between 3rd and 30th May 2015 and tested on 300 artificial pipe burst events and under 9 different HSCs, as explained in detail in Section 6.2.4. The performance of the BDLM is evaluated based on the number of detected bursts and false alarms raised. In addition, the BDLM capabilities were also evaluated based on the time taken to detect the artificial pipe bursts.

Figure 6.4 shows the summary of the burst detection rates between 3rd and 30th May 2015 using different HSCs, as shown in Table 6.2. As it can be seen from Figure 6.4, BDLM detected more than 66% of 300 artificial pipe burst events regardless of the HSC used. HSC with 5, 4 and 3 artificial flow meters resulted in BDLM detecting bursts with 82%, 68% and 66% detection rates, respectively.

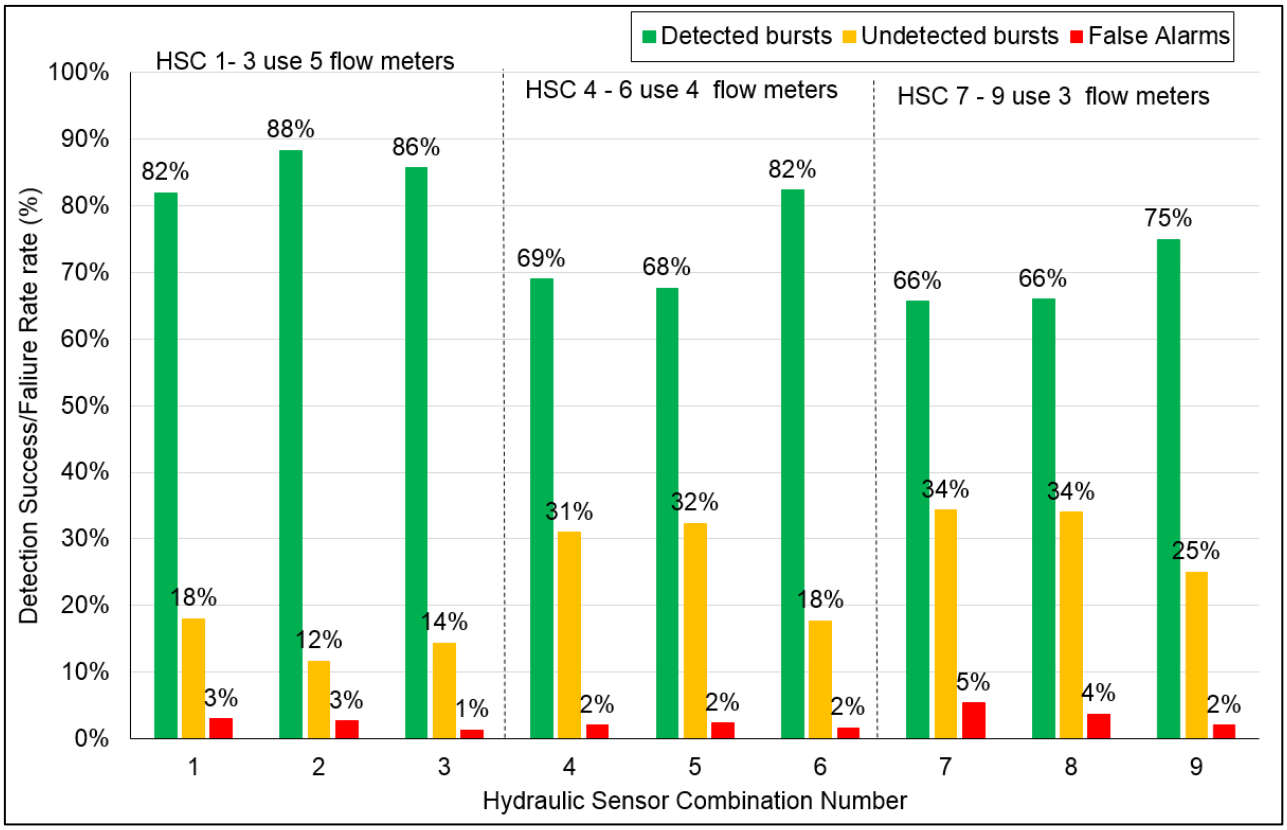


Figure 6.4: The overall burst detection rates between 3rd and 30th May 2015

Figure 6.4 also reveals that a reduction of a number of artificial flow meters is likely to reduce the BDLM bursts detection success rate and/or increase the number of false alarms. The results in Figure 6.4 further show that BDLM with HSC 9 detected more pipe bursts compared to BDLM with HSC 4 and 5 (refer to Figure 6.1 in Section 6.2.2 and Table 6.2 in Section 6.2.3). Yet, HSC 9 has 3 flow meters and HSC 4 and 5 used 4 flow meters. BDLM with HSC 6 outperformed HSC 1 due to lower false alarm rate despite both HSCs have the same burst detection success rate (82%). The above indicates that the number, location and type of the hydraulic meters used is critical for monitoring a DMA and detecting pipe burst events.

Figure 6.5 shows the burst detection rate for 3 different time periods between 3rd and 30th May 2015 for each HSC. The number of pipe burst events is 60 including different 5 burst magnitudes and 5 observation noises.

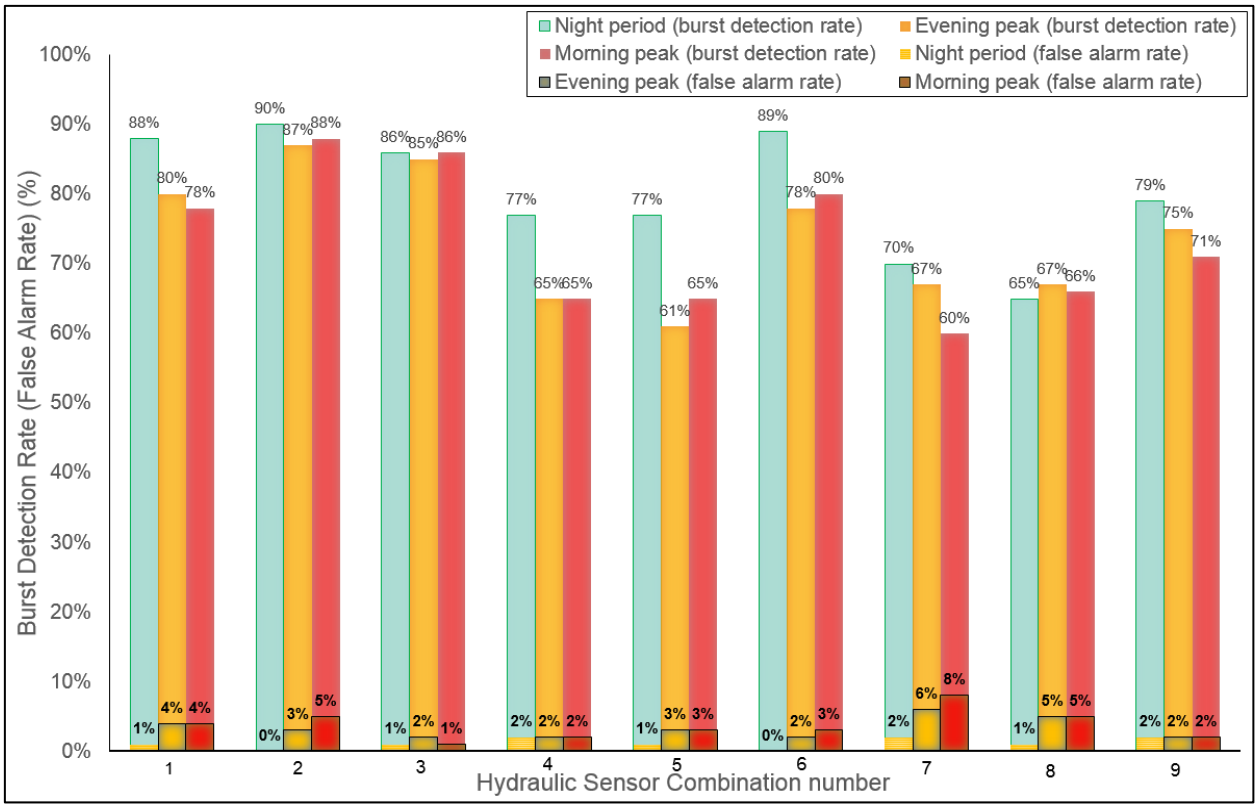


Figure 6.5: Alarm success rate for burst detections at different time periods between 3rd and 30th May 2015

The results illustrate that artificial pipe bursts at night are more likely to be detected by the BDLM compared to other time periods (morning and evening). The burst detection rate during the night is 65% which is larger than the minimum of 60% of pipe burst events detected in the morning and evening. The reason for higher detection rate at night is the small consumption during that time which makes it easier for the BDLM to distinguish bursts from actual consumption.

Table 6.11 shows the burst detection rates for different burst magnitudes between 3rd and 30th May 2015. As it can be seen from Table 6.11, at least 91% of bursts that have magnitudes above 20% of the average daily DMA demand are detected by the BDLM. The burst detection rate decreases when the burst magnitudes decrease. This indicates that detecting small-sized artificial pipe bursts (i.e., 10% of average daily demand) is likely to be difficult. In the case of using BDLM with HSC 4, 5, 7 and 9, it is difficult to detect pipe burst events below 10% of the average daily DMA demand due to poor coverage of the DMA07.

Table 6.11: The burst detection rates for different burst magnitudes between 3rd and 30th May 2015 (number of burst events for each HSC is 75)

Hydraulic Sensor Combination (HSC)	**5% of average daily DMA demand**	**10% of average daily DMA demand**	**20% of average daily DMA demand**	**30% of average daily DMA demand**
HSC 1	42 (56%)	54 (72%)	75 (100%)	75 (100%)
HSC 2	54 (72%)	61 (81%)	75 (100%)	75 (100%)
HSC 3	44 (59%)	63 (84%)	75 (100%)	75 (100%)
HSC 4	20 (27%)	40 (53%)	72 (96%)	75 (100%)
HSC 5	18 (24%)	36 (48%)	74 (99%)	75 (100%)
HSC 6	36 (48%)	61 (81%)	75 (100%)	75 (100%)
HSC 7	10 (13%)	39 (52%)	73 (97%)	75 (100%)
HSC 8	19 (25%)	36 (48%)	68 (91%)	75 (100%)
HSC 9	17 (23%)	58 (77%)	75 (100%)	75 (100%)

Table 6.11 also shows all the HSCs detected all the pipe burst events with a burst magnitude (30% of the average daily DMA demand). BDLM with HSC 1 – 3 have high burst detection rate of artificial pipe burst events with burst magnitude below 20% of the average daily DMA demand. This is due to 5 artificial flow meters were used. In the case of using HSC 7, it is difficult to detect pipe burst events below 10% of the average daily DMA demand due to low number of (or 3) artificial flow meters.

Figure 6.6 shows the burst detection rate for different observational noises between 3rd and 30th May 2015.

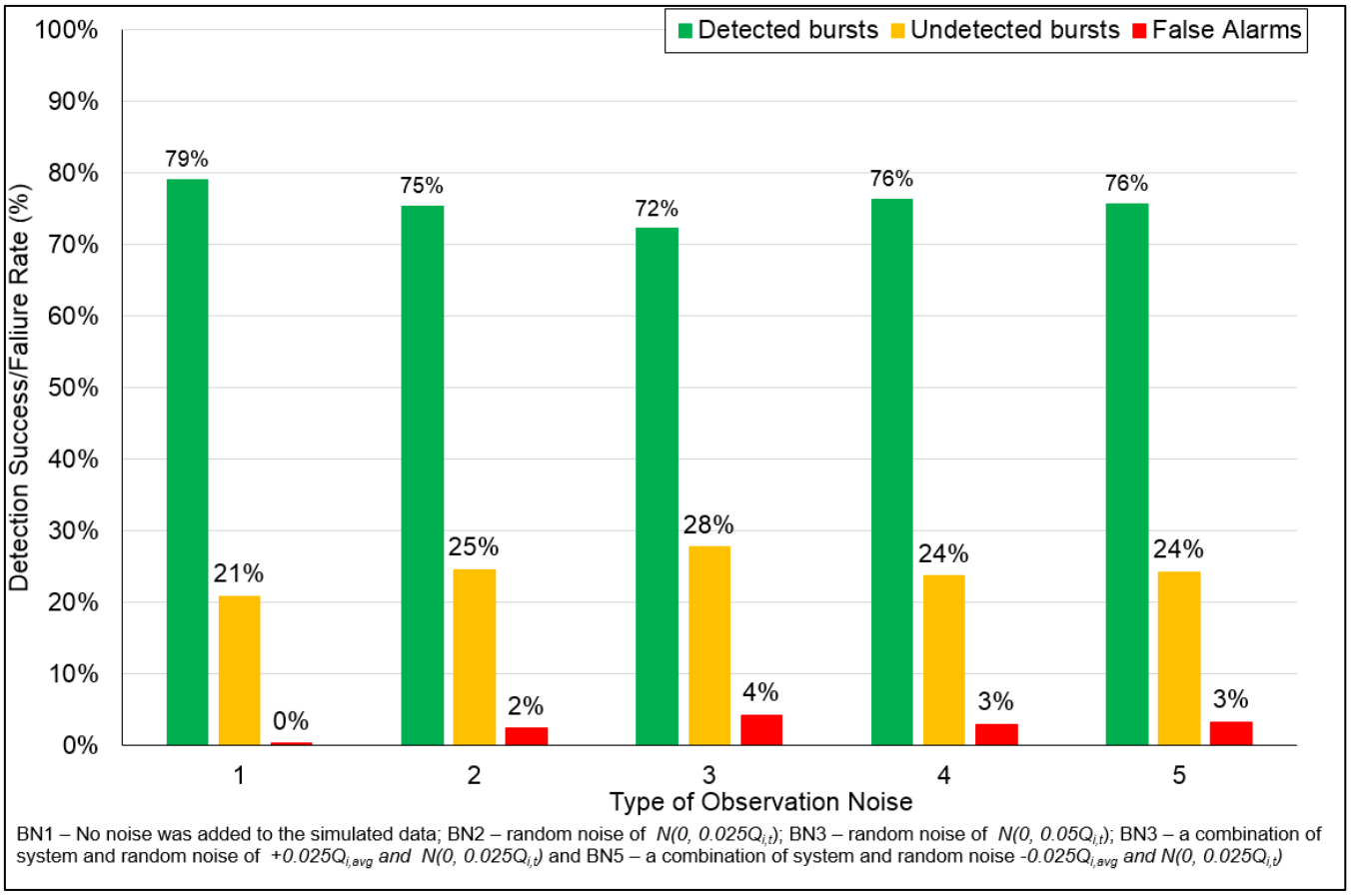


Figure 6.6: Alarm success rate for burst detections at different observational noise cases between 3rd and 30th May 2015

As it can be seen from Figure 6.6, the developed BDLM detected 79% of 540 pipe burst events (between 3rd and 30th May 2015) in the first case, i.e. when the flow and pressure observation are '*perfect observations*'. The 540 artificial pipe burst cases analysed are based on 4 different burst magnitudes, 3 different burst time periods, 5 different burst locations and 9 HSCs. Each of these 540 cases was analysed against 5 different noise cases (refer to Figure 6.6) totalling 2,700 cases analysed. The observation noise (refer to Table 6.1) in cases 2-5 was added to '*perfect observations'*.

Figure 6.6 shows that burst detection rates decrease when the observational noise level increases (from observational noise case 1 to 3). This is because additional measurement noise tends to affect changes in flow and pressure due to bursts, especially for lower magnitude bursts. Indeed, the data noise on low burst magnitudes (i.e., 5% of the average daily DMA demand) disturb the MAR/P-MAR data to develop a consecutive pattern to meet the developed ISDA/PSDA Control Rules. This

interference of develop consecutive pattern makes flow meters and pressures insensitive to low burst magnitudes especially when there is a large data noise. Still, the results shown in Figure 6.6 indicate that BDLM has a relatively good capability of detecting pipe burst events despite noisy flow/pressure observations.

Table 6.12 shows the burst detection rates for different burst locations shown in Figure 6.1, Section 6.2.2.

Table 6.12: The burst detection rates for individual burst locations between 3rd and 30th May 2015 (total number of bursts in each time period is 60)

Hydraulic Sensor Combination (HSC)	**Burst location 1**	**Burst location 2**	**Burst location 3**	**Burst location 4**	**Burst location 5**
HSC 1	47 (78%)	51 (85%)	47 (78%)	46 (77%)	55 (92%)
HSC 2	52 (87%)	54 (90%)	53 (88%)	51 (85%)	55 (92%)
HSC 3	55 (92%)	49 (82%)	51 (85%)	46 (77%)	56 (93%)
HSC 4	44 (73%)	39 (65%)	35 (58%)	37 (62%)	52 (87%)
HSC 5	32 (53%)	45 (75%)	37 (62%)	40 (67%)	49 (82%)
HSC 6	53 (88%)	48 (80%)	47 (78%)	43 (72%)	55 (92%)
HSC 7	42 (70%)	28 (47%)	42 (70%)	36 (60%)	49 (82%)
HSC 8	38 (63%)	37 (62%)	34 (57%)	42 (70%)	47 (78%)
HSC 9	49 (82%)	41 (68%)	45 (75%)	37 (62%)	53 (88%)

As it can be seen from Table 6.12, BDLM with HSC 2 is the best performing HSC in all burst locations and when compared to other HSCs. It detected above 85% of pipe burst events at each burst location compared to other HSCs. This is because flow meters F3 and F5 are at located at network locations that are most sensitive to 5 burst locations. The combination of F3 or F5 with other flow meters in PSDA Control Charts also plays a factor in detecting pipe burst events at 5 locations.

Table 6.12 also shows that BDLM has detected more pipe burst events at burst location 5 compared to other burst locations. This is because the observed flow data from the unaffected flow meters contribute in outputting corrected DMA demands (via the data correction) close to ‘normal’ corrected DMA demands. These ‘’normal’ corrected DMA

demands are inserted into the hydraulic model which output lower predicted flow rates at the same flow meters' location that are affected by the bursts. Hence, a larger discrepancy between the predicted and observed flow data at the same flow meters' location that are affected by the bursts. This large discrepancy data easily violates both ISDA/PSDA SPC-based Control Rules.

The results in Table 6.12 also show the BDLM is capable of detecting pipe burst at various locations if there is a sufficient number and distribution of hydraulic (flow and pressure) meters in a DMA. For example, BDLM with HSC 2 (the best performing HSC among 9 HSCs) detected most pipe bursts at 3 different burst locations (location 2, 3 and 4). BDLM with HSC 3 detected most pipe bursts at 2 different burst locations (location 1 and 5).

Figure 6.7 shows an average number of time steps required to detect pipe burst events for different burst magnitudes at 3 different time periods.

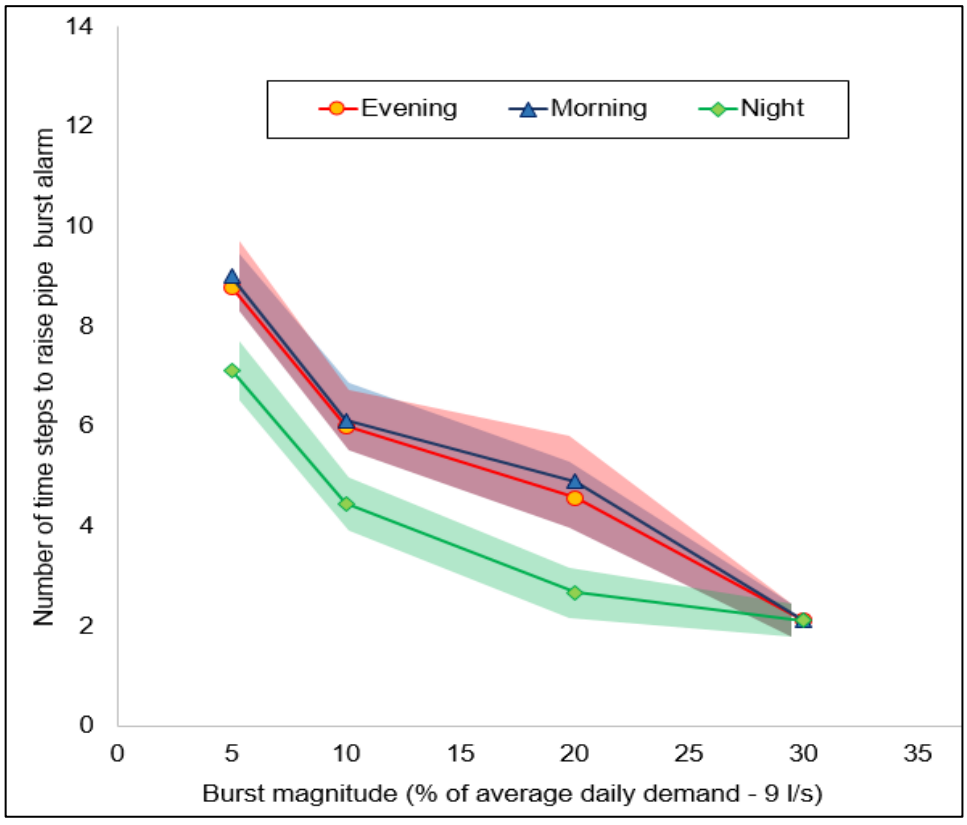


Figure 6.7: Average number of time steps to detect a pipe burst event. The shaded areas illustrate the range between the upper and lower limit (in time steps) to detect a pipe burst event.

Figure 6.7 shows that all pipe burst magnitudes that are equal to 30% average of average demand in the DMA are detected in approximately 2-time steps (i.e. within 30 mins) regardless of the time period when the burst occurred. The BDLM detects medium-sized bursts (20% to 30% of average daily DMA demand) within 6-time steps (i.e. within 1 hours 30 mins). The increased number of time steps required when compared to the large burst case is due to the fact that it is more difficult to detect smaller bursts hence additional time step are required to confirm these. The pipe burst events below 10% of average daily DMA demand during the morning and evening peak are detected between 6 and 9 time steps (1 hour 30mins – 2hours 15 mins). Figure 6.7 also shows that all bursts are detected earlier during the night when compared to other time periods (morning and evening peak). This is because flow due to consumption during the day tends to mask the bursts lot more than during the night time.

Comparing the performance of the developed BDLM to other model-based burst detection technologies are difficult to make due to the size and type of the water network models used. Also, most of the developed model-based burst detection technologies results are too few to be compared to the developed BDLM in this thesis. However, the development of a new BDLM in this thesis is relevant because it considers the uncertainty in WDS observations and hydraulic model for online application. The running time to calibrate the BDLM take less than an hour depending on a DMA configuration and volume of data used to calibrate the burst detector. BDLM takes nearly one minute to correct demand forecast every time and less than five seconds to raise burst alarms. The developed BDLM can also be used to suggest the preferable hydraulic sensors combination for an online burst detection.

False Burst Alarms Analysis

Table 6.13 shows the flow meters that are likely to generate false alarms during non-burst period.

Table 6.13 shows F2, F2, F6, F7 and F6 in HSC 1, HSC 2, HSC 5, HSC 7 and HSC 8 respectively were the primarily reason why the false alarms are generated compared to other hydraulic meters. When the flow meters in Table 6.13 are exempted from the

burst detection analysis, the burst detection rate remains unchanged and the number of false alarms decreased. However, the flow data from the exempted flow meters were still part of the flow observation data that correct the predicted DMA demand at every time step.

Table 6.13: False alarms generation result

HSC	**Flow meter**	**Number of false alarms generated by the flow meter**	**Total number of false alarms generated by the HSC**
1	F2	5	9
2	F2	6	8
5	F6	4	7
7	F7	7	16
8	F6	7	11

The results from the detection analysis also reveals that increasing the number of artificial flow meters can still generate a few false burst alarms. This means the location of the hydraulic meters is another critical factor to effectively monitor a DMA. Therefore, results from the meters performance during the detection analysis can be used to suggest the number and location of the hydraulic meters for burst detections.

A comparison of the closest and most sensitive hydraulic meters for detecting pipe bursts

Further analysis was carried out to analyse the relationship between the closest and most sensitive hydraulic meters to the burst locations. The meters that have largest difference between the predicted value and observed signal value during the burst period is regarded as the most sensitive hydraulic meters (flow/pressure). The hydraulic meter with the shortest flow direction (or path) between its location and burst location is regarded as the closest hydraulic meters.

Figure 6.8 displays a comparison of the closest and most sensitive hydraulic meters to the burst locations.

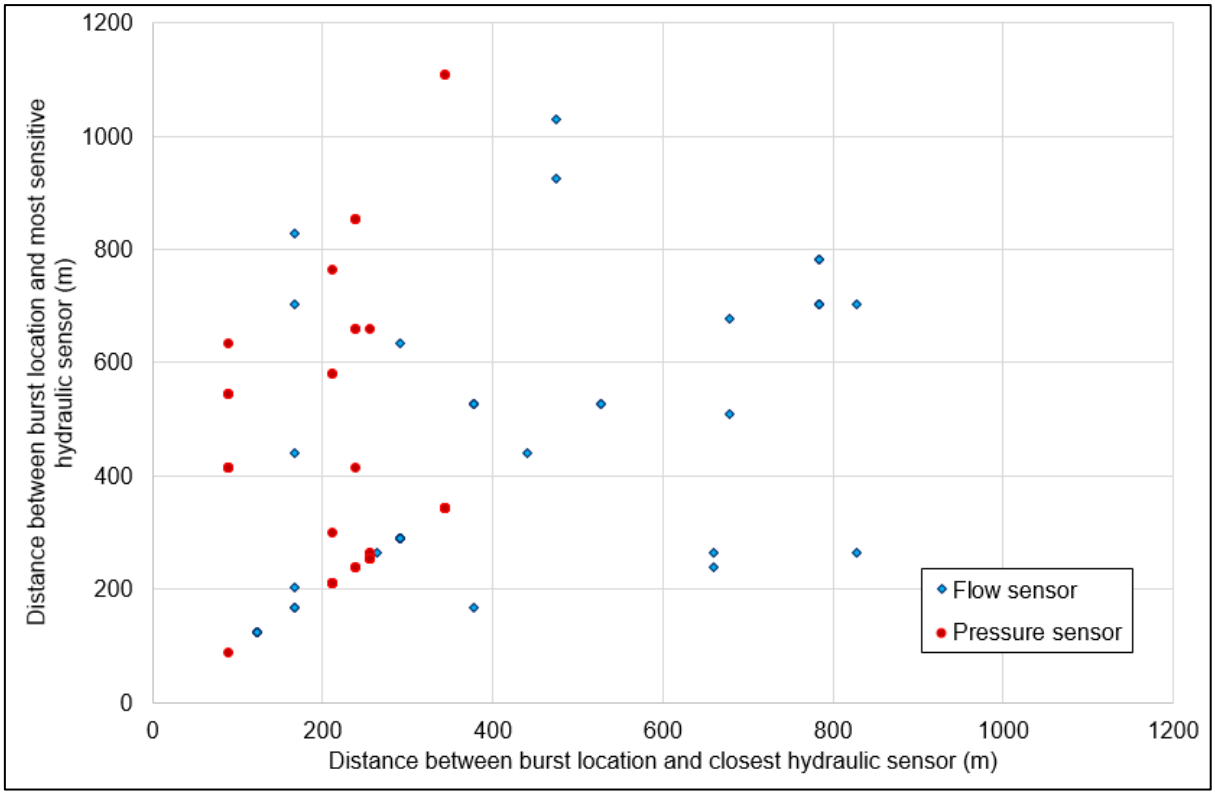


Figure 6.8: A comparison between the closest and most sensitive hydraulic meters to the burst locations.

As it can be seen from Figure 6.8, the most sensitive hydraulic meters usually have the largest burst metric value. The chart shows no strong correlation between the closest and most sensitive hydraulic meters. Almost 48% of the artificial flow meters match the closest flow meters to the burst locations while 42% of the artificial pressure meters match the closest pressure meters to the burst locations. This is due to the complexity of the DMA configuration. Therefore, using the most sensitive hydraulic meter may not be a good indicator to suggest the approximate area of the burst locations.

The BDLM can assist a leakage technician to locate bursts if sufficient number of hydraulic meters is well distributed across the chosen. Hence, a burst localisation methodology is required to complement the developed BDLM to locate bursts efficiently and in timely fashion. Therefore, the next section reviews the performance of the developed BDLM in term of locating the artificial burst events.

6.2.7 Testing of Burst Localisation Capabilities in the DMA07 Water Network

In the data analyses carried out here, the methodology behind the BDLM is tested on the artificial pipe bursts. The BDLM capabilities were also evaluated based on the approximated burst areas output. The BDLM was tested on with 3 HSCs (HSC2, HSC 6 and HSC 9). The three HSCs were selected because they were best performing HSC for detection with 5, 4 and 3 artificial flow meters respectively.

Table 6.14 shows the total number of burst locations located within the approximated area and number of burst locations among the top 10 recurring candidate burst nodes. It is pertinent to mention that 52 nodes (or 10%) of the total DMA model nodes were used to form the approximate burst area. The top 52 recurring candidate burst nodes were determined after 6-time steps (1hour 30 mins) since the initial burst alarm. Out of the top 52 recurring candidate burst nodes, the top 10 ranked candidate burst nodes are used to indicate likely location of the burst events.

Table 6.14: Number of burst locations located

HSC	Number of detected bursts*	Number of burst locations within the approximated area**	Number of burst locations among top 10 recurring candidate burst nodes**
HSC2	265 (88%)	223 (74%)	184 (69%)
HSC6	247 (82%)	193 (64%)	161 (65%)
HSC9	225 (75%)	167 (56%)	114 (51%)

*Proportion of total artificial pipe burst events (300)

**Proportion of pipe burst events detected

As it can be seen from Table 6.14, a minimum of 74% of artificial burst events detected were located within the approximated area. Whilst a minimum of 51% of artificial burst events detected were among the top 10 recurring candidate burst nodes.

Table 6.15 to Table 6.17 show the number of burst locations within the top 10 recurring nodes in accordance with observational noise type, burst magnitude, burst period and burst location respectively.

Table 6.15: The number of burst locations within the top 10 recurring nodes in accordance with observational noise type (total number of burst events is 300)

HSC	Observational Noise 1 (perfect data)	Observational Noise 2	Observational Noise 3	Observational Noise 4	Observational Noise 5
HSC2	48 (80%)	44 (73%)	28 (47%)	36 (60%)	38 (63%)
HSC6	49 (82%)	39 (65%)	14 (23%)	31 (52%)	28 (47%)
HSC9	33 (55%)	28 (47%)	11 (18%)	21 (35%)	21 (35%)

Table 6.16: Number of burst locations within the top 10 recurring nodes in accordance with burst magnitude (total number of artificial burst events is 300)

HSC	BM1- 5% of average DMA daily demand	BM2 - 10% of average DMA daily demand	BM3 - 20% of average DMA daily demand	BM4 - 30% of average DMA daily demand
HSC2	18 (24%)	32 (43%)	66 (88%)	68 (91%)
HSC6	12 (16%)	27 (36%)	54 (73%)	68 (91%)
HSC9	4 (5%)	16 (21%)	38 (51%)	56 (75%)

Table 6.17: Number of burst locations within the top 10 recurring nodes in accordance with burst periods (total number of artificial burst events is 300)

HSC	Morning (6:00am – 11:00am)	Evening (04:00pm – 9:30pm)	Night (0:30am – 5:30am)
HSC2	51 (51%)	54 (54%)	79 (79%)
HSC6	40 (40%)	43 (43%)	78 (78%)
HSC9	24 (24%)	22 (22%)	68 (68%)

Table 6.15 shows increasing level of noise (from noise type 1 to 3) do affect the performance of BDLM to locate the pipe burst events. However, HSC 2 perform better than other two HSCs due to 5 flow meters. This indicates that, in this case, more than 4 flow meters and 6 pressure meters are required to minimise the impact of noisy hydraulic data.

Table 6.16 demonstrates the capability of BDLM to locate more pipe burst events as the burst magnitude increase. Table 6.17 indicates BDLM locate more pipe burst events in the night compared to other periods (morning and evening). Table 6.15 to

Table 6.17 prove the BDLM works better with more flow meters as HSC2 located more artificial pipe bursts compared to the other 2 HSCs (HSC 6 and HSC 9).

Table 6.18 shows the BDLM with a specific HSC determines the location success rate of each pipe burst events. BDLM with HSC 2 located burst event 1 and 5 compared to the other two HSCs. Whilst BDLM with HSC 6 located more burst events at burst location 2, 3, and 4. This proves the location of the hydraulic meters is another critical factor to effectively locate pipe bursts in a DMA.

Table 6.18: Number of burst locations within the top 10 recurring nodes in accordance with burst locations (total number of artificial burst events is 300)

HSC	Burst location 1	Burst location 2	Burst location 3	Burst location 4	Burst location 5
HSC2	44 (73%)	40 (67%)	34 (57%)	27 (45%)	39 (65%)
HSC6	31 (52%)	39 (65%)	28 (47%)	32 (53%)	31 (52%)
HSC9	19 (32%)	24 (40%)	18 (30%)	22 (37%)	31 (52%)

Table 6.19 shows the average simulated and estimated burst flows

Table 6.19: The average simulated and estimated burst flows

Average simulated burst flow (l/s)	Average estimated burst flow (l/s)	Absolute Difference (l/s)
0.44	0.63	0.19
0.89	0.73	0.16
1.83	1.44	0.39
2.64	3.09	0.45

As it can be seen from Table 6.19, the estimated average burst flows are close to the average simulated burst flows.

Figure 6.9 to Figure 6.11 show the burst areas (shown in sky blue) identified for simulated bursts with a flow of 2.7 l/s taking place during the night, morning and evening periods, respectively. A total of 5 different burst locations are also shown in the figures. As it can be seen from the below figures, each burst assumed is located

by the BDLM within the identified burst area (based on top 52 recurring candidate burst locations). A set of top 10 recurring candidate burst locations (yellow dots with green outlines) are displayed for each of the 5 bursts simulated. The purpose of the below figures is to display the proximity of the identified burst nodes generated by the BDLM to the pipe burst locations.

Figure 6.9 to Figure 6.11 show the BDLM can approximate the burst areas. The actual burst nodes are among the top 10 ranked (or recurring) nodes displayed. Note that these can be used only to indicate the approximate area of a pipe burst as it is very difficult to pinpoint the actual burst location in near real time. The top 10 ranked (or recurring) candidate nodes determined 6 time steps (i.e. 90 minutes) after the burst has been detected.

It takes between 3 and 12 minutes to run the BDLM to determine burst flow and approximate burst area every time step. The time taken for the BDLM to output the suspected burst areas depends on 3 factors:

(1) the size and configuration of the DMA;
(2) Initial estimated burst flow.
(3) Burst flow increment.

The above simulations were performed on a personal computer with Intel core i5 processor @ 2.30GHz and 6.0 Gb RAM memory.

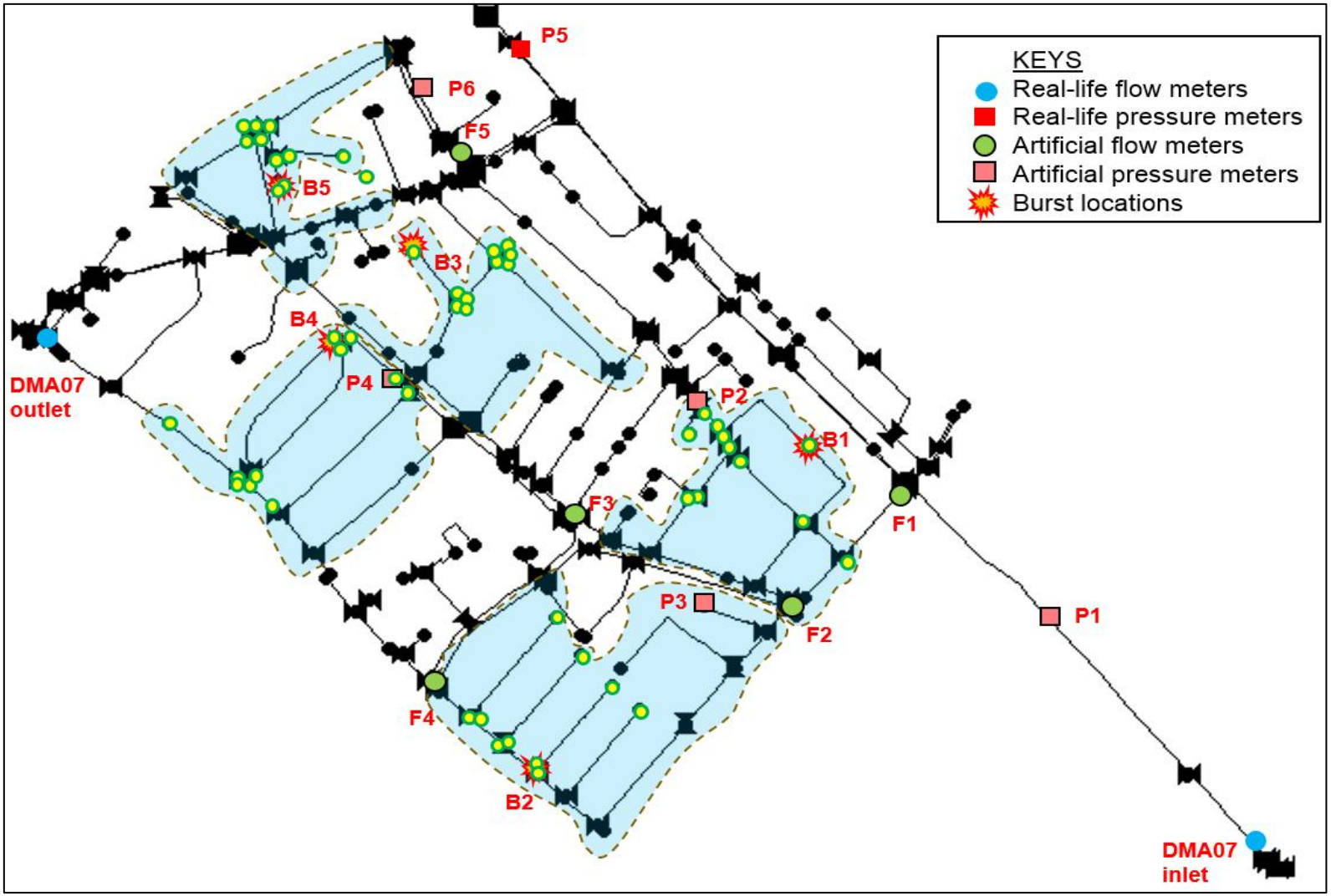


Figure 6.9: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for each burst location during the night period (burst flow - 2.7 l/s)

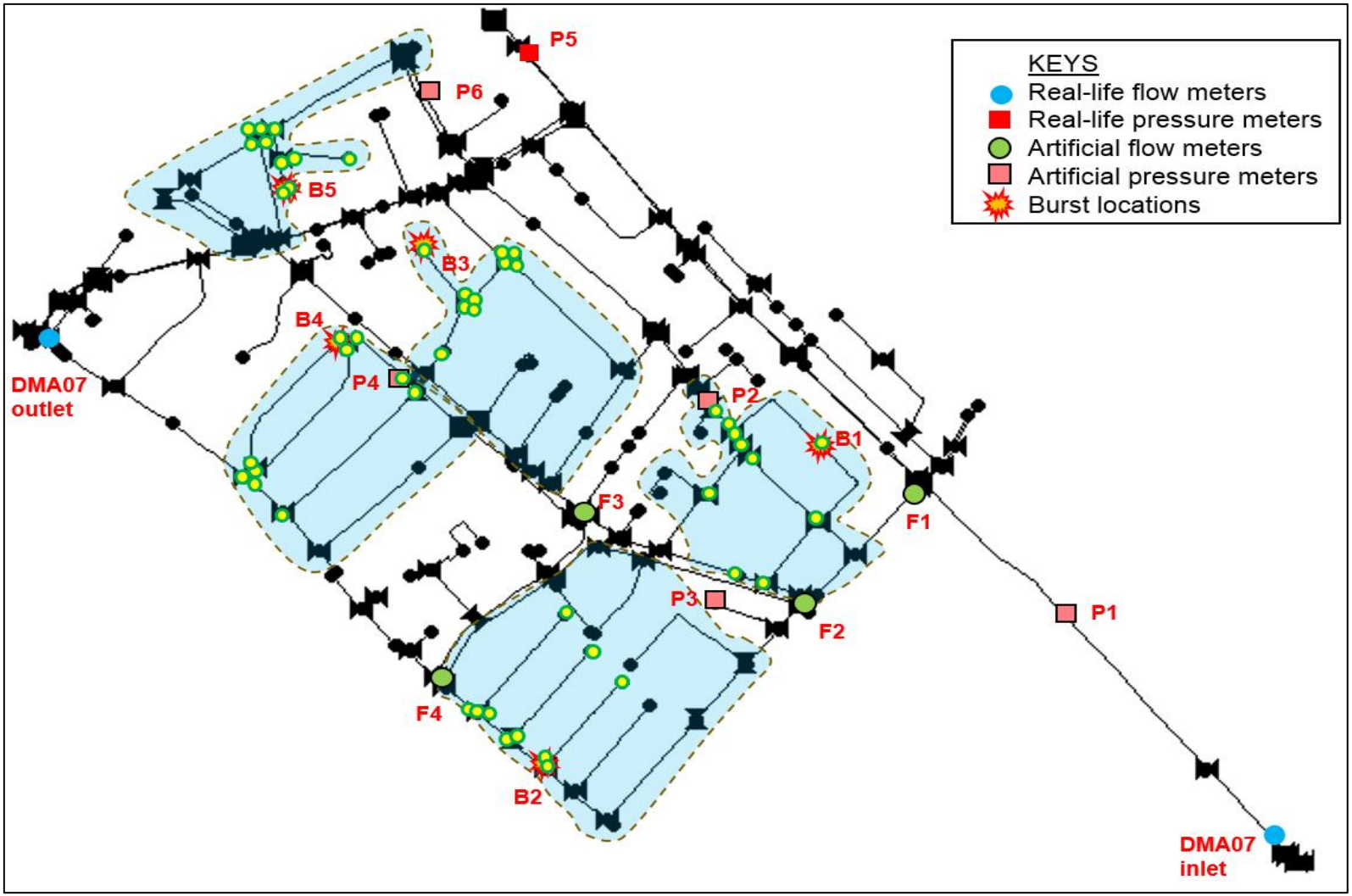


Figure 6.10: Illustration of the burst areas and 10 most sensitive nodes (red dots with yellow outline) for each burst location during the evening period (burst flow - 2.7 l/s)

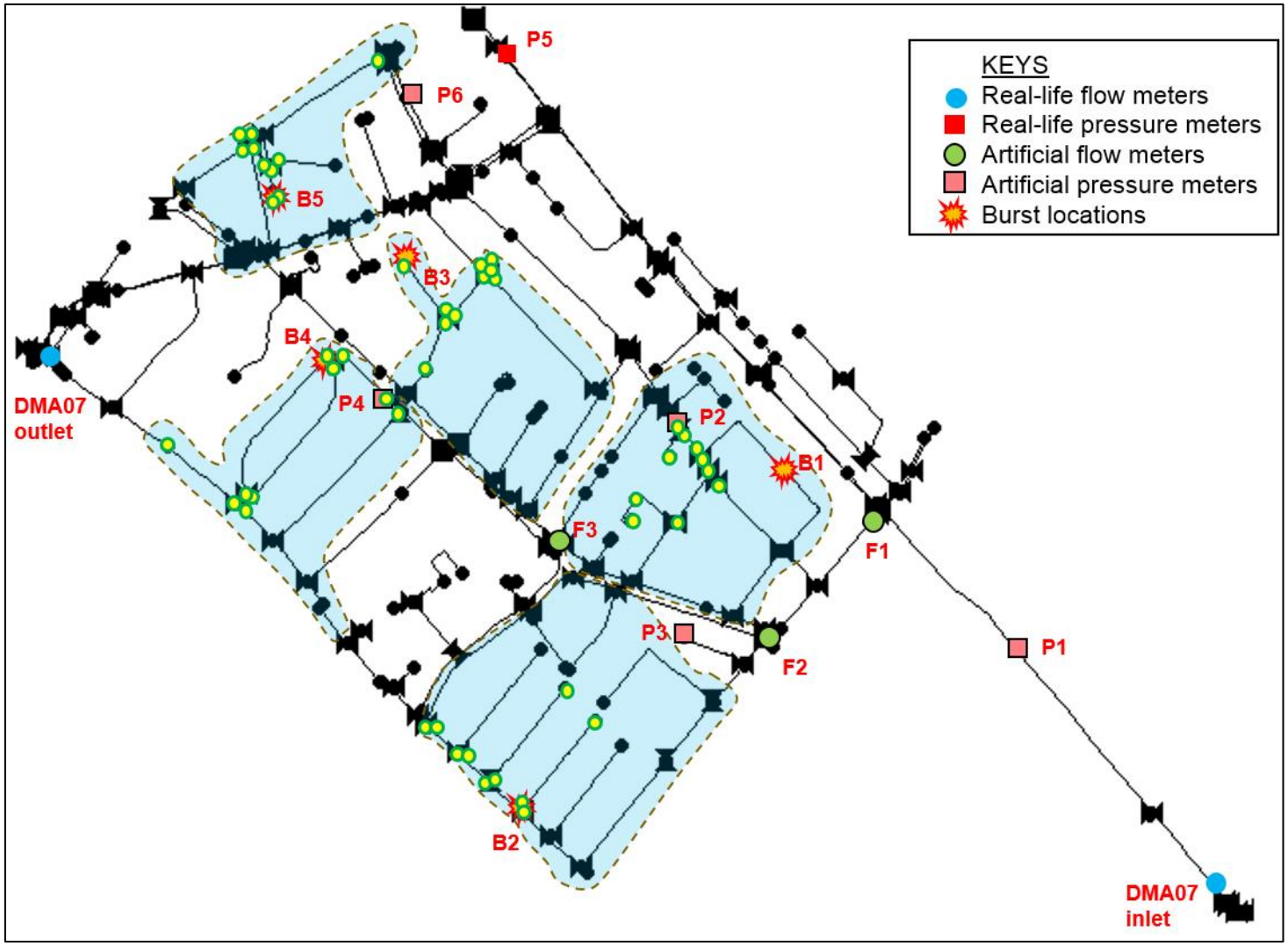


Figure 6.11: Illustration of the burst areas and 10 most sensitive nodes (red dots with yellow outline) for each burst location during the morning period (burst flow - 2.7 l/s)

6.2.8 Summary

This section presented BDLM that enables detection and localisation of pipe burst events at the DMA level. This is achieved by a combination of data algorithms that make use of flow and pressure residuals between the online hydraulic model predictions and corresponding WDS observations. The methodology comprises BDM and BLM which are used for pipe burst detection and location analysis respectively.

The BDLM includes WDFM which forecast DMA demand every 15mins for 15mins lead time. The predicted DMA demand is inserted into hydraulic model to obtain the predicted hydraulic state of the DMA. The i-KF method corrects the predicted demand of the DMA from the WDFM using the flow residual between the artificial flow observations and flow predictions from the hydraulic model. The BDLM then makes of use ISDA/PSDA SPC–based Control Charts and Rules to detect artificial pipe bursts in the near real-time.

The BDLM make use of an online hydraulic model, sensitivity matrix and various empirical statistical analyses to determine the approximate burst areas.

The BDLM was tested on a series of simulated pipe burst events in a real-life UK DMA (DMA07). The methodology is developed to make use of multiple hydraulic meters (flow and pressure) to detect and locate artificial pipe bursts online in a real-life DMA.

The case study used in this section have a limited number of pressure and flow meters in the DMA, hence all the flow and pressure observations are artificially generated from the hydraulic model. A total of 300 burst events are also artificially generated using the hydraulic model.

The results obtained show that the proposed BDLM is effective in detecting and locating burst locations in near real-time and satisfactorily estimating the burst flows. The results have also shown that the BDLM can detect pipe burst events in a timely fashion. In addition, the results also show the BDLM with a dynamic hydraulic model and reveals:

- Using a HSC with smaller number of hydraulic meters generally have lower burst detection rate compared to HSCs with larger number of hydraulic meters.
- More artificial pipe burst events are detected in the night compared to other period (morning and evening)
- There is no strong relationship between the closest and most hydraulic meters relating to the burst location.
- Low burst magnitudes (i.e. below 20% average of daily DMA demand) take longer time to detect (between 1 hour 45 minutes and 2 hours 15 minutes) while any pipe burst events above 20% average of daily DMA demand are detected more quickly (within 15 minutes).
- HSC have a good coverage of the DMA produced a favourable list of top 10 recurring candidate burst nodes
- HSC with 4 or more flow meters detect more burst events compared to HSC with fewer flow meters if observational noise level increased.
- More pipe burst events are located by the BDLM in the night compared to other periods (i.e., morning and evening).
- The location success rate of pipe burst events depend on the HSC used.

The next section shows the performance of the developed BDLM in a real-life DMA.

6.3 CASE STUDY #2

6.3.1 Section Overview

This section performs further testing and evaluation of BDLM detection and localisation capabilities on a real DMA with engineered burst events. After a general description of the case study area (Section 6.3.2), hydraulic meters data used are outlined (Section 6.3.3) and chosen parameters used for the BDLM (Section 6.3.5). Section 6.3.4 explains how the engineering events were simulated.

The specific objectives of the analyses performed and presented here were as follows:

- The first analysis aims at evaluating the detection performance of the BDLM at 5 different engineering events.
- The second analysis aims at evaluating the detection performance of the ISDA and PSDA SPC based charts during the engineered events.
- The third analysis aims at testing and evaluating the methodology's location capabilities on 5 different engineering events.

The major objective of the analyses is to review the robustness of the proposed BDLM on engineered events with pressure data only.

Details of above analyses and the results obtained are reported in Sections 6.3.6 and 6.3.7. The analyses are performed on Microsoft Visual C++ along with the EPANET 2.0 on a HP laptop with Intel core i5 processor @ 2.30GHz and 6.0 Gb RAM memory.

6.3.2 Case Study 2 Area Description

The BDLM was tested on a real-life water network called E13. It is located in Yorkshire, England. The E13 supplies water to approximately 897 domestic properties and 28 commercial properties under gravity with an average daily demand of 11 l/s. The E13 model consists of 373 nodes and 390 pipes and is fed from a trunk main. The E13 network model is shown in Figure 6.12.

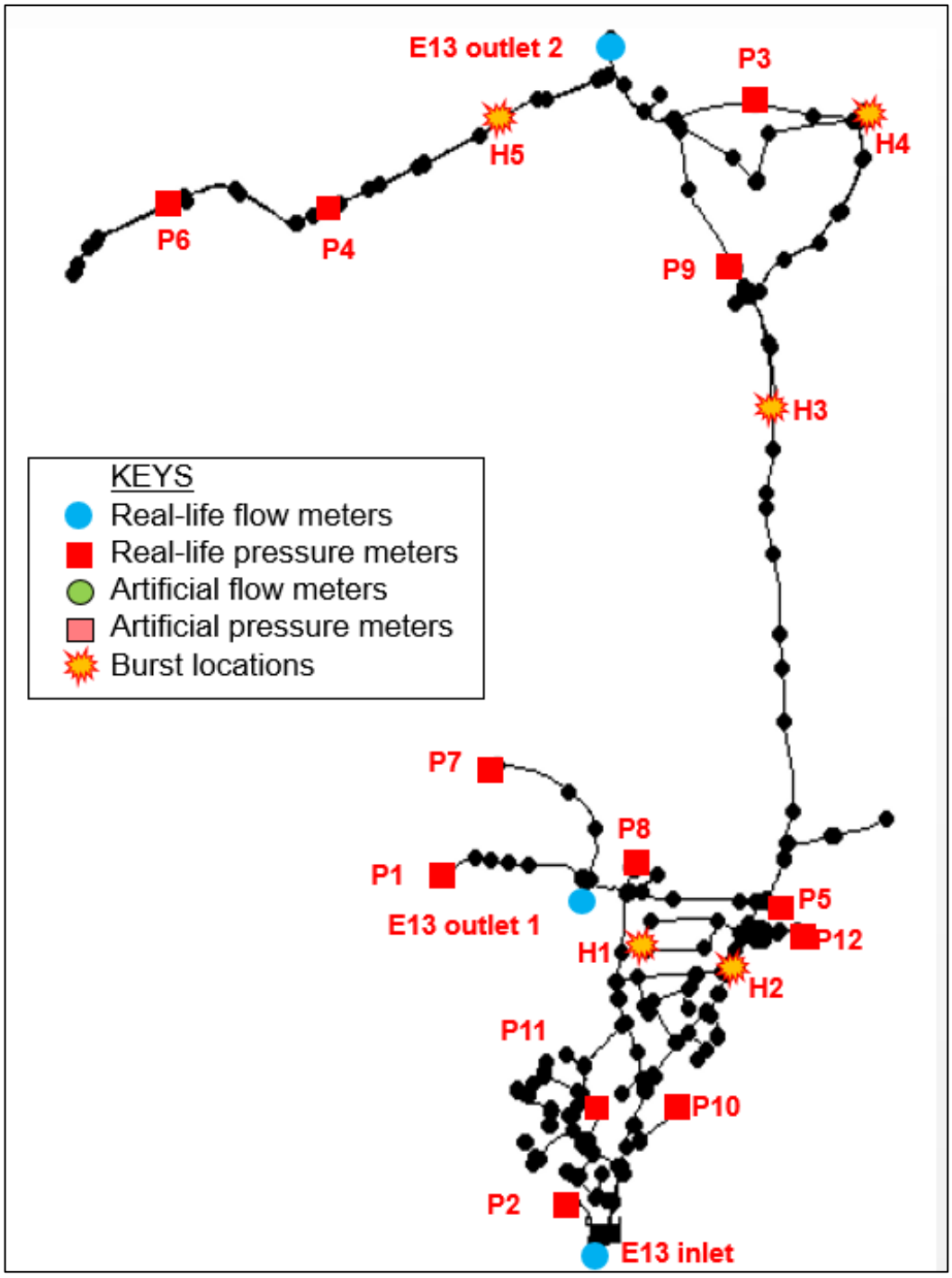


Figure 6.12: The overview of the studied water network – E13.

6.3.3 Sensor Data

The available real-life data are the E13's inflow and outflow data, and pressure data gathered at various point of the E13 (refer to Figure 6.12). The flow data is gathered between 1st April 2008 and 30st August 2008 (22 weeks). The pressure data is gathered

from 12 different pressure points (refer to Figure 6.12) within the DMA on the 6th, 7th and 8th of August 2008. For the purposes of the burst detection analyses presented in this section, only the 6 pressure meters data out of the 12 pressure meters were used. These six pressure meters are the most sensitive sensors among the available pressure meters (P1 – P6) determined using the DarwinSampler software (Bentley software) (Farley, et al., 2010). Six pressure meters were used in this case study to mimic the current situation in United Utilities where an average of 6 pressure meters are used to monitor a DMA.

However, there is no additional flow meter within the E13 water network. The spatial (i.e. nodal) demand allocation is done based on the fraction of E13 water demand allocated to each network node and the total DMA demand.

6.3.4 E13 Engineered Events

The engineered events in DMA E13 were carried out on the 7th and 8th of August 2008. The events involved opening several fire hydrants during the day to simulate pipe bursts fire hydrants at different times. Table 6.20 shows 5 engineering events that occurred on the 7th of August 2008. The fire hydrants were opened for approximately an hour.

Table 6.20: E13 Engineered Events time schedule on the 7th of August 2008.

Fire Hydrant location/index	**Time**	**Avg. Flow from fire hydrant (l/s)**	**Avg. Flow Magnitude (%)**
H1	08:30 - 09:15	5.833	54%
H2	09:30 - 10:30	6.911	64%
H3	10:45 - 11:45	6.615	64%
H4	12:00 - 13:00	6.611	64%
H5	13:15 - 14:15	5.191	49%

The average DMA demand which is 10.8l/s.

Engineered events carried out on the 8th of August 2008 are shown in Table 6.21. On the 8th of August 2008, each fire hydrant was open for approximately two hours. Note

that some fire hydrants (H4/H5 and H1/H2) ran simultaneously for approximately an hour.

Table 6.21: E13 Engineered Events time schedule on the 8th of August 2008.

Fire Hydrant location/index	Time	Avg. Flow from fire hydrant (l/s)	Avg. Flow Magnitude (%)
H5	08:45 - 10:45	2.00	19%
H4	09:45 - 11:45	2.00	19%
H1	12:00 - 14:00	2.00	19%
H2	13:00 - 15:00	2.00	19%

6.3.5 Burst Detection and Localisation Method Parameters

6.3.5.1 Offline Water Demand Forecasting Model Calibration

The difference between the inflows and outflows of the E13 water network at intervals of 15 minutes between May 2008 and June 2008 are assumed to be the total DMA water demand. The time series of water demand data are used to develop and calibrate the WDFM offline. Based on the cross-correlation results, the WDFM was developed using a combination of water demands from time step, *t*, t-670, *t-671, t-672* and *t - 1343*. The MLE is used to derive the weight coefficients. The WDFM uses previous demands at the DMA level as input to forecast demand and it can be written as:

$$d_{t+1} = 0.655d_t + 0.229d_{t-670} + 0.241d_{t-671} - 0.294d_{t-672} + 0.167d_{t-1343} \quad (42)$$

where d_{t+1} is the forecast demand; d_t is the current demand at current time step, t; d_{t-672} is the previous demand from a week + 15 minutes ago, t; d_{t-670} is the previous corrected demand from 167.75 hours ago, d_{t-671} is the previous corrected demand from a week ago and d_{t-1343} is the previous demand from 2 weeks ago.

6.3.5.2 Offline Hydraulic Model Calibration

The E13 hydraulic model is provided by Yorkshire Water via the University of Exeter. Due to lack of historic multiple pressure data before 5th August 2008 E13 hydraulic model is assumed to be calibrated.

6.3.5.3 BDLM Parameters

The pre-defined convergence/error conditions of i-KF method is the same as parameters used in the case study 1 (refer to Section 6.2.5.3).

The initial iterative step for each time step, $\boldsymbol{P}^{c}_{t,k-1}$ is a zero matrix while the diagonal element of $\boldsymbol{P}^{Q}_{t,\ k-1}$ is the product of average relative flow error and observed flow. The final demand estimates in equation (16) (refer to Section 3.4) are taken as the corrected water demands.

Prior to the engineered events detection analysis, the value of $\bar{p}$- and $\bar{r}$-*value* for inflow observation in equation (18) and (19) (refer to Section 3.4) respectively are determined by running the BDLM offline for a week (using July 2008 flow data) under normal conditions.Table 6.22 shows the initial $\bar{p}$-value and $\bar{r}$-*value* for E13 network model. The $\bar{p}$-*value is* the average weekly normalised error between the observed and predicted flow observations and while the $\bar{r}$-*value* is the average weekly normalised error between the observed and corrected flow observations.

Table 6.22: The value of $\bar{p}$-*value* and $\bar{r}$-*value* for each flow observations

Flow meter ID	**F1**	**F2**	**F3**
$\bar{p}$***-value***	0.055	0.065	0.057
$\bar{r}$***-value***	0.022	0.027	0.026

Since the BDLM makes use of the ISDA and PDSA SPC-based Control Charts/Rules to detect DMA burst pipes. The generic Control Rules for (ISDA and ASDA) SPC-based Control Charts are the same in Section 5.3. The Control Limits (mean and

standard deviation (sigma) value) are estimated using the previous one day of burst detection metric data at each time.

The flow observations from the DMA boundary (inflow/outflow) are used as the observed flow data to correct the forecast DMA demands. This is because there are no flow meters within the DMA.

The selected value of flow increment (*in step 4, Section 5.4.3*) is 0.5 l/s during the DMA demand and burst flow estimation process. This increment operator is chosen as small as possible to ensure the burst flow is accurately determined bearing in mind the computational constraints associated with the iterative procedure used. The lowest total absolute flow/pressure at initial iteration step is set to zero every time step once a pipe burst is detected. In the BDLM methodology step 4, 10% of the DMA nodes (i.e. a total of 33 nodes here) are used as the maximum number of nodes to identify the burst area. The top 10 recurring candidate burst nodes is obtained after 3-time steps (45mins) due to the limited period of the engineered events.

6.3.6 Testing of Burst Detection Capabilities in the E13 Water Network

The BDLM was run between 6th August 2008 and 8th August 2008, and tested on the engineered events. The performance of the BDLM is evaluated based on the number of detected bursts; false alarms raised and undetected bursts. In addition, the BDLM capabilities were also evaluated based on the time taken to detect pipe bursts.

Figure 6.13 shows the summary of the burst detection rate between 6th Aug 2008 and 8th August 2008 using six pressure meters (see P1 – P6 in Figure 6.12). Figure 6.14 and Figure 6.15 show the burst alarm status of each pressure meters (via ISDA and PSDA SPC-based charts) during the engineered events on the 7th and 8th of August 2008 respectively.

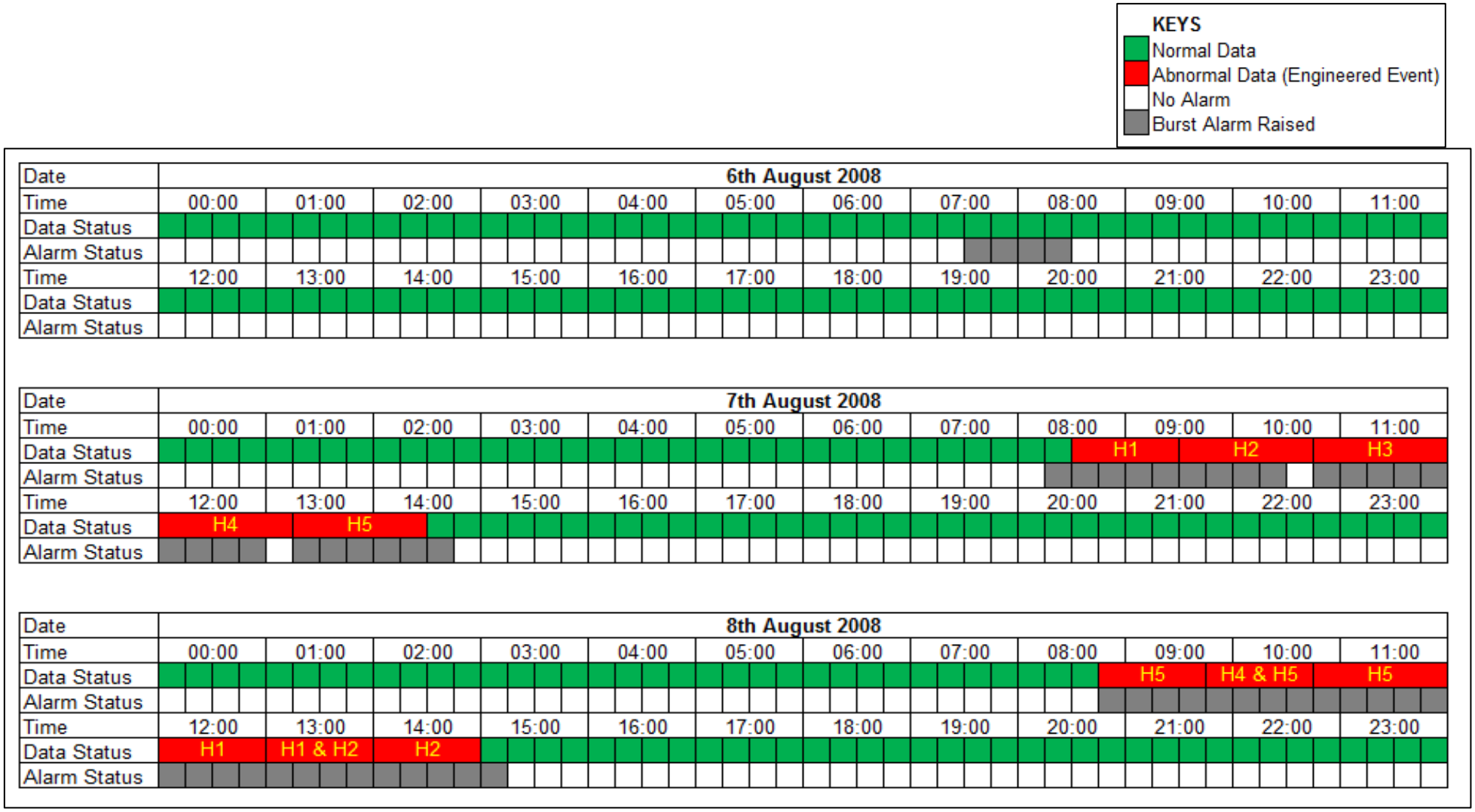


Figure 6.13: The dashboard of the detection results of E13 using the developed BDLM between 6th and 8th August 2008 (based on ISDA SPC-based charts only). Every square represents a 15 minutes flow/pressure data.

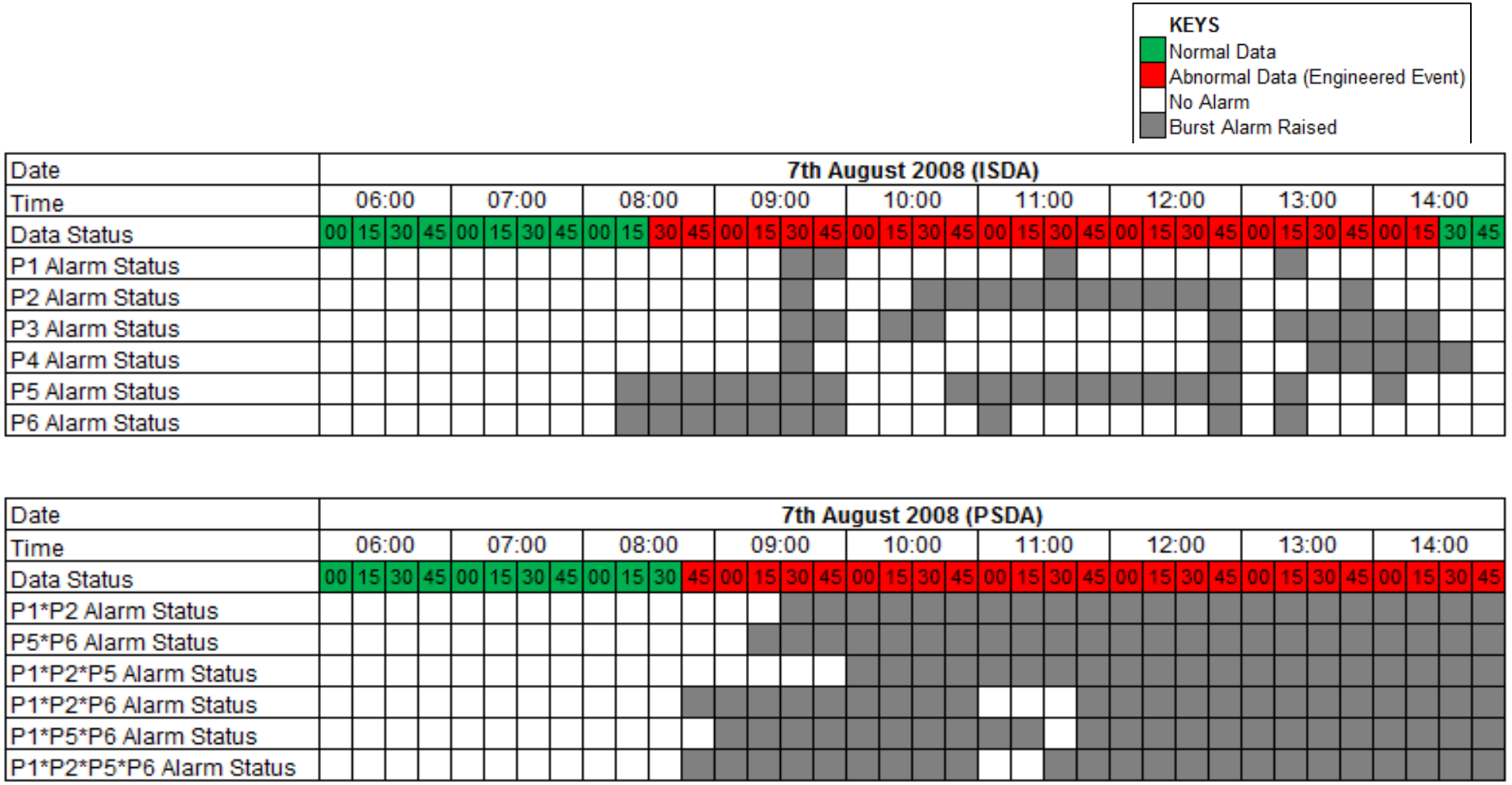
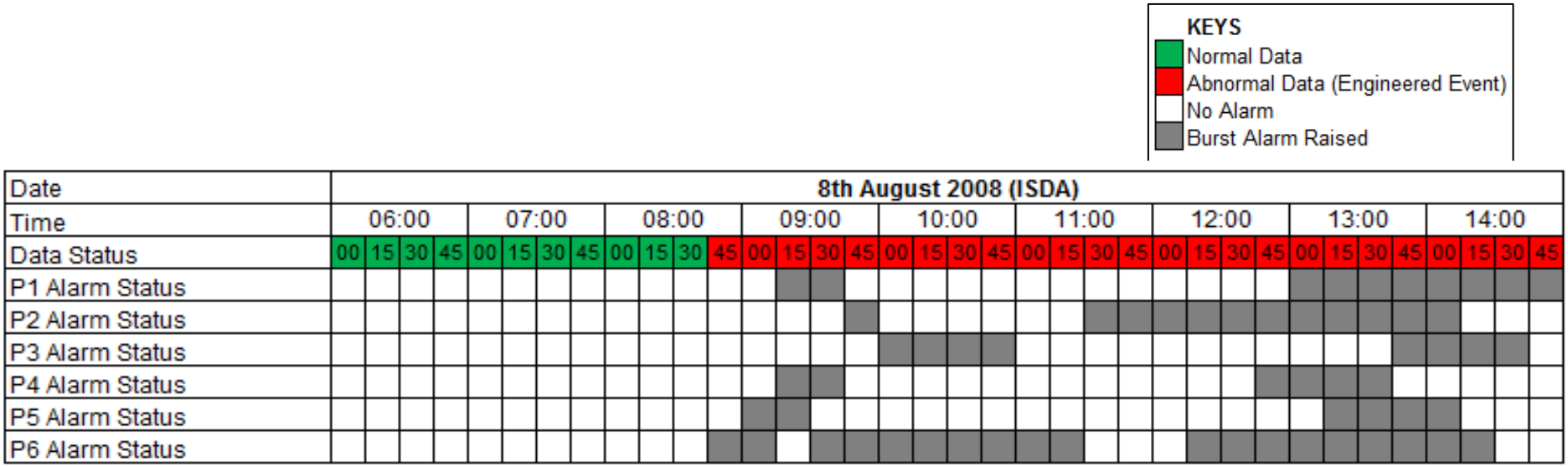


Figure 6.14: The alarm status of each pressure meters (via ISDA and PSDA SPC-based charts) during the engineered events on the 7th of August 2008. In the PSDA chart, the meters combinations are randomly selected.



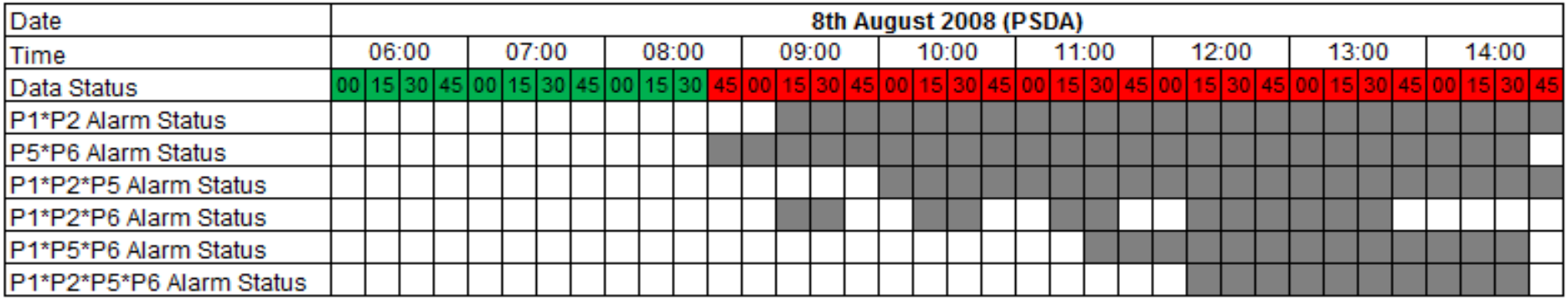


Figure 6.15: The alarm status of each pressure meters (via ISDA and PSDA SPC-based charts) during the engineered events on the 8th of August 2008. In the PSDA chart, the meter combinations are randomly selected

In Figure 6.13, each square box represents 15 minutes hydraulic data, red rectangular box represented engineered events and grey square box represents raised burst alarm at that time step. Figure 6.13 also shows all the engineered events were detected. A false burst alarm was raised on the 6th of August 2008 (between 07:30 and 08:30). This is because the pressure data is unusually lower than the historic pressure data.

Figure 6.13 and Figure 6.14 prove that the ISDA and PSDA complement each other. For example, on the 7th August 2008 at 10:00 am and 13:00, ISDA SPC-based charts could not detect engineered events. These missed engineered events were detected by PSDA SPC-based charts.

6.3.7 Testing of Localisation Capabilities in the E13 Water Network

The methodology behind the BDLM is tested on the E13 hydraulic model, a real-life DMA. The BDLM capabilities were also evaluated based on the approximated burst areas and top 10 recurring bust nodes.

Figure 6.16 to Figure 6.20 show the individual engineered event (7th Aug 2008) located within the approximated area and top 10 recurring candidate burst nodes. The results shown in these figures refer to the fourth-time step after the engineered events was detected (e.g., 09:30 hour. for the first engineered events, 10:30 hour. for the second engineered events and so on.). In each figure, the locations of the boundary flow meters and deployed pressure meters are indicated by using sky blue circles and red squares respectively. The real location of the engineered events is indicated by a yellow explosion symbol.

Figure 6.16 to Figure 6.20 show the approximated burst area for each engineered event. 33 nodes (or 10%) of the total DMA model nodes were used to form the approximate burst area. Whilst the top 10 recurring candidate burst nodes were determined after 4 time steps (45 minutes) since the initial burst alarm. As it can be seen from Figure 6.16 to Figure 6.20, the engineered events are located within their respective approximated burst area. The engineering event nodes (or hydrants) are among their respective top 10 recurring burst nodes except in Figure 6.20 (engineering

event, H5). However, engineering event node H5 is not far from the top 10 recurring candidate nodes. The results show BDLM works well in a real-life DMA and high engineered event flow certainly help in locating the burst areas.

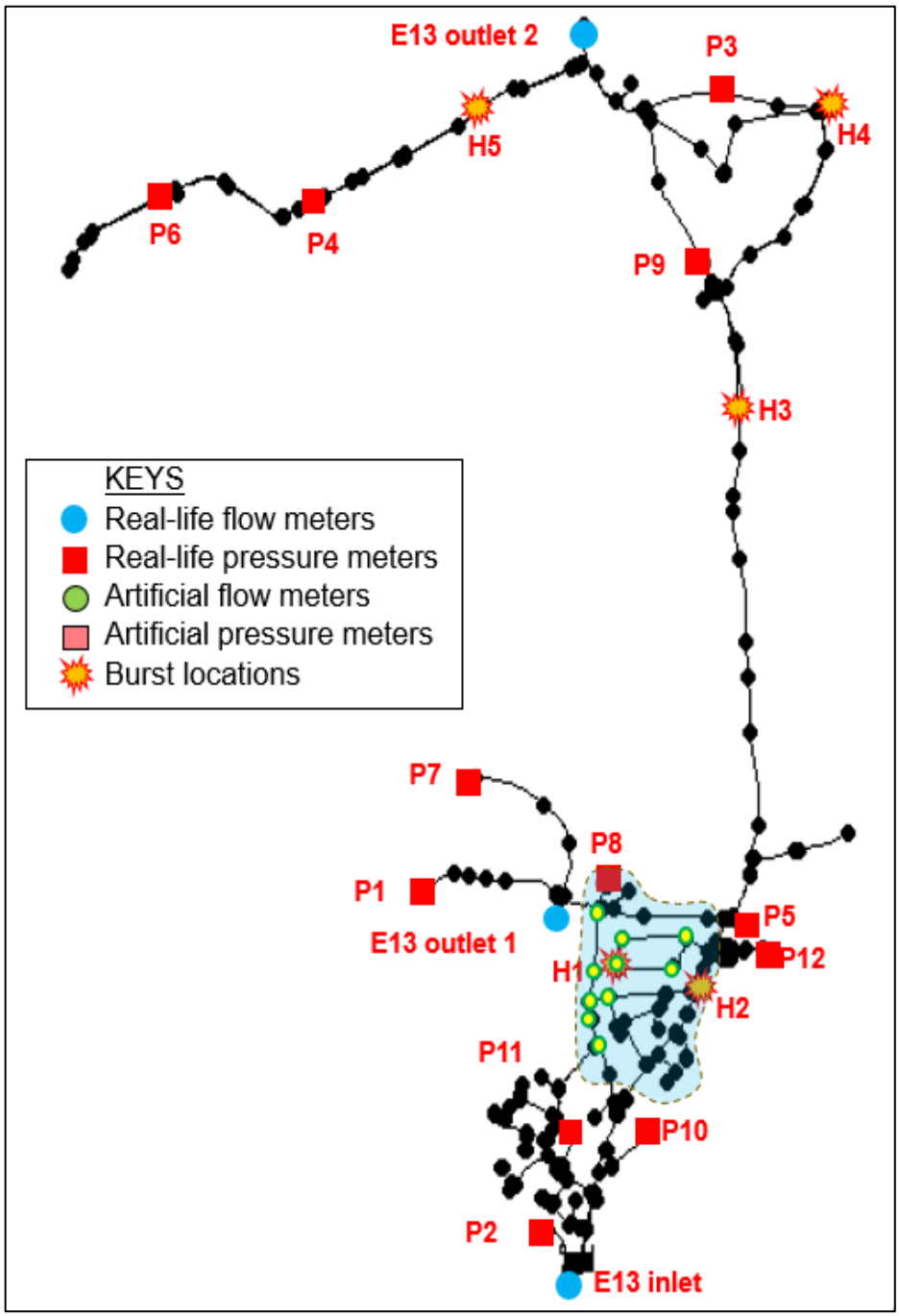


Figure 6.16: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for an engineering event at fire hydrant, H1 on the 7th Aug 2008 (burst flow ~ 5.833 l/s)

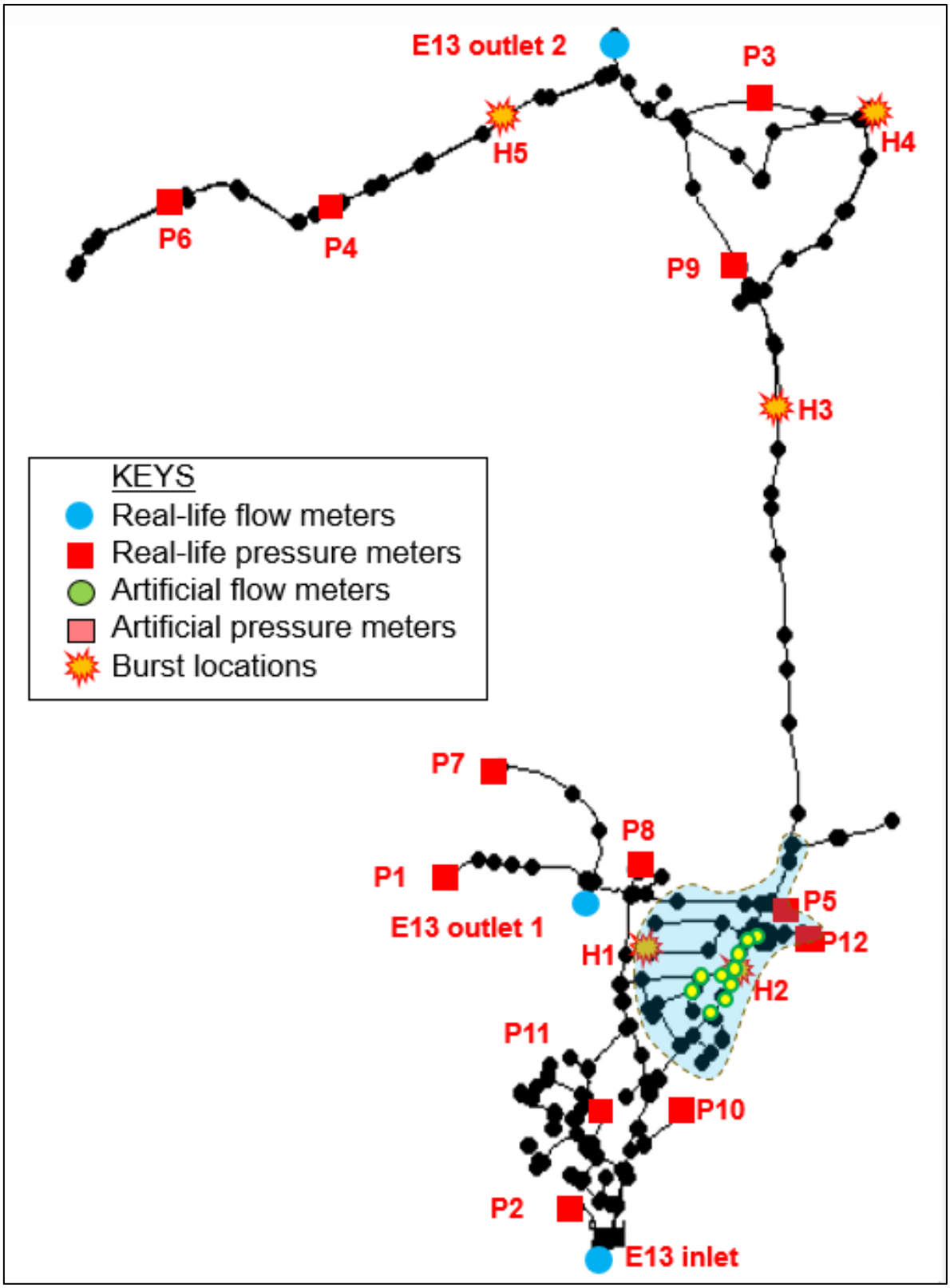


Figure 6.17: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for an engineering event at fire hydrant, H2 on the 7th Aug 2008 (burst flow ~ 5.911 l/s)

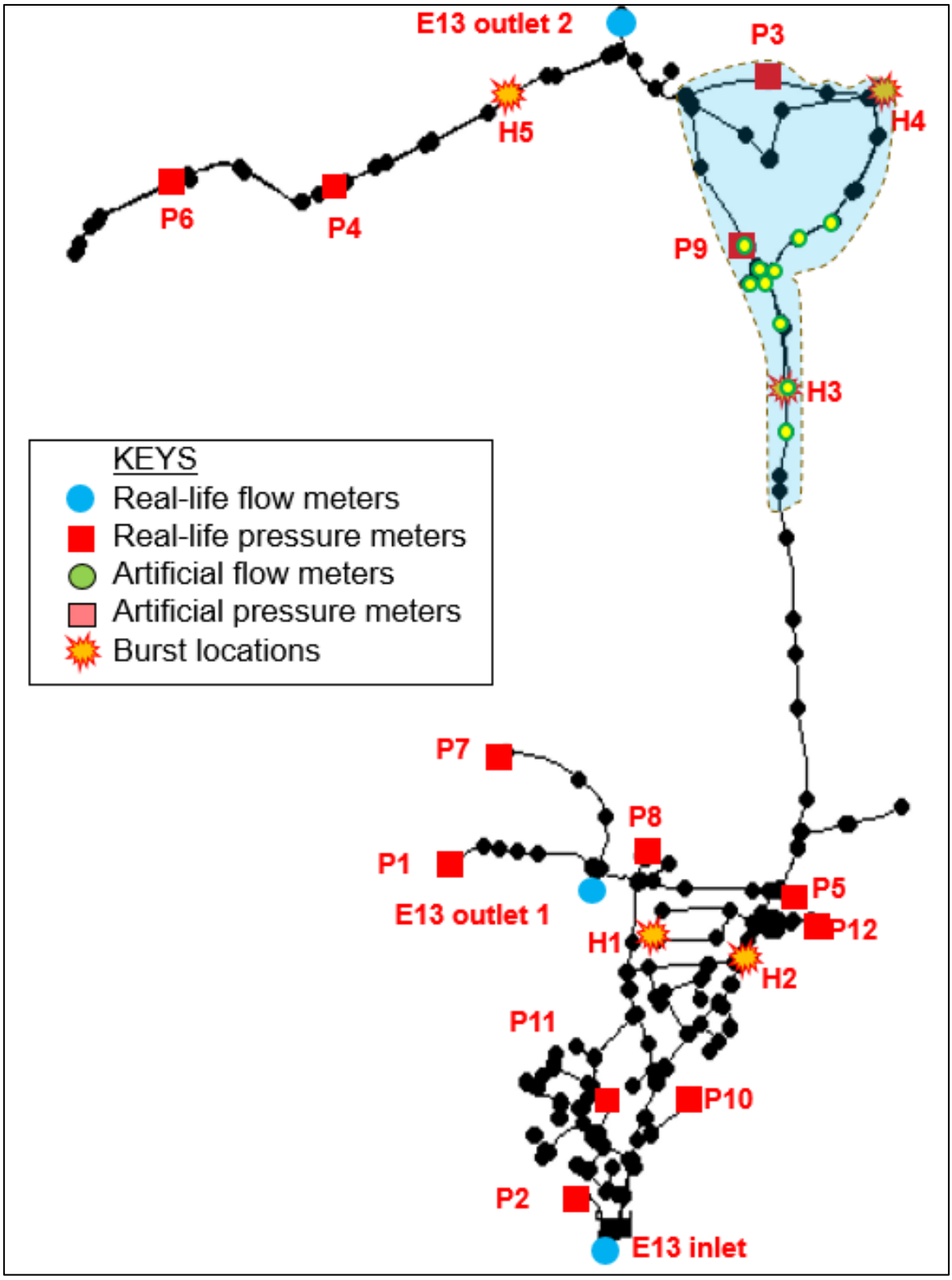


Figure 6.18: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for an engineering event at fire hydrant, H3 on the 7th Aug 2008 (burst flow ~ 6.615 l/s)

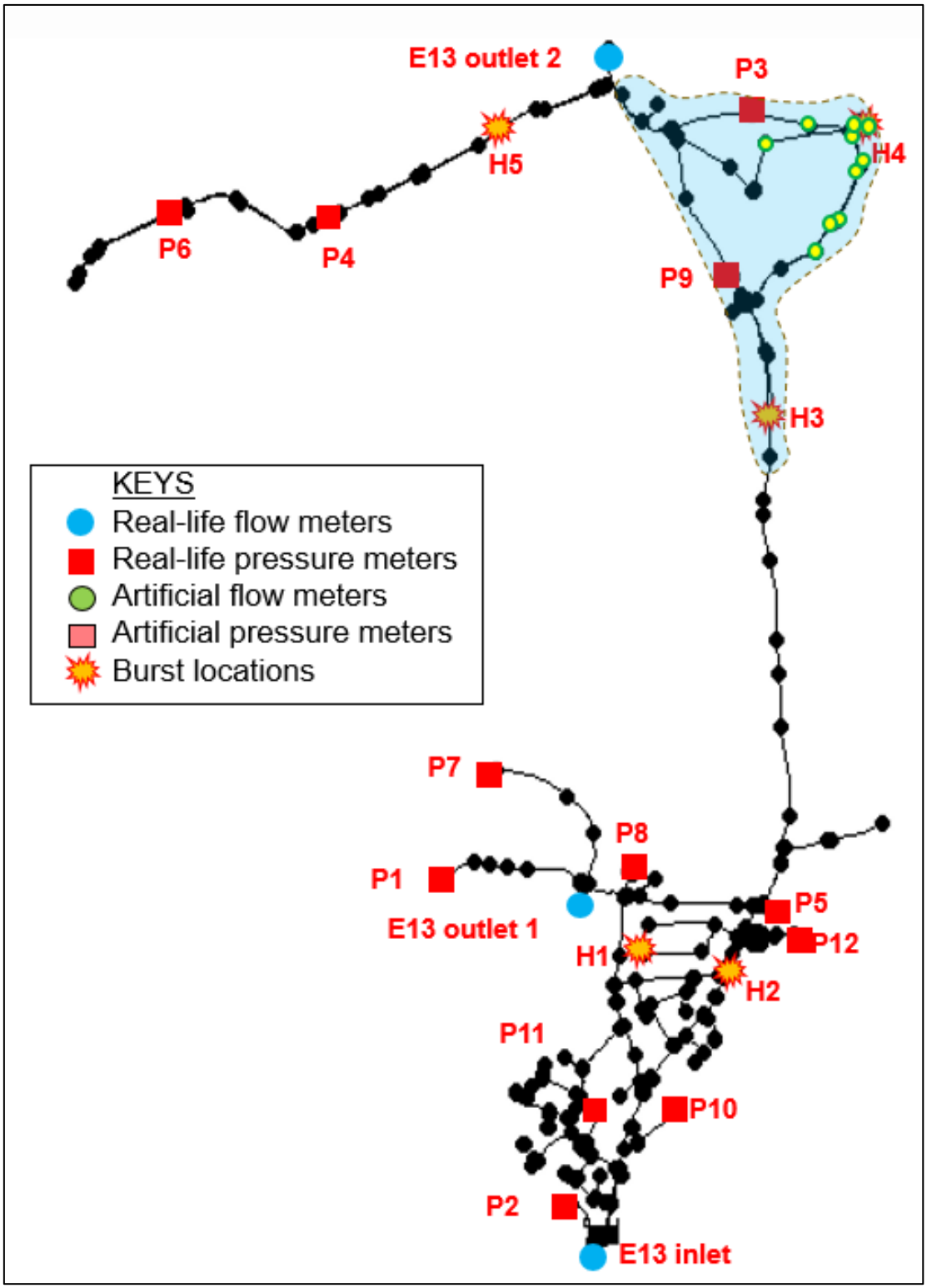


Figure 6.19: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for an engineering event at fire hydrant, H4 on the 7th Aug 2008 (burst flow ~ 6.611 l/s)

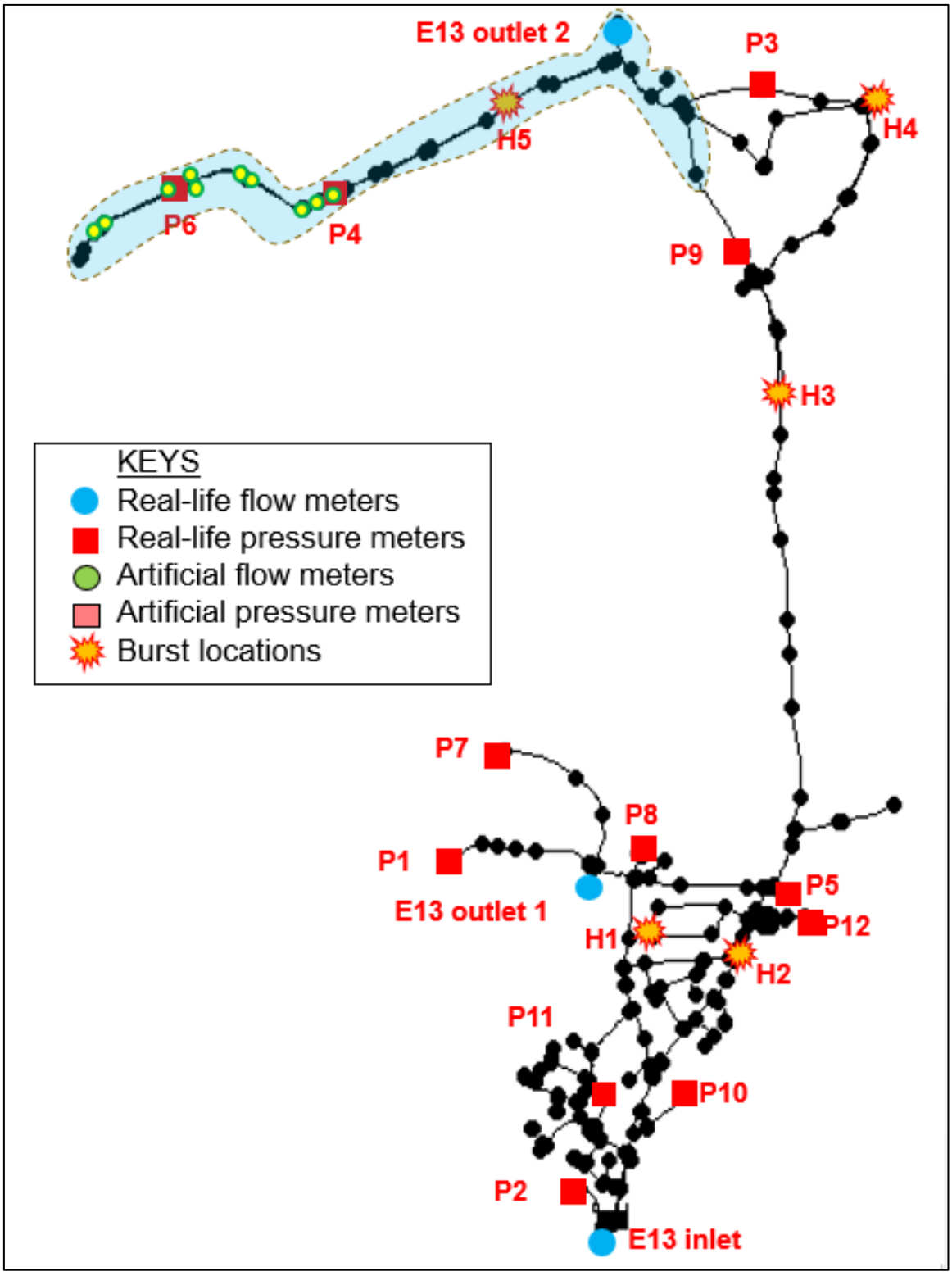


Figure 6.20: Illustration of the burst areas and 10 most sensitive nodes (yellow dots with green outline) for an engineering event at fire hydrant, H5 on the 7th Aug 2008 (burst flow ~ 5.191 l/s)

Since the engineering event node H5 (refer to Figure 6.20) is not far from the top 10 recurring candidate nodes. This could be because the 3 time steps were not sufficient to produce a favourable list of top 10 recurring candidate node for such engineered event, H5.

Table 6.23 shows the average simulated and estimated burst flows

Table 6.23: The average simulated and estimated engineering event flows

Location	Average simulated engineering event flow (l/s)	Average estimated engineering event flow (l/s)	Engineering event flow difference (l/s)
H1	5.883	6.731	0.848
H2	5.911	6.874	0.963
H3	6.615	8.379	1.764
H4	6.611	7.866	1.255
H5	5.191	6.433	1.242

As it can be seen from Table 6.23, the average estimated burst flows via the BDLM are larger than the average engineering event flows. This could be the hydraulic model may not be perfectly calibrated or the estimated engineering event flow could be underestimated. Nevertheless, the BDLM prove it is capable of detecting and locating engineering events with large flow.

6.3.8 Section Summary

The methodology's event detection and localisation capabilities have been tested and evaluated in the above section. The methodology was tested on a real-life DMA and real-life flow/pressure data. The main findings from the analyses performed in this section are briefly summarised below.

All the engineering events between 7th and 8th August 2008 were detected on time. The BDLM raised a false alarm because the pressure data on the 6th of August 2008 (between 07:30 and 08:30) were usually low. The detection results proved the BDLM

can detect abnormal events reliably and on time when real-life pressure data were at every time step (i.e., 15 minutes.).

The detection performance of ISDA and PSDA SPC-based Control Charts were also reviewed. It was found that two SPC-based Control Charts complement each other especially the PSDA SPC-based Control Charts.

The use of BDLM for locating engineered events has also been investigated. The BDLM was able determine the approximate location of all simulated pipe burst events at different location within the DMA being studied. The BDLM was able to produce a good list of top 10 recurring (or ranked) candidate burst node reliably. A total of 4 out of 5 bursts were micro located in a sense that respective network nodes where bursts where engineered were selected by the BDLM as one of the top ten recurring candidate nodes.

By using pressure meters only, the developed methodology successfully detected and determined the approximate location of all five different engineered events within the DMA being studied.

6.4 SUMMARY

There is an increasing number of accurate and cheap flow/pressure meters been installed in various DMAs by the UK WCs. The installation of multiple meters in a DMA would allow WC/water researchers to monitor their water network behaviour. This chapter aims at testing, validating and demonstrating the methodology presented in Chapter 5. The methodology aims to mimic the process of Network Modelling Engineer and Leakage Technician to detect pipe burst events.

The capabilities of the developed BDLM that enables performing event detection and location based on the analysis of observed flow/pressure data and online hydraulic model predictions were assed in this chapter.
Two case studies, both based on real pipe networks were used to test the methodology, one with numerically simulated bursts and the other with engineered

burst events. The numerically simulated bursts were generated via a real-life DMA hydraulic model (Section 6.2) while the engineered events were generated via opening fire hydrants (Section 6.3).

The BDLM methodology was tested and evaluated in the first case study by generating a number of artificial bursts as follows: -

- 5 different artificial burst locations
- 4 different burst magnitudes
- 3 different burst time periods (morning, evening and night)
- 5 observational noise levels.

Furthermore, issues such as the use of burst threshold (new Control Rules between 1 and 2 sigma) for pressure data to detect more bursts online have been investigated. The time taken for the BDLM to raise burst alarms when the burst event occurred was reviewed. Additional sensitivity type data analyses that evaluated the methodology' performances using a real-life DMA and pressure data (engineered events) have been investigated.

After the introduction, Section 6.1 discussed the objectives behind the detection and localisation analyses performed and presented in this chapter. Section 6.2 and 6.3 provided a description of the case study area and hydraulic meters data used. The details of the methodology's implementation or user defined parameters are provided in Section 6.2.5 and 6.3.5. Section 6.2.6 and 6.2.7 showed the results of the BDLM investigations and sensitivity type data analyses on artificial pipe burst events. Section 6.3.6 and 6.3.7 demonstrated the performance of the BDLM investigations and sensitivity type data analyses on engineered events (simulated burst by opening fire hydrants). Section 6.2.8 and 6.3.8 provided the research work summary and key conclusions.

This chapter has demonstrated the capabilities of the developed BDLM described in Chapter 5. It showed how the BDLM can detect in a reliable and timely fashion and then successfully approximate the burst areas few time steps later within the DMA. This has the potential to enable the WCs' personnel to locate bursts without resorting

to their expensive and laborious hardware tools in a DMA. Early detection of pipe burst events via the BDLM would also improve WCs' SIM score, reduce potential damages to third parties (i.e., residential properties or highway structures) and also improve their customer service.

The next chapter provides a summary of the research work carried out and presented in this thesis. The thesis main conclusions and a discussion about the directions of the future research work were outlined.

7 SUMMARY, CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS

7.1 THESIS SUMMARY

The world population growth means the demand for potable water use will continue to rise in the future. This puts immense pressure on the limited water resources available and the existing supply and distribution infrastructure. Most of these water supply infrastructures are ageing, especially in the developed countries (i.e., some pipeline in the UK are more than 150 years old). As a consequence, the existing water distribution systems are leaking substantial amounts of water, especially following pipe bursts. The issue is also present in the developing countries. The work done in this thesis aims to help reduce water loss (and overall demand for water) via an effective detection and localisation of pipe bursts/leaks in real time.

The proposed online detection and localisation of WDS pipe bursts/leaks is possible due to the advancement of hydraulic meters technologies. The hydraulic meters technologies are capable of transferring large quantity of data to the control room in the near real-time due to improved wireless technologies. Hence, an appropriate data assimilation scheme based on the up-to-date hydraulic model and data analysis methodologies provide the capability to detect and locate the pipe bursts online in timely ad reliable fashion.

After the introduction (Chapter 1), the thesis first reviewed the existing burst and localisation techniques (Chapter 2.2) and then the existing data assimilation schemes for online hydraulic modelling (Chapter 2.3). The concept of an offline and online hydraulic models in this thesis were developed and explained (Chapter 3). The developed offline and online hydraulic models were tested and evaluated on a real-life WDS and its hydraulic meters data (Chapter 4). The results showed online hydraulic models outperformed the offline model. Among the online hydraulic models analysed, the EnKF method was identified as the best performing online hydraulic model in terms of prediction accuracy. However, given relatively good prediction accuracy and much more computationally faster performance than the EnKF method, the i-KF method and was chosen as the online hydraulic model in this thesis.

The research gaps presented in this thesis (Chapter 2.2) resulted in the development of the proposed model-based detection and localisation methodology (Chapter 5). The proposed model-based methodology can make use of a large quantity of sensor observations and model detecting the pipe bursts and identifying their approximate location online. The developed methodology work in the following steps: Firstly, it instantaneously forecasts DMA water demand and then predicts the corresponding WDS states (i.e. system pressures and flows). Secondly, it assimilates the on-line WDS observations (flow/pressure signals) from the hydraulic meters installed in a DMA thus correcting the forecasted demands and predicted WDS states. Thirdly, it compares the online corrected WDS state with the incoming pressure and flow observations and uses this information to raise an alarm if a pipe burst is detected. Finally, the proposed localisation methodology based on statistical analysis is used to determine the approximate location of the pipe bursts. The developed online hydraulic model-based methodology for burst/leak detection and localisation is tested and demonstrated on two case study studies (Chapter 6). The first case study evaluates the performance of the novel model-based methodology on a real-life network with artificial hydraulic meters (flow and pressure) data and burst events (Chapter 6.2). The second case study further validated and demonstrated the novel methodology on a real-life DMA model with real-life pressure meters data (Chapter 6.3).

The results obtained have shown that the proposed BDLM can detect and locate pipe bursts events in a timely and reliably manner. The results obtained have also shown that developed methodology can be a useful tool to assist the leakage technicians as early identification of pipe bursts would help them make a prompt and reliable decision for an appropriate interventions and repairs.

7.2 SUMMARY OF THESIS CONTRIBUTIONS

The main contributions of this thesis to the research field of water distribution systems are as follows:

1. **A new methodology for an online burst detection in the WDS** is developed and presented. The proposed methodology makes use of online flow and pressure meters data, the online hydraulic model, newly developed burst detection metric and the Statistical Process Control Type Charts combined with new control rules for raising alarms. The new burst detection metric, based on the moving average residuals between the predicted and observed hydraulic states proved to be effective for detecting burst events. Two new SPC-based Charts with associated generic set of control rules for analysing burst detection metric values over consecutive time steps were introduced to raise burst alarms in a reliable and timely fashion.
2. **A new methodology for an online burst localisation in the WDS** is developed and presented. The new methodology integrates a new sensitivity matrix developed offline, information on burst detection metric values obtained during the detection stage and hydraulic model runs with simulated potential bursts to identify the most likely burst location in the pipe network. A new data algorithm for estimating the 'normal' DMA demand and burst flow during the burst period is developed and used for localisation. A new data algorithm for statistical analysis of flow and pressure data is developed and used to determine the approximate burst area and produce top 10 suspected burst location nodes.
3. **A novel comparative analysis of three online (EnKF, i-KF and PF based methods) and an offline hydraulic modelling technique is conducted**. The results show that all three online hydraulic models have outperformed the offline hydraulic model in terms of prediction accuracy. The main reason for this is the continuous updating of online hydraulic model parameters by using hydraulic meters data and assimilation techniques which, in turn, enables more accurate system state predictions to be made under continually changing demand and other conditions in the WDS. A comparison of three online hydraulic modelling techniques has also identified the EnKF method as best performing online hydraulic model in terms of prediction accuracy. However, online hydraulic model based on i-KF method was ultimately selected for burst detection and localisation due

to its relatively good prediction performance combined with substantially lower computational cost when compared to the EnKF method.

As it can be seen from the above, the methodologies presented in this thesis are based largely on the application of existing methods and concepts such as data assimilation, online hydraulic WDS models with demand forecasting, sensitivity analysis and statistical process control charts and associated detection rules. The key element of novelty comes from the innovative **integration of these techniques into a unique novel methodology** for burst/leak detection and localisation in WDS.

7.3 CONCLUSIONS

The main conclusions that summaries the research of the model-based burst detection and localisation methodology presented in this thesis.

The case study results obtained demonstrate that the proposed BDLM developed and presented in this thesis can be used in near real-time to reliably detect and locate pipe bursts and in timely manner. Early burst detection allows water company response and send out their leakage technicians to the approximate burst location suggested by the proposed BDLM. If the water from the pipe burst is not visible on the surface, the leakage technicians can use the hardware-based techniques to locate the burst (e.g. sounding equipment). Once the burst location is found, the repair team can repair the pipeline with minimum water supply interruption to water customers.

It is pertinent to mention that some of the UK water companies are still using the current best practice approach that is based on the combined application of water audits/MNF monitoring and hardware-based techniques. One of the UK WCs has trailed and adopted an advanced detection and localisation technique called Event Recognition System (ERS), developed by the University of Exeter. Even though ERS has proven its ability to effectively detect pipe bursts/leaks in ear-real time and by processing hydraulic meters data only, its main criticism is that, under the current low spatial density of flow and pressure meters in a DMA, it has limited ability to locate pipe burst/leaks. This is a consequence of the fact that ERS does not make use of the

hydraulic model. Therefore, in addition to this, any operational change (i.e, DMA boundary change or valve closure) requires periodic retraining of data driven models in the ERS.

Many UK WCs are beginning to install multiple hydraulic meters in a DMA which in turn collate hydraulic data such flows, pressures and tank levels. These hydraulic data are stored in SCADA systems. However, the UK water companies are still not making the most of this data. The flow/pressure data in the SCADA systems are mostly used to calibrate (offline) hydraulic models, for contingency planning and strategic planning. Majority of these hydraulic data are still collated to comply with OFWAT data collection regulations or regulatory reporting. The proposed BDLM aims to combine live hydraulic meters' data and online hydraulic model to detect and locate pipe burst in any WDS in timely and reliably fashion.

Furthermore, the proposed BDLM presented in this thesis is particularly suited to work in a live environment or near real-time. This is due to the following key benefits:

(1) The online hydraulic model makes use of a data assimilation method which correct DMA demand forecasts with WDS observations and model predictions along with hydraulic meters and model uncertainty. Therefore, the BDLM is already performing an automate task of demand analysis to estimate DMA demand every time interval. It is also extracting information from the imperfect hydraulic meters data to detect sudden increase of flow.
(2) The proposed BDLM is capable of re-calibrating WDFM online when its performance deteriorates over time due to change of DMA demand or behaviour.
(3) The proposed BDLM can be applied to any DMA networks without retorting to the DMA historic burst data.

The above-mentioned advantages allow the proposed BDLM to adapt to different the DMA operating conditions. The proposed BDLM can makes use of the existing available pressure and flow data along the hydraulic model. It can also process additional flow and pressure data if more meters are deployed in a DMA. Unlike most

model-based detection tools, the BDLM can assimilate a large quantity of flow data to estimate DMA demand despite noise being present in the data.

7.4 FUTURE WORK RECOMMENDATIONS

The future work should involve further testing and validation of the proposed BDLM in additional DMAs.

Some of the BDLM components presented in this thesis could also be improved upon. These include:

- Improvement of the generic Control Rules for pressure meters to detect small-sized pipe bursts. The BDLM has difficulty detecting low burst magnitudes with the current Control Rules in this thesis. The Control Rules could be improved by developing a method that output a localised SPC-based Control Rules for each DMA. Detecting low burst magnitudes with pressure meters only would help WCs who can't afford to install multiple flow meters in a DMA/WDS.
- Improvement of the BDLM detection component at the DMA level. The BDLM is currently not capable of differentiating real pipe burst events from other events that lead to increased local flow (e.g. hydrant flushing events, demand increases, equipment failures). This could be improved by studying the relationship between the abnormal events and flow/pressure data. Differentiating the abnormal events would help leakage technicians to make more informed intervention decisions.
- Improvement of the BDLM localisation component at the DMA level. The BDLM has been tested in a real-life case study to approximate the engineered event locations with a relative high flow in a DMA using pressure data only. The BDLM section could be improved if tested against real-life pipe bursts with smaller burst magnitudes (i.e., less than 20% of the average daily DMA demand) with a combination of flow and pressure.
- Development of a method that ranks the raised alarms from severe to low level. The BDLM does not currently provide the risk level of each pipe bursts detected

if it is used in multiple DMAs. This area of the BDLM could be improved to help WCs to prioritise the events to deal with and minimise structural damages in timely fashion. A combination of Eisenhower's Urgent/Important Principle (Covey (2004); McKay (2013)) and risk-based approach could be used to rank the burst alarms. The risk-based approach could be calculated as a function of these factors such as number and type of properties in a DMA, the burst alarm period, number of hydraulic meters raising the burst alarms and MAR values.

- Improvement of the selection of hydraulic meters' combinations for PSDA SPC-based Chart to detect pipe bursts online. For example, a DMA with 5 pressure meters results in 24 PSDA SPC- based charts which comprises ten 2–pressure meter PSDA charts, ten 3–pressure meter PSDA charts, three 4–pressure meter PSDA charts and one 5–pressure meter PSDA chart. This number of PSDA charts could be reduced by selecting fewer PSDA charts without losing the effectiveness of the BDLM to detect pipe bursts online. Reduction of PSDA SPC-based charts would also reduce the computational time to process the detection analyses.

Finally, the knowledge of the BDLM framework could be potentially applied to other water networks, wastewater networks and gas networks providing a number of flow meters and pressure meters are installed in the distribution system good number.

This page is intentionally left blank

APPENDIX A. HYDRAULIC MODEL OF WATER DISTRIBUTION SYSTEM

A1. HYDRAULIC NETWORK ANALYSIS (EPANET2.0)

A typical WDS model consists of a large set of pipes and nodes that have to be solved with a number of continuity and energy (head) equations simultaneously. Traditionally, Hardy Cross method (Cross, 1936) was initially used to solve the flow-head loss problems in WDS. The method assumes the internal flow and head of WDS are unknown whereas the inputs and outputs are known. This method is now redundant due to the rise of high-performance computer processor. Newton-Raphson method, a gradient method is mostly used to solve the WDS node-loop problems computationally (Boulos, et al., 2006). The Newton-Raphson method for the steady-state conditions are generalised by (Todini & Pilati, 1987). The nodal balance and head loss-flow relationship are expressed in the following matrix form:

$$\begin{bmatrix} \boldsymbol{A_{11}} & \boldsymbol{A_{12}} \\ \boldsymbol{A_{21}} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{H} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{A_{10}H_0} \\ \boldsymbol{q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F_p}(\boldsymbol{Q},\boldsymbol{H}) \\ \boldsymbol{F_Q}(\boldsymbol{Q},\boldsymbol{H}) \end{bmatrix} \quad (1)$$

where $\boldsymbol{A_{11}}$ matrix(n_{pu}, n_{pu}) that depends on the particular head loss-flow relationship, $\boldsymbol{A_{12}}$ is the topological matrix (n_{pu}, n_{nu}); $\boldsymbol{A_{21}}$ is the transposed of the $\boldsymbol{A_{12}}$ matrix (n_{nu}, n_{pu}); $\boldsymbol{A_{10}H_0}$ is the known vector (n_{pu}, 1); $\boldsymbol{q}$ is the nodal demand; $\boldsymbol{Q}$ is the flow rate; $\boldsymbol{H}$ is the nodal head; $\boldsymbol{F_p}(Q,H)$ is the system that represents the head loss-flow relationship; $\boldsymbol{F_Q}(Q,H)$ represents to the nodal flow (mass) balance.

n_{nu} is the number of nodes with unknown head; n_{nk} is the number of nodes with fixed head and n_{pu} is the number of pipes with unknown flow.

The head loss-flow relationship in WDS is non-linear and their generalised relationship is described the law of energy conservation for fluid passing through two nodes and a pipe is (Boulos, et al., 2006): The $\boldsymbol{F_p}(Q,H)$ matrix can also represent the head loss of the pipe between the nodes and $\boldsymbol{A_{11}}$ matrix can be written as:

$$\boldsymbol{A_{11}} = \left[\begin{pmatrix} \boldsymbol{K_1}|\boldsymbol{Q_1}|^{\boldsymbol{m}} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{K_{n_{pu}}}|\boldsymbol{Q_{n_{pu}}}|^{\boldsymbol{m}} \end{pmatrix}\right] \tag{2}$$

where $\boldsymbol{A_{11}}$ matrix that describe the head loss-flow relationship; K_n is the minor loss coefficient; Q_n is the flow rate at the node; n is the node index and m is the iterative index.

Obtaining solution from the system represented by equation (2) is almost impossible (Ahmed, 1997) because of its highly non-linear system. Therefore, an iterative gradient method is feasible to solve the non-linear system of WDS (Boulos, et al., 2006):

$$\begin{bmatrix} \boldsymbol{NA_{11}} & \boldsymbol{A_{12}} \\ \boldsymbol{A_{21}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\partial Q} \\ \boldsymbol{\partial H} \end{bmatrix}^{\boldsymbol{m}} = \begin{bmatrix} \boldsymbol{\partial \mathrm{E}} \\ \boldsymbol{\partial q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\partial F_p(Q,H)} \\ \boldsymbol{\partial F_Q(Q,H)} \end{bmatrix} \tag{3}$$

where $\boldsymbol{N}$ is the diagonal matrix (n_{pu}, n_{pu}) of the exponents *m* of the head loss-flow relationship; $\boldsymbol{A_{11}}$ matrix that describe the head loss-flow relationship; $\boldsymbol{A_{12}}$ is the topological matrix; $\boldsymbol{A_{21}}$ is the transposed of the $\boldsymbol{A_{12}}$ matrix; $\mathbf{0}$ is the matrix full of zero; $\partial\boldsymbol{Q}$ is the change in flow rate; $\partial\boldsymbol{H}$ is the change in nodal head; *m* is the iterative index; $\boldsymbol{F_p}(Q,H)$ is the system that represents the head loss-flow relationship; $\boldsymbol{F_Q}(Q,H)$ represents to the nodal flow (mass) balance.

Therefore, the solution to solve the system of imbalance equation is:

$$\begin{bmatrix} \boldsymbol{\partial Q} \\ \boldsymbol{\partial H} \end{bmatrix}^{\boldsymbol{m}} = \begin{bmatrix} \boldsymbol{NA_{11}} & \boldsymbol{A_{12}} \\ \boldsymbol{A_{21}} & \mathbf{0} \end{bmatrix}^{-\mathbf{1}} \begin{bmatrix} \boldsymbol{\partial \mathrm{E}} \\ \boldsymbol{\partial q} \end{bmatrix} \tag{4}$$

and the systems of equations are updated every iterative step and can be expressed as:

$$\begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{H} \end{bmatrix}^{\boldsymbol{m}} = \begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{H} \end{bmatrix}^{\boldsymbol{m-1}} + \begin{bmatrix} \boldsymbol{\partial Q} \\ \boldsymbol{\partial H} \end{bmatrix}^{\boldsymbol{m}} \tag{5}$$

When the convergence is not achieved at the first iteration, the energy and flow imbalance equation is then recalculated until the m^{th} iteration are equal or lower than the predefined convergence accuracy. The full details of the Newton-Raphson procedures for WDS are explained in (Boulos, et al., 2006).

A2. DATA COLLECTION

The information needed for hydraulic modelling of WDS section are hydraulic parameters and pipe physical properties. The main WDS components and their properties of the WDS are listed in Table A2.1.

Table A2.1: Water Distribution System Model Properties

Components	**Static Properties**	**Dynamics Properties**
Storages - (Tanks/Reservoirs)	Maximum capacity, Maximum height	Level (pressure head)
Pipes (links)	Length, diameter, roughness	Flow rate, Head loss
Nodes (junctions)		Pressure head, Nodal demand
Valves	Tau value	Flow rate, Head loss, Tau value
Pump		Pressure head, Pump speed

The components' dynamics properties in Table A2.1 can be measured either directly or indirectly providing telemetry devices read and record hydraulic properties. The pipe roughness is difficult to measure because it requires field test that involved expensive

technological equipment and human intervention. The number of telemetry devices within a WDS section is limited by budget constraint and restricted location. Some of the observation processes may be interfering with the normal condition of WDS and often corrupted due to either background noise or poor device calibration.

A3. HYDRAULIC MODEL CALIBRATION

In the water industry, hydraulic or operational engineer require up-to-date WDS HM to predict WDS behaviour under number of operational conditions (Walski, (1983); Ormsbee, (1989)). The concerned WDS HM has to undergo model calibration to be updated. The main goal of calibration is to change the WDS hydraulic parameters to reduce the error between the WDS HM state (outputs) and its corresponding system observations. The calibrated WDS hydraulic parameters (physical attributes) give information concerning the physical state of the WDS HM studied. The common WDS hydraulic parameters that are calibrated are:

- Unmetered demand coefficients or nodal base demand
- Pipe roughness
- Valve setting or status
- Pump status or flow/head graph (pump curve)

The acceptable difference between the WDS HM states and observed values (flow and pressure) must be within 5%-10% in the water industry (Ormsbee & Lingireddy, 1997).

APPENDIX B. DATA ASSIMILATION METHODS

B1. KALMAN FILTER METHOD

The Kalman filter (KF) technique (Kalman, 1960) is a recursive estimator that combines measurements and related errors to produce the best estimates of unknown hydraulic state estimates. KF method is commonly applied to navigation and control of projectiles such as missile weapons, aircraft and spacecraft. It is also used for time series system such as signal processing, automation control of large stochastic dynamical system and econometrics. KF methods aim to estimate the hydraulic state vector of WDS model via linear stochastic process (Greg & Gary, 2001) and the prediction steps of KF in WDS context are as follows:

$$\boldsymbol{x_t^f} = \boldsymbol{M}(\boldsymbol{x_{t-1}^a}) + \boldsymbol{\omega_t} \tag{6}$$

$$\boldsymbol{y_t^f} = \boldsymbol{h}\left(\boldsymbol{x_t^f}\right) \tag{7}$$

$$\boldsymbol{z_t} = \boldsymbol{H}\left(\boldsymbol{y_t^f}\right) + \boldsymbol{\varepsilon_t} \tag{8}$$

where $\boldsymbol{x_t^f}$ is the forecast hydraulic state vector; $\boldsymbol{M}$ is the hydraulic state model; $\boldsymbol{\omega}_t$ is the hydraulic state model error; $h(.)$ models the nonlinear network hydraulics of WDS; $\boldsymbol{y_t^f}$ is the total predicted hydraulic state vector of all flows and heads at step, t of WDS model based on $\boldsymbol{x_t^f}$; $\boldsymbol{H}$ is the observation model which map the prior vector of hydraulic state vector to observed states; $\boldsymbol{\varepsilon}_t$ is the observation error and $\boldsymbol{z}_t$ is the system observations (flow rates and pressure heads) of the WDS.

The errors associated with the hydraulic state vector and system observations are assumed to be independent and zero-mean white noise with process error covariance matrix, $\boldsymbol{Q}_t$ and observations covariance matrix, $\boldsymbol{R}_t$ respectively:

$$\boldsymbol{\omega_t} \sim \boldsymbol{N}(\boldsymbol{0},\ \boldsymbol{Q_t}) \tag{9}$$

$$\boldsymbol{\varepsilon_t} \sim \boldsymbol{N}(\mathbf{0},\ \boldsymbol{R_t}) \tag{10}$$

where $\boldsymbol{\omega}_t$ is the hydraulic state model error; $\boldsymbol{\varepsilon}_t$ is the observation error (due to the quality of calibration within the hydraulic meters); *N* is the normal distribution with mean and variance to generate random mean.

The forecast hydraulic state vector consists of WDS hydraulic model states and parameters:

$$\boldsymbol{x_t^f} = (\boldsymbol{s_{t-1}^{a,T}}, \boldsymbol{p_t^f}) \tag{11}$$

where $\boldsymbol{x}_t^f$ is the forecast hydraulic state vector; $\boldsymbol{s}_{t-1}^{a,T}$ is the vector of posterior essential hydraulic state at previous time step (which are used as initial boundary conditions of WDS model); $\boldsymbol{p}_t^f$ is the forecast WDS parameter vector (mostly demand coefficients); T represents the transpose.

The WDS hydraulic model parameters are unknown dynamic variables such as the valve tau values, status of the valves and pump and water demand coefficients (e.g. unmetered or metered domestic users, farm usage, metered 10 hours or 24 hours user and leakage). The posterior essential hydraulic states at previous time step are used as the initial boundary conditions of WDS. Example of the essential hydraulic states are flow rates (source) and pressure heads (including water level in tanks/reservoirs) which are used to drive the WDS hydraulic model forward in time.

For modelling the uncertainties of the WDS hydraulic model, the initial boundary conditions are corrected with high accuracy and relative errors are likely to be small. WDS hydraulic parameters are likely to have large errors due to the true WDS parameters are uncertain and unknown. The errors in hydraulic state model are also partly caused by imperfect and heuristic representation (or model) of hydraulic state estimates.

The initial prior error covariance matrix, $\boldsymbol{P}_0^f$ can either be guessed if there is no observation available or derived from past system observations. The forecast error covariance measures the uncertainty associated with the predicted WDS model state estimates before the system observations are available (Kahl & Ledolter, 1983). The conservative approach for selecting the values for $\boldsymbol{P}_0^f$ is to use a large value or use the variance between the predicted and observed hydraulic state. The forecast error covariance for time step greater zero is given by:

$$\boldsymbol{P_t^f} = \boldsymbol{M P_{t-1}^a M^T} + \boldsymbol{Q_t} \tag{12}$$

where $\boldsymbol{P}_t^f$ is the prior (or forecast) error covariance matrix; $\boldsymbol{P}_{t-1}^a$ is posterior error covariance matrix at previous time step; $\boldsymbol{M}$ is the hydraulic state model that forecast the hydraulic state estimates; $\boldsymbol{Q}_t$ is the hydraulic state model error covariance.

The correction step involved combination of system observations and hydraulic state vector to produce corrected vector of hydraulic state vector and posterior error covariance matrix:

$$\boldsymbol{x_t^a} = \boldsymbol{x_t^f} + \boldsymbol{K_t}(\boldsymbol{z_t} - \boldsymbol{H}(\boldsymbol{y_t^f})) \tag{13}$$

$$\boldsymbol{P_t^a} = (\boldsymbol{I} - \boldsymbol{K_t H})\boldsymbol{P_t^f} \tag{14}$$

where $\boldsymbol{x}_t^a$ and $\boldsymbol{x}_t^f$ are the corrected and forecast hydraulic state vector respectively; $\boldsymbol{K}_t$ is the Kalman gain; $\boldsymbol{H}$ is the observation model; $\boldsymbol{y}_t^f$ is the total predicted hydraulic state vector of all flows and heads at the time step, t ; $\boldsymbol{z}_t$ is the vector of system observations; $\boldsymbol{I}$ is the identity matrix; $\boldsymbol{P}_t^a$ and $\boldsymbol{P}_t^f$ are the posterior and prior error covariance matrix respectively.

The difference $(\boldsymbol{z}_t - \boldsymbol{H}(\boldsymbol{y}_t^f))$ in equation (13) is called the Kalman innovation which reflects the discrepancy between the predicted WDS hydraulic model state vector and

the system observations. When the Kalman innovation approaches zero, it means both predicted WDS hydraulic model state vector and the system observation vector are reaching complete agreement. The Kalman gain, $\boldsymbol{K}_t$ is viewed as the weight factor based on the prior and observation error covariance matrix:

$$\boldsymbol{K_t = P_t^f H^T \left(H P_t^f H^T + R_t\right)^{-1}} \quad (15)$$

where $\boldsymbol{K}_t$ is the Kalman gain; $\boldsymbol{P}_t^f$ is the prior error covariance matrix; T indicates the matrix is transposed; $\boldsymbol{H}$ is the observation model; $\boldsymbol{R}_t$ is the observation error covariance matrix.

B2. Extended Kalman Filter Method

Extended Kalman Filter (EKF) (Jazwinski, 1970) is the advancement of KF to accommodate the nonlinear system like WDS. EKF expand the nonlinear system of WDS via Taylor expansion about prior hydraulic state estimates to forecast hydraulic state estimates at the next time step. Thus, the prediction steps of EKF are:

$$\boldsymbol{x_t^f = f\left(x_{t-1}^a, \omega_t\right)} \quad (16)$$

$$\boldsymbol{y_t^f = h\left(x_t^f\right)} \quad (17)$$

$$\boldsymbol{z_t = H y_t^f + \varepsilon_t} \quad (18)$$

where $\boldsymbol{x}_t^f$ is the forecast hydraulic state vector at the time step, t which consists of WDS model states and parameters; $f(.)$ models nonlinear function of forecast hydraulic state vector; $\boldsymbol{y}_t^f$ is the total predicted hydraulic state vector of all flows and heads at step, t; $h(.)$ models nonlinear network hydraulics; $\boldsymbol{z}_t$ is the vector of system observations; $\boldsymbol{H}$ is observation model that relate the predicted WDS hydraulic model state vector to system observations; $\boldsymbol{\omega}_t$ and $\boldsymbol{\varepsilon}_t$ are the vector of hydraulic state model

and observation error (or noises) which are both assumed to be zero mean multivariate Gaussian noises with hydraulic state model error covariance matrix $\boldsymbol{Q}_t$ and observation error covariance matrix, $\boldsymbol{R}_t$ respectively.

The prior error covariance matrix becomes:

$$\boldsymbol{P}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}_{\boldsymbol{t-1}}^{\boldsymbol{a}})\, \boldsymbol{P}_{\boldsymbol{t-1}}^{\boldsymbol{a}}\, (\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}_{\boldsymbol{t-1}}^{\boldsymbol{a}}))^{\boldsymbol{T}} + \boldsymbol{Q}_{\mathrm{t}} \quad (19)$$

where $J_f(.)$ is the Jacobian of the hydraulic state model; $\boldsymbol{P}_t^f$ is the prior error covariance matrix; $\boldsymbol{P}_{t-1}^a$ is the posterior error covariance at previous time step; $\boldsymbol{Q}_{\mathrm{t}}$ is the hydraulic state model (or process) error covariance matrix at current time step and T indicates the matrix is transposed.

The correction step of EKF is:

$$\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{a}} \approx \boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{f}} + \boldsymbol{K}_{\boldsymbol{t}}(\boldsymbol{z}_{\boldsymbol{t}} - \boldsymbol{H}\left(\boldsymbol{y}_{\boldsymbol{t}}^{\boldsymbol{f}}\right)) \quad (20)$$

$$\boldsymbol{P}_{\boldsymbol{t}}^{\boldsymbol{a}} = (\boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{t}}\boldsymbol{J}_{\boldsymbol{h}}\left(\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{f}}\right))\boldsymbol{P}_{\boldsymbol{t}}^{\boldsymbol{f}} \quad (21)$$

where $J_h(.)$ is the Jacobian of the hydraulic network (or the sensitivity of the system observation to the changes in hydraulic state estimates); $\boldsymbol{K}_t$ is the Kalman is gain; $\boldsymbol{P}_t^f$ is the prior error covariance matrix; $\boldsymbol{P}_t^a$ is the posterior error covariance; $\boldsymbol{y}_t^f$ is the total predicted WDS hydraulic model state vector of all flows and heads at the time step, t ; $\boldsymbol{z}_t$ is the system observation vector; $\boldsymbol{I}$ is the identity matrix; $\boldsymbol{P}_t^a$ and $\boldsymbol{P}_t^f$ are the posterior and prior error covariance matrix respectively.

The Jacobian of the WDS hydraulic state model and WDS hydraulic model nonlinear function is defined as:

$$\boldsymbol{J}_{\boldsymbol{f}}(\boldsymbol{x}_{\boldsymbol{t-1}}^{\boldsymbol{a}}) = \frac{\boldsymbol{\partial f}}{\boldsymbol{\partial x}}|_{\boldsymbol{x}_{\boldsymbol{t-1}}^{\boldsymbol{a}}} \quad (22)$$

$$\boldsymbol{J_h\left(x_t^f\right)} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}|_{\boldsymbol{x_t^f}} \quad (23)$$

where $J_f(.)$ is the Jacobian of the hydraulic state model with respect to the corrected hydraulic parameter estimates at previous time step and $J_h(.)$ is the Jacobian of the hydraulic network with respect to the forecast hydraulic parameter estimates;

The higher-order terms of Taylor series are considered negligible (Terejanu, 2008) due to the large variant of observation error. Even the higher-order terms are included the obtained hydraulic state estimates are still biased. This is because the observation nonlinear functions are approximated without other higher-order terms. Hence, the EKF is limited the first-order of Taylor series.

The Kalman gain of EKF is:

$$\boldsymbol{K_t} = \boldsymbol{P_t^f J_h^T\left(x_t^f\right)}\left[\boldsymbol{J_h\left(x_t^f\right)P_t^f J_h^T\left(x_t^f\right)} + \boldsymbol{R_t}\right]^{-1} \quad (24)$$

The disadvantage of EKF is EKF neglects mapping of observations of the system that has higher-order derivatives (Pauwels & De Lannoy, 2009) in the analysis step. Therefore, the hydraulic state model does not take account of WDS state and parameter estimates with inherent uncertainty. This makes both the forecast hydraulic state and parameter estimates biased. It is possible to account hydraulic state and parameter uncertainty from higher-order terms of Taylor series if the observation error is very small. However, the higher-order term of EKF is computational expensive which make it unfeasible for practical usage in cases of both real time applications and high dimensional systems (Evensen, 2003; Terejanu, 2008). The Jacobian matrix of the hydraulic state model and observation nonlinear function is restricted to a nonlinear system that has small non-linearity (Julier & Uhlmann, 1996).

The other disadvantage of EKF is it is not regarded as optimal estimator of hydraulic state estimates because EKF linearises the system. Like KF, EKF tends to underestimate the corrected forecast error covariance which causes the filter divergence. Hence, EKF relies heavily on the addition of Gaussian random variable

(white noise) to prevent the underestimation of the true forecast error covariance matrix. Due to the firs order approximation of EKF, EKF introduce large errors in the corrected hydraulic parameter and state estimates which lead to the divergence of the filter. This flaw is addressed by Unscented Kalman Filter (UKF) in section 4.3. Nevertheless, EKF is still capable of providing a good performance of estimating hydraulic states and parameter estimates in a small WDS but implementing EKF in a large WDS would be challenging with high computational cost.

B3. UNSCENTED KALMAN FILTER METHOD

The Unscented Kalman Filter (UKF) (Uhlmann, 1995) was developed to address the limitation of the EKF. EKF loses information concerning the hydraulic state model when EKF tries to linearise the system. UKF uses unscented transformation, a deterministic sampling technique, to capture the minimal set of sample points around the hydraulic state vector. These minimal sets of sample points are called sigma points. UKF updates the hydraulic state vector and error covariance at previous time step by using (*2L + 1*) sigma points, a set of weights given by:

$$\boldsymbol{X_{t-1}^{0}} = \boldsymbol{x_{t-1}^{a}} \tag{25}$$

$$\boldsymbol{X_{t-1}^{i}} = \boldsymbol{x_{t-1}^{a}} + (\sqrt{(\boldsymbol{L} + \boldsymbol{\lambda})\boldsymbol{P_{t-1}^{a}}})_{\boldsymbol{i}} \tag{26}$$

$$\boldsymbol{X_{t-1}^{i+L}} = \boldsymbol{x_{t-1}^{a}} - (\sqrt{(\boldsymbol{L} + \boldsymbol{\lambda})\boldsymbol{P_{t-1}^{a}}})_{\boldsymbol{i+L}} \tag{27}$$

$$\mathbf{X_{t-1}^{a}} = \left[\boldsymbol{X_{t-1}^{0}} \;\; \boldsymbol{X_{t-1}^{i}} \;\; \boldsymbol{X_{t-1}^{i+L}}\right] \tag{28}$$

where $\boldsymbol{x_{t-1}^{a}}$ and $\boldsymbol{P_{t-1}^{a}}$ are vector of corrected hydraulic state estimates and forecast error covariance at previous time step respectively; $\lambda = \alpha^2(L + k)$ is the fitting parameter that is derived empirically from initial distribution; $(\sqrt{(L + \lambda)\boldsymbol{P_{t-1}^{a}}})_i$ is the *i*th column of the matrix square root; L is the scaling parameter; α is the spread of the

sigma points around hydraulic state estimation (usually set to a small positive value, e.g. 0.001) and k is a secondary scaling parameter which is usually set to 0.

The (*2L+ 1)* sigma points (or 3 sigma points when L is 1) is based on empirical analysis conducted by (Zhang, et al., 2009) which shown as for Gaussian distribution, minimal set points of *2L + 1* is a good rule of thumb. The forecast step of UKF is:

$$\mathbf{X}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{f}(\mathbf{X}_{\boldsymbol{t-1}}^{\boldsymbol{a}},\ \boldsymbol{w}_{\boldsymbol{t}}) \tag{29}$$

$$\boldsymbol{Y}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{h}\left(\mathbf{X}_{\boldsymbol{t}}^{\boldsymbol{f}}\right) \tag{30}$$

where $\mathbf{X}_t^f$ is the ensemble forecast hydraulic state vector; $\mathbf{X}_{t-1}^{a,}$ is the ensemble hydraulic state estimates approximated by sigma points at the previous time step; $\boldsymbol{w}_t$ is the hydraulic state model error matrix; $\boldsymbol{Y}_t^f$ is the total forecast hydraulic state vector of all flows and heads at the time step, t; and $f(.)$ models the nonlinear function of hydraulic state estimates and $h(.)$ models hydraulic network.

The obtained sigma points are selected by the system observation matrix is calculated as follows:

$$\boldsymbol{Z}_{\boldsymbol{t}} = \boldsymbol{H}\boldsymbol{Y}_{\boldsymbol{t}}^{\boldsymbol{f}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \tag{31}$$

where $\boldsymbol{Z}_t$ is the system observations; $\boldsymbol{Y}_t^f$ is the total forecast hydraulic state vector of all flows and heads at the time step, t; $\boldsymbol{H}$ is the observational matrix and $\boldsymbol{\varepsilon}_t$ is the observational error matrix.

The ensemble mean of the forecast hydraulic state estimates and error covariance approximated by *2L+1* sigma points are:

$$\widetilde{\boldsymbol{Y}_{\boldsymbol{t}}^{\boldsymbol{f}}} = \boldsymbol{W}_{\boldsymbol{t}}^{\boldsymbol{i,y}}\left(\boldsymbol{Y}_{\boldsymbol{t}}^{\boldsymbol{f}}\right)^{\boldsymbol{T}} \tag{32}$$

$$\boldsymbol{P_t^{yy,f}} = \boldsymbol{W_t^{i,p}} \left(\boldsymbol{Y_t^f} - \widetilde{\boldsymbol{Y_t^f}}\right)\left(\boldsymbol{Y_t^f} - \widetilde{\boldsymbol{Y_t^f}}\right)^{\boldsymbol{T}} + \boldsymbol{Q_t} \quad (33)$$

where $\boldsymbol{Y}_t^f$ is the total forecast hydraulic state vector of all flows and heads at the time step, t; $\widetilde{\boldsymbol{Y}_t^f}$ is the ensemble mean of the forecast hydraulic state estimates; $\boldsymbol{W}_t^{i.y}$ and $\boldsymbol{W}_t^{i,p}$ are the weight matrix for hydraulic state ensemble (and the system observations) and forecast error covariance when L-dimension is not zero; $\boldsymbol{P}_t^{yy,f}$ is the forecast error covariance and $\boldsymbol{Q}_t$ is the hydraulic state model process covariance.

The sigma (weight) points for the hydraulic state estimates and forecast error covariance are given by:

$$\boldsymbol{W_t^{0,y}} = \frac{\boldsymbol{\lambda}}{\boldsymbol{L+\lambda}} \quad (34)$$

$$\boldsymbol{W_t^{0,p}} = \frac{\boldsymbol{\lambda}}{\boldsymbol{L+\lambda}} + (\mathbf{1} - \boldsymbol{\alpha^2} + \boldsymbol{\beta}) \quad (35)$$

$$\boldsymbol{W_t^{i,y}} = \boldsymbol{W_t^{i,p}} = \frac{\boldsymbol{\lambda}}{\mathbf{2}(\boldsymbol{L+\lambda})} \quad (36)$$

where $\boldsymbol{W}_t^{0,y}$ is the sigma points (weights) for both forecast hydraulic state estimates and system observation when L-dimension is zero; $\boldsymbol{W}_t^{0,p}$ is the sigma points (weights) for forecast error covariance when L-dimension is zero; $\boldsymbol{W}_t^{i.y}$ and $\boldsymbol{W}_t^{i,p}$ are the weight matrix for hydraulic state ensemble (and the system observations) and forecast error covariance when L-dimension is not zero; β incorporate prior knowledge of the hydraulic state distribution (for Gaussian distributions, $\beta = 2$ is optimal).

The predicted system observations result from:

$$(\widetilde{\boldsymbol{HY}})_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{W_t^{i,y}}\left(\boldsymbol{HY_t^f}\right)^{\boldsymbol{T}} \quad (37)$$

with the auto-covariance of system observations and cross-covariance between the predicted hydraulic states and system observations are calculated as:

$$\boldsymbol{P_t^{zz,f}} = \boldsymbol{W_t^{i,p}} \left(\boldsymbol{HY_t^f} - (\widetilde{\boldsymbol{HY}})_{\boldsymbol{t}}^{\boldsymbol{f}}\right)\left(\boldsymbol{HY_t^f} - (\widetilde{\boldsymbol{HY}})_{\boldsymbol{t}}^{\boldsymbol{f}}\right)^{\boldsymbol{T}} + \boldsymbol{Q_t} \quad (38)$$

$$\boldsymbol{P_t^{yz,f}} = \boldsymbol{W_t^{i,p}} \left(\boldsymbol{Y_t^f} - \widetilde{\boldsymbol{Y_t^f}}\right)\left(\boldsymbol{Z_t} - (\widetilde{\boldsymbol{HY}})_{\boldsymbol{t}}^{\boldsymbol{f}}\right)^{\boldsymbol{T}} \quad (39)$$

where $\boldsymbol{Z_t^f}$ is the system observations; $\widetilde{\boldsymbol{Z_t}}$ is the ensemble mean of system observations (via perturbations); $\boldsymbol{W_t^{i.y}}$ and $\boldsymbol{W_t^{i,p}}$ are the weight matrix for hydraulic state ensemble (and the system observations) and forecast error covariance when L-dimension is not zero respectively; $\boldsymbol{P_t^{zz,f}}$ is the forecast error covariance; $\boldsymbol{R_t}$ is the observation covariance; $\boldsymbol{P_t^{yz,f}}$ is the error cross-covariance between the forecast hydraulic state estimates and system observations; $\boldsymbol{Y_t^f}$ is the total forecast hydraulic state vector of all flows and heads at the time step, t; $\widetilde{\boldsymbol{Y_t^f}}$ is the ensemble mean of the predicted hydraulic state estimates; *T* is the transpose of the designated matrix and $\boldsymbol{H}$ is the observation matrix;

The analysis step of UKF is:

$$\mathbf{X_t^a} = \widetilde{\boldsymbol{X_t^f}} + \boldsymbol{K}\left(\boldsymbol{Z_t} - \widetilde{\boldsymbol{HY_t^f}}\right) \quad (40)$$

$$\boldsymbol{P_t^{yy,a}} = \boldsymbol{P_t^{yy,f}} - \boldsymbol{K}\,\boldsymbol{P_t^{zz,f}}\boldsymbol{K^T} \quad (41)$$

and

$$\boldsymbol{K} = \boldsymbol{P_t^{yz,f}}\left[\boldsymbol{P_t^{zz,f}}\right]^{-1} \quad (42)$$

where $\mathbf{X}_t^a$ is the corrected hydraulic state estimates; $\widetilde{\mathbf{X}_t^f}$ is the ensemble mean of forecast hydraulic state estimates; $\boldsymbol{P}_t^{yy,a}$ is the corrected forecast error covariance; $\boldsymbol{P}_t^{yz,f}$ is the cross-covariance between the hydraulic state vectors and system observations; $\boldsymbol{Y}_t^f$ is the total forecast hydraulic state vector of all flows and heads at the time step, t; $\widetilde{\boldsymbol{Y}_t^f}$ is the ensemble mean of the predicted hydraulic state estimates; $\boldsymbol{Z}_t^f$ is the system observations; $\boldsymbol{P}_t^{zz,f}$ is the system observation error variance; *T* is the transpose of the designated matrix and $\boldsymbol{K}$ is the Kalman gain.

B4. ENSEMBLE KALMAN FILTER METHOD

Ensemble Kalman Filter (EnKF) (Evensen, 1994) is a suboptimal estimator which is suitable for nonlinear system with a large number of state variables. EnKF is widely used for spatial-temporal phenomena evaluation like ocean modelling (Evensen, 2009) and weather forecasting (Myrseth, et al., 2009).

Unlike UKF and EKF, EnKF update the ensemble of forecast hydraulic state parameters individually without the need of covariance matrices or integrating backward in time (Mandel, 2007; Evensen, 2009; Myrseth, et al., 2009). The ensemble of forecast hydraulic state vectors and system observations can be written as:

$$\boldsymbol{X}_t^f = [\boldsymbol{x}_t^{1f}, \quad \ldots\ldots \quad , \boldsymbol{x}_t^{Nf}] \tag{43}$$

$$\boldsymbol{Y}_t^f = \boldsymbol{h}(\boldsymbol{X}_t^f) \tag{44}$$

$$\boldsymbol{Z}_t = [\boldsymbol{y}_t^1, \quad \ldots\ldots \quad , \boldsymbol{y}_t^N] \tag{45}$$

where $\boldsymbol{X}_t^f$ is the ensemble matrix of forecast hydraulic state vectors; $\boldsymbol{Y}_t^f$ is the ensemble of hydraulic state of flows ad pressures from WDS model and $\boldsymbol{Z}_t$ is the perturbed system observations and *N* is the ensemble size.

The ensemble mean of forecast hydraulic state vectors and system observation are:

$$\boldsymbol{\mu_t^y = \frac{1}{N}\sum_{t=1}^{N} y_t^{if}} \quad (46)$$

$$\boldsymbol{\mu_t^z = \frac{1}{N}\sum_{t=1}^{N} z_t^{i}} \quad (47)$$

where μ_t^y and μ_t^z are the ensemble mean of forecast hydraulic state vector and system observation respectively; y_t^{if} and z_t^i are the ensemble members of the forecast hydraulic state vector and system observation respectively.

The principle of EnKF is to approximate the hydraulic state vectors and forecast error covariance from the ensemble statistics (equations (48) and (49)). The ensemble hydraulic state vectors are subjected to hydraulic state model error to represent the possible hydraulic state vectors and observation error covariance is represented by an ensemble of possible observations of the system: The ensemble statistics are:

$$\boldsymbol{C_Q = \frac{E_y {E_y}^T}{N-1}} \quad (48)$$

$$\boldsymbol{C_R = \frac{E_z {E_z}^T}{N-1}} \quad (49)$$

$$\boldsymbol{E_y = Y_t - \mu_y} \quad (50)$$

$$\boldsymbol{E_z = Z_t - \mu_z} \quad (51)$$

where $\boldsymbol{C_Q}$ and $\boldsymbol{C_R}$ are the ensemble hydraulic state vectors and system observation error covariance respectively; T is the transpose of the designated matrix. $\boldsymbol{E_y}$ and $\boldsymbol{E_z}$ are the hydraulic forecast and observation errors respectively**.**

The analysis step of EnKF is:

$$\boldsymbol{X_t^a} = \boldsymbol{X_t^f} + \boldsymbol{K_t}(\boldsymbol{Z_t} - \boldsymbol{H}(\boldsymbol{Y_t^f})) \quad (52)$$

where $\boldsymbol{X_t^a}$ and $\boldsymbol{X_t^f}$ are the analysis and forecast hydraulic state estimates of WDS model; $\boldsymbol{K_t}$ is the Kalman gain; $\boldsymbol{H}$ is the observation operator; $\boldsymbol{Y_t^f}$ is the ensemble of hydraulic state of flows ad pressures from WDS model and $\boldsymbol{Z_t}$ is the perturbed system observations and N is the ensemble size.

The Kalman gain matrix is defined as:

$$\boldsymbol{K_t} = \boldsymbol{C_Q H^T} (\boldsymbol{H C_Q H^T} + \boldsymbol{C_R})^{-1} \quad (53)$$

where $\boldsymbol{C_Q}$ and $\boldsymbol{C_R}$ are the ensemble hydraulic state estimates and system observation covariance respectively, T is the transpose of the designated matrix; $\boldsymbol{K_t}$ is the Kalman gain and $\boldsymbol{H}$ is the observation operator**.**

When the number of system observations gets too large, the Sherman–Morrison–Woodbury Identity (Golub & Van Loan, 1996) is used instead (Kang & Lansey, 2009). It replaces the denominator of Kalman gain, $(\boldsymbol{H C_Q H^T} + \boldsymbol{C_R})^{-1}$ in equation (54) with $\boldsymbol{D_t^{-1}}$ in equation (55). The Sherman–Morrison–Woodbury formula avoids inverting of the sum innovation covariance matrix and observation error covariance, $\boldsymbol{D_t}$:

$$\boldsymbol{D_t} = \boldsymbol{H C_Q H^T} + \boldsymbol{C_R} = \boldsymbol{U V^T} + \boldsymbol{C_R} \quad (54)$$

$$\boldsymbol{D_t^{-1}} = [\boldsymbol{C_R^{-1}} - \boldsymbol{C_R^{-1} U}(\boldsymbol{I} + \boldsymbol{V^T C_R^{-1} U})^{-1} \boldsymbol{V^T C_R}]^{-1} \quad (55)$$

with

$$\boldsymbol{U} = \frac{\boldsymbol{1}}{\boldsymbol{N-1}} (\boldsymbol{H E_y})^{\boldsymbol{T}} \quad (56)$$

$$\boldsymbol{V = HE_y} \tag{57}$$

where $\boldsymbol{C_Q}$ and $\boldsymbol{C_R}$ are the ensemble covariance of hydraulic state estimates and system observation respectively; $\boldsymbol{D_t}$ is the sum innovation covariance matrix and observation error covariance; T is the transpose of the designated matrix; $\boldsymbol{K_t}$ is the Kalman gain; $\boldsymbol{H}$ is the observation operator and $\boldsymbol{E_y}$ is the observational errors**.**

The major key issue in implementing EnKF still persist is quantifying the covariance error matrices. EnKF relies on sampling design to generate both ensemble hydraulic state vectors and system observations. The most common method is the perturbed observations method (Burgers, et al., 1998). The perturbed observations method (Burgers, et al., 1998) involves adding random perturbations to both hydraulic state vectors (equation (58)) and system observations (equation (59)) which introduces sampling errors (Evensen, 2004). This is to prevent the underestimation of analysis error covariance (Chena, et al., 2013):

$$\boldsymbol{X_t^a = \overline{X_t^a} + N(0,\ C_Q)} \tag{58}$$

$$\boldsymbol{Z_t = z_t + N(0,\ C_R)} \tag{59}$$

where $\boldsymbol{X_t^a}$ is the corrected ensemble hydraulic state vectors; $\boldsymbol{\overline{X_t^a}}$ is the ensemble mean of the corrected hydraulic state vectors; $\boldsymbol{z_t}$is the vector of system observations; $\boldsymbol{Z_t}$ is the ensemble perturbed system observations; N is the probability distribution function; $\boldsymbol{C_Q}$ and $\boldsymbol{C_R}$ are the ensemble hydraulic state vectors and system observation error covariance respectively.

The problem with the perturbed observations method is it affects the prior knowledge between the hydraulic state estimates which cause EnKF divergence (Sun. et al., 2009; Sakov and Oke, 2008; Anderson, 2001). Therefore, various deterministic ensemble filters are developed to overcome the limitations of the perturbed observations method. Most of the EnKF variants combine the ensemble mean of

corrected hydraulic state estimates (equation (60)) with the analysis ensemble perturbations:

$$\boldsymbol{X_t^a} = \overline{\boldsymbol{X_t^a}} + \boldsymbol{A_t^a} \quad (60)$$

where $\boldsymbol{X_t^a}$ is the ensemble corrected hydraulic state vectors; $\overline{\boldsymbol{X_t^a}}$ is the ensemble mean of the corrected hydraulic state vectors; $\boldsymbol{A_t^a}$ is the analysis ensemble perturbations matrix;

The analysis ensemble perturbations matrix is derived from transformation of forecast ensemble perturbations through a transform matrix. There are various options of transform matrix:

1) Singular 'Evolutive' Interpolated Kalman (SEIK) (Pham, 2001);
2) Ensemble Square Root Filter (EnSRF) (Whitaker & Hamill, 2001);
3) Deterministic Ensemble Kalman Filter (DEnKF) (Sakov & Oke, 2008);
4) Ensemble Adjustment Kalman Filler (EAKF) (Andersons, 2001);
5) Ensemble Transform Kalman Filter (ETKF) (Bishop, et al., 2001) and
6) Local Ensemble Transform Kalman Filter (LETKF) (Ott, et al., 2004).

*B4.1. S*ingular Evolutive Interpolated Kalman Filter

Singular Evolutive Interpolated Kalman (SEIK filter) (Pham, 2001) is initially developed by Pham et al. (1998). SEIK uses a procedure called minimum second-order exact sampling which preserve the ensemble mean and error covariance of hydraulic state vectors whenever new ensembles are generated (Kang & Lansey, 2009). The efficient analysis ensemble perturbation of SEIK is calculated as:

$$\boldsymbol{A_t^a} = \sqrt{\boldsymbol{N}}\boldsymbol{X_t^f}\mathbb{T}_{\boldsymbol{t}}\mathbb{C}_{\boldsymbol{t}}^{-\mathbf{1}}\boldsymbol{\Omega_t^T} \quad (61)$$

and

$$\mathbb{T}_{\boldsymbol{t}} = \boldsymbol{I} - \frac{\mathbf{1}}{\boldsymbol{N}} \quad (62)$$

$$\mathbb{U}_t^{-1} = \boldsymbol{\rho_u} N \mathbb{T}_t^T \mathbb{T}_t + (\boldsymbol{H X_t^f} \mathbb{T}_t)^T \boldsymbol{R_t^{-1}} (\boldsymbol{H X_t^f} \mathbb{T}_t) \quad (63)$$

where $\boldsymbol{A_t^a}$ is the analysis ensemble perturbations matrix; *N* is the ensemble size; $\boldsymbol{X_t^f}$ is the ensemble of hydraulic state vectors; *N* is the ensemble size; $\boldsymbol{I}$ is the identity matrix; $\mathbb{T}_t$ is a matrix with zero column sums; $\mathbb{C}_t^{-1}$ comes from Cholesky decomposition of $\mathbb{U}_t^{-1}$ (equation (63)); $\boldsymbol{\Omega_t^T}$ is the random matrix whose column are orthonormal and orthogonal to the vector$[\mathbf{1} \quad ... \quad \mathbf{1}]^T$; ρ_u is a forgetting factor which act like covariance inflation error (usually between 0 and 1) to account for other sources of hydraulic state model error; $\boldsymbol{H}$ is the observation operator and $\boldsymbol{C_R}$ is the observation error covariance.

B4.2. Ensemble Square Root Filter

The basic idea of Ensemble Square Root Filter (EnSRF) (Whitaker & Hamill, 2001) is to update hydraulic model state vectors by updating the hydraulic ensemble mean and analysis ensemble perturbations matrix separately. The analysis perturbation ensemble is found from:

$$\boldsymbol{A_t^a} = \left(\boldsymbol{I} - \widetilde{\boldsymbol{K_t}} \boldsymbol{H}\right) \boldsymbol{A_t^f} \quad (64)$$

and

$$\boldsymbol{A_t^f} = \boldsymbol{M_t}\, \boldsymbol{A_{t-1}^a} \quad (65)$$

$$\widetilde{\boldsymbol{K_t}} = [\mathbf{1} + (\, \boldsymbol{R_t}/\boldsymbol{D_t})^{\mathbf{1/2}}]^{-\mathbf{1}} \boldsymbol{K_t} \quad (66)$$

where $\boldsymbol{A_t^a}$ is the analysis ensemble perturbations matrix; $\boldsymbol{A_{t-1}^a}$ is the analysis ensemble perturbations matrix at previous time step; $\boldsymbol{A_t^f}$ is the forecast ensemble perturbations matrix; $\boldsymbol{I}$ is the identity matrix; $\boldsymbol{H}$ is the observation operator; $\boldsymbol{K_t}$ is the Kalman gain; $\widetilde{\boldsymbol{K_t}}$ is the reduced Kalman gain due to nonlinear equation; $\boldsymbol{M_t}$ is the hydraulic state model

operator; $\boldsymbol{R_t}$ is the observation error covariance; $\boldsymbol{D_t}$ is the sum innovation covariance matrix and observation error covariance.

B4.3. Deterministic Ensemble Kalman Filter

Sakov and Oke, (2008) develop a simple Deterministic Ensemble Kalman Filter (DEnKF) that combines the EnSRF's robust performance with the EnKF's simplicity and versatility. The aim of DEnKF is to make the updated error covariance asymptotically match the theoretical value. The analysis perturbation ensemble of DEnKF is calculated as:

$$\boldsymbol{A_t^a} = \left(\boldsymbol{I} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{K_t H}\right)\boldsymbol{A_t^f} \tag{67}$$

where $\boldsymbol{A_t^a}$ is the analysis ensemble perturbations matrix; $\boldsymbol{A_{t-1}^a}$ is the analysis ensemble perturbations matrix at previous time step; $\boldsymbol{A_t^f}$ is the forecast ensemble perturbations matrix; $\boldsymbol{I}$ is the identity matrix; $\boldsymbol{H}$ is the observation operator and $\boldsymbol{K_t}$ is the Kalman gain.

B4.4. Ensemble Adjustment Kalman Filter

Ensemble Adjustment Kalman Filter (EAKF) (Andersons, 2001) is another method to derive the analysis ensemble perturbations matrix:

$$\boldsymbol{A_t^a} = \mathbb{A}_t^T \boldsymbol{A_t^f} \tag{68}$$

the ensemble adjustment matrix $\mathbb{A}_t$ is defined by:

$$\mathbb{A}_t = ((\boldsymbol{F_t^T})^{-1}\boldsymbol{G_t^T})(\mathbb{U}_t^T)^{-1}\boldsymbol{B_t^T}((\boldsymbol{G_t^T})^{-1}\boldsymbol{F_t^T}) \tag{69}$$

where $\boldsymbol{F_t}$ is the rotation matrix of forecast hydraulic state error covariance; $\boldsymbol{B_t}$ is the square root of the ensemble mean of forecast hydraulic state vectors and identity

matrix; $\boldsymbol{G}_t$ is the diagonal matrix with $\boldsymbol{B_t}$; $\mathbb{U}_t$ is the rotation matrix of $\boldsymbol{B_t}$ (not to be confused with $\boldsymbol{U}$ in equation (69) and T is the transpose.

B4.5. Ensemble Transform Kalman Filter

Ensemble Transform Kalman Filter (ETKF) is developed by Bishop, et al., (2001). ETKF requires square root of the covariance matrix via the Sherman–Morrison–Woodbury Identity (Golub & Van Loan, 1996) to update the analysis ensemble perturbations matrix:

$$\boldsymbol{A_t^a} = \boldsymbol{A_t^f}(\boldsymbol{I} + \boldsymbol{V^T R_t^{-1} U})_t^{-1/2} \quad (70)$$

where $\boldsymbol{A_t^a}$ is the analysis ensemble perturbations matrix; $\boldsymbol{A_t^f}$ is the forecast ensemble perturbations matrix; $\boldsymbol{I}$ is the identity matrix; $\boldsymbol{R_t}$ is the observation error covariance; $\boldsymbol{V}$ and $\boldsymbol{U}$ are the product of $\boldsymbol{HE_y}$ and $\frac{1}{N-1}(\boldsymbol{HE_y})^{\boldsymbol{T}}$ respectively.

B4.6. Local Ensemble Transform Kalman Filter

Local Ensemble Transform Kalman Filter (LETKF) (Ott, et al., 2004) performs analysis locally in ensemble space by using only nearby observations within local region. The analysis ensemble perturbations are the product of forecast hydraulic state estimates and local analysis ensemble perturbations:

$$\boldsymbol{A_t^a} = \boldsymbol{X_t^f}\ (\boldsymbol{L_t^a} + \bar{\boldsymbol{L}}_t^{\boldsymbol{a}}) \quad (71)$$

and

$$\bar{\boldsymbol{L}}_t^{\boldsymbol{a}} = \acute{\boldsymbol{P}}_t^{\boldsymbol{a}} \boldsymbol{E_y^T R_t^{-1}} (\boldsymbol{Z_t} - \boldsymbol{\mu_x}) \quad (72)$$

$$\boldsymbol{L_t^a} = [(\boldsymbol{N}-\boldsymbol{1})\acute{\boldsymbol{P}}_t^{\boldsymbol{a}}]^{\boldsymbol{1/2}} \quad (73)$$

$$\acute{\boldsymbol{P}}_{\boldsymbol{t}}^{\boldsymbol{a}} = [(\boldsymbol{N}-\mathbf{1})\mathbf{I} + \boldsymbol{E}_{\boldsymbol{y}}^{\boldsymbol{T}}\boldsymbol{R}_{\boldsymbol{t}}^{-\mathbf{1}}\boldsymbol{E}_{\boldsymbol{y}}]^{-\mathbf{1}} \quad (74)$$

where $\boldsymbol{L}_{\boldsymbol{t}}^{\boldsymbol{a}}$ is the local analysis ensemble perturbations; $\bar{\boldsymbol{L}}_{\boldsymbol{t}}^{\boldsymbol{a}}$ is the ensemble mean of the local analysis ensemble perturbations; z is the ensemble mean of the system observations; $\boldsymbol{Z}_{\boldsymbol{t}}$ is the ensemble system observations and N is the ensemble size; T is transpose; $\boldsymbol{E}_{\boldsymbol{z}}$ is the observation error; $\boldsymbol{R}_{\boldsymbol{t}}$ is the observation error covariance matrix; $\boldsymbol{X}_{\boldsymbol{t}}^{\boldsymbol{f}}$ is the forecast hydraulic state estimates; $\boldsymbol{A}_{\boldsymbol{t}}^{\boldsymbol{a}}$ is the analysis ensemble perturbations; where $\acute{\boldsymbol{P}}_{\boldsymbol{t}}^{\boldsymbol{a}}$ is the local corrected forecast error covariance and $\boldsymbol{E}_{\boldsymbol{y}}$ are the ensemble errors respectively.

B5. PARTICLE FILTER METHOD

Particle Filter (PF) method is a Sequential Monte Carlo method that updates predicted hydraulic state vectors (Doucet, et al., 2001) described by the discrete-time nonlinear state-space model:

$$\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{f}(\boldsymbol{x}_{\boldsymbol{t-1}}^{\boldsymbol{a}}, \boldsymbol{\omega}_{\boldsymbol{t}}) \quad (75)$$

$$\boldsymbol{y}_{\boldsymbol{t}}^{\boldsymbol{f}} = \boldsymbol{h}\left(\boldsymbol{x}_{\boldsymbol{t}}^{\boldsymbol{f}}\right) \quad (76)$$

$$\boldsymbol{z}_{\boldsymbol{t}} = \boldsymbol{H}\boldsymbol{y}_{\boldsymbol{t}}^{\boldsymbol{f}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \quad (77)$$

where $\boldsymbol{x}_t^f$ is the forecast hydraulic state vector at the time step, t which consists of WDS model states and parameters; $f(.)$ models nonlinear function of forecast hydraulic state vector; $\boldsymbol{y}_t^f$ is the total hydraulic state vector of all flows and heads at step, t; $h(.)$ models nonlinear network hydraulics; $\boldsymbol{z}_t$ is the vector of system observations; $\boldsymbol{H}$ is observation operator that relate hydraulic state estimates to system

observations; $\boldsymbol{\omega}_t$ and $\boldsymbol{\varepsilon}_t$ are the vector of hydraulic state and observation noises which are both assumed to be zero mean multivariate Gaussian noises with hydraulic model process covariance $\boldsymbol{Q}_t$ and observation covariance, $\boldsymbol{R}_t$ respectively.

In prediction stage, PF obtains the prior probability distribution for time nonlinear state-space mode:

$$\boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{Z_{t-1}}) = \int \boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{Z_{t-1}})\, \boldsymbol{P}(\boldsymbol{X_{t-1}}|\boldsymbol{Z_{t-1}})\, \boldsymbol{dx_{t-1}} \quad (78)$$

where $P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})$ is the forecast (or prior) probability distribution; $P(\boldsymbol{Y}_t|\boldsymbol{X}_{t-1}, \boldsymbol{Z}_{t-1})$ is the probabilistic model obtained from prediction stage (equation (78) and $P(\boldsymbol{X}_{t-1}|\boldsymbol{Z}_{t-1})$ is the posterior probability distribution from the previous time step.

The main principle of PF is to use Bayesian models (equation (79)) to approximate the posterior probability distribution of forecast hydraulic state vectors at a time step, t when the system observations become available. The posterior probability distribution of forecast hydraulic state estimates given by system observations at time step, t is calculated as follows:

$$\boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{Z_t}) = \frac{\boldsymbol{P}(\boldsymbol{Z_t}|\boldsymbol{X_t})\, \boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{Z_{t-1}})}{\boldsymbol{P}(\boldsymbol{Z_t}|\boldsymbol{Z_{t-1}})} \quad (79)$$

where $P(\boldsymbol{Z}_t|\boldsymbol{X}_t)$ is the probability of system observations given by forecast hydraulic state estimates; $P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})$ is the prior probability distribution of forecast hydraulic state estimates given by the prior observations and $P(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})$ is the normalization factor which can be expressed as $\int P(\boldsymbol{Z}_t|\boldsymbol{X}_t)\, P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})\, d\boldsymbol{x}_t$.

Hutton, et al. (2012) highlighted that it is generally difficult to sample directly from the posterior probability distribution itself, hence, sequential importance sampling (SIS) is used instead (Orhan, 2012). The idea of SIS is to use samples drawn from a proposal probability distribution. In other word, PF approximates the posterior probability distribution at previous time step (t-1) with the weighted set of particles. Unfortunately, there is a discrepancy between the posterior probability distributions at

time step, t and posterior probability distribution from the previous time step, t-1. To compensate for this discrepancy, each sample is weighted:

$$\boldsymbol{P}(\boldsymbol{X_{t-1}}|\boldsymbol{Z_{t-1}}) \approx \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{w_{t-1}^{i}} \boldsymbol{\delta}_{\boldsymbol{X_{t-1}^{i}}} \quad (80)$$

where $\delta_{X_{t-1}^{i}} = \delta(\boldsymbol{X}_{t-1} - \boldsymbol{X}_{t-1}^{i})$ and it is a delta function centred at $\boldsymbol{X}_{t-1}^{i}$; N is the number of ensemble or particle members; i is the index number and w_{t-1}^{i} is the particle weight relative to the posterior distribution of the particles at previous time step satisfying $\sum_{i=1}^{N} w_{t-1}^{i} = 1$ and $w_{t-1}^{i} \geq 0$.

To recursively update these particles (hydraulic state vectors) and the associated weights, the approximation of posterior probability distribution must be obtained. Hence, equation (81) is replaced the posterior probability distribution in equation (80) to get the Monte Carlo approximation (Zuo, et al., 2013):

$$\boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{Z_{t-1}}) \approx \int \boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{X_{t-1}}, \boldsymbol{Z_{t-1}})\, \boldsymbol{P}(\boldsymbol{X_{t-1}}|\boldsymbol{Z_{t-1}})\, \boldsymbol{dx_{t-1}} \quad (81)$$

where $P(\boldsymbol{X}_t|\boldsymbol{Z}_{t-1})$ is the prior probability distribution; $P(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}, \boldsymbol{Z}_{t-1})$ is the probabilistic model; $P(\boldsymbol{X}_{t-1}|\boldsymbol{Z}_{t-1})$ is the posterior probability distribution from the previous time step.

The posterior probability distribution of hydraulic state estimates is further simplified to use the most recent hydraulic state vectors and observations:

$$\boldsymbol{P}(\boldsymbol{X_t}|\boldsymbol{Z_{t-1}}) \approx \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{w_{t-1}^{i}} \boldsymbol{\delta}_{\boldsymbol{Y_{t-1}^{i}}} \boldsymbol{P}\left(\boldsymbol{X_t}|\, \boldsymbol{X_{t-1}^{i}}, \boldsymbol{Z_{t-1}}\right) \quad (82)$$

The prior probability function is now the weighted sum of transition probability distribution with the associated weights. Supposing the equation (82) is substituted into equation (79), the posterior probability distribution becomes the weighted sum of posterior probability distribution with the associated weights:

$$\boldsymbol{P(X_t|Z_t)} \approx \frac{\mathbf{1}}{\boldsymbol{P(Z_t|Z_{t-1})}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{w_{t-1}^i P\left(Z_t|\, X_{t-1}^i\right) P\left(X_t|\, X_{t-1}^i, Z_{t-1}\right)} \quad (83)$$

where $P\left(\boldsymbol{Z}_t|\, \boldsymbol{X}_{t-1}^i\right)$ is the forecast probability distribution; $P\left(\boldsymbol{X}_t|\, \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the probabilistic model of hydraulic states; $P(\boldsymbol{Z}_t|\boldsymbol{Z}_{t-1})$ is the normalisation factor and $w_{t-1}^i$is the associated weight of the particles.

Each particle (forecast hydraulic state estimates) is sampled from the probabilistic model of hydraulic states based on the importance sampling theory (Arulampalam, et al., 2002):

$$\boldsymbol{P\left(X_t|\, X_{t-1}^i, Z_{t-1}\right)} = \frac{\boldsymbol{Q\left(X_t^j|\, X_{t-1}^i, Z_{t-1}\right)}}{\boldsymbol{P\left(X_t^j|\, X_{t-1}^i, Z_{t-1}\right)}} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \frac{\boldsymbol{P\left(X_t^j|\, X_{t-1}^i, Z_{t-1}\right)} \boldsymbol{\delta}_{\boldsymbol{YX_{t-1}^i}}}{\boldsymbol{Q\left(X_t^j|\, X_{t-1}^i, Z_{t-1}\right)}} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{w_t^j \delta}_{\boldsymbol{X_{t-1}^i}} \quad (84)$$

where $Q\left(\boldsymbol{X}_t^j|\, \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the proposal probability distribution where the particles are drawn from; w_t^j is the normalised weight of $\boldsymbol{X}_t^j$.

The particles and the associated weights are updated as follows (Arulampalam, et al., 2002):

$$\boldsymbol{P(X_t|Z_t)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{w_{t-1}^i P\left(Z_t|\, X_{t-1}^i\right) w_t^i \delta}_{\boldsymbol{X_{t-1}^i}} \quad (85)$$

$$\boldsymbol{w_t^i} = \boldsymbol{w_{t-1}^i} \frac{\boldsymbol{P(Z_t|\, X_t^i) P\left(X_t^j|\, X_{t-1}^i, Z_{t-1}\right)}}{\boldsymbol{Q(X_t|\, X_{t-1}^i, Z_{t-1})}} \quad (86)$$

where w_t^i is the updated particle weight; w_{t-1}^i is the particle weight at the previous time step; $(\boldsymbol{X}_t|\boldsymbol{Z}_t)$ is the posterior probability distribution at time step, t; $P\left(\boldsymbol{X}_t^j|\, \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the probabilistic model of hydraulic (or transition probability distribution); $P\left(\boldsymbol{Z}_t|\, \boldsymbol{X}_{t-1}^i\right)$ is the probability of system observations given by forecast hydraulic state estimates and $Q\left(\boldsymbol{X}_t|\, \boldsymbol{X}_{t-1}^i, \boldsymbol{Z}_{t-1}\right)$ is the proposal probability distribution.

In SIS, the probabilistic model of hydraulic (or the transition probability distribution) is selected as proposal probability distribution:

$$\boldsymbol{Q}\left(\boldsymbol{X_t}|\, \boldsymbol{X_{t-1}^i}, \boldsymbol{Z_{t-1}}\right) = \boldsymbol{P}\left(\, \boldsymbol{X_t^j}|\, \boldsymbol{X_{t-1}^i}, \boldsymbol{Z_{t-1}}\right) \tag{87}$$

which further simplified the weight update equation (Zuo, et al., 2013):

$$\boldsymbol{w_t^i} = \boldsymbol{w_{t-1}^i P}\left(\boldsymbol{Z_t}|\, \boldsymbol{X_t^i}\right) \tag{88}$$

Therefore, the updated hydraulic state estimates are approximated from:

$$\boldsymbol{X_t^a} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{X_t^i w_t^i} \tag{89}$$

BIBLIOGRAPHY

Publications and poster presentations by the Candidate

Okeya, I., Hutton, C. and Kapelan, Z., (2015), "Locating pipe bursts in a district metered area via online hydraulic modelling", *Procedia Engineering*, vol. 119, p. 101–110.

Okeya, I., Kapelan, Z., Hutton, C. J., and Naga, D., (2014), “Online Burst Detection in a Water Distribution System Using the Kalman Filter and Hydraulic Modelling”, *Procedia Engineering*, vol. 89, p. 418-427

Okeya, I., Kapelan, Z., Hutton, C. and Naga, D., (2014), "Online Modelling of Water Distribution System using Data Assimilation", *Procedia Engineering*, vol. 70, p. 1261-1270.

LIST OF REFERENCES

Adamowski, J. F., 2008. Peak Daily Water Demand Forecast Modelling Using Artificial Neural Networks. *Journal of Water Resources Planning and Management,* pp. 119-128.

Adamowski, J. & Karapataki, C., 2010. Comparison of Multivariate Regression and Artificial Neural Networks of Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms. *Journal of Hydrologic,* pp. 729-742.

Ahmed, I., 1997. *Application of the Gradient Method for The Analysis of Unsteady Flow in Water Networks.* [Online] Available at: http://arizona.openrepository.com/arizona/bitstream/10150/192098/1/azu_td_hy_e9791_1997_307_sip1_w.pdf [Accessed 17 12 2012].

Aksela, K., Aksela, M. & Vahala, R., 2009. Leakage Detection in a Real Distribution Network Using a SOM. *Urban Water Journal,* 6(4), p. 279–289.

Andersen, E. B., 1970. Asymptotic Properties of Conditional Maximum Likelihood Estimators. *Journal of the Royal Statistical Society,* B(32), p. 283–301.

Andersen, J. & Powell, R., 2000. Implicit State-Estimation Technique for Water Network Monitoring. *Urban Water,* Volume 2, pp. 123-130.

Andersons, J., 2001. An Ensemble Adjustment Ensembled Kalman Filter for Data Assimilation. *Monthly Weather Review,* Volume 129, pp. 2884-2903.

Anjana, G. R., Sheetal Kumar, K. R., Mohan Kumar, M. S. & Amruthur, B., 2015. *A Particle Filter-Based Leak Detection Technique for Water Distribution Systems.* Leicester, England, Proceedings of 13th International Conference on Computing and Control for the Water Industry.

Arsene, T. C., Gabrys, B. & Al-Dabass, D., 2012. Decision Support System for Water Distribution System-based on Neural Networks and Graphs Theory for leakage detection. *Expert Systems with Applications,* Volume 39, pp. 13214-13224.

Arulampalam, M., Maskell, S., Gordon, N. & Clapp, T., 2002. A Tutorial on Particle Filters for Online Non-linear/non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing,* 50(2), pp. 174-188.

Atkinson, A. C., Donev, A. N. & Tobias, R. D., 2007. *Optimum Experimental Designs, with SAS.* 1st ed. Oxford, UK: Oxford University Press.

AWWA, 1999. *Water Audits and Leak Detection.* Denver, USA, American Water Works Association.

Bakker, M. et al., 2013. *Detecting Pipe Bursts Using Heuristic and CUSUM Methods.* Bari, Italy, Proceedings of 12th International Conference on Computing and Control for the Water Industry.

Bargiela, A. & Hainsworth, G. D., 1989. Pressure and Flow Uncertainty in Water Systems. *Water Resources, Planning and Management,* 114(2), pp. 212-229.

Barnes, S. L., 1964. A Technique for Maximizing Details in Numerical Weather Map Analysis. *Journal of Applied Meteorology,* Volume 3, pp. 396-409.

Bertino, L., Evensen, G. & Wackernagel, H., 2003. Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review,* 71(2), pp. 223-241.

Bicik, J., 2010. *A Risk-Based Decision Support System for Failure Management in Water Distribution Networks (A PhD Thesis).* Exeter, UK: University of Exeter.

Bicik, J., K. Z. & Savic, D. A., 2009. *Operational Perspective of the Impact of Failures in Water Distribution Systems.* Kansas City, Missouri, USA, Proceedings of the World Environmental & Water Resources Congress.

Bicik, J., Makropoulos, C., Kapelan, Z. & Savic, D. A., 2009. *The Application of Evidence Theory in Decision Support for Water Distribution Operations.* Concepcion, Chile, Proceedings of the 8th International Conference on Hydroinformatics.

Bishop, C. H., Etherton, B. J. & Majumdar, S. J., 2001. Adaptive Sampling with the Ensemble Transform Kalman Filter. *Monthly Weather Review,* Volume 129, pp. 420-436.

Bishop, C. M., 1994. *Mixture Density Networks. Technical Report, Department of Computer Science and Applied Mathematics,* Birmingham, UK.: Aston University.

Borovik, I., Ulanicki, B. & Skworcow, P., 2009. Bursts identification in water distribution systems. *World Environmental and Water Resources Congress, ASCE.*

Boulos, P. F., Lansey, K. E. & Karney, B. W., 2006. Chapter 5: Network Hydraulic. In: *Comprehensive Water Distribution Systems Analysis Handbook for Engineers and Planners.* 2nd ed. California, USA: MWH Soft, Inc, p. 62.

Bouttier, F. & Courtier, P., 1999. *Data Assimilation Concepts and Methods.* Toulouse, France: ECMWF.

Box, G. P. & Jenkins, G. M., 1976. *Times Series Analysis: Forecasting and Control.* 1st ed. Hoboken, New Jersey: Wiley.

Broad, D. R., Maier, H. R. & Dandy, G. C., 2010. Optimal Operation of Complex Water Distribution Systems Using Metamodels. *Journal of Water Resources Planning and Management, ASCE,* 136(4).

Burgers, G., van Leeuwen, p. J. & Evensen, G., 1998. Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review,* Volume 126, pp. 1719-1724.

Caputo, A. C. & Pelagagge, P. M., 2002. An Inverse Approach for Piping Networks Monitoring. *Journal of Loss Prevention in the Process Industries,* 15(6), pp. 497-505.

Caputo, A. C. & Pelagagge, P. M., 2003. Using Neural Networks to Monitor Piping Systems. *Process Safety,* 22(2), p. 119–127.

Carpentier, P. & Cohen, G., 1993. Applied Mathematics in Water Supply Network Management. *Automatica,* 29(5), pp. 1215-1250.

Chen, H., Hong, Y., Gourley, J. J. & Zhang, Y., 2013. Hydrological Data Assimilation with the Ensemble Square-Root-Filter: Use of Streamflow Observations to Update Model States for Real-time Flash Flood Forecasting. *Advances in Water Resources,* Volume 59, pp. 209-220.

Ching, J., Beck, J. L., Porter, K. A. & Shaikhutdinov, R., 2006. Bayesian State Estimation Method for Nonlinear Systems and Its Application to Recorded Seismic Response. *Journal of Engineering Mechanics, ASCE,* 132(4).

Colombo, A. F., Lee, P. J. & Karney, B. W., 2009. A Selective Literature Review of Transient-based Leak Detection Methods. *Journal of Hydro-environment Research,* 2(4), p. 212–227.

Covas, D. & Ramos, H., 2010. Case Studies of Leak Detection and Location in Water Pipe Systems by Inverse Transient Analysis. *Journal of Water Resources Planning and Management,* 136(2), p. 248–257.

Covey, S. R., 2004. *The 7 Habits of Highly Effective People: Powerful Lessons in Personal Change.* 1st ed. New York: Simon & Schuster, Inc.

Cressman, G. P., 1959. An Operational Objective Analysis System. *Monthly Weather Review,* Volume 87, pp. 367-374.

Cross, H., 1936. *Analysis of Flow in Networks of Conduits or Conductors.* 1st ed. Illinois, USA: University of Illinois.

Darvini, G. & Soldini, L., 2014. *Calibration of Numerical Model of WDS in a Real Case.* Bari, Proceedings of 16th Conference on Water Distribution System Analysis.

Davidson, J. W. & Bouchart, F. J. C., 2006. Adjusting Nodal Demands in SCADA Constrained Real-Time Water Distribution Network Models. *Journal of Hydraulic Engineering, ASCE,* 132(1), pp. 102-110.

Deza, E. & Deza, M. M., 2009. *Encyclopaedia of Distances.* Berlin, Germany: Springer-Verlag.

Doosun, K. & Lansey, K., 2014. Novel Approach to Detecting Pipe Bursts in Water Distribution Networks. *Journal of Water Resources Planning and Management,* pp. 121-127.

Doucet, A. & A., J., 2008. *A Tutorial on Particle Filtering and Smoothing: 15 Years Later.* [Online]
Available at: https://www.stats.ox.ac.uk/~doucet/doucet_johansen_tutorialPF2011.pdf
[Accessed 21 03 2013].

Doucet, A., Freitas, N. & Gordon, N., 2001. *Sequential Monte Carlo Methods in Practice.* 1st ed. New York, USA: Springer.

Duan, H. F., Lee, P. J., Ghidaoui, M. S. & Tung, Y. K., 2012. Leak Detection in Pipes by Frequency Response Method. *Journal of Hydraulic Engineering,* 138(2), pp. 143-153.

Ellul, I. R., 1989. Advances in Pipeline Leak Detection. *Pipes and Pipelines International,* 34(2), pp. 7-12.

Environment Agency, OFWAT & DEFRA, 2012. *Review of the Calculation of Sustainable Economic Level of Leakage and its Integration with Water Resource Management Planning Contract 26777,* Bristol, UK: Strategic Management Consultants.

Evensen, G., 1994. Sequential Data Assimilation with A Non-linear Quasi-Geostrophic Model using Monte Carlo Methods to Forecast Error Statistics. *Geophysics Research,* 49(C5), pp. 10143 - 10162.

Evensen, G., 2003. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. *Ocean Dynamics,* Volume 53, pp. 343-367.

Evensen, G., 2004. Sampling Strategies and Square Root Analysis Schemes for the EnKF. *Ocean Dynamics,* Volume 54, pp. 539-560.

Evensen, G., 2009. The Ensemble Kalman Filter for Combined State and Parameter Estimation: Monte Carlo Techniques. *IEEE Control Systems Magazine,* pp. 83-104.

Farley, B., Boxall, J. B. & Mounce, S. R., 2008. *Optimal Locations of Pressure Meter for Burst Detection.* Proceedings of the 10th Annual International Symposium on Water Distribution Systems Analysis, Kruger National Park, South Africa.

Farley, B., Mounce, S. R. & Boxall, J. B., 2010. Field Testing of an Optimal Sensor Placement Methodology in an Urban Water Distribution Network for Event Detection. *Urban Water,* 7(6), pp. 345-356.

Ferrante, M. & Brunone, B., 2003a. Pipe System Diagnosis and Leak Detection by Unsteady-State Tests. *Advances in Water Resources,* 26(1), p. 95–105.

Ferrante, M. & Brunone, B., 2003b. Pipe System Diagnosis and Leak Detection by Unsteady-State Tests 2 - Wavelet Analysis. *Advances in Water Resources,* 26(1), p. 107–116.

Gabrys, A. & Bargiela, A., 1999. Neural Networks Based Decision Support in Presence of Uncertainties. *Journal of Water Resources, Planning and Management ,* pp. 272-280.

Gabrys, B. & Bargiela, A., 1995. *Neural Simulation of Water Systems for Efficient State Estimation.* Prague, Czech Republic, Proceedings of the Modelling and Simulation Conference.

Gabrys, B. & Bargiela, A., 1999. Neural Networks Based Decision Support in Presence of Uncertainties. *Journal of Water Resource Planning and Management,* 125(5), pp. 272-280.

Golub, G. H. & Van Loan, F. C., 1996. *Matrix Computations.* 3rd ed. Baltimore, Maryland, USA: The Johns Hopkins University Press.

Hardwick, C. P., 2014. *Practical Statistical Process Control: SPC Made Easy! Statistics for Engineers.* United States: CreateSpace Independent Publishing Platform.

Hartley, J. K. & Bargiela, A., 1995. *Parallel Simulation of Large Scale Water Distribution System.* Prague, Czech Republic, Proceedings of the Modelling and Simulation Conference.

Hatchett, S. et al., 2009. *Real-Time Hydraulic Modelling: Open source Software and Results from A Long-Term Field Study.* Cincinnati, USA: University of Cincinnati.

Haykin, S., 1994. *Neural networks: a comprehensive foundation.* New York, USA.: Macmillan.

Herrera, M., Torgo, L., Izquierdo, J. & Perez-Garcia, R., 2010. Predictive Models for Forecasting Hourly Urban Water Demand. *Journal of Hydrology,* pp. 141-150.

Holland, H. J., 1975. *Adaptation in Natural and Artificial Systems.* Michigan: University of Michigan Press.

Hosseinzaman, A., 1995. *The Parallel and Distributed Simulation of Network Systems (A PhD Thesis).* Nottingham, UK: Nottingham Trent University.

Hoteit, I., Korres, G. & Triantafyllou, G., 2005. Comparison of Extended and Ensemble based Kalman Filters with Low and High Resolution Primitive Equation Ocean Models. *Nonlinear Processes in Geophysics,* 12(5), pp. 755-765.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, D. A., 2011. Uncertainty Quantification and Reduction in Urban Water Systems Modelling: Evaluation Report. *PREPARED: Enabling Change, Work Package 3.6, European Commission 7th Framework Programme.*

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, D. A., 2012. Dealing with Uncertainty in Water Distribution Systems' Models: A Framework for Real-Time Modelling and Data Assimilation. *Journal of Water Resources Planning and Management, ASCE,* 140(2), pp. 169-183.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, D. A., 2012. *Real-Time Demand Estimation in Water Distribution Systems Under Uncertainty.* Adelaide, South Australia, Proceedings of 14th Water Distribution Systems Analysis Conference.

Hutton, C. J., Vamvakeridou-Lyroudia, L. S., Kapelan, Z. & Savic, D. A., 2010. Real-Time Modelling and Data Assimilation Techniques for Improving the Accuracy of Model. *Journal Des Sciences Hydrologiques,* 48(5), pp. 679-692.

Hutton, C. & Kapelan, Z., 2015. *Real-time Detection in Water Distribution System Using a Bayesian Demand Forecasting Methodology.* Leicester, England, Proceedings of 13th International Conference on Computing and Control for the Water Industry, pp. 13-18.

Ishido, Y. & Takahashi, S., 2014. *A New Indicator for Real-Time Leak Detection in Water Distribution Networks: Design and Simulation Validation.* Bari, Italy, Proceedings of 16th Conference on Water Distribution System Analysis.

IWA, 2000. *Manual of Best Practice: Performance Indicators for Water Supply Services.* London, UK, International Water Association.

Izquierdo, J., López, P. A., Martínez, F. J. & Pérez, R., 2007. Fault Detection in Water Supply Systems Using Hybrid (Theory and Data-Driven) Modelling. *Mathematical and Computer Modelling,* 46(3-4), pp. 341-350.

Jardak, M., Navon, I. M. & Zupanski, M., 2013. *Comparison of Ensemble Data Assimilation methods for the Shallow Water Equations Model in the Presence of Nonlinear Observation Operators,* Paris, France: Tellus A.

Jazwinski, A. H., 1970. *Stochastic Processes and Filtering Theory.* New York: Academic Press.

Julier, J. S. & Uhlmann, K. J., 1996. *A General Method for Approximating Non-linear Transformation of Probability Distribution,* Oxford, UK: CiteSeer.

Jung, D. & Lansey, K., 2013. *Burst Detection in Water Distribution System using the Extended Kalman Filter.* Bari, Italy, Proceedings of 12th International Conference on Computing and Control for the Water Industry.

Kahl, D. R. & Ledolter, J., 1983. A Recursive Kalman Filter Forecasting Approach. *Institute of Management Sciences,* 29(11).

Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of American Society of Mechanical Engineers,* Series D(82), pp. 35-45.

Kang, D. & Lansey, K., 2009. Real-time Demand Estimation and Confidence Limit Analysis Distribution Systems. *Journal of Hydraulic Engineering,* pp. 825-837.

Kang, D. & Lansey, K., 2009. Real-Time Demand Estimation and Confidence Limit Analysis for Water Distribution Systems. *Journal of Hydraulic Engineering, ASCE,* Volume 135(10), pp. 825-837.

Kang, D. & Lansey, K., 2010. Optimal Meter Placement for Water Distribution System State Estimation. *Water Resources Planning and Management,* Volume 136, pp. 337-447.

Kapelan, Z., Savić, D. A. & Walters, G. A., 2003. A Hybrid Inverse Transient Model for Leakage Detection and Roughness Calibration in Pipe Networks. *Journal of Hydraulic Research,* 41(5), pp. 481-492.

Kapelan, Z., Savić, D. A. & Walters, G. A., 2005. Optimal Sampling Design Methodologies for Water Distribution Model Calibration. *Journal of Hydraulic Engineering,* 131(3), pp. 190-200.

Kitagawa, G., 1996. Monte Carlo Filter and Smoother for Non-Gaussian and Non-linear State Space Models. *Journal of Computational and Graphical Statistics,* Volume 5, pp. 1-25.

Kohonen, T., 1990. *The Self-Organizing Map.* Washington DC, USA, Proceedings of the IEEE, p. 1464–1480.

Kreuzinger, T., Bitzer, M. & Marquardt, W., 2008. State Estimation of a Stratified Storage Tank. *Control Engineering Practice,* Volume 16, pp. 308-320.

Lambert, A., 2012. *Pressure - Bursts Relationships: Prediction, Validation of Results and Extension of Asset Life.* Llandudno, UK: Water Loss Research & Analysis Ltd,

Lambert, A. O., 2002. *International Report: Water Losses Management Techniques,* London, UK: International Water Association.

Lee, P. J., Lambert, M. F., Simpson, A. R. & Vítkovský, J. P., 2006. Experimental Verification of the Frequency Response Method of Leak Detection. *Journal of Hydraulic Research,* 44(5), pp. 451-468.

Lee, P. J. et al., 2007. Leak location in Single Pipelines Using Transient Reflections. *Australian Journal of Water Resources,* 11(1), p. 53–65.

Lee, S. J., Lee, G., Suh, J. C. & Lee, J. M., 2016. Online Burst Detection and Location of Water Distribution Systems and Its Practical Applications. *Journal of Water Resources Planning and Management,* 142(1).

Liggett, J. A. & Chen, L. C., 1994. Inverse Transient Analysis in Pipe Networks. *Journal of Hydraulic Engineering,* Volume 120, pp. 934-955.

Li, R., Haidong, H., Xin, K. & Tao, T., 2015. A Review of Methods for Burst/Leakage Detection and Location in Water Distribution Systems. *Water Science & Technology: Water Supply,* 15(3), pp. 429-441.

Machell, J., Mounce, S. R. & Boxall, J. B., 2010. Online Modelling of Water Distribution Systems. *Drinking Water, Engineering and Science.*

Mandel, J., 2007. *A Brief Tutorial on the Ensemble Kalman Filter,* Denver, USA: Cornell University.

Mashford, J. et al., 2009. An Approach to Leak Detection in Pipe Networks Using Analysis of Monitored Pressure Values by Support Vector Machine. *Network and System Security,* pp. 534 - 539.

McKay, B. & McKay, K., 2013. *The Eisenhower Decision Matrix: How to Distinguish Between Urgent and Important Tasks and Make Real Progress in Your Life". A Man's Life, Personal Development.* [Online] Available at: https://www.artofmanliness.com/2013/10/23/eisenhower-decision-matrix/ [Accessed 14 10 2016].

Mehra, R. & Peschon, J., 1970. An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems. *Automatica,* Volume 7, pp. 637-640.

Misiunas, D., 2004. *Burst Detection and Location in Pipelines and Pipe Networks.* Lund, Sweden: Lund Institute of Technology.

Misiunas, D., Lambert, M., Simpson, A. & Olsson, G., 2005. Burst Detection and Location in Water Distribution Network. *Water Science and Technology: Water Supply,* 5(3), pp. 71-81.

Misiunas, D. et al., 2006. Failure Monitoring in Water Distribution Networks. *Water Science and Technology: A Journal of the International Association on Water Pollution Research,* Volume 53, p. 503–511.

Misiunas, D. et al., 2004. *Burst Detection and Location in Pipe Networks Using a Continuous Monitoring Technique.* Chester, UK, Proceedings of the 9th International Conference on Pressure Surges.

Montgomery, D. C., 2005. *Introduction to Statistical Quality Control.* 5th Ed. ed. Hoboken, New Jersey: John Wiley & Sons.

Mounce, S. R., 2005. *A Neural Network Approach to Burst Detection (A PhD Thesis).* Sheffield, UK: University of Sheffield.

Mounce, S. R. & Boxall, J. B., 2010. Implementation of an On-line Artificial Intelligence District Meter Area Flow Meter Data Analysis System for Abnormality Detection: A Case Study. *Water Science & Technology: Water Supply—WSTWS,* Volume 10.3, pp. 437 - 444.

Mounce, S. R., Boxall, J. B. & Machell, J., 2007. *An Artificial Neural Network/Fuzzy Logic System for DMA Flow Meter Data Analysis Providing Burst Identification and Size Estimation.* Leicester, UK, Proceedings of the 9th International Conference on Computing and Control in the Water Industry.

Mounce, S. R. et al., 2001. A Neural Approach to Burst Detection. *1st IWA Conference on Instrumentation, Control and Automation,* Volume 1, pp. 349-356.

Mounce, S. R. et al., 2002. A Neural Network Approach to Burst Detection. *Water Science and Technology,* 45(4), pp. 237-246.

Myrseth, I. & Omre, H., 2009. Hierarchical Ensemble Kalman Filter. *SPE Journal.*

Myrseth, I., Saetrom, J. & Omre, H., 2009. Resampling the Ensemble Kalman Filter.

Nelder, J. A., 1990. *The Knowledge Needed to Computerise the Analysis and Interpretation of Statistical Information in Expert Systems and Artificial Intelligence: The Need for Information about Data.,* London: Library Association Report.

OFWAT, 2011. *Innovation Priorities for the Water Sector: Innovation priorities for the water sector,* Birmingham, UK: Crown.

Orhan, E., 2012. *Particle Filtering.* [Online] Available at: http://www.bcs.rochester.edu/people/eorhan/notes/particle-filtering.pdf [Accessed 09 02 2013].

Ormsbee, L. E., 1989. Implicit Network Calibration. *Water Resources Planning and Management,* Volume 115.

Ormsbee, L. E. & Lingireddy, S., 1997. Calibrating Hydraulic Network Models. *AWWA: Network Modelling,* 89(2), p. 42.

Ott, E. et al., 2004. A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus,* 56(5), pp. 415-428.

Pauwels, V. R. N. & De Lannoy, G. J. M., 2009. Ensemble-Based Assimilation of Discharge into Rainfall-Runoff Models: A Comparison of Approaches to Mapping Observational Information to State Space. *Water Resources Research,* Volume 45.

Pham, D. T., 2001. Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems. *Monthly Weather Review,* Volume 129, pp. 1194-1207.

Pham, D. T., Verron, J. & Gourdeau, L., 1998. Singular Evolutive Kalman Filters for Data Assimilation in Oceanography. *C. R. Academic Science Ser. II,* 326(4), p. 255–260.

Poulakis, Z., Valougeorgis, D. & Papadimitriou, C., 2003. Leakage Detection in Water Pipe Networks Using a Bayesian Probabilistic Framework. *Probabilistic Engineering Mechanics,* Volume 18, pp. 315-327.

Powell, R. S., 1992. *On-line Monitoring for Operational Control of Water Distribution Networks (A PhD Thesis).* Durham, UK: University of Durham.

Preis, A., Whittle, A., Ostfeld, A. & Perelman, L., 2011. Efficient Hydraulic State Estimation Technique Using Reduced Models of Urban Water Networks. *Water Resources Planning and Management.*

Pudar, R. S. & Liggett, J. A., 1992. Leaks in Pipe Networks. *Journal of Hydraulic Engineering,* Volume 118, pp. 1031-1046.

Puust, R., Kapelan, Z., Savić, D. A. & Koppel, T., 2006. *Probabilistic Leak Detection in Pipe Networks Using the SCEM-UA Algorithm.* Cincinnati, USA., Proceedings of the 8th Annual International Symposium on Water Distribution Systems Analysis.

Puust, R., Kapelan, Z., Savic, D. A. & Koppel, T., 2010. A Review of Methods for Leakage Management in Pipe Networks. *Urban Water,* Volume 7, pp. 25-45.

Qiu, P., 2013. *Introduction to Statistical Process Control.* Oxford, UK: Taylor and Francis.

Quinlan, J. R., 1992. *Learning with Continuous Classes.* Singapore, Proceedings 5th Australian Joint Conference on Artificial Intelligence. World Scientific.

Rao, Z. F. & Salomons, E., 2007. Development of a real-time, near-optimal control process for water-distribution networks. *Journal of Hydroinformatics,* 9(1), pp. 25-37.

Romano, M., 2012. *Near Real-Time Detection and Approximate Location of Pipe Bursts and Other Events in Water Distribution Systems (A PhD Thesis).* Exeter, UK: University of Exeter.

Romano, M., Kapelan, Z. & Savic, D. A., 2010. *Bayesian-Based Online Burst Detection in Water Distribution System.* Sheffield, UK, Integrating Water Systems: Proceedings of the 10th International Conference on Computing and Control in the Water Industry.

Romano, M., Kapelan, Z. & Savic, D. A., 2013. Geostatistical Techniques for Approximate Location of Pipe Burst Events in Water Distribution Systems. *Hydroinformatics,* pp. 634-651.

Romano, R., Kapelan, Z. & Savic, D. A., 2013. Geostatistical Techniques for Approximate Location of Pipe Burst Events in Water Distribution Systems. *Hydroinformatics,* pp. 634- 651.

Russell, S. J. & Norvig, P., 2009. *Artificial Intelligence: A Modern Approach.* 2nd Edition ed. Upper Saddle River, New Jersey: Prentice Hall.

Sakov, P. & Oke, P., 2008. A Deterministic Formulation of the Ensemble Kalman Filter: An Alternative to Ensemble Square Root Filters. *Tellus - Series A,* Volume 60A, p. 361–371.

Salamon, P. & Feyen, L., 2010. Disentangling Uncertainties in Distributed Hydrological Modelling Using Multiplicative Error Models and Sequential Data Assimilation. *Water Resources Research,* Volume 46.

Savic, D. A. et al., 2008. *Project Neptune: Improved Operation of Water Distribution Networks.* Kruger National Park, South Africa, Proceedings of the 10th Annual International Symposium on Water Distribution Systems Analysis.

Schilling, W. & Martens, J., 1986. Recursive state and parameter estimation with applications in water resources. *Applied Mathematical Modelling,* 10(6), pp. 433-437.

Schlatter, T. W., 1988. Past and present trends in the objective analysis of meteorological data for nowcasting and numerical forecasting. *American Meteorological Society,* pp. 9-25.

Shamloo, H. & Haghighi, A., 2011. Transient Generation in Pipe Networks for Leak Detection. *Proceedings of the Institution of Civil Engineers: Water Management,* 164(6), pp. 311-318.

Shang, F. et al., 2006. *Real Time Water Demand Estimation in Water Distribution System.* Cincinnati, Ohio, USA, The 8th Annual International Symposium on Water Distribution Systems Analysis.

Shinozuka, M., Liang, J. & Feng, M. Q., 2005. Use of Supervisory Control and Data Acquisition for Damage Location of Water Delivery Systems. *Journal of Engineering Mechanics,* 131(3), pp. 225-230.

Siemen, 2010. *Flow Measurement.* Cambridge , UK: Siemen AG.

Skworcow, P. & Ulanicki, B., 2011. *Burst Detection in Water Distribution Systems via Active Identification Procedure.* Exeter, UK, 11th International Conference on Computing and Control for the Water Industry.

Sousa, J., Ribeiroa, L., Muranhob, J. & Marques, A. S., 2015. *Locating Leaks in Water Distribution Networks with Simulated Annealing and Graph Theory.* Leicester, England, Proceedings of 13th International Conference on Computing and Control for the Water Industry, pp. 63-71.

Srirangarajan, S. et al., 2010. *Water Main Burst Event Detection and Localization.* Tucson, USA, Proceedings of 12th International Conference on Water Distribution Systems Analysis.

Srirangarajan, S. et al., 2012. Wavelet-Based Burst Event Detection and Localization in Water Distribution Systems. *Journal of Signal Processing Systems,* pp. 1-16.

Stigler, S. M., 1989. Francis Galton's Account of the Invention of Correlation. *Statistical Science,* 4(2), pp. 73-79.

Sun, A. Y., Morris, A. & Mohanty, S., 2009. Comparison of deterministic Kalman filters for assimilating hydrogeological data. *Advances in Water Resources,* Volume 32, pp. 280-292.

Tao, T., Huang, H., Li, F. & Xin, K., 2014. Burst Detection Using an Artificial Immune Network in Water-Distribution Systems. *Journal of Water Resources Planning and Management, ASCE,* 140(10).

Terejanu, G. A., 2008. *Extended Kalman Filter Tutorial,* New York, USA: University at Buffalo.

Tippet, M. K. et al., 2003. Ensemble Square Root Filters. *Monthly Weather Review,* Volume 131, pp. 1485-1490.

Todini, E., 1999. *Using a Kalman Filter Approach for Looped Water Distribution Network Calibration." Water Industry Systems: Modelling and Optimization Applications".* Hertfordshire, England, Research Studies Press.

Todini, E. & Pilati, S., 1987. *A Gradient Method for the Analysis of Pipe Networks.* Leicester Polytechnic, UK, Proceedings of Computer Applications for Water Supply and Distribution.

Todini, E. & Pilati, S., 1988. *A Gradient Algorithm for the Analysis of Pipe Networks, A Computer Applications in Water Supply.* London: John Wiley & Sons.

Tureyen, O. I. & Onur, M., 2011. Investigation of the Use of the Ensemble Kalman Filter (EnKF) for History Matching Pressure Data from Geothermal Reservoirs. *Thirty-Sixth Workshop on Geothermal Reservoir Engineering,* 31 January.

Uhlmann, J., 1995. *Dynamic Map Building and Localization: New Theoretical Foundations (A PhD Thesis).* Oxford, UK: University of Oxford.

UKWIR, 1994. *Managing Leakage 1994,* London, UK: UKWIR.

UKWIR, 2011. *Managing Leakage,* London, UK: UKWIR.

Ulanicki, B., Zegnpfund, A. & Martinez, F., 1996. *Simplification of Water Distribution Network Models.* Rotterdam, Netherlands, International Conference on Hydroinformatics, pp. 493-500.

van der Merwe, R., Doucet, A., de Freitas, N. & Wan, E., 2000. *The Unscented Particle Filter,* Portland, Oregon, USA: Oregon Graduate Institute.

van Leeuwen, P. J., 2009. Particle Filtering in Geophysical Systems. *Monthly Weather Review,* 137(12), pp. 4089-4114.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters. *Water Resources Research,* 39(8).

Walski, T. M., 1983. Technique for Calibrating Network Models. *Water Resources Planning and Management,* Volume 109.

Walski, T. M., Chase, D. V. & Savic, D. A., 2001. *Water Distribution Modelling.* 1st ed. Chicago, USA: Haestad Methods Inc.

Wan, E. A. & van der Merwe, R., 2000. *The Unscented Kalman Filter for Nonlinear Estimation,* Portland, Oregon, USA: Oregon Graduate Institute.

Wang, G., Dong, D. & Fang, C., 1993. Leak Detection for Transport Pipelines based on Autoregressive Modelling. *IEEE Transactions on Instrumentation,* 42(1), pp. 68-71.

Weerts, A. H. & El Serafy, G. Y. H., 2006. Particle Filtering and Ensemble Kalman Filtering for State Updating with Hydrological Conceptual Rainfall-runoff Models. *Water Resources Planning and Management,* Volume 42.

Welch, G. & Bishop, G., 2001. *An Introduction to the Kalman Filter.* UNC-Chapel Hill(Orange County): University of North Carolina.

Welch, G. & Bishop, G., 2006. *An Introduction to the Kalman Filter.* [Online] Available at: http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf [Accessed 04 12 2014].

Western Electric Company, 1956. *Statistical Quality Control Handbook,* Indianapolis, Indiana, USA: Western Electric Company.

Whitaker, J. S. & Hamill, T. M., 2001. Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review,* Volume 130, pp. 1915-1924.

Wu, Z. Y., 2013. *Bentley Applied Research - Darwin Sampler.* [Online] Available at: https://communities.bentley.com/communities/other_communities/bentley_applied_research/w/bentley_applied_research__wiki/6843/darwin-sampler
[Accessed 12 08 2013].

Wu, Z. Y. & Sage, P., 2006. *Water Loss Detection via Genetic Algorithm Optimization-based Model Calibration.* Cincinnati, Ohio, USA, Proceedings of the 8th Annual International Symposium on Water Distribution Systems Analysis.

Wu, Z. Y. & Sage, P., 2007. *Pressure Dependent Demand Optimization for Leakage Detection in Water Distribution Systems.* Leicester, UK., Proceedings of the 9th International Conference on Computing and Control in the Water Industry.

Wu, Z. Y., Sage, P. & Turtle, D., 2010. Pressure-Dependent Leak Detection Model and its Application to a District Water System. *Journal of Water Resources, Planning and Management,* pp. 116-128.

Wu, Z. Y. et al., 2002. *Calibrating Water Distribution Models via Genetic Algorithms.* Kansas City, USA, Proceedings of the Water Resources Planning and Management Conference.

Yang, J., Wen, Y., Li, P. & Wang, X., 2013. Study on an Improved Acoustic Leak Detection Method for Water Distribution Systems. *Urban Water Journal,* 10(2), pp. 71-84.

Ye, G. & Fenner, R. A., 2011. Kalman Filtering of Hydraulic Measurements for Burst Detection in Water Distribution Systems. *Journal of Pipeline Systems Engineering and Practice (ASCE), ,* pp. 14-22.

Ye, G. & Fenner, R. A., 2013. Study of Burst Alarming and Data Sampling Frequency in Water Distribution Networks. *Journal of Water Resources Planning and Management, ASCE.*

Zhang, J., 2001. Statistical Pipeline Leak Detection for All Operating Conditions. *Pipeline and Gas Journal,* 229(2), p. 42–45.

Zhang, W., M. Liu, M. & Zhao, Z., 2009. *Accuracy Analysis of Unscented Transformation of Several Sampling Strategies.* Daegu, South Korea , Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing.

Zuo, J., Liang, Y., Zhang, T. & Pan, Q., 2013. Particle filter with Multi-Mode Sampling Strategy. *Signal Processing,* Volume 93, p. 3192–3201.