FastDetect: DESIGN OF A PERFORMANCE MONITORING METHODOLOGY BASED ON DATA ANALYTICS FOR LARGE-SCALE WATER DISTRIBUTION SYSTEMS IN INDUSTRIAL SETTINGS

By

Hafiz Hashim

BE, MSc,

A thesis submitted to the College of Science and Engineering, National University of Ireland, Galway, in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

2022

Academic Supervisors: Dr. Eoghan Clifford, Dr. Paraic Ryan

Head of Civil Engineering: Prof. Xinmin Zhan

Declaration

I, the undersigned, hereby declare that this thesis, entitled, ‘Design of a performance monitoring methodology based on data analytics for large-scale water distribution systems in industrial settings’, is entirely my own work. The thesis has not been submitted in whole or in part to any other university or institution. All sources used have been acknowledged and referenced in the text.

Hafiz Hashim

Acknowledgements

I would sincerely like to thank Dr. Eoghan Clifford for providing me with the opportunity to work in this field of research. I would also like to thank him for his unwavering support, guidance, and boundary-less enthusiasm for my research.

I would also like to thank Dr. Paraic Ryan for agreeing to co-supervise me. He showed true interest and dedication to this project and pushed me to try and achieve the highest standard and nothing less only ever. I am grateful to the Energy System Integration Partnership Programme (ESIPP) for funding the study.

My parents, family and friends were the last pieces in the puzzle to help me through my PhD. For their unwavering belief in me and their support, I will always be grateful. Brother and sister, for making me feel less alone when I needed somebody on the other end of the phone the most.

I would like to thank the College of Engineering and Informatics in NUI Galway, Ireland for supporting this research, my graduate research committee Prof. Padraic O'Donoghue, Dr. Martin Glavin and Dr. Rory Monaghan and my colleagues in NUI Galway for making my experience so enjoyable.

My beloved mother was one of the greatest sources of joy and happiness in my life growing up, especially in my college years. She was a great inspiration to me, and this thesis is dedicated to her memory.

Abstract

Large non-residential buildings can contain complex and often inefficient water distribution systems. As requirements for water increase due to water scarcity and industrialization, it has become increasingly important to effectively detect and diagnose faults in water distribution system in large buildings. For many of these faults, if water supply is not impacted, water loss can go unnoticed for long periods. This can lead to unnecessary increases in water usage and associated energy losses arising from water pumping, treating, and heating. The majority of fault detection and diagnosis (FDD) studies in the water sector are limited to municipal water supplies and leakage detection. The application of FDD in building water networks remains largely unexplored. While considerable attention has been given to data driven methods that analyse and control energy systems in buildings, the same cannot be said for building water systems. This is a relatively complex and challenging research area, as the non-stationarity (variations in statistical properties over time) of water usage in non-residential buildings makes it challenging to distinguish between routine (normal) and non-routine (anomalous) water uses. As a result, methodologies which support enhanced efficiency in building water consumption are somewhat underdeveloped, particularly in industrial settings.

In such scenarios, FDD methods that leverage multivariate statistical process control with, for example, principal component analysis can be successfully used to identify faults. This research leverages two case-studies, to develop and demonstrate a methodology based on component analysis (PCA) and a multi-class support vector machine (SVM) to detect and classify faults for non-residential building water networks. In the absence of a process model (which is typical for such water distribution systems), PCA is proposed as a data-driven fault detection technique for building water distribution systems for the first time herein. PCA detection indices (T^2-statistics and Q-statistics) were employed to detect faults in the incoming data, and a multi-class SVM was trained for fault classification.

However, even with these methodologies, the non-stationarity of water uses can lead to false alarms being generated. Historically, issues such as false alarm prevalence has led to a relatively low industry uptake of FDD systems. Thus, to efficiently detect and diagnose faults in non-residential building water distribution systems in a manner which

is practical, false alarms should be controlled through false alarm moderation approaches so that building managers/operators only need to focus on true faults.

This research utilises two statistical non-parametric false alarm moderation approaches (namely window-based, and trial-based) that generate a second control limit for T^2-statistics and Q-statistics. The implementation of these false alarm moderation approaches was combined with PCA to detect true faults. Using both the window-based and trial-based approaches false alarms were reduced greatly, and the overall performance and reliability of the overall approach was improved. The PCA model with the window-based approach was shown to be particularly effective in reducing false alarms during fault detection process.

Despite the relatively limited training data available from the case-studies (which can often reflect the situation in many buildings), meaningful faults were detected. The designed FDD methodology (FastDetect) proved successful in discriminating between various types of faults in the case-studies. The effectiveness of FastDetect was compared to a conventional univariate threshold technique through comparison of their respective performance in the detection of faults that occurred in the case-study sites. FastDetect demonstrated promising capabilities when compared to the conventional, in-situ, fault detection systems. The multi-class SVM also allowed these faults to be classified, providing a greater level of information to building managers, which may avoid unnecessary emergency shutdown in industrial applications.

Finally, a comprehensive investigation of the energy and economic impact of water use in the case-study site 1 was conducted. This assessment was then used to develop a cost-benefit analysis for the implementation of FastDetect in industrial settings. Given the links between water and energy systems, faults in a water distribution system can impact overall building energy use and associated carbon footprint. Thus, addressing water and associated energy losses resulting from faults and inefficiencies in water distribution systems can positively impact upon the overall sustainability of non-residential buildings, reducing life-cycle carbon footprint.

Table of Contents

List of Figures

List of Tables

Nomenclature

ANN	Artificial neural network
CPV	Cumulative percent variance
CWS	Cold-water system
DMA	District metering area
ELM	Extreme learning machine
ECOC	Error-correcting output code
FDD	Fault detection and diagnosis
HVAC	Heat ventilation and air conditioning
GWS	Grey-water system
IWA	International water association
m^3/hr	Cubic metre per hour
MSPC	Multivariate statistical process control
RWHS	Rainwater-harvesting system
GWRS	Greywater recycling system
WES	Water end-user segment
WWAP	World water assessment programme
PCA	Principal component analysis
PVC	Polyvinyl chloride
SVM	Support vector machine
NIST	National institute of standards and technology
NPV	Net present value
BCR	Benefit cost ratio
WHAM	Water heater analysis model

List of Notations

X	Data matrix
U	Score matrix
P	Loading matrix
E	residual matrix
R	Real numbers
S	Covariance matrix
Λ	Diagonal matrix
r	Retained principal components
t	Time period
λ	Non-negative eigenvalues
m	Number of features
x	Observation
I	Identity matrix
α	Confidence level for first control limit
β	Confidence level for second control limit
c	Standard normal deviation
s_i	System alarm (monitoring statistic beyond α-control limit)
s	Total number of system alarms (s is the summation of s_i)
v	Observation window
w	Length of observation window
θ	False alarm probability
n	Number of observations
M_i	Monitoring statistics

k	Second control limit
T	Independent trials
$p(s_i)$	Computed probabilities of system alarms
E_{tot}	Total energy use for pumping water
P_{elec}	Power withdrawn from the electricity supply
P_t	Actual power transferred to the water
P_{tot}	Sum of power used for pump start-up and operation
t_{op}	Duration of pump operation
V_t	Total volume of water consumed
Qp	Pump capacity
η_s	System efficiency
V_s	Percentage of total volume pumped during start-up phase
N_s	Number of start-ups
f_s	Percentage of energy consumed during start-up phase
P_{sb}	Power consumed during stand-by
V_{op}	Volume pumped during pump operation
$V_{T,u}$	Upper threshold volume of header tank
$V_{T,l}$	Lower threshold volume of header tank
n_{hw}	Number of header tank
E_{losses}	Energy losses for water heating
V_{hw}	Daily hot-water volume heated
ρ	Density of water
C_p	Specific heat of water
T_{st}	Set temperature (^oC)

T_{in}	Inlet temperature (^oC)
μ	Water heater recovery efficiency
UA	Standby heat loss coefficient
Tamb	Ambient temperature (^oC)
H_{in}	Rated input power of the heater (kW)
H_{out}	Heat content of the water drawn from the heater (kW)
E_{factor}	Energy factor
C_t	Cost cash flow in a year
B_t	Benefit cash flow in a year
Y	Number of years
d	Discount rate

1. INTRODUCTION

1 Introduction

1.1 Background

Increasing water consumption worldwide is primarily due to population growth, increased economic activities and improved standards of living (Ociepa et al., 2019). Water abstraction has increased at double the rate of population growth and presently one-third of the world's population lives in countries, experiencing medium to high water stress (Makaya and Hensel, 2015). By 2025, it is believed that three billion people will suffer from water stress or scarcity, and by 2050 this number will rise to four billion which is about 40% of the expected world's population (San Antonio Water System, 2017).

Many countries worldwide are grappling with the problem of increasing water demand and diminishing water resources. The irony is that the world's water resources are finite and can no longer sustain this increasing water demand unless used wisely. These challenges are exacerbated by water loss from water distribution systems (Mutikanga, 2012). Globally, the amount of water lost through water distribution systems is estimated at 48 billion m^3/year. This generally results from poor management and the condition of water distribution systems (Ociepa et al., 2019).

In Europe, water is also becoming a critical issue and a point of concern. Water stress and scarcity progressively affect several regions of Europe with climate change and population growth predicted to exacerbate the issue directly and through increases in agricultural water uses and industrialization (Eurostat, 2019; British Water, 2019). According to the European Commission, more than half of the European cities are extracting water more rapidly than it can be replenished (European Commission, 2012). By 2030, it is estimated that in Europe water consumption by industries, municipalities, and agriculture will increase by 16% when compared to 2010 figures (European Commission, 2013).

In a water distribution system, water loss is primarily related to the leakage. However, other factors can lead to water loss or inefficient water usage in water distribution systems. These include metering errors, inadequate operational and maintenance practices, high pressures, aging infrastructure, network pipeline corrosion, etc. Water loss through leakage and inefficient operation of water distribution systems often leads to service interruption and can also result in the excess usage of energy resources (Ociepa et al., 2019; Makaya and Hensel, 2015). Operating problems associated with water distribution systems can also include consumer behaviour, degraded equipment, faulty sensors or actuators, poor quality installations, poor maintenance, and improperly implemented controls. These issues can all lead to inefficient operation and the wastage of resources (water and energy) with consequent economical loss and increased equipment wear that can reduce reliability and equipment life (Cugueró et al., 2016).

1.2 Challenges in developing fault detection and diagnosis methods for building water distribution systems

As requirements for water increase due to water scarcity and industrialization, it has become crucial to effectively detect and diagnose faults in water distribution systems in large buildings. Fault detection and diagnosis (FDD) typically utilise inference processes to map system alarms (non-routine water uses or faults) in industrial processes (Zhao et al., 2019). FDD is an area of system engineering which has become essential in the management of many processes, systems, and technology. In the water sector, FDD has focused mainly on leak detection at a municipal water supply network level (Seyoum et al., 2017), while application of FDD in building water distribution systems remains largely unexplored. A range of detection approaches have been studied at municipal level and considered accurate for leak detection (Robles et al., 2016; Escofet et al., 2016; Pérez et al., 2015; Makaya and Hensel, 2015). These approaches are not suitable when it comes to building water distribution systems due to being relatively high cost, labour-intensive and time consuming. In many cases, existing studies in the literature utilise generated or experimental data sets. Due to the presence of non-stationarity and uncertainties in non-residential building water distribution systems, producing such models (building water models) is often difficult and expensive. Non-stationarity may be a feature that exists in residential building water distribution systems, usually linked with the seasonal variation in water consumption and can also be a challenge in non-residential buildings due to season factors, production variations, occupancy variations, etc.

Building water distribution systems are relatively complex systems that comprise pipe networks, meters/sensors, actuators, controllers, and other devices such as control valves and are generally equipped with some (though often the minimum necessary for overall public supply consumption) metering and sensing equipment. For most building managers/owner's initial investment costs of sensors/meters is always a concern and generally install essential equipment required for system monitoring. In most cases, such as non-residential water distribution systems, existing meters/sensors are insufficient to perform FDD and for performance monitoring purposes (Patabendige et al., 2018). The key gaps in developing FDD methodologies for buildings water distribution systems can be summarized as.

- Data availability has been identified as an obstacle to achieve satisfactory operation and maintenance strategies in building water distribution systems. Generally, high-level water use statistics are used to drive alarms which are often not logged to building management systems - unlike energy monitoring for example. Also, historical data can have a significant portion of missing data for a range of reasons including malfunctioning of monitoring equipment and often comprise outlier measurements which can results in unreliable or false detection during the process.

- Developing FDD methodology for non- residential buildings is a relatively complex and challenging research area, as the non-stationarity (variations in statistical properties over time) of water usage makes it challenging to distinguish between routine and non-routine water uses and can incorrectly report one-off or non-routine water uses as faults. The non-stationarity of water usage can also drive false alarms or lead to poor FDD.

- Sensors/meters can be of inadequate accuracy due to the limitations of initial capital investments and poor maintenance. Thus, sensor/meter faults along with the water network faults need to be considered, which adds complexity in developing FDD methodologies (as those same sensors/meters provide the data to drive the FDD process).

- Building water distribution systems are generally equipped with sensors, controllers, meters, and equipment from different manufacturers. There are no such standards for installing sensor/meter exists, which makes much more difficult to develop generalized FDD methodologies for building water distribution systems.

1.3 Research objectives

There has been limited work to date specifically targeting the development of FDD methodologies for non-residential building water distribution systems. However, methodologies have been developed for other engineered systems that can be leveraged in the development of FDD methodologies to improve the reliability and efficiency of the building water distribution systems. The specific aims undertaken to address these research objectives are listed below.

- Investigate the use of temporally sensitive unsupervised and supervised learning models (namely PCA and SVM), capable of identifying non-routine water uses and faults of different types such as equipment malfunctioning, high or low-level imperceptible process faults in addition to water leakage. These approaches have the potential to address the shortcomings in the literature associated with conventional methods, addressing industry needs.

- Combine the proposed above FDD methodology with an outlier localization approach, which can robustly identify outliers within the water time series and considers the non-stationarity of water use as state of things rather than process disturbances.

- Demonstrate the use of non-parametric false alarm moderation approaches which can detect true faults while maintaining a low false alarm probability to limit false detection. The addresses the high incidence of false alarms during FDD process.

- Demonstrate the above using two case-study sites and investigate the costs and benefits associated with implementing these approaches into building water networks.

- Develop an overall framework for the implementation of the above fault detection and false alarm moderation approaches into building water networks.

1.4 Structure of dissertation

The Thesis is organised into seven chapters and one appendix. Chapter 1 presents an introduction to research, an overview of some of the important challenges in existing knowledge in this area, the research objectives, and the structure of dissertation are presented.

Chapter 2 provides essential background information relating to water consumption by different sectors and by different building types with a focus on FDD and its role in sustaining water resource. Finally, Chapter 2 presents a detailed and focused discussion on the FDD approaches used in the literature for the performance monitoring of water distribution systems to improve the water efficiency. The shortcomings in the existing literature are presented and the way in which these shortcomings helped form the research questions themselves is discussed.

Chapter 3 provides comprehensive discussion on the methodology development of FastDetect for non-residential water distribution systems which gives an overall understanding of the more sophisticated machine learning models used in this research.

Chapter 4 of this thesis presents the FDD methodology developed (FastDetect) based on the machine learning models. The primary aim of this chapter is to assess the efficacy of the designed FDD methodology in detecting and diagnosing among different types of faults in a case-study building water systems. Subsequently, the FDD results are discussed in detail and the results are compared to known faults in the building and the results from a conventional FDD approach. This chapter has been published as a peer reviewed journal paper (see Chapter 1.5.1).

Chapter 5 of this thesis presents the utilization of non-parametric false alarm moderation approaches to control the prevalence of false alarms during fault detection process. The work in this chapter is aimed at addressing the issue of false alarms usually resulted due to non-stationarity of water use in non-residential buildings. Again, this is done through utilising real case-study sites data. Subsequently, the false alarm moderation and FDD results are discussed, before being compared to a conventional monitoring system output.

Chapter 6 of this thesis presents the discussion on the economic aspect of water-energy interaction linked with faults in non-residential water distribution systems and quantifying the energy use for pumping water. The first section of this chapter analyses the energy and economic impact of water use in the case-study 1. The later section utilizes the analysis and develop a cost-benefit analysis to assess the cost and benefits associated with the implementation of the FastDetect in industrial settings.

Chapter 7 of this thesis presents the discussion on the potential benefits from application of the FastDetect in the water industry with a focus on applying the FastDetect at a national level in an Irish setting.

Chapter 8 concludes the thesis by outlining how the work presented in this thesis has addressed the research questions. In addition, recommendations, and some promising future directions for research are presented.

1.5 Contributions to knowledge

1.5.1 Invention disclosure

Designed performance monitoring methodology build on data-analytics and machine learning approaches for non-residential water distribution systems (**FastDetect**) – Invention Disclosure Form # Tech-2020-014.

1.5.2 Journals

Hashim, H., Clifford, E., Paraic R., 2021. False alarm moderation for performance monitoring in industrial water distribution systems. Journal of Advanced Engineering Informatics (under review).

Hashim, H., Paraic R., Clifford, E., 2020. A statistically based fault detection and diagnosis approach for non-residential water distribution system. Journal of Advanced Engineering Informatics (https://doi.org/10.1016/j.aei.2020.101187).

Nair, S., Hashim, H., Hannon, L., Clifford, E. 2018. End use level water and energy interactions: A large non-residential building case study. Journal of Water (https://doi.org/10.3390/w10060810).

1.5.3 Conferences

Hashim, H., Paraic R., Clifford, E., Principal component analysis-based fault detection and isolation for building water distribution system: A case-study. Civil Engineering Research in Ireland, Cork, Ireland, 27-28 August 2020.

Hashim, H., Paraic R., Clifford, E., An event-triggered fault detection for non-residential water distribution system: A case-study. ENVIRON 2020 conference, Dublin, Ireland 20 October 2020.

Hashim, H., Paraic R., Clifford, E., An event-based fault detection and isolation approach for non-residential water distribution system: A case-study. 9th International Young Water Professionals Conference, Toronto, Canada 23-27 June 2019.

Hashim, H., Paraic R., Clifford, E., An event-triggered fault detection for non-residential water distribution system: A case-study. ENVIRON 2019 conference, Carlow, Ireland 15-17 April 2019.

Nair, S., Hashim, H., Mulligan, S., Ryan, P., Keane, M., Clifford, E., Hannon, L., Quantifying energy losses associated with faults in building water networks using a novel fault detection, diagnosis, and optimization approach. 12th SDEWES Conference, Dubrovnik, Croatia 4-8 October 2017.

1.5.4 Industry article

Fault detection and diagnosis in building water systems: Overview of challenges and the application of supervised and unsupervised learning models for two case-study buildings, 2021 (https://issuu.com/patlehane/docs/building_services_engineering_may_june_2021_web_fi).

2. LITERATURE REVIEW

2 Literature Review

2.1 Overview

Water is essential for life and a fundamental resource for human development (Houngbo et al., 2018). Water is not only necessary for everyday human needs, but also for economic development, energy production, and for industrial and agricultural processes (UNESCO, 2015) and is conveyed through municipal water supply network to end users (**Figure 2.1**). It is well known in the municipal sector that water losses through leakage is a substantial barrier to sustainable water uses (European Commission, 2019). This has been the subject of significant research. However, much of this water is destined for non-residential buildings which in themselves can be a significant source of inefficient water use (Patabendige et al., 2018). Water distribution systems (and end uses) within this sector have received significantly less attention (Mannan and Al-Ghamdi, 2020).

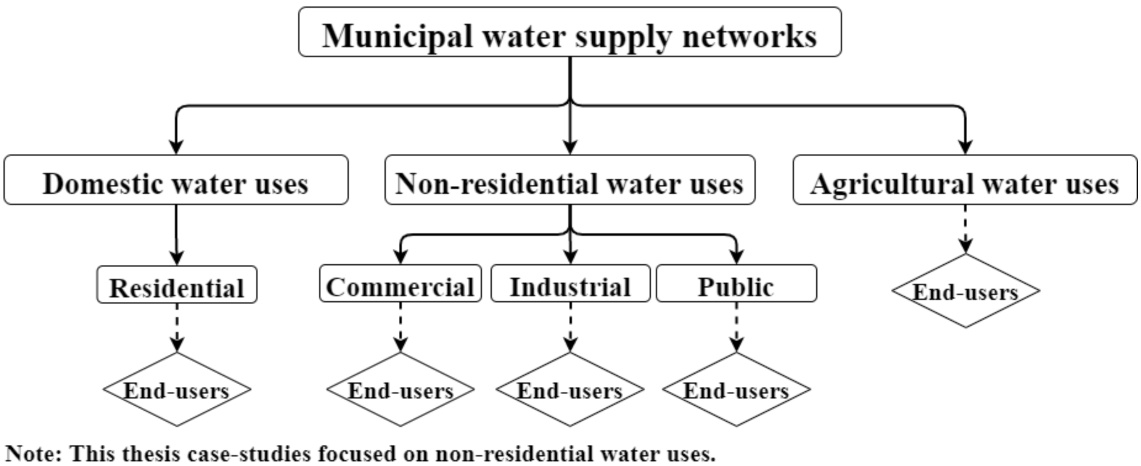


Figure 2.1: Municipal water supply network.

There are numerous factors that can lead to water losses or ineffiicient water use in water distribution systems such as metering errors, inadequate maintenance practices, leakage, end-user behaviour, high pressures, etc. Since water losses are inevitable, various tools and methods have been developed and implemented for reducing this in water distribution systems with a particular focus on municipal water supply networks and domestic water uses. However, perhaps due to their complexity, non-residential water distribution systems such as large-scale buildings or industrial facilities have received limited attention (Prabuchandran et al., 2019; Mannan and Ghamdi, 2020). Thus, the development of a sustainable and robust performance monitoring methodology for non-residential water distribution systems is required to tackle the issues linked to water losses (faults), and to ensure the optimal operation of the non-residential water distribution systems. This is a necessary step towards maintaining and improving water resource management in non-residential buildings.

This chapter outlines the present state of non-residential building water distribution systems performance monitoring strategies and how this research aims to contribute to its development. To establish context for this research, the initial sections discuss challenges with water resources globally and in Europe and the extent to which water is used by sectors and by different buildings in Europe with a emphasizes on fault detection and diagnosis (FDD) in sustaining water resource. The topic of FDD is introduced and the FDD approaches used at a municipal and building level water systems reviewed and provide a background for the gaps in existing literature and practice. Later sections then provide discussion on the FDD methods utilized in context of performance monitoring of building water distribution systems.

2.2 Water use by sectors globally and in Europe

Water use has increased globally by an annual rate of about 1% since the 1980s (Houngbo, 2019). By 2050, global water demand is projected to rise by 55% compared to 2000, primarily due to growing demand from energy generation, industrial processes, and domestic water uses (UNESCO, 2015). Indeed, the agricultural sector through irrigation, livestock husbandry, and aquaculture is considered to be the largest water consumer, accounting for about 69% of annual water abstraction (**Figure 2.2**). The non-residential sector is considered to be one of the major water consumers worldwide ranking only second to the agricultural sector (Seiler et al., 2019; WWAP, 2018). The non-residential sector including manufacturing, energy generation, etc. accounts for about 19% and domestic for around 12% **- Figure 2.2** (Houngbo, 2019; UNESCO, 2015). Managing the efficient use of water is an increasingly challenging, yet urgent task worldwide as inefficient water use can bring various adverse economic, and environmental impacts (Makaya and Hensel, 2015; Washali et al., 2020).

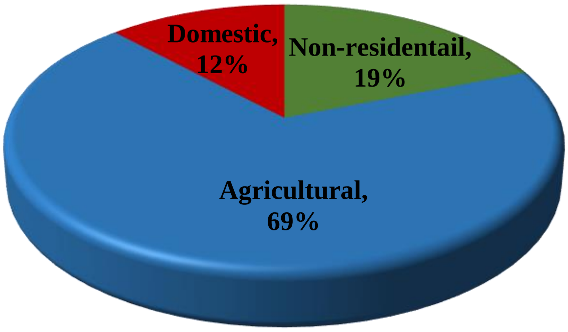


Figure 2.2: Global water abstraction per sector (Houngbo, 2019).

Work to improve the efficiency of water abstraction and consumption generally focuses on two areas i.e., reducing the water use by end-users and improving the efficiency of infrastructure that supplies water including rapidly detecting and resolving faults that can appear within water distribution systems (UNESCO, 2015; European Commission, 2019). Globally in water distribution systems, faults can lead to leakage within the network and are projected to account for up to 30% of the total water extracted (Gertler et al., 2010; Makaya & Hensel, 2015; Farley & Trow, 2003). In Europe it is estimated that 20% of water abstracted is wasted through leakage which could be reduced by improving the water efficiency by 10-30% through technological advancement (European Commission, 2019). Indeed, for water distribution systems leakage is considered to be the main source of water loss (European Commission, 2019; Gupta and Kulat, 2018).

Water loss through leakage varies widely among European nations (**Figure 2.3**), ranging from 4% to nearly 44%, with an average value of 23% (EurEau, 2017). The Netherlands and Germany have the lowest leakage rates in Europe of between 5-7% (EurEau, 2017; European Commission, 2019). While Ireland's leakage is one of the highest at 41% (Irish Water, 2019; EurEau, 2017). Improvements in leakage rates are not only a prerequisite for making water distribution systems sustainable but also an environmental necessity. Leakage poses a risk to public health and safety (e.g., leaking flow may damage the foundations of a building). In fact, it is not only a matter of an efficient use of water resource but also of energy, since a water distribution system with reduced water loss requires less pumping energy to serve end-users (Cavazzini et al., 2020; Ociepa et al., 2019).

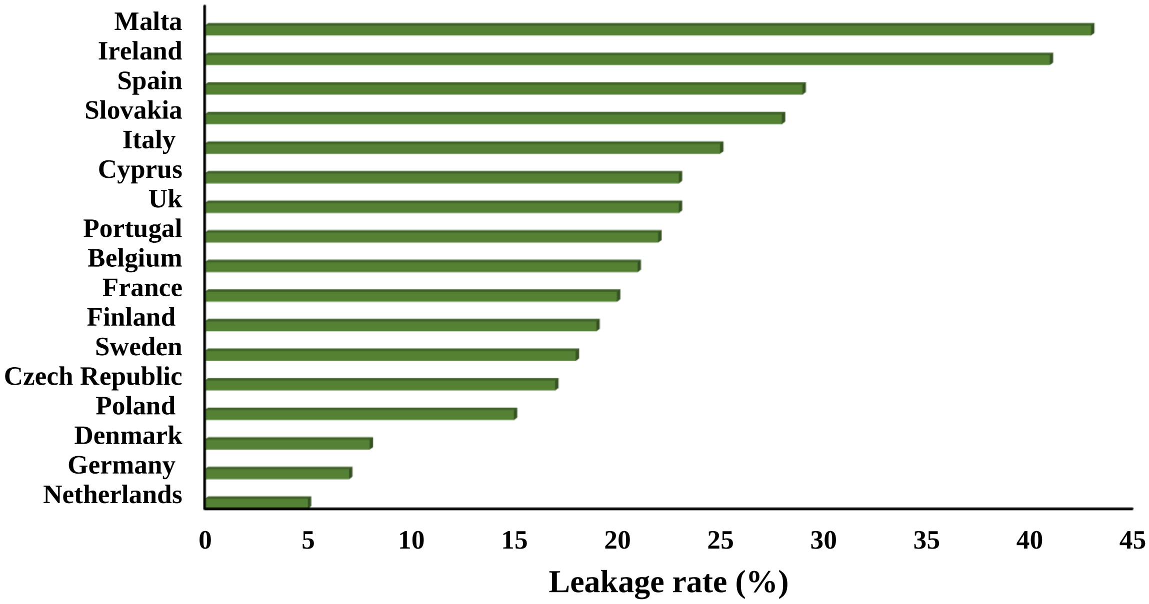
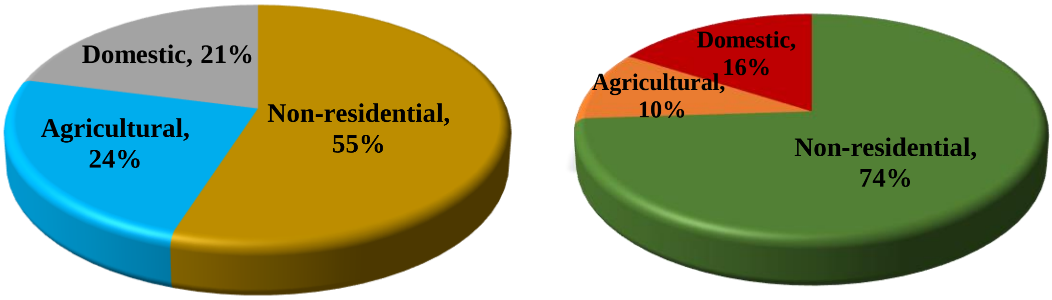


Figure 2.3: Average leakage rates among European countries (Irish Water, 2019; European Commission, 2019; EurEau, 2017).

2.2.1 Water uses by sectors in Europe

In Europe, the non-residential sector consumes about 28% of total water withdrawn (European Commission, 2012) – this being part of the 55% withdrawn for industrial uses. More recently, it has been estimated that about 30% of water withdrawn is consumed by non-residential buildings in Europe (Mannan and Ghamdi, 2020). While the agricultural and domestic sectors represent 24% and 21% of the total abstracted water, respectively (**Figure 2.4a**).



(a) **(b)**

Figure 2.4: (a) water abstraction per sector within Europe, and (b) water abstraction per sector within Ireland (European Commission, 2012; European Commission, 2013).

However, the water abstraction of each sector varies greatly depending on the location and economic development of a country as does whether water is ultimately returned to a water body. For instance, in Ireland, the non-residential sector (including manufacturing processes, and energy generation) returns most of the abstracted water (74%) directly to a water body, while in the domestic and agricultural sector it is not the case. The domestic sector represents 16% and agricultural sector represents 10% of the total abstracted water, respectively - **Figure 2.4b** (European Commission, 2013). The increased water abstraction in non-residential sector could be due to manufacturing activities, as Ireland is home to 18 of the world's top 25 medical device manufacturers and is considered as the second largest country for exporting medical devices in Europe (Irish Medtech Association, 2019). Strategies to manage water uses have been developed in many European countries, and have focused mainly on municipal water supply networks (Washali et al., 2020; European Commission, 2019). However, due to the increased pressure of urbanization and improved living standards, the amount of water use in all buildings is also projected to rise substantially but this has received considerably less attention (Ociepa et al., 2019; European Commission, 2019).

2.2.2 Water uses by buildings in Europe

In general, data and research has focused on the residential sector which is more uniform than the non-residential sector due to the array of different non-residential building types

(**Figure 2.5**) which are a function of consumer needs and production processes (EurEau, 2017; Alsaydalani, 2017).

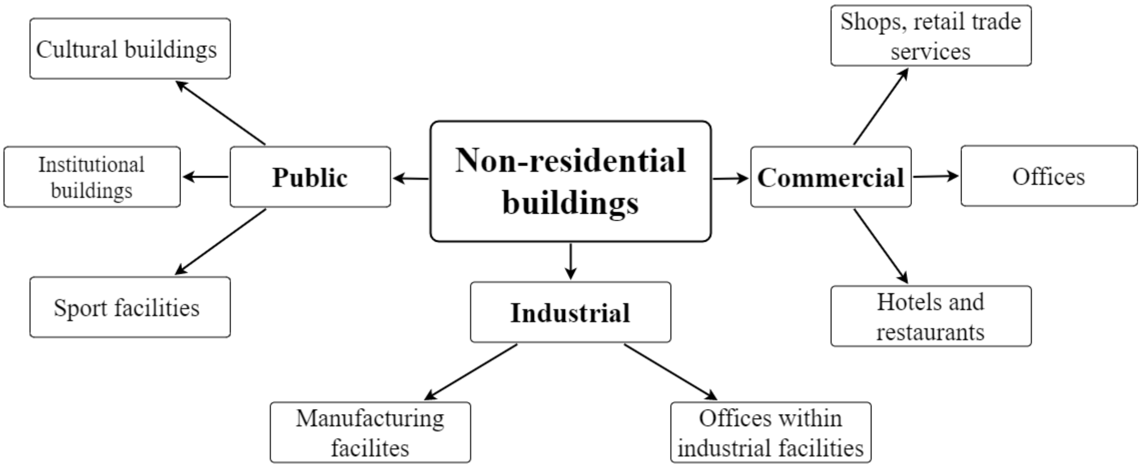


Figure 2.5: Non-residential building types (BIS and Cranfield University, 2009).

Water use in non-residential buildings can be categorized under various headings as summarised in **Figure 2.6**. For instance in an Irish context, toilet and urinals in office buildings consumes about 37% of their water use, while restaurants and cafés use about 52% mostly in kitchen/dishwashing activities (Ireland Environment, 2016). This is somewhat in contrast to European wide reports which estimate that in non-residential buildings, toilets and urinals account for 70-95% of the total water use (European Commission, 2012; EPA, 2013).

Figure 2.6: Water usage in non-residential building (Mannan and Ghamdi, 2020; European Commission, 2013).

There is limited data on the quantity of water which is lost in non-residential buildings due to leaks or other faults or simply inefficient operation (Ihasalo and Karjalainen, 2014). To the best of the author's knowledge there are no widespread statistics available in the literature regarding the prevalence or impact of inefficient water use specifically in non-residential water distribution systems. A small number of studies have respectively estimated that in non-residential buildings leaks account for up to 28% of total water use (Sydney Water, 2011) or that between 10-30% of water is lost through the building water distribution system which includes pipe network, plumbing fixtures, etc. (Aşchilean et al., 2017; EPA, 2013). It has also been reported that water loss in buildings is primarily linked to leakage within the building water distribution systems (Datta & Sarkar, 2016; Skworcow et al., 2013). However, even in systems without significant losses due to leakage, building water distribution systems particularly in developing countries continue to operate in an inefficient manner resulting in water and revenue losses (Mutikanga, 2012). This has resulted in a growing focus on optimising water (and associated energy) usage across all building types using building management system to detect problems (Sousa et al., 2019), particularly given the water scarcities impacting many countries today (Irish Water, 2017). Water is often an ignored resource in terms of research in the building sector; while there has been a significant body of research regarding embodied and recurring energy use in buildings, limited attention has been drawn to water use in buildings particularly in case of non-residential buildings (Ociepa et al., 2019; Coelho et al., 2014) - indeed, there are on average 80% more studies on energy use in buildings than water use in buildings on Scopus for instance[1].

Furthermore, there have been a number of recent regulatory and policy interventions including BREEAM (Building Research Establishment Environmental Assessment Method), LEED (Leadership in Energy and Environmental Design) which are designed to ensure improved sustainability of water consumption in buildings. Sustainability criteria usually include different aspect of water efficiency and management such as water usage, water use reduction, monitoring leakage and inefficient water use, use of high-performance water-using fittings, and recycling or harvesting water, etc. (Al-Qawasmi et al., 2019). This sustainability drive is resulting in a growing focus on optimising water (and associated energy) use across all building types by detecting and diagnosing faults at an early stage (Sousa et al., 2019). It is also important to note that data availability for non-residential building water distribution systems represents a significant and persistent challenge when compared to municipal water supply networks (Matos et al., 2020). In many such cases reliable and objective information about water use and network performance is poor, lacking or otherwise unavailable (Houngbo, 2019). This underlines the need of a robust performance monitoring methodology which is capable of operating under these practical data constraints (poor data quality, missing data, etc.) and assessing the effectiveness of a designed FDD methodology in identifying the faults of different nature in addition to leakage. Where possible data from non-residential water distribution systems is used, however, much of the literature has focused on municipal water supply networks and given its relevance this has also been reviewed (Coelho e al., 2014; Mulligan et al., 2020; Cody and Narasimhan, 2020). The application of fault detection approaches developed for municipal water supply networks may be limited within non-residential buildings as key faults not only involve leaks but also comprise system related faults such as equipment malfunctioning, operational errors, etc. or non-routine water uses caused by inefficient water usage and infrequent sudden changes in demand such as hosting a one-off event or increased production for a short

[1] Search term used (2016 - 2020):
Fault detection and diagnosis in building water distribution systems
Fault detection and diagnosis in water supply networks
Performance monitoring of water distribution systems

period of time, etc. (Patabendige et al. 2018; Prabuchandran et al., 2019). While these can occur at a municipal level they may not be as prevalent when compared to the overall capacity of the system.

2.2.3 Broad impacts of leakages in water distribution systems

Water loss resulting from leakage and inefficient use has been considered as a serious issue since of the 1980's. (Gupta and Kulat, 2018). Leakage issues in building water networks can result from aging infrastructure and inadequate maintenance practices which in turn cause undesirable service disruptions and significant water losses (European Commission, 2012; Merzi and Özkan, 2015). Inefficient operational performance of the water distribution systems can result directly in leakages but also in poor infrastructure condition. This is the case when water distribution systems are not protected against pressure surges (unusual pump start-ups) which cause significant pressure fluctuations within a network and lead to network pipe degradation causing leaks and burst pipes (Arregui and Carlos, 2012). Water loss through leakage can cause damage to the building itself (e.g., if low-level leakage is not get detected and remains there for a period of time it can result in structural damage to a building foundations). Leakage not only cause added operating cost but also has negative social and ecological impact (European Commission, 2018). Currently available methods for detecting and identifying leaks in buildings are often not considered economical and require detailed information about the water use and network performance (Matos et al., 2020). Non-residential building water distribution systems are usually studied in either assessing water efficiency performance or providing financial assessment, limited attention has been given in detecting leaks in buildings. (Mannan and Ghamdi, 2020; Sousa e al., 2019; Kern et al., 2016; Gois et al., 2015; Cook et al., 2014; Matos et al., 2013). This is in contrast to municipal water networks where considerable work in detecting and identifying leaks targeted at municipal water supply networks has been carried out (Cavazzini et al., 2020; Seyoum et al., 2017; Robles et al., 2016; Escofet et al., 2016; Makaya and Hensel, 2015; Moors et al., 2018; Abdulshaheed et al., 2017; Moser et al., 2015).

In building water distribution systems, energy cost linked to pumping water represent a significant fraction of the overall operational cost of the system (Fletcher et al., 2018). A number of studies have explored the interaction between water and energy use in municipal water supply networks (Ercole et al., 2016; Siddiqi and Weck, 2013;

Skworcow et al., 2013; Colombo & Karney, 2005; Colombo & Karney, 2003). For example, Colombo & Karney, (2003) published an initial approach to analyse the impact of leaks on municipal water supply network performance (impact on design, cost, and performance). This study discussed the cost and trade-off of water and energy use associated with typical municipal water supply networks by highlighting the key elements (water demand, treatment, and pumping capacities), interactions, and life cycle assessment issues. At a laboratory scale, a single leaky pipe segment was considered to analyse the impact of leaks on energy and water consumption and the energy efficiency of a leaky pipe was studied with respect to leak location and size. A follow-on study by (Colombo & Karney, 2005) analysed various topological configurations of the domestic water network, and leak intensities by analysing the energy influence of leakage on pumping with and without storage and associated energy cost. In the study five distinct scenarios of different leakage levels were evaluated and it was concluded that the relative increase in energy cost was found to be greater than the cost incurred due to leakage when identical pressure requirements were considered.

More recently a model-based study was performed by Ercole et al., (2016) to analyse municipal water supply network performance using resource input-output analysis. A water balancing approach (i.e., ensuring that water input is equal to the water output) was used to reduce the water losses through leakage and reduce energy use within the water supply network. Additionally, the associated environmental impact and the cost involved in optimising the water supply network management were also estimated. Siddiqi & Weck, (2013) proposed an analytical framework that can be utilized at the planning stages of new municipal developments to assess future building level water demand and associated energy use such as for heating water, recycling water and on-site pumping. The study focused on the water and energy uses in residential buildings and found that the energy intensity in the water end-use segment (WES) can be significant in households and buildings. It was estimated that for every unit reduction (in litre/person/day) of indoor residential water use up to 225 MWh could be saved annually. Specifically, the energy requirement for water heating dominates when compared to on-site booster pumping and grey-water recycling systems. However, it is still challenging to obtain comprehensive information on water and associated energy use in non-residential buildings, since it entails more detailed monitoring and measurements than what is often available (Kern et al., 2016; Bylka and Mroz, 2019). In non-residential buildings energy cost related to water distribution systems can rise non-linearly with the faults (e.g., leakage) (Abdulshaheed et al., 2017). In this manner, improvements in performance monitoring of non-residential water distribution systems can result in substantial water-energy resource conservation.

2.3 Leak detection in water distribution systems - state-of-the-art

Faults in water distribution systems mainly arise due to leaks in transmission/distribution mains, leak/overflow in storage facilities or leak in service connections up to the point of meters at user-end (Bernieri et al., 2017; European Commission, 2019). Conventionally, water authorities conduct assessments and audits periodically at a municipal level to estimate leakage from water distribution systems. High level audits are supplemented by targeted leak detection and identification to validate the presence of leaks (Puust et al., 2010).

Several leak detection approaches have been formulated and studied to help identify leaks with research, almost exclusively, focusing at municipal water supply network level (Seyoum et al., 2017). A range of hardware-based and software-based leak detection approaches such as leak-noise correlation, acoustic sensing, water balancing, pressure management, etc. have been studied for these municipal systems, for both steady-state and transient-state water flow conditions - **Figure 2.7** (Xing and Sela, 2019; Zhou et al., 2019; Aghda et al., 2018; Robles et al., 2016; Escofet et al., 2016; Pérez et al., 2015; Makaya & Hensel, 2015). The hardware-based approaches are usually linked to physical aspect of leak detection (i.e., instruments used for leak detection), while software-based approaches rely on computing and data acquisition techniques (Stevenson, 2020).

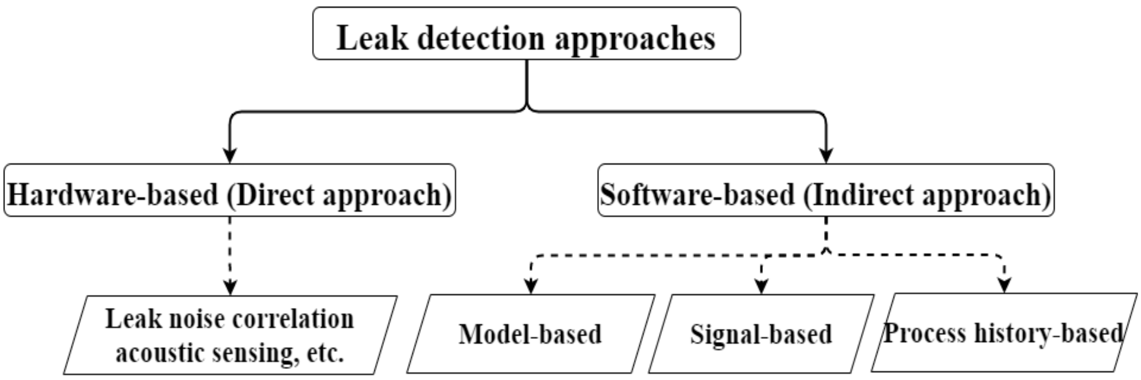


Figure 2.7: Leak detection approaches.

In a number of studies leak detection approaches were also categorised either as direct means of detecting leaks (hardware-based) and indirect means of detecting leaks (software-based) (Merzi and Özkan, 2015). Direct methods generally involve the use of commercially available leak detection tools (instruments or sensors) such as acoustic devices, leak noise correlation devices and ground penetrating radar imaging devices (Lai et al., 2018; Moser et al., 2015; Liu and Kleiner, 2013) - discussed further in Section 2.4.1. These methods rely on the noise or vibration signals resulting from leakage within a network. Whereas indirect methods (which are the focus of this research) are based on the mathematical/software-based techniques. Software-based approaches can be divided into model-based approaches, signal-based approaches, and process history-based approaches. Model-based approaches are typically more straightforward, and use mathematical operations to detect fault from pressure measurements (Meseguer et al., 2015; Soldevila et al., 2016) whereas signal-based approaches and process history-based approaches rely on various types of data which depend on the method employed. These approaches then use mathematical and computational techniques used to extract the leak information from the measured data such as flowrate data, pressure data, etc. (Stevenson, 2020).

Two commonly used model based approaches include district metered area and water balancing, which can be categorized as steady-state approaches (Moser et al., 2015). District metered area (or “step testing”) is a conventional technique utilized by water authorities in detecting leaks in which the water network is subdivided into smaller sections and by systematically closing valves and the variation in water flow in each section can be analysed. Any inconsistent drop in water flow can be indicative of a leak in a water supply network (Boulos and Aboujaoude, 2011). The principal of water balancing is to audit the water supply network to force equality between the water input to the water distribution system and end uses (Moseret al., 2015). Whereas transient-state approaches generally refer to those in which pressure measurements from the water network are used to measure pressure fluctuations. These techniques such as inverse transient analysis, pressure residual vector method, etc. use measured transient signals to detect leaks within a water supply network (Makaya and Hensel, 2015).

Each leak detection approach has a certain degree of performance efficiency depending on the condition of the water distribution system and limitations of the approaches itself.

Recent studies have focused on addressing on leveraging the above techniques to locate potential leak-causing faults and thus improve the structural reliability of the water supply networks through preventative maintenance (Wong et al., 2018).

2.3.1 Hardware-based approaches

Due to the size and complexity of building water distribution systems it can be challenging to monitor their condition physically and can mean temporary interruptions to supply or require some level of on-site works. Hardware-based approaches usually employ non-intrusive or external devices including acoustic sensors, fiber optic sensors and ultrasonic flow meters which were some of the first types of methods developed for leak inspection in municipal water supply networks (Silva et al., 2011). External devices detect leaks within a water supply network by sensing specific properties such as acoustic signals, vibration signals, etc. from the pipeline outer surface based on different principles of operation (Cataldo et al., 2014). These devices can be permanently installed within a network or can be used temporarily when needed. Among various hardware-based approaches, the acoustic leak detection approach is the most used approach in the water sector at a municipal level (Perez et al., 2014; Wong et al., 2018). Acoustic leak detection devices capture changes in acoustic properties due to leaks in the pipeline which travel along the pipe wall or escape through the cracks. These devices generally perform reasonably well in detecting leaks in metal pipes but have limited capability of detecting leaks in plastic pipes (Silva et al., 2011). Although these approaches are considered the most accurate for leak detection, they are not suitable when it comes to large-scale non-residential building water distribution systems due to being relatively high cost, labour-intensive and time consuming. Furthermore, the use of hardware-based approaches can be time consuming and require skilled and experienced human resources to operate, detect and identify leaks in the water supply network (Puig et al., 2017; Cody and Narasimhan, 2020). Moreover, faults which give rise to small leaks or inefficient water use such as low-level continuous flows in complex non-residential water distribution systems can remain undetected by using conventional hardware-based approaches (Tidriri et al. 2016).

2.3.2 Software-based approaches

The presence of a leak within a water supply network can be assessed by employing steady-state approaches such as district metered area and water balancing which analyse the input flow data and end-user meter readings during low flow periods (typically night time) (Boulos and Aboujaoude, 2011). Software-based approaches have been suggested as a promising platform in this respect towards providing a near-complete solution for fault management of water distribution systems. A number of software-based approaches have been developed and studied at a municipal level for leak detection. However, these approaches are still not widely used in water sector as they are enormously data intensive and require sensors/meters to be installed in the water distribution system in order to compile adequate information from the water distribution system (Makaya and Hensel, 2015; Wu et al., 2016).

A case-study performed by (Ruzza et al., 2014) leveraged a data assimilation technique (Kalman filter) combined with a hydraulic model. A Kalman filter was used along with known end-user demand to calibrate a hydraulic model and to perform leak detection within a municipal water supply network. The hydraulic model combined with a Kalman filter was shown to be effective in monitoring and identifying leak locations and could be used for the calibration of the water supply network when considering pipe roughness and flow rate distribution. The operational performance of a municipal water supply network was evaluated by (Muranho et al., 2014) based on performance indicators (i.e., state variable slack, service pressure above required pressure and constraint violation) to identify problematic regions in a water supply network and assist in mapping out the burst pipes. These performance indicators were coupled with a pressure-driven hydraulic model. The integrated model showed usefulness in identifying model elements such as inadequately functioning pipelines and in detecting burst pipe events within the water supply network. Similarly, Günther et al., (2014) conducted a comparative analysis of leak detection approaches based on a laboratory scale water supply network. The aim of the analysis was to develop and evaluate new approaches to hydraulic modelling of water supply networks and to validate a proposed algorithm for leak detection. Leakage in a single pipe section was analysed by measuring pressure within the pipe. During experiments measurement and demand uncertainties were considered to improve leak detection and localization within a water supply network.

Model-based approaches are usually preferred when detailed information from the physical water distribution system is readily available to develop a hydraulic model but day to day operational data may be less available. These approaches have relied on hydraulic modelling techniques such as transient analysis, pressure residual vector method, etc. which have been validated using modelled or synthetic data. These approaches have focused on comparisons of measurements with predictions obtained from hydraulic models (Moors et al., 2018; Abdulshaheed et al., 2017; Alsaydalani, 2017; Soldevila et al., 2017; Perfido et al., 2016; Moser et al., 2015; Sedki & Ouazar, 2012; Mashford et al., 2012; Salam et al., 2015). Creating such models is often difficult and expensive due to the presence of non-stationarity and uncertainties in complex non-residential water distribution systems.

Where process history data for the water distribution system is accessible but available resources and/or the system complexities make it difficult to model, process history-based approaches are considered more appropriate to perform leak detection (Stevenson, 2020; Yipeng Wu & Shuming Liu, 2017). In the subsequent section, a brief overview of the fault detection and diagnosis (FDD) is presented. Following that software-based approaches (model-based, signal-based and process history-based) are reviewed (Section 2.5.2) in context of FDD in building water distribution systems.

2.4 Fault detection and diagnosis

FDD can be applied in various systems where faults can compromise system operation and/or efficiency (Frank et al., 2019). In engineering systems this can vary from small and simple to large-scale and complex systems and technological advancements have resulted in increasingly complex and interrelated systems. Therefore systems can suffer from wide variety of potential faults which can cause sub-optimal operation and result in performance deficiency, unnecessary downtime and whole-system malfunction (Zare, 2018). This increasing complexity of industrial systems such as non-residential water distribution systems and their related performance requirements. It is therefore critical to detect faults in their early stage and to diagnose their underlying cause, severity, and consequences (Lazarova et al., 2016). This need has driven the development of robust FDD methodologies to supervise systems. In this context, supervision involves performance monitoring tasks that aim to govern a system's operational state in real-time or almost real-time. Various FDD methodologies have been developed for various applications that depend on a system's nature and operational/maintenance requirements (Bakdi and Kouadri, 2017; Chambers, 2017).

In relation to municipal water supply networks FDD research is relatively well developed in the municipal water sector as comprehensively summarised in a recent review (Cody and Narasimhan, 2020). It is recognized that the non-residential building sector lags behind other industries where FDD is crucial such as nuclear power plants, manufacturing industries, etc. (Houngbo, 2019; UNESCO, 2015). Though a considerable amount of research has been carried out on energy use in buildings (and associated FDD methodologies), the implementation of FDD for non-residential water distribution systems is still sparse (Geng et al., 2019; Benndorf et al., 2018). Indeed, in a recent review aforementioned review (Cody, 2020), it is notable that there were no studies which focused on non-residential buildings. In the following discussion, FDD research in the water sector is discussed with a particular focus on statistical analysis and machine learning approaches. The quantitative performance evaluation of these approaches and their applicability to FDD in non-residential water distribution systems are studied.

2.4.1 Overview of fault detection and diagnosis approaches

FDD can be divided into two distinct but complementary steps (Tidriri et al., 2016).

- **Detection** - identify the presence of fault or non-routine water uses in the system. This step involves fault detection and isolation phases, while reducing the prevalence of false alarms.

- **Diagnosis** - determine the root cause of the detected fault. This step involves the classification phase which enable characterization of the fault type.

In FDD literature, a system is defined as faulty when at least one of its characteristic feature or parameters deviates from its routine operating condition (Nowicki et al., 2012). During routine operation, all features (measurable characteristics) and parameters of the system operate in a tolerable range. From a data processing perspective, FDD based on software-based approaches (similarly to leak detection approaches) can be classified into three distinct broader categories: (1) signal-based, (2) model-based, and (3) process history-based approaches (Dai and Gao, 2013). These approaches are summarised in **Figure 2.8** and **Table 2.1**.

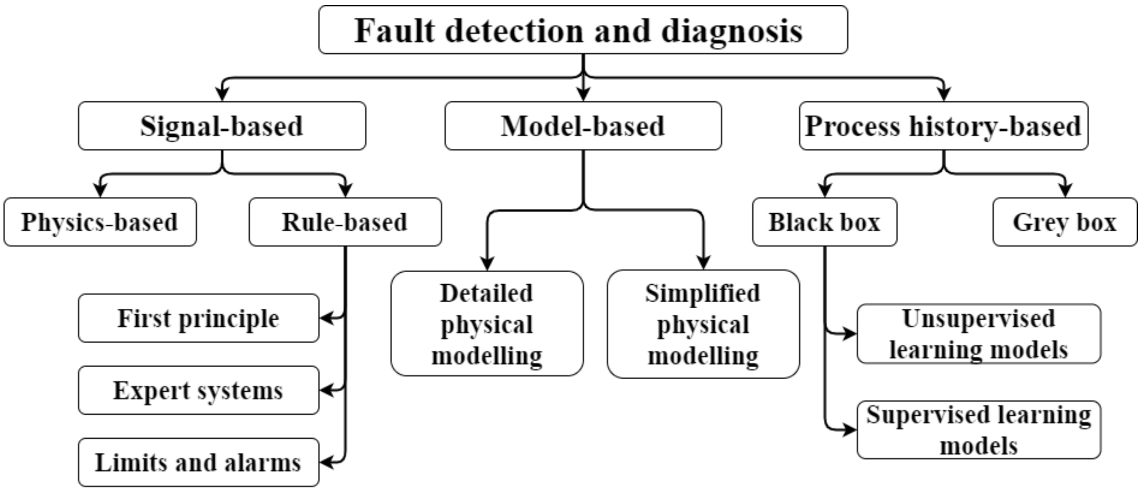


Figure 2.8: Outline of fault detection and diagnosis categories (Dai and Gao, 2013).

Table 2.1: Brief overview of fault detection and diagnosis categories (Chambers, 2017; Dai & Gao, 2013; Tidriri et al., 2016).

Model-based	• Requires a-priori known models. • Model-based approaches typically leverage mathematical functional correlations based on a fundamental understanding of the system. • These models use mathematical equations as the knowledge representation schemes to predict the output of the healthy system. • These models can be time intensive and computationally expensive to develop. • Algorithms within model-based approach continually monitor actual system behaviour with respect to the system model for inconsistency to detect and diagnose a fault.
Signal-based	• Signal-based approaches based on the analysis of the output signal and the knowledge of the system. • They do not involve an explicit input-output model.

	• In a faulty condition, generated residuals fluctuate from those of healthy system. • Typically, these approaches analyse signals either in time-domain or frequency-domain. However, some methods utilise both time and frequency domains.
Process history-based	• Require sizable amount of historic data and can learn from empirical data to extract the underlying knowledge that represents the information redundancy among the system's variable (correlation between the measurable characteristics - statistical features). • This approach is employed on systems which are too complex to have an implicit/explicit system model. • The performance of the system can be observed in real-time or almost real-time and compared with the knowledge obtained from the historical data to detect possible faults and conduct fault diagnosis. • This approach can be sub-divided into a qualitative method that rely on expert system which compare real-time data to a set of rules, which are typically derived from the knowledge of an expert human operator. Quantitative methods typically leverage pattern recognition for diagnosis and include statistical methods such as principal component analysis, support vector machine and neural networks.

2.4.2 Fault detection and diagnosis approaches in water distribution systems

In many sectors, FDD systems have been developed to optimise system operation and to detect and diagnose system abnormalities. Examples include heating ventilation and air conditioning (HVAC) systems, water and wastewater treatment processes, automotive industries, chemical and petrochemical processes, wind farms and some manufacturing facilities (Hu et al., 2019; Xu et al., 2017; Zhang et al., 2017; Naderi & Khorasani, 2016; Zhou et al., 2015; Yu et al., 2014; Bruton et al., 2014; Zhang et al., 2017). For example, the analysis of energy consumption in buildings as a key driver of FDD has been extensively studied (Hu et al., 2019; Gunay et al., 2019; Agostino et al., 2017; Balaras et al., 2017; Danacova et al., 2016; Stavset & Kauko, 2015). As previously mentioned, due to the lack of awareness and complexity of water distribution systems, the implementation of FDD has received more limited attention (Stetco et al., 2019; Geng et al., 2018). This is particularly the case for non-residential buildings (Adeyeye, 2014; Cosgrove et al., 2015) where it is acknowledged significant challenges remain in developing FDD systems for building water distribution systems (Liu et al., 2016; Nezhad et al., 2014). FDD studies in the water sector have focused mainly on leak detection at a municipal water supply network level (Seyoum et al., 2017). Against this backdrop FDD in water distribution systems is, relative to FDD in other engineering systems, an emerging topic in the water sector and certainly in relation to building water networks.

FDD generally depends on sensor input or derived measures of performance. In non-residential building water distribution systems "sensors" typically include process monitoring instrumentation for flow, water level in tanks, pipe pressure, temperature of the distribution/purification equipment and energy use of the equipment (Stanley, 2010).. However in non-residential building water distribution systems sensor failures are considered the most common equipment fault (Sharma et al., 2010). Thus, differentiating between faults associated with sensors and process faults is challenging in non-residential water distribution systems. The inherent system complexity and the increasing size of non-residential water distribution systems complicate the nature of faults, with faults occurring at different levels such as low-level leakage faults, excess usage faults, etc. and impacting the system with unique degree of severity (i.e., making it difficult to distinguish between routine and non-routine water uses, etc.) (Liu et al., 2015).

Basic fault detection systems comprise alarms which are triggered when relatively high levels of water is consumed (Mulligan et al., 2020; Clifford et al., 2018; Quevedo et al., 2014; Perfido et al., 2016) and provide high level statistics on water consumption. Typically, in such systems when water consumption exceeds a predefined threshold, an alarm is sent automatically (Clifford et al., 2018; Chambers et al., 2015; Perfido et al., 2016). While the alarm may not be designed to detect excessive water consumption due to improper use (e.g., a tap left running) or other non-routine water uses, it can be adapted to do so. However, such systems are unable to identify complex patterns of water usage such as distinguishing between actual faults or anomalous/inefficient water uses. Furthermore, basic fault detection systems are embedded with alarm notification features but still require assistance from experts with day-to-day operational insights to choose thresholds for faults (Mulligan et al., 2020; Perfido et al., 2016). More effective FDD approaches must be capable of detecting inevitable deviations in water time series (flow deviations due to inefficient water uses). As mentioned earlier, the prediction of faults for non-residential building water distribution systems is made more difficult by the non-stationarity of water usage (e.g., water consumption can vary between seasons and depend on working hours, holiday periods, etc.) (Prabuchandran et al., 2019). Furthermore, FDD methods must be cognisant with the real data related issues such as outlier measurements and data with incomplete/missing information that cause difficulty in identifying complex pattern of non-routine water uses or faults in non-residential building water distribution systems.

In terms of software-based approaches model-based approaches are not only capable of detecting and isolating faults in real time but can also detect faults associated with lower flow readings (leakage). Model-based approaches which primarily depend on a process model and the analytical relationships among the measured variables (statistical features), can be used to obtain information about probable anomalies that may cause faults in a system (Ahmad et al., 2018; Tardioli et al., 2015). Model-based approaches typically consist of three fundamental phases - (i) model building, (ii) model calibration and validation and (iii) leak detection and identification (Stevenson, 2020). Model-based approaches rely heavily on a mathematical model for simulating the real water distribution system and their performance depends on how precise the model is. They require an explicit input-output model of the real water distribution system (a complete physical representation of water distribution system) which can make it difficult to obtain complete and accurate mathematical process model in case of a large-scale complex infrastructure such as non-residential building water distribution systems.

In contrast, signal-based and process history-based approaches do not require an explicit input-output model of the system. Signal-based methods mainly focus on analysis of output signals from the system with less attention to the input dynamics of the system. For instance, in non-residential building water distribution systems, signal-based FDD may not be effective and may degrade FDD performance in the events of non-routine water uses resulting from inefficient water use or infrequent sudden variation in water demand (e.g., varying production demand in manufacturing facility for a short period of time or a conference event in institutional building, etc.) (Hashemi et al., 2016). As it would not be able to distinguish between actual faults and inefficient water use and incorrectly report inefficient water use as fault. That makes signal-based approaches less effective in distinguishing between non-routine water uses and faults in non-residential building water distribution systems.

Process history-based approaches are relatively recent in terms of their application to process monitoring and control, have received significant attention in recent years (Xu et al., 2017; Gharsellaoui et al., 2020; Thomas et al., 2018; Zhang et al., 2016; Ammiche and Kouadri, 2017). Process history-based approaches employ a 'learn by example' approach and leverage process history data (Mulligan et al., 2018). In the context of building water networks, they would not require comprehensive information regarding non-routine water uses and faulty events of the water distribution system. Process history-based approaches are often enabled by statistical analysis-based models (unsupervised learning models) and machine learning-based models (supervised learning models) which acquire knowledge from empirical data to determine routine operating conditions and subsequently monitor system anomalies.

These approaches can overcome the technical complexities associated with the model-based approaches and signal-based approaches. In process history-based approaches historical process data is needed to extract and map correlations between faulty features and operational modes of the system (e.g., water usage baseline and diurnal patterns in building water distribution systems) (Venkatasubramanian et al., 2003; Ahmad et al., 2018). In contrast, model-based approaches are based on physical a-priori information of the system (first principles and expert experience) and may provide less fault detection accuracy and lack in modelling inherent complexities of the system (Zhao et al., 2019). Process history-based models can always have advantages such as relative ease of implementation and higher FDD accuracy than the model-based and signal-based approaches (Bakdi and Kouadri, 2017). Moreover, process history-based models are always sensitive to minor deviations in patterns of monitored data and are more capable of distinguishing between non-routine water uses and faulty events when compared to model-based and signal-based approaches (Garcia-Alvarez, 2014). Process history-based approaches typically comprise the following phases (**Figure 2.9**).

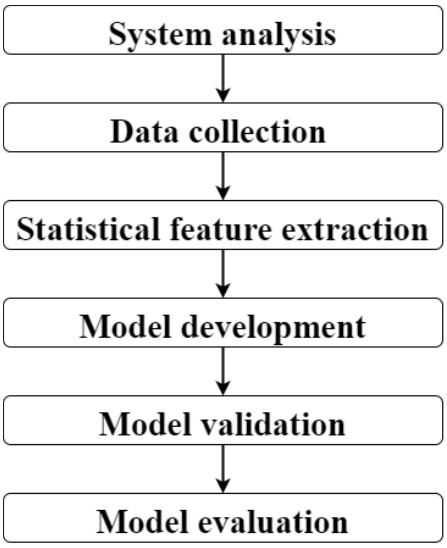


Figure 2.9: Process history-based modelling phases.

The structural and operational complexity of non-residential building water distribution systems do not impact process history-based approaches applicability. In this era of exploratory data analysis, rational as well as promising conclusions can be drawn from the process history data. In this regard, several process history-based approaches rely on statistical analysis such as principal component analysis (PCA), principal component regression (PCR), partial least squares (PLS), canonical correlation analysis (CCA), canonical variate and subspace state space models have been proposed.

2.4.3 Supervised learning models

Supervised learning models involve leveraging known information to learn existing patterns or distribution in the data under the supervision of practitioners or professionals. In this category, models are trained to predict the output based on the available information (Mao et al., 2019). Supervised learning models (machine learning-based models) include support vector machines (SVM), artificial neural network (ANN), partial least square (PLS) and auto regression models. In recent times, the use of SVM learning has increased relative to other methods such as ANN when modelling the relationship between the input and the output in classification and regression problems in various engineering sectors (Xu et al., 2020; Kouziokas, 2020; Gautam et al., 2020; Cody, 2020;

Akil et al., 2019; Grueiro et al., 2018; Liu et al., 2019; Samer et al., 2018). This is due to SVM's strong fault diagnosis ability in multi-class classification problems, high training efficiency, and its effectiveness when working with low-dimensional (small sample dataset) process data (Xu et al., 2017). In contrast, ANN represents a more challenging pattern learning problem with random sampling of datapoints arising when working with smaller datasets (Kouziokas, 2020; Gautam et al., 2020). Supervised learning models such as SVM usually require labelled data or involves manual labelling steps (results in higher labelling cost in terms of time resources) to develop a FDD model. In case of non-residential water distribution systems, it is difficult to anticipate a-priori information about all the possible conditions in which the fault can occur. Examples of the application of supervised approaches in water distribution systems are discussed below. It should be noted that previous applications are limited to municipal water supply networks with none, to the best knowledge of the authors, focusing on non-residential buildings.

Mashford et al., (2009) presented a supervised learning-based methodology for leak detection within a modelled water supply network using pressure sensor data. This study examined some of the key drawbacks of hardware-based approaches - in this case incorrect detection of multiple leaks and undetection of leak with larger size using acoustic leak detection techniques. SVM was used as a supervised learning approach. The SVM was trained by using pressure values at a number of nodes within a water supply network. Several leaks were virtually generated through hydraulic modelling by assigning distinct emitter coefficients (coefficients related to the magnitude of a virtual leak) at multiple nodes to train the SVM. The trained SVM was then tested on the data obtained from the hydraulic model with different virtual leaks and showed accuracy in predicting leak size within the water supply network. A similar study based on a SVM combined with the optical fibre interferometer sensing approach (hardware-based approach) was presented by Qu et al., (2010) to detect underground pipeline leaks. Vibration signals from a real 35 km municipal water supply pipeline were used to train a SVM and test it against vibration data with known leaks. The SVM performed well in detecting pipeline leaks with an accuracy of 95% and identifying the location of leaks to an accuracy of ±200 m. Nasir et al., (2014) proposed an approach for leak detection and identification, and for estimating the magnitude of leaks in municipal water supply networks. In this study a hydraulic model was developed, and simulated data was obtained using pressure sensors installed at various intervals along a pipe network and with different leak magnitudes. The modelled data was then used to train two supervised learning approaches (SVM and ANN). A comparison of both supervised learning approaches was developed with the results showing that the SVM functioned well but was found to be less sensitive than ANN in terms of performance and robustness in analysing both routine and faulty conditions. Similarly, Salam et al., (2015) carried out a comparative analysis of supervised learning approaches, in which SVM and extreme learning machine (ELM) methods were used to perform leak detection and predict the magnitude of leaks in a municipal water supply network. In this case the input-output data to the SVM and ELM methods were obtained through hydraulic modelling (i.e., real data was not used). The prediction accuracy for the leak magnitude and location of each method was analysed by the value of normalized root mean square error (NRMSE), the lower the NRMSE value, the higher the prediction accuracy. The results demonstrated improved prediction accuracy of SVM over ELM supervised learning approaches. Comparative analyses by Grueiro et al., (2018) of different machine learning approaches such as SVM, ANN, nearest neighbours, etc. was conducted to detect leaks in a modelled municipal water supply network. Training and testing data were generated via a hydraulic model by introducing leaks virtually (i.e., no real data was used). The study showed the robustness and improved performance of SVM when compared to other approaches in identifying municipal water supply network leaks in the presence of different uncertain conditions such as non-stationarity of water use, sensor noise, etc. A similar study by Akil et al., (2019) in which an analysis-decision making methodology was developed for an institutional building by monitoring statistical features derived from meter readings. Water, electricity, and gas consumption data for one year was used to train SVM developed to detect anomalies in water, electricity, and gas consumption. SVM showed good performance in identifying atypical behaviour such as increased electricity or increased water uses, etc. within the building, though the study did not analyse or develop FDD strategies.

2.4.4 Unsupervised learning models

Unsupervised learning models attempt to extract the underlying measurable characteristics (statistical features) or distribution in the data without any prior knowledge of the existing pattern in the data. These can play a vital role in discriminating between routine, non-routine water uses, and faults in a system. Unsupervised learning models (statistical analysis approaches) include principal component analysis (PCA), hierarchical clustering techniques (Dendrograms), non-hierarchical clustering techniques (K-means), etc. Unsupervised learning models have certain advantages in that they do not require a-prior information of the system being analysed (Mao et al., 2019).

Unsupervised learning models such as PCA are a versatile and well-known technique for data analysis (Izem et al., 2018; Wei Li et al., 2018). It has been adopted in various disciplines such as in damage detection of civil engineering structures, data compression, face recognition, image analysis, visualization as well as in fault detection in petroleum/chemical industries (Laory et al., 2011; Saitta et al., 2005; Posenato et al., 2008; Harrou et al., 2013; Abdi & Williams, 2010; Li et al., 2000). PCA is a vector space transformation predominately used to reduce data dimensionality by extracting correlation between variables in sets of independent variables that explain the trend of the process while optimising the variance of the original data in a reduced number of dimensions (Garcia-Alvarez, 2014; Laory et al., 2011; Villegas et al., 2010). A PCA model leverages process history data and extracts a linear combination of variables explaining the major trend of the process being analysed (Izem et al., 2018). Some examples of the PCA application in engineering systems given below provide insights into the wider applicability of PCA in diverse multivariate processes.

Damage assessment of a piezoelectric sensor by means of the application of PCA and detection indices (T^2 and Q-statistics) was studied by (Mujica et al., 2011), to detect and isolate structural damage using vibration signals. Known vibration signals were applied and their corresponding dynamical responses were analysed. A number of different damage conditions were experimentally analysed with an increased difficulty level (from low frequency to high frequency). A PCA model was built utilizing undamaged data as a reference base line. The results demonstrated the feasibility and potential of using PCA with detection indices (T^2 and Q-statistics) in detecting and isolating piezoelectric sensor faults. Another study regarding FDD was performed by (Villegas et al., 2010) on two real plant water tanks based on PCA that works at different operating points. A PCA model was built considering normal operation data and tested under different faulty conditions such as tank outlet clogging, tank level sensor fault. The PCA model performed well in detecting variations across the operating point assuring PCA scheme robustness for system's non-linearities. Ahmed et al., (2012) studied time domain vibration features for detecting and diagnosing different fault types from a multistage reciprocating compressor. Several diagnostic features were extracted and a FDD method based on PCA was developed to detect various leakage faults in the system. The PCA model showed good agreement in detecting single and multiple faults in a reciprocating compressor.

In a statistical FDD framework, the result statistics plays a central role in identifying potential variations in the process. Zhang et al., (2017) performed a study on detecting multiplicative and additive faults and was evaluated in terms of fault detection rate. A fault detection method based on PCA, and detection indices (T^2 and Q-statistics) was developed to model process data and to detect faults (multiplicative and additive faults) in the system. Multiplicative faults are usually parametric faults resulted from malfunctioning of the internal factors such as excess usage or leakage within a network. Whereas additive faults are normally resultant of sensor faults or actuator related faults. The results demonstrated that T^2-statistics performs better than Q-statistics in detecting multiplicative and additive faults as it contains explicit distribution information. While Q-statistics contained the approximate probability distribution. Similarly, a brief survey of PCA was conducted by (Zhang et al., 2016) for detecting multiplicative faults in multivariate statistical process monitoring. PCA detection indices (T^2 and Q-statistics) and their extensions based on Wishart distribution and mutual information Kullback-Leibler divergence were studied using numerical simulations to describe the measurement variance and covariance characteristics. Among different extensions studied, conventional PCA showed better performance in detecting multiplicative faults by reducing the measurement variance.

In one of the first studies on the application of PCA to water distribution data, Gertler et al., (2010) leveraged PCA to develop a leak detection and isolation approach for water supply networks using pressure measurements. Fault-free system data was obtained through hydraulic modelling of a municipal water supply network. The developed model was tested by introducing leaks of known magnitude and location into a modelled water supply network. The PCA-based approach was found to be useful in detecting and isolating leaks and represent an alternative to other approaches such as district metered area approaches. A similar study by (Villegas et al., 2010) presented the application of the PCA for fault detection and isolation in a laboratory-scale water plant. In total, five PCA models were developed, one encompassing routine operating condition and four with different faulty conditions (i.e., two faults related to each tank). Level sensor and outlet clogging faults were considered in this study. Fault detection and isolation was carried out by evaluating the detection indices results, for each faulty condition against the T^2 and Q α-control limits calculated at a routine operating condition. Correct detection and isolation of different defined faults were obtained by using the PCA approach. Arregui and Carlos, (2012) proposed a methodology to prevent burst pipes by controlling inflow for a real municipal water supply network. The water supply network was divided into six independent regions using a district metering area technique. PCA with detection indices (T^2-statistics) was used in this study. The T^2 α-control limit was developed based on the real flow rate data. Evaluation of the developed PCA model was done by employing artificial fault scenarios (i.e., burst pipes or abnormal water flow). The PCA model showed effectiveness in detecting abnormal flow behaviour within 5% of the average flow with a probability of 30-95% depending on the hour of occurrence. Another study by (Quiñones et al., 2017) in which a demand model was developed for fault detection and isolation in a modelled municipal water supply network. PCA, independent component analysis (ICA), and support vector data description (SVDD) methods were used to develop separate fault detection and isolation models. Leaks of varying magnitude were simulated across each node of the hydraulic model to analyse the sensitivity of the fault detection and isolation model. The model based on components analysis (PCA and ICA) showed overall better performance than the SVDD based model in detecting leaks.

At present, there are a small number of unsupervised learning model-based fault detection systems for water distribution systems that are commercially available, such as Outpost[2] and Greensense View[3]. These provide high level statistics on water uses and are designed to detect excessive water usage. Univariate systems are embedded with alarm notification features which trigger when water usage exceeds a predefined threshold, however they are unable to detect low-level imperceptible (low flow reading) faults such as tap left running, etc. and require assistance from experts to choose thresholds for faults.

2.4.5 Summary of learning models

It is evident from the published literature that there has been significant reseach into developing techniques to preserve the integrity of large-scale municipal water supply

[2] http://www.outpostcentral.com/.
[3] http://www.greensense.com.au/.

networks. Various forms of performance monitoring technologies have been developed, mostly using modelled data, to mitigate the effects of faults mainly leakage.

Table 2.2 provides a comprehensive summary of relevant published studies on the various different approaches developed (i.e., hardware-based, software-based, supervised learning model-based and unsupervised learning model-based) for performance monitoring of water distribution systems (leak detection mainly at a municipal level).

Previous literature has generally focused on stationary systems (with stationary water use pattern) and often utilized modelled data (i.e., testing and validation has usually been performed using hydraulic models or laboratory-scale water supply networks). In some cases, studies at municipal level have suffered from lack of data. Limited research has been done related to the area of performance monitoring of non-residential building water distribution systems (Stevenson, 2020; Xing and Sela, 2019; Seyoum et al., 2017).

Table 2.2: Literature review.

References	Objective	Key features and outcomes of study
	Hardware-based studies	
Wong et al., 2018	Monitoring of underground **municipal water supply network** subjected to hydraulic pressure transient excitation	Optical fibre sensors were used to detect the structural damage and the presence of leakage. Findings showed promising potential for the pipeline health monitoring using transient excitation.
Li et al., 2018	Study on leak detection of a **municipal water supply network** subjected to failure of socket joint based on acoustic emission and pattern recognition	Acoustic features of leak signals in socket and spigot pipe segments were analysed. The extracted features were used to establish an artificial neural network-based classifier for leak detection.
Gao et al., 2017	Monitoring of buried plastic **municipal water supply network** using acoustic signals	Leak noise signals were analysed using a cross-correlation function. Manual interpretation of leak signals, potentially restrict the practical application of acoustic leak noise correlation function.
Cataldo et al., 2014	Analysis of an underground **water pipe leak detection** methodology utilizing time domain reflectometry, ground penetrating radar and electrical resistivity tomography techniques	Comparative analysis of alternative techniques was performed for their application in leak detection. Time domain reflectometry technique allowed detection of the position of the leak with a very low error.
Mooney and Johnson, 2014	Review of robotic devices for **municipal water supply network** in-pipe inspection	Reviewed in-pipe robotic devices suitable for performance monitoring of pipes within water supply network.
Liu et al. 2013	Review of pipe inspection technologies for condition monitoring of underground **municipal water supply network**	Smart pipes and intelligent robotic methods are discussed. Reviewed suitable performance monitoring methods for different pipe materials.
	Software-based studies	
Moser et al., 2015	Performance comparison of modelled **water supply network** reduction strategies for leak detection	A model-based data interpretation methodology was developed by using five different reduction categories to detect leakage in a water supply network. Up to 20% computational time can be reduced by using modelled water supply network reduction strategies

Colombo et al., 2009	Review of leak detection techniques based on transient and frequency domain methods for a **municipal water supply network**	All major transient-based and frequency domain-based techniques were reviewed. Related mathematical schemes for analysis were also discussed.
Makaya and Hensel, 2015	**Water loss management** strategies for developing countries	Integrated water leakage management framework using artificial neural network and district metering area. Artificial water demand pattern was used to simulate leakage in a water distribution network. Artificial neural network can be used to forecast flow with up to 99% confidence. The study concluded that 33% of water loss through leakage.
Yamijala et al., 2009	Analysis of pipe break data of a real **municipal water supply network**	Comparative analysis of statistical regression models to estimate the reliability of pipes in water supply network. The goal was to predict the likelihood of pipe breaks and linked parameters for future. Statistical methods are not exclusively studied.
Sanz et al., 2016	Development of leak detection and identification through demand components calibration for a real **municipal water supply network**	Pearson correlation method was used to predict the probability of leak by analysing pressure values. Performance of the proposed methodology was validated against real water supply network using synthetic data. Leak through burst pipe only considered in the methodology.
Perfido et al., 2016	Development of leak detection system for the improvement of a real **non-residential water supply network** management	Real time water flow and pressure data were used to infer performance and operational faults in a water supply network. Two simulated leak scenarios were used to validate the model performance. The model showed good detection rates, low false positives, and good leak detection accuracies.
Anjana et al., 2015	Study of particle filter-based leak detection technique for real **municipal water supply network**	Several anomalies within a modelled pipe network were simulated. Proposed method used for user-end flow estimation in the water supply network. Anomaly detection studied for application on burst pipes only.

Sun et al., 2015	Development of leak detection approach for drinking real **municipal water supply network** using pressure residuals and classifiers	Proposed integrated fault detection approach based on the pressure sensors and classifiers. Hydraulic model of the water supply network was used to generate data for validation against sensor values.
Meseguer et al., 2015	Study of model-based monitoring techniques for leak detection in **real municipal water supply network**	Hydraulic simulation was performed at a district metered area level. Simulated leak situations were used to assess leak detection model performance. Pressure sensor value from real water supply network were used to validate the leak detection model.
Ferrandez-Gamot et al., 2015	Study of leak detection in **real municipal water supply network** using pressure residuals and classifiers	Leak detection approach based on the hydraulic model, pressure sensors and classifier were proposed. Modelled water supply network was used to generate residuals (difference between the real measurements and their estimation using hydraulic model). Generated residuals were validated against real water supply network using classifiers to detect leak.
Perez et al., 2014	Development of leak detection and isolation in **real municipal drinking water supply network**	Model-based leakage detection and isolation methodology was proposed using real pressure measurements. Hydraulic simulation was performed at a district metered area level to characterize fault situations in a water supply network.
Ruzza et al., 2014	Study of leak detection in a **real municipal water supply network**	Data assimilation (Kalman filter) technique used for the calibration of the hydraulic model. Artificially created leaks was used and compared with water audit analysis.
Muranho et al., 2014	Technical performance evaluation of **municipal water supply network** based on hydraulic modelling	The operational performance of a water supply network was evaluated based on technical performance indexes to identify problematic regions in the water supply network. Pressure driven approaches were used to calibrate the model.
Günther et al., 2014	Analysis of leakage outflow in **municipal water supply network**	A comparative analysis of leak detection based on a simple laboratory water supply network was performed to validate hydraulic modelling approaches and the performance accuracy of leak detection algorithm (differential evolution).

Sophocleous et al., 2017	Study of two-stage calibration for detection of leakage hotspots in a real **municipal water supply network**	A Genetic Algorithm was used to calibrate a hydraulic model and detect leak hotspots in a water supply network. Leaks were modelled using a hydraulic modelling tool.
Ercole et al., 2016	Development of an integrated modelling approach to optimize the management of a **municipal water distribution system**	To optimize water distribution system management pressure driven approaches were used for hydraulic simulation. Pumping energy analysis was performed based on pipe failures such as burst pipes, leakage, etc.
Almandoz et al., 2005	Analysis of leakage through virtual **municipal water supply network**	Water balancing approach used to identify the non-revenue consumption in water supply network. Analytical approach used to estimate the leakage and demand within a network.
Colombo and Karney, 2005	Study of leakage impacts on energy consumption in **municipal water supply network**	Leakage impacts on pumping energy costs was analysed. Hydraulic simulations were performed to determine different parameters (i.e., system pressures, storage tank levels, energy cost, power consumption and leakage volumes) within a water supply network.
Colombo and Karney, 2003	Study of pipe breaks and the role of leaks in **municipal water supply network** from an economic perspective	Analytical approaches presented to sizing and locating leaks in a pipe in relation to the energy consumption and hydraulic transients. Pressure management and life cycle analysis used to improve the leak detection and system efficiency.
Seyoum et al., 2017	Development of Shazam-like household **water leakage detection** method	Acoustic sound signals produced by water activities were analysed to detect leakages within a household water supply network.
Unsupervised learning-based studies		
Horrigan et al., 2018	Development of statistical fault detection scheme utilizing **commercial building energy performance** data	Development of proposed statistical-based performance prediction model utilizing univariate data time series. Actual fault scenarios were not studied to validate the fault detection method. Fault detection methodology developed based on PCA and detection indices (T^2 and Q-statistics).
Li et al., 2018	Development of sensor fault detection in a **nuclear power plant** using PCA method	
Zhang et al., 2017	Assessment of T^2 and Q-statistics for detecting additive and multiplicative faults in multivariate statistical **process monitoring**	

Ahmed et al., 2012	Development of fault detection and diagnosis approach based on PCA using **vibration data from reciprocating compressor**	The efficiency of PCA in detecting additive and multiplicative faults was studied in the context of fault detection rates. It was found T^2-statistics performs better than Q-statistics in detecting additive and multiplicative faults.
Arregui and Carlos, 2012	Development of burst detection for **municipal water supply network** using PCA	Several diagnostic features were extracted and analysed by PCA to detect leakage faults in the system.
Gertler et al., 2010	Development of PCA-based approach for the detection and isolation of leaks in **municipal water supply network**	The PCA model showed good accuracy in providing detected fault information.
Villegas et al., 2010	Study of PCA for fault detection and diagnosis, experience with a **pilot plant (two-tanks system**)	PCA model of the fault-free system is obtained from simulated data and tested against simulated faults. PCA methodology developed to identify changes across operating point, assuring model robustness for system's non-linearities.
Eliades and Polycarpou, 2012	Leakage fault detection in district metered areas of **water supply network**	Periodic water usage and leakage dynamics were estimated using Fourier series. Simulated leak scenarios were used and compared with the real data measured in district metered areas. Proposed methodology demonstrated ability to detect leak with smaller magnitude.
Jung et al., 2015	Study of burst detection based on statistical process control methods for a real **municipal water supply network**	Comparative analysis of six univariate and multivariate statistical process control methods. Synthetic flow and pressure data were used for the model calibration. Results indicate that system pressure are more sensitive to disturbances and provide higher detectability than flow meters.
Wu et al., 2016	Development of burst pipe detection methodology for a real **municipal water supply network**	Water supply network was divided into regions by using district metered area technique. Clustering algorithm was used to detect burst pipe and to cope with non-stationarity of water usage data. Artificial burst conditions (flushing with short duration) were implemented to evaluate the effectiveness of proposed approach.

Grueiro et al., 2018	Development of PCA-based leak detection and identification approach for a real **municipal water supply network**	The study used measured data from a district metering area under normal operating conditions. Results showed good accuracy of detecting leak with a magnitude of 2.5% of the total water demand with an average detection rate of 72%.
	Supervised learning-based studies	
Escofet et al., 2016	Study of model vs. process history-based fault diagnosis approaches for a real **municipal water supply network**	SVM based methodology developed for detecting leakage and to diagnose anomalies in a system. Use of simulated data rather than measured data. Support vector machines shows superiority over artificial neural network in terms of performance, robustness in both normal and noisy conditions. SVM exhibit insensitiveness in detecting leaks in a noisy environment as compared to artificial neural network. Measurement/modelling uncertainties were not studied. Signal decomposition techniques along with support vector machines to extract features for classification. Classification accuracies of discrete wavelet at different levels were calculated and compared.
Silva et al., 2011	Development of computer aided leak location and sizing approach for real **municipal water supply network** using SVM and neural networks	
Muralidharan et al., 2014	Analysis of SVM-based fault diagnosis for **monoblock centrifugal pump**	
Nasir et al., 2014	Analysis of measurement error sensitivity for detecting and locating leaks in modelled **municipal water supply network** using ANN and SVM	
Saberi et al., 2011	Study on performance comparison and robustness of SVM and ANN for fault diagnosis in a **centrifugal pump**	
Qu et al., 2010	Analysis of SVM-based underground **pipeline** leakage detection and pre-warning system	
Mashford et al., 2009	Development of leak detection approach based on support vector machine for **municipal water supply network** using monitored pressure values by	
Patabendige et al., 2018	Development of detection and interpretation of anomalous **water use** for non-residential customers	Nearest neighbour machine learning approach was used to detect anomaly in water usage from smart water readings. Validated against non-residential site water use data.
Soldevila et al., 2016	Development of leak detection and identification method for real **municipal water supply network** using a mixed model-based/process history-based approach	Single leak scenario was considered under different leak size and complexity.

Cody, 2020; Grueiro et al., 2018	Study on performance comparison and robustness of different classifiers for detecting leaks in **water supply network**	SVM, ANN, nearest neighbours, and Bayes classifier were used. Use of simulated data rather than measured data. SVM shows superiority over other classifiers in terms of performance, robustness in in the presence of different uncertain conditions such as non-stationarity of water use, sensor noise, etc.
Samer et al., 2018	Development of SVM-based leak detection model using accelerometers for pressurized **municipal water supply network**	Use of experimental data at laboratory scale setup. Experiments were conducted using one inch and two-inch pipelines of material ductile iron and polyvinyl chloride (PVC). SVM performed with an accuracy of 98% in identifying leaks.
Mounce and Machell, 2006	Analysis of ANN-based burst detection using hydraulic data for real **municipal water distribution systems**	Two ANNs (static and time delay) were developed to learn patterns of leak/burst. Both networks tested against data recorded from a real water distribution system were effective in classifying leak/burst events. The method requires sufficient leak/burst exemplars (training data points) to train the network.
Akil et al., 2019	Monitoring of SVM-based of an institutional building using gas, electricity, and water use data	Use of measured water, electricity, and gas usage data. SVM shows effectiveness in identifying atypical behaviour in an institutional building such as water leak, increased electricity use, etc.
Gautam et al., 2020	Study of SVM-based monitoring and forecasting for leak detection in household **water storage**	Use of measured water usage data. Comparison of actual value and predicted value to identify leaks.

Table 2.2 summarises the efficacy and applicability of the unsupervised learning model (PCA) and the supervised learning model (SVM) in various engineering domains (municipal water supply networks, nuclear power plants, rotary machines, etc.). A key challenge with the use of PCA is the presence of noise in the data. Principal components are often attracted towards outlying data points and may not be able to capture the realistic behaviour of routine variation in a non-residential water distribution system, resulting in unreliable fault detection. This can be improved by using robust approaches that can deal with biased uncertainties and non-stationarity of water use which are typical in non-residential building water distribution systems. The major drawback of using SVM is that the trained classifier will only be able to deal with fault alarms that have been experienced by the non-residential building water distribution system previously. In practical applications, the available data can be limited to a subset of the potential fault alarms that the system has experienced. However, this could be improved by utilizing knowledge of experienced faults or anomalous situations to update the FDD model and provide a more reliable performance monitoring system. As the number of faults or anomalous situations in the FDD model database increases, the system becomes more valuable and reliable.

Previous studies have tended to describe one FDD method at a time (either PCA or SVM) often formulated as an optimization task. For instance, PCA has been used to compress data or to identify the anomalies using detection indices (T^2 or Q-statistics), whereas the goal of using SVM was to minimise the difference between the measurements taken on the system and predicted values from physics-based models (Grueiro et al., 2018; Escofet et al., 2016; Nasir et al., 2014). A significant challenge impacting the performance of these approaches is the time varying characteristics of real time processes such as change in mean, variance over time, process noise, etc. which can be a particular characteristic of water consumption in non-residential buildings.

It can also be noted from **Table 2.2** that a significant gap exists in published literature relating to FDD methods that utilize real case-study data. Existing studies in the literature tend to focus on the use of generated or experimental data sets (Cody & Narasimhan, 2020; Grueiro et al., 2018; Moser et al., 2015; Villegas et al., 2010). The use of idealised data means the effect of random variations and outliers which undoubtedly occur in industry are not taken into account in the assessment of proposed methodologies. This can lead to an over-estimation of the effectiveness of proposed approaches if applied to real case-studies.

The authors did not find any studies which applied PCA and SVM methods to building water networks (or indeed industry facilities). The use of more advanced statistical techniques such as PCA in conjunction with SVM has not been widely explored in water sector. Given that building water distribution systems can often exhibit non-stationarity (change in statistical properties over time), the combination of PCA and SVM techniques has potential to offer robust applications in water distribution system in non-residential buildings.

The findings underpin the objectives for this PhD and provide a focus for the remainder of the literature review. It is evident that if water resource management is to be improved within the built environment, then the performance monitoring methodologies which can detect and diagnose faults in building water distribution systems need to be developed. The following section reviews the area of statistical process control techniques and how it can be applied to building water distribution systems to design FDD methodology which will counteract the deficiencies outlined hitherto.

2.5 Statistical process control

2.5.1 Overview

Application of the statistical methods to control and monitor process is commonly known as statistical process control (Yan, 2015). Statistical process control comprises the application of tools to increase the quality of a process by reducing its variability, thus making each process conform to a certain standard (Dong and Qin, 2018). Engineering process control problems such as large-scale non-residential water distribution systems are multivariate in nature since water distribution systems involve several statistical features (measurable characteristics). These statistical features may have collinearity that can be highly correlated. In practice, process data can often be recorded irregularly, may contain missing data points such as missing water use values caused by metering error or comprise some data that is corrupted by process and measurement noise such as equipment failure, operational error, etc. (Zhu et al., 2018).

Conventional statistical process control methods such as Shewhart chart, cumulative sum and exponentially weighted moving average have been widely used for monitoring processes based on univariate approaches that monitor and analyse a single variable at a time such as water use at a particular point in a water supply network (Jung et al., 2015). These conventional statistical process control methods do not function well when introduced to multivariable processes such as non-residential water distribution systems (Dhini and Surjandari, 2016). Other challenges associated with the application of classical statistical process control approaches to building water distribution systems include the considerable variation in water consumption levels in buildings, and the possible mismatches between the proposed statistical model and the process being monitored (Chen, 2010). In such cases multivariate statistical process control (MSPC) provides an alternative approach to conventional statistical process control and can overcome the limitation of univariate statistical process control methods (Yan, 2015). Most univariate statistical process control shortcomings can be addressed through the application of MSPC techniques which consider multiple variables of interest simultaneously and can extract behavioural information for each variable or characteristic relative to the others. MSPC research can thus add significant value in theoretical as well as in practical applications and is conducive to process FDD (Ragab et al., 2018; Damarla, 2011).

2.5.2 Challenges to multivariate statistical process control to non-residential water distribution systems

MSPC was designed to monitor process variability and can offer several advantages to building water networks. A key advantage in the use of MSPC is its potential application in distinguishing between routine, non-routine water uses and fault events. Routine water uses are the events which are the deviations from the nominal pattern that usually present and are natural for the process to have. Non-routine water uses are the events which are not the results of faults but resulted from sudden and sporadic changes in water use due to specific events occurring in the building. On the other hand, fault events are the variations caused by some errors or malfunctioning in the building water distribution system which usually represents either equipment malfunctioning, metering error or operational errors. Though, MSPC poses some challenges in detecting critical faults at their early stage crucial to operational condition of the process such as low-level imperceptible faults. These challenges need to be addressed in developing FDD methodologies for non-residential water distribution systems.

- Most MSPC methods were designed based on the assumption that data follows the normal distribution. The conditions in which non-residential water distribution systems operates are of varying nature such as diurnal, weekly, seasonal patterns, etc. often representing non-stationarity in water use and biased uncertainties (measurement errors, incomplete information) which tend to have a dominant effect on the process performance. These factors can cause complications and impede the ability of FDD model to distinguish between routine, non-routine water uses, and fault events. This can also drive false alarms (discussed in Section 2.6.3.) during FDD process due to the violation of model assumption (data normality) (Zhao et al., 2019). Therefore, these factors need to be considered as routine process variation rather than process disturbances to robustly detect and diagnose faults in non-residential water distribution systems.

- Non-residential building water distribution systems may contain several measurable characteristics (statistical features) which are highly inter-dependant. That could produce collinearity in the dataset and result in modelling and interpretation errors. Thus, it is crucial to capture the main features that drives the building water distribution system operation.

- In the presence of outliers within the data, MSPC methods may not be able to capture the realistic behaviour of routine variation in building water distribution systems. This can result in unreliable or false fault detection during the process. It is imperative that the FDD methodology is capable of considering process disturbances such as non-routine water uses which are typical in building water distribution system modelling challenges and is in practice seen as state of things.

- The implementation of MSPC-based performance monitoring is complex and time-consuming. Several supervised learning models such as SVM for FDD have been proposed and have shown promising results in municipal water supply networks. The challenges linked to MSPC in predicting the hidden risk associated with faults can be overcome by integrating these supervised learning models. However, there are no studies yet that stated a specific supervised learning model is suited or designed for any particular system (e.g., non-residential water distribution systems) (Dhini and Surjandari, 2016).

- The derivation of control limits is more challenging and time-consuming when the data comprises a larger number of dimensions (increased number of statistical features) which increases the difficulty of theoretically estimating the distribution of data. The time required for deriving control limits could increase proportionally as the number of statistical features increases (dimensionality of data). Therefore, extraction of statistical features that could provide complete and useful information regarding a building water distribution system is critical and needs to be considered when developing a FDD model.

2.5.3 Prevalence of false alarms during process monitoring

Fault detection and diagnosis for water distribution system is made more difficult by the non-stationarity of water uses in non-residential buildings. This non-stationarity can lead a high prevalence of false alarms when conventional fault detection approaches are used which has led to a relatively low industry uptake of FDD systems. The prevalence of false alarms and the fault detection sensitivity and the prevalence of false alarms have not been thoroughly investigated in non-residential buildings (Touzani et al., 2019; Ihasalo and Karjalainen, 2014). A relatively high prevalence of false alarms will reduce confidence in any FDD system and cause unnecessary downtime and associated cost. Thus, can cause hesitation in building managers when responding to true faults (i.e., faults and non-routine water usage events) during performance monitoring of non-residential water distribution systems. Setting up a balance between sensitivity in detecting faults and robustness in minimizing false alarms is a challenging task. There is a real risk of incorrect detection when faced with high levels of uncertainty, and the cost of failing to detect faults must be weighed against the cost of having to respond to a false alarm. Building managers may even turn-off or ignore the fault detection system if there are too many false alarms and may respond to faults based on existing knowledge or experience or only if the fault causes perceptible impacts. This may be exacerbated where the economic cost of water is relatively low and thus, for example, inefficient water use or leaks that don't cause perceptible impacts may be ignored or not prioritised. Eventually this may lead to building managers ignoring alarms once their systems are operating (however inefficiently). It is difficult to specify the appropriate fault sensitivity, since the conflicting economic cost of failing to detect a fault or having to deal with a false alarm is usually unknown (Dexter and Pakanen, 2017). This emphasises the need for false alarm moderation approaches to control false alarms and non-detection effectively, while retaining the capability of detecting and diagnosing faults in non-residential water distribution systems.

A number of false alarm moderation methods that utilise machine learning-based classifiers such as ANN, SVM, decision trees, etc. have been applied in different engineering domains such as air handling units, ethylene cracking, medicinal informatics, etc. (Fernandes et al., 2020; Bruton et al., 2014; Kämpjärvi et al., 2008; Malsburg and Angele, 2017; Zhang et al., 2015; Wang et al., 2016; Hubballi and Suryanarayanan, 2014). Although effective in some sectors, these methods usually require a considerable amount of training data with "false alarm" data points, resulting in higher computational and labelling cost (Zhao et al., 2019; Hubballi and Suryanarayanan, 2014). This may be impractical in non-residential water distribution systems, where data from known false alarm situations can be difficult to obtain due to the non-stationarity in water uses over time (e.g., variability in production activities over time). In contrast, statistical approaches like PCA have been shown to be effective in some sectors when dealing with false alarm issues where limited training data is available based on studies conducted in the context of health monitoring systems and building energy systems (Xu et al., 2017; Zhao et al., 2019). However, the non-stationarity of water consumption in many non-residential water distribution systems (or indeed in other large facilities such as large public buildings) can cause difficulties and impede the ability of PCA methods to distinguish between false alarms and true faults (Prabuchandran et al., 2019). Thus while the application of PCA as a FDD method for non-residential water systems can reduce false alarms when compared to more basic univariate methods, challenges relating to false alarms persist even with such improvements (Wei et al., 2017; Chen, 2010).

These challenges and related false alarms can be due to a mismatch between the trained PCA model and the incoming data being tested (which in practice can be due to incorrect or inaccurate measurements) (Li et al., 2018; Chen, 2010). Borderline flow fluctuations which are close to the control limits in the developed PCA model to detect fault can also lead to false alarms or undetected faults (Wei et al., 2017). Finally, faults resulting in false alarms can be linked to the inherent definition of the control limit in process monitoring (Chen, 2010). For instance, a 95% confidence level infers that 5% of routine operating data will exceed the control limit and could result in false alarms. While the importance of minimising false alarms from a building manager's and water conservation perspective is clear, little work has been done in this area for the non-residential water sector.

2.5.4 Conclusion

FDD techniques have been applied (albeit relatively recently) to water distribution systems and studied extensively from leak detection perspective mainly at a municipal supply level (Escofet et al., 2016; Robles et al., 2016). Initial work in this area has mainly focused on leak detection based on software-based approaches using modelled data to detect leakage or burst pipes in water supply networks (Stevenson 2020; Wu et al., 2016; Almandoz et al., 2005). Currently, most building managers react to faults on an ad-hoc basis, responding to obvious faults and repairing infrastructure as required (Moser et al., 2015). This can result in low-level imperceptible faults persisting within a water distribution system, compromising system's efficiency. The application of existing approaches developed for municipal water supply networks may be limited within non-residential buildings as key faults not only involve leaks but also comprise system related faults such as equipment malfunctioning, operational errors, etc. and non-routine water uses caused by inefficient water usage and infrequent sudden changes in demand. It is advantageous within buildings to distinguish between faults, inefficient water usage and non-routine water uses which may or may not be the result of a fault.

Building water distribution system studies have not leveraged the more sophisticated FDD approaches used in other sectors such as PCA or SVM (Gharsellaoui et al., 2020; Zhao et al., 2019). PCA has advantages over classical statistical methods (Shewhart chart, cumulative sum, exponentially weighted moving average) which while accurate in many cases, have been shown to produce inaccurate results when introduced to multivariate systems such as non-residential building water distribution systems (Maione et al., 2019; Xiao et al., 2017). Thus, the development and validation of robust FDD systems that can detect system alarms, distinguish between faults and non-routine water uses in non-residential water distribution systems is lacking, but such a system has the potential to significantly increase the efficiency of water consumption (Pelz, 2003; Danacova et al., 2016). Thus, non-residential water sector can benefit from unsupervised and supervised learning models (PCA and SVM). The present work aims to fill these gaps by developing decision models based on real sites data that can detect faults in non-residential buildings without the use of geometrical information.

Customarily, modern non-residential buildings are commissioned with building management systems that can gather large amount of water use data in real-time which is generally only used at a macro level (high-level water use statistics such as daily flow) (Sousa et al., 2019). A question arises as to how to further improve building water efficiency by utilizing the monitored data. Performance monitoring based FDD has demonstrated to be one of the most effective ways to achieve such efficiency (Zhao et al., 2019). It can play a crucial role in reducing equipment downtime, added water-energy use and associated maintenance cost in the non-residential water distribution systems. Older buildings may lack such systems but can often be relatively easily retrofitted with water meters. Thus, in many cases data (or potential for data) may exist to enable sophisticated and practical FDD with resulting improvements in water consumption efficiency if appropriate methodologies exits.

This study utilizes PCA and SVM to develop a FDD methodology for non-residential building water distribution systems for the first time. To this end, PCA and SVM have not been applied to non-residential water distribution systems. To date the designed FDD methodology will not only detect and diagnose faults in the non-residential water distribution systems bridging the gap in the existing literature but will also comprise a robust means of identifying outliers within the water time series, leading to an advancement in use of these computational approaches without explicit knowledge of the system. The successful implementation of PCA and SVM in other engineering sectors motivates this research to adopt these approaches in water sector for the performance monitoring of large-scale water distribution systems in industrial setting. The following chapter describes the practical interpretation of the designed FDD methodology termed as “FastDetect”, and the steps involved in detail.

3. FastDetect METHODOLOGY DEVELOPMENT

3 FastDetect Methodology Development

3.1 Overview

This chapter details the development of FastDetect for non-residential water distribution systems. FastDetect is based on PCA combined with false alarm reduction approaches and SVM techniques to reduce false alarms and categorise faults, respectively. The methodologies were developed and validated using leveraged time-series water flow data from case-study sites, which is discussed in Chapters 4 and 5. FastDetect was developed using MATLAB 2021a, a computer programming platform widely used among engineers, chemometricians, statisticians, etc. both in industry and academia.

3.2 Principal Component Analysis

PCA is based on an orthogonal decomposition of the covariance matrix of the original statistical features along the direction that explains the maximum variance within the dataset (Creaco et al., 2019). In principle, PCA linearly transforms input data to new orientations (a set of statistical features) known as principal components in such a way that they are uncorrelated (Silva et al., 2019; Mujica et al., 2011). The statistical features are measurable characteristics that are defined and calculated through statistical analysis (Qin and Chiang, 2019). The uncorrelated statistical features are created as a linear combination of the original statistical features ordered by an amount of explained variance in component direction known as principal components. Thus, the first principal component holds most of the inherent variance of the original data (Kamiel, 2015). The outline of PCA models is shown in **Figure 3.1**.

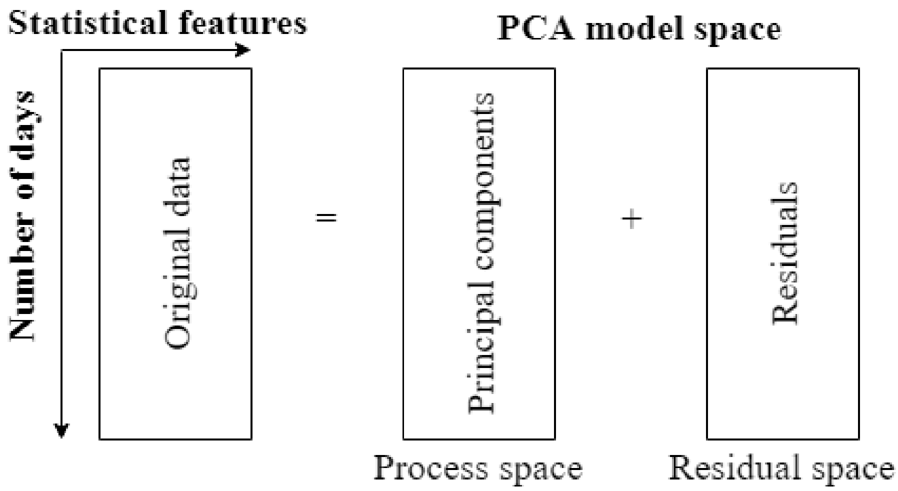


Figure 3.1: Outline of PCA model.

Each principal component is orthogonal to the other principal components to explain the variation that is not already explained by other statistical features used to describe (in this context) baseline water usage and diurnal patterns and results in a set of transformed principal components. The reduced number of statistical features (data dimensionality) simplifies interpretation of the data, making the following analysis more convenient. It can also significantly reduce the computational cost depending on the data dimensionality (the lower the data dimensionality the more computational cost can be saved). Thus, leveraging the process history data by eliminating multicollinearity (inter correlation among the statistical features).

PCA can not only be used for dimensional reduction of original datasets but also for FDD, through leveraging PCA detection indices namely Hotelling's T^2 and Q-statistics (Sgorbissa, 2018). T^2-statistics measure the variability of a score matrix and can detect abnormality within new data by comparing it to variation in the parameters defined during baseline or "routine" conditions. Q-statistics evaluate the variability of a matrix which is the projection of the original data onto a residual subspace (Benaicha et al., 2013; Ahmed et al., 2012; Zumoffen & Basualdo, 2008; Sgorbissa, 2018).

In the context of this work, the observations in **Figure 3.1** could refer to mains daily water consumption data, the statistical features could refer to average daily consumption, minimum flow rate, maximum flow rate during the day (for example). PCA summarizes these statistical features (which can often be correlated) into a set of fewer uncorrelated statistical features (principal components) that define the majority of the original inherent variation in the dataset. Thus, PCA eradicates the redundancy or correlation that could exist in the original dataset between several different statistical features.

3.2.1 Interpretation of T^2-statistics and Q-statistics

To measure the variation of samples within the PCA model and detect the abnormality of the new incoming data, Hotelling T^2-statistics, and Q-statistics (squared prediction error) have been applied in other engineering domains (Li et al., 2018; Horrigan et al., 2018; Zhang et al., 2017; Villegas et al., 2010). The concept underlying the use of PCA detection indices (Hoteling T^2-statistics and Q-statistics) for fault detection (for a dataset with three dimensions (i.e., 3 features)) is shown as an example in **Figure 3.2**, whereby the data points lie primarily in a plane, described by two principal components. The first principal component aligns with the maximum variation in the data, where the second principal component aligns with the maximum variation (i.e., orthogonal to the first principal component) (Sgorbissa, 2018). In **Figure 3.2**, a fault is indicated by the red circles which are beyond the control limit boundary (black solid line) and can be detected by analysing Hoteling T^2-statistics and Q-statistics. In the context of FDD, the range or normal operation of a system as defined by the T^2-statistics α-control limit (T^2_α) is generally larger than that as defined by the Q-statistics α-control limit (Q_α). Therefore, minor variations in a system can more easily result in Q_α being exceeded, whereas T^2-statistics have large variance and require larger variance for T^2_α to be exceeded. In the case of a non-residential water distribution system, low-level imperceptible faults such as continuous flow via low level leakage or a tap left running, etc. would result in minor variation of "routine" operating conditions and may be detected by Q-statistics. Whereas larger faults such as excess usage, metering error, larger leakages, etc. would result in more variation in the system and can be detected by T^2-statistics.

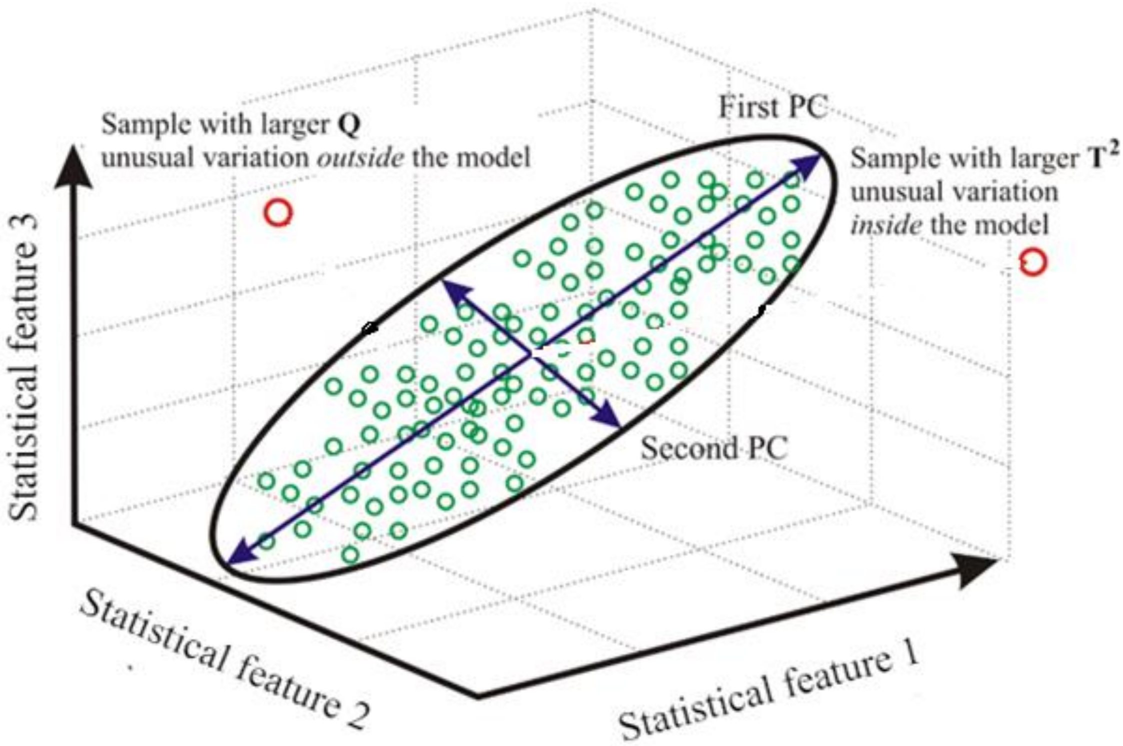


Figure 3.2: PCA model of three dimensional data set showing T^2-statistic and Q-statistic outliers (Sgorbissa, 2018).

The overall layout of FastDetect developed for the non-residential water distribution systems is shown in **Figure 3.3**. The practical interpretation of FastDetect is discussed individually in five different phases below.

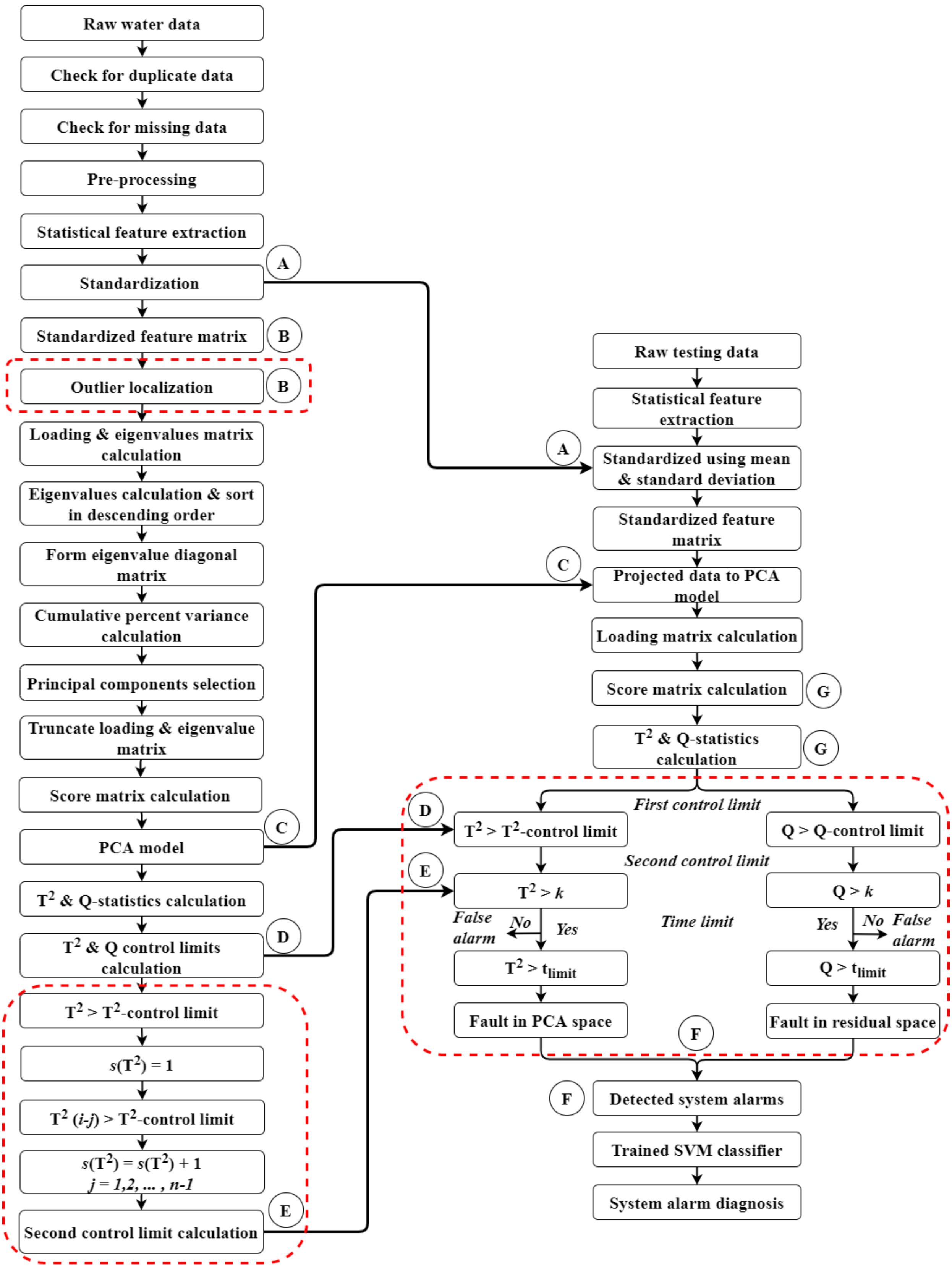


Figure 3.3: FastDetect layout. The red dashed lines indicate the key contributions to knowledge – letters denote connections between different stages of the FDD process.

3.3 Practical interpretation of FastDetect for non-residential water distribution systems

Prior to statistical feature extraction, and for the purpose of model development, pre-processing of the raw data can be done to resolve known data issues such as missing data (which may be confused with "zero" flow events) or overlap/repeated readings to ensure the quality of data used for FDD model development as these can have a significant impact on the resulting PCA model and thus reduce the performance of the resulting FDD model. To illustrate the methodology, as an example real raw water meter readings at an interval of 7.5 minutes are shown in **Table 3.1** and used to show how the data is analysed. Prior to analyse the water data, the recorded water meter readings were aggregated into hourly flow traces (8 readings at 7.5-minute intervals averaged over one hour).

Table 3.1: Example of a real raw water flow readings.

Date	Meter No.	Water flow readings (7.5 minutes interval)						
Time		00:00	00:07:30	…	23:30:00	23:37:50	23:45:00	23:52:50
07/01/2016	1024	0	0		1.6	4.8	0	7.2
08/01/2016	1024	0.8	0	…	0	0	5.6	1.6
09/01/2016	1024	0	0	…	0	0.8	4.8	0
10/01/2016	1024	0	0	…	1.6	5.6	0	0
11/01/2016	1024	1.6	8.8	…	0	4.8	4	0
12/01/2016	1024	3.2	4	…	5.6	4.8	0	0
13/01/2016	1024	0	7.2	…	0	0	6.4	0
14/01/2016	1024	0	0	…	9.6	0	3.2	8.8

3.3.1 Statistical features extraction

Statistical features of the water distribution system (derived from flow meter readings) were extracted to analyse baseline water usage and diurnal patterns (or indeed patterns at any chosen time interval for which adequate data is available). Such features could include average daily flow, maximum daily flowrate (over a certain period), average flows over various parts of the day, etc. To extract statistical features for various time periods, knowledge of the site is preferable as these periods will likely coincide with working day hours, non-working day hours, daily production cycles, etc. The use of the statistical features to describe overall patterns is a well-accepted approach where the data variation is non-stationary (and reflects the situation in many non-residential water distribution systems) (Fernández et al., 2018; Ding et al., 2019). The example of a statistical feature matrix and standardized statistical feature matrix are shown in **Table 3.2** and **Table 3.3**. The goal is to extract the useful information which may be inundated in a large number of redundant water distribution system data and eliminate the influence of overlap or repetitive readings. **Table 3.2** indicates the features extracted over a certain period such as maximum flow is the highest flow recorded in 24h (from midnight to midnight), minimum flow is the lowest flow recorded in 24h (from midnight to midnight), average flow is the average flow recorded in 24h (from midnight to midnight), working hours flow is the average flow recorded during the working hours (from 6 a.m. to 6 p.m.), non-working hours flow is the average flow recorded during the non-working hours (from 6 p.m. to midnight), night time flow is the average flow recorded during the night time (from midnight to 6 a.m.) (discussed in Chapter 4 - Section 4.3.3.).

Table 3.2: Example of a statistical feature matrix with water flow data (m^3/day).

Days	Maximum flow	Average flow	Working hours flow	Non-working hours flow	Night time flow	Minimum flow
1	2.00	0.80	0.63	0.92	0.70	0.81
2	3.10	1.93	2.70	1.56	0.73	0.97
3	3.50	1.75	2.58	1.87	1.13	0.94
4	3.20	1.83	2.23	1.63	1.25	0.87
5	4.10	1.45	2.33	1.87	0.88	0.97

Table 3.3 represent the standardized form of statistical feature matrix (**Table 3.2**) obtained by subtracting the population mean from each feature matrix value, divided by the population standard deviation. The values in **Table 3.3** describes the position of a feature value in terms of its distance from the mean in terms of standard deviation. For example, negative value of maximum flow (-0.67) at day 1 (**Table 3.3**) indicates that it is at -0.67 standard deviation far from the mean. Likewise, positive value of minimum flow (1.00) at day 3 (**Table 3.3**) indicates that it is at +1 standard deviation far from the mean.

Table 3.3: Example of a standardized feature matrix.

Days	Maximum flow	Average flow	Working hours flow	Non-working hours flow	Night time flow	Minimum flow
1	-0.67	-0.90	-0.79	-0.29	-0.36	-1.11
2	-0.74	-1.39	-1.09	-1.04	-0.97	-1.07
3	0.04	0.89	0.66	-0.48	-0.73	1.00
4	0.32	0.76	0.66	0.83	0.77	0.68
5	0.11	0.37	0.18	0.42	0.57	0.82

The statistical features matrix *X* comprises *n* rows x *m* columns, where *n* rows represent the values for all features within a single day and *m* columns represent a single feature such as maximum flow, average flow, etc. over the number of days. The statistical features matrix is then used as an input to build the PCA model for fault detection. A standardised statistical feature matrix is developed that transformed data to have zero-mean and unit-variance which comprised the time-series data for each of these statistical features (as per **Table 3.3**). The matrix is then standardized using a mean (μ) and a standard deviation (σ) calculated for each feature to give equal influence of original data on the PCA model and eliminate bias that can occur due to the use of different units (e.g., one statistical feature measured in m^3/hour, while other is in litre/minute, etc.) (Appendix A-Eq. 1-4). The standardized parameter vectors (the mean and standard deviation outlined above) were stored for later use (for standardizing new incoming data in the fault detection process – Section 3.3.3.). The steps involved in obtaining the statistical feature matrix from raw water data are shown in **Figure 3.4**.

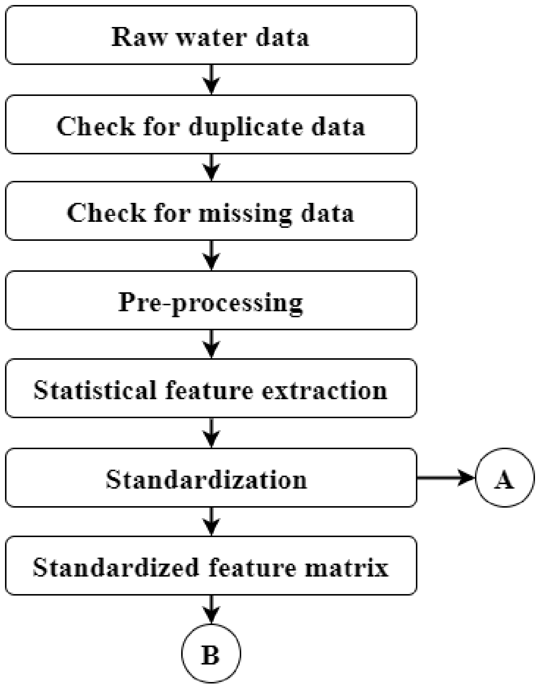


Figure 3.4: Statistical feature extraction process – letters denote connections between different stages of the FDD process (see Figure 3.*3*).

3.3.2 Fault detection model development

The standardized statistical feature matrix obtained was used to initialize the PCA modelling process. The statistical feature matrix was then transformed by utilizing singular value decomposition (Appendix A-Eq. 5) to calculate the loading matrix P of n rows x m columns and eigenvalues vector λ of 1 rows x m columns containing the true variation in the standardized dataset. Where each column of loading matrix P is populated by the eigenvectors associated with the covariance matrix. The covariance matrix quantifies the amount of linear correlation between all possible combinations of statistical features within the dataset (Appendix A-Eq. 6-7). The diagonal of the covariance matrix defines the variance explained by those statistical features and non-diagonal elements are the pairwise covariance. The decomposition of standardized datasets to PCA model is shown in **Figure 3.5**.

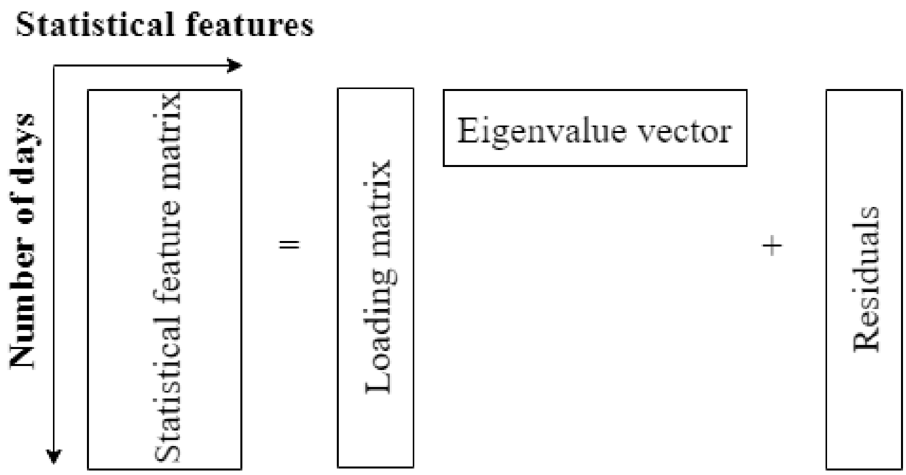


Figure 3.5: Decomposition of data to PCA model space (Appendix A-Eq. 5).

The computed eigenvalue vector is then further transformed into diagonal matrix Λ of *m* columns x *m* rows square matrix containing non-negative eigenvalues on the main diagonal. The loadings describe how much each feature contributes to a particular principal component. The sign (positive or negative) of a loading indicates whether a feature and a principal component are positively or negatively correlated (Jolliffe, 2002). The eigenvalues in the main diagonal represent variance correspond to the eigenvalues vector (statistical features in data matrix). Hence, arranging the eigenvalues in the order of highest to lowest provides an ordered orthogonal basis (greatest variance to smallest in the original dataset). The example of a loading matrix *P* and eigenvalues vector λ is shown in **Table 3.4**.

Table 3.4: Example of a loading matrix (left), and eigenvalues vector (right).

0.36	-0.23	0.03	0.11	0.28	0.75
0.38	-0.13	0.23	-0.22	-0.01	-0.04
0.36	-0.25	0.12	0.11	0.55	-0.29
0.35	-0.09	-0.31	0.52	-0.14	-0.47
0.35	0.07	-0.44	0.06	-0.54	0.30

6.07	1.17	0.78	0.39	0.25	0.16

3.3.2.1 Outlier localization

Measurements in water distribution systems contain outliers arising from issues such as sampling, measurement, and data logging issues or from faults. PCA is based on the covariance structure of the data in PCA space and is sensitive to outliers (Rousseeuw et al., 2018; Hubert et al., 2016). Such data points may not represent realistic process behaviour and can have an adverse influence on the accuracy of the model output when developing the model. Caution should be exercised when removing outliers and an understanding of the process being modelled will help ensure that outliers are removed carefully at this stage (Xu et al., 2020; Hubert et al., 2016). For example, some non-routine water use events may result in higher-than-expected water usage such as conferences hosted by a facility, short-term increases in product production, unscheduled cleaning activities, etc. and may be thought of as outliers and removed at model development stage. However, these may contain valuable information which could be lost if the outlier is removed and can improve the sensitivity of the PCA model.

In this research outlier localization was carried out by combining PCA with a distance-distance approach for outlier localization which computes the distances in the PCA plane and orthogonal directions of each data point to the centroid of the dataset in the PCA space (Appendix A-Eq. 10 & 11). In this context, data points can be categorized into four data types as (i) routine data points, (ii) good leverage data points, (iii) orthogonal outliers and (iv) bad leverage data points (Rousseeuw and Hubert, 2018; Harris et al., 2014).

- The data points with low score and orthogonal distance values are considered as the routine data points that follow routine water use: illustrated in **Figure 3.6** using a green dashed line.

- The data points with a large score distance but a low orthogonal distance (data points 1 & 4 shown in **Figure 3.6** using a blue dashed line), can be considered as good leverage points such as non-routine water uses. These data points lie close to the PCA space and not far from the routine pattern. This implies that they differ from the majority of the routine data points. These data points can be attributed to non-routine water uses resulting from an event during that period, could contain useful information and may improve the accuracy of the PCA model.

- The data points (2 & 3 shown in **Figure 3.6** – red dashed line) with a large score and orthogonal distance values represent outliers (orthogonal outliers). These data point’s statistical properties differ significantly from the other data points in the dataset. These data points can have a strong influence on the PCA model output and can be eliminated during the PCA model development.

- The data points with a low score distance but large orthogonal distance (data point 5 shown in **Figure 3.6** - orange dashed line), can be expressed as bad leverage data points. These data points lie far from the PCA space and could be caused by the noise in the data and could affect the accuracy of the PCA model.

The outlier localization process is shown in **Figure 3.6** by means of distance-distance plot using artificial data.

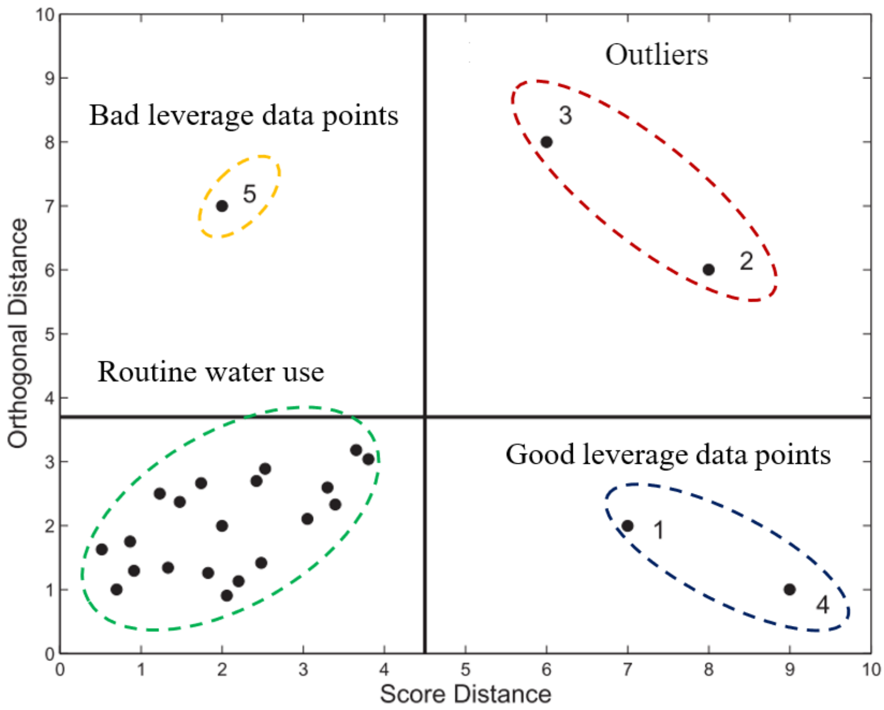


Figure 3.6: Distance-distance plot.

3.3.2.2 Selecting principal components

Selecting the number of principal components that represents the original dataset is one of the key characteristics in PCA model development as it can affect PCA model adaptability and sensitiveness to the borderline flow fluctuation which are close to the α-control limits (T^2_α and Q_α). For example, selecting too many principal components can trigger system alarms even in the case of routine water use and thus reduce the reliability and limiting the applicability in of FastDetect in industrial settings. It should also be noted that previous modelling experience and process knowledge of water distribution systems can also be used to make decisions on the number of principal components to select.

In literature, the most widely used techniques to choose number of principal components are the Eigenvalue-One-Criterion, Scree Test, and Cumulative Percent Variance (Silva et al., 2019). To illustrate the techniques, the eigenvalues, and their corresponding percentage variability for datasets from the case-studies used in this research are given in **Table 3.5**. The nine eigenvalues each correspond to a statistical feature (maximum flow, average flow, average flow during working or non-working hours, etc.). They are shown in order of the variability (from highest to lowest) of the original dataset. In other words, **Table 3.5** shows the variance distribution in the original dataset for each statistical feature.

In Eigenvalue-One-Criterion, the number of principal components is usually selected as those which have eigenvalues greater than one. Since in the covariance matrix obtained during the decomposition process, the variance of each statistical feature is one. Therefore, principal components having a variance of less than one are presumed to have less information than the original dataset and are generally ignored in principal components selection. Based on Eigenvalue-One-Criterion, case-study 1 (Alice Perry Engineering building NUI Galway) has two principal components with values greater than one. These account for about 80% of the total variability (see cumulative variance column - **Table 3.5**). Case-study 2 (Irish food and drinks company) has three principal components with values greater than one and these account for about 74% of the total variability.

Table 3.5: Eigenvalues and associated variance for case-study 1, and case-study 2.

Case-study 1				Case-study 2		
PCs	Eigenvalues	Component variance (%)	Cumulative variance (%)	Eigenvalues	Component variance (%)	Cumulative variance (%)
1	6.0704	67.449	67.45	3.6236	40.263	40.263
2	1.1726	13.029	80.48	1.8392	20.436	60.699
3	0.7777	8.641	89.12	1.2051	13.390	74.088
4	0.3850	4.278	93.40	0.9139	10.155	84.243
5	0.2486	2.763	96.16	0.6078	6.753	90.996
6	0.1649	1.833	97.99	0.3889	4.322	95.318
7	0.1039	1.155	99.15	0.2722	3.025	98.342
8	0.0628	0.697	99.84	0.0968	1.076	99.418
9	0.0140	0.156	100	0.0524	0.582	100

This can be visualised using a Scree Test, whereby eigenvalues of principal components are plotted in descending order. The number of selected principal components are generally those that occur before the elbow (Liliana and Ordóñez, 2008; Silva et al., 2019). The Scree plot for two case-study sites is shown in **Figure 3.7**. As can be seen from **Figure 3.7**, in case-study 1 (Alice Perry Engineering building NUI Galway) an elbow occurs at principal component 2 and in case-study 2 (Irish food and drinks company) an elbow occurs at principal component 2 or 3, hence one can pick either 2 or 3 principal components for case-study 2.

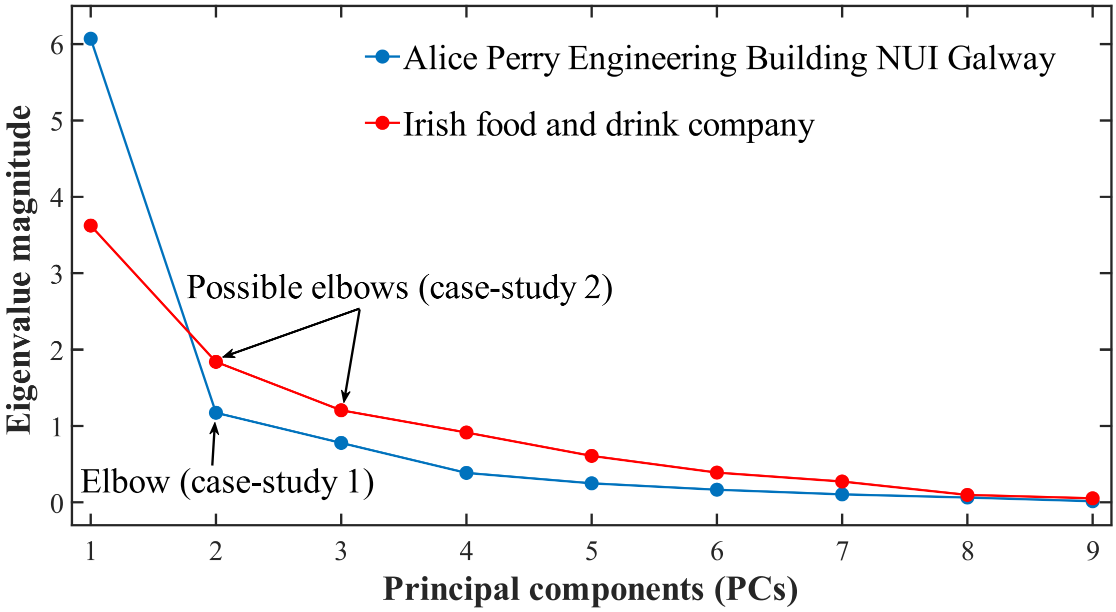


Figure 3.7: Scree plot for non-residential case-studies (Alice Perry Engineering building NUI Galway, and Irish food and drinks company).

In this research, the number of principal components that had practical significance was determined by using Cumulative Percent Variance (CPV) technique to account for as much of the variance as possible (when compared to other techniques), while retaining the defined amount of variation of the original dataset having the first *r* largest eigenvalues (Appendix A-Eq. 8). One of the simple but arbitrary rules-of-thumb is to consider the principal components that explain greater than 80-90% of the CPV as having practical significance (Silva et al., 2019). In the context of this work different CPVs (80-95%) were checked for both case-study sites by virtue of faults detected and sensitiveness to variation in the system. Having CPV 80%, the model triggered alarm on minor variation of “routine” operating conditions and resulted in increased false alarms. Whereas having CPV 95%, the model overlooked low-level variation resulted from non-routine water uses which could be due to inefficient water usage, infrequent sudden changes in demand or due to low-level imperceptible fault and triggered on significant variation in the system. To retain the maximum variability of the process, the cut-off of having CPV greater than 90% has been set in this research - CPV(r) ≥ 90%; where r is the retained principal components.

3.3.2.3 Principal component analysis model formation

Once the number of principal components to be retained were defined, the loading matrix P and eigenvalue matrix Λ were reduced – i.e., the original loading matrix P of size n rows x m columns was reduced to n rows x r columns, where $r < m$ and r are the retained principal components in the PCA model (Appendix A-Eq. 9). Similarly, the eigenvalue matrix Λ was reduced to r rows x r columns. The reduced loading matrix and reduced eigenvalue matrix were denoted $\hat{\hat{P}}$ and $\hat{\hat{\Lambda}}$ matrix, respectively. To compute monitoring statistics (Hoteling T^2 and Q-statistics) score matrix U of n rows x r columns was obtained by multiplication of statistical feature matrix X and reduced loading matrix $\hat{\hat{P}}$. The score matrix represents the projection of each day on each of the retained principal components (Fuentes-García et al., 2018).

The α-control limits (T_α^2 and Q_α) for PCA detection indices (Hoteling T^2 and Q-statistics) were obtained by means of F-distribution and approximate distribution respectively (Appendix A-Eq. 13,15-18). Lastly, the T^2 and Q-statistics were calculated (Appendix A-Eq. 12, 14) and further analysed to obtain a second control limit k. The second control limit was developed as part of the false alarm moderation approaches and is further discussed in Chapter 5. The PCA model development process is depicted in **Figure 3.8**.

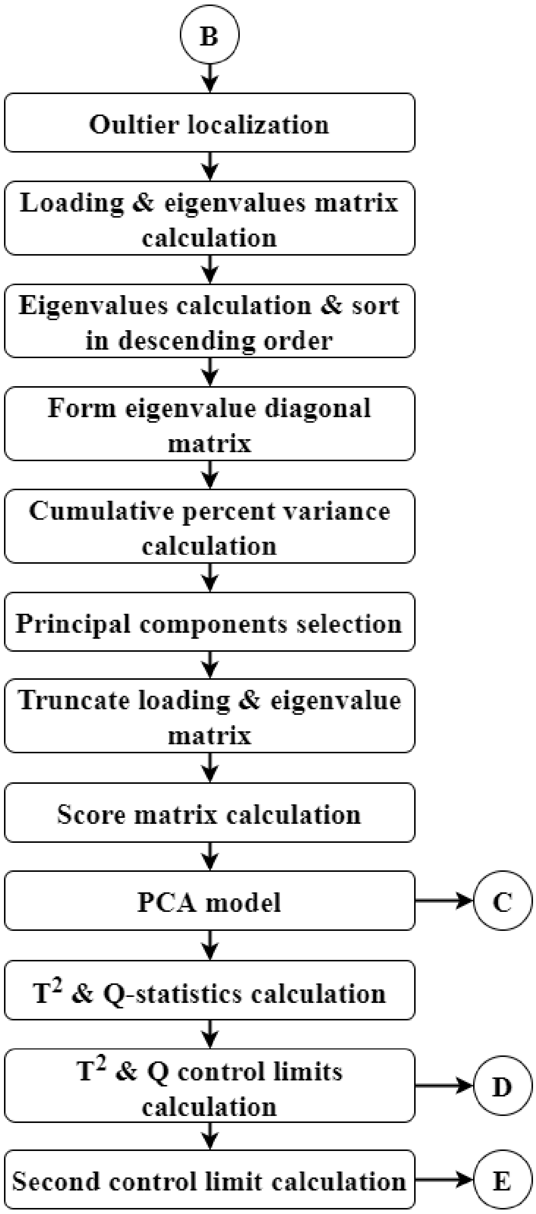


Figure 3.8: PCA model development process – letters denote connections between different stages of the FDD process (see Figure 3.3).

3.3.3 Fault detection model testing

After projection, statistical feature matrix of testing data X (n rows x m columns) transformed into the reduced loading matrix $\widehat{P}$ of n rows x r columns. The scores matrix T of n rows x r columns was obtained by the multiplication of statistical feature matrix X and $\widehat{P}$ of the testing data. The score matrix U was then used to compute the monitoring statistics (Hoteling T^2 and Q-statistics) and analyse testing data in relation to FDD. The steps encompass testing new data are summarised in **Figure 3.9**.

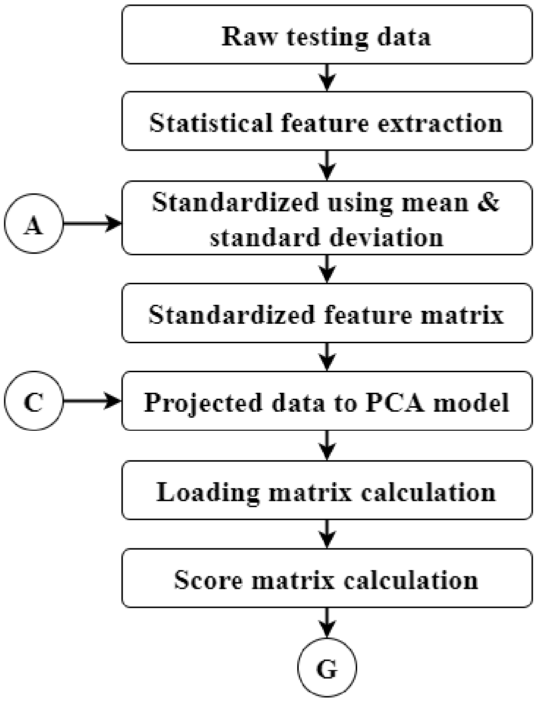


Figure 3.9: Projection of testing data to PCA model – letters denote connections between different stages of the FDD process (see Figure 3.*3*).

The T^2 and Q-statistics based on the score matrix *U* were then calculated. The monitoring statistics were then compared with the (T_α^2 and Q_α) α-control limits calculated during the FDD model development – Section 3.3.2. If ($T^2 > T_\alpha^2$ or $Q > Q_\alpha$) then the samples are considered as system alarms (warning stage) or alarms (Chapter 4) and if (T^2 or $Q > k$) (where *k* is the second control limit designed to control false alarms during fault detection process), system alarms were confirmed to have occurred (discussed in Chapter 5 in detail). In the last step, if ($T^2 > t_{limit}$ or $Q > t_{limit}$), where t_{limit} is the time series nature of the data designed in a way that when system alarms persist over two or more consecutive days fault alarm in the water distribution system is triggered (discussed in Chapter 4 in detail).

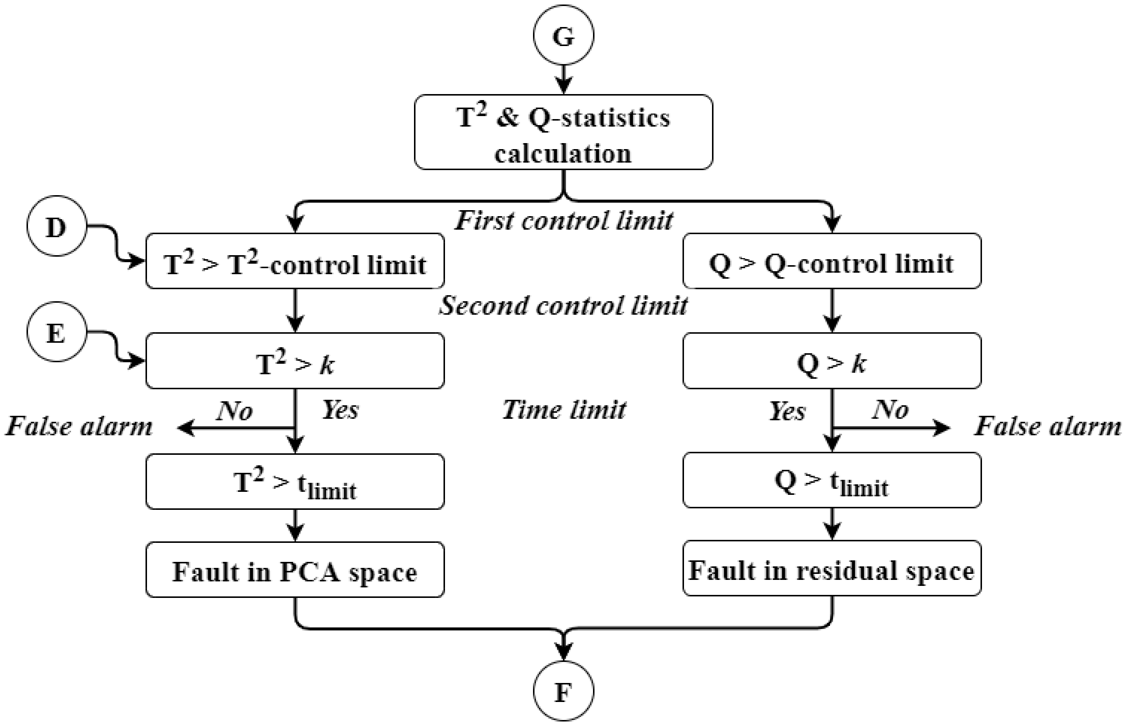


Figure 3.10: False alarm moderation and fault isolation – letters denote connections between different stages of the FDD process (see Figure 3.*3*).

3.3.4 Fault diagnosis

Once the system alarms were detected, the last phase of FastDetect was to classify the system alarms based on a trained classifier. Multi-class SVM was used in this study and has been used in different engineering sectors such as thermal power plants, centrifugal pumps, bearing, etc. in the context of classification (Sabri et al., 2017; Xiao, 2016; Bayar et al., 2015; Chen et al., 2011, Hmeidi et al., 2008). To date, PCA and SVM have not been applied to water distribution systems. In this work the multi-class SVM has been trained by considering labelled flow water data (data points with different faulty conditions outlined in the **Table 4.2**) (Appendix A-Eq. 21-22). The goal of the multi-class classifier was to classify the sample data into classes which can include “routine operation” or various fault headings (Zhao et al., 2019). Data with different system alarm labels were obtained from the training data (it should be noted that fault data may be limited in the training data due to lack of occurrence of faults in that period). The fault diagnosis process is summarized in **Figure 3.11**.

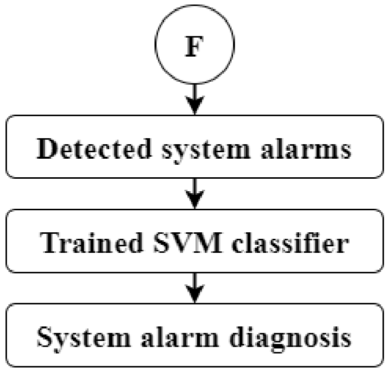


Figure 3.11: Fault diagnosis – letter denote connection between different stages of the FDD process (see Figure 3.3).

3.4 Conclusion

The literature review in Chapter 2 established that very limited research has been carried out to date within non-residential buildings in context of performance monitoring of water distribution system. This is likely due to the fact that the literature is dominated by basic fault detection systems which are not effective when introduce to non-stationarity and complexity of non-residential water distribution systems such as distinguishing between true faults or non-routine water uses, presence of outliers in real water usage data, prevalence of false alarms, etc. The objective of FastDetect was to obtain an alternative approach to physics-based models and reduce the complexities linked with the hydraulic modelling of non-residential water distribution systems. The key contribution of FastDetect is an improvement in the performance monitoring of non-residential water distribution systems through the identification of true faults at an early stage for building managers under realistic conditions. FastDetect introduces distance-distance approach combined with PCA that provides unsupervised localization of outlying data points. FastDetect also introduces false alarm moderation approaches that control false alarm occurrences during performance monitoring without the need for labelled “false alarms” data points. The datasets used in this study are typical of what is available in non-residential buildings, and thus the aim was to develop a FDD methodology which is capable of operating under these practical data constraints and assessing the effectiveness of FastDetect under these practical data constraints. Moreover, by utilizing multi-class SVM classification the faults can be classified quickly and effectively with a relatively small amount of training data.

4. A STATISTICALLY BASED FAULT DETECTION AND DIAGNOSIS METHODOLOGY FOR NON-RESIDENTIAL WATER DISTRBUTION SYSTEMS

4 A Statistically based Fault Detection and Diagnosis Methodology for Non-residential Water Distribution Systems

4.1 Overview

The aim of the chapter was to quantify the efficacy of FastDetect. Regardless of the relatively limited training data available from the case-study (which would reflect the situation in many buildings) meaningful faults were detected and the technique proved successful in classifying among different types of faults in the water distribution system. The effectiveness of FastDetect is compared to a univariate threshold technique by comparison of their respective performance in the detection of faults that occurred in the case-study site.

The study has been published in Advanced Engineering Informatics (Hashim, H., Paraic, R., Clifford, E., 2020) (https://doi.org/10.1016/j.aei.2020.101187). This chapter is arranged as follows. Section 4.2 provide an overview of conventional fault detection systems and key aspect of FastDetect to deal with the different faults in the non-residential water distribution system. In Section 4.3 materials and methods are presented which include the case study and pumping system details, measurable characteristics followed by a fault alarm reduction scheme is outlined. Later results and discussion are presented in Section 4.4. Section 4.5 summarizes the study and provides conclusions.

4.2 Introduction

Conventional fault detection systems in water distribution systems comprise alarms which trigger when relatively high levels of water are being used (Mulligan et al., 2020; Clifford et al., 2018; Quevedo et al., 2014; Perfido et al., 2016) and provide high level statistics on water consumption. Such systems are unable to identify complex patterns of water usage such as differentiating between actual faults or non-routine water uses. In non-residential building water distribution systems key faults not only involve leakage but also comprise system related faults such as equipment malfunctioning, operational errors, etc. or non-routine water uses (**Figure 4.1**).

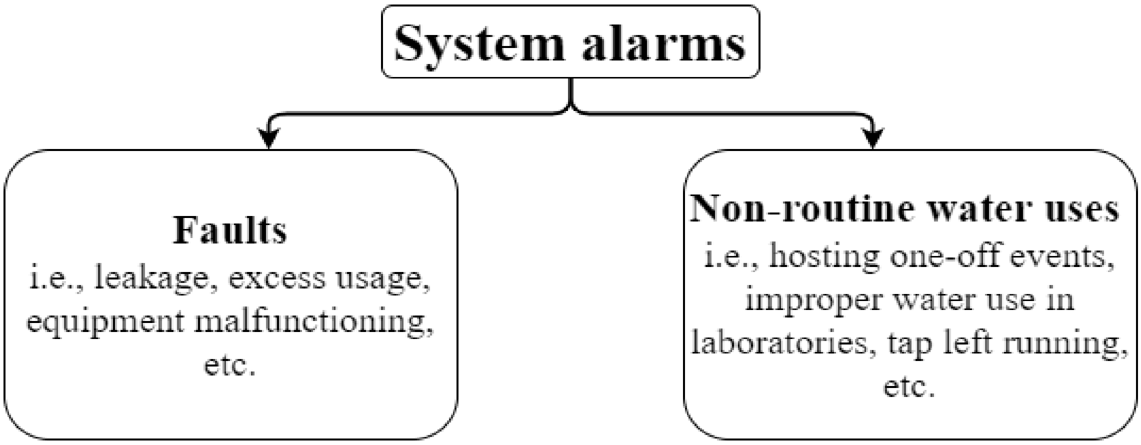


Figure 4.1: System alarm characterization.

FastDetect involves analysing water time series and monitoring inevitable changes in a system without the use of geometrical information. In the context of performance monitoring, to differentiate between features of routine and non-routine or anomalous water uses and to limit false detection. Moreover, to address the issue of outlying data points, the FDD methodology combines PCA with a distance-distance approach for outlier localization (as discussed in Section 4.3.2.1) creating a robust approach where principal components are not influenced by outliers. FastDetect was assessed using real site water distribution system data with associated outliers and process noise leading to a more robust assessment of the machine learning techniques suitability to detect and diagnose faults in water distribution system.

This chapter leverages a case-study building with real-time water consumption data to illustrate the effectiveness of using PCA detection indices (Hotelling T^2-statistics and Q-statistics) in identifying faults in a non-residential building water network. FastDetect applies the PCA technique to measured time-series water data to (i) identify faults and (ii) classifying the faults into various categories represented by prediction model output data to reset and guide operational performance.

4.3 Methods and materials

4.3.1 Case Study

The case-study site was the Alice Perry Engineering Building at the National University of Ireland, Galway (NUI Galway) (**Figure 4.2**). The building is one of the most state-of-the-art university buildings in Ireland and has been operational since 2010. The building is located on the west coast of Ireland in an area with a temperate marine climate. The building houses approximately 1,100 students and 100 staff during teaching and exam terms (approximately 26-28 weeks a year in total) and about 150 research, academic, administrative, and technical staff and 50-100 postgraduates during the rest of the year. The building includes lecture halls, classrooms, offices, laboratory facilities, a café, and shower and toilet facilities spread across 14,000 m^2 of floor space on four storeys. Thus, it has a variety of end-uses of its water and hot water systems. The building is managed through a building management system that collects data on building performance and operational efficiency - including 11 water meters and a number of energy meters. Some of the key water uses include showers and hand wash basins, grey water from a rainwater-harvesting system for toilets and urinals and potable water for the water fountains and the café. This study deals primarily with mains water usage within the building (i.e., hot water system is not considered).

Figure 4.2: Alice Perry Engineering Building at the National University of Ireland, Galway (NUI Galway).

4.3.2 Building Pumping System

The mains water system (MWS) in the building is divided into a cold-water system (CWS) and a potable water supply for drinking fountains and for the café. There are two sets of booster pumps (**Figure 4.3**), that deliver (i) the cold-water system and (ii) potable water supply. The cold-water system water is then divided into water for grey-water system (GWS) applications and water for other uses such as laboratories, hot water mixing in showers and faucets and non-potable uses in the café. The grey-water system supplies water to toilets and urinal flushing as required by gravity. The building has a large rainwater-harvesting system (RWHS) for supplying grey water. When the rainwater-harvesting system cannot supply sufficient grey water, two holding tanks located on the roof of the building (8 m^3 capacity each) are topped up by the cold-water system. During the study period, the building rainwater-harvesting system utilised the cold-water system even during periods when rainwater was available due to system faults that remained undetected prior to this study.

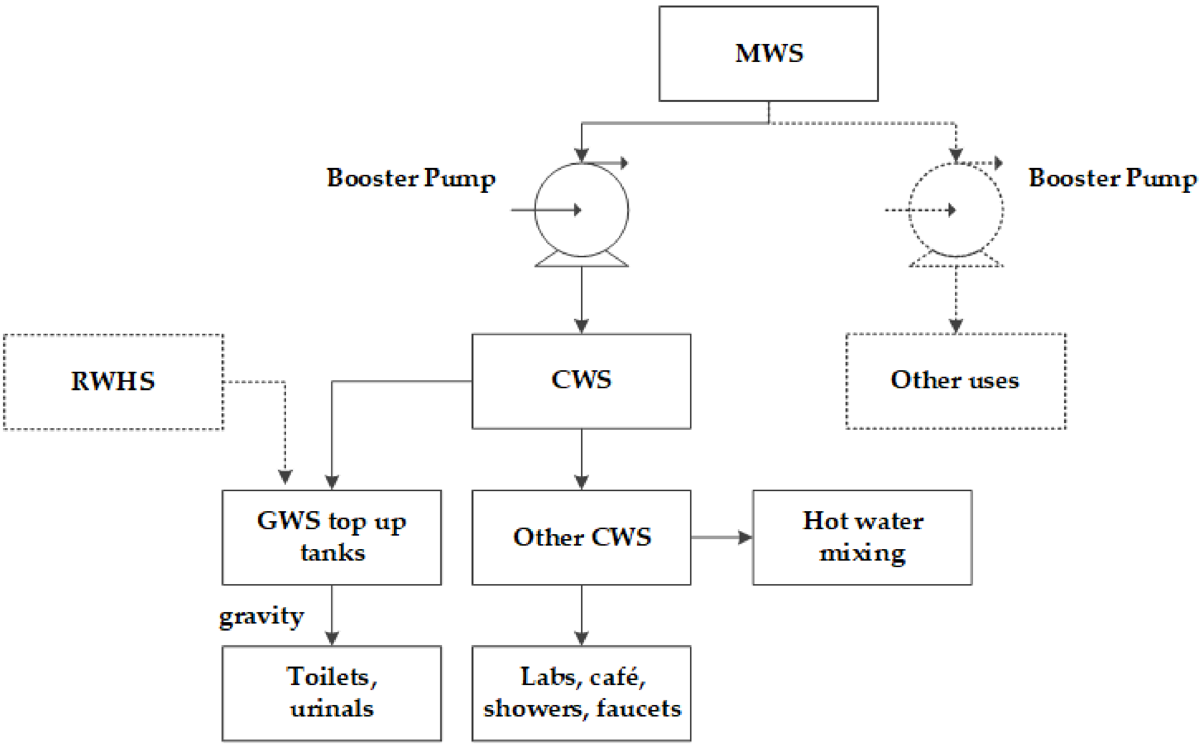


Figure 4.3: Simplified schematic of water network in the building. The dashed line indicates systems not considered in this study.

4.3.3 Statistical feature extraction

To perform FDD in water distribution systems, measurable characteristics can play a vital role in discriminating between routine and abnormal events in the water network. The characteristics of the system, derived from meter readings, can be used to analyse the water usage baseline and diurnal patterns. For instance, analysing night time flow can assist in detecting continuous flow which can be due to leaks; analysing peak flows can assist in detecting exceptional events. Analysing water consumption at specific daily intervals (working and non-working hours) can help in identifying disruptions to normal routines. The details of the features utilized in this study are listed in **Table 4.1**. The daily flow and maximum flow features summarize the water usage over a 24-hour period from midnight to midnight. Whereas the remaining features (working hours, non-working hours, and night time flow) focus on water consumption for particular periods of the day which are linked to how the building is used. In this case, these periods comprise of four working hour time intervals, two non-working hour time intervals and a night time interval. Medium resolution data recorded at 7.5-minute intervals was collected in the case-study building from an inline displacement water meter equipped with a magnetic pulse output and analysed in developing the FDD methodology.

The selection of the historical data to train the FastDetect model can be a factor in optimizing the model sensitiveness to true alarm detection. For instance, if the building has a significant amount of historical data this may include previous faults or non-routine water uses which can be used to enhance the training of the FastDetect model. The information regarding faults or non-routine water uses can be obtained by conducting a preliminary water audit (described in detail in Section 7.2.1 - Chapter 7) which can be used in developing the FastDetect model. Such information provides in-depth description of non-routine water use events and the nature of faults that had occurred. This can also ensure FastDetect will be more robust in identifying unseen or new types of faults.

Table 4.1: Statistical features (measured in m^3/hr) used to characterise daily water demand.

	Feature	Description
Total	Daily flow	Average flow in 24 h
	Maximum flow	Highest flow in 24 h
Time of day	Working hours	Average flow between 6 a.m. – 9 a.m.
		Average flow between 9 a.m. – 12 p.m.
		Average flow between 12 p.m. – 3 p.m.
		Average flow between 3 p.m. – 6 p.m.
	Non-working hours	Average flow between 6 p.m. – 9 p.m.
		Average flow between 9 p.m. – 12 a.m.

Night time flow	Average flow between 12 a.m. – 6 a.m.

For a given year the case-study monitoring period data was divided into Semester 1 - September to December, Semester 2 - January to May and a summer period (June-August). In general, water consumption would be higher in Semesters 1 and 2 than during the summer periods as undergraduate students would not be present during the summer periods. **Figure 4.4a** demonstrates the flow pattern at different intervals (**Table 4.1**) during Semester 2 and the summer period under the routine operating condition considered for PCA model training. Limited variation can be observed between working and non-working hour water usage in the summer due to reduced building occupancy as compared to Semester periods. In **Figure 4.4b** water consumption patterns are visualized during the summer period. The flow pattern during night time, when the building was not in use, is primarily due to the routine urinal flushing which operate irrespective of the building occupancy. These patterns derived from the water time series, can assist in analysing different faults in the water distribution system.

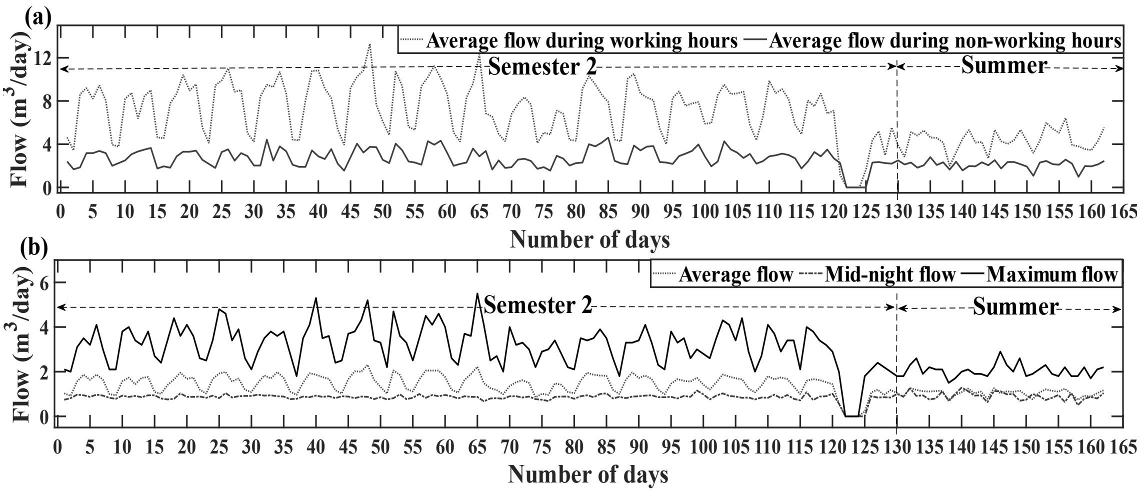


Figure 4.4: Flow characteristics within the periods of interest (a) Average 7.5-minute interval flow pattern during the 12 working hours and the 6 non-working hours, (b) Maximum, 7.5-minute interval average over 24 hours and night time flow pattern over 6 hours.

This study utilises multivariate statistical approaches (PCA and SVM) to attempt to detect patterns in data, to develop a new FDD methodology for non-residential building water distribution systems. The theory behind the statistical methods used are outlined in the next section. It is noted that the primary advantages of the combined PCA and SVM approaches are the ability to compress large data sets without losing data definition. In this case it can enable analysis of faults across a number of water usage features, both of which reduce the likelihood of false positives and false negatives within the FDD system. In addition, SVM facilitates classification of faults, providing a greater level of information to building managers.

4.3.4 Fault alarm reduction

To isolate actual fault alarms from system alarms and subsequently reduce false alarms in the FDD system, the time series nature of the data over periods of two days or more were considered. Non-residential public use buildings such as the one explored herein, are occasionally used for large events which result in significant increases in water consumption such as conferences, seminars, etc. on a given day. Thus, fault alarms in the proposed framework are raised only when system alarms persist over two or more consecutive days. Thus, when the analysed data exceeded the α-control limits (Eq. 23) for more than two days a fault alarm is triggered. The two-day time period was selected based on a historical analysis of the events within the building (conferences, seminars, workshops, etc.). The nature of the time series alarm system could vary for different buildings based on water consumption patterns and event patterns and can be established from a high-level water audit. It is also noted that this framework step can be removed for buildings where occasional large occupancy events do not occur (i.e., food processing facilities, manufacturing facilities, etc.).

$$\begin{cases} t_{acceptable} < t_{limit} --- Normal \\ t_{acceptable} \geq t_{limit} --- Indicative\ of\ fault \end{cases} \tag{23}$$

In this case a classification accuracy of 80.95% was achieved during training. The summary of the FDD methodology is outlined in **Figure 4.5**.

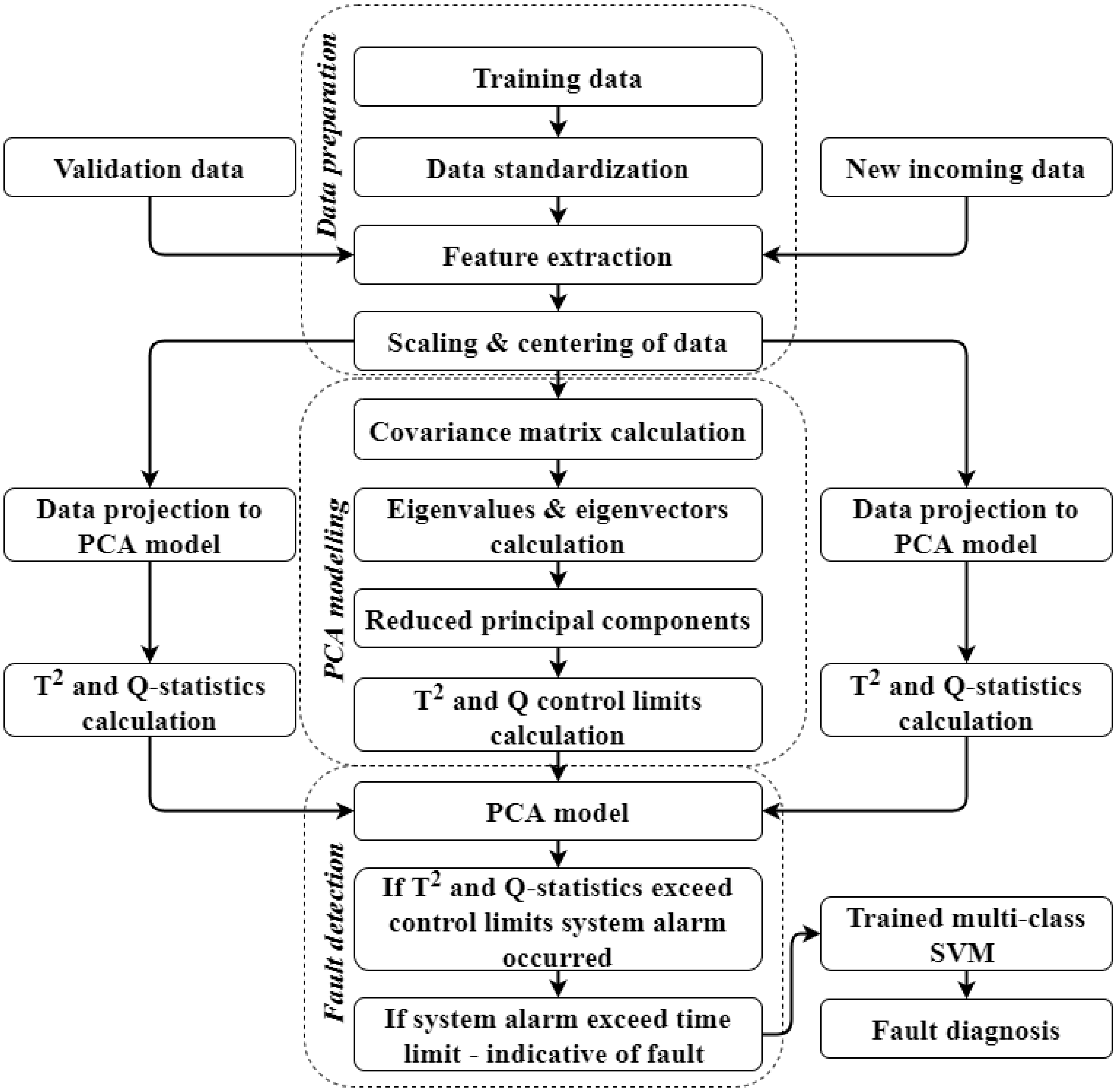


Figure 4.5: Summary of the designed fault detection and diagnosis methodology.

4.4 Results and discussion

4.4.1 Training

The water usage data was aggregated into hourly flow traces (8 readings at 7.5-minute intervals averaged over one hour) to analyse the usage characteristics at the pilot site. In a number of cases, data points had overlapping timestamps due to metering errors, and these were removed from the training dataset. Real water time series data often contains values that exhibit atypical usage characteristics and do not follow expected or routine patterns. Such data points are normally identified as outliers (Harris et al. 2014). These outliers can be due to measurement error, variation in the water use, and process noise, but could assist in stabilizing the sensitivity of the PCA model. Thus, some outliers in the dataset may contain valuable information which could be lost if all outliers are removed, while others (i.e., measurement error and process noise) can negatively affect the model accuracy and could be removed. To localize outliers in this context, the dataset was analysed by plotting distances (score and orthogonal distances) from the centroid of the covariance data structure (**Figure 4.6**). The score distance was measured within a PCA space, while the orthogonal distance was measured between data each point and its projection in the *r*-dimensional space (Rousseeuw et al., 2018).

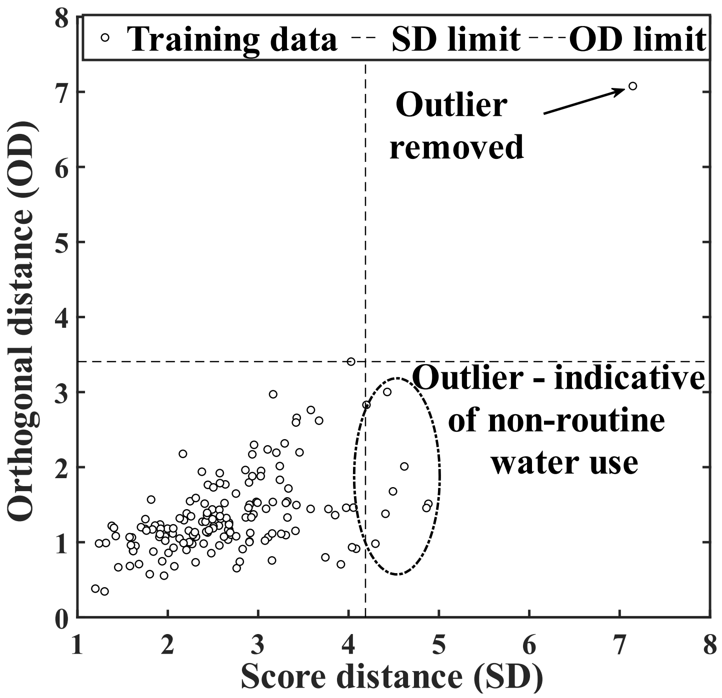


Figure 4.6: Distance-distance plot for the training dataset.

To localize the outliers within the dataset, the training dataset was analysed by computing the distance values from each data point to the centroid of the dataset in the PCA space. The basic idea of distance-distance plot is to map the outlying data points that exhibit unusual correlation with respect to the covariance structure of the data, results in larger distance values. As can be seen from the **Figure 4.6**, during training period a number of data points breached the score distance cut-off constituting to larger score distance values, indicative of non-routine water uses data points that do not follow the routine pattern. On further manual inspection of the training dataset exhibit that the average flow at the data points with large score distance values was around 48 m^3/day which was higher than the routine average flow (i.e., ~43 m^3/day). Where no orthogonal data points were observed during training period. The average flow at data point detected as outlier was recorded around 56 m^3/day, which was higher than the routine average flow, hence ignored during the PCA model development.

The PCA model and the SVM model were trained using six months of data from January to June (which incorporated a Semester of teaching and exams); the calculated PCA monitoring statistics (T^2-statistics and Q-statistics) are shown in **Figure 4.7** (dashed lines). As can be seen from **Figure 4.7**, a number of points lie above the α-control limits, constituting system alarms. These are due to a combination of non-routine water uses in the building (e.g., event at 122 days) and routine statistical variation related to the inherited definition of the α-control limit, which is linked to false alarm probability (Li et al., 2019). Importantly, none of these system alarms in the training data would trigger a fault alarm in the FDD framework as none of the system alarms occur on consecutive days (as discussed in Section 4.3.4). It is also noted that metering fault data and one obvious fault data point were removed from the training data set, in line with standard threshold development procedure.

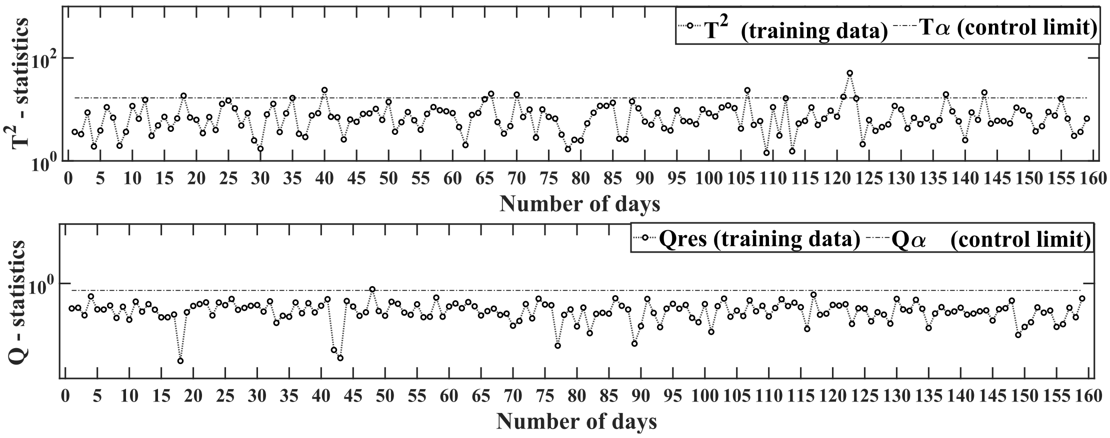


Figure 4.7: The time evolution of T^2-statistics and Q-statistics on a semi-logarithmic scale for the training data (PCA model).

4.4.2 Validation

Having trained the PCA model, a model validation was conducted on academic Semester 1 water consumption data (September-December) to assess the model's ability to detect faults. Importantly, it was known that this Semester 1 data contained an actual fault in the building water distribution system. A previous study (Clifford et al., 2018) focused on the use of low-resolution meter data to detect various flow signatures associated with different end-use cases in this building. This study demonstrated faults had occurred in the rainwater-harvesting system, which was eventually found to be due to a faulty valve on the inlet to one of the storage tanks on the roof top of the building. Further inspection of meter data from the period concerned indicated that the fault caused excess metered water consumption of 3.5 m^3/day in the building (as the cold-water system supply was engaged to fill the rainwater top-up tanks even though rainwater was available). **Figure 4.8a** illustrates the total daily usage of water over the training period and during the Semester 1 period (validation period) where the fault occurred. **Figure 4.8b & c** show the PCA model output for the validation period which utilised the Hotelling T^2-statistics and Q-statistics α-control limits calculated during the training phase.

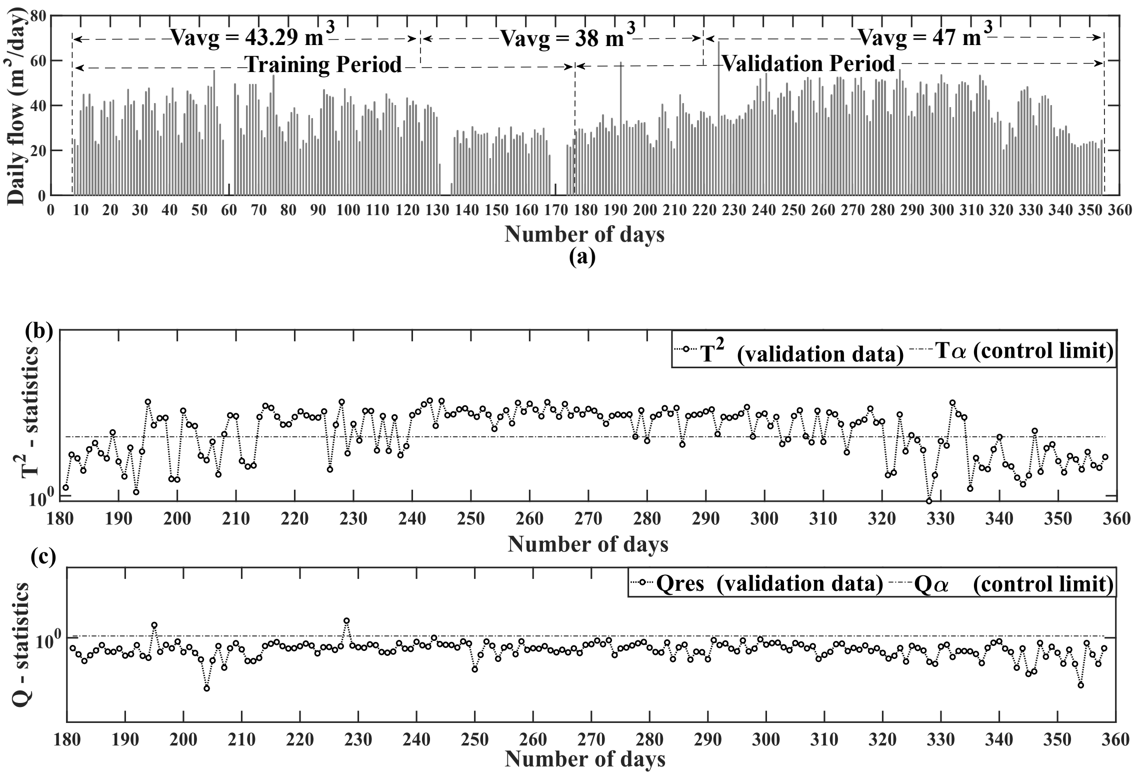


Figure 4.8: (a) Daily usage of total water use in the building, (b & c) The time evolution of T^2-statistics and Q-statistics on a semi-logarithmic scale for the testing data (PCA model validation).

As can be seen from the plots, the T^2-statistics for the Semester 1 data breach the α-control limit T_α^2 for a large portion of the monitoring period, resulting in a significant number of system alarms. As many of these system alarms occurred on consecutive days, fault alarms were triggered in the PCA FDD framework in accordance with the false alarm reduction step (Section 3.2.3). While the T^2-statistics were effective in detecting the water distribution system faults (**Figure 4.8b**), the Q-statistic did not pick up the fault in question with few Q values breeching Q_α. This is due to the fact that the T^2-statistics is more suited to detecting faults relating to higher flows, while the Q-statistics is targeted towards detecting lower flow reading faults. To gain further insight into the PCA model performance, the results were compared to those obtained using the univariate method in (Clifford et al., 2018). The PCA model outperformed the univariate model over the validation period, detecting 25% more system alarms and triggering four more fault alarms. When these additional system alarms were checked back against the historical records for the building, they correctly corresponded to known faults or known non-routine water uses. Comparisons between the performance of the univariate FDD approach and the PCA and SVM FDD approach proposed herein is explored in more detail in Section 4.4.4. of this chapter.

4.4.3 Evaluation

Having trained and validated the fault detection model, its performance was evaluated by considering data for the year following the validation period. As can be seen in **Figure 4.9**, a large number of data points (272 data points) exceeded the α-control limit and were thus labelled as system alarms (i.e., non-routine water uses or due to faults in the system). Of these 272 system alarms, sixteen sets of fault alarms were raised (dashed line boxes in **Figure 4.9**). As per the Section 3.2.3, these fault alarms occur when system alarms (caused by values above the α-control limit) persisted for two consecutive days or more. Several non-routine water uses were also observed over the 1-year period (whereby there was a single 1-day above α-control limit). These high usage peaks raised system alarms but did not trigger a fault alarm in this FDD methodology.

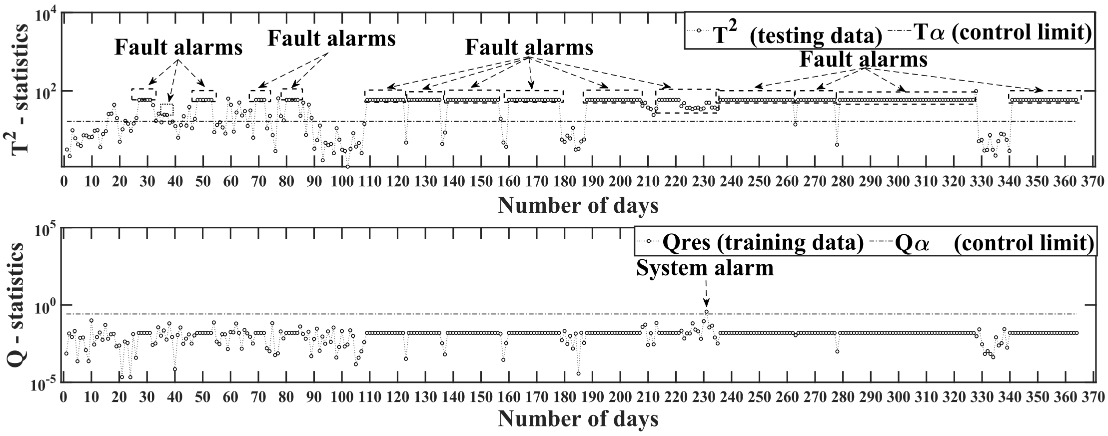


Figure 4.9: The time evolution of T^2-statistics and Q-statistics on a semi-logarithmic scale for the new testing data (PCA fault detection).

Having detected the system alarms using the PCA approach, the model then used SVM to classify each system alarm. The SVM classifier checked system alarms to ensure they were not in fact routine flow events (i.e., some events which are close to the α-control limit may be classified as routine flow). The remaining system alarms were classified as 1) non-routine water uses (non-consecutive system alarms), 2 metering error, or 3) excess usage indicatives of fault. This is shown in **Table 6**. It is noted that, in this case, the SVM could not be trained to classify continuous flow events for this case study as no such events occurred during the training data period.

Table 6: System alarm classification labels.

S.No.	Labels	Description
1	Non-routine water uses	Unusual water uses in working or non-working hours
2	Metering error	Error due to equipment malfunctioning / zero water use during the time period
3	Excess usage	Exceptional water uses during the time period indicative of water distribution system fault

Importantly, the SVM classification identified fifteen of the sixteen fault alarm sets (sequential α-control limit breeches). Of sixteen fault alarm sets detected, one fault alarm set was due to excess usage, this was due to a fault in the rainwater-harvesting system which had been previously undetected. This significant fault resulted in additional mains water usage to maintain a constant supply of cold-water to the roof top tanks for the grey-water uses. Fifteen were due to metering errors (the previous fault identified with the rainwater-harvesting system was explored as a possible issue but was found to have been resolved). Metering error faults were verified by manual interrogation of data sets to ensure the correct detection of metering error faults that had occurred. On further investigation the “metering error” was in fact found to be the result of flow data not being logged and thus there was no data for these periods (which in this approach are equivalent to “zero” flow days). For days with system alarms these were attributed to workshops, infrequent laboratory events, conferences, etc. that occurred on the given days. In retrospect it was not possible to identify the exact nature of the error.

These results are reflective of situations where such data sets can have a significant portion of missing data for a range of reasons including malfunctioning of monitoring equipment (the replacement of which is seldom a short-term investment priority), lapses in monthly or annual contracts between building owner and third-party data monitoring companies, etc. It is noted that if FastDetect was in place over the evaluation data period the faulty meter would have been identified and subsequently replaced.

Figure 4.10 represents the performance of the classifier across the three system alarm classifications via a confusion matrix. Confusion matrices (aka contingency tables) reveal how the classifier mislabels (or confuses) system alarms and can be used to summarise the performance of a classifier with respect to test data (Vitter and Webber 2018). In this matrix, the main diagonal elements represent how often a certain system alarm label is classified correctly (green), while the other boxes show the classification results for misclassification. It is noted that the correct classification of the data for this comparison was determined from time consuming manual interrogation of the historical test data for all system alarms over the monitoring period. In general, **Figure 4.10** shows that the classifier performs well given the relatively short training dataset of six months. The extent of error in classification is a function of a) the difficulty in identification of a particular class, and b) the small sample sizes of the different system alarms present in the training data. For instance, the training data contained only three metering errors. The classifier can be easily trained for this error however, as metering error readings nearly always show zero flow or a series of days with the same flow (which is generally highly unlikely). This results in 100% correct classification for metering error as shown in

Figure 4.10. The training data also contained three non-routine water use events; however, this data set was only sufficient to ensure 50% classification accuracy for non-routine water uses. This is due to the fact that many non-routine water use events are borderline flow events, which are close to the T^2 and Q α-control limits and are thus highly susceptible to misclassification. For excess usage, which is perhaps the most important classification as it is indicative of a fault in the water distribution system, the classifier was 67% accurate. This success rate was obtained with only two excess usage data events in the training data. This is partially due to the fact that excess usage events tend to have easily distinguishable characteristics such as high flows and are thus easier to distinguish from routine operation than say non-routine water uses. While overall classifier performance was good, the analysis clearly indicated that a greater number of system alarms in the training data would facilitate improved classification going forward. Thus, from a practical viewpoint, a classifier can be continually improved during the performance monitoring period, through continued updating of the system alarm training dataset. It should be noted that missed alarms were included in the fault diagnosis process. During the fault detection process, fewer missed detected events were observed which were classified as non-routine water uses in the fault diagnosis process. These system alarms did not persist over two consecutive days as verified by manual interrogation of the historical data. Moreover, the metering error and excess usage are considered as real faults where excess flows were not considered faults as it is indicative of high flow on any given day.

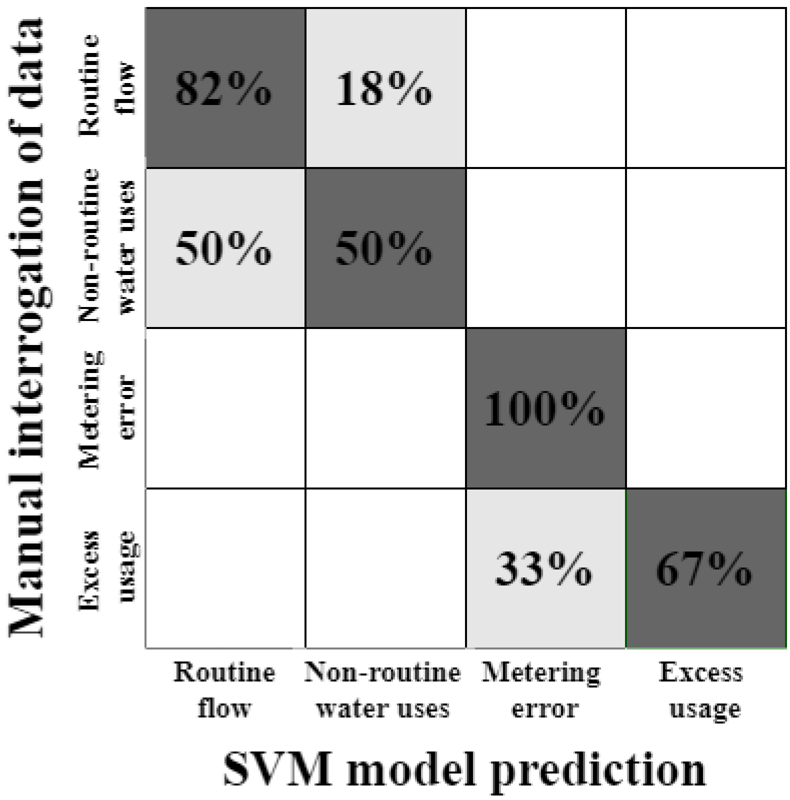


Figure 4.10: Confusion matrix of an actual and predicted system alarm labels.

4.4.4 Effectiveness of fault detection and diagnosis

In order to gain further insight into the performance of FastDetect, the results over the evaluation period were compared to those obtained from the univariate approach which have been recently reported widely in water sector (Gois et al., 2015; Mulligan et al., 2020; Sousa et al., 2019) as presented in Section 2.4.2 and Section 2.5.1 – Chapter 2. The results were further verified by manual interrogation of data set and through discussion with technical staff to investigate if the correct detection of faults had occurred. For any new non-residential building, a validation process will be required to ensure good performance of FastDetect and appropriate accuracy in identifying faults. The findings from this comparison are presented graphically in **Figure 4.11a**. As discussed, 272 system alarms were detected by FastDetect. As shown in **Figure 4.11**, 248 of these mapped directly onto the univariate model results, meaning the PCA approach detected an extra 24 system alarms over the univariate method. Manual interrogation of the historical data revealed that of these 24 additional system alarms, 13 corresponded to non-routine water uses, while 11 constituted false system alarms (i.e., examination of the historical data indicated the water distribution system in the building was operating correctly at the time of the system alarm). As shown in **Figure 4.11b** however, the performance of FastDetect was further enhanced by the SVM component of the model. The SVM classification resulted in 11 of the false system alarms being re-classified as routine flow. Consequently, overall, the method proposed herein identified all the univariate model system alarms, detected 13 additional system alarms successfully, and revealed 3 were found to be false alarms. FastDetect also facilitated accurate classification of the system alarms (over 90% accuracy), providing a greater level of information to the building manager. Again, it is noted that continuous flow alarms could not be classified due to a lack of these faults in the training data, and while FastDetect has been shown to have advantages over the standard univariate approach, its performance would be even further enhanced through provision of larger training data sets.

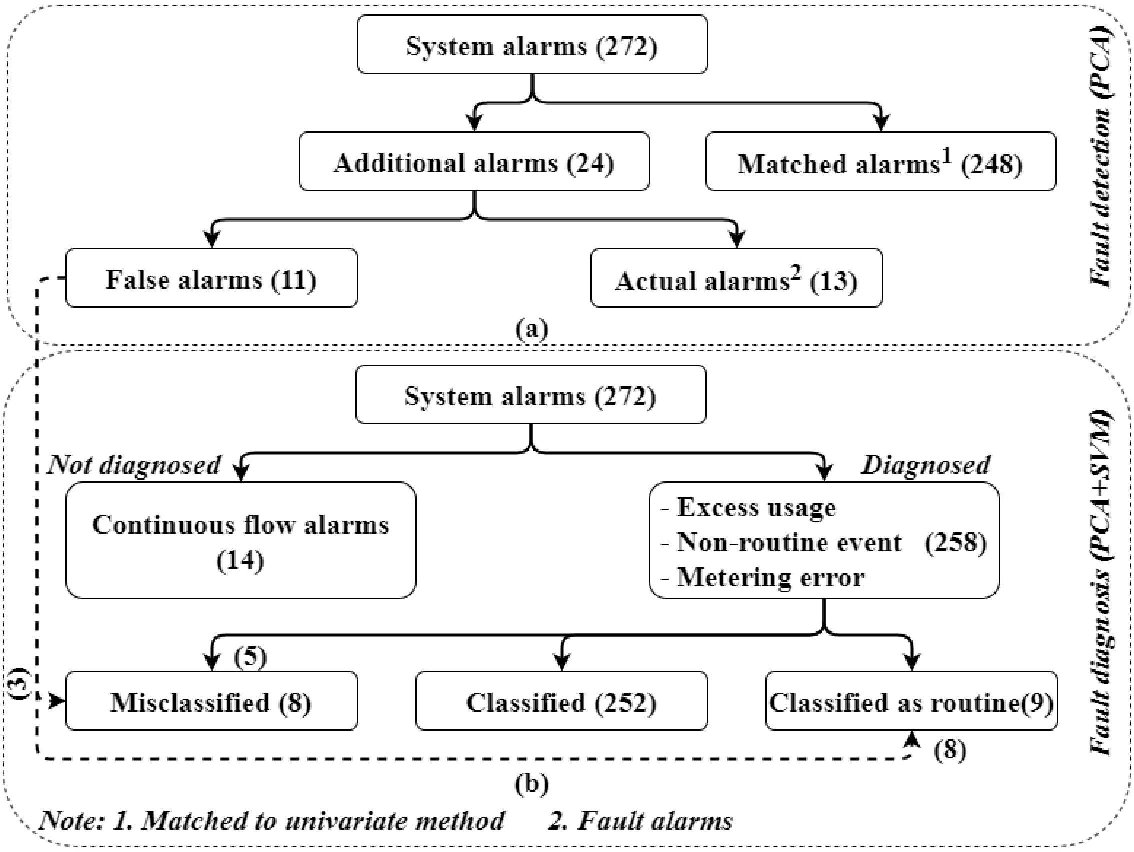


Figure 4.11: Description of (a) system alarm detection results, and (b) system alarm diagnosis results.

As mentioned earlier, FastDetect detected additional alarms (9% additional system alarms) when compared to the univariate approach, out of which 5% were true system alarms and 4% were false alarms (verified by manual interrogation of the data set). As can be seen from **Figure 11**, 91% of FastDetect detected alarms were mapped directly onto the univariate approach results (99% univariate approach results were matched with FastDetect results). However, FastDetect resulted in comparatively more false alarms (4%) than univariate approach but outperformed in detecting true low-flow alarms which the univariate approach failed in. These low-flow system alarms could cause imperceptible faults in the building water distribution system, compromising the system's efficiency. Hence, there is a trade-off associated with the number of false alarms that can be tolerated by the building manager and the accuracy of detecting true system alarms.

A significant advantage of FastDetect over univariate approaches is that these approaches tend to detect faults that have distinguishable characteristics such as high flows which are easier to identify from routine water uses. This could lead to faults, resulting from low flow readings, remaining undetected. FastDetect showed excellent capability in detecting faults with distinct characteristics such as high flow fault, low-level imperceptible faults and equipment malfunctioning faults. Moreover, FastDetect can identify the root cause of the detected faults which a univariate approach cannot.

4.5 Conclusion

PCA together with a multi-class classifier SVM were found to be useful in detecting and diagnosing faults in building water networks. Classical and advanced approaches have resolved many FDD problems in water networks, but when it comes to building level multivariate systems or special cases (for instance substantial variation in the water consumption data), these approaches are not entirely effective. This study investigated a FastDetect for non-residential building water distribution system which combines the detection indices (T^2 and Q-statistics) and multi-class SVM. Hotelling T^2-statistics and Q-statistics were used to detect system alarms in the incoming data and the latter multi-class SVM along with error-correcting output code was trained for system alarm classification. The results demonstrated promising capabilities of FastDetect. When compared to the standard univariate approach, a greater number of system alarms were detected and found to have occurred. The multi-class SVM also allowed these system alarms to be classified, providing a greater level of information to building managers, which may avoid unnecessary emergency shutdown in industrial applications. Thus, the comparative study has shown that FastDetect performs better than standard univariate approach. While FastDetect has good capability in detecting and diagnose faults in complex non-residential water distribution systems.

5. FALSE ALARM MODERATION FOR PERFORMANCE MONITORING IN INDUSTRIAL WATER DISTRIBUTION SYSTEMS

5 False Alarms Moderation for Performance Monitoring in Industrial Water Distribution Systems

5.1 Overview

The aim of this chapter was to address the issue of high prevalence of false alarms during fault detection process leading to low industry uptake of FDD systems developed, or where in place, alarms can be ignored. To efficiently detect and diagnose water distribution system faults, false alarms have to be controlled through false alarm moderation approaches so that building managers/operators only need to focus on critical system alarms or system alarms with high risk levels. In this part of work two statistical false alarm moderation approaches (window-based, and trial-based) were presented that generate a second control limit for monitoring statistics (T^2-statistics and Q-statistics). The implementation of these false alarm moderation approaches was combined with PCA and SVM to detect and diagnose faults with real water time series data from two non-residential case-study sites.

The study has been submitted in Advanced Engineering Informatics (Hashim, H., Clifford, E., Paraic, R., 2021) This chapter is arranged as follows. Section 5.2 provides an overview of false alarm moderation approaches used in the context of performance monitoring followed by the details of two case-study sites and their pumping systems are outlined in Section 5.3. In Section 5.4, the false alarm moderation approaches were integrated into a PCA-based fault detection discussed in detail. The methodology applies the PCA technique combined with false alarm moderation approaches on two case-study sites. The results which showed improved performance of PCA-based fault detection through the application of false alarm moderation are presented in Section 5.5. Section 5.6 summarises the study and provides conclusions.

5.2 Introduction

Modern industrial facilities require large-scale and complex water distribution systems which in turn experience faults that produce undesirable water consumption within a facility that can often go undetected (Hashim et al., 2020). A relatively high prevalence of false alarms will reduce confidence in any fault detection and diagnosis system and cause unnecessary downtime and cost. Thus, can cause hesitation in building managers when responding to true system alarms (faults and non-routine water usage events) during performance monitoring of industrial water distribution systems (**Figure 5.1**). This study seeks to address this important gap in the literature through application of two false alarm moderation approaches to the PCA fault detection methods for the first time.

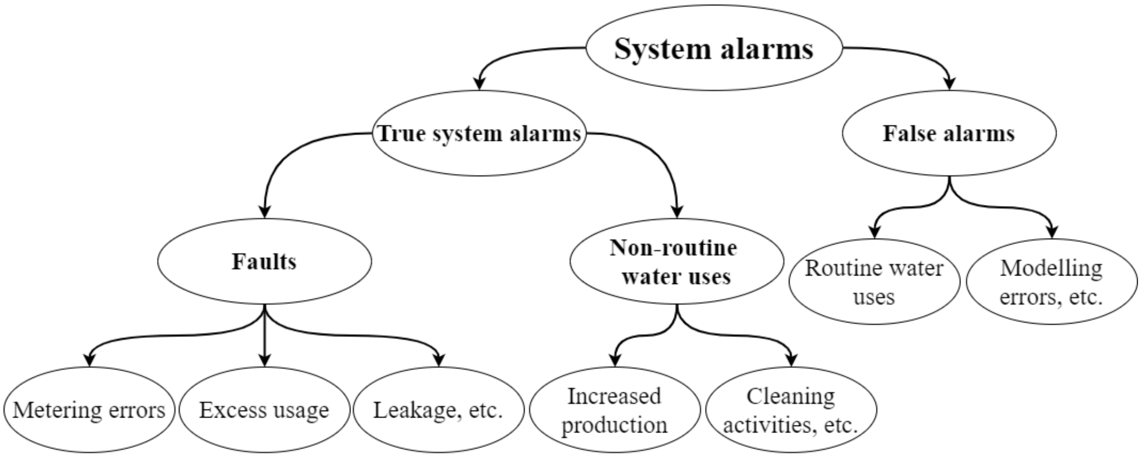


Figure 5.1: System alarm characterization.

Having established a PCA-based fault detection approach for industrial buildings (Chapter 4 & Hashim et al., (2020). This paper builds on this work through utilization of false alarm moderation techniques using real sites data. The two false alarm moderation approaches used in this study are the window-based, and the trial-based methods. These are non-parametric statistical approaches that apply false alarm moderation for monitoring statistics (T^2-statistics and Q-statistics) in the developed PCA model. These approaches do not require additional information from the training data and can be integrated into the developed PCA model without any computational and labelling cost (making them ideally suited to non-residential water usage applications). For the trial-based approach, probabilities of the monitoring statistics are analysed based on independent trials. This approach was originally proposed for binary integration in radar and electronic warfare problems to detect contiguous or successive signals (Bell, 1996; Gilliam et al., 2012). The work herein represents the first application of the approach to performance monitoring. For the window-based approach, false alarms are controlled by employing an "observation window". The window-based approach originates within time-series outlier detection research developed to isolate the occurrence of randomly induced outliers resulting from systematic process variation (Abraham and Chuang, 1993; Chen et al., 2008; Singhal and Seborg, 2000). It was initially proposed by (Li et al., 2018; Chen, 2010) in performance monitoring of modelled chemical processes, but has not been used with real site data to date. Existing false alarm moderation approaches for water distribution system in the literature utilise generated or trial data sets. The use of idealised data means does not facilitate examination of the effect of random variations and outliers, which undoubtedly occur in industry. This can lead to an over-estimation of the effectiveness of false alarm moderation approaches when testing against generated data (Blázquez-García et al., 2021; Hu et al., 2021; Jung et al., 2015). However, in this study the proposed approaches are assessed using real case-study site data (described in detail in Section 5.4.2) for the first time in the literature, with associated outliers and process noise. This leads to a more robust assessment of the false alarm moderation technique's ability to reduce false alarms in water distribution systems. The dataset used in this study is directly from operational industrial water distribution systems, facilitating achievement of the aim of developing and evaluating false alarm moderation approaches within the practical data constraints frequently encountered in industry. Evaluation metrics were defined to study the effects of these approaches on the PCA model performance. This study proposes and demonstrates that with the application of the false alarm moderation approaches, false alarms can be controlled greatly, enhancing the fault detection and diagnosis performance of the PCA model. Overall, the goal of the proposed false alarm approaches was to obtain a simpler integrable method to control false alarm occurrences during performance monitoring without the use of labelled "false alarms" data points.

5.3 Case-study sites data

The case-study 1 industrial water distribution system (within an Irish food and drinks company) supplied water to production and distribution facilities, offices, cleaning facilities, a canteen and toilet facilities spread across 9,300 m^2 of floor space (**Figure 5.2**). The case-study site is one of the most modern and highly automated production facilities in Ireland. The company is located near the west coast of Ireland in an area with a temperate marine climate. The company operates six days a week (closed Sundays). The mains water supply (public supply) is currently monitored using a third-party monitoring system that collects data, provides high level usage statistics, and sends notifications when water usage exceeds a predefined threshold based on a univariate approach. Some of the key water uses include transport vehicle washing, scheduled cleaning activities, toilets and urinals, and potable water for the water fountains and the canteen. The case-study site is managed by a third-party monitoring system which is based on a conventional threshold system which triggers alarms when relatively high levels of water are being used.

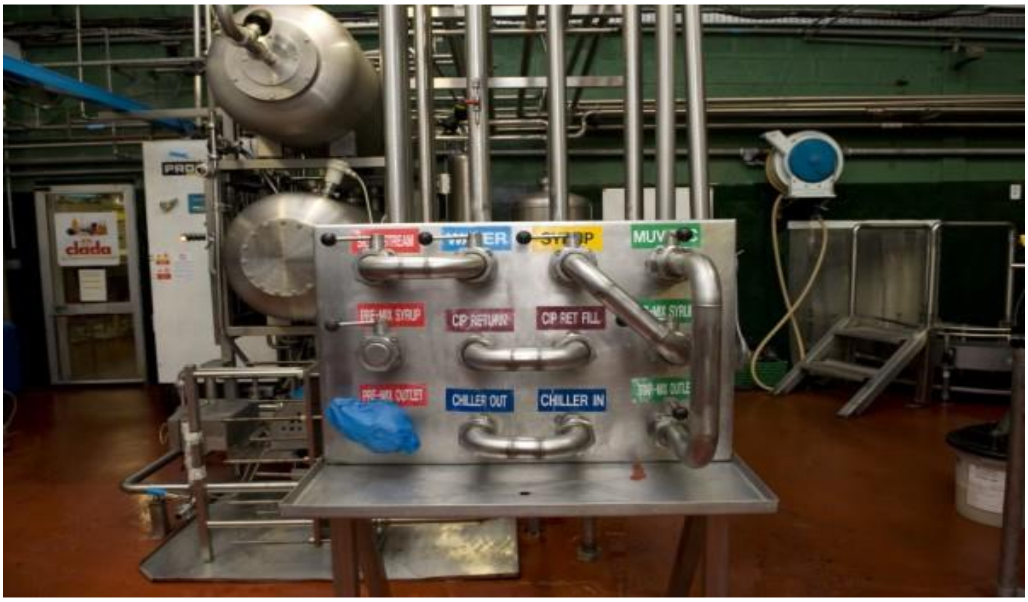


Figure 5.2: production facilities - within an Irish food and drinks company.

The case-study 2 water distribution system was the within Alice Perry Engineering Building at the National University of Ireland, Galway (NUI Galway) (**Figure 4.2** – Chapter 4). The building is located on the west coast of Ireland in an area with a temperate marine climate. The building is one of the most state-of-the-art university buildings in Ireland and has been operational since 2010. The building includes lecture halls, classrooms, offices, laboratory facilities, a café, and shower and toilet facilities spread across 14,000 m^2 of floor space on four storeys. Some of the key water uses include showers and hand wash basins, grey water from a rainwater-harvesting system for toilets and urinals and potable water for the water fountains and the café. The building is managed through a building management system that collects data on building performance and operational efficiency - including 11 water meters and several electricity meters. This study deals primarily with the overall mains water usage within the case-study sites.

To initialize the false alarm moderation process for this study, statistical features from the mains meter readings of two case-study sites were derived to describe baseline water usage and diurnal patterns (or indeed patterns at any chosen time interval for which adequate data was available). The daily flow and maximum flow features summarize water consumption over a 24-hour period from midnight to midnight. Whereas the remaining features (working hours, non-working hours, and night time) focus on water consumption for specific periods of the day which are linked to how the case-study sites are used. These periods comprise four working hour time intervals, two non-working hour time intervals and night time interval. For the case-study 1, water flow data recorded at 30-minute intervals, was collected from the on-site third-party monitoring system, and analysed in developing the integrated methodology with false alarm moderation approaches. Case-study 2 applies fault moderation to the application of PCA based fault detection previously published (Hashim et al., 2020). For the case-study 2, medium resolution data recorded at 7.5-minute intervals was collected in the case-study building from an inline displacement water meter equipped with a magnetic pulse output and analysed in developing the integrated methodology with false alarm moderation approaches.

5.4 Methodology

In this study, PCA and detection indices namely, Hotelling T^2-statistics and Q-statistics were applied to extract the useful information from the dataset by identifying the data noise and thus distinguishing between the features of routine, non-routine, and faulty operation.

5.4.1 Principal component analysis and detection indices

PCA is used to extract the correlation between variables in sets of independent variables, that explain the trend of the process while optimizing the variance of the original data in a reduced number of dimensions. PCA determines the orthogonal vectors which are defined by a linear combination of the original features ordered by an amount of explained variance in component direction known as principal components. Moreover, to detect the anomaly of the new incoming data and to measure the variation of the samples within the PCA model detection indices (namely, Hotelling T^2-statistics and Q-statistics) are utilized in this study. A more detailed explanation of PCA and detection indices can be found in (Hashim et al., 2020).

5.4.2 False alarm moderation approaches

5.4.2.1 Window-based approach

In this false alarm moderation approach, when the number of system alarms s (data points beyond α-control limit) in a particular observation window v of length w is greater than a second control limit k, a fault is confirmed (not a false alarm). In the ongoing analysis, the monitoring statistics (T^2 or Q-statistics) are referred as M_i, where i is the i^{th} value (i = 1, 2, 3…n). A "sliding" observation window v (moving linearly at a fixed interval) is considered along the full length of the monitoring statistics such that $w < n$. For instance, if window "1" = v_1 and has a window length "w" then v_1 starts at i = 1 and encompasses data points from x_1 to x_w. For window "2" = v_2, the observation window then slides forward by one position to start at i +1 (v_2 starts at i = 2) and encompasses data points from x_2 to x_{w+1} and the process continues till all M_i are checked in probable observation windows. The outline of the window-based approach is shown in **Figure 5.3**.

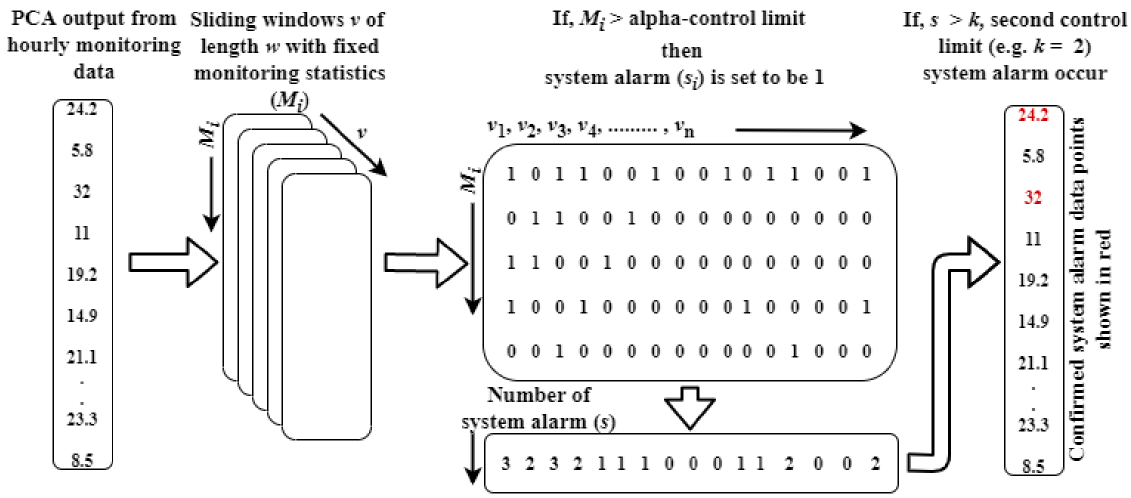


Figure 5.3: Outline of window-based approach.

According to probability theory (Papoulis, 1980), if M_i are independent and identically distributed, then M_i can be approximately regarded as obeying Bernoulli distribution. Thus, the sum s is a binomial random variable which is the summation of s_i which represents M_i beyond the α-control limit with a probability density distribution which can be expressed as (Chen, 2010; Li et al., 2018; Smith & Aretxabaleta, 2007; Barczy et al., 2010).

$$p(s;\ \alpha, w) = \binom{w}{s} \alpha^{w-s}(1-\alpha)^{s} \tag{24}$$

A key parameter of the window-based approach is the length of the observation window which is selected at the training data stage. Where possible the selection of window length w should enable as many windows as possible to have at least one system alarm. However, if the selection of a suitable window length w is based on the frequency of system alarm occurrence. If the window selected is too large there will be increased likelihood of false alarm occurrence. For instance, a large observation window could contain a number of false alarms, thus increasing the potential for total system alarms s to be greater than the second control limit k ($s \geq k$), triggering a true system alarm. Thus, too large a window length could result in the FDD system incorrectly reporting true system alarms, reducing the system's reliability and performance. However, if the window length selected is too small there may be an insufficient number of time dependent monitoring statistics within the window to facilitate detection of true system alarms in practice (s cannot rise above k) (Li et al., 2018; Chen, 2010). Given the challenges associated with window length selection, extensive trials and subsequent analysis was conducted in the training data stage of this study to identify the most appropriate window length.

For a given data point x_i, the second control limit k for an observation window v can be computed by a cumulative distribution function (Abraham and Chuang, 1993; Chen et al., 2008; Chen, 2010).

$$F(k;\ \alpha, w) = \sum_{i=1}^{s} \binom{w}{i}\ \alpha^{w-i}\ (1-\alpha)^{i}\ < \beta \tag{25}$$

where β is the confidence level for a second control limit k. In accordance with the multivariate fault detection in process industries, β is usually set between 0.9 and 0.99 (Chen, 2010; Li et al., 2018). Therefore, it can be concluded that if the system alarm number is within the second control limit k in the observation window v, the time dependent monitoring statistics can be assumed to be representative of the routine system operating condition (not resulting from a fault or non-routine water uses). It is noted that the steps above for determination for the control limit k and window length w were completed in the training data stage and are thus not required for the implementation phase. In this approach, the number of system alarms s in each observation window v was compared to the second control limit k to detect true system alarm. Overall, the application of the window-based approach can be summarised as follows.

1. Calculate monitoring statistics M_i with n observations from the training dataset.
2. Selection of a suitable observation window length w.
3. Compute a second control limit k utilising a cumulative distribution function.
4. Generate sliding observation windows.
5. If M_i is beyond the α-control limit at a given timestep i, then the corresponding system alarm s_i is set to be 1 for that time step, otherwise s_i remains zero.
6. All time dependent monitoring statistics M_i in the current observation window are examined using the approach Step 5.
7. s is then calculated as $s = \Sigma\, s_i$ for the observation window.
8. s is then compared to the second control limit k. If $s \leq k$, then the starting data point in the observation window be considered to be representative of routine water usage.
9. The window then slides forward by one position and the process starts again from Step 5.

5.4.2.2 Trial-based approach

In this approach, false alarm probabilities were analysed based on T independent trials. Under the prior assumption of monitoring statistics can be regarded as obeying Bernoulli distribution (Section 5.3.2.1), false alarm detection in each individual observation is an independent Bernoulli trial with probability θ. This can be referred to as the ratio of number of successes to the total number of trials. The outline of the trial-based approach is shown in **Figure 5.4** and a detailed description of this can be read in (Papoulis, 1980; Bell, 1996).

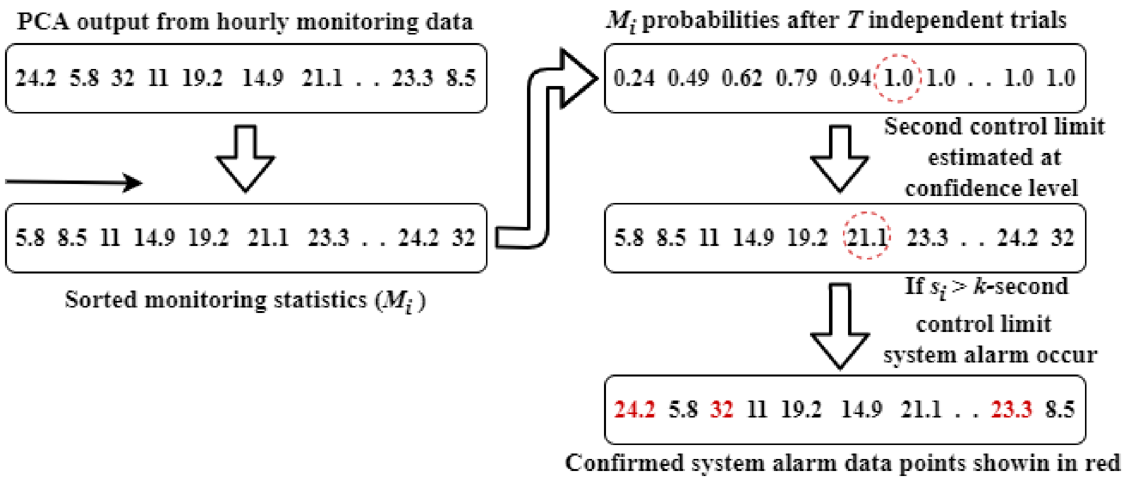


Figure 5.4: Outline of trial-based approach.

The false alarm probability can be determined by a sufficient number of trials based on the confidence level often defined as the binomial probability mass function or the binomial discrete density function. The number of successes in T independent trials conforms to a binomial distribution, if in each trial either success or failure stage is achieved with a predefined confidence level. The minimum number of trials T can be determined as (Gilliam et al., 2012).

$$T = \frac{\log(1-\alpha)}{\log(1-\theta)} \tag{26}$$

where α is the confidence level and θ is the estimated false alarm probability of the system alarms. According to binomial binary integration rule, at a specified confidence level the sequence of probabilities of an individual observation can be interpreted as the binomial cumulative distribution function (Papoulis, 1980).

$$F(s_i;\ \theta, T) = \sum_{i=1}^{s} p(s_i;\ \theta, T) \tag{27}$$

where s_i represents the system alarm in T independent trials with $(0 \leq \theta \leq 1)$ given that the probability for success is predefined. By analysing the minimum number of trials adequate results can be obtained at a specified false alarm probability and at a desired confidence level (Gilliam et al., 2012). The false alarm probabilities for s_i events can be attained by computing cumulative the distribution function of the monitoring statistics. Thus, the second control limit k can be determined at the maximum value of the computed false alarm probabilities, such that $k \geq 1$. In this approach, the individual system alarm s_i was compared to the second control limit k to detect true system alarm. The trial-based approach can be summarized as follows.

1. Calculate monitoring statistics M_i with n observations from the training dataset.
2. Analyse the system alarms s_i (data points beyond α-control limit).
3. Estimate the false alarm probability θ from the system alarms.
4. Determine the number of independent trials T with a predefined confidence level.
5. Compute false alarm probabilities $p(s_i)$ with a predefined confidence level.
6. Calculate the second control limit k at a false alarm probability, such that $k \geq 1$.
7. s_i is then compared to the second control limit k. If $s_i \leq k$, then the data point can be considered to be representative of routine water usage.

5.4.2.3 Evaluation indices

The performance of the proposed false alarm moderation approaches was evaluated in terms of (1) the false alarm rate (P_{FAR}) during the fault detection process and (2) misdetection rate (P_{MDR}) of system alarms of the PCA model. The false alarm rate (P_{FAR}) – for monitoring statistics in the PCA model can be expressed as (Li et al., 2018).

$$P_{FAR} = \frac{N_{false.alarm}}{N_{normal}} \tag{28}$$

where $N_{false.alarm}$ is the number of data points falsely identified as system alarms during the fault detection process and N_{normal} is the number of data points resulting from routine operating conditions (within the α-control limit). Lower values of P_{FAR} demonstrate that a greater number of false alarms are controlled by the false alarm moderation approach indicating better performance of the proposed approach.

Misdetection rats has previously been used to compare the fault detection performance of various detection methods (Gajjar et al., 2018; Qin and Chiang, 2019; Zhao et al., 2019; Bakdi and Kouadri, 2017). The misdetection rate can be expressed as the ratio of system alarms that are not detected as system alarms (identified as normal - $N_{mis.alarm}$) to the total system alarms $N_{sys.alarm}$. It can be expressed as (Bakdi and Kouadri, 2017).

$$P_{MDR} = \frac{N_{mis.alarm}}{N_{sys.alarm}} \tag{29}$$

Lower values of P_{MDR} indicate that fewer system alarms go undetected during the fault detection process. The objective of an industrially reliable fault detection and diagnosis approach is to achieve the lowest P_{FAR} and P_{MDR} values (Zhao et al., 2019; Qin and Chiang, 2019). While limiting the prevalence of false alarms during fault detection and diagnosis processes.

The F1-score has previously been utilised to evaluate the machine learning-based model performance in different sectors such as construction, cybersecurity, neuroscience, etc. (Zhao and Obonyo, 2021; Zhao and Obonyo, 2020; Jung and Lee, 2019). The F1-score is defined as the harmonic mean of precision which is given by $N_{sys.alarm}/\left(N_{sys.alarm} + N_{false.alarm}\right)$ and recall which is given by $N_{sys.alarm}/\left(N_{sys.alarm} + N_{mis.alarm}\right)$ (Zhang et al., 2021). It can be expressed as (Zhao et al., 2021).

$$F_{1-score} = \frac{N_{sys.alarm}}{N_{sys.alarm} + \frac{1}{2}\left(N_{false.alarm} + N_{mis.alarm}\right)} \tag{30}$$

5.4.3 Fault detection

To identify faults within the true system alarms category (see **Figure 5.1**), the time series nature of the data over periods of two days or more was considered. This practical approach, which was also used by (Hashim et al., 2020), recognises the fact that industrial facilities such as the one explored herein are occasionally used for events which result in significant increases in water consumption such as transport vehicle washing, scheduled site cleaning activities, etc. on a given day. The two-day time period was selected based on a high-level water audit of historical data (12-month period prior to the data used in this study) within the facility which demonstrated that events such as vehicle washing, or site cleaning occurred on a single day (sometimes scheduled and sometimes not) but did not occur on consecutive days. Thus, an unusual water pattern might be observed for one day due to vehicle washing but was highly unlikely to be observed for two consecutive days (this is highlighted in **Figure 5.6** - red solid data points). Thus, faults in the proposed framework were raised when true system alarms persisted over two or more consecutive days. The nature of the time series alarm system could vary between industrial sites based on water use event patterns within the facility and can be established by a high-level water audit. The summary of the false alarm moderation is outlined in **Figure 5.5**.

$$\begin{cases} t_{acceptable} < t_{limit} --- \text{Normal} \\ t_{acceptable} \geq t_{limit} --- \text{Indicative of fault} \end{cases} \tag{31}$$

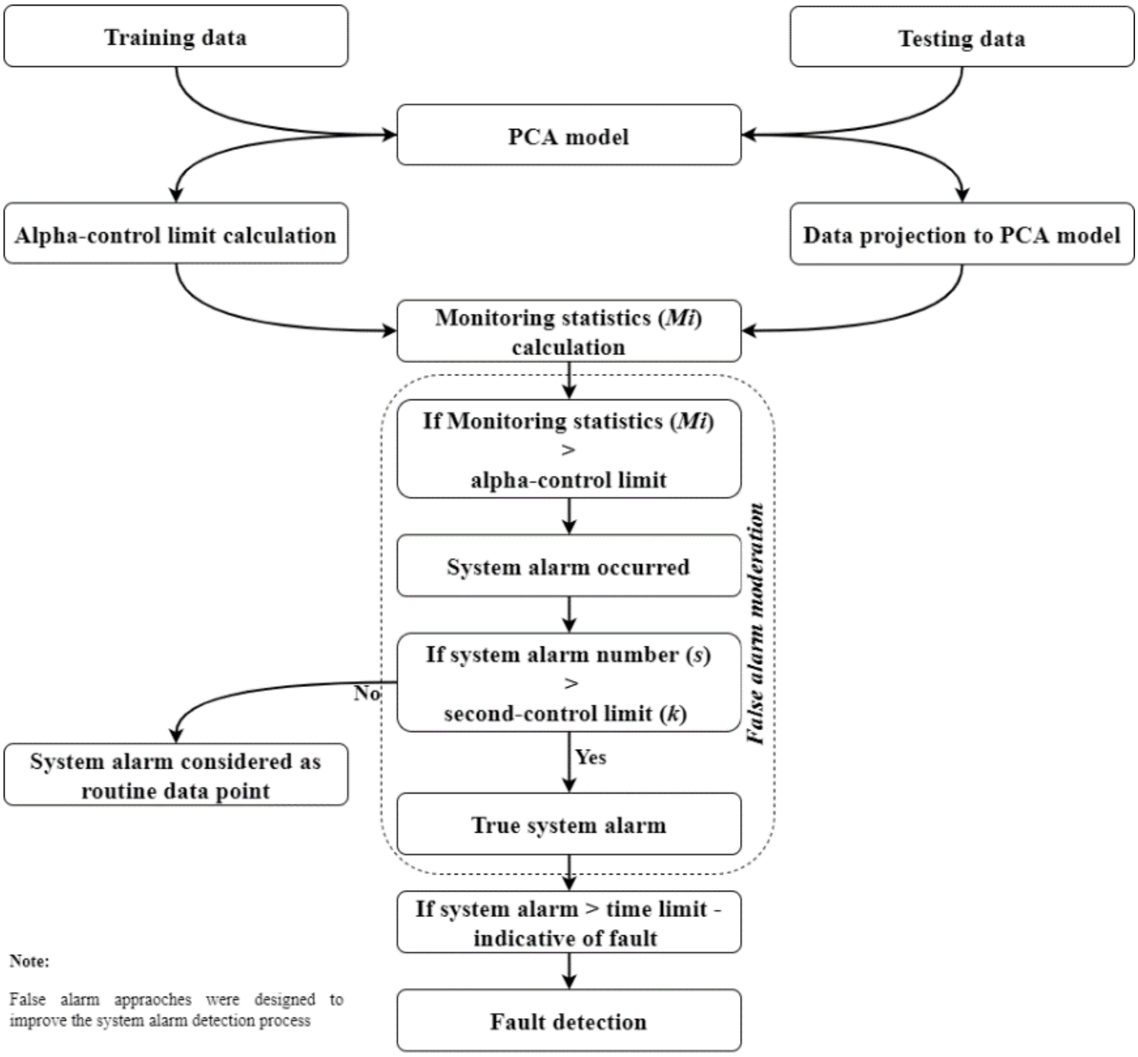


Figure 5.5: Summary of PCA with false alarm moderation approaches.

5.5 Results and discussion

5.5.1 Training and model development – Case-study 1

Water usage data was aggregated into hourly flow traces to analyse the usage characteristics and to facilitate application of the methodology for false alarm moderation at the case-study 1 site (the original data comprised a volume reading every 30 minutes).

In a number of cases data points had overlapping timestamps due to sensor and/or communication errors and these were removed from the training set. The PCA model was trained using nearly six months of data (174 data points). The calculated PCA monitoring statistics (T^2-statistics and Q-statistics) are shown in **Figure 5.6a**. It is noted that the water usage data does not reflect normally distributed data. It can be seen from **Figure 5.6a** that a number of data points lie above the α-control limits, constituting system alarms (12 system alarms constituting 7% of the data points). These were due to a combination of on-site non-routine events (which inevitably form part of the training data) and routine statistical variation related to the inherited definition of the α-control limit, which is linked to false alarm probability (Hashim et al., 2020; Li et al., 2019). While none of these system alarms in the training data would result in a fault in this integrated fault detection and diagnosis framework (even prior to false alarm moderation), as none of the system alarms occur on consecutive days (as discussed in Section 5.4) - they would result in alarms at the case study 1 facility's current third-party monitoring system. To identify the false alarms in the training data, these 12 system alarms were examined manually against the facility records for the training period. Examination indicated that of the 12 system alarms, 8 were false alarms events (events at days 46, 53, 91, 131, 133, 137, 143, 151) at which the water distribution system in the facility was operating normally (checked against the historical record for the site) but alarms were raised (shown in **Figure 5.6a** - black solid data points). The system alarms at days 19, 88, 104, 162 days resulted from non-routine events (cleaning activities at site) which were manually checked against the historical record for the site to ensure they were not in fact routine flow events (shown in **Figure 5.6a** - red solid data points). Thus, only one third of detected alarms were true alarms (170 of 174 data points were routine data points). It should be noted that a manual check of the data, for verification purposes, showed the PCA-based fault detection detected all of the system alarms (those 4 system alarms were the only ones within this dataset).

5.5.1.1 Determination of observation window length w (window-based approach)

A key factor in applying the window-based approach is the choice of the appropriate observation window length *w*, which can be somewhat subjective. In this study, the impact of an observation window length *w* on both the sensitivity to true system alarm detection and the prevalence of false alarms, was studied. Six different observation window lengths were analysed to select a sensible length *w* of an observation window

(**Table 5.1**). For observation window lengths of *w* = 20 and *w* = 25, false alarms were reduced by 50% (4 of 8 false alarms eliminated); for *w* = 30 false alarm occurrence was reduced by 25% (2 out of 8 false alarms eliminated). However, in each of these cases the true system alarms were not reported, and thus these window lengths were not suitable.

For observation windows of length *w* = 5 to *w* = 15 all system alarms were correctly reported. However, there was a significant difference in the reporting of false alarms. For *w* = 5 the number of false alarms increased to 10 (from the original figure of 8) – thus providing no advantage over the based PCA-based method. A marginal reduction of 1 false alarm was observed for *w* = 10 but for *w* = 15 the number of false alarms reduced by 37.5% (from 8 to 5 false alarms). It should be noted that the final number of false alarms reported for each window length is the sum of the original false alarms that are now correctly categorised (as routine data points) and the new false alarms "created". For example, in **Table 5.1**, for *w* = 15, 5 of the original false alarms were correctly categorised through implementation of the window-based approach, but 2 "additional" false alarms arose. In this case, the detection of true system alarms is the key criterion and thereafter the ability of a method to reduce false alarms. The selection of the appropriate observation window length *w* can be further enhanced through provision of larger training data sets – particularly so where a dataset contains more system alarms. The above analysis would be specific to any building/scenario, but the same steps could be followed to establish the appropriate window length regardless of the building.

Table 5.1: Selection of an observation window length w (Case-study 1).

<table>
<tr><th></th><th>Observation window length w</th><th>PCA only</th><th>w=5</th><th>w=10</th><th>w=15</th><th>w=20</th><th>w=25</th><th>w=30</th></tr>
<tr><td></td><td>Data point considered</td><td colspan="7">174</td></tr>
<tr><td></td><td>Number of routine data points</td><td colspan="7">170</td></tr>
<tr><td></td><td>Number of true system alarms*</td><td colspan="7">4</td></tr>
<tr><td>a</td><td>Number of original false alarms</td><td colspan="7">8</td></tr>
<tr><td></td><td>Number of system alarms detected</td><td>4</td><td>4</td><td>4</td><td>4</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>b</td><td>Number of original false alarms reduced</td><td>N/A</td><td>4</td><td>5</td><td>5</td><td>6</td><td>6</td><td>6</td></tr>
<tr><td>c</td><td>False alarms produced by method</td><td>0</td><td>6</td><td>4</td><td>2</td><td>2</td><td>2</td><td>0</td></tr>
<tr><td>(a+c)-b</td><td>Final number of false alarms</td><td>N/A</td><td>10</td><td>7</td><td>5</td><td>4</td><td>4</td><td>2</td></tr>
</table>

*Verified by manual analysis

5.5.1.2 Application of window-based and trial-based approaches.

To reduce the prevalence of false alarms, the second control limit *k* for false alarm moderation was determined, for both the window-based and trial-based methods (see Eq. 25 and 27), using the basic-PCA results (**Figure 5.6a**). False alarm probability was calculated from the false alarms identified in the training data during the PCA modelling process in comparison to those identified subsequently from manual interrogation (8 false alarms from a total 12 system alarms). The specific procedures involved in determining the second control limits based on false alarm probability presented in Section 5.4 were utilised. Having identified the appropriate *k*-values, the false alarm moderation techniques were implemented. The results are shown in **Figure 5.6**, with the output from the basic PCA model presented in **Figure 5.6a**, and the output from the PCA model with the trial-based approach and the window-based approach shown in **Figure 5.6b** and **Figure 5.6c**, respectively. **Table 5.2** provides further insight into the performance of false alarm moderation approaches through presentation of the false alarm rate - P_{FAR}. As can be seen from (**Table 5.2**), the application of PCA with trial-based approach resulted in a 16.7% reduction in false alarms. The PCA with window-based approach resulted in a false alarm reduction of reduction of 37.5%.

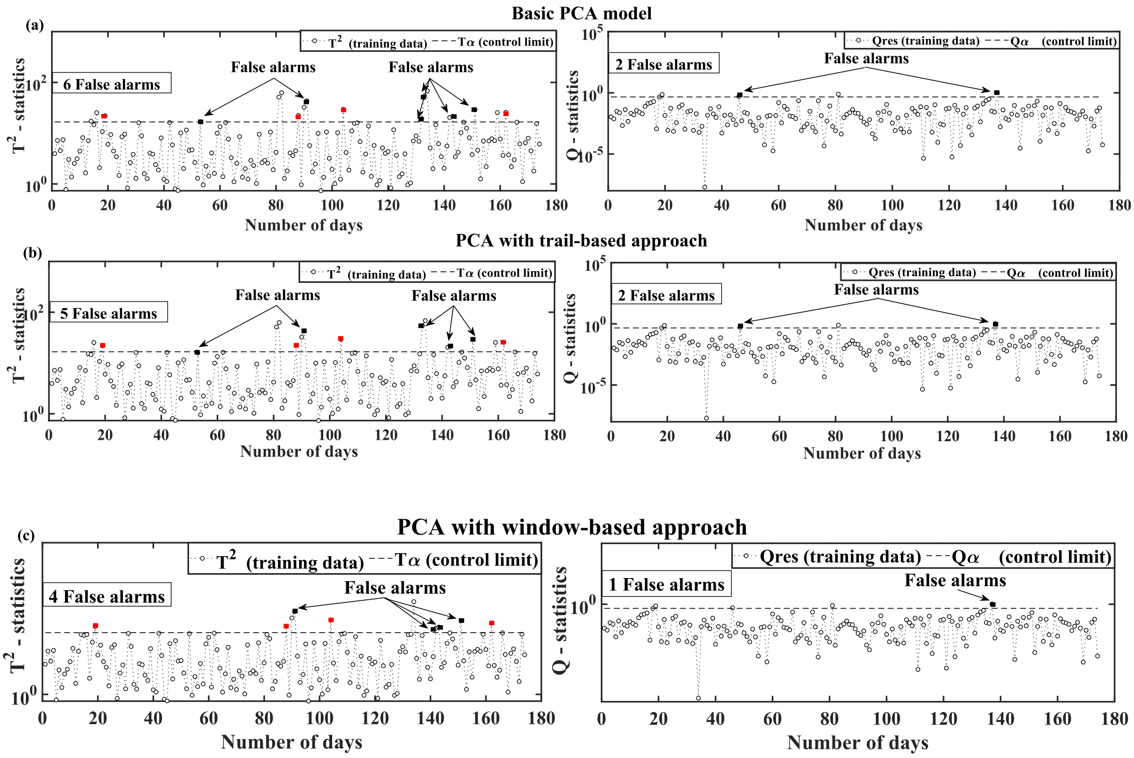


Figure 5.6: (a) The time evolution of T^2-statistics and Q-statistics in basic PCA, (b) PCA with trial-based approach, and (c) PCA with window-based approach for the training period.

Table 5.2: False alarm rate for PCA with false alarm moderation (training period).

	Basic PCA		PCA with trial-based	PCA with window-based
	System alarms	*False alarms*	*False alarms*	*False alarms*
P_{FAR}	12	8	16.7% less (i.e., 7 false alarms)	37.5% less (i.e., 5 false alarms)

5.5.2 Testing period: effectiveness of false alarm moderation

Having trained the basic PCA model and integrated the false alarm moderation approaches, the performance of the window-based approach and trial-based approach were compared to the basic PCA model using 12 months of data (date from the 12-month period following the training data period). As can be seen in **Figure 5.7**, a significant number of data points (181 data points) exceeded the α-control limit and resulted in system alarms (non-routine events or faults in the system for the basic PCA model). It is noted here that the majority of system alarms over the monitoring period resulted from a recurring fault at the case-study site. This will be discussed and analysed in detail in Section 5.5.3. The focus of this sub-section is however to compare the performance of the basic PCA model to the PCA model with false alarm moderation. To do this it was necessary to manually check the 181 system alarms detected using the basic PCA model to identify which of these were false alarms. In total 13 of the 181 system alarms were found to be false alarms (indicated in **Figure 5.7a** - solid data points). These false alarms were generally due to PCA modelling inaccuracies or due to borderline flow fluctuation (flow events close to the α-control limits, and thus more likely to present as a false alarm). Having obtained this information, the effectiveness of the PCA model with false alarm moderation approaches was represented graphically, as shown in **Figure 5.7b** and **Figure 5.7c**. It can be seen from the plot that both the trial-based approach and the window-based approach resulted in the reduction of false alarm prevalence.

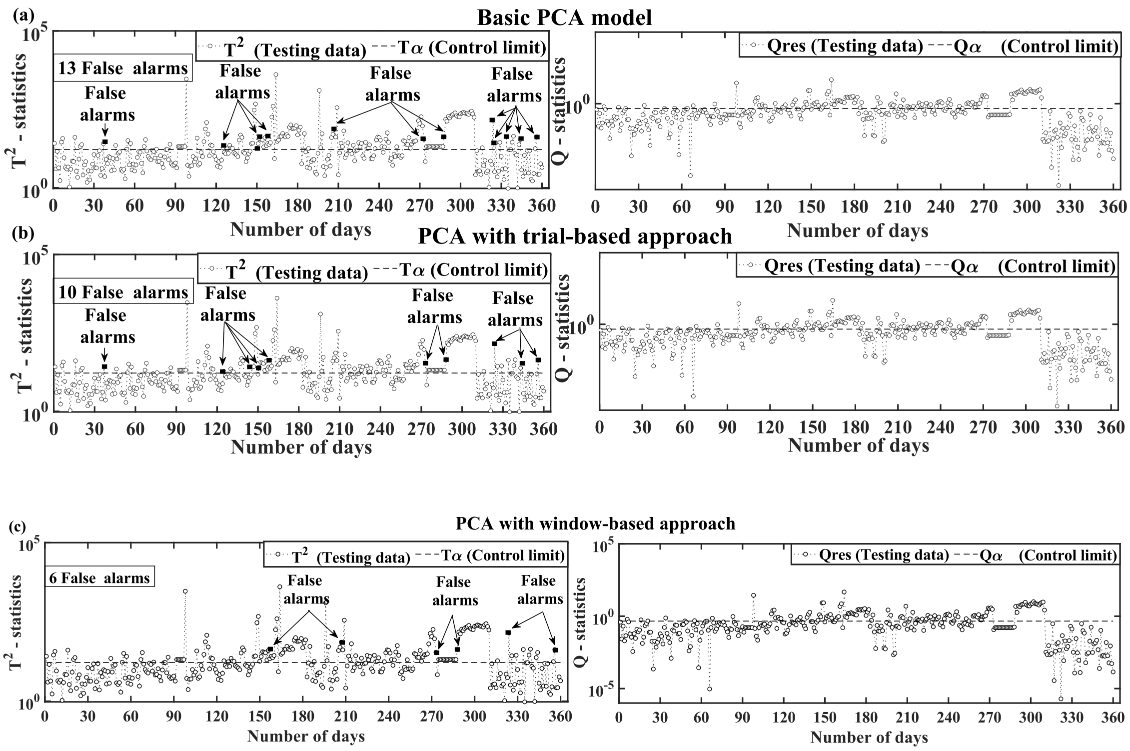


Figure 5.7: (a) The time evolution of T^2-statistics and Q-statistics in basic PCA, (b) PCA with trial-based approach, and (c) PCA with window-based approach for the testing period.

Table 5.3 facilitates further comparison through presentation of the three evaluation indices: false alarm rate - P_{FAR}, misdetection rate - P_{MDR} and F1-score. As can be seen, the PCA with the window-based approach performed best, with a 53.8% reduction in false alarms, and only a 2.4% missed detection rate (i.e., 4 of the 168 system alarms went undetected), when compared to the basic PCA approach. The PCA with trial-based approach was also effective, resulting in a 23.1% reduction in false alarms, but an 11.9% misdetection rate when compared to the basic PCA approach. This means that the trial-based approach had three less false alarms (compared to the basic PCA approach), but this came at the expense of 20 missed system alarms. In addition, the F1-score for the PCA with window-based approach yielded the highest F1-score of 97% when compared to PCA with trial-based approach (91%). Considering this, the following sections focus on the performance of the window-based approach model in more detail.

Table 5.3: Evaluation indices for PCA with false alarm moderation approaches (testing period).

	Basic PCA		**PCA with trial-based**	**PCA with window-based**
	System alarms	*False alarms*	*False alarm reduced*	*False alarm reduced*
P_{FAR}	168	13	23.1% less (i.e., 10 false alarms)	53.8% less (i.e., 6 false alarms)
P_{MDR}	-	-	11.9% (20 system alarms not detected)	2.4% (4 system alarms not detected)
F1-score			91%	97%

5.5.3 Detailed results evaluation for PCA with window-based approach

As mentioned above, the monitoring period data obtained for the industry partner represented a period over which a recurring fault occurred at the case-study site. Consultation with the building manager revealed that these faults had occurred due to leakage in the water network in one of the mains underground transmission pipes for a period of five months (from 10th May 2018 (day 146) to 10th October 2018 (day 310)) as indicated in **Figure 5.8** (fault period). The fault initially occurred in one of the mains water pipes and was repaired in mid-June 2018 (day 184) as can be seen in **Figure 5.8** (fault repaired initially). Approximately two weeks later the fault re-occurred at the same location (indicated in **Figure 5.8** - 'fault occurred again' (day 196)). This failure was caused by heavy truck loads in the area over the weakened repaired section. For the second repair the whole pipe section was replaced, with no further faults occurring to date in this area of the site.

The occurrence of this actual fault over the monitoring period provided an excellent opportunity to conduct a detailed analysis of the efficacy of the PCA model with window-based approach. As mentioned in Section 5.4.2 there were 174 system alarms over the 360-day monitoring period. As per the model protocol (Section 5.3.3), system alarms that occurred on consecutive days triggered faults in the FDD framework. Of the 174 system alarms in the testing dataset, twelve sets of faults were triggered by the T^2 α-control limit (dashed line boxes in **Figure 5.8a**) and nine sets of these faults were also triggered by Q α-control limit (dashed line boxes in **Figure 5.8b**). Two additional sets of faults (low flow faults) were detected by Q α-control limit. These would not have been detected by a conventional threshold FDD system.

The pipework fault which occurred at the site, resulted in nonstop lower flow volume loss, and thus went undetected for some time by the facility's existing third-party monitoring system, which is based on the conventional univariate FDD approach. By contrast the PCA with window-based approach detected the pipework fault almost immediately and triggered an alarm at Day 146, over two weeks earlier than the third-party data monitoring system at the site, which triggered fault at Day 162, as illustrated in **Figure 5.8**. For the pipework fault re-occurrence event, the model proposed herein again significantly outperformed the facility's third-party data monitoring system, detecting the fault at Day 196, 9 days earlier than the facilities existing FDD system (which triggered a fault alarm at Day 205). Inspection of metered data from the period concerned, confirmed that the fault occurred. To contextualise the water loss caused by the delayed detection, it is noted that during the fault period the facilities water consumption increased by of 3.4 m^3/day.

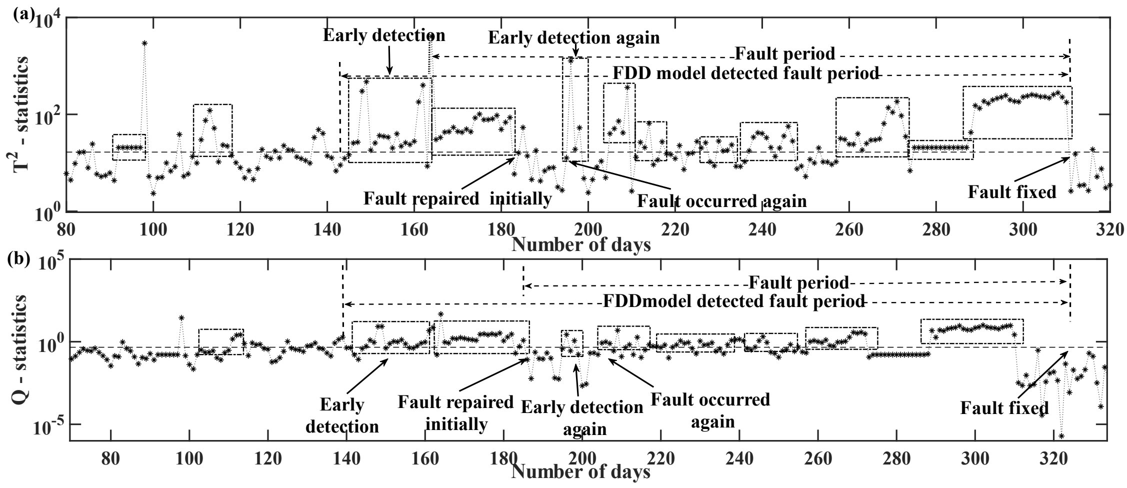


Figure 5.8: The time evolution of T^2-statistics and Q-statistics on a semi-logarithmic scale for the testing data (PCA with window-based approach).

An additional comparative study was also conducted over the testing period contrasting the output of the various models examined herein (basic PCA, PCA with trial-based approach, and PCA with window-based approach), with a conventional univariate approach, which forms the basis for the fault detection and diagnosis system currently in use at the first case-study site. The findings from this comparison are presented in **Table 5.4**. As previously mentioned, 181 system alarms were detected by the basic PCA approach. Of these, 132 mapped directly onto the system alarms from the univariate approach (i.e., the basic PCA approach detected 49 additional system alarms or 27% additional system alarms). Interrogation of the data and site records revealed that of these 49 additional system alarms detected through PCA analysis, 13 were false alarms and 36 were system alarms comprising of either non-routine events or faults (**Table 5.4**). Importantly, 7 of the 13 additional false alarms arising from the PCA method were removed by the window-based approach (53.8%), as discussed in the previous section. It should be noted that in Figure 5.7 each point represents the monitoring statistics (results in PCA space). Only the data points above the control limit are considered as system alarm; these in turn can be due to the non-routine water uses or faults. The monitoring statistics below the control limit represent the routine variation in the data set. In practice, if FastDetect was in place over the testing period, fault alarms would have been raised when system alarm persisted over two consecutive days. The fault would have been subsequently fixed at an earlier stage, which would in turn, result in fewer false alarms and less system alarms that had occurred in the remaining testing period.

Table 5.4: Comparative analysis of detected system alarms (case-study 1).

System alarms	**Matched alarms**	**Additional alarms**		**False alarm reduced**	**False alarm remains**
		True alarms	False alarms		
181	132	36	13	***Trial-based approach***	
				3 (23.1%)	10
				Window-based approach	
				7 (53.8%)	6

5.5.4 Verification of false alarms moderation approaches - Case-study 2

In a previous study on case-study 2 (Hashim et al., 2020), a basic-PCA fault detection model was trained using flow data over a 162 day period. The resulting PCA-based fault detection model was then tested on data from the following 360-days. For the purposes of this fault moderation study, Days 111 - 360 from that testing period were excluded due to an ongoing long-term fault (which resulted in daily system alarms until the fault was eventually fixed). Thus, Days 1 - 110 were analysed herein in the context of fault moderation using the window-based approach only.

As with case-study 1, the observation window length (w) for the window-based approach data was optimised using the data from the (162-day) training period. The training data comprised 162 data points of which the original PCA model detected 9 system alarms. Of these 9 system alarms, 6 were false alarms and 3 were true system alarms meaning that 159 of 162 data points were routine data points. The PCA model combined with the false alarm moderation approaches for the case-study 2 site were developed by utilising the protocols outlined in Section 5.4. The selection of an observation window w for window-based approach was carried out as described in Section 5.5.1.1 and an observation window of length (w = 15) was also found to be most appropriate for case-study 2 (**Table 5.5**).

For observation windows of length w = 5 and w = 10, false alarms were reduced by 66% (2 out of 6 false alarms); for w = 20 false alarm occurrence was reduced by 50% (3 of 6 false alarms). However, in each of these cases all 3 true system alarms were not reported and thus these window lengths would not be suitable. For observation windows of length w = 15, w = 25 and w = 30 all system alarms were correctly reported. However, there was a difference in the reporting of false alarms. For w = 25 and w = 30, the number of original false alarms reduced by 33% (from 6 to 4 false alarms) but 1 “additional” false alarm was created. For *w* = 15 the number of false alarms reduced by 83% (from 6 to 1 false alarm) with no “additional” false alarm was created. In view of these facts an observation window of length *w* = 15 was selected for case-study 2. For both case-study sites, the impact of an observation window length *w* in detecting true system alarms and reducing false alarms were discussed in Section 5.5.1 and Section 5.5.4, respectively. Based on that, an observation window of length (*w* = 15) was found most appropriate for both case-study sites with a detection delay of about 2 to 7 hours. A window length that can raise alarms within 12-hour period of time or less can be used as an initial value when selecting length of an observation window for other industrial sites. However, the exact length of the window may need to be amended to reflect (i) the frequency of data inputs to the PCA model (and thus the windows-based fault moderation system), (ii) the delay that can be tolerated by the user and (iii) the level of false alarms that is considered acceptable. With further research it may be possible to generalise this for various types of water distribution systems.

Table 5.5: Selection of an observation window length *w* (Case-study 2).

	Observation window length *w*	***PCA only***	***w*=5**	***w*=10**	***w*=15**	***w*=20**	***w*=25**	***w*=30**
	Data point considered	162						
	Number of routine data points	159						
	Number of true system alarms*	3						
a	Number of original false alarms	6						
	Number of system alarms detected	3	1	2	3	2	3	3
b	Number of original false alarms reduced	N/A	5	5	5	3	3	3
c	False alarms produced by method	0	1	1	0	0	1	1
(a+c)-b	Final number of false alarms	N/A	2	2	1	3	4	4

*Verified by manual analysis

The results for the case-study 2 are presented using Hotelling T^2-statistics as no system alarm was observed in the Q-statistics (**Figure 5.9**). During the testing period of case-study 2 site the faults resulted from excess usage and metering error faults, detailed discussion can be found in (Hashim et al., 2020). The performance of the PCA model combined with false alarm moderation approaches are represented graphically in **Figure 5.9**. It can be noted that during the period considered 42 data points (about 38% of data points) exceeded the α-control limit resulting in system alarms as shown in **Figure 5.9a**. These system alarms were caused by some on-site non-routine water uses or faults in a system. Of these 42 system alarms raised by basic PCA-based fault detection, 11 system alarms (26% of system alarms) were false alarms (black solid data points - **Figure 5.9a**). These false alarms were verified manually against the historical records for the building to ensure they were not in fact system alarms. **Figure 5.9b** and **Figure 5.9c** shows the impact of the fault moderation on reducing false alarms. **Table 5.6** shows the presentation of false alarm rates (P_{FAR}), misdetection rate (P_{MDR}) and F1-score for the various approaches. The PCA with the window-based approach performed best, with a 54.5% reduction in false alarms (from 11 in the original model to 5 when the window-based approach was adopted - **Figure 5.9c**), and only a 3.2% missed detection rate (i.e., 1 of the 31 system alarm went undetected), when compared to the PCA with trial-based approach in which 18.2% of false alarms were reduced (2 of the 11 original false alarms were no longer detected as system alarms - **Figure 5.9b**) with a misdetection rate of 9.7% (i.e., 3 of the 31 system alarms went undetected). In terms of performance, the F1-score of the PCA with window-based approach score of 92% reflects its enhanced performance when compared to PCA with trial-based approach (F1-score = 84%). It can be noted that all system alarms both case-study sites were treated to be equal in calculating missed detection rate, but in fact they will have varying financial and sustainability implications. Previous work has also shown that PCA-based models can be effective in detecting system alarms related to low flow faults, metering error and excess usage faults (Hashim et al., 2020). System alarms resulting from low flow and metering error faults are difficult to detect due to the complex infrastructure of industrial water distribution systems (as illustrated in Section 5.5.3). The costs associated with low flow and metering error faults can vary widely depending on the process, water costs and the complexity of the water network (or even the location of the fault). While excess usage faults can often be more easily detected due to having distinguishable/detectable characteristics when compared to other faults, the associated repair costs may be minor if the leak is easily accessible and can be rapidly isolated to significant if it causes infrastructure damage and requires civil engineering works. Furthermore, research into the sensitivity of both fault detection and diagnosis systems and supporting false alarm moderation techniques could determine the degree to which these methods can reduce misdetection rates and false alarms for various types of faults (e.g., low volume continuous faults, high volume continuous faults or one-off events). The findings from the integrated PCA models with false alarm moderation were thus effective in reducing false alarms while maintaining sensitivity in relation to detecting true system alarms. This was particularly the case for window-based approaches when appropriate window sizes were selected. In the two case-study sites previous false alarms were found to be due to the combination of non-routine events and routine statistical variation related to the inherited definition of the α-control limit, which is linked to false alarm probability (Li et al., 2019). In case-study 1, such false alarms resulted from non-routine events were mainly due to unscheduled (but relatively frequent) cleaning activities on site, which consumed significantly more water than routine water usage. In case-study 2, the non-routine events which caused false alarms were mainly conferences, seminars or particular laboratory activities which resulted in increased water consumption. The optimized PCA model with window-based approach was found to be sensitive to water distribution system faults whether the faults are of total failure such as metering error or low flow faults as illustrated in the case-study sites. Furthermore, the length of an observation window and delay in detecting fault are inter-dependent (Najah et al., 2021). Thus, the selection of an observation window length is critical, inappropriate window length (i.e., too small, or too large as discussed in Section 5.4.2.1) could result in increased detection delay. For example, for case-study 1, the original data comprised of a volume reading every 30 minutes. Given that in case-study 1, w = 15, a potential delay of 7.5 hours could result in terms of the fault being detected. This could be considered a small fault detection delay when compared to third-party monitoring systems, which resulted in the detection delay of over two weeks (as discussed in Section 5.5.3). Thus, there is a trade-off associated with the detection delay and the accuracy of detecting faults. Too small a window length will reduce detection delay but can result in increased number of false alarms. While too large a window length can lead to relatively high detection delay but can result in increased misdetection (as discussed in Section 5.5.1) or vice versa. A key consideration in this regard is the time interval of the data being considered i.e., data with a sub-hourly time series is ideally suited to the window-based approach, however daily data would be less suitable given that practical window lengths could result in detection delays of 7 to 14 days. The detection delay as performance indicator would be more effective in dealing with short-length time series (i.e., minutely, or hourly) of water consumption data which would be analysed in future work.

Table 5.6: Evaluation indices for PCA with false alarm moderation approaches (Case-study 2).

	Basic PCA		PCA with trial-based	PCA with window-based
	System alarms	*False alarms*	*False alarms*	*False alarms*
P_{FAR}	42	11	18.2% reduction (9 false alarms)	54.5% reduction (9 false alarms)
P_{MDR}	-	-	9.7% (3 system alarms not detected)	3.2% (1 system alarms not detected)
F1-score			84%	92%

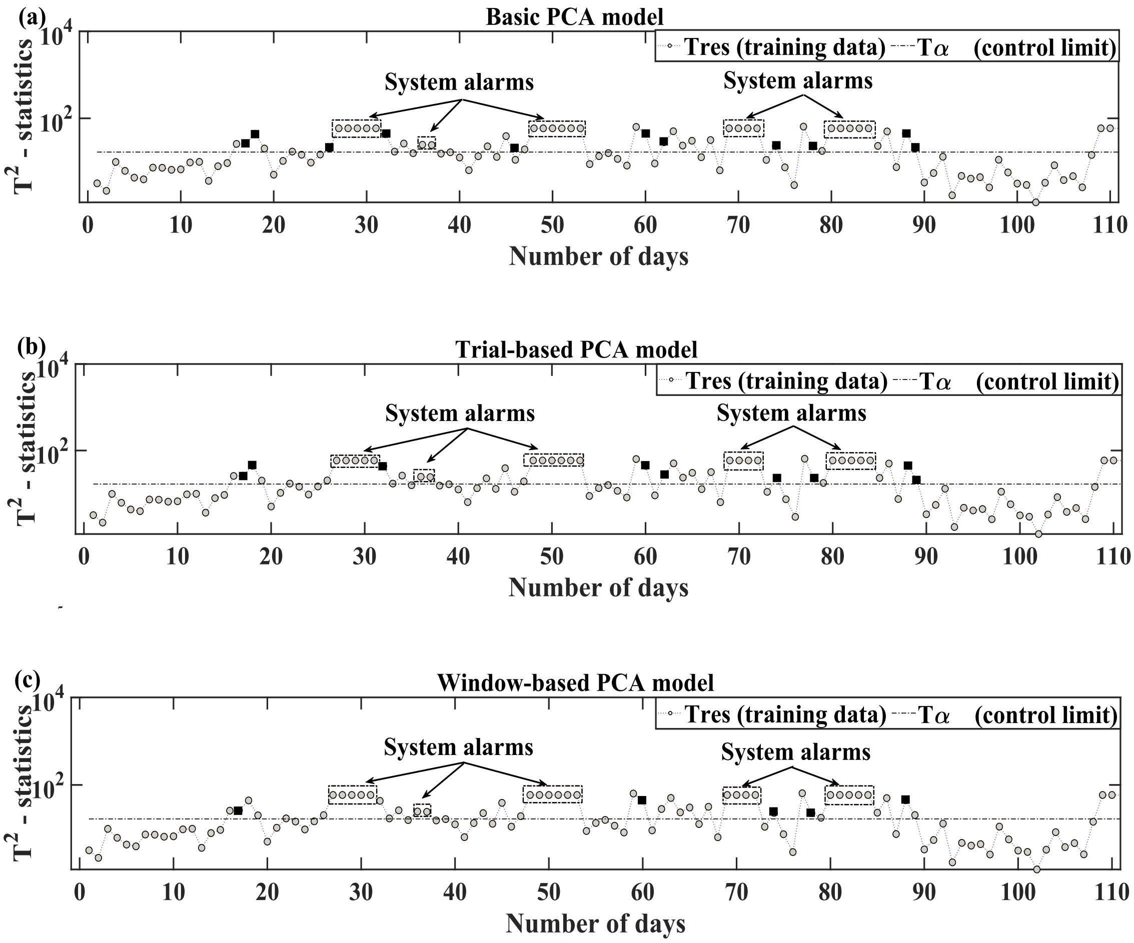


Figure 5.9: (a) The time evolution of T2-statistics in basic PCA, (b) PCA with trial-based approach, and (c) PCA with window-based approach for the case-study 2 (Alice Perry Engineering Building at the National University of Ireland, Galway).

5.5 Conclusion

The prevalence of false alarms is one of the primary reasons for building managers reluctance to integrate and utilise fault detection and diagnosis methodologies as part of building monitoring systems. Given the scale of water usage by the industrial sector, this reluctance has considerable sustainability consequences. This part of work aims to help address this important issue through false alarm reduction in building water distribution systems fault detection and diagnosis using a combination of principal component analysis (PCA) and false alarm moderation approaches, which have not been used in this sector to date.

Two statistical false alarm moderation approaches were examined, namely a window-based approach and a trial-based approach, for their efficacy in reducing false alarms resulting from PCA-based fault detection model. The window-based approach model was found to be most effective in reducing false alarms arising during fault detection and diagnosis of data from the two case-study facilities. For the case-study 1, the model resulted in a 53.8% reduction in false alarms, with a 2.4% increase in missed detection rate and a F1-score of 97%, when compared to a PCA approach without false alarm moderation. In contrast, the trial-based approach delivered a 23.1% reduction in false alarms, but this came at the expense of an 11.9% missed detection rate and a F1-score of 91%. Further evaluation of the window-based approach revealed it detected more faults than the currently applied conventional univariate approach, and importantly for the real site data it detected leaks in the system earlier (up to 16 days earlier for a pipeline failure at the case-study site). For the case-study 2, the PCA with window-based resulted in a 54.5% reduction in false alarms with a 3.2% increase in missed detection rate and a F1-score of 92%, and the PCA with trial-based delivered a 18.2% reduction in false alarms, but this came at the expense of an 9.7% missed detection rate and a F1-score of 84% when compared to the basic PCA approach. The PCA model combined with false alarm moderation approaches showed good capability of detecting true alarms when appropriate window sizes were selected.

The false alarm moderation performance, particular using a window-based approach, is promising given require large amounts of fault training data (an issue with machine learning approaches for industrial buildings) are not. Furthermore, both approaches can be integrated into PCA-based fault detection and diagnosis systems and does not require additional human resources or input once commissioned.

6. THE ECONOMIC ASPECTS OF WATER-ENERGY INTERACTION LINKED WITH FAULTS IN NON-RESIDENTIAL WATER DISTRIBUTION SYSTEMS

6 The Economic Aspects of Water-energy Interaction Linked with Faults in Non-residential Water Distribution Systems

6.1 Overview

This chapter analyses the energy and economic impact of water use in the case-study 1 site (Alice Perry Engineering Building NUI Galway). This analysis is then used to develop a cost-benefit analysis for the implementation of FastDetect in industrial settings. For Case-Study 1, pumping energy used for water supply, hot-water consumption, and associated energy use (e.g., pump start-ups) were analysed. The work in part of this chapter (Chapter 6.3) has been published in Water (Nair et al., 2018; https://doi.org/10.3390/w10060810). Chapter 6.3 details the analysis of water related energy use (both pumping and heating) in a non-residential building which is a fundamental part of the over resource consumption associated with water use. Furthermore, ongoing faults are likely to induce increased water and energy costs. Later in the chapter, a cost-benefit analysis is presented for the implementation of FastDetect by considering case-study 1 as a pilot site. In this chapter the “testing periods” mentioned equate to the testing period for FastDetect presented in Chapters 4, respectively and comprises two Semesters and a summer period. Thus (as per Chapter 4) this equates to a full year (January to December). The “training period” refers to the training period used for FastDetect model development in Chapter 4 and equates to Semester 2 (January to May) from the previous year.

6.2 Energy use for water pumping in buildings

Energy is used in every stage of the water cycle, from water abstraction, treatment, distribution and end-use (Lam et al., 2017). In recent years, increasing energy use for provision of water to end-users, rising energy cost, and the need to mitigate climate change have been drivers for better management of energy use and associated carbon footprint in the water sector (Bylka and Mroz, 2019; Kenway et al., 2015). The electricity consumed to extract, pump and treat water represents 2.6% of European Union energy use (European Commission, 2019). However, analysing water related energy use can be complex given that end-uses are highly varied, and scenarios range from a single household with single occupancy to large non-residential (commercial and industrial) estates and various municipal buildings. In a European context, buildings constitute around 40% of the energy use and 36% of CO_2 emissions in European Union and its directive on energy performance of buildings requires all new buildings to be nearly zero-energy buildings by 2020 (European Commission, 2019). A methodology aimed at demand side energy efficiency activities has been developed by the United Nations Framework Convention on Climate Change that includes water conserving devices in buildings (European Commission, 2012).

In Ireland, buildings are the second largest consumers of energy after the transport sector (SEAI, 2015). In 2014, buildings consumed 35% of total energy and 59% of electricity in Ireland. To-date, energy use by water end-users in buildings has received limited attention in Ireland (Dubuisson, 2019). The ratio of water extraction to water availability is low (2%) in Ireland when compared to most other European Countries (Walsh et al., 2015). Perhaps as a result of this, historically water conservation has not been a major issue and the hidden economic and energy costs involved in pumping, treatment, end-use and disposal of water are often neglected (Walsh et al., 2015). In building water distribution systems, energy cost associated with pumping water represents a significant proportion of the overall operational costs. This is exacerbated by leakage and other faults (Fletcher et al., 2018).

In buildings, energy use in water pumping and heating is directly associated with the water end uses. In case of residential water end use, there is generally no requirement for onsite pumping as the mains water supplied from the utility usually has sufficient pressure to maintain the flow (Siddiqi and Weck, 2013). However, in non-residential buildings that comprise several floors and facilities such as laboratories, production lines, offices, toilets, canteens, etc. The pressure supplied by the utility is unlikely to be sufficient and may require additional "booster pumping" to maintain the water pressure at the user-end. In literature, analysis of the energy used for pumping water considers pump operation only and does not include pump start-ups (Beghi et al., 2016; Vieira et al., 2014). In non-residential buildings, frequent pump start-ups can constitute significant energy use (as high as 60% increase in total pumping energy consumption). Energy use during pump standby mode is considerably less significant at around 0.2% of total pumping energy consumption (Nair et al., 2018; Beghi et al., 2016; Vieira et al., 2014). Similarly, additional energy use associated with faults may vary according to the water distribution system infrastructure, pumping configuration and operating policy in a building. Water loss or inefficient water use as a result of faults implicitly leads to the loss of energy resource representing an increased emission of greenhouse gases. Even low-level, normally imperceptible faults can lead to higher dynamic losses requiring pump upgrades or replacement due to non-routine pumping operation (Abdulshaheed et al., 2017).

6.3 Case-study 1 - Alice Perry Engineering Building NUI Galway

For this case-study, water usage data was aggregated into diurnal flow volumes to analyse energy use for pumping. The mains water system in the building is divided into a cold-water system and potable water supply for drinking fountains, a café, and other similar uses (as described in Chapter 4 - Section 4.3.2). Two booster pumps were used to maintain pressure in the cold-water system water and other water supply such as for drinking fountains and for the café. Detailed data for drinking fountains and for the café was not available due to metering error for the study period analysed. The drinking fountains and café uses were relatively small (about 3% - established based on available data from previous years) compared to the total cold-water system water use within the case-study site and were not considered for energy use analysis. The building also comprised a large rainwater-harvesting system for supplying grey-water. During the testing period, the building rainwater-harvesting system was not working due to some system faults. Hence the pumping energy quantification involved booster pumping for grey-water uses and other cold-water uses.

6.3.1. Building hot-water system – Case-study 1

In Case-study 1, water was heated to a required temperature (usually 60ºC) using a centralized natural gas heating system located in an adjacent building (energy centre) and circulated through a calorifier system as presented in **Figure 6.1**. One calorifier (Cal 2001) is used to transfer heat to the cold-water coming from the cold-water system. The water is then further heated in a second calorifier (Cal 2002) using solar energy (when available), before being transferred to Cal 2001 for heating to a set temperature as required for end uses. The cold-water feed is brought into the solar cylinder to heat the cold-water to maximise the use of renewable energy. As per the operation manual, the solar energy potential of the system installed on the building roof is 20 kWh/day from an area of 12 m^2 solar panels. The current study did not disaggregate hot-water end users. It should be noted that the calorifiers are electric and serve showers and hot water taps. They are supported by solar panels which were not operational during the testing period. Likewise, the space heating is via gas boilers and were not considered in this study.

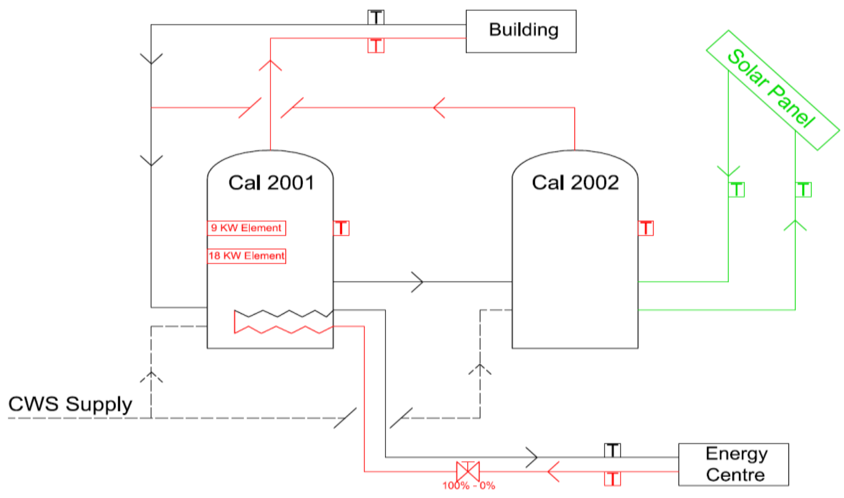


Figure 6.1: Hot-water system in the building (Alice Perry Engineering Building NUI Galway).

6.3.2. Energy consumption in the building – Case-study 1

The overall methodology adopted for the estimation of energy associated with water end uses is outlined in **Figure 6.2**. Water heating and pumping constituted the energy consumption associated with building water end use. The water heating energy calculation approach is explained in Section 6.3.2.1, while the same for pumping energy is described in Section 6.3.2.2.

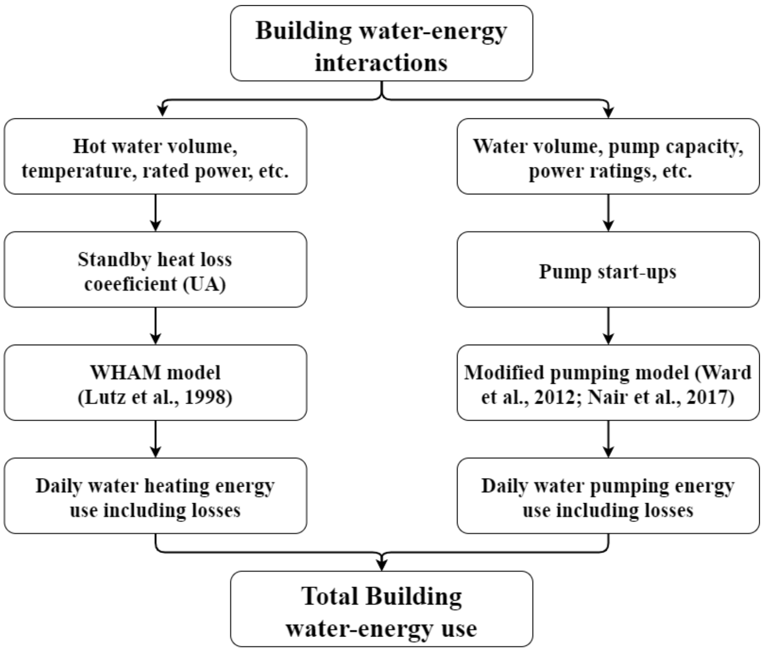


Figure 6.2: Building water-energy use estimation methodology.

6.3.2.1. Water heating energy use in the building – Case-study 1

The energy consumed on a daily basis for water heating in the building was analysed using the water heater analysis (WHAM) model expression. In the storage water heater, as water temperature was constantly maintained, distribution and standby heat losses were key. Given the building in the case-study 1 had a relatively high level of occupancy (particularly at certain times of the year), standby heat losses were included, and the total energy used for water heating was estimated using the WHAM by (Nair et al., 2018; Sterman, 2006).

$$E_{hw} = \frac{V_{hw} * \rho * C_p * (T_{st} - T_{in}) * 0.00028}{\mu} \times (1 - \frac{E_{loss.standby} * (T_{st} - T_{amb})}{P_{in}}) + 24 * E_{loss.standby} * (T_{st} - T_{amb}) \quad (32)$$

where E_{hw} is the total energy used for water heating per day (kWh/day), V_{hw} is the daily volume of water heated in the building (m^3), ρ is the density of water (1000 kg/m^3), C_p is the specific heat of water (4.187 kJ/kg-^oC), T_{st} is the set temperature (^oC), T_{in} is the inlet temperature (^oC), μ is the water heater recovery efficiency, $E_{loss.standby}$ is the standby heat loss coefficient, T_{amb} is the ambient temperature (^oC) and P_{in} is the rated input power of the heater (kW).

The WHAM model accounts for a range of operating conditions and energy efficiency characteristics of water heaters as compared to the general expression for calculating the water heating energy use. Recovery efficiency (μ), standby heat loss coefficient ($E_{loss.standby}$) and rated input power (P_{in}) were the efficiency parameters used in the model. Daily hot water usage, water temperature at heater inlet, water heater tank temperature and ambient air temperature around the heater were taken as the operating parameters. The recovery efficiency (μ) can be defined as the ratio of energy added to the water compared to the energy input to the water heater. This accounts for the losses transferring energy to the heater and then into the water. In this study, the efficiency of the heater was assumed to be 76% based on the average of efficiencies for gas heaters at various efficiency levels as given by DoE - Design of Experiments (AWWA, 2016). The rated input power (P_{in}) was taken as the nominal power rating of the water heater assigned by the manufacturer in kW (96 kW in this case). The inlet water temperature was assumed to be 5^{o}C more than the mains water temperature (for the case-study 1, the mains water temperature was around 3^{o}C) (Howley and Holland, 2016) and the average set temperature in offices and class room spaces within the building was 21.5^{o}C. The set temperature of the heater output was 60^{o}C.

Standby heat loss coefficient $E_{loss.standby}$ (kW/^{o}C) is the rate at which energy must be added to the water heater to maintain the water at desired temperature when it is not heating water for delivery and can be calculated as follows (Sterman, 2006).

$$E_{loss.standby} = \frac{(\frac{1}{E_{factor}} - \frac{1}{\mu})}{(T_{st} - T_{amb}) \times (\frac{24}{P_{out}} - \frac{1}{\mu * P_{in}})} \tag{33}$$

where E_{factor} is the energy factor of the heater given by the manufacturer, μ is the efficiency of the heater and P_{out} is the heat content of the water drawn from the heater in kW. The average energy factor E_{factor} or energy efficiency level for gas-fired storage water heaters was taken as 65% (AWWA, 2016).

In the water heating system of the case-study site (Alice Perry Engineering Building NUI Galway), the pipes distributing hot water to various parts of the building were well insulated using mineral wool (with a thermal conductivity of 0.037 W/m.K) to avoid distribution energy losses. The heat losses were estimated and were about 0.004 kWh/day. Hence, heat losses in the water network were not represented in the above expression. The estimated energy use per day for the water heating system is shown in **Figure 6.3a**. The days with the metering error (days with no data points) were not included in estimating the energy use for water heating (138 non-consecutive days with data points). Subsequently the metering error was in fact found to be the result of flow data not being logged (which in this study are equivalent to “zero” flow days). The average energy used for water heating in Semester 1 was 219 kWh/day, 68 kWh/day in the summer period and 419 kWh/day in Semester 2. An average energy input of 236 kWh/day was observed to overcome stand-by energy losses (excluding energy losses in the pipework) and maintain the water temperature in the water network. There was a marked reduction in energy use for water heating in Semester 1 compared to Semester 2. The major share of hot-water use in the building was attributed to showers. The reduction in hot-water use was attributed to stringent demand management measures that was implemented lately in the building (e.g., push button showers and lower pressure shower heads). Though the effect of temperature might not be significant on water use quantity, it is paramount to the impact on energy use, especially energy intensity. This was evident from the energy intensities for Semester 1, summer, and Semester 2 analysed. The water heating energy intensities of Semester 1, summer, and Semester 2 were estimated and were 65, 60, and 58 kWh/m^3, respectively (an average of 61 kWh/m^3 was observed across the year). The hot water consumption of Semester 1, summer, and Semester 2 were also analysed **(Figure 6.3b)** and were around 7, 2, and 3 m^3/day, respectively (an average of 5 m^3/day was observed across the year). The variation in energy intensities could be due to reduced heating requirements across different periods (Semester 1, summer, and Semester 2) and reduced energy losses in warmer months. The findings indicated that water heating energy consumption accounted for about 18% of the total heating energy use in the building.

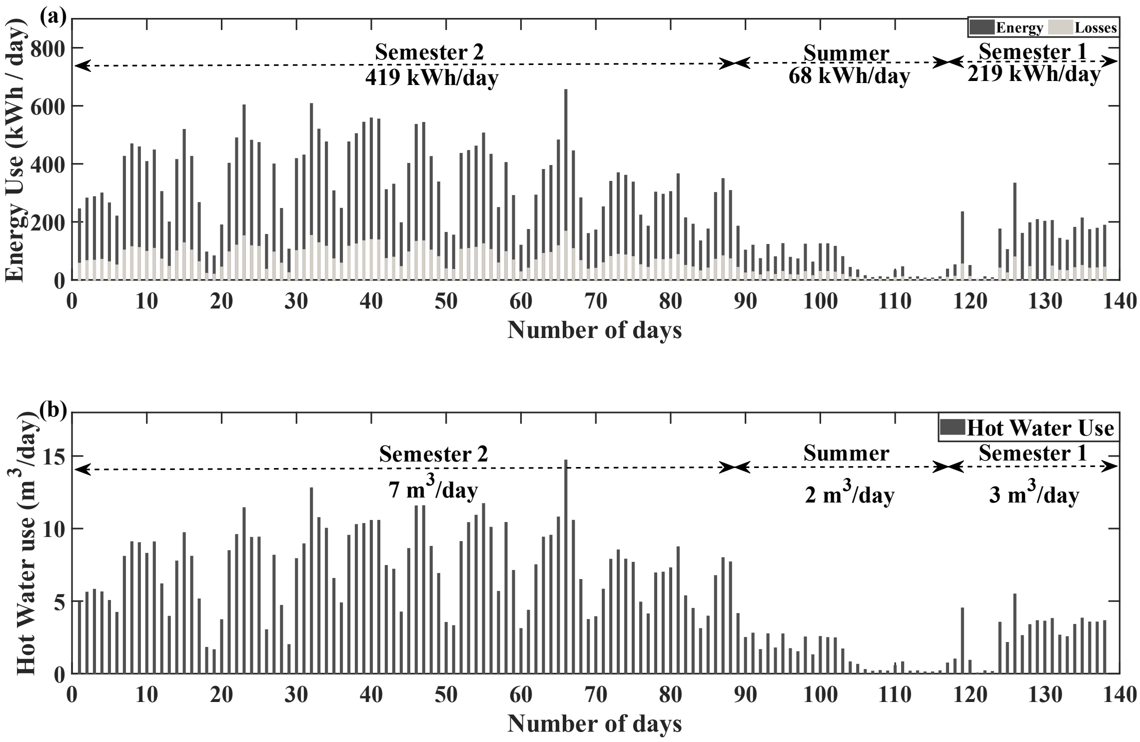


Figure 6.3: (a) Energy use for water heating in the building (non-consecutive days) and (b) hot water use in the building (non-consecutive days).

The variations in energy intensities across different study periods could be due to reduced heating requirements and reduced energy uses in warmer months (Fuentes et al., 2018).

There does not appear to be literature available regarding the energy intensities of hot water use in large buildings; thus, a direct comparison with this study is difficult. However, Fuentes et al., (2018) have reported the hot water use and energy use per day for households for various countries in Europe, US, and Canada. The average European (Germany, Spain, Portugal and Switzerland) household hot water use energy intensity was around 52 kWh/m^3, with the exception of UK (35 kWh/m^3) and Finland (42 kWh/m^3). The energy intensities of US and Canadian households were reported as 23 and 46 kWh/m^3, respectively. The energy intensities obtained in this study are high compared to other stages of the urban water cycle, as mentioned in (Walsh et al.,, 2015). For example, the energy intensity of seawater desalination is usually considered to be the highest, falling in the range of 3-8.5 kWh/m^3 (Walsh et al., 2015). Therefore, the water heating energy intensity of this building was around 7.5 times higher than that of a desalination plant. Analysis of losses associated with water heating showed that the losses for standby heating in this building were about 18% of the total water heating energy. This equated to 56 kWh/day during the testing period, with an average energy intensity of about 11 kWh/m^3.

6.3.2.2. Energy consumption for water pumping in the building – Case-study 1

The pumping energy calculation approach has been outlined in a paper published as part of this research (Nair et al., 2018). The total pumping energy required including the pump start-up energy, and operating energy can be given as (Beghi et al., 2016; Nair et al., 2018).

$$E_{tot} = P_{tot} - P_t/P_{elec} \qquad (34)$$

where E_{tot} is the total energy use for pumping water (kWh), P_{elec} is the power withdrawn from the electricity supply (kW), P_t is the actual power transferred to the water (kW) and P_{tot} is the sum of energy use for pump start up and operation. The above expression does not consider standby energy usage of the pump when water is not being pumped. Therefore, considering that aspect, the above expression can be modified as follows (Ward et al., 2012; Nair et al., 2018).

$$E_{tot} = \left[\frac{P_t V_t}{Q_p}(f_s\ V_s\ N_s + 1) + \ P_{sb}(24 - \frac{V_{op}}{Q_p})\right]\left(2 - \frac{P_t}{P_{elec}}\right) \qquad (35)$$

where V_t is the total volume of water consumed (m^3), Q_p is the pump capacity (m^3/hr), V_s is the percentage of total volume pumped during the start-up phase, N_s is the number of start-ups in a day, f_s is the percentage of energy consumed during the start-up phase, P_{sb} is the power consumed during stand-by (kW), V_{op} is the volume pumped during pump operation. The number of pump start-ups can be given as (Nair et al., 2018).

$$N_s = \frac{V_t}{n_{ht}\ (\ V_{T,u} - V_{T,l})} \tag{36}$$

where $V_{T,u}$ is the upper threshold volume of header tank (m^3), $V_{T,l}$ is the lower threshold volume of header tank (m^3), n_{ht} is the number of header tanks. The pump power withdrawn from electricity supply is given as (Nair et al., 2018).

$$P_{elec} = \frac{P_t}{\eta_s} \tag{37}$$

where η_s is the system efficiency. The energy consumption for pumping operation E_{op} is given by a simple expression (Nair et al., 2018).

$$E_{op} = P_{elec} \times t_{op} \tag{38}$$

where t_{op} is the duration of pump operation. The variables and parameters for the above set of equations were acquired from the pump and water distribution system specifications in the Alice Perry Engineering Building at NUI Galway. The parameters required and used to estimate the energy use (Eq. 31-35) are summarised in **Table 6.1**. The pump differential head and the associated frictional losses were estimated based on measured (using engineering as-built drawings) of the pipe network length within the building.

Table 6.1: Pumping energy parameters - Case-study 1.

Parameter	Description	Value
Pr*	Rated power	5.5 kW
Qp*	Pump capacity	21 m^3/hr
H*	Pump differential head	80 m
Vs**	Start-up volume	0.001 m^3
Fs**	Start-up factor	0.6 %
Psb**	Stand-by power factor	0.002

Rated values from pump motor name plate*, Literature** (Nair et al., 2018; Ward et al., 2012; Vieira et al., 2014).

Figure 6.4a depicts the energy expended for pumping water to top up the grey-water system and the energy used for pumping water for the other cold-water uses. During this testing period, around 74% of flow datapoints resulted in system alarms, of which 17% of alarms resulted from non-routine water uses and excess usage as discussed in Chapter 4. The increased energy consumption in Semester 2 for grey-water uses and other cold-water uses resulted from a fault (which occurred in Semester 1 but continued into Semester 2), changes in routine water uses and non-routine water uses attributed to hosting one-off events, improper water use in laboratories, etc. It should be noted that 62% of the data points were zero flow readings (resulting from a metering error) and were not included in estimating the energy use for water pumping.

The energy use for pumping is directly related to the water volume used on that given day. In the building, water meters were installed in only one large bathroom (with male and female rooms) to monitor the grey-water uses. Thus, the grey-water use for the entire building was analysed measuring water consumption of this bathroom (with male and female rooms) and extrapolating to other bathrooms in the building. Based on that grey-water use for urinal flushing and toilets for the whole building was estimated by considering parameters such as volume used for flushing, number of urinals and toilets within the building, cyclic flushing routine, etc.

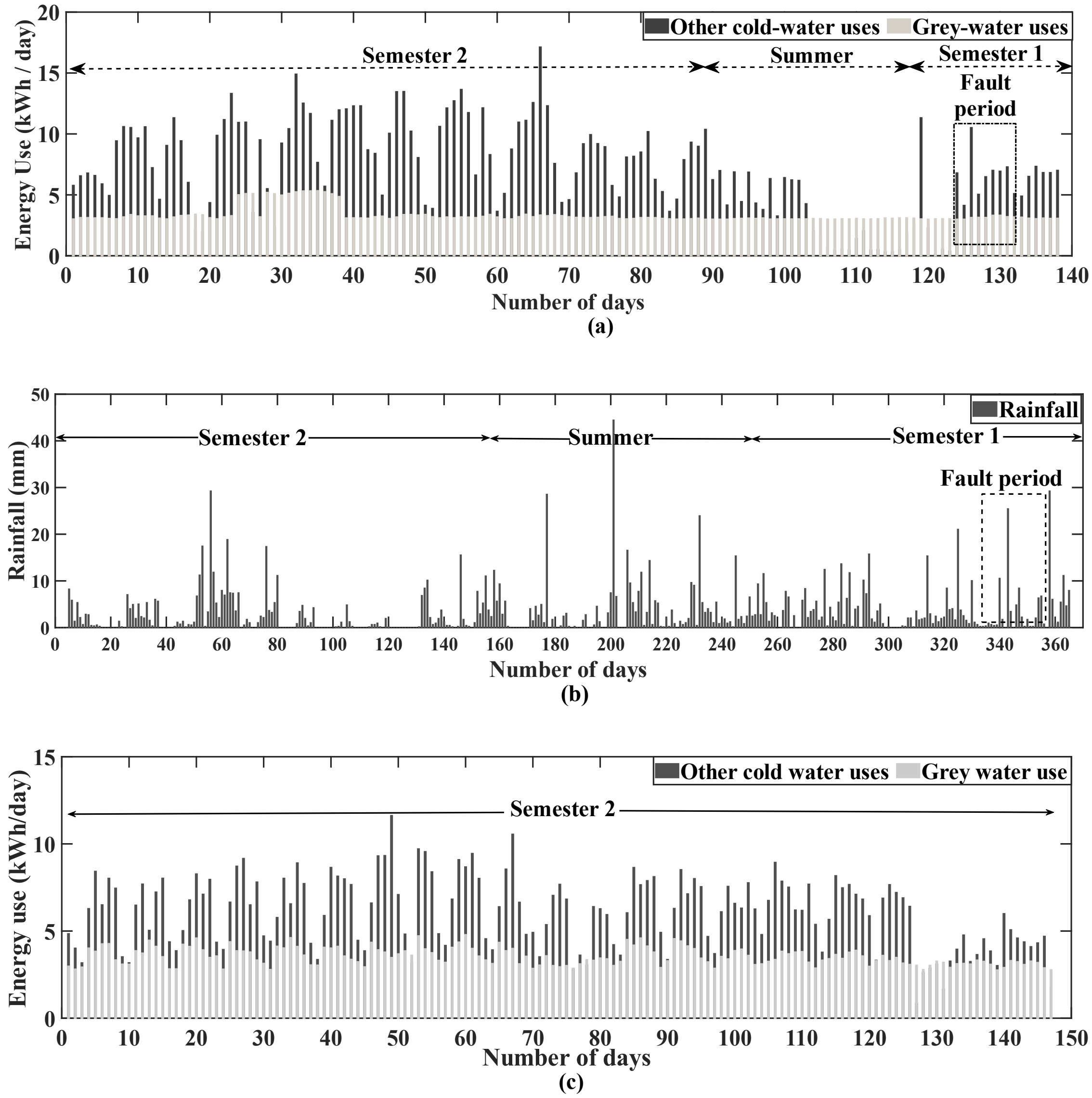


Figure 6.4: Energy used for water pumping in the building, for grey water uses and other cold water uses, (a) testing period (b) testing period rainfall, and (c) training period used for FastDetect model development in Chapter 4.

During the testing period (**Figure 6.4a**), the average energy use (excluding days with metering errors) for grey-water consumption was 3.33 kWh/day and for other cold-water consumption was 7.06 kWh/day. There was a relative consistency observed in energy use for grey-water uses throughout the testing period for which the data was available, due to urinal flushing which occurs irrespective of the building occupancy. The average energy use for these cold-water uses during the testing period was 29% higher than the average pumping energy use for cold-water consumption during the training period (which averaged 5 kWh/day). This was due to the fault in the rainwater harvesting system (discussed in Chapter 4). On average, a 38% increase in energy use for pumping was observed for grey-water uses and other cold-water uses compared to energy use in the training period (where this fault was not observed). **Figure 6.4b** shows the rainfall data obtained from the weather station in NUI Galway for the whole testing period (360-days) includes the days with no data points. Moderate rainfall was observed during the period of interest (the period in which the fault had occurred). During the testing period, the building grey-water system was not operating due to some system faults, added energy was used by booster pumps to maintain the constant supply of cold-water to the roof top tanks for the grey-water uses (resulting in additional pumping energy use for cold-water uses during the testing period).

6.3.2.3. Energy consumption for pump start-ups

The energy consumed during pump start-ups for grey-water and other cold-water uses was also assessed at a monthly interval. The water usage data was aggregated into monthly flow volumes to study the energy use during start-ups as a percentage of corresponding total pumping energy use in each month (Eq. 36) – **Figure 6.5**.

The start-up energy $E_{start\text{-}up}$ can be calculated as follows (Nair et al., 2018).

$$E_{startup} = \frac{P_t\, V_t\, V_s\, N_s\, (1 + f_s)}{Q_p} \qquad (39)$$

where V_t is the total volume of water consumed (m^3), Q_p is the pump capacity (m^3/hr), V_s is the percentage of total volume pumped during the start-up phase, N_s is the number of start-ups in a day, f_s is the percentage of energy consumed during the start-up phase, and P_t is the actual power transferred to the water (kW).

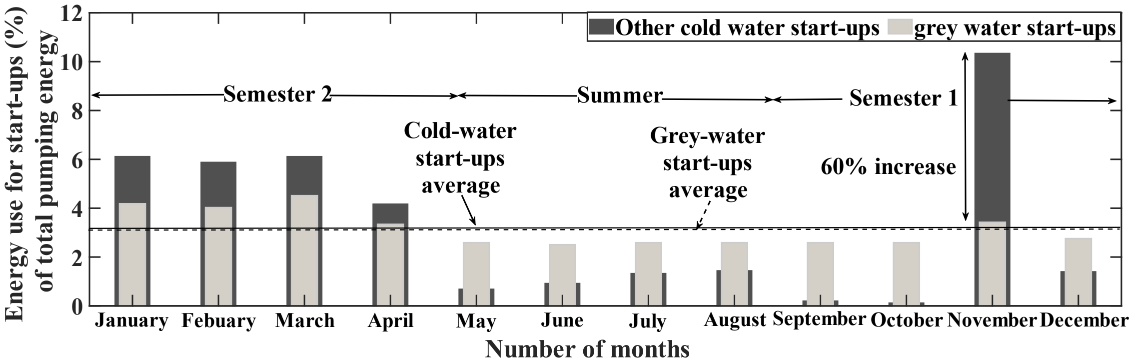


Figure 6.5: Monthly pump start-up energy usages for grey-water uses and other cold-water uses for the testing period.

For the testing period pump start-ups contributed averaged 6% of the total water pumping energy use and pump start-ups for grey-water use alone was responsible for around 3% of total water pumping energy use. During Semester 2, average pumping energy due to start-ups for total water (cold-water and grey-water) was 5% which decreased to 2% in the summer period. In November it then increased to 10% for cold-water uses alone before decreasing to 3% in December. A change of roughly 60% increased start-ups for other cold-water uses was observed (**Figure 6.5**) and can be linked with the excess usage fault (rainwater harvesting system fault) that occurred in the testing period (Fault period – **Figure 6.4a**); less variation was observed in the case of start-ups for grey-water uses. On average, a 3% increase in total energy use due to start-ups was observed for the testing period when compared to the training period. During the testing period, the number of pump start-ups per day for other cold-water uses averaged 3 which increased to an average of 5 start-ups per day when the fault occurred.

6.4 Economic aspects of water-energy interaction - Case-study 1

To study the economic aspects of water-energy interaction, the costs of energy and water were estimated based on Irish energy and water price statistics for non-residential buildings which averages €0.2/kWh[4] and €1.10/m^{3}[5] (Irish Water, 2019; Kenway et al., 2015). The estimated water cost and electricity cost associated with water pumping for the testing period are shown in **Figure 6.6a,** and **Figure 6.6b**. On average, a 41% increase in water-energy cost was observed during the testing period when compared to the

[4] https://www.seai.ie/data-and-insights/seai-statistics/key-statistics/prices/.
[5] https://www.water.ie/about-us/our-company/annual-reports/.

training period. The increased water-energy cost in Semester 2 could be attributed to increased number of occupants in the building and increased outdoor activities of the occupants during Semester 2 (especially in Spring). The increased water-energy costs in Semester 1 were due to energy used in booster pumps to maintain a constant supply of cold-water to the roof top tanks for the grey-water uses (as discussed in Section 6.3.2.2). During the testing period, the water cost observed roughly ten times greater than the energy cost.

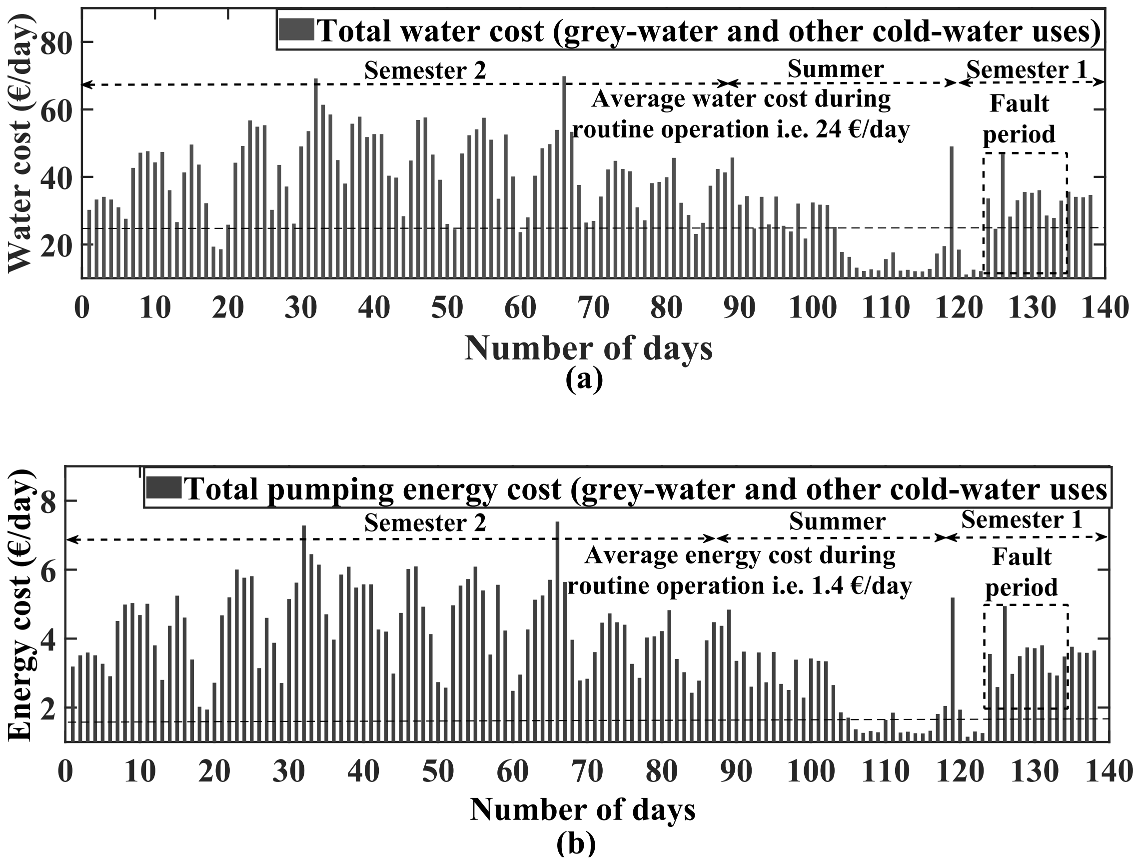


Figure 6.6: (a) Water cost for the testing period, and (b) Energy cost for the testing period (non-consecutive days).

The economic aspects of water-energy interaction associated with faults in a large-scale non-residential building's water distribution system are often overlooked and need to be incorporated into performance monitoring frameworks. Extrapolation of such losses on a larger scale would suggest substantial impacts on economy and the environment. Moreover, extrapolation of such losses offers water-energy savings opportunities. Costs associated with the operation of water distribution system in non-residential buildings are not limited to the water and energy consumed by the water distribution system.

Unnecessary wear, leading to premature component failure, increases costs through the embodied water-energy and material resources in the replacement of equipment and the indirect costs associated with the restoration process (e.g., transport). Leakage or inefficient water distribution systems operation gives rise to global and local environmental issues. All of which suggests the need for other indices, besides direct water, and energy costs, when assessing the performance of water distribution system in non-residential buildings. In the following section, the cost-effectiveness of FastDetect was examined to evaluate the costs and benefits associated with the application of FastDetect in non-residential water distribution systems.

6.5 Cost-benefit analysis

A cost-benefit analysis was conducted to assess the costs and benefits associated with implementing FastDetect. The cost model was developed for the case-study 1 (Alice Perry Engineering Building NUI Galway) and evaluated the expected life cycle costs for the use of FastDetect in a continuous performance monitoring mode. The life cycle cost was estimated over a 5-year time horizon. Case-study 1 was considered given the levels of information available to analyse costs and benefits linked with implementing FastDetect in large non-residential buildings. The mode assumed FastDetect would be installed in a building and then monitored as a service.

The cost-benefit analysis conducted provides economic feasibility for the application of FastDetect in non-residential water distribution systems. Improving building water distribution system’s efficiency provides direct financial benefits to stakeholders/building managers by reducing their costs for water, and energy, can also lead to lower operation and maintenance costs.

6.5.1 Cost-benefit analysis methodology

The cost-benefit analysis methodology used herein simply assumes that by increasing the early detection of potential (critical) faults, the likelihood of faults will decrease and lead to a reduction in overall maintenance time with a concomitant increase in profitability of the non-residential facilities. There are two approaches to assess the annualized performance monitoring costs, labour costs, and economic efficiency based upon cost-benefit analysis. The approaches most typically used for such applications are net present value (NPV) and benefit cost ratio (BCR) (Hsu, 2021; Liu et al., 2020; Petkov et al., 2020; Makonin et al., 2018; Rajan and Roylance, 1996). NPV represents the difference between the total discounted benefits and the total discounted costs, and BCR represents the ratio of total discounted benefits to total discounted costs, can be defined as follows (Reniers et al., 2016; Liu et al., 2020).

$$NPV = \left(\sum_{t}^{Y} \frac{B_t}{(1+d)^t}\right) - \left(\sum_{t}^{Y} \frac{C_t}{(1+d)^t}\right) \tag{40}$$

$$BCR = \left(\sum_{t}^{Y} \frac{B_t}{(1+d)^t}\right) \Bigg/ \left(\sum_{t}^{Y} \frac{C_t}{(1+d)^t}\right) \tag{41}$$

where B_t represent the benefits in the year t, C_t represent the cost in the year t, Y is the time horizon in years, and d is the discount rate. The former term in Eq. 40 represents the total present benefits in years Y, and the latter term represents the total present costs in years Y. Likewise, the numerator in Eq. 41 represents the total present benefits in years Y, and the denominator represent the total present costs in years Y. In this analysis, year "1" represents the investment year in which benefits, and costs were not discounted to present value. Thereafter the yearly benefits and costs cash flows for the 5-year time horizon were discounted to present day values (i.e., year 1 values). The investment may be considered economically viable if the NPV is greater than zero (NPV > 0) or BCR is greater than 1 (BCR > 1) (Hsu, 2021). The NPV and BCR values for FastDetect can provide an indication of the economic viability and profitability and allows comparison of up-front and medium-term costs and benefits of investment (Ahopelto and Vahala, 2020; Reniers et al., 2016).

There can be different types of costs associated with installing a system such as FastDetect including one-off investment costs, installation costs, inspection costs, operation, and maintenance costs. These costs are represented by negative cash flows. Some costs such as one-off investment and installation costs occur in the present and thus do not have to be discounted, whereas other costs such as inspection costs, operation, and maintenance costs occur throughout the remaining life cycle timeframe and are discounted to present value. Similarly, there can be different types of benefits linked to a profitable investment of a FastDetect in this context such as process efficiency benefits, reduced water-energy costs, potential environmental benefits, and benefits which can be harder to monetise such as improved sustainability reputation, etc. The benefits can represent positive cash flows which occur throughout the remaining life cycle timeframe and were discounted to the present. The NPV and BCR values were calculated based on the total costs (non-recurring and recurring costs) and benefits acquired by the non-residential building. A discount rate of 3.5% was used for NPV and BCR calculations as per relatively recent energy and water related analysis (O'Callaghan and Prior, 2018; Lavappa and Kneifel., 2015; Medina et al., 2017). The capital cost of FastDetect was divided into different implementation phases (discussed in Chapter 7 in detail – Section 7.2).

The factors which were considered as inputs to the cost-benefit analysis were:

- Cost of maintenance operations.
- Cost of hardware and equipment required for FastDetect.
- Cost of labour to implement, operate, and maintain FastDetect.
- Other recurring costs such as software cost, water-energy costs, etc.

The following factors were not included as part of this cost-benefit analysis as they were either not applicable to the case-study or there was insufficient data to enable an accurate assessment.

- Benefits due to reduced performance monitoring equipment inventory (potentially relevant in this case but could not be estimated).
- Benefits due to environmental impacts (relevant but linked to reputational rather than direct financial benefits).
- Benefits due to system safety increase (potentially relevant in this case but could not be estimated).

6.5.2 Quantitative costs

The capital cost of FastDetect was considered as being €15,000 (broken down in **Table 6.*2*** - dashed line box) which included the cost and time involved in implementing FastDetect model in the non-residential site (**Figure 6.7**). Installation, set-up assistance, and software training were included in the capital cost of FastDetect. Based on the research experience with industrial facilities, regardless of the site size if data is available, it was assumed development takes the same amount of time (**Figure 6.7**) to develop and implement FastDetect model.

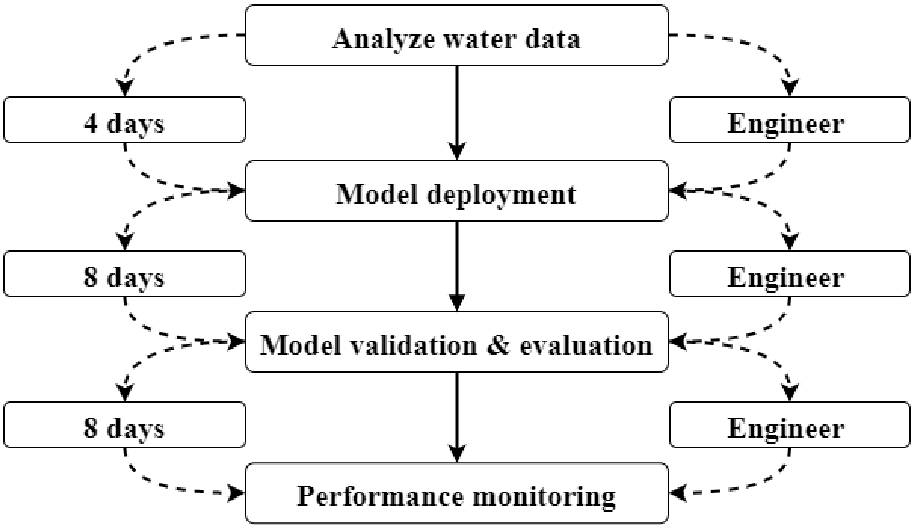


Figure 6.7: FastDetect implementation outline.

The impact that FastDetect will have on the overall cost of running the building water distribution system and the case-study is not easily quantifiable. For this purpose, the investment cost of FastDetect was divided into non-recurring and recurring costs.

6.5.2.1 Non-recurring costs

Non-recurring costs are one-time investment costs that are not expected to continue over time or at least not on a regular basis[6]. A one-off site inspection cost to perform feasibility and was taken as €300 (based on local consultation with NUI Galway Facilities Management and a commercial provider of water metering solutions[7] and existing literature where appropriate (Medina et al., 2017)). The cost was assumed as being a day at a rate of €300 per day[8] (i.e., it could take minimum one day or maximum 5 days depending on the scale and operational complexity of non-residential buildings) for this case-study. Site inspection activities involve physical inspection and inventory of all water use points within the facility and metering availability. The capital cost of FastDetect includes data acquisition, data analysis, model development and validation steps (**Figure 6.7**), while the cost for the sensors/meters varies with the size of the

[6] https://www.investopedia.com/.
[7] Personal communication with the building managers of non-residential buildings.
[8] Personal communication with the building managers of non-residential buildings.

facility[7]. For the considered case-study site (Alice Perry Engineering Building NUI Galway), the mains water network length (not including the pipes going to individual outlets) was estimated around 320 m in length per storey, covers four storeys of building. Based on the case-study layout, four sensors/meters installed at each floor would adequately cover the key water users. The capital cost of the water meter/sensor was taken as €1,200[9] which includes the purchase of the sensor/meter and other related hardware components required such as adaptors, cable ties, etc. The cost of the labour (technician) required to install meters/sensors where needed was taken at 2 hours per sensor/meter[7] based on the average salary of qualified technicians in Ireland at a cost of €33 per hour[10]. The non-recurring costs considered for the cost-benefit analysis are summarized in **Table 6.2**.

Table 6.2: Annualized non-recurring costs (Case-study 1).

S.No.	Description	Cost/annum
1.	Site inspection (1 day)	€300
2.	Sensor/meter - 4 (€1,200 each)	€4,800
3.	Sensor/meter installation (1.5 hours/sensor)	€264
4.	Data acquisition and analysis (4 days)	€2,000
5.	FastDetect model development (8 days)	€4,000
6.	FastDetect model validation (8 days)	€4,000
7.	Performance monitoring model (FastDetect)	€5,000
	Total estimated non-recurring cost per annum	€20,364

Note: Cost values in dashed line box represents the capital cost of FastDetect (€15,000).

6.5.2.2 Recurring costs

Recurring costs are usually general and administrative operating costs that occur at regular intervals, ongoing cost required to operate a system[11]. The cost of the engineering professional required to implement, update, or tune FastDetect model was considered as

[9] https://micronicsflowmeters.com/product/u1000mkii-fm-fixed-clamp-on-ultrasonic-flow-meter/.
[10] https://www.cso.ie/en/statistics/earnings/earningsandlabourcosts/.
[11] https://www.investopedia.com/.

€400 per day based on the average salary of engineers in Ireland[12]. FastDetect model maintenance and upgrades also comprise water meter calibration activities where required which can be achieved via temporary clamp-on meters to ensure correct reporting of water meters. This on-spot verification of water meters would assist in cutting down routine equipment maintenance cost which is usually taken as equivalent to 5% of the capital cost of the sensor/meter per annum (Rajan and Roylance, 1996; Hardeman, 2008). The annualized recurring cost for FastDetect model maintenance and upgrades considered twice a year as €800.

6.5.3 Quantitative benefits (cost savings)

Routine operation and maintenance of a non-residential water distribution system usually varies between sites from being rarely conduction to conducted at 3 monthly intervals that last for 3-4 days per interval (e.g., in the case of biomedical production facilities)[13]. For the considered case-study 1, operation and maintenance usually conducted twice a year (at 6 monthly interval). This could result in operation and maintenance costs of around €4,752 per annum based on the average salary of qualified technician in Ireland of €33 per hour[3]. With the fault detection approach, the routine operation and maintenance cost would be expected to be reduced by at least 50%, which in the scenario above would accrue benefits of about €2,376 per annum. The methodology used herein assumes that by increasing the early detection of faults, the likelihood of serious faults will decrease and lead to a reduction in overall operation and maintenance cost by at least 50%. Typically, large-scale non-residential facilities such as medical device manufacturers, production facilities, etc. require staff once a week[13] (8 hours per day) to perform data acquisition tasks and to examine the operating condition of water distribution systems which cost roughly €12,672 per annum based on the average salary of qualified technician in Ireland of €33 per hour[3]. Due to the unsupervised monitoring of non-residential water distribution system the staffing cost would be expected to be reduced by at least 50% and the benefits from reduced staffing turnover of roughly €6,336 per annum can be achieved as this process would be automated (once FastDetect is

[12] http://www.salaryexplorer.com/salary-survey.php?loc=104&loctype=1.
[13] Personal communication with Facilities Management in NUI Galway, a commercial provider of water metering solutions and facilities manager at a biomedical production facility. Personal communication with the maintenance engineers in non-residential water distribution systems.

implemented no labour would be required to perform manual data acquisition and inspection activities).

In a recent study, it was estimated that on average leakage, faults accounts for up to 3 faults annually depending on the size of the municipal water distribution system (Medina et al., 2017). In Ireland, the average cost to repair a fault in water distribution systems is around €250 (Irish Water, 2015). For the case-study sites investigated in this research, the number of fault occurrences varied between 1 and 6 faults per annum excluding metering error faults (faults from zero flow readings). It could be assumed that, on average, 3 faults per annum can occur in the water distribution system which cause unnecessary shutdown or unscheduled maintenance operation. Thus, this performance monitoring cuts maintenance costs by half and thus cuts faults by half through there being continuous oversight of the system. Hence, by early fault detection efforts and continous monitoring through FastDetect it is expcted that the fault occurances will be mitigated to minimium and the benefits of approximately €750 per annum can be achieved by reducing the downtime of the water distribution system.

As stated in Section 2.2, water loss through leakage could reduce water consumption by 10-30% through technological advancement and could increase water-energy benefits (European Commission, 2019; Ahopelto and Vahala, 2020). To analyse the water-energy benefits, the annualized water-energy cost was estimated by utilizing the case-study 1 site water-energy consumption data for the training period excluding faults (annualized water volume and associated energy use multiplied by the water-energy price statistics for non-residential buildings) which was around €14,000 per annum. This period was chosen as it contains minimal faults and more water-energy data with no missing data points. During the testing period, the fault caused 41% increased water-energy cost (as discussed in Section 6.4). If FastDetect was installed over the testing period in case-study 1, the excess usage fault would have been identified at an early stage and a 41% reduction in water-energy costs (benefits of around €5,740 – 41% of water-energy cost) would have been achieved over the course of that year. Other benefits would include (depending on the building) reduced monitoring equipment inventory, benefits due to process improvement, benefits due to reduced water network renovation, consumer behavioural benefits and environmental benefits, etc. Thus, improved water-energy efficiency can be achieved particularly in the period after the commissioning of FastDetect and repair of current faults. Thereafter as FastDetect becomes more robust additional faults and inefficiencies would be detected and new faults identified rapidly; all of which would improve non-residential water distribution system performance and sustainability. Furthermore, it can be assumed that in the absence of FastDetect the previous conditions that applied whereby faults were left unattended or undetected would reoccur and in this analysis are included as potential annual savings. The estimated annualized benefits for case-study 1 (Alice Perry Engineering Building NUI Galway) are summarized in **Table 6.3**.

Table 6.3: Annualized quantitative benefits (cost savings) - Case-study 1.

S.No.	Description	Benefits/annum	One-off avoided costs
1.	Estimated water-energy benefits	€5,740	
2.	Operation and maintenance benefits	€2,376	-
3.	Benefits due to reduced staffing	€6,336	€6,336
4.	Benefits due to reduced fault occurrences	€750	-
	Total estimated benefits per annum	€15,202	

Lastly, the NPV and BCR values were calculated to analyse the economic viability of a projected investment of FastDetect for case-study 1. The cost-benefit analysis and the result of the cost-benefit analysis is presented in **Table 6.4**, and a BCR of 1.98 and a NPV of €23,665 over the analysed testing period was estimated. The economic viability of FastDetect was analysed based on the cost-benefit analysis which includes potential savings each year as a result of a fault being fixed in year 1. Thus, if the fault were to persist the expected impact of the fault has been included as a saving. In some cases, in industry the cost benefit analysis may only include such savings in the year they occur. If such an analysis were applied the payback to users would increase to more than 2 years.

Table 6.4: Cost-benefit analysis (Case-study 1).

Cost-benefit analysis					**[FastDetect]**
	Quantitative costs				
Non-recurring costs	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Data acquisition and analysis costs	€2,000				
FDD model development costs	€4,000				
FDD model validation & evaluation costs	€4,000				
FDD model cost	€5,000				
Sensors/meters cost	€4,800				
Sensor/meter installation costs	€264				
Site inspection cost	€300				
Total non-recurring costs	€20,364				
Recurring costs	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
FDD model maintenance/upgrades costs	€800	€800	€800	€800	€800
Total costs	€21,164	€800	€800	€800	€800
Benefits	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Water-energy benefits	€5,740	€5,740	€5,740	€5,740	€5,740
Benefits from less faults	€750	€750	€750	€750	€750
Annual O&M benefits	€2,376	€2,376	€2,376	€2,376	€2,376
Reduced staffing turnover benefits	€6,336				
Total benefits	€15,202	€8,866	€8,866	€8,866	€8,866
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Total present costs	€21,164	€773	€747	€723	€697
Total present benefits	€15,202	€8,566	€8,276	€7,997	€7,726
Net present value (NPV)	€23,665				
Benefit cost ratio (BCR)	1.98				

The estimated positive values of NPV and BCR indicates that, in line with personal communication with facility managers, the BCR for fault detection in water systems may be low in an Irish context given the low cost of water. However, facility managers may wish to invest in such technology as they are required to report on water consumption and demonstrate efficiencies are being achieved.

6.5.4 Sensitivity analysis

To analyse the impact of relevant economic parameters on the calculated BCR and NPV, and the uncertainty built into some of the estimated costs and benefits discussed previously, a sensitivity analysis was carried out. For each parameter analysed an increase and decrease of 10% and 20% in the parameter value was analysed. The parameters analysed include the assumed recurring cost of FastDetect maintenance and upgrades, water-energy savings, operation, and maintenance benefits, benefits due to reduced fault occurrences and benefits due to reduced staffing costs. The results of the sensitivity analysis are shown in **Figure 6.8**.

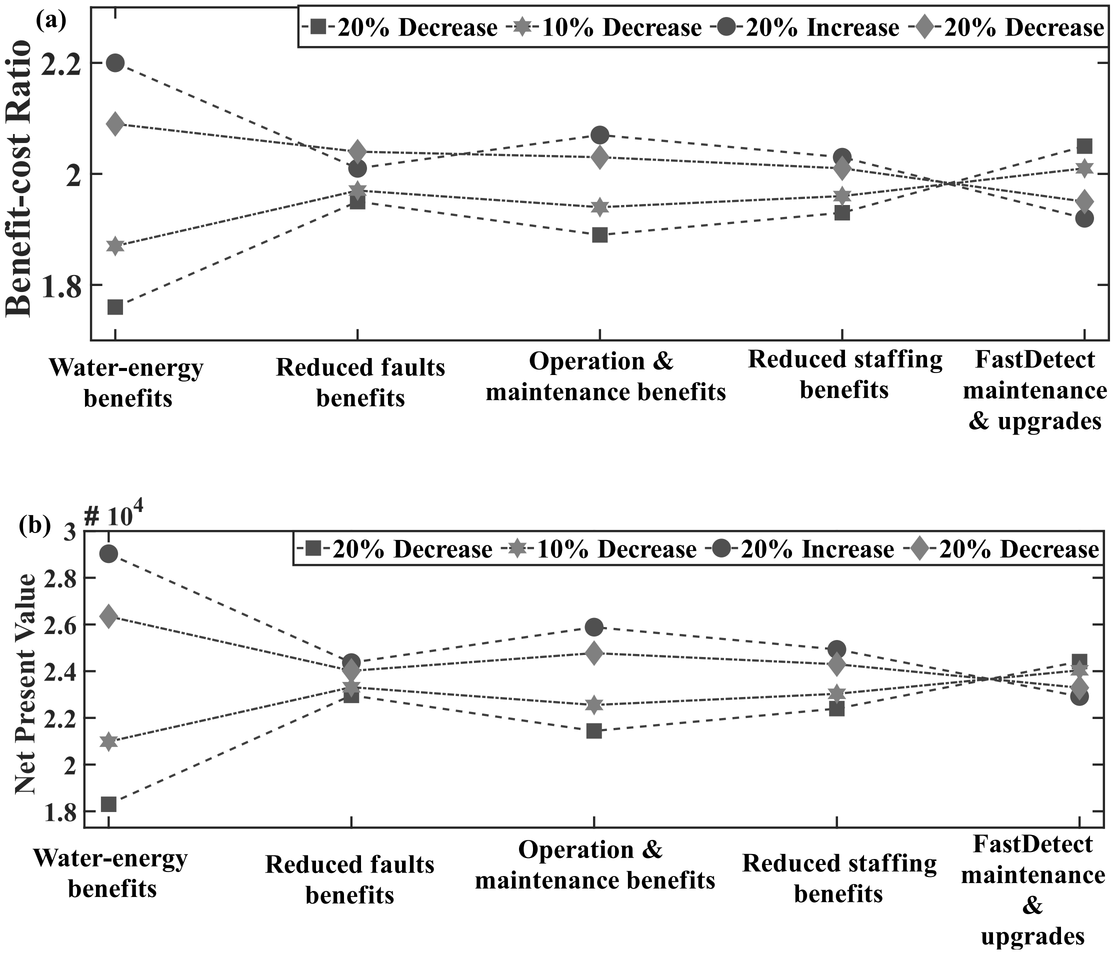


Figure 6.8: Sensitivity analysis, (a) BCR and (b) NPV - Case-study 1.

The assumed water-energy benefits were observed to have the most influence on BCR and NPV, followed by operation and maintenance benefits and FastDetect maintenance and upgrades cost parameters. A 20% increase in operation and maintenance benefits would lead to a 5% increase in BCR (2.08), whereas a 20% decrease would lead to a 4%

drop in BCR (1.89). Correspondingly, a 10% increase in operation and maintenance benefits would lead to a 3% increase in BCR (2.04), whereas a 10% decrease would lead to a 2% drop in BCR (1.94). A 20% increase and decrease in water-energy benefits would lead to ±11% change in BCR (from 2.2 to 1.76). Likewise, a 10% increase and decrease in water-energy benefits would lead to a ±6% change in BCR (from 2.09 to 1.87). A 20% decrease in recurring cost of FastDetect maintenance and upgrades would lead to a 4% increase in BCR (2.05), whereas a 20% increase would lead to a 3% drop in BCR (1.92). Similarly, a 10% decrease in recurring cost of FastDetect maintenance and upgrades would lead to a 2% increase in BCR (2.01), whereas a 10% increase would lead to a 2% drop in BCR (1.95). Whilst in the parameters (reduced fault occurrences and reduced staffing benefits), a relatively small change in BCR (approx. ±3%) has been observed.

Correspondingly, a 20% increase in operation and maintenance benefits would lead to a 10% increase in NPV (€25,885), whereas a 20% decrease would lead to a 9% drop in NPV (€21,441). A 10% increase and decrease in operation and maintenance benefits would lead to a ±5% change in NPV (from €24,773 to €22,553). A 20% increase and decrease in water-energy benefits would lead to ±22% change in NPV (from €29,029 to €18,300). Likewise, a 10% increase and decrease in water-energy benefits would lead to a ±11% change in NPV (from €26,347 to €20,982). A 20% increase and decrease in recurring cost of FastDetect maintenance and upgrades would lead to a ±3% change in NPV (from €22,917 to €24,412). Similarly, a 10% increase and decrease in recurring cost of FastDetect maintenance and upgrades would lead to a ±2% change in NPV (from €23,291 to €24,038). Whereas, in the parameters (reduced fault occurrences and reduced staffing benefits), a relatively small change in BCR (approx. ±4%) has been observed. The results of sensitivity analysis indicate that the water-energy benefits, operation and maintenance benefits and recurring cost of FastDetect maintenance and upgrades play a crucial role in the economic viability of FastDetect.

6.6 Conclusion

This chapter conducted a comprehensive investigation of the hot water energy use, on-site pumping energy use and economic aspects of water-energy interaction in a large building. This analysis is then used to develop a cost-benefit analysis for the implementation of FastDetect in industrial settings. During the testing period, an average of 236 kWh/day water heating energy was utilised to overcome the stand-by energy losses and to maintain the water temperature in the water network. The study found that water heating accounted for up to 18% of the total heating energy use in the building. The fault in the testing period caused a 38% increase in on-site pumping energy use resulted in overall 41% increase in water-energy cost which was due to the continuous operation of booster pumps to maintain a constant supply of cold-water to the roof top tanks for the grey-water uses in the building.

To examine the economic feasibility of FastDetect implementation in non-residential water distribution systems, a cost-benefit analysis was carried out. The estimated NPV and BCR were €23,665 and 1.98 respectively over a 5-year time horizon indicates the economic viability of FastDetect in this case-study. Furthermore, a sensitivity analysis was conducted to study the impact of economic parameters and uncertainty linked with calculated costs and benefits. Water-energy benefits, operation and maintenance benefits, and recurring costs associated with FastDetect maintenance and upgrades were found to be key in impacting the economic viability of FastDetect. However, some of this is due to the low cost of water in Ireland compared to other countries and efficiencies in implementation such as installation and running costs would improve the overall NPV and BCR. The NPV and BCR results indicated that adopting FastDetect in non-residential water distribution systems is economically viable and could introduce significant economic benefits by increasing the overall feasibility and profitability of the system, followed by the mitigation in the environmental impacts.

7. APPLICATION OF FastDetect IN INDUSTRIAL SETTINGS

7 Application of FastDetect in Industrial Settings

7.1 Overview

The chapter discusses how FastDetect can be applied in the context of performance monitoring in non-residential facilities including those with limited/poor quality data and those with plentiful/accurate data. FastDetect is particularly suited to non-residential buildings which tend to have complex water distribution systems. The methodology utilises existing energy management principals or newly proposed water management principles such as Assess-Plan-Do-Check-Act protocols. It integrates the principles of ISO 50001 (Energy Management Programs), ISO 50002 (Energy Audits/Diagnosis), IPMVP (International Performance Measurement & Verification Protocol) and ISO 14046 (Water footprint).

7.2 Application of FastDetect

The application of FastDetect is a four-phase process (**Figure 7.1**) (i) assessment and feasibility study, (ii) planning measures to implement FastDetect, (iii) FastDetect implementation and performance evaluation, and (iv) undertaking performance optimization using performance monitoring outcomes.

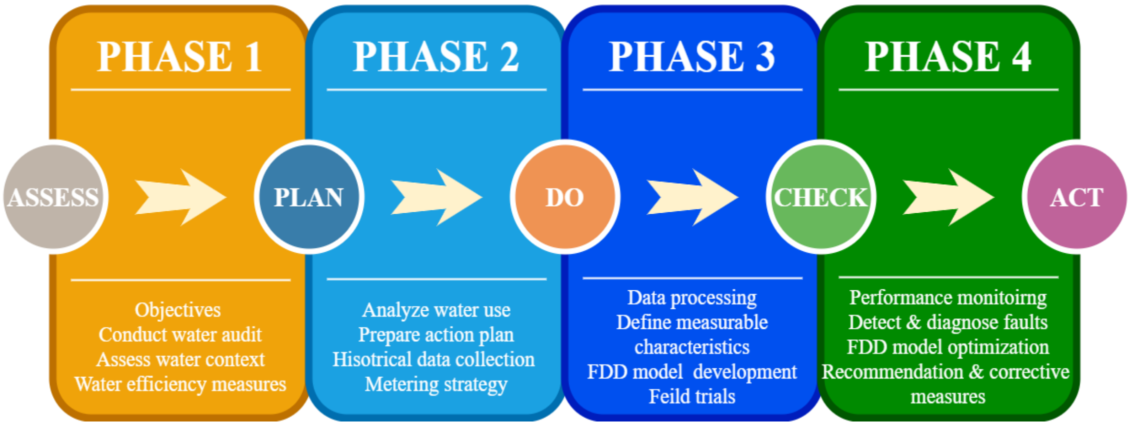


Figure 7.1: Application phases of FastDetect.

7.2.1 Phase 1 - Assessment and feasibility study

The foremost goal of this phase is to become familiar with the building water network in question and associated systems. This phase enables understanding of the issues in the water distribution systems as a first step towards engaging with fault management and water efficiency. The activities that make up this phase include information gathering by conducting a water audit, defining objectives, and outlining potential water efficiency measures.

A water audit may involve physical inspection of the water network and system to develop an inventory of all water use points within the facility and metering availability. This can be carried out by conducting a preliminary water audit in correspondence with the building or facility manager. Such a water audit can provide decision makers and stakeholders with a detailed profile of their water use pattern and system, and highlight the areas endure higher fault probability (i.e., where equipment malfunctioning, network faults, and procedural errors could lead to inefficiencies). Thus, a water audit serves as an important step towards improved performance monitoring, water efficiency, optimization, and their corresponding economic savings. Some of the high-level sources that can be queried while conducting a water audit include.

- Facilities managers or other technical staff responsible for the building water distribution system.
- Previous studies conducted on the building water distribution system.
- As-built drawings for the building water distribution system and a review of any updates to the system since its installation.
- Existing metering and data collection services in place.
- Any common issues with the building water distribution system such as recurring faults, or increased water bills at certain times of year.

The queries that can be made are not limited to those stated above as each non-residential building differs in its management, water uses, etc. A water audit may uncover faults, which were not identified previously. The faults found during the water audit would be considered when preparing action plan for the implementation of FastDetect so that at the very least, the potential for recurring faults will be eliminated. Also, employing a water audit across non-residential water distribution systems may streamline regulatory data reporting processes. By the application of an audit, it is expected that data collection methods would not only improve but would also improve throughout the duration of performance monitoring due to the inclination for data to be collected more frequently.

Through standardised reporting of data, it may be possible to create an automated means of reporting data to regulatory bodies.

In conclusion, this phase will assist the decision makers, stakeholders to understand the real issues and potential opportunities in achieving improved performance monitoring and water efficiency. Once engaged in improving performance monitoring and water efficiency, the end-user enters a continuous iterative cycle of improvement guided by the PDCA cycle (Plan-Do-Check-Act).

7.2.2 Phase 2 - Planning measures to implement FastDetect

The goal of this phase is to prepare an action plan for the implementation of FastDetect based on the water audit outcomes conducted in Phase 1. The plan must consider the complexity, and operational/water usage cycles of the non-residential facility. For instance, water use for an industrial facility may be independent of external factors such as seasonal effect, staffing, etc., a short monitoring plan of 3-5 months with spot measurements could be sufficient to develop a FastDetect model. In contrast, an institutional facility such as school, universities, etc. that has seasonal effects due to both weather and staffing (school in or out of session) may require a longer monitoring plan (often a year) to develop a FastDetect model.

This will provide an accurate representation of water usage patterns for FastDetect model development and will inform the degree of equipment (water meters) installation required at a building water distribution system (additional metering if required). As many non-residential buildings do not have extensive (or any) metering, retrofitting of meters in strategic places of the water network will likely be necessary. If additional metering is required a decision on meter type is needed between using intrusive or non-intrusive meters. The common trade-off between intrusive and non-intrusive ultrasonic water meters is that intrusive meters tend to cheaper to purchase but more expensive and disruptive to install. While non-intrusive meters are relatively expensive but easier to install as a retrofit. For this reason, the minimum number of meters that allow access to most amount of information is desirable (Clifford et al., 2018). A crucial aspect of installing water meters is ensuring good data acquisition, however, as has been shown in Chapters 4 and 5, FastDetect can operate successfully with gaps in training data – accurate data measurement, and collection will improve overall efficiency. In the meantime, historical data collected during the water audit can be utilised where available to provide additional information (historical data can only be incorporated if the accuracy of the datasets is assured in each non-residential water distribution system).

7.2.3 Phase 3 - FastDetect implementation and performance evaluation

The goal of this phase is to develop and implement FastDetect for performance monitoring of building water distribution systems. Where it is established that performance, assessment is feasible, FastDetect can be implemented based on the action plan developed in Phase 2. In this phase, collected historical time series data is sampled into working and non-working hours' time series to get deeper understanding of building water distribution system's performance and factors which may affect it. The data is then pre-processed to detect any invalid or missing data points. In some cases, it may be possible, and desirable interpolate incorrect data or missing data using statistical techniques such as regression estimates, nearest neighbourhood, moving average, etc. Thereafter, statistical features (measurable characteristics) can be derived from the collected data based on the protocols defined in FastDetect. The statistical features play a vital role in analysing the water usage baseline and diurnal patterns (or indeed patterns at any chosen time interval) and can assist in distinguishing routine and non-routine water uses in the building water distribution system. FastDetect can be developed and evaluated using data with known faults or non-routine water uses and model sensitiveness in detecting and diagnosing faults can be tuned accordingly.

FastDetect can adapt to changing on-site conditions in detecting and diagnosing non-routine water uses or faults arising during routine operation. In particular, practical issues associated with fault classification can be investigated. Usually, the amount of fault information (labelled data points or measurements) needed for fault diagnosis is of a magnitude higher than for detection only. which and in what form information should be provided to the end-users/building managers.

7.2.4 Phase 4 - Performance monitoring optimization

The goal of this phase is to develop innovative measures for optimizing performance monitoring of building water distribution systems. Performance monitoring of non-residential water distribution systems can be improved by utilizing the knowledge of previously experienced fault situations. That can be stored in a FastDetect database and can be referred as a fault together with the details of the action that was taken to respond to the fault situation. In FastDetect model development phase, usually a limited number of faults are foreseen and can be stored in the database as fault models. However, in practice additional and unforeseen faults can occur. Being able to monitor unforeseen faults and transform them into new fault models in the database provides stakeholders with an opportunity of getting a more robust performance monitoring system. As the number of fault situations in the database increases, the system becomes more reliable and valuable in time. A FastDetect engineer would be able to collect the monitoring data and can derive new fault models. The detailed process cycle of designed performance monitoring framework is shown in **Figure 7.2**.

As such it is intended not only as a performance monitoring measure but as an opportunity to further improve stakeholder's awareness regarding water challenges and issues. According to the ISO50001 standard, a fundamental principle is that organizations are not only capable of identifying and fixing the problems, but also taking actions to eradicate the source that caused the problem. In this regard, this phase may also involve a description of follow-up activities, monitoring, and measurement of results as well as a description of responsibilities.

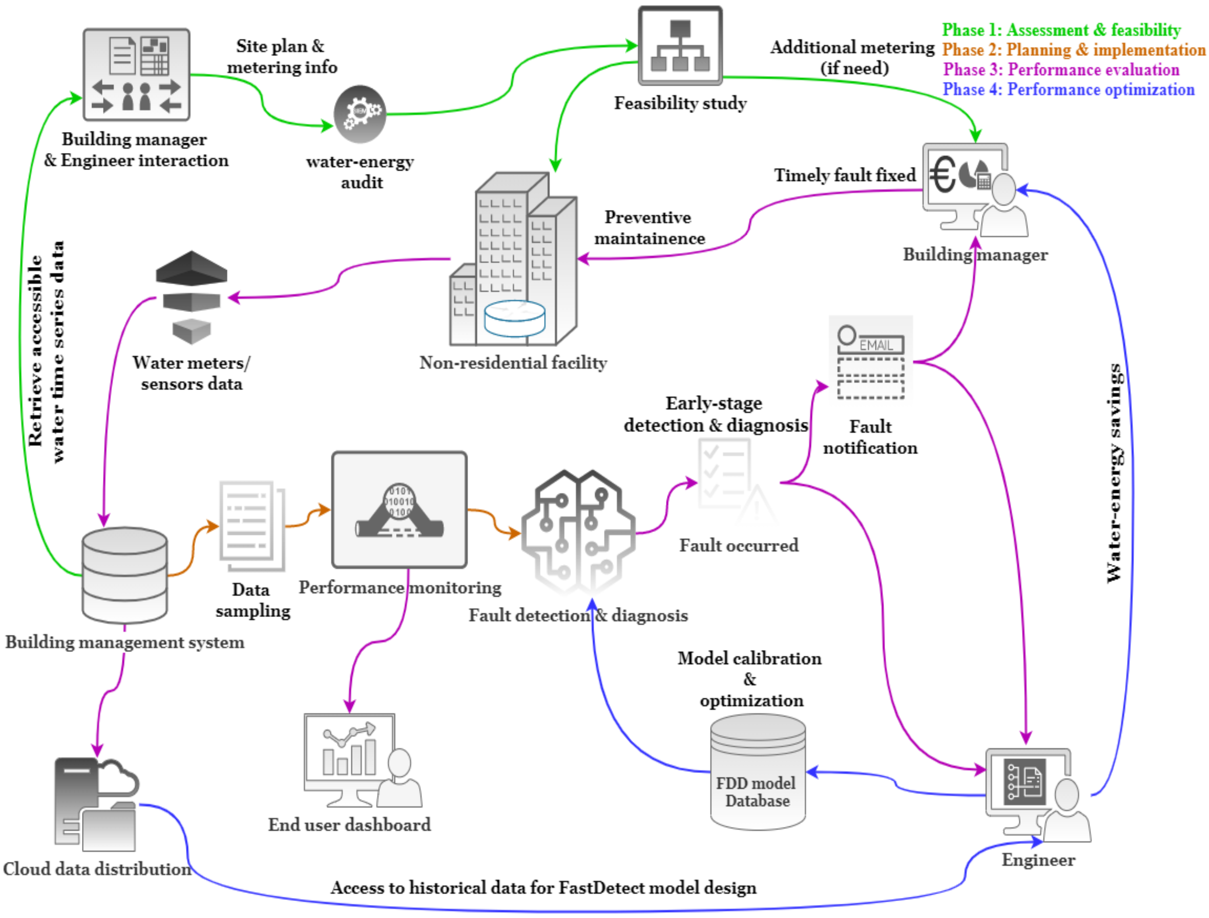


Figure 7.2: Designed performance monitoring framework - process cycle.

7.3 Platforms for FastDetect application

FastDetect was developed and implemented in MATLAB during this study, however, the implementation of FastDetect at a wider level would require more automation and capability in terms of storing databases and reporting performance monitoring results in a user-friendly fashion.

7.3.1 Standalone software package

The most straightforward application of FastDetect would be within a bespoke standalone software application. In this case, the modules used in FastDetect can be programmed into a software application without requiring adaptation (which may be required for integration into existing building management systems). Development of a standalone software package can increase the usability and robustness of FastDetect when compared to its application in an existing platform such as MATLAB which was used in this study. The high-level standalone FastDetect architecture is shown in **Figure 7.3**.

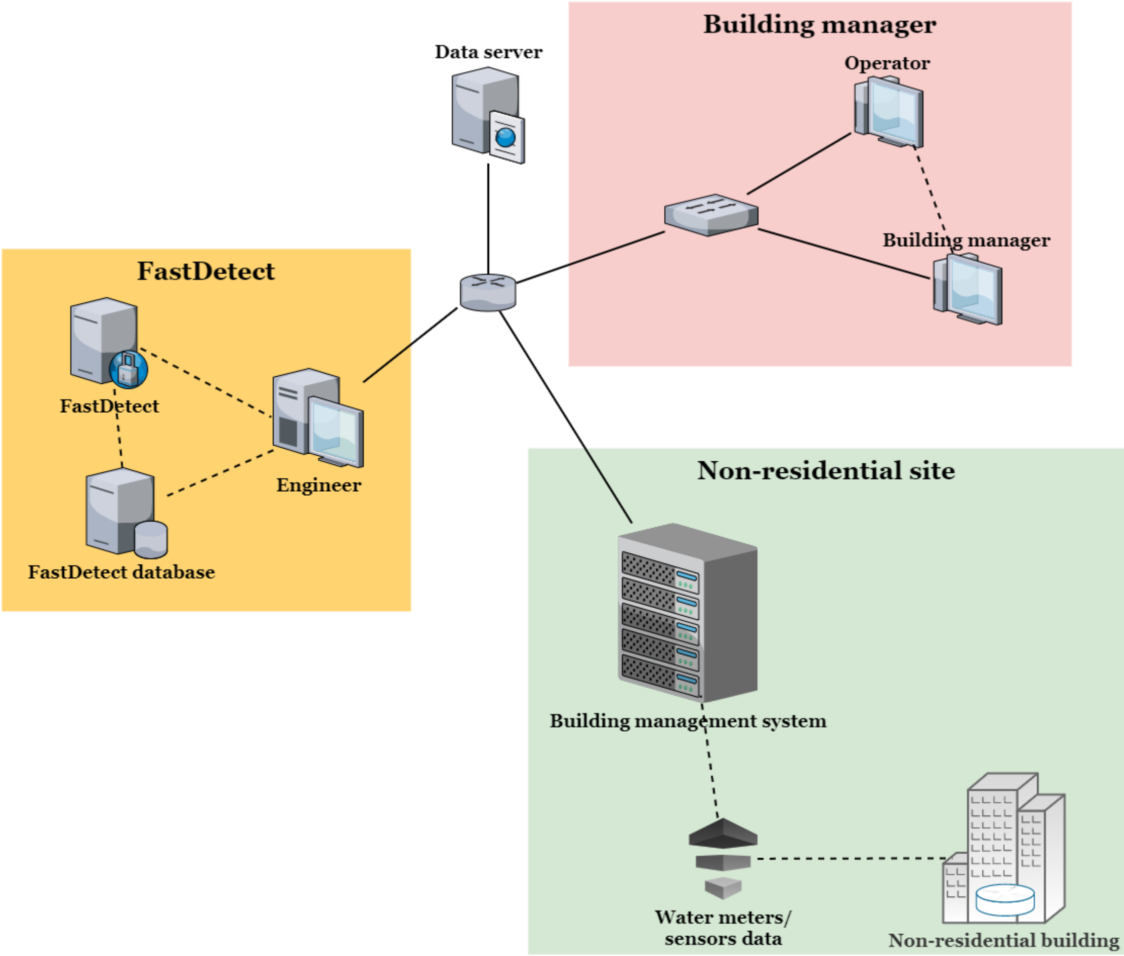


Figure 7.3: High-level standalone FastDetect architecture.

While MATLAB provides a suitable platform for the implementation of FastDetect, it could be difficult for end-users to import and analyse data, it may require additional training for the end-user, may not provide the expected levels of user interface and can require the purchase of additional software licences. The drawbacks of developing bespoke standalone software application include the costs and the subsequent increase in software training needs at a facility level. Therefore, it is important to adapt it to the needs of the end-users. The integration of FastDetect into existing building management systems may be more practical in buildings with a preinstalled building management system.

7.3.2 Integration to existing building management systems

In many cases stakeholders need to be convinced regarding benefits of adapting performance monitoring tools for water systems (as water is not often economically costed unlike electricity). Besides, many end-users presume that a building management system should already provide FDD tools, even if currently they do not[14]. Therefore, the first step is to make the stakeholder confident about using the performance monitoring platform. Secondly, the platform should be presented as something that will help to transform large amounts of data usually stored in the building management system, and which are not currently used effectively, into more useful knowledge that can have positive commercial, operational and sustainability impacts. As building management systems can be usually found in non-residential buildings, the integration of FastDetect can be the most effective means of applying performance monitoring on building level (**Figure 7.4**).

[14] Personal communication with the maintenance engineers in non-residential water distribution systems.

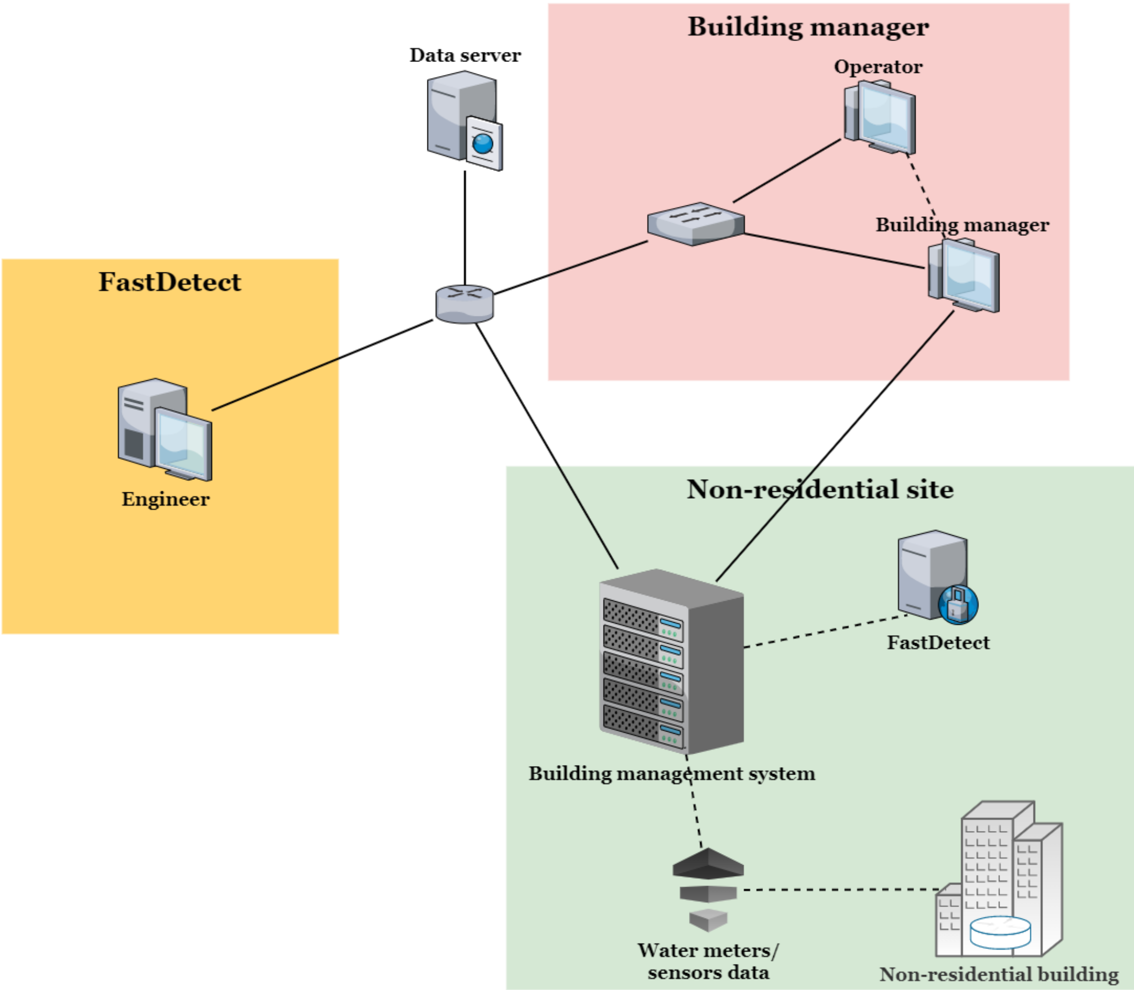


Figure 7.4: High-level integrated FastDetect architecture.

The benefits of integrating FastDetect into existing building management systems include data sharing between the building management system and FastDetect potentially eradicating the need to manually retrieve water data. Typically, data from different meters such as water meters, energy meters, etc. is often fed automatically into a building management system, thus, it acts as the first step in accumulating operational data into a single database. These characteristics can be of great benefit to performance monitoring, data and can be automatically fed into FastDetect. Building management system integrated with FastDetect potentially offer a means of better understanding the overall building water distribution system operation, the equipment used for performance monitoring, and could reduce the human resource involved in operation and maintenance practices.

7.3.3 Advantages of FastDetect over existing performance monitoring methods

FastDetect proposed some of the key advantages over existing methods.

- FastDetect is simple and/or easy to understand, to explain and to commission. The parametrisation of FastDetect requires only a short training period and training data (historical water use data) to develop a FastDetect model for performance monitoring.

- The configuration of FastDetect to a specific site or system is simple and straightforward. FastDetect relies on the use of water data from data management system or from installed water meters and eradicates the need for detailed knowledge of the physical layout of the building water distribution system. FastDetect will also incorporate existing operator knowledge of the building water distribution system (where available), historical data and knowledge of previous faults. Such knowledge is very common among building managers in large-scale facilities and buildings.

- FastDetect can be tailored to each non-residential facility and to the specific user (building managers, practitioners, etc.). The user can interface/interfere actively or passively with a limited training.

- FastDetect is easy to embed and to integrate in a building management system. FastDetect only needs to access data sets that already exist (water consumption data and overall knowledge of the facility as provided by the building management system) and can be used to train FastDetect model. Then as a facility develops more data, FastDetect model can be adjusted or can self-adjust.

- FastDetect is adaptable with new fault situations in a modular way. Once a new fault or anomalous situation is identified, FastDetect model can be updated to integrate this additional knowledge. This adaptation will make FastDetect reactive to planned changes in water usage on site.

- FastDetect integrated with false alarms moderation offers reduced false alarm rate when compared to traditional methods during performance monitoring of non-residential water distribution systems making FastDetect more valuable and reliable in time.

7.4 Conclusion

In this chapter application of FastDetect in industrial settings and two different platforms (standalone software package and integration to building management systems) that can be utilized to implement it at a commercial scale were discussed. The process described utilises concepts from the energy sector to systematically setup a standardized protocol so that FastDetect can be implemented in a wide range of non-residential facilities to aid in the conservation of water. Utilizing the basic steps of IS0 50001 and IS0 50002 the process facilitates the crucial initial steps of developing a performance monitoring platform; notably “Assess”, “Plan”, and “Do”. The application of FastDetect will form a versatile, robust, and adaptable performance monitoring system to mitigate the effects of faults within building water distribution systems in Irish settings. The application of FastDetect will also facilitate improvement in the data management and data collection practices.

FastDetect can be applied at different platforms representing promising opportunities. A standalone software package can increase the robustness of the performance monitoring of non-residential water distribution systems without the need of adaptation (which may be required for integration to existing technologies like building management system). Wherever, integration of FastDetect into building management systems can increase the usability and automation of non-residential water distribution system performance monitoring.

8. CONCLUSIONS, AND RECOMMENDATIONS

8 Conclusions, and Recommendations

8.1 Overview

Large non-residential buildings can contain complex and often inefficient water distribution systems. As public and private sectors strive towards more sustainable use of water increase it has become increasingly important to effectively detect and diagnose faults in water distribution system in large buildings. In many cases, if water supply is not impacted, water loss can go unnoticed for long periods. This can lead to unnecessary increases in water usage and associated energy losses arising from water pumping, treating, and heating. Most of the FDD studies in the water sector are limited to municipal water supplies and leakage detection. The application of FDD in building water networks remains largely unexplored.

Chapter 2 revealed a number of important shortcomings and challenges in existing literature related to the performance monitoring of the non-residential water distribution systems in identifying non-routine water uses and faults of different types (not only leakage) within the system. The non-stationarity of water use data in non-residential buildings results in high incidence of false alarms when conventional fault detection approaches are used. This high incidence of false alarms is one of the main reasons that makes building managers and stakeholders reluctant to add FDD methodologies to existing monitoring system.

Existing and newly developed FDD technologies to detect and diagnose faults in building water distribution systems are not entirely effective and useful, require capital investment, are labour-intensive, and time consuming. In this research, a novel FDD methodology termed as “FastDetect” was designed and validated against two different non-residential case-study sites, addressing shortcomings and challenges in the existing literature, respectively.

8.2 Conclusions

The main conclusions are as follows.

- FastDetect was developed and demonstrated on two non-residential case-study sites and proved to be effective in detecting and diagnosing faults at their early stage which conventional systems missed. FastDetect was also effective in limiting false alarms (even before the false alarm moderation).

- One of the most significant current challenges in the performance monitoring of non-residential water distribution systems is related to limited data availability and data with incomplete information. Thus, it is important to implement robust FDD methodologies that can operate under these practical data constraints. The performance of FastDetect is promising given that it can be applied to sites with limited data and with incomplete information and be successful.
- FastDetect was capable of handling highly variable water consumption data and outliers using a distance-distance approach (described in Section 3.3.2.1 – Chapter 3). In FastDetect, the data with large variations and outliers were assessed and localized using the distance-distance approach which computes the distances from each data point to the centroid of the dataset in the PCA space both in the PCA plane and orthogonal directions (shown in Section 4.4.1 – Chapter 4). FastDetect demonstrated a robust means of dealing with outlier measurements by extracting valuable information from the data, while stabilizing the over sensitiveness of the PCA model.
- This study utilized temporally sensitive unsupervised and supervised learning models (namely PCA and multi-class SVM), capable of identifying complex patterns within data, whereby PCA was combined with detection indices (T^2 and Q-statistics) to detect faults in the incoming data and the multi-class SVM was trained for fault classification. A greater number of true faults were detected (in some cases faults were detected earlier) when compared to existing in-situ conventional fault detection system (univariate approaches). The PCA model outperformed the conventional systems, detected averaged 26% more faults in two case-studies. The multi-class SVM also allowed the faults to be classified (showed over 90% accuracy in classifying faults), providing a greater level of information to building managers, which can avoid unnecessary emergency shutdown in industrial applications. FastDetect showed good capability in detecting and diagnosing faults in complex non-residential water distribution systems.

- To control the prevalence of false alarms, this study utilises two non-parametric false alarm moderation approaches (namely window-based approach and trial-based approach) for the first time in water sector. The implementation of these false alarm moderation approaches was combined with PCA to detect true faults. Using PCA with the window-based approach averaged 54% false alarms were reduced in two case-studies, and the overall performance and reliability of FastDetect was improved.

- The energy use for water heating and water pumping for case-study 1 were examined. On average, 236 kWh/day water heating energy was used to overcome the stand-by energy losses and to maintain the water temperature in the water network. The study found that water heating accounted for up to 18% of the total heating energy use in the building. For water pumping energy, an increase of 38% was observed for grey-water and other cold-water uses due to the fault that had occurred in the case-study 1.

- The economic aspects of water-energy interaction associated with the faults in a non-residential water distribution system were investigated. The 41% increase in water-energy costs due to the faults in case-study 1 demonstrated the significance of considering energy use related to water end uses in large-scale non-residential building water distribution systems and the need for using holistic approach to integrate water-energy losses.

The cost-benefit analysis was conducted to examine the economic feasibility of FastDetect implementation in non-residential water distribution systems. The NPV and BCR values of €23,665 and 1.98 over a 5-year time horizon was estimated provides positive economic viability of FastDetect. To examine the impact of economic parameters and uncertainty linked with estimated costs and benefits, a sensitivity analysis was carried out. The results indicated that the water-energy benefits, operation and maintenance benefits, and recurring costs associated with FastDetect maintenance and upgrades were found to be key in impacting the economic viability of FastDetect. The conclusions above point to the promising capabilities of FastDetect when compared to conventional fault detection systems. FastDetect has been developed using data from case-study sites with various faults and outliers whereas, typically existing studies utilise generated or experimental data sets. FastDetect is capable of considering biased uncertainties and non-stationarity of water use, addressing the key challenges in literature. FastDetect showed usefulness in identifying non-routine water uses and faults of different types in addition to leakage as shown in two case-study sites, while considerably reducing the false alarms during the fault detection process when compared to conventional univariate approach.

The early detection of faults provides a greater level of system information to building managers making such facilities more sustainable. Another important characteristic is adaptability. FastDetect is easy to embed and can integrate into the existing water data collection/monitoring systems. Besides, the water loss as a resultant of faults implicitly leads to the loss of energy resource representing an increased life cycle carbon footprint. The economic aspects of extrapolation of such losses on a larger scale would suggest substantial economic and environmental impacts.

8.3 Recommendations for future work

The work carried out in this thesis has emphasized a number of areas where future research is recommended as follows.

- The temporally sensitive unsupervised and supervised learning models (PCA and SVM) showed advantages in working without the geometrical information of the system and showed high classification accuracy. Future research should involve improving some practical aspects of the model, such as improving fault classification through enhanced training datasets. This will likely be more challenging as data from known faults could be difficult to obtain due to the infrequency of such events and the fact that introducing faults to functional systems for the purposes of data collection is impractical.

- To enhance the robustness and applicability of FastDetect, further investigation should involve testing and validating the approach against water distribution system data from different non-residential buildings and projects. Another avenue would be to explore relationships within the data that can be used to define additional statistical features for water event disaggregation.

- FastDetect is adaptable in that it can diagnose new faults which have not been experienced previously which can be verified in future work. Once a new fault or anomalous situation is identified, the FDD model can be updated to incorporate this additional information. Adaptation will make the system more reliable and valuable in time and allow for detection and diagnosis of further anomalies in the building water distribution systems. This can enable building operators to target predictive maintenance and reduce the occurrence of false alarms, while also developing strategies to reduce unnecessary or inefficient consumption of water.

- Further research could support the evolution of FastDetect into a set of standards, or trusted evaluation procedures. Evaluation results from distinct non-residential water distribution systems can be used to develop a performance evaluation framework, or a set of standards by synthesizing with industry domain expertise. It would enhance the industry's understanding of the trade-offs inherent in FastDetect performance evaluation and the desired form and content of outcomes.

- High priority longer-term efforts involve research to estimate fault prevalence, environmental impact, and the associated cost.

- To this end, the further investigation should involve into user and stakeholder expectations for FastDetect performance and comparative analysis, development of universally available domain data-driven performance monitoring platform that can facilitate individual authorities to analyse their water distribution system performance.

REFERENCES

References

Abdi, Hervé, and Lynne J. Williams. 2010. “Principal Component Analysis.” *Wiley Interdisciplinary Reviews: Computational Statistics* 2(4): 433–59.

Abdulshaheed, A., F. Mustapha, and A. Ghavamian. 2017. “A Pressure-Based Method for Monitoring Leaks in a Pipe Distribution System: A Review.” *Renewable and Sustainable Energy Reviews* 69(May 2015): 902–11. http://dx.doi.org/10.1016/j.rser.2016.08.024.

Ahmad, Tanveer, Huanxin Chen, Yabin Guo, and Jiangyu Wang. 2018. “A Comprehensive Overview on the Data Driven and Large Scale Based Approaches for Forecasting of Building Energy Demand: A Review.” *Energy and Buildings* 165: 301–20.

Ahmed, M., M. Baqqar, F. Gu, and A. D. Ball. 2012. “Fault Detection and Diagnosis Using Principal Component Analysis of Vibration Data from a Reciprocating Compressor.” *Proceedings of the 2012 UKACC International Conference on Control, CONTROL 2012* (September 2012): 461–66.

Ahopelto, Suvi, and Riku Vahala. 2020. “Cost-Benefit Analysis of Leakage Reduction Methods in Water Supply Networks.” *Water (Switzerland)* 12(1).

Ait-Izem, Tarek, M. Faouzi Harkat, Messaoud Djeghaba, and Frédéric Kratz. 2018. “On the Application of Interval PCA to Process Monitoring: A Robust Strategy for Sensor FDI with New Efficient Control Statistics.” *Journal of Process Control* 63: 29–46.

Akil, Mostafa, Pierre Tittelein, Didier Defer, and Frédéric Suard. 2019. “Statistical Indicator for the Detection of Anomalies in Gas, Electricity and Water Consumption: Application of Smart Monitoring for Educational Buildings.” *Energy and Buildings* 199: 512–22.

Al-Qawasmi, Jamal, Muhammad Asif, Ahmed Abd El Fattah, and Mohammad O. Babsail. 2019. “Water Effciency and Management in Sustainable Building Rating Systems: Examining Variation in Criteria Usage.” *Sustainability (Switzerland)*

11(8).

AL-Washali, Taha et al. 2020. “Assessment of Water Losses in Distribution Networks: Methods, Applications, Uncertainties, and Implications in Intermittent Supply.” *Resources, Conservation and Recycling* 152(January): 104515. https://doi.org/10.1016/j.resconrec.2019.104515.

Almandoz, Javier et al. 2005. “Leakage Assessment through Water Distribution Network Simulation.” *Journal of Water Resources Planning and Management* 131(6): 458–66.

Alsaydalani, M O A. 2017. “Influence of Cavitation Phenomenon on the Hydraulic Behavior of Leaks in Water Distribution Systems.” *Journal of Water Supply: Research and Technology - Aqua* 66(5): 327 LP – 339. http://aqua.iwaponline.com/content/66/5/327.abstract.

Ammiche, Mustapha, and Abdelmalek Kouadri. 2017. “Constant False Alarms Rate for Fault Detection.” *2017 5th International Conference on Electrical Engineering - Boumerdes, ICEE-B 2017* 2017-Janua: 1–7.

Anjana, G. R., K. R. Sheetal Kumar, M. S. Mohan Kumar, and Amrutur Bharadwaj. 2015. “A Particle Filter Based Leak Detection Technique for Water Distribution Systems.” *Procedia Engineering* 119(1): 28–34. http://dx.doi.org/10.1016/j.proeng.2015.08.849.

Arregui, F J, and M Carlos. 2012. “Burst Detection in Water Networks.” 138: 47–54.

Aşchilean, Ioan et al. 2017. “Determining Priorities Concerning Water Distribution Network Rehabilitation.” *Energy Procedia* 112(March): 27–34. http://dx.doi.org/10.1016/j.egypro.2017.03.1055.

AWWA. 2016. “Residential End Uses of Water, Version 2.” : 16.

Bagheri, Mohammad Ali, Gholam Ali Montazer, and Sergio Escalera. 2012. “Error Correcting Output Codes for Multiclass Classification: Application to Two Image Vision Problems.” *AISP 2012 - 16th CSI International Symposium on Artificial Intelligence and Signal Processing* (Aisp): 508–13.

Bakdi, Azzeddine, and Abdelmalek Kouadri. 2017. “A New Adaptive PCA Based Thresholding Scheme for Fault Detection in Complex Systems.” *Chemometrics and Intelligent Laboratory Systems* 162: 83–93. https://www.sciencedirect.com/science/article/pii/S0169743916302015?via%3Dihub (August 29, 2019).

Balaras, Constantinos. 2017. “Energy Use Intensities for Non-Residential Buildings.” *Proceedings of the 48th International HVAC&R Congress* (December): 369–89. https://izdanja.smeits.rs/index.php/kghk/article/view/3315.

Ballabio, Davide. 2015. “A MATLAB Toolbox for Principal Component Analysis and Unsupervised Exploration of Data Structure.” *Chemometrics and Intelligent Laboratory Systems* 149(November): 1–9. http://dx.doi.org/10.1016/j.chemolab.2015.10.003.

Barczy, Mátyás et al. 2010. “Innovational Outliers in INAR(1) Models.” *Communications in Statistics - Theory and Methods* 39(18): 3343–62.

Bayar, Nawel, Saber Darmoul, Sonia Hajri-Gabouj, and Henri Pierreval. 2015. “Fault Detection, Diagnosis and Recovery Using Artificial Immune Systems: A Review.” *Engineering Applications of Artificial Intelligence* 46(November 2017): 43–57. http://dx.doi.org/10.1016/j.engappai.2015.08.006.

Beghi, A. 2016. “Benchmarking Energy Consumption and CO2 Emissions from Rainwater-Harvesting Systems: An Improved Method by Proxy.” *Procedia Engineering* 89(2): 184–90. http://linkinghub.elsevier.com/retrieve/pii/S1877705815025345 (March 7, 2017).

Bell, Mark R. 1996. “Contiguous Pulse Binary Integration Analysis.” *IEEE Transactions on Aerospace and Electronic Systems* 32(3): 923–33.

Benaicha, Anissa, Gilles Mourot, Kamel Benothman, and José Ragot. 2013. “Determination of Principal Component Analysis Models for Sensor Fault Detection and Isolation.” *International Journal of Control, Automation and Systems* 11(2): 296–305.

Benndorf, Gesa A., Dominik Wystrcil, and Nicolas Réhault. 2018. “Energy

Performance Optimization in Buildings: A Review on Semantic Interoperability, Fault Detection, and Predictive Control." *Applied Physics Reviews* 5(4).

Bernieri, Giuseppe, Miciolino, Estefanía Etchevés, Setola, Roberto. 2017. "Fault Diagnosis and Network Anomaly Detection in Water Infrastructures." *IEEE Design&Test* 2356(c): 44–51.

Bishop, Christopher M. 2013. 53 Journal of Chemical Information and Modeling *Pattern Recognition and Machine Learning*.

Blázquez-García, Ane, Angel Conde, Usue Mori, and Jose A. Lozano. 2021. "Water Leak Detection Using Self-Supervised Time Series Classification." *Information Sciences* 574: 528–41.

Boulos, Paul F., and Adel S. Aboujaoude. 2011. "Managing Leaks Using Flow Step-Testing, Network Modeling, and Field Measurement." *Journal / American Water Works Association* 103(2): 90–97.

Brendan Wall, Dr Jonathan Derham and Mr Tadhg O'Mahony. 2016. *Ireland's Environment*.

Bruton, Ken, Raftery, Paul, Kennedy, Barry, Keane, Marcus M., and O'Sullivan, D. T. J. 2014. "Review of Automated Fault Detection and Diagnostic Tools in Air Handling Units." *Energy Efficiency* 7(2): 335–51.

Burak Gunay, H., Weiming Shen, and Guy Newsham. 2019. "Data Analytics to Improve Building Performance: A Critical Review." *Automation in Construction* 97(June 2018): 96–109.

Bylka, Jedrzej, and Tomasz Mroz. 2019. "A Review of Energy Assessment Methodology for Water Supply Systems." *Energies* 12(23).

Carretero-Ayuso, Manuel J., Alberto Moreno-Cansado, and Justo García-Sanz-Calcedo. 2020. "Occurrence of Faults in Water Installations of Residential Buildings: An Analysis Based on User Complaints." *Journal of Building Engineering* 27(August 2019).

Cataldo, A. et al. 2014. "Time Domain Reflectometry, Ground Penetrating Radar and

Electrical Resistivity Tomography: A Comparative Analysis of Alternative Approaches for Leak Detection in Underground Pipes.” *NDT and E International* 62: 14–28. http://dx.doi.org/10.1016/j.ndteint.2013.10.007.

Cavazzini, Giovanna, Giorgio Pavesi, and Guido Ardizzon. 2020. “Optimal Assets Management of a Water Distribution Network for Leakage Minimization Based on an Innovative Index.” *Sustainable Cities and Society* 54(May 2019).

Chen, Kai Ying, Long Sheng Chen, Mu Chen Chen, and Chia Lung Lee. 2011. “Using SVM Based Method for Equipment Fault Detection in a Thermal Power Plant.” *Computers in Industry* 62(1): 42–50.

Chen, Tao.. 2010. “On Reducing False Alarms in Multivariate Statistical Process Control.” *Chemical Engineering Research and Design* 88(4): 430–36. http://linkinghub.elsevier.com/retrieve/pii/S0263876209002299.

Clifford, Eoghan, Sean Mulligan, Joanne Comer, and Louise Hannon. 2018. “Flow-Signature Analysis of Water Consumption in Nonresidential Building Water Networks Using High-Resolution and Medium-Resolution Smart Meter Data: Two Case Studies.” *Water Resources Research* 54(1): 88–106.

Cody, Roya. 2020. “Acoustic Monitoring for Leaks in Water Distribution Networks By.”

Cody, Roya A., and Sriram Narasimhan. 2020. “A Field Implementation of Linear Prediction for Leak-Monitoring in Water Distribution Networks.” *Advanced Engineering Informatics* 45(April).

Coelho, B., and A. Andrade-Campos. 2014. “Efficiency Achievement in Water Supply Systems—A Review.” *Renewable and Sustainable Energy Reviews* 30: 59–84. https://www.sciencedirect.com/science/article/pii/S1364032113006692?via%3Dihub (October 8, 2019).

Colombo, A., and B. W. Karney. 2003. “Pipe Breaks and the Role of Leaks from an Economic Perspective.” *Water Science and Technology: Water Supply* 3(1–2): 163–69.

Colombo, Andrew F., and Bryan W. Karney. 2005. “Impacts of Leaks on Energy Consumption in Pumped Systems with Storage.” *Journal of Water Resources Planning and Management* 131(April): 146–55.

Colombo, Andrew F., Pedro Lee, and Bryan W. Karney. 2009. “A Selective Literature Review of Transient-Based Leak Detection Methods.” *Journal of Hydro-Environment Research* 2(4): 212–27. http://dx.doi.org/10.1016/j.jher.2009.02.003.

Commission, European. 2013. “Resource and Economic Efficiency of Water Distribution Networks in the EU Final Report.” (October).

Commission, European. 2019. *Jrc Science for Policy Report Water Energy Nexus in Europe*. http://ec.europa.eu/jrc.

Cook, Stephen, Ashok K. Sharma, and Thulo Ram Gurung. 2014. “Evaluation of Alternative Water Sources for Commercial Buildings: A Case Study in Brisbane, Australia.” *Resources, Conservation and Recycling* 89: 86–93. http://dx.doi.org/10.1016/j.resconrec.2014.05.003.

Cortes, Corinna, and Vladimir Vapnik. 1995. “Support-Vector Networks.” *Machine Learning* 20(3): 273–97.

Creaco, E., Campisano, A., Fontana, N., Marini, G. 2019. “Real Time Control of Water Distribution Networks: A State-of-the-Art Review.” *Water Research* 161: 517–30. https://www.sciencedirect.com/science/article/pii/S0043135419305366 (July 18, 2019).

Cuguero Escofet, Miquel À. Quevedo J., Alippi, Roveri C., Puig, V. 2016. “Control Engineering Practice A Methodology and a Software Tool for Sensor Data Validation / Reconstruction : Application to the Catalonia Regional Water Network.” *Control Engineering Practice* 49: 159–72. http://dx.doi.org/10.1016/j.conengprac.2015.11.005.

Cuguero Escofet, Miquel À. Quevedo J., Alippi, Roveri C., Puig, V. 2016. “Model- vs. Data-Based Approaches Applied to Fault Diagnosis in Potable Water Supply Networks.” *Journal of Hydroinformatics* (May): 1–20.

D'Agostino, Delia, Barbara Cuniberti, and Paolo Bertoldi. 2017. "Energy Consumption and Efficiency Technology Measures in European Non-Residential Buildings." *Energy and Buildings* 153(2017): 72–86. http://dx.doi.org/10.1016/j.enbuild.2017.07.062.

D'Ercole, Marianna, Righettia, Maurizio, Ugarelli, Rita Maria, Berardi, Luigi, and Bertola, Paolo. 2016. "An Integrated Modeling Approach to Optimize the Management of a Water Distribution System: Improving the Sustainability While Dealing with Water Loss, Energy Consumption and Environmental Impacts." *Procedia Engineering* 162: 433–40.

Dai, Xuewu, and Zhiwei Gao. 2013. "From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis." *IEEE Transactions on Industrial Informatics* 9(4): 2226–38.

Damarla, Seshu Kumar. 2011. "Multivariate Statistical Process Monitoring and Control." (May).

Danacova, Michaela, Robert Fencik, and Radovan Nosko. 2016. "Historical Development of the Permanent Gully Erosion - Case Study Tura Luka." *Water, Resources, Forest, Marine and Ocean Ecosystems Conference Proceedings, Vol I* (June): 391–98.

Datta, Shantanu, and Shibayan Sarkar. 2016. "A Review on Different Pipeline Fault Detection Methods." *Journal of Loss Prevention in the Process Industries* 41: 97–106. http://dx.doi.org/10.1016/j.jlp.2016.03.010.

Dexter, A, and J Pakanen. 2017. "Demonstrating Automated Fault Detection and Diagnosis Methods in Real Buildings." *VTT Symposium: 217*: 381. http://www2.vtt.fi/inf/pdf/symposiums/2001/S217.pdf%5Cnhttp://www.iea-ebc.org/fileadmin/user_upload/images/Pictures/EBC_Annex_34_VTT_Symposium_217.pdf.

Dhini, Arian, and Isti Surjandari. 2016. "Review on Some Multivariate Statistical Process Control Methods for Process Monitoring." *Proceedings of the International Conference on Industrial Engineering and Operations Management*

8-10 March(iid): 754–59.

Ding, Sheng, and Li Chen. 2010. “Intelligent Optimization Methods for High-Dimensional Data Classification for Support Vector Machines.” *Intelligent Information Management* 02(06): 354–64. http://www.scirp.org/journal/doi.aspx?DOI=10.4236/iim.2010.26043.

Ding, Y., Ma, Liang, Ma, Jian, Suoa, Mingliang, Taoa, Laifa, Chenga,Yujie, and Lu, Chen. 2019. “Intelligent Fault Diagnosis for Rotating Machinery Using Deep Q-Network Based Health State Classification: A Deep Reinforcement Learning Approach.” *Advanced Engineering Informatics* 42(January): 100977. https://doi.org/10.1016/j.aei.2019.100977.

Dong, Yining, and S. Joe Qin. 2018. “A Novel Dynamic PCA Algorithm for Dynamic Data Modeling and Process Monitoring.” *Journal of Process Control* 67: 1–11. https://www.sciencedirect.com/science/article/pii/S095915241730094X?via%3Dihub (December 19, 2019).

Dubuisson, Xavier, and X D Consulting. 2019. “Local Building Energy Standards in Ireland : Report on the Experience of Irish Local Communities In.”

Eliades, D. G., and M. M. Polycarpou. 2012. “Leakage Fault Detection in District Metered Areas of Water Distribution Systems.” *Journal of Hydroinformatics* 14(4): 992–1005.

EPA. “Water Use Efficiency in Buildings.” In *EPA of USA*, , 48–71. http://www.afedonline.org/water efficiency manual/PDF/5Chapter 4_Buildings.pdf.

EurEau - The European Federation of National Water Services. 2017. “Europe’s Water in Figures - 2017 Edition.” https://www.danva.dk/media/3645/eureau_water_in_figures.pdf.

European Commission. 2018. European Union *Artificial Intelligence: A European Perspective*. http://www.sciencedirect.com/science/article/pii/B0122272102000388.

European Commission - DG Environment. 2012. “Water Performance of Buildings.” (August): 154.

Farley, Malcolm, and Stuart Trow. “Losses in Water Distribution Networks.”

Fatemi Aghda, S. M., K. Ganjalipour, and K. Nabiollahi. 2018. “Assessing the Accuracy of TDR-Based Water Leak Detection System.” *Results in Physics* 8: 939–48. https://doi.org/10.1016/j.rinp.2018.01.027.

Fernandes, Chrystinne, Simon Miles, and Carlos José Pereira Lucena. 2020. “Detecting False Alarms by Analyzing Alarm-Context Information: Algorithm Development and Validation.” *JMIR Medical Informatics* 8(5): e15407.

Ferrandez-Gamot, Lise P. Busson, J. Blesa, S. Tornil-Sin, V. Puig, E. Duviella, A. Soldevila. 2015. “Leak Localization in Water Distribution Networks Using Pressure Residuals and Classifiers.” *IFAC-PapersOnLine* 28(21): 220–25.

Fletcher, Sarah, Energy Intensity, End-uses Curr Sustainable, and Renewable Energy. 2018. “Energy Intensity of Water End-Uses The MIT Faculty Has Made This Article Openly Available . Please Share Citation Springer International Publishing Publisher Version Accessed Citable Link Terms of Use Detailed Terms Heading for SpringerLink : Energy-Water.”

Frank, Stephen, Lin, Guanjing, Jin, Xin, Singla, Rupam, Farthing, Amanda, and Granderson, Jessica. 2019. “A Performance Evaluation Framework for Building Fault Detection and Diagnosis Algorithms.” *Energy and Buildings* 192: 84–92. https://www.sciencedirect.com/science/article/pii/S0378778818335680 (July 18, 2019).

Fuentes-García, Marta, Gabriel Maciá-Fernández, and José Camacho. 2018. “Evaluation of Diagnosis Methods in PCA-Based Multivariate Statistical Process Control.” *Chemometrics and Intelligent Laboratory Systems* 172(March 2017): 194–210.

Fuentes, E., L. Arce, and J. Salom. 2018. “A Review of Domestic Hot Water Consumption Profiles for Application in Systems and Buildings Energy Performance Analysis.” *Renewable and Sustainable Energy Reviews* 81(May

2017): 1530–47. http://dx.doi.org/10.1016/j.rser.2017.05.229.

Gajjar, Shriram, Murat Kulahci, and Ahmet Palazoglu. 2018. “Real-Time Fault Detection and Diagnosis Using Sparse Principal Component Analysis.” *Journal of Process Control* 67: 112–28. https://www.sciencedirect.com/science/article/pii/S0959152417300677 (July 30, 2019).

Gao, Yan et al. 2017. “Improving the Shape of the Cross-Correlation Function for Leak Detection in a Plastic Water Distribution Pipe Using Acoustic Signals.” *Applied Acoustics* 127: 24–33.

Garc\ia-Alvarez, D. 2014. “Fault Detection Using Principal Component Analysis (PCA) in a Wastewater Treatment Plant (WWTP).” (January). http://sntk09en.guap.ru/sntk09en/main/docs/Alvarezisa.pdf.

Gautam, Jyoti, Chakrabarti, Amlan , Agarwal, Shruti, Singh, Anushka, Gupta, Shweta and Singh, Jatin. 2020. “Monitoring and Forecasting Water Consumption and Detecting Leakage Using an IoT System.” *Water Supply* 20(3): 1103–13.

Geng, Yang, Ji, Wenjie, Wang, Zhe, Lin, Borong, and Zhu, Yingxin. 2018. “A Review of Operating Performance in Green Buildings: Energy Use, Indoor Environmental Quality and Occupant Satisfaction.” *Energy and Buildings* 183: 500–514. https://linkinghub.elsevier.com/retrieve/pii/S0378778818315378.

Geng, Yang, Ji, Wenjie, Wang, Zhe, Lin, Borong, and Zhu, Yingxin. 2019. “Response to the Commentary on ‘A Review of Operating Performance in Green Buildings: Energy Use, Indoor Environmental Quality and Occupant Satisfaction’ by John H. Scofield.” *Energy and Buildings* 194: 366–68.

Gertler, J., J. Romera, V. Puig, and J. Quevedo. 2010. “Leak Detection and Isolation in Water Distribution Networks Using Principal Component Analysis and Structured Residuals.” *Conference on Control and Fault-Tolerant Systems, SysTol’10 - Final Program and Book of Abstracts*: 191–96.

Gharsellaoui, Sondes, Mansouri, Majdi, Trabelsi, Mohamed, Refaat, Shady S., and Messaoud, Hassani. 2020. “Fault Diagnosis of Heating Systems Using

Multivariate Feature Extraction Based Machine Learning Classifiers." *Journal of Building Engineering* 30(September 2019): 101221.

Gilliam, David, Stefan Leigh, Andrew Rukhin, and William Strawderman. 2012. "Pass-Fail Testing: Statistical Requirements and Interpretations." *Journal of Research of the National Institute of Standards and Technology* 114(3): 195.

De Gois, Eduardo H.B., Carlos A.S. Rios, and Ricardo N. Costanzi. 2015. "Evaluation of Water Conservation and Reuse: A Case Study of a Shopping Mall in Southern Brazil." *Journal of Cleaner Production* 96: 263–71. http://dx.doi.org/10.1016/j.jclepro.2014.08.097.

Günther, M., D. Steffelbauer, M. Neumayer, and D. Fuchs-Hanusch. 2014. "Experimental Setup to Examine Leakage Outflow in a Scaled Water Distribution Network." *Procedia Engineering* 89: 311–17.

Gupta, Aditya, and K. D. Kulat. 2018. "A Selective Literature Review on Leak Management Techniques for Water Distribution System." *Water Resources Management* 32(10): 3247–69.

Gupta, Shivani. 2019. "Dealing with Noise Problem in Machine Learning Data-Sets.Pdf."

Hardeman, S. 2008. "A Cost-Benefit Analysis of Leak Detection and the Potential of Real Water Savings for New Mexico Water Systems." https://repository.unm.edu/handle/1928/7536.

Harris, Paul Brunsdon, Chris, Charlton, Martin, Juggins, Steve, and Clarke, Annemarie. 2014. "Multivariate Spatial Outlier Detection Using Robust Geographically Weighted Methods." *Mathematical Geosciences* 46(1): 1–31.

Harrou, Fouzi, Mohamed N. Nounou, Hazem N. Nounou, and Muddu Madakyaru. 2013. "Statistical Fault Detection Using PCA-Based GLR Hypothesis Testing." *Journal of Loss Prevention in the Process Industries* 26(1): 129–39. http://dx.doi.org/10.1016/j.jlp.2012.10.003.

Hashemi, Saeed, Filion Yves R., and Speight, Vanessa L. 2016. "A Review on

Different Pipeline Fault Detection Methods." *Applied Energy* 11(1): 79–91. http://linkinghub.elsevier.com/retrieve/pii/S1474667016447403 (March 7, 2017).

Hashim, Hafiz, Paraic Ryan, and Eoghan Clifford. 2020. "A Statistically Based Fault Detection and Diagnosis Approach for Non-Residential Building Water Distribution Systems." *Advanced Engineering Informatics* (September): 101187. https://doi.org/10.1016/j.aei.2020.101187.

Hmeidi, Ismail, Bilal Hawashin, and Eyas El-Qawasmeh. 2008. "Performance of KNN and SVM Classifiers on Full Word Arabic Articles." *Advanced Engineering Informatics* 22(1): 106–11.

Houngbo, Gilbert F. 2018. UNESCO Digital Library *The United Nations World Water Development Report 2018: Nature-Based Solutions for Water*.

Houngbo, Gilbert. 2019. UNESCO Digital Library *The United Nations World Development Report 2019. Leaving No One Behind.* www.unesco.org/water/wwap.

Howley, Martin, and Mary Holland. 2016. "Energy in Ireland 1990 - 2015." *Seai*: 84. http://www.seai.ie/resources/publications/Energy-in-Ireland-1990-2015.pdf.

Hsu, Esher. 2021. "Cost-Benefit Analysis for Recycling of Agricultural Wastes in Taiwan." *Waste Management* 120: 424–32.

Hu, R. L., J. Granderson, D. M. Auslander, and A. Agogino. 2019. "Design of Machine Learning Models with Domain Experts for Automated Sensor Selection for Energy Fault Detection." *Applied Energy* 235(May 2018): 117–28.

Hu, Zukang et al. 2021. "Review of Model-Based and Data-Driven Approaches for Leak Detection and Location in Water Distribution Systems." *Water Supply*: 1–25.

Hubballi, Neminath, and Vinoth Suryanarayanan. 2014. "False Alarm Minimization Techniques in Signature-Based Intrusion Detection Systems: A Survey." *Computer Communications* 49: 1–17.

Hubert, Mia, Tom Reynkens, Eric Schmitt, and Tim Verdonck. 2016. "Sparse PCA for High-Dimensional Data With Outliers." *Technometrics* 58(4): 424–34.

Hubert, Mia, Peter J. Rousseeuw, and Karlien Vanden Branden. 2005. “ROBPCA: A New Approach to Robust Principal Component Analysis.” *Technometrics* 47(1): 64–79.

Ihasalo, Heikki, and Sami Karjalainen. 2014. “Challenges in Getting Building Performance Monitoring Tools for Everyday Use: User Experiences with A New Tool.” *Buildings* 4(2): 222–43.

“Improved Water Resource Management with the Development of Maintenance Technologies for the Built Environment By.” 2017. (January).

Irish Medtech Association. 2019. “Irish Medtech Association Statement of Strategy 2016-2020.” 1(1): 1–36.

Irish Water. 2013. “Irish Water.” https://www.water.ie.

J., Prabuchandran K., Nitin Singh, Pankaj Dayama, and Vinayaka Pandit. 2019. “Change Point Detection for Compositional Multivariate Data.” http://arxiv.org/abs/1901.04935.

Jackson, J Edward, and Govind S Mudholkar. 1979. “Control Procedures for Residuals Associated With Principal Component Analysis.” *Technometrics* 21(3): 341–49. https://www.tandfonline.com/doi/abs/10.1080/00401706.1979.10489779.

Jarrah Nezhad, Aylin, Tri Kurniawan Wijaya, Matteo Vasirani, and Karl Aberer. 2014. “SmartD: Smart Meter Data Analytics Dashboard.” *e-Energy 2014 - Proceedings of the 5th ACM International Conference on Future Energy Systems*: 213–14.

Joe Qin, S. 2003. “Statistical Process Monitoring: Basics and Beyond.” *Journal of Chemometrics* 17(8–9): 480–502. http://doi.wiley.com/10.1002/cem.800.

Johnson, Richard A, and Dean W Wichern. 2007. “Applied Multivariate Statistical Analysis.”

Jolliffe, I T. 2002. *Principal Component Analysis, Second Edition*. Springer International Publishing.

Jung, Donghwi, Doosun Kang, Jian Liu, and Kevin Lansey. 2015. “Improving the

Rapidity of Responses to Pipe Burst in Water Distribution Systems: A Comparison of Statistical Process Control Methods." *Journal of Hydroinformatics* 17(2): 307–28.

Jung, Namcheol, and Ghang Lee. 2019. "Automated Classification of Building Information Modeling (BIM) Case Studies by BIM Use Based on Natural Language Processing (NLP) and Unsupervised Learning." *Advanced Engineering Informatics* 41(September 2018): 100917. https://doi.org/10.1016/j.aei.2019.04.007.

Kamiel, Berli Paripurna. 2015. "Vibration-Based Multi-Fault Diagnosis for Centrifugal Pumps Berli Paripurna Kamiel." (February).

Chambers, N., Coakley, D., Keane, M, and E Clifford. 2015. "Water Conservation with Novel Application of Fault Detection Diagnostics (FDD) Applied to a Rain Water Harvesting System in Ireland." (May).

Kenway, S. J., Binks, A., Lane, J., Lant, P.A., Lam, K.L., and Simms, A.. 2015. "A Systemic Framework and Analysis of Urban Water Energy." *Environmental Modelling and Software* 73: 272–85. http://dx.doi.org/10.1016/j.envsoft.2015.08.009.

Kern, Andrea Parisi, Antoniolli, Cibele Bossa, Wander, Paulo Roberto, Mancio, Maurício. 2016. "Energy and Water Consumption during the Post-Occupancy Phase and the Users' Perception of a Commercial Building Certified by Leadership in Energy and Environmental Design (LEED)." *Journal of Cleaner Production* 133: 826–34.

Kouziokas, Georgios N. 2020. "SVMkernelbasedonparticleswarmoptimizedvector.Pdf."

Lam, Ka Leung, Steven J. Kenway, and Paul A. Lant. 2017. "City-Scale Analysis of Water-Related Energy Identifies More Cost-Effective Solutions." *Water Research* 109: 287–98. http://dx.doi.org/10.1016/j.watres.2016.11.059.

Laory, Irwanda, Thanh N. Trinh, and Ian F.C. Smith. 2011. "Evaluating Two Model-Free Data Interpretation Methods for Measurements That Are Influenced by Temperature." *Advanced Engineering Informatics* 25(3): 495–506.

http://dx.doi.org/10.1016/j.aei.2011.01.001.

Lavappa, Priya D., and Joshua D. Kneifel. 2015. “Energy Price Indices and Discount Factors for Life-Cycle Cost Analysis – 2015 Annual Supplement to NIST Handbook 135.” *National Institte of Standards and Technology . U.S Depatrtement of Commerce* (April 2005): 67.

Lazarova-molnar, Sanja, Hamid Reza Shaker, Nader Mohamed, and Bo Nørregaard Jørgensen. “Fault Detection and Diagnosis for Smart Buildings : State of the Art , Trends and Challenges.”

Li, Suzhen, Yanjue Song, and Gongqi Zhou. 2018. “Leak Detection of Water Distribution Pipeline Subject to Failure of Socket Joint Based on Acoustic Emission and Pattern Recognition.” *Measurement: Journal of the International Measurement Confederation* 115(October 2017): 39–44. http://dx.doi.org/10.1016/j.measurement.2017.10.021.

Li, Wei, Minjun Peng, and Qingzhong Wang. 2018. “False Alarm Reducing in PCA Method for Sensor Fault Detection in a Nuclear Power Plant.” *Annals of Nuclear Energy* 118: 131–39. https://doi.org/10.1016/j.anucene.2018.04.012.

Li, Wei, Minjun Peng, and Qingzhong Wang. 2019. “Improved PCA Method for Sensor Fault Detection and Isolation in a Nuclear Power Plant.” *Nuclear Engineering and Technology* 51(1): 146–54. https://doi.org/10.1016/j.net.2018.08.020.

Li, Weihua, H. Henry Yue, Sergio Valle-Cervantes, and S. Joe Qin. 2000. “Recursive PCA for Adaptive Process Monitoring.” *Journal of Process Control* 10(5): 471–86.

Liliana, Magda, and Ruiz Ordóñez. 2008. “Multivariate Statistical Process Control and Case-based Reasoning for Situation Assesment of Sequencing Batch Reactors.”

Lin, Tsun-Kuo. 2018. “PCA/SVM-Based Method for Pattern Detection in a Multisensor System.” *Mathematical Problems in Engineering* 2018: 1–11.

Liu, Caijie, Qin Zhang, and Hai Wang. 2020. “Cost-Benefit Analysis of Waste Photovoltaic Module Recycling in China.” *Waste Management* 118: 491–500.

Liu, Xiufeng, Lukasz Golab, and Ihab F. Ilyas. 2015. “SMAS: A Smart Meter Data Analytics System.” *Proceedings - International Conference on Data Engineering* 2015-May: 1476–79.

Liu, Xiufeng, Nadeem Iftikhar, Per Sieverts Nielsen, and Alfred Heller. 2016. “Online Anomaly Energy Consumption Detection Using Lambda Architecture.” In *Big Data Analytics and Knowledge Discovery*, eds. Sanjay Madria and Takahiro Hara. Cham: Springer International Publishing, 193–209.

Liu, Yang et al. 2019. “Water Pipeline Leakage Detection Based on Machine Learning and Wireless Sensor Networks.” *Sensors (Switzerland)* 19(23): 1–21.

Liu, Zheng, and Yehuda Kleiner. 2013. “State of the Art Review of Inspection Technologies for Condition Assessment of Water Pipes.” *Measurement: Journal of the International Measurement Confederation* 46(1): 1–15.

Maione, Camila, Fernando Barbosa, and Rommel Melgaço Barbosa. 2019. “Predicting the Botanical and Geographical Origin of Honey with Multivariate Data Analysis and Machine Learning Techniques: A Review.” *Computers and Electronics in Agriculture* 157(January): 436–46.

Makaya, Eugine, and Oliver Hensel. 2015. “Water Loss Management Strategies for Developing Countries: Understanding the Dynamics of Water Leakages.” : 1–97. https://d-nb.info/1112580042/34.

von der Malsburg, Titus, and Bernhard Angele. 2017. “False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in Reading.” *Journal of Memory and Language* 94: 119–33.

Mannan, Mehzabeen, and Sami G. Al-Ghamdi. 2020. “Environmental Impact of Water-Use in Buildings: Latest Developments from a Life-Cycle Assessment Perspective.” *Journal of Environmental Management* 261(January): 110198. https://doi.org/10.1016/j.jenvman.2020.110198.

Mao, Xiangke, Yang, Hui, Huang, Shaobin, Liu, Ye, and Li, Rongsheng. 2019. “Extractive Summarization Using Supervised and Unsupervised Learning.” *Expert Systems with Applications* 133: 173–81.

https://doi.org/10.1016/j.eswa.2019.05.011.

Mashford, John, Dhammika De Silva, Stewart Burn, and Donavan Marney. 2012. “Leak Detection in Simulated Water Pipe Networks Using SVM.” *Applied Artificial Intelligence* 26(5): 429–44.

Mashford, John, Dhammika De Silva, Donavan Marney, and Stewart Burn. 2009. “An Approach to Leak Detection in Pipe Networks Using Analysis of Monitored Pressure Values by Support Vector Machine.” *NSS 2009 - Network and System Security* (Figure 1): 534–39.

Matos, C. Santos, C., Pereira, S., Bentes, I., and Imteaz, Monzur. 2013. “Rainwater Storage Tank Sizing: Case Study of a Commercial Building.” *International Journal of Sustainable Built Environment* 2(2): 109–18. http://dx.doi.org/10.1016/j.ijsbe.2014.04.004.

Alsaydalani, M. O. A. 2020. “Environmental Impact of Water-Use in Buildings: Latest Developments from a Life-Cycle Assessment Perspective.” *Journal of Environmental Management* 261(2): 110198. http://ec.europa.eu/jrc.

Medina, Victor F, John Matthews, and Ryan Stowe. 2017. “Final Report : Innovative Acoustic Sensor Technologies for Leak Technical Report TR-NAVFAC-EXWC-EV-1706 D ECEMBER 2016 Final Report : Innovative Acoustic Sensor Technologies for Leak Detection in Challenging Pipe Types.” (June).

Merzi, N, and T Özkan. 2015. “Water Leakage and Detection in Municipal Water Distribution Networks.” *DAAAM International Conference* (1).

Meseguer, Jordi, Josep M. Mirats-Tur, Gabriela Cembrano, and Vicenç Puig. 2015. “Model-Based Monitoring Techniques for Leakage Localization in Distribution Water Networks.” *Procedia Engineering* 119(1): 1399–1408. http://dx.doi.org/10.1016/j.proeng.2015.08.1000.

MOCCAE. 2016. “Water Statistics.” *Abu Dhabi Water Statistics* 3(July): 1–8.

Mooney, John G., and Eric N. Johnson. 2014. “A Comparison of Automatic Nap-of-the-Earth Guidance Strategies for Helicopters.” *Journal of Field Robotics* 33(1):

1–17. http://onlinelibrary.wiley.com/doi/10.1002/rob.21514/abstract.

Moors, J., L. Scholten, J. P. van der Hoek, and J. den Besten. 2018. “Automated Leak Localization Performance without Detailed Demand Distribution Data.” *Urban Water Journal* 15(2): 116–23. https://www.tandfonline.com/doi/full/10.1080/1573062X.2017.1414272.

Moser, Gaudenz, Stephanie German Paal, and Ian F.C. Smith. 2015. “Performance Comparison of Reduced Models for Leak Detection in Water Distribution Networks.” *Advanced Engineering Informatics* 29(3): 714–26. http://dx.doi.org/10.1016/j.aei.2015.07.003.

Mounce, Stephen R., and John Machell. 2006. “Burst Detection Using Hydraulic Data from Water Distribution Systems with Artificial Neural Networks.” *Urban Water Journal* 3(1): 21–31.

Mudry, Albert, and Anders Tjellström. 2011. “Historical Background of Bone Conduction Hearing Devices and Bone Conduction Hearing Aids.” *Advances in Oto-Rhino-Laryngology* 71: 1–9.

Mujica, L. E., J. Rodellar, A. Fernández, and A. Güemes. 2011. “Q-Statistic and T2-Statistic Pca-Based Measures for Damage Assessment in Structures.” *Structural Health Monitoring* 10(5): 539–53.

Muranho, J., Ferreirad, A., Sousac, J., Gomesa, A., and Sá Marquesc, A.. 2014. “Technical Performance Evaluation of Water Distribution Networks Based on EPANET.” *Procedia Engineering* 70: 1201–10. http://dx.doi.org/10.1016/j.proeng.2014.02.133.

Mutikanga, Harrison E. 2012. *WATER LOSS MANAGEMENT: Tools and Methods for Developing Countries*.

Naderi, Esmaeil, and Khashayar Khorasani. 2016. “A Data-Driven Approach to Actuator and Sensor Fault Detection, Isolation and Estimation in Discrete-Time Linear Systems.” 85: 165–78. http://arxiv.org/abs/1606.06220.

Nair, Sudeep, Mulligan, Sean, Hafiz Hashim, Louise Hannon, Ryan, Paraic, and

Eoghan Clifford. 2018. “Quantifying Pumping Energy Losses Associated with Faults in Building Water Networks Using a Novel Fault Detection and Diagnosis Approach.” : 1–13.

Nair, Sudeep, Hafiz Hashim, Louise Hannon, and Eoghan Clifford. 2018. “End Use Level Water and Energy Interactions: A Large Non-Residential Building Case Study.” *Water (Switzerland)* 10(6).

Najah F. Ghalyan a, B, and Asok Ray. 2021. “2021_Measure Invariance of Ergodic Symbolic Systems for Low-Delay.Pdf.” *Mechanical Systems and Signal Processing*.

Nalepa, Jakub, and Michal Kawulok. 2018. “Selecting Training Sets for Support Vector Machines: A Review.” *Artificial Intelligence Review*: 1–44. https://doi.org/10.1007/s10462-017-9611-1.

Nasir, Mohammad Tariq, Mysorewala, Muhammad, Cheded, Lahouari, Siddiqui, Bilal, and Sabih, Muhammad. 2014. “Measurement Error Sensitivity Analysis for Detecting and Locating Leak in Pipeline Using ANN and SVM.” *2014 IEEE 11th International Multi-Conference on Systems, Signals and Devices, SSD 2014*: 7–10.

Nayak, Janmenjoy, Bighnaraj Naik, and H. S. Behera. 2015. “A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges.” *International Journal of Database Theory and Application* 8(1): 169–86. http://www.sersc.org/journals/IJDTA/vol8_no1/18.pdf.

Nowicki, Adam, Michał Grochowski, and Kazimierz Duzinkiewicz. 2012. “Data-Driven Models for Fault Detection Using Kernel PCA: A Water Distribution System Case Study.” *International Journal of Applied Mathematics and Computer Science* 22(4): 939–49. http://www.degruyter.com/view/j/amcs.2012.22.issue-4/v10006-012-0070-1/v10006-012-0070-1.xml.

O’Callaghan, Daniel, and Seán Prior. 2018. “Staff Paper 2018 Central Technical Appraisal Parameters Discount Rate, Time Horizon, Shadow Price of Public Funds and Shadow Price of Labour.” https://igees.gov.ie/wp-content/uploads/2019/07/Parameters-Paper-Final-Version.pdf.

Ociepa, Ewa, Maciej Mrowiec, and Iwona Deska. 2019. “Analysis of Water Losses and Assessment of Initiatives Aimed at Their Reduction in Selected Water Supply Systems.” *Water (Switzerland)* 11(5): 1–18.

Papoulis, Athanasios. 1980. *Probability, Random Variables and Stochastic Processes*. Third Edit. Polytechnic Institute of New York: McGraw-Hill Series in Electrical engineering.

Patabendige, Samitha, Rachel Cardell-Oliver, Rui Wang, and Wei Liu. 2018. “Detection and Interpretation of Anomalous Water Use for Non-Residential Customers.” *Environmental Modelling and Software* 100: 291–301.

Pelz, Georg. 2003. Library *Mechatronic Systems*. http://doi.wiley.com/10.1002/0470867906.

Perez, Ramon, Sanz, Gerar, Puig, Vicenc. 2014. “Leak Localization in Water Networks: A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona [Applications of Control].” *IEEE Control Systems* 34(4): 24–36.

Pérez, Ramon, Gerard Sanz, Miquel-Àngel Cugueró, Joaquim Blesa,. 2015. “Parameter Uncertainty Modelling in Water Distribution Network Models.” *Procedia Engineering* 119(September): 583–92. http://linkinghub.elsevier.com/retrieve/pii/S1877705815025813.

Pérez, Ramon, Gerard Sanz, Miquel À Cugueró, P Ramon, et al. 2015. “Parameter Uncertainty Modelling in Water Distribution Network Models Models.” (September).

Perfido, Domenico, Messervey, Thomas, Zanotti, Chiara, Raciti, Massimiliano and Costa, Andrea. 2016. “Automated Leak Detection System for the Improvement of Water Network Management.” *Proceedings* 1(2): 28. http://www.mdpi.com/2504-3900/1/2/28.

Petkov, Nikola, Huapeng Wu, and Roger Powell. 2020. “Cost-Benefit Analysis of Condition Monitoring on DEMO Remote Maintenance System.” *Fusion Engineering and Design* 160.

Posenato, Daniele, Francesca Lanata, Daniele Inaudi, and Ian F.C. Smith. 2008. "Model-Free Data Interpretation for Continuous Monitoring of Complex Structures." *Advanced Engineering Informatics* 22(1): 135–44.

Puig, Vicenç, Carlos Ocampo-Martínez, Deneb Robles, and Luis Eduardo Garza-Castañón. 2017. "Fault-Tolerant Model Predictive Control of Water Transport Networks." *Advances in Industrial Control* (9783319507507): 291–319.

Puust, R., Z. Kapelan, D. A. Savic, and T. Koppel. 2010. 7 Urban Water Journal *A Review of Methods for Leakage Management in Pipe Networks*.

Qin, S. Joe, and Leo H. Chiang. 2019. "Advances and Opportunities in Machine Learning for Process Data Analytics." *Computers & Chemical Engineering* 126: 465–73. https://www.sciencedirect.com/science/article/pii/S0098135419302248?via%3Dihub (August 29, 2019).

Qin, S Joe. 2012. "Annual Reviews in Control Survey on Data-Driven Industrial Process Monitoring and Diagnosis." 36: 220–34.

Qu, Zhigang, Feng, Hao, Zeng, Zhoumo, Zhuge, Jingchang, and Jin, Shijiu. 2010. "A SVM-Based Pipeline Leakage Detection and Pre-Warning System." *Measurement: Journal of the International Measurement Confederation* 43(4): 513–19. http://dx.doi.org/10.1016/j.measurement.2009.12.022.

Quevedo, J., Chen, H., Cugueró, M.À., Tino, P., Puig, V., García, D., Sarrate, R., Yao, X.. 2014. "Engineering Applications of Artificial Intelligence Combining Learning in Model Space Fault Diagnosis with Data Validation / Reconstruction : Application to the Barcelona Water Network." *Engineering Applications of Artificial Intelligence* 30: 18–29. http://dx.doi.org/10.1016/j.engappai.2014.01.008.

Marcos Qui˜nones-Grueiro, Jos´e M. Bernal-de L´azaro, Cristina, Verde, and Alberto, Prieto. 2018. "Comparison of Classifiers for Leak Location in Water Distribution Networks." *IFAC-PapersOnLine* 51(24): 407–13.

Quiñones-Grueiro, Marcos, Cristina Verde, and Orestes Llanes-Santiago. 2017. "Demand Model in Water Distribution Networks for Fault Detection." *IFAC-*

PapersOnLine 50(1): 3263–68.

Quinõnes-Grueiro, Marcos, Cristina Verde, Alberto Prieto-Moreno, and Orestes Llanes-Santiago. 2018. “An Unsupervised Approach to Leak Detection and Location in Water Distribution Networks.” *International Journal of Applied Mathematics and Computer Science* 28(2): 283–95.

Ragab, Ahmed et al. 2018. “Fault Diagnosis in Industrial Chemical Processes Using Interpretable Patterns Based on Logical Analysis of Data.” *Expert Systems with Applications* 95: 368–83. https://doi.org/10.1016/j.eswa.2017.11.045.

Rajan, B S, and Brian J Roylance. 1996. “The Development of a Cost Benefit Analysis Method for Monitoring the Condition of Batch Process Plant Machinery.” *Technology Showcase: Integrated Monitoring, Diagnostics and Failure Prevention* (AD-A325 558): 13.

Reniers, G, L Talarico, and N Paltrinieri. 2016. “Cost-Bene Fi t Analysis of Safety Measures.” In , 195–205.

Robles, Deneb, Vicenç Puig, Carlos Ocampo-Martinez, and Luis E. Garza-Castañón. 2016. “Reliable Fault-Tolerant Model Predictive Control of Drinking Water Transport Networks.” *Control Engineering Practice* 55: 197–211. http://dx.doi.org/10.1016/j.conengprac.2016.06.014.

Rosen, C. 2001. Department of Industrial Electrical Engineering and Automation *A Chemometric Approch to Process Monitoring and Control\n- with Application to Wastewater Treatment Operation-*.

Rousseeuw, Peter J., and Mia Hubert. 2018. “Anomaly Detection by Robust Statistics.” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8(2): 1–14.

Rousseeuw, Peter J., Jakob Raymaekers, and Mia Hubert. 2018. “A Measure of Directional Outlyingness With Applications to Image Data and Video.” *Journal of Computational and Graphical Statistics* 27(2): 345–59.

Ruzza, V., E. Crestani, G. Darvini, and P. Salandin. 2014. “Losses Identification in

Uncalibrated Water Distribution Networks: A Case Study.” *Procedia Engineering* 89: 1545–52.

Sabri, Mohamad, Bin Sinal, and Eiji Kamioka. 2017. “Adaptive Threshold based Approach to Perfectly Detect Heart Cycle in ECG Data.” (44): 492–98.

Saitta, S., B. Raphael, and I. F.C. Smith. 2005. “Data Mining Techniques for Improving the Reliability of System Identification.” *Advanced Engineering Informatics* 19(4): 289–98.

Salam, A Ejah Umraeni, Muh Tola, Mary Selintung, and Farouk Maricar. 2015. “Application of SVM and ELM Methods to Predict Location and Magnitude Leakage of Pipelines on Water Distribution Network.” (19): 7970.

Samer El-Zahab. 2018. “An Accelerometer-Based Leak Detection System _ Elsevier Enhanced Reader.Pdf.” : Mechanical Systems and Signal Processing Volume 10.

San Antonio Water System. 2017. “Your Water.” http://www.saws.org/Your_Water/.

Sánchez-Fernández, Baldan, A. F.J., Sainz-Palmero, G.I., Benítez, J.M., and Fuente, M.J. 2018. “Fault Detection Based on Time Series Modeling and Multivariate Statistical Process Control.” *Chemometrics and Intelligent Laboratory Systems* 182(July): 57–69.

Sanz, Gerard, Ramon Pérez, Zoran Kapelan, and Dragan Savic. 2016. “Leak Detection and Localization through Demand Components Calibration.” *Journal of Water Resources Planning and Management* 142(2).

Sean Mulligan, Louise Hannon, Paraic Ryan, Sudeep Nair, Eoghan Clifford. 2020. “Development of a Data Driven FDD Approach for Building Water Networks: Water Distribution System Performance Assessment Rules.” *Journal of Building Engineering*: 102248. https://doi.org/10.1016/j.janxdis.2020.102248.

Sedki, A., and D. Ouazar. 2012. “Hybrid Particle Swarm Optimization and Differential Evolution for Optimal Design of Water Distribution Systems.” *Advanced Engineering Informatics* 26(3): 582–91.

Seiler, Dominik, Dan Donovan, and Garret E. O’Donnell. 2019. “Resource

Transparency of Industrial Systems: A Model to Demonstrate the Total Costs of Water Purification." *Procedia CIRP* 80: 281–86. https://doi.org/10.1016/j.procir.2019.01.013.

Sengupta, Raghu Nandan, and Debasis Kundu. 2016. "Statistical Methods." : 413–520.

Seyoum, Solomon, Alfonsoa, Leonardo, Andela, Schalk Jan van, Koole, Wouter, Groenewegen, Ad , and Giesen, Nick van de. 2017. "A Shazam-like Household Water Leakage Detection Method." *Procedia Engineering* 186: 452–59.

Seyoum, Zewdu, Alemu Zewdu, Shimelis Dagnachew, and Basazinew Bogale. 2017. "Anthelmintic Resistance of Strongyle Nematodes to Ivermectin and Fenbendazole on Cart Horses in Gondar, Northwest Ethiopia." *BioMed Research International* 2017(February): 1–6. https://www.hindawi.com/journals/bmri/2017/5163968/.

Sgorbissa, Antonio. 2018. "Integrated Robot Planning, Path Following and Obstacle Avoidance in 2D and 3D." (March): 1–13. http://arxiv.org/abs/1804.06557.

Sharma, Abishek B, Leana Golubchik, and Ramesh Govindan. 2010. "Sensor Faults : Detection Methods and Prevalence in Real-World Datasets." *ACM transactions on Sensor Networks* 6(3): 23.

Sheriff, Mohammed Ziyan, Botre, Chiranjivi, Mansouri, Majdi, Nounou, Hazem, Nounou, Mohamed and Karim, Mohammad Nazmul. 2017. "Process Monitoring Using Data-Based Fault Detection Techniques: Comparative Studies." *Fault Diagnosis and Detection* (December). http://www.intechopen.com/books/fault-diagnosis-and-detection/process-monitoring-using-data-based-fault-detection-techniques-comparative-studies.

Siddiqi, Afreen, and Olivier L de Weck. 2013. "Quantifying End-Use Energy Intensity of the Urban Water Cycle." *Journal of Infrastructure Systems* 19(4): 474–85.

De Silva, Chamila C., Scott P. Beckman, Shuaishuai Liu, and Nicola Bowler. 2019. "Principal Component Analysis (PCA) as a Statistical Tool for Identifying Key Indicators of Nuclear Power Plant Cable Insulation Degradation." *Minerals, Metals and Materials Series*: 1227–39.

Silva, Dhammika De, John Mashford, and Stewart Burn. 2011. "Computer Aided Leak Location and Sizing in Pipe Networks Urban Water Security Research Alliance Technical Report No . 17. Urban Water Security Research Alliance Technical Report NO. 17, 2011(17).De." *Urban Water Security Research Alliance Technical Report No. 17* 2011(17).

Skworcow, P. 2013. "Pressure , Leakage and Energy Management in Water Distribution Systems." *Water Resources Management* i(1): 266. http://dx.doi.org/10.1016/j.proeng.2015.08.855.

Smith, K. W., and A. L. Aretxabaleta. 2007. "Expectation-Maximization Analysis of Spatial Time Series." *Nonlinear Processes in Geophysics* 14(1): 73–77.

Soldevila, Adrià, Fernandez-Cantia, Rosa M., Blesa, Joaquim, Tornil-Sina, Sebastian, and Puig, Vicenc. 2016. "Leak Localization in Water Distribution Networks Using a Mixed Model-Based/Data-Driven Approach." *Control Engineering Practice* 55: 162–73.

Soldevila, Adrià, Fernandez-Cantia, Rosa M., Blesa, Joaquim, Tornil-Sina, Sebastian, and Puig, Vicenc. 2017. "Leak Localization in Water Distribution Networks Using Bayesian Classifiers." *Journal of Process Control* 55: 1–9.

Sophocleous, Sophocles, Dragan A. Savić, Zoran Kapelan, and Orazio Giustolisi. 2017. "A Two-Stage Calibration for Detection of Leakage Hotspots in a Real Water Distribution Network." *Procedia Engineering* 186(0): 168–76. http://dx.doi.org/10.1016/j.proeng.2017.03.223.

Sousa, Vitor, Cristina Matos Silva, and Inês Meireles. 2019. "Performance of Water Efficiency Measures in Commercial Buildings." *Resources, Conservation and Recycling* 143(October 2018): 251–59. https://doi.org/10.1016/j.resconrec.2019.01.013.

Stanley, Greg. 2010. "A Guide to Fault Detection and Diagnosis." *Greg Stanley and Associates*. http://gregstanleyandassociates.com/whitepapers/FaultDiagnosis/faultdiagnosis.htm.

Stavset, Ole, and Hanne Kauko. 2015. *Report -Possibilities for Smart Energy Solutions*.

Sterman, John D. 2006. “Domestic Water Heating and Water Heater Energy Consumption in Canada.” *American Journal of Public Health* 96(3): 505–14.

Stetco, Adrian et al. 2019. “Machine Learning Methods for Wind Turbine Condition Monitoring: A Review.” *Renewable Energy* 133: 620–35. https://doi.org/10.1016/j.renene.2018.10.047.

Stevenson, Mike. 2005. “Engineering Failure Analysis.” *Journal of Failure Analysis and Prevention* 5(6): 5–6.

Sun, Congcong, Benjamí Parellada, Vicenç Puig, and Gabriela Cembrano. 2020. “Leak Localization in Water Distribution Networks Using Pressure and Data-Driven Classifier Approach.” *Water (Switzerland)* 12(1).

Sydney Water. 2011. “Best Practice Guidelines for Water Management in Aquatic Leisure Centres.”

Tardioli, Giovanni et al. 2015. “Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level.” *Energy Procedia* 78: 3378–83. http://dx.doi.org/10.1016/j.egypro.2015.11.754.

Thomas, Michael C., Wenbo Zhu, and Jose A. Romagnoli. 2018. “Data Mining and Clustering in Chemical Process Databases for Monitoring and Knowledge Discovery.” *Journal of Process Control* 67: 160–75. http://dx.doi.org/10.1016/j.jprocont.2017.02.006.

Tidriri, Khaoula, Nizar Chatti, Sylvain Verron, and Teodor Tiplica. 2016. “Bridging Data-Driven and Model-Based Approaches for Process Fault Diagnosis and Health Monitoring: A Review of Researches and Future Challenges.” *Annual Reviews in Control* 42: 63–81.

Touzani, Samir, Baptiste Ravache, Eliot Crowe, and Jessica Granderson. 2019. “Statistical Change Detection of Building Energy Consumption : Applications to Savings Estimation.” 4: 1–22.

UNESCO. 2015. *The United Nations World Water Development Report 2015: Water* *for a Sustainable World. UNESCO Publishing, Paris*.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, and Kewen Yin. 2003. "A Re v Iew of Process Fault Detection and Diagnosis Part I : Quantitati v e Model-Based Methods." 27: 293–311.

Version, Document. 2008. "Fault Detection and Isolation of an On-Line Analyzer for an Ethylene Cracking Process."

Vieira, A S, C D Beal, E Ghisi, and R A Stewart. 2014. "Energy Intensity of Rainwater Harvesting Systems: A Review." *Renewable and Sustainable Energy Reviews* 34: 225–42. http://www.sciencedirect.com/science/article/pii/S1364032114001695%5Cnhttps://www.scopus.com/inward/record.uri?eid=2-s2.0-84897391469&partnerID=40&md5=5f28501c2f1d33d22df9f24b6c6ea94d.

Vieira, Abel S., Cara D. Beal, Enedir Ghisi, and Rodney A. Stewart. 2014. "Energy Intensity of Rainwater Harvesting Systems: A Review." *Renewable and Sustainable Energy Reviews* 34(June): 225–42.

Villegas, Thamara, M.J. Fuente, and Miguel Rodríguez. 2010. "Principal Component Analysis for Fault Detection and Diagnosis. Experience with a Pilot Plant." *Proceedings of the 9th WSEAS International conference on computational intelligence, man-machine systems and cybernetics*: 147–52.

Vitter, J. S., and M. E. Webber. 2018. "A Non-Intrusive Approach for Classifying Residential Water Events Using Coincident Electricity Data." *Environmental Modelling and Software* 100: 302–13. https://doi.org/10.1016/j.envsoft.2017.11.029.

Wai-Lok Lai, Wallace, Xavier Dérobert, and Peter Annan. 2018. "A Review of Ground Penetrating Radar Application in Civil Engineering: A 30-Year Journey from Locating and Testing to Imaging and Diagnosis." *NDT and E International* 96: 58–78.

Walsh, Brendan P., Sean N. Murray, and D. T.J. O'Sullivan. 2015. "The Water Energy Nexus, an ISO50001 Water Case Study and the Need for a Water Value System."

Water Resources and Industry 10: 15–28. http://dx.doi.org/10.1016/j.wri.2015.02.001.

Wang, Mengyu, Meng Wang, Lars J. Grimm, and Maciej A. Mazurowski. 2016. “A Computer Vision-Based Algorithm to Predict False Positive Errors in Radiology Trainees When Interpreting Digital Breast Tomosynthesis Cases.” *Expert Systems with Applications* 64: 490–99.

Ward, Sarah, David Butler, and Fayyaz Ali Memon. 2012. “Benchmarking Energy Consumption and CO2 Emissions from Rainwater-Harvesting Systems: An Improved Method by Proxy.” *Water and Environment Journal* 26(2): 184–90.

Wei, L I, Peng Minjun, L I U Yongkuo, and X I A Hong. 2017. “CONDITION MONITORING AND FALSE ALARM REDUCING OF SENSORS IN NPP.” : 914–26.

Wong, Leslie, Wing Kong Chiu, and Jayantha Kodikara. 2018. “Using Distributed Optical Fibre Sensor to Enhance Structural Health Monitoring of a Pipeline Subjected to Hydraulic Transient Excitation.” *Structural Health Monitoring* 17(2): 298–312.

Wu, Yipeng, Liu, Shuming, Wu, Xue, Liu, Youfei, and Guan, Yisheng. 2016. “Burst Detection in District Metering Areas Using a Data Driven Clustering Algorithm.” *Water Research* 100: 28–37. http://dx.doi.org/10.1016/j.watres.2016.05.016.

WWAP (United Nations World Water Assessment Programme). 2018. *Natuer-based Solutions for Water World Water Assessment Programme United Nations Educational, Scientific and Cultural Organization Sustainable Development Goals Water and Sanitation Ter Ent e Sustainable Development Goals Water and Sanitation*. www.unesco.org/open-access/.

Xiao, Sinan, Zhenzhou Lu, and Liyang Xu. 2017. “Multivariate Sensitivity Analysis Based on the Direction of Eigen Space through Principal Component Analysis.” *Reliability Engineering and System Safety* 165(March): 1–10.

Xiao, Wei. 2016. “A Probabilistic Machine Learning Approach to Detect Industrial Plant Faults.” *International Journal of Prognostics and Health Management* (c):

1–11.

Xing, Lu, and Lina Sela. 2019. “Unsteady Pressure Patterns Discovery from High-Frequency Sensing in Water Distribution Systems.” *Water Research* 158: 291–300. https://doi.org/10.1016/j.watres.2019.03.051.

Xu, Lixiang et al. 2020. “Probabilistic SVM Classifier Ensemble Selection Based on GMDH-Type Neural Network.” *Pattern Recognition* 106.

Xu, Luoliang, Mackenzie Mazur, Xinjun Chen, and Yong Chen. 2020. “Improving the Robustness of Fisheries Stock Assessment Models to Outliers in Input Data.” *Fisheries Research* 230(November 2019).

Xu, Xiaowei, Wangi, Hongxia, Zhangi, Nan, Liui, Zhenxing, and Wang, Xiaqing. 2017. “Review of the Fault Mechanism and Diagnostic Techniques for the Range Extender Hybrid Electric Vehicle.” *IEEE Access* 5: 14234–44.

Xu, Yan, Sun, Yanming, Wan, Jiafu, Liui, Xiaolong, and Song, Zhiting. 2017. “Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications.” *IEEE Access* 5: 17368–80.

Yamijala, Shridhar, Seth D. Guikema, and Kelly Brumbelow. 2009. “Statistical Models for the Analysis of Water Distribution System Pipe Break Data.” *Reliability Engineering and System Safety* 94(2): 282–93.

Yan, Lipeng. 2015. “Statistical Analysis and Optimization To Batch Processes.”

Yu, Yuebin, Denchai Woradechjumroen, and Daihong Yu. 2014. “A Review of Fault Detection and Diagnosis Methodologies on Air-Handling Units.” *Energy and Buildings* 82: 550–62.

Zare, Shokoofeh. 2018. “Scholarship at UWindsor Fault Detection and Diagnosis of Electric Drives Using Intelligent Machine Learning Approaches.” : 4–27. https://scholar.uwindsor.ca/etd/7436.

Zenobi, Renato. 2011. “Introduction to Multivariate Analysis.” *J Am Chem Soc* 8(5): 1–14.
http://books.google.com/books?hl=en&lr=&id=frOGrJTXwKcC&oi=fnd&pg=PA

3&dq=Introduction+to+Multivariate+Data+Analysis&ots=8TWX5iDGWp&sig=oBhpXW7ZWGtp2a8Dmu6eWnu-pLc%5Cnhttp://dx.doi.org/10.1016/j.pid.2012.12.004.

Zhang, Haili, Qia, Yongsheng, Wanga, Lin, Gao, Xuejin, and Wang, Xichang. 2017. "Fault Detection and Diagnosis of Chemical Process Using Enhanced KECA." *Chemometrics and Intelligent Laboratory Systems* 161(December 2016): 61–69. http://linkinghub.elsevier.com/retrieve/pii/S0169743916301873.

Zhang, Hangkui, Zhu, Li, Xu, Senwei, Cao, Jianting, and Kong, Wanzeng. 2021. "Two Brains, One Target: Design of a Multi-Level Information Fusion Model Based on Dual-Subject RSVP." *Journal of neuroscience methods* 363(April): 109346. https://doi.org/10.1016/j.jneumeth.2021.109346.

Zhang, Jing, James I. Silber, and Maciej A. Mazurowski. 2015. "Modeling False Positive Error Making Patterns in Radiology Trainees for Improved Mammography Education." *Journal of Biomedical Informatics* 54: 50–57.

Zhang, Kai, Shardt, Yuri A. W., Chen, Zhiwen, Ding, Steven X., and Peng, Kaixiang. 2017. "Assessment of T^2- and Q-Statistics for Detecting Additive and Multiplicative Faults in Multivariate Statistical Process Monitoring." *Journal of the Franklin Institute* 354(2): 668–88.

Zhang, Kai, Shardt, Yuri A. W., Chen, Zhiwen, Ding, Steven X., and Peng, Kaixiang. 2016. "A Brief Survey of Different Statistics for Detecting Multiplicative Faults in Multivariate Statistical Process Monitoring." *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (Cdc): 2152–57.

Zhao, Jinxiong, Sensen Guo, and Dejun Mu. 2021. "DouBiGRU-A: Software Defect Detection Algorithm Based on Attention Mechanism and Double BiGRU." *Computers and Security* 111: 102459. https://doi.org/10.1016/j.cose.2021.102459.

Zhao, Junqi, and Esther Obonyo. 2020. "Convolutional Long Short-Term Memory Model for Recognizing Construction Workers ' Postures from Wearable Inertial Measurement Units." 46(December 2019).

Zhao, Junqi, and Obonyo, Esther. 2021. "Applying Incremental Deep Neural Networks-

Based Posture Recognition Model for Ergonomics Risk Assessment in Construction." *Advanced Engineering Informatics* 50(May): 101374. https://doi.org/10.1016/j.aei.2021.101374.

Zhao, Yang, Tingting Li, Xuejun Zhang, and Chaobo Zhang. 2019. "Artificial Intelligence-Based Fault Detection and Diagnosis Methods for Building Energy Systems: Advantages, Challenges and the Future." *Renewable and Sustainable Energy Reviews* 109(February): 85–101. https://doi.org/10.1016/j.rser.2019.04.021.

Zheng, K. 2020. "A SVM Based Ship Collision Risk Assessment Algorithm." *Journal of Chemical Information and Modeling* 53(9): 287.

Zhou, Anmei, Dejie Yu, and Wenyi Zhang. 2015. "A Research on Intelligent Fault Diagnosis of Wind Turbines Based on Ontology and FMECA." *Advanced Engineering Informatics* 29(1): 115–25. http://dx.doi.org/10.1016/j.aei.2014.10.001.

Zhou, Xiao, Tang, Zhenheng, Xu, Weirong, Meng, Fanlin, Chu, Xiaowen, Xin, Kunlun, and Fu, Guangtao. 2019. "Deep Learning Identifies Accurate Burst Locations in Water Distribution Networks." *Water Research* 166: 115058. https://doi.org/10.1016/j.watres.2019.115058.

Zhu, Jinlin, Zhiqiang Ge, Zhihuan Song, and Furong Gao. 2018. "Review and Big Data Perspectives on Robust Data Mining Approaches for Industrial Process Modeling with Outliers and Missing Data." *Annual Reviews in Control* 46: 107–33. https://www.sciencedirect.com/science/article/pii/S1367578818301056?via%3Dihub (December 19, 2019).

Zhu, Mu, and Ali Ghodsi. 2006. "Automatic Dimensionality Selection from the Scree Plot via the Use of Profile Likelihood." 51: 918–30.

Zumoffen, David, and Marta Basualdo. 2008. "From Large Chemical Plant Data to Fault Diagnosis Integrated to Decentralized Fault-Tolerant Control: Pulp Mill Process Application." *Industrial and Engineering Chemistry Research* 47(4): 1201–20.

APPENDIX

Appendix

The mechanics of principal component analysis

Computation of all feature values for water consumption, gives a data matrix of $X \in \mathrm{R}^{\mathrm{n} \times \mathrm{m}}$ whereby *n* rows represent the values for all features within a single day and *m* columns represent a single feature such as maximum flow, average flow, etc. over the observation with mean zero and unit variance.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{im} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{nm} \end{bmatrix} = (w_1 \quad w_2 \quad \cdots \quad w_j \ldots \quad w_m) \tag{1}$$

where row vector X_i represents all feature measurements (i.e., daily flow, etc.) at specific time intervals and column vector w_j represents single feature measurements (average flow, night time flow, etc.) in a series of days. Generally, data matrix $X \in \mathrm{R}^{\mathrm{n} \times \mathrm{m}}$ contain different magnitude and scales of the physical features. The feature matrix was standardized using a z-score standardizing technique (also known as auto-scaling), transforming the data to have a mean zero and a unit variance (Kamiel, 2015).

$$\mu_{wj} = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \tag{2}$$

$$\sigma_{wj} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(x_{ij} - \mu_{wj}\right)^2} \tag{3}$$

$$\overline{x_{ij}} = \frac{x_{ij} - \mu_{wj}}{\sigma_{wj}} \tag{4}$$

where $\overline{x_{ij}}$ is the data point re-scaled to $\mu_{wj} = 0$ and $\sigma_{wj} = 1$, μ_{wj} is the mean and σ_{wj} is the standard deviation of the variable w_j. The standardized data matrix $X \in \mathrm{R}^{n\times m}$ is then transformed by utilizing singular value decomposition as per (Jackson & Mudholkar, 1979; Zenobi et al., 2011; Sheriff et al., 2017).

$$X = UP^T + E \tag{5}$$

where $U = [u_1, u_2, u_3, \ldots u_m] \in \mathrm{R}^{n\times m}$ is a matrix of transformed variables, where each column represents the score vectors and the i^{th} eigenvalue equals the square of the i^{th} singular value (i.e., $\lambda_i = \sigma_i^2$). $P = [p_1, p_2, p_3, \ldots p_m] \in \mathrm{R}^{m\times m}$ is the matrix of orthogonal vectors, where each column is populated by the eigenvectors associated with the covariance matrix of the data matrix $X \in \mathrm{R}^{n\times m}$. E is the residual matrix which ideally contain noise in the data. Typically, most of the data variance is contained in the principal components with larger eigenvalues, while the remaining principal components are considered as measurement noise which can be removed by reducing the data dimension (Grueiro et al., 2018; Laory et al., 2011). The covariance matrix S quantifies the amount of linear correlation between all possible combinations of features within the dataset and can be computed as (Sengupta & Kundu, 2016; Rosen, 2001; Jackson & Mudholkar, 1979).

$$S = \frac{1}{n-1}X^T X = P\Lambda P^T \tag{6}$$

where Λ is the diagonal matrix $\Lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_m) \in \mathrm{R}^{m\times m}$ containing non-negative eigenvalues related to m principal components and magnitudes in a descending order $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0$. The covariance matrix is a symmetric and square matrix and can also be expressed as.

$$S = \begin{bmatrix} w_1^T w_1 & w_1^T w_2 & \ldots & w_1^T w_j & \ldots & w_1^T w_m \\ w_2^T w_1 & w_2^T w_2 & \ldots & w_2^T w_j & \ldots & w_2^T w_m \\ \ldots & \ldots & & \ldots \quad \ldots & \ldots & \ldots \\ w_j^T w_1 & w_j^T w_2 & \ldots & w_j^T w_j & \ldots & w_j^T w_m \\ \ldots & \ldots & & \ldots \quad \ldots & \ldots & \ldots \\ w_m^T w_1 & w_n^T w_2 & \ldots & w_n^T w_j & \ldots & w_n^T w_m \end{bmatrix} \tag{7}$$

where Λ is the diagonal matrix Λ = diag (λ_1, λ_2, λ_3 … λ_m) $\in \mathrm{R}^{m\times m}$ containing non-negative eigenvalues related to *m* principal components and magnitudes in a descending order $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0$. The model built by principal component analysis contains the same number of principal components as the original number of features in the data matrix $X \in \mathrm{R}^{n\times m}$. In the case of water distribution systems, it may contain features such as average flow, average flow during working hours, average flow during non-working hours, etc. that are highly correlated. A small number of principal components can be used to detect flow deviation in the water consumption data. However, it is critical to choose the number of principal components *r*. Overestimation could introduce false alarm and noise (misdetection) in the model, while, underestimating could decrease the detection capability of the PCA model (Jolliffe, 2002.; Zhu & Ghodsi, 2006).

In this study, the number of components is selected based on the cumulative percent variance (CPV) due to its computational simplicity and wide adoptability (discussed in Chapter 3 – Section 3.3.2.2) (Sheriff et al. 2017). It provides a good estimate of the number of principal components that need to be retained for most practical applications. Cumulative percent variance is a measure of the percentage variance CPV(r)≥ 90% captured by the first *r* principal components (Zumoffen & Basualdo, 2008; Johnson & Wichern, 2007).

$$CPV(r) = \frac{\sum_{i=1}^{r} \lambda_i}{trace(S)} 100 \tag{8}$$

The number of principal components (PCs) is selected in a way that CPV is greater than the minimum amount of variation the model should explain (i.e., CPV(r)≥ 90%). Once the number of principal components *r* were retained, the data matrix $X \in \mathrm{R}^{n\times m}$ can be expressed as (Jolliffe, 2002; Jackson & Mudholkar, 1979).

$$X = UP = \left[\hat{U}\ \tilde{U}\right]\left[\hat{P}\ \tilde{P}\right] = \overbrace{X\hat{P}\hat{P}^T}^{\hat{X}} + \overbrace{X(I_m - \hat{P}\hat{P}^T)}^{E} \tag{9}$$

where $\hat{U} \in \mathrm{R}^{n\times r}$ represents the matrices containing *r* retained principal components and $\tilde{U} \in \mathrm{R}^{(n\times m)-r}$ represents the ignored *m-r* principal components. Similarly, $\hat{P} \in \mathrm{R}^{m\times r}$ represents the matrices containing the *r* retained eigenvectors and $\tilde{P} \in \mathrm{R}^{(n\times m)-r}$ represents the ignored *m-r* eigenvectors, respectively. $\hat{X} \in \mathrm{R}^{n\times r}$ is the modelled matrix established by projecting matrix *X* onto the reduced loading matrix *P* containing retained principal components (principal component subspace) and E is the residual matrix computed as the difference between X and $\hat{X}$, captures the variations associated with n-r singular values. The matrix $\hat{X} \in \mathrm{R}^{\mathrm{n}\times\mathrm{r}}$ is the variation of $X \in \mathrm{R}^{\mathrm{n}\times\mathrm{m}}$ calculated utilizing r retained principal components.

Score and orthogonal distance

To analyse the linear combination of water time series that restrain valuable information or outliers; the score distance SD_i and the orthogonal distance OD_i of each data point to the PCA subspace are given by (Hubert et al., 2005; Harris et al., 2014; Rousseeuw et al., 2018).

$$SD_i = \sqrt{\sum_{j=1}^{r} \frac{u_{ij}^2}{\lambda_j}} \tag{10}$$

$$OD_i = \left\| x_i - \hat{\mu} - P_{m,r} u_i' \right\| \tag{11}$$

where u_{ij} is the score value of each data point, λ_j is the non-negative eigenvalues and $P_{m,r}$ is the eigenvectors matrix related to r principal components, $\hat{\mu}$ is the mean of the covariance matrix. Assuming the data follows a multivariate normal distribution. The cut-off for the data points were estimated using Chi-square distribution $\sqrt{X_{r,0.975}^2}$. Where $X_{r,0.975}^2$ is the 97.5% quantile of the Chi-squared distribution with r principal components (Rousseeuw and Hubert, 2018).

Fault detection indices

Hotelling T^2-statistics

T^2-statistics represents the major variation in the data (Garcia-Alvarez, 2014). The T^2-statistics of the i^{th} sample or observation x can be expressed as (Mujica et al., 2011; Qin, 2003).

$$T_i^2 = x_i^T P \, \Lambda_r^{-1} P^T x_i \tag{12}$$

where Λ_r^{-1} is the diagonal matrix containing the eigenvalues related to retained principal components, x_i is the data vector of the i^{th} observation and P contains the loading vector associated with the r columns. Under routine process conditions, the data follow a multivariate normal distribution, the T^2-statistics is related to an F-distribution considering that the population mean, and covariance are estimated from data (Qin, 2003).

The control limit T_α^2 can be obtained by F-distribution as follows.

$$T_\alpha^2 = \frac{r(n-1)}{n-r} F_{\alpha(r,n-r)} \tag{13}$$

where n is the number of observations in the data, r is the number of retained principal components, $F_{\alpha\,(r,\,n-r)}$ is the F-distribution. In the current study, the α-control limits are calculated at a confidence level of 95%. The T^2-statistics can be interpreted as measuring the systematic variations of the process. Violation of routine conditions would indicate that the systematic variations are outside normal operating bounds, and thus the data set may indicate a fault or non-routine process condition (Vieira et al., 2014). The new observation is considered to be normal if it satisfies the following condition.

$$\begin{cases} T_i^2 < T_\alpha^2 --- Normal \\ T_i^2 \geq T_\alpha^2 --- System\ alarm \end{cases} \tag{14}$$

T^2-statistics is based on the first r retained principal components so that it provides a test for derivations in the principal score vectors that are of greatest importance to the variance of the process. When the T_i^2 of the new observation exceeds the control limit, T_α^2, this triggers an alarm. For instance, if new incoming data contains a fault or non-routine water uses (conference, workshop, etc.) which consumes more water than the routine consumption, the data will reflect substantial deviation from normal operating values for the statistical features (**Table 4.1**). The T^2-statistics is used to evaluate the variation in the statistical feature values predefined by the baseline condition. T^2-statistics will only detect an alarm if the variation in the latent variables is greater than the variation explained by common causes. False negatives can occur, whereby the variation in data output (i.e., daily flow rate) is not enough to be detected by the selected control limit dictated by T^2_α. To reduce false negatives, and thus increase the sensitivity of the water distribution system fault detection methodology, the Q-statistics can be utilised. This Q-statistics is based on the observation residuals not detected by T^2-statistics, which is the sum of squares of the residuals as discussed below (Zenobi et al., 2011).

Q-statistic or squared prediction error

The Q-statistics also known as squared prediction errors, measure the variability of the observation that violates the routine process correlation (with small or moderated magnitudes) not accounted by the principal component subspace. In case of water distribution systems, continuous flow usually does not impact on the overall water consumption in a short period and so may not be detected by conventional methods, however a low-grade fault, such as minor continuous flows, can result in undesirable water loss and accumulate to significant losses. The portion of the measurement space corresponding to the lowest (*m* - *r*) eigenvalues are thus monitored (Ballabio, 2015; Qin, 2012).

$$Q_i = x_i^T (I - P_i P_i^T) x_i \tag{15}$$

where *I* is the identity matrix. The control limit Q_α can be computed from its approximate distribution (Ballabio, 2015; Qin, 2012).

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 \ \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \tag{16}$$

$$\theta_i = \sum\nolimits_{j=r+1}^{n} \lambda_j^{2i} \tag{17}$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \tag{18}$$

where *r* is retained principal components inside the model and c_α is the standard normal deviation with the upper (1 - α) percentile. When an unusual event occurs, and it produces a change in the covariance structure of the model, it will be detected by a high value of Q. The new observation is considered to be normal if it satisfies the following condition.

$$\begin{cases} Q_i < Q_\alpha --- Normal \\ Q_i \geq Q_\alpha --- \text{System alarm} \end{cases} \tag{19}$$

when the Q_i of the new experimental trial violates the Q_α control limit, this is indicative of a fault, or a non-routine water uses. The value of Q_α is defined with an assumption that the observation data is multivariate normally distributed and time-independent (Harrou et al., 2013). The value Q-statistics are small and consequently more sensitive than the T^2-statistics. This characteristic makes the Q-statistics ideally suited to detecting minor variation within the system behaviour. It evaluates the variation of the new incoming water data which are not accounted for by the principal component subspace. On the contrary, the T^2-statistics require significant variation in the system behaviour to be measurable (Mujica et al., 2011).

Support vector machines

Support vector machine (SVM) is a supervised learning technique widely used for classification and regression. The SVM technique has been used herein due to its wide use in different engineering sectors such as thermal power plants, centrifugal pumps, bearing, etc. in context of classification (Sabri et al., 2017; Xiao, 2016; Bayar et al., 2015; Chen et al., 2011, Hmeidi et al., 2008). SVM creates a decision boundary in between classes by mapping the training data (through kernel function) onto a higher dimensional space, and then obtaining the maximum margin hyperplane within that space. Thereafter, this hyperplane can be regarded as a classifier. To achieve multi-class classification, error-correcting output code (ECOC) was used. ECOC represents a powerful framework to deal with multi-class classification problems combining binary classifiers such as support vector machine, neural networks, decision trees, etc. (Bagheri et al., 2012; Lin, 2018).

The classification problem in SVM can be confined by considering the two-class problem which aims to separate the two classes by a function induced from available training data (Zheng, 2020). An example of linearly separable problem in a 2D space is shown below. The SVM is based on solving the optimization problem which tends to maximize the margin width as shown in **Figure** and minimize the classification error. SVM usually consider the hyperplane with maximum margin (hyperplane of larger distance as possible). A detailed explanation of the SVM technique can be found in (Bishop, 2013).

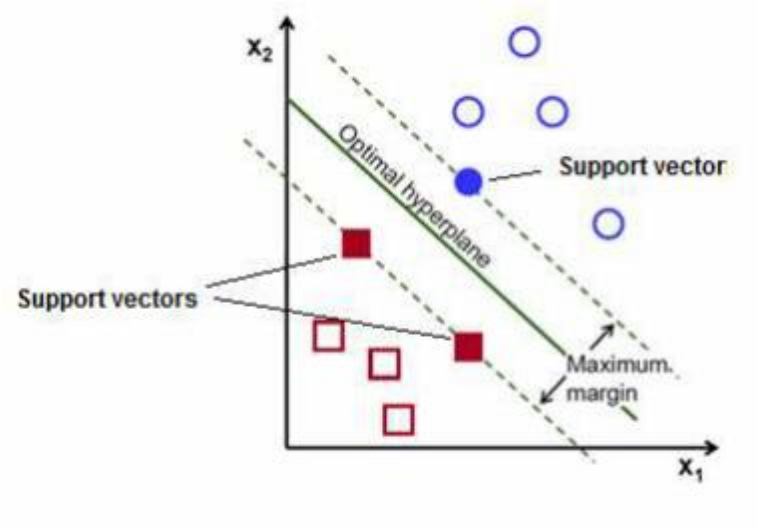


Figure: An example of a linearly separable problem in a 2-D space. The support vectors, marked with red squares, define the margin of largest separation between the two classes (Cortes and Vapnik, 1995).

Prior to classification among routine or faulty condition, cross-validation was performed to optimize classifier's hyperparameters and to assess the performance of a classifier. Cross-validation ensures that SVM classifier can be trained rapidly with less computational effort while maintaining a good accuracy of the classifier. It is noted that SVM is less effective when the data contain noise and overlapping data points (Nalepa & Kawulok, 2018; Nayak et al., 2015).

Cross-Validation

Cross validation is a statistical method for estimating the prediction accuracy of the classifier. The goal of the cross-validation is to gauge the generalizability of an algorithm (SVM) and to prevent overfitting by minimizing the influence of noisy data points (Gupta, 2019; Mudry & Tjellström, 2011). During classifier training, suitable kernel, regularization (gamma) and soft margin (cost) parameters are required to determine whether the classifier predicts unknown data accurately. However, the optimal parameters are generally not known beforehand; but can be developed using a cross-validation algorithm (Ding and Chen, 2010). In this study, the *f*-fold cross-validation technique was used to evaluate the accuracy of the classifier. To achieve multi-class classification, ECOC was used. ECOC represents a powerful framework to deal with multi-class classification problems based on combining binary classifiers such as SVM, neural networks, decision trees, etc. (Bagheri et al., 2012; Lin, 2018).

In the process of *f*-fold cross-validation, the original training dataset was partitioned into *f* equally sized segments (i.e., f = 1,2,3…). Subsequently, *f* iterations of training and validation were performed such that within each iteration a different set of the data was held-out for validation while the remaining (*f*-1) folds were utilized for learning (Wei Li et al., 2019). Data is generally stratified before being split into *f*-folds. Stratification is the process of reorganizing the data ensuring each fold is a good representation of the whole dataset (Mudry and Tjellström, 2011). In this study, the SVM classifier was trained by conducting stratified 10-fold cross validation to minimize the generalization error (minimizing the error associated with the bias and variance of data).