College of Engineering, Mathematics and Physical Sciences

Events Recognition System for Water Treatment Works

Submitted by Gerald Riss to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Engineering
in November 2019

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

Signature: …………………………………………………………...

ABSTRACT

The supply of drinking water in sufficient quantity and required quality is a challenging task for water companies. Tackling this task successfully depends largely on ensuring a continuous high quality level of water treatment at Water Treatment Works (WTW). Therefore, processes at WTWs are highly automated and controlled. A reliable and rapid detection of faulty sensor data and failure events at WTWs processes is of prime importance for its efficient and effective operation. Therefore, the vast majority of WTWs operated in the UK make use of event detection systems that automatically generate alarms after the detection of abnormal behaviour on observed signals to ensure an early detection of WTW's process failures. Event detection systems usually deployed at WTWs apply thresholds to the monitored signals for the recognition of WTW's faulty processes.

The research work described in this thesis investigates new methods for near real-time event detection at WTWs by the implementation of statistical process control and machine learning techniques applied for an automated near real-time recognition of failure events at WTWs processes. The resulting novel Hybrid CUSUM Event Recognition System (HC-ERS) makes use of new online sensor data validation and pre-processing techniques and utilises two distinct detection methodologies: first for fault detection on individual signals and second for the recognition of faulty processes and events at WTWs.

The fault detection methodology automatically detects abnormal behaviour of observed water quality parameters in near real-time using the data of the corresponding sensors that is online validated and pre-processed. The methodology utilises CUSUM control charts to predict the presence of faults by tracking the variation of each signal individually to identify abnormal shifts in its mean. The basic CUSUM methodology was refined by investigating optimised interdependent parameters for each signal individually. The combined predictions of CUSUM fault detection on individual signals serves the basis for application of the second event detection methodology. The second event detection methodology automatically identifies faults at WTW's processes respectively failure events at WTWs in near real-time, utilising the faults detected by CUSUM fault detection on individual signals beforehand. The method applies Random Forest classifiers to predict the presence of an event at WTW's processes.

All methods have been developed to be generic and generalising well across different drinking water treatment processes at WTWs. HC-ERS has proved to be effective in the detection of failure events at WTWs demonstrated by the application on real data of water quality signals with historical events from a UK's WTWs. The methodology achieved a peak F_1 value of 0.84 and generates 0.3 false alarms per week. These results demonstrate the ability of method to automatically and reliably detect failure events at WTW's processes in near real-time and also show promise for practical application of the HC-ERS in industry. The combination of both methodologies presents a unique contribution to the field of near real-time event detection at WTW.

ACKNOWLEDGEMENTS

I would like to thank these people, without whom this work would not have been possible:

Prof. Zoran Kapelan and Dr Michele Romano

for being my supervisors over the last four years and their guidance through this research work with great support and patience, but also for always encouraging me to take the next step and most importantly for their relentless commitment to provide always valuable advice.

Prof. Fayyaz Ali Memon

for opening my mind and enabling me with great spirit to the research on this issue and also for choosing to supervise me through the last years of this work.

Prof. Gino Hrkac

for being my mentor and for listening to all kinds of questions throughout the four years of mentoring.

Dr Diego Gomez and Kevin Woodward

for being my supervisors over the first stages of the work and for the valuable advice provided throughout the time of supervision.

Laura Conroy and the whole Innovation Team of United Utilities

for their willingness to support and help on all questions and queries so promptly during my stage at United Utilities.

My family for their support and patience experiencing all of the ups and downs of the research work.

I also wish to extend my thanks to all the people who cooperated with me in this regard.

Finally, I would like to thank the Centre for Water Systems at the University of Exeter and STREAM IDC as well as the Engineering and Physical Sciences Research Council and United Utilities, not only for providing the funding which allowed me to undertake this research, but also for giving me the opportunity to attend events and meet so many interesting people.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

TABLE OF ABBREVIATIONS

ACM	Abnormal Condition Management
AEM	Abnormal Event Management
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARL	Average Run Length
BED	Binomial Event Discriminator
CBR	Case-Based Reasoning
DIGRAPH	Directed Graph
DO	Dissolved Oxygen
DOC	Dissolved Organic Carbon
EC	Electrical Conductivity
ERS	Event Recognition System
ETA	Event-Tree Analysis
EWMA	Exponentially Weighted Moving Average
FDD	Fault Detection and Diagnosis
FDI	Fault Detection and Isolation
FTA	Fault tree Analysis
HCERS	Hybrid CUSUM Event Recognition System
ICA	Independent Component Analysis
k-NN	k-Nearest Neighbour
LPCF	Linear Prediction-Correction Filter
MLP	Multilayer Perceptron
MSPC	Multivariate Statistical Process Control
MSPCA	MultiScale PCA
MVNN	Multivariate Nearest Neighbour

NN	Neural Networks
ORP	Oxidation Reduction Potential
PCA	Principle Component Analysis
PI	Performance Indicator
PLS	Partial Least Squares
QPT	Qualitative Process Theory
QSIM	Qualitative Simulation
QTA	Qualitative Trend Analysis
RF	Random Forest
SPC	Statistical Process Control
SVM	Support Vector Machine
TC	Total Chlorine
TOC	Total Organic Carbon
TU	Turbidity
UU	United Utilities Group PLC
WDS	Water Distribution System
WTW	Water Treatment Works
WWTW	Wastewater Treatment Works

Chapter 1: Introduction

1 INTRODUCTION

1.1 Motivation

Water as source of life is crucial for all development and progressing on earth. For this reason, a safe drinking water supply is essential not only for human health and well-being, but also for the sake of economic and social development of the whole mankind. However, for a safe drinking water supply, producing water in the required quality and quantity has to be ensured. Nowadays, with a world population of over 7 billion people, water is scare and in many parts of the world considerable polluted. In 2015 around 660 million people lack access to clean water (WHO/UNICEF, 2015) and the World Health Organization estimated that in 2008 worldwide 2.5 million people died from water-related diseases (WHO, 2012). In order to ensure that drinking water can be consumed safely with a high level of health protection, certain standards, based on World Health Organisation's guidelines for Drinking-Water Quality, have been defined by several authorities in Europe and the UK.

The Drinking Water Directive (Council Directive 98/83/EC on the quality of water intended for human consumption with recently revised Annex II 'Monitoring' and Annex III 'Specifications for the analysis of parameters') aims to protect human health from different effects of any contamination of water by monitoring and frequently testing of 48 microbiological, chemical and indicator parameters. The requirements of the Drinking Water Directive have been transferred into the Water Supply (Water Quality) Regulations 2010 in England and Wales, recently revoked and replaced by the Water Supply (Water Quality) Regulations 2018.

However, it is always a great challenge for water utilities to produce safe drinking water in an efficient and effective way. In addition to the issue of water scarcity, that makes major changes in the use, management and distribution of water necessary (United Nations, 2015), many factors can affect the quality of drinking water. For example, natural events such as heavy rainfalls, storms, hurricanes, droughts, flooding, earthquakes or manmade threats like accidents, operation or maintenance errors, point or non-point pollution to water sources and related infrastructure, but also threats of terrorist attacks or chemical, biological, and radiological contamination pose potential risks for the drinking water quality.

Moreover, the realisation that the impact of climate change will lead to a rise in major natural disasters and hydrological variability, evinced by an increased frequency and amplitude of droughts, floods and hurricanes raised the awareness of water industry in this regard.

The growing concern over water pollution has triggered regulatory action by the revision of the mentioned Drinking Water Directive with the focus on establishing specifications for the controlling of water quality parameters as well as on monitoring of microbiological and chemical parameters. Since the legislation always lag behind the recent technological developments, online monitoring technologies currently play a limited role here, although future developments may lead to technologies which can adequately and cost-effectively monitor these parameters (EIP Water Action Group, 2015). It is expected that online monitoring technologies will be enshrined in law in near future, latest by the next review of the European Drinking Water Directive in 2020.

Although early detection of water contamination and pollution events is the most important, but it is not the only reason that motivates water companies to pursue a reliable and timely detection of failure events at their WTWs. Water companies are not only challenged with the day by day task to produce water in required quantity and quality by WTW's operation at minimum costs but also by the handling of exceptional situations, e.g. in case of WTWs processes need to be shut down to prevent failures before affecting customers and/or environment. Especially in these situations, ultimate attention is demanded from water companies, because they will be judged by their customers and the regulative authority how well they manage such adverse situations. Frequent interruptions in water supply will cause not only a rise in operational costs, but also will lead to fines from the regulator (i.e., DWI) and a bad image in customer's eyes. Precisely these aspects, i.e. the increase of WTW's operational efficiency, improvement of customer service and avoidance of regulatory fines are the main drivers that the research work described in this thesis was conducted.

Early detections of failure events at WTWs processes offer water companies the opportunity to reduce the number of unplanned WTW's shutdowns and corresponding interruptions in water supply to customers and enable them to carry out proactive interventions, e.g. to address issues before they reach a critical point where the WTWs may need to be shut down.

The mentioned increase of WTW's operational efficiency, improved customer service associated with less fines from the regulator will likely result in significant financial savings for water companies. But apart from the monetary aspects of cost savings, water companies will encourage and speeding up the cultural change in the water sector required for the implementation of smart water technologies. Furthermore, the new technology will give water companies a technology lead on UK's and worldwide water engineering markets. The latter aspects are also very important drivers for the research carried out in this thesis, since the benefits from a positive image in the public resulting from company's promotion to be 'innovative' counts maybe more than any monetary benefits.

1.2 Background

Near real time detection of faulty sensors or processes at WTWs is of greatest significance for water supply companies. Due to several factors, such as frequently varying water demand, changing influent conditions, dynamics in water treatment processes and imperfect, missing or incorrect sensor data, controlling of WTWs is a difficult task for water companies to manage.

The importance of controlling WTWs processes in a timely manner can be illustrated by following example. Rapidly changing influent conditions, e.g. caused by a sudden rainstorm event, have instantaneous impact on WTW's processes. Once a rainstorm appears, a fundamental increase of influent raw turbidity will follow, combined with a raise of Dissolved Organic Carbon (DOC) compounds as shown in Figure 1-1. These changing conditions have significant impact on WTW's treatment processes (Parsons and Jefferson, 2006) and therefore for WTW's operation. Coagulation and flocculation processes need to be adjusted rapidly by elevating the dosage of chemical coagulants and thus reduce turbidity levels to finally meet compliance with the standards on drinking water quality.

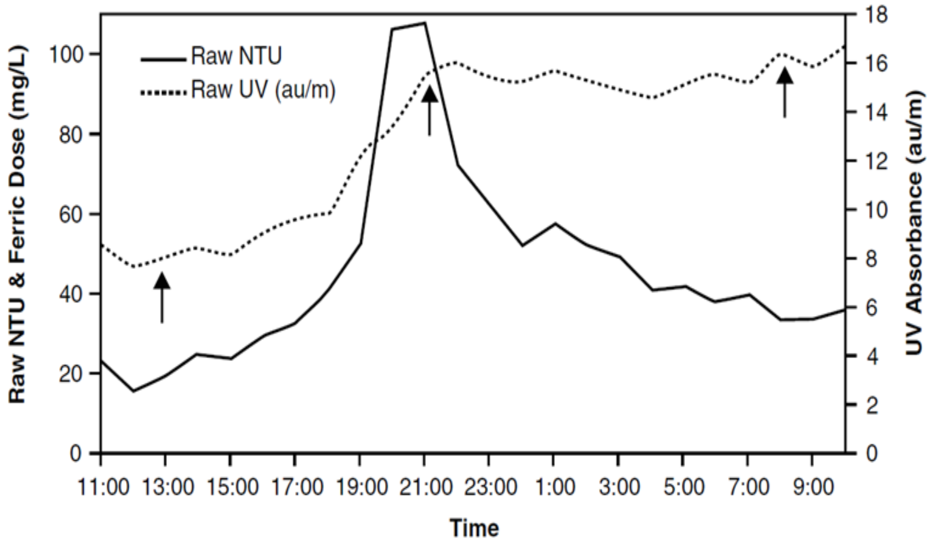

Figure 1-1 *Raw water quality changes during rainstorm event (Parsons and Jefferson, 2006).*

This example illustrates one of the challenges water companies have to face in their day by day operation to produce water in the required quantity and quality at lowest expenditure. To achieve this, WTWs are already heavily monitored and automated with a high level of process optimisation and are therefore heavily dependent on reliable and accurate sensor data. Usually, Supervisory Control and Data Acquisition Systems (SCADA) are used to control treatment processes by monitoring of the critical water quality and flow parameters in near real-time, but the data is often imperfect collected by SCADA systems (Romano, Kapelan and Savic, 2014).

In UK, most systems used for the recognition of failure events at WTWs usually apply thresholds to generate alarms after detecting abnormal behaviour on observed signals. However, similar to various near real-time applications threshold-based event detection systems have this major drawback that they are either frequently robust to minor degree or sensitive to high frequency influences associated with a high level of false alarms, since quick fault detection and robustness of the system are two conflictive goals (Venkatasubramanian et al., 2003a). Results of the ‘Water quality event detection system challenge’ report published by U.S. Environmental Protection Agency (EPA) (EPA, 2013) have shown that event detection performances of the participating detection systems vary greatly, and the number of false alarms produced be these systems is generally still high. New and more efficient technologies need to be designed to address this issue.

A broad range of fault detection techniques have been already developed (Venkatasubramanian et al., 2003a,b,c) (Miljkovic, 2011) (Maiti and Banerjee, 2012) (Sin et al., 2012), but only a few were applied for event detection at WTWS in industry. This first generation of applications suffers from a range of shortcomings (Bernard et al., 2015), i.e. none of these methods is optimal and therefore only a limited number is practiced by water companies. In consequence - and also motivated by recently occurred events, e.g. Cryptosporidium contamination in 2015 - an increased interest on improved event detection technologies has induced the need of further research with the focus set on innovative, cost-effective near real-time event recognition systems for water treatment processes. For this reason, the development of a new technology for event recognition at WTWs is a strategic priority for water utilities.

Several methods already exist with great potential to address above issues. Statistical Process Control (SPC) and Artificial Intelligence (AI), especially machine learning techniques seem to be particularly suitable for promising improvements, since they can extract information useful for operational decisions and are usually able to deal efficiently with imperfect sensor data collected by SCADA (Romano, Kapelan and Savic, 2014). Although first investigations of Artificial Neural Networks (ANNs) for monitoring and controlling WTWs filtration processes started already in 2001 (Lennox et al., 2001), most of machine learning techniques applied for event detection in the water sector began only recently to appear (Oliker and Ostfeld, 2014) (Liu, Smith and Che, 2015a) (Liu et al., 2015) with demonstrating continuously promising results.

Therefore, the focus of the research work that is conducted in this thesis is set on the development of a new technology for improved near real-time recognition of faulty sensor data and faulty processes at WTWs which combines well-established SPC methods for fault detection with new machine learning techniques for event classification with the aim to address particularly following key research questions:

1) What is the best way to identify faulty WTW sensor data and how and when this data can be trusted?

2) How reliably and how quickly can degradations in effluent quality, abnormal variations in effluent quality and faulty processes at WTW be detected?

3) Is it possible and how best to distinguish a faulty sensor from a faulty process?

4) What are the likely benefits of the new technology and can the new technology easily and efficiently be integrated with current industry practices?

1.3 Scope and Objectives

The overall aim of this work is to develop and validate new methodology and a near real-time event recognition system to detect faulty sensor data and faulty processes at Water Treatment Works (WTWs). The new technology will enable water companies to carry out more proactive interventions in operation the WTWs with the potential to prevent failures before they impact customers or environment and will lead to an increased efficiency in WTW's operation, safe money and improved customer service as consequence from reduced water supply interruptions.

The developed methodology should be able to identify faulty WTWs sensor data and distinguish between faulty sensor data and faulty WTW processes in a timely and reliable manner. Furthermore, the new technology has to be cheap in implementation and operation.

The specific objectives are as follows:

1) To investigate the quality and adequacy of historical data collected from various sensors deployed at different WTWs treatment stages, observing water quality parameters, such as, pH, turbidity, iron and Chlorine. This will be done to define normal and abnormal WTW's process conditions and identify data streams that will be used in the thesis In addition, the objective is to use the processed data to define a set of minor and major failure type events that will be used for the development and testing of a new detection method;

2) To establish the baseline by assessing the performance of the existing, threshold-based event detection system by using selected WTW's historical sensor data and events.

3) To develop new methodology capable for WTW's sensor data validation in near real-time by checking, detecting and rectifying erroneous sensor readings, missing data and unusual spikes (outliers) as well as by identifying sequenced constant sensor measurements (flat line faults).

4) To develop methodology for the automated near real-time recognition (i.e. detection) of failure events at WTWs processes. The new technology should detect faulty events at WTWs processes in a reliable and timely manner associated with low false alarm rates.

5) To test, validate and demonstrate the above methodologies on unseen historical data and failure events. These demonstrations should display the effectivity and efficiency of each technique by the assessment of its detection performance.

1.4 Thesis Structure

The remainder of this thesis is structured as follows:

Chapter 2: Literature Review

The literature review on fault detection methods in general and on specific fault detection approaches in the water sector explores key gaps in knowledge in the field of detection of faulty events at water treatment works.

Chapter 3: Case Study Description and Data

Introduces the demonstration site and datasets used throughout this thesis. The chapter describes the methods applied for investigating the quality and adequacy of historical data collected from various sensors deployed at different WTWs treatment stages and the data streams identified this way used for defining a set of minor and major failure type events utilised for the development and testing of a new detection method.

Chapter 4: Event Recognition Methodology

The chapter outlines the methodologies used for (a) establishing the baseline of the existing, threshold-based event detection system, (b) developing the new WTW's sensor data validation methodology in near real-time and (c) developing the novel Hybrid CUSUM ERS (HC-ERS).

Chapter 5: Case Study Applications

In the fifth chapter, the event detection capabilities resulting from the testing and validation of threshold based, modified and novel hybrid CUSUM detection systems are demonstrated on corresponding case studies. This chapter also describes the assessment of the detection performance of the Hybrid CUSUM method by comparing the results with those of the well-established CANARY event detection method.

Chapter 6: Conclusion

The final chapter summarises the developments and results obtained in the course of the research and presents the main conclusions and contributions of this work to the research field. Finally, this final chapter outlines the potential for recommended further research.

Chapter 2: Literature Review

2 LITERATURE REVIEW

2.1 Introduction

This review of literature introduces fault detection methods in general and outlines previous work done in the area of fault detection in the water sector aiming to identify key gaps in knowledge in detection of faulty events at water treatment works.

Against the background of a long experience in the field of process control, automated fault detection has become of increasing interest to water industry since the past 20 years. Various techniques for fault detection at water treatment works have been developed and presented (see, e.g., Lennox et al., 2001; George, Chen and Shaw, 2009; Housh and Ostfeld, 2015). Fault detection methodologies have constantly been evolved, in particular by the application of machine learning techniques. This chapter provides a broad overview of general fault detection methodologies and specific developments for the water sector. However, before continuing with the fault detection methodologies it is important to point out some general definitions used in the remainder of this thesis and their specific interpretation for fault detection in the water sector.

The term *fault* is generally defined as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process (Himmelblau, 1978). By transferring this fault definition to water treatment processes, a fault is defined in this context as abnormal deviation of a water quality parameter from its normal process condition. Since various water quality parameters, such as pH, turbidity, Iron, etc. undergo significant value changes during the treatment processes influenced by diverse factors, e.g. retention times after dosing, the occurrence of abnormal process conditions within the water treatment processes is very likely. Water quality parameters desired to be observed for fault detection at different WTWs processes/stages are provided in Table 2-1.

Table 2-1 *Water quality parameters desired to be observed.*

Parameter	Raw Water/Inlet	Pre-Flocculation	Post-Flocculation	Post-Clarification	Outlet
pH	X	X	X	X	X
Temperature	X	X	X	X	X
Turbidity	X		X	X	X
DOC	X		X	X	
Iron	X	X	X	X	X
Conductivity	X				X
Chlorine				X	X

Faults are caused by certain events related to errors, malfunctions, failures and disturbances or perturbations in the system. Typical faults arising in water treatment processes are sensor data errors, sensor or actuator malfunctions, equipment faults, e.g. pump failures, and contamination events. *Fault detection* means that a problem occurred in the system has been identified, even the event or cause of the problem is not known. Faults can be detected either by model-based methods, where a priori quantitative or qualitative knowledge about the process is needed or by process history based methods, also referred to as data driven methods which require a large amount on historical process data (Venkatasubramanian et al., 2003a). Based on commonly accepted definitions agreed by the SAFEPROCESS Technical Committee (Isermann and Balle, 1997) supervisory functions can be determined as follows:

Fault isolation is pinpointing the type, location and time of a fault, whereas *Fault identification* is determining the size and time-dependent behaviour of the fault.

Fault diagnosis comprises both fault isolation and fault identification and leads generally to the root cause(s) of the problem.

Monitoring is been defined as a continuous real-time task of determining the conditions of a physical system, by recording information, recognizing and indicating anomalies in the behaviour.

In the specific literature the term *Fault Detection and Isolation* (FDI) is often used, whilst by *isolation* usually fault *diagnosis* is meant. For better distinction between fault isolation and fault diagnosis, the terminology *Fault Detection and Diagnosis* (FDD) is preferred and used in the subsequent sections of this review.

The overall processes including fault detection, diagnosis and correction of faults including the return to normal process operations is determined as *Abnormal*

Condition Management (ACM) or *Abnormal Event Management* (AEM). In the remainder of this thesis the term event management is used in this context.

It should be noted that in the course of the work presented in this thesis, the term *Event Recognition System* (ERS) is used in addition to *Fault Detection System*. ERS in this context describes a framework to detect potential events/faults at WTW sensors and/or processes and to provide a basis for taking decisions and actions to bring back the faulty water treatment process to normal operation. A general schematic of an ERS for application at WTWs is shown in Figure 2-1.

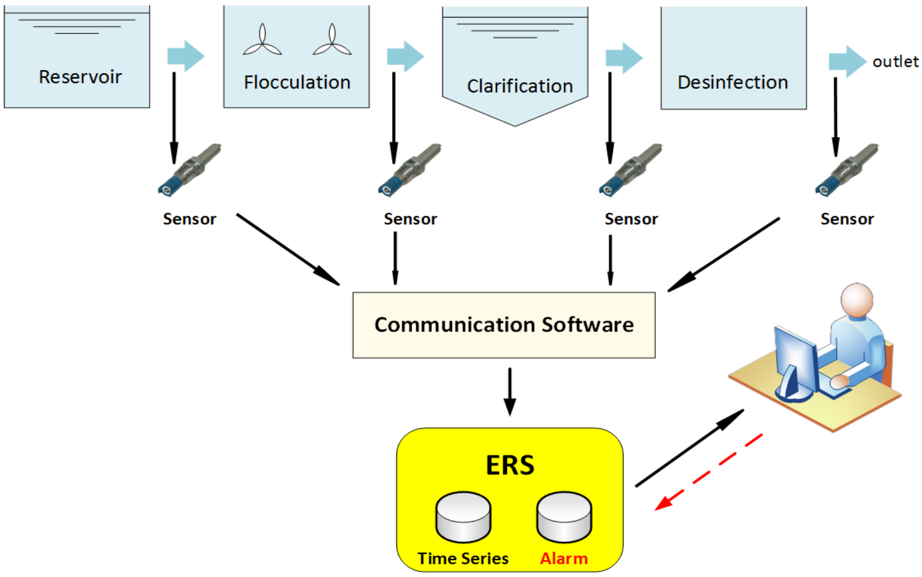

Figure 2-1 *General ERS schematic for WTW application.*

The development of a new ERS for water treatment works requires that (i) the characteristics to be met by the new system are determined, (ii) a prioritisation of these characteristics is carried out, (iii) investigation of fault detection methods, reflecting the recent state of art in research and, (iv) the adaptation of the event recognition methodology to the requested WTW application.

Essential characteristics desired for a successful application of a new ERS at WTWs have been identified as follows:

1) Near real time detection – most crucial for fault detection in timely manner.
2) Reliability - required functions need to be performed in a reliable and safe manner.

3) Adaptability of the system - the system needs to be adaptable to different water treatment processes and locations.
4) Robustness - the system needs to perform effectively even in the presence of diverse process noise.
5) Low computing time and storage capacity - to meet the requirement of the first point, the applied algorithms need to be performed quickly and the processed data volume should be manageable.
6) Cost effectiveness - the system needs to be cheap in implementation and operation.

Implementing all mentioned characteristics in one single system with acceptable functionality is a difficult task, since some of them are usually counteractive. For example, a system that is designed for a quick fault detection, i.e. near real-time, will be highly sensitive and tends to produce a high number of false alarms. Although all above characteristics will not usually be met by any single detection method, they are useful to benchmark different methods in terms of reliability, generality and efficiency in computing etc. (Venkatasubramanian et al., 2003a). The following literature review aims to provide a wide overview about the state of art fault detection methods and to explore their capabilities respective limitations as well as their potential to improve shortcomings of fault detection applications used to date at WTWs.

The literature review report is organised as follows: After this introduction an overview of general fault detection methods is given in Section 2.2. This section provides a review about the various methodologies of model-based, process history based and hybrid fault detection techniques. Section 2.3 presents an overview of specific methods developed for fault detection in the water sector, structured according to approaches used at WTWs and techniques applied to water infrastructure followed by the review of latest developer independent and dependent software applications for fault detection at WTWs. Section 2.5 contains a summary of previous publications and discusses capabilities and limitations of the main findings as well as key challenges for the achievement of further improvements. The final section 2.6 provides a concluding summary of this chapter and presents the key gaps in knowledge identified in detection of faulty events at water treatment works.

2.2 General Fault Detection and Diagnosis Approaches

In literature, various fault detection techniques were presented and several reviews have been published over the last 20 years (Isermann and Balle, 1997; Frank and Ding, 1997; Russell and Braatz, 2000b; Venkatasubramanian et al., 2003a,b,c; Quin, 2003; Isermann, 2005; Hwang et al., 2010; Alcala and Qin, 2011; Maiti and Banerjee, 2012; Yin et al., 2012). The vast majority of developed methods have found their first application in mechanical and electrical processes at this time. Fault detection for chemical processes was only slightly developed, but the number of applications was growing (Isermann and Balle, 1997). Nowadays new technologies including hybrid systems, i.e. combinations of different methods to control even more complex systems and processes are favoured, but still need to be further developed. Therefore, general fault detection methods and strategies are briefly described in the course of the literature review focussing on main principles and fundamentals of the wide variety of fault detection techniques and strategies with no claim to completeness.

The various methodologies reviewed are classified into three groups: model-based, process history based and hybrid approaches. For the model-based techniques a distinction is made between qualitative and quantitative model-based methods, in literature frequently also referred to as analytical and process knowledge-based methods (Frank, 1996). This distinction made between the different methodologies is based on the process knowledge and information required for each single method. In quantitative model-based approaches, where real process data are compared to calculated data from the model, an in-depth knowledge about the (usually dynamic) relationships in the process is required to describe the process behaviour in mathematical terms. Similar to the quantitative model-based methods, the qualitative model-based techniques demand knowledge about the physics or chemistry of the system as well as their relationships within the process, but in contrast to qualitative methods their process behaviour is expressed in qualitative terms by qualitative functions such as causalities or rules.

Contrary to the model based approaches, where a priori knowledge is needed, process history based methods require only the availability of a large amount of historical process data (Venkatasubramanian et al., 2003c). Some methodologies found in literature were developed as combination of several fault detection techniques. Such fault detection strategies, referred to as hybrid models are able to complement one another resulting in better performing systems. Integrating the complementary features of multiple methods into a single approach is one way to develop hybrid methods that could overcome the limitations of individual solution strategies (Sin et al., 2012). The hybrid approach, which is still at a nascent stage, is an amalgamation of data-based and/or model based approaches (Maiti and Banerjee, 2012).

In general, above classification of fault detection methodologies provides a feasible and useful scheme where all reviewed fault detection methodologies, strategies and techniques can be categorised into. The classification of the different fault detection methodologies is illustrated by the scheme shown in Figure 2-2.

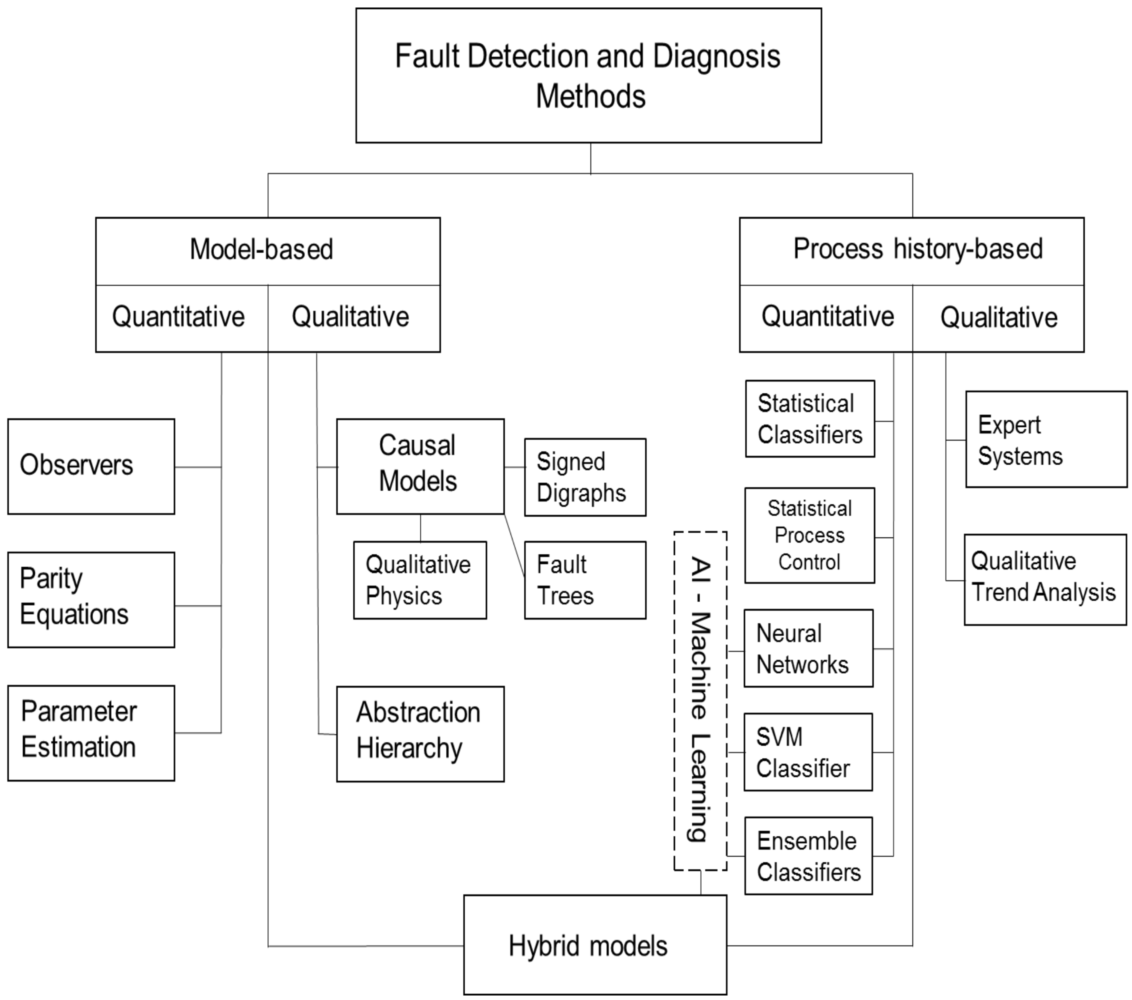

Figure 2-2 *Categorisation of fault detection and diagnosis methods.*

2.2.1 Quantitative Model-based Methods

Quantitative model-based techniques make use of static and dynamic relationships across system variables and parameters to describe system's behaviour in quantitative mathematical functions. The concept of model based methods is always the same, whilst in a first step inconsistences, i.e. residuals between measured values at the actual operating state and values obtained from the model at the expected operating state are generated, a decision rule for diagnosis is selected in the second step. This is achieved by comparing the signal that has been artificially generated by the analytical model to the real measured signal of the system, where a large difference between measured value and calculated prediction infers the existence of a fault. Several methods for residual generation have been presented in literature, such as observers, parity equations and parameter estimation.

Diagnostic observers utilise ‘observable’ residuals generated by mathematical functions that enable the direct detection of faults in a system. Filtering is the most used observer technique, especially Kalman filters have been widely applied as observers for state estimation (Frank and Wuennenberg, 1989). Due to the their restrictive application to linear models, various non-linear observers approaches have been presented, including unknown input observers (Frank and Ding, 1997), sliding mode observers (Edwards, Spurgeon and Patton, 2000) and extended Kalman filters (An and Sepehri, 2005).

Parity equations are rearranged and usually transformed variants of the input-output or state-space models of a system (Gertler, 1991; Gertler and Singer, 1990). In that case, where the generated residuals equal to zero or are correspondingly small, i.e. the output of the process matches with the model equations (parity), an error-free system is estimated, otherwise the occurrence of a fault is presumed. To diagnose multiple faults and to improve fault isolation in processes with more than one input and one output, it is possible to design a set of structured residuals, so that faults do not affect all residuals (Isermann, 2005).

Parameter estimation was introduced by Filbert and Metzger (1982) as a fault detection strategy for technical systems, which is most frequently used for fault detection in mechanical and electrical processes (Isermann and Balle, 1997). The basic concept of this method is to construct a reference model that is designed for fault-free operation using previously measured data. Fault detection is enabled by calculating the residuals from deviations of on-line estimated parameters from their corresponding reference values.

2.2.2 Qualitative Model-based Methods

While in quantitative model-based methods the behaviour of a system is expressed by mathematical relations between input and output variables, in qualitative model-based techniques these relations are described by qualitative functions across different variables in a process. A fundamental process knowledge is required to describe these qualitative relationships or causalities, by means of causal (cause-effect) modelling or rules, i.e. rule-based or fault-symptom modelling. As shown in Figure 2-2, the qualitative fault detection models are divided into two types: causal models and abstraction hierarchies. Causal models are subdivided into signed digraphs, fault trees and qualitative physics.

The idea applying signed digraphs (SDG) for fault detection was first introduced by (Iri et al., 1979). This method uses directed graphs (Digraphs), i.e. graphs with directed arcs by connecting nodes with positive or negative signs to express cause-effect relationships in a system. In this method, the connecting nodes represent variables or events, and the edges describe the relationship between the nodes, while the direction of change is indicated by the signs. SDG used as a rule-based method for fault diagnosis was introduced by Kramer and Palowitch (1987). Notable work was presented by Raghuraj, Bhushan and Rengaswamy (1999) with the introduction of a new algorithm for sensor network design in chemical plants. SDG models are suitable to be used in hybrid models, especially in conjunction with expert systems (Severson, Chaiwatanodom and Braatz, 2015).

Fault trees originally developed in 1962 by Bell Laboratories are logic diagrams (logic trees) representing causal relations by nodes as states and edges as relations between fault, events and symptoms. Causalities are expressed by condition–conclusion rules and operations (and-or) at each node whilst their conclusion extends to other events connected by logic nodes. After a framework of causalities is established a fault tree analysis (FTA) follows by determining the causal pathways in hierarchy from the faults over events to the symptoms

(Fussell, 1974). The inverse way of diagnosis from symptoms to faults, presented by Rasmussen in 1975, is known as event-tree analysis (ETA). Ulerich and Powers (1988) introduced a fault detection tree using real-time data to verify events in the fault tree.

With Qualitative Physics, also known as common reasoning about physical systems, first a model is developed starting from the description of physical mechanisms in a system. Without precise knowledge of the parameters and functional relationships, an algorithm is used in a second step to determine the overall behaviour of the system (Venkatasubramanian et al., 2003b). Resulting qualitative behaviour, as source of knowledge, can be used for fault detection. Qualitative physics methods for fault detection are most commonly used in Qualitative Simulation (QSIM) and in Qualitative Process Theory (QPT) to build a prototype first-principles troubleshooting system (Grantham and Ungar, 1990, 1991).

Another technique to derive knowledge for qualitative models is the concept of abstraction hierarchies based on decomposition. Depending on how the system is decomposed into subsystems, structural and functional hierarchies are distinguished. (Rasmussen, 1986). Further details on abstraction hierarchies and their applications for fault detection can be found in Venkatasubramanian et al. (2003b).

All model-based methodologies, whether quantitative or qualitative require a precise a priori knowledge of the modelled system. Therefore, designing a model for complex systems, e.g. non-linear chemical processes is a difficult task and often runs the risk to be incomplete or flawed. Furthermore, such models usually require an immense computational effort.

2.2.3 Process History-based Methods

While fundamental knowledge of the physical system is necessary for model-based fault detection methods, the application of history-based fault detection methodologies only requires the availability of a large amount of previous process data. Therefore, process history-based methods are also referred to as data-driven methods. In these methods, fault detection is carried out by identifying deviations in the behaviour of the observed processes in comparison to their behaviour in normal or abnormal operating conditions explored from historical process data. According to their way of how the used historical data is provided and transformed (feature extraction), process history based detection methods can be classified into two groups: methods based on statistical techniques and those based on Artificial Intelligence (AI) techniques (see Figure 2-3). Statistical approaches are widespread particularly in the field of statistical process control (SPC), which covers univariate and multivariate methods (Corominas et al., 2010). For large systems, where the generation of detailed reliable analytical models is difficult or not possible, the application of data driven detection models, which are mostly quantitative models based on rigorous statistical development of the process data is preferred (Verron, Tiplica and Kobi, 2008).

Statistical methods for fault detection cover statistical classifiers and SPC techniques including dimension reduction methods and control charts. The following review presents most important developments and progress made in the application of those methods.

2.2.3.1 Statistical Classifiers

Statistical classifiers make use of pattern recognition methods to achieve fault classification. Initial work on the statistical classification of a two class problem was carried out by Fisher's linear discriminant function (Fisher, 1938) assuming multivariate normal distribution for values within each of the two classes. Bayes classifiers, based on the Bayes' theorem, assuming Gaussian distributions with equal covariance matrices are considered as optimum classifiers (Fukunaga, 1972) if the classes are Gaussian distributed. Distance based classifiers, which calculate the distance of patterns from the means of various classes are frequently used as baseline classifiers for pattern recognition problems. The K-Nearest Neighbour (k-NN) classifier (Cover and Hart, 1967) as probably the most popular distance-based classifier generally uses the Euclidean distance function. Similar to Bayes classifier, Euclidean distance based classifiers require Gaussian distributions. Quadratic or piecewise classifiers are other types of distribution-free classifiers using a quadric surface for the classification of two or more classes.

2.2.3.2 Principal Component Analysis

Principal Component Analysis (PCA), including its extension Independent Component Analysis (ICA) and Partial Least Squares (PLS) are the most commonly used multivariate data analysis techniques for monitoring and modelling dynamic processes, based on dimension reduction of the process variables. PCA was invented in 1901 by Pearson (1901) and discovered as a basic SPC methodology to manage a large number of process and quality variables for continuous process (Kresta, MacGregor and Marlin, 1991) and at the same time the oldest multivariate method (Abdi and Williams, 2010). PCA for sensor fault detection was introduced by Dunia and Qin (1998) using the method for analysing the fault subspace via reconstruction. ICA introduced by (Hérault and Ans, 1994) as an extension of PCA is applicable to non-Gaussian multivariate processes and, therefore, ICA plays an important role in real-time monitoring and diagnosis for practical industrial processes (Zhiwei, Cecati and Ding, 2015). PLS developed by Wold (1966) made its breakthrough in the mid-1980s in SPC applications for complex processes. The ability of handling two sets of data, namely predictor and response variables represent a major advantage compared to the PCA method. PLS is frequently used to date in the chemometrics field, where often a large number of process variables for both input and output have to be managed (Montgomery, 2009).

2.2.3.3 Shewhart and Exponentially Weighted Moving Average Control Chats

Control charts have been widely used to reduce deviations in manufacturing processes. The introduction of Shewhart control charts in 1932 (Shewhart, 1931) as first concept for quality control represents the origin in the field of SPC. With its variations, such as x-bar charts, s-charts and R-charts, SPC has evolved and made further progress with the introduction of the exponentially weighted moving average (EWMA) charts, invented by Roberts (1959). The main disadvantage of the above control charts is their limitation to control only one single parameter in a process. As a multivariate extension of the X-bar chart, the first multivariate control chart was introduced with the Hotelling control chart, also known as Hotelling's T2 control chart (Hotelling, 1947). Only much later followed the multivariate exponentially weighted moving average charts (MWMA) (Lowry et al., 1992) as a further development of the univariate EWMA variant. Due to a growing interest in measuring product quality and process reliability and especially because of their easy implementation, SPC charts have been extensively used until now.

2.2.3.4 Cumulative Sum (CUSUM) Control Charts

Cumulative sum (CUSUM) charts monitor the drifts of cumulative sums of the observed quality characteristic against time to locate statistically significant abnormalities (‘out of control’ points or sequences). CUSUM control charts were first introduced by Page (1954) who developed the method to filter general changes from random noise and proposed a limit criterion from which to intervene in the process. CUSUM charts have been early studied by many authors, e.g. Barnard (1959), Johnson and Leone (1962), Ewan (1963) and, have been successfully applied especially in the chemical and process industry. There are two different ways of presenting CUSUM control charts, namely tabular (algorithmic) CUMSUM and V masks (Barnard, 1959). Tabular CUSUM, also known as two-sided CUSUM control charts, are preferable for monitoring the process mean either for an individual observation or rational subgroup. Rather than examining the mean of each subgroup independently as it done by other control charts, the CUSUM chart incorporates the information of the entire observations. Therefore, the CUSUM chart is more efficient in detecting small and moderate shifts in the mean of a process than other control charts, e.g. Shewhart charts, X-bar or Hotelling’s T^2 charts which are better able to detect large process shifts. However, CUSUM control charts tend to respond only slowly to large process shifts. Furthermore, the identification and analysis of trend patterns is reasonable difficult.

Lucas and Crosier (1982) have proposed important modifications to the CUSUM chart that allow early detection of process shifts using the Fast Initial Response (FIR) function for CUSUM quality control schemes. In their study, Lucas and Crosier have shown that if the process mean is not at the desired level, an out-of-control signal will be given faster when the FIR feature is used. Multivariate cumulative sum methods have been introduced by Woodall and Ncube (1985), who proposed using a set of univariate CUSUMs on principle component to test shifts in the mean of a multivariate normal. This has the disadvantage that such diagrams would then be less effective in detecting shifts compared to the use of the original quality characteristics.

Crosier (1988) suggested to replace the scalar values of a univariate CUSUM control chart by vectors. This multivariate CUSUM scheme, known as MCUSUM, has demonstrated the ability to design CUSUM charts in such a way that it detects a specific shift in the mean vector and thus overcomes the disadvantage of multivariate Shewhart control charts. Pignatiello and Runger (1990) compared multiple univariate and multivariate CUSUM approaches using the average run length (ARL) to determine their performance and demonstrated that one-sided univariate CUSUM schemes using successive values do not show strong ARL characteristics. However, for multivariate control charts, a multivariate normal distribution is an important assumption that is used to describe the behaviour of the quality characteristics observed. This assumption cannot always be met in real situations, for example in systems such as water treatment processes with multiple signals, which are observed under harsh conditions and cause a large number of sensor errors and faulty sensor measurements.

Nidsunkid, Borkowski and Budsaba (2017) examined the effect of violations of the multivariate normality assumption on the performance of MCUSUM control charts using the average run length (ARL) and the standard deviation of run length (SDRL). In this research, the authors emphasise the sensitivity (lack of robustness) of the MCUSUM control charts to multivariate non-normal distributions and consider this a pitfall for process engineers. When applied under these conditions, the level of false alarms is likely to increase. Therefore, multivariate CUSUM charts lose their practical use in cases where the assumption is violated.

Based on the experience that CUSUM control chats used to monitor the process mean are sensitive to outliers, Yang, Pai and Wang (2010) presented a median-based CUSUM control chart to overcome the sensitivity of usual CUSUM charts in response to slight deviations from normality and make the CUSUM scheme more robust against outliers. When comparing the performance of the median CUSUM charts with the performance of the CUSUM, Shewhart and EWMA control charts, it was found that the CUSUM median diagram performs best in terms of outlier resistance, whereas the EWMA-X diagram is able to detect process shifts faster.

Recent approaches in the literature also move towards median CUSUM control schemes. Cheng and Wang (2018) investigated the effect of measurement errors to the performance of median based CUSUM control charts and EWMA median charts. This study showed that both chart types are significantly influenced by existing measurement errors and demonstrated that the CUSUM median chart performs better for small deviations and EWMA for larger deviations. Rahman, Yahaya and Atta (2018) examined the impact of using median and median of pairwise averages as location estimators in combination with the median absolute deviation about the median (MADn) as scale estimator (denoted as duo median based CUSUM charts) on the performance of the CUSUM control schemes. The results presented in this work have demonstrated that, in contrast to the standard CUSUM control card, the performance of the two duo-median-based CUSUM cards remains stable even in the presence of outliers. The last two studies have clearly shown the better performance of robust univariate CUSUM control schemes compared to the multivariate CUSUM variants when used in non-normal environments.

Compared to above methodologies, AI based fault detection is still an emerging technology, which made progressive progress in the mid-1980s with increasing computing resources for data acquisition, processing and storage. First attempts in this field were presented by Henley (1984) and Chester, Lamb and Dhurjati (1984) using expert systems for fault detection. Similar to above statistical fault detection methodologies, AI based strategies have found their primary applications in systems where an in-depth knowledge of the processes is not present or modelling of these processes/systems is too complex. But also, their potential to handle imperfect sensor readings make these technologies attractive for many fault detection applications. Most promising AI technologies, such as expert systems, neural networks (NN), fuzzy logic, Support Vector Machine (SVM) and Ensemble Classifiers (Bagged and Boosted Trees) are presented below, while knowing that other methods exist.

2.2.3.5 Expert systems

An expert system is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise (Feigenbaum, 1982). Expert systems are widely used in fault detection applications since they are easy to develop, accurate, and have the ability to reason under uncertainties and provide explanations for decision systems (Venkatasubramanian et al., 2003c). But also, Expert Systems have a number of shortcomings, for example the lack of generality, poor handling of novel situations, inability to represent time-varying phenomena and to learn from their errors as well as their sumptuous development and maintenance (Agneli, 2010). However, by using expert knowledge, rules and training in conjunction with new techniques, i.e. soft computing, models for a broad class of nonlinear systems can be represented with arbitrary accuracy (Khoukhi et al., 2012).

2.2.3.6 Artificial Neural Networks

Artificial Neural Networks (ANN) are systems of interconnected nodes (i.e. artificial neurons) which act as computational elements exchanging data between each other arranged in layers. ANNs have been used with a variety of architectures, i.e. as single layer feedforward, multi-layer feedforward or recurrent networks (Haykin, 2009) for fault classification, fault isolation and fault diagnosis. In particular, complying above mentioned multiple approaches, the ANN requires data on all possible process conditions at residual generation stage for an accurate fault classification in order to learn from system's behaviour. This can be a disadvantage if such required information is not available. On the other hand, ANNs are able to model nonlinear system behaviour and complex functional relationships across processes without deep knowledge of the physics of the system. Furthermore, ANNs are successfully used for classification and nonlinear function approximation problems (Sin et al., 2012). In the light of above, ANNs are very promising for fault detection and have been extensively applied to fault detection and diagnosis (Sobhani-Tehrani and Khorasani, 2009).

2.2.3.7 Fuzzy Logic

Fuzzy Logic techniques are widely used in AI for converting quantitative data into qualitative categories by segmenting them into fuzzy sets and applying rules of reasoning, that mimic comparatively human reasoning. Fuzzy logic approaches are widely used in the field of process control (Shaocheng, Bin and Yongfu, 2005), but are rarely utilised as fault detection strategy, usually in conjunction with other qualitative models (Sokoowski, 2004).

2.2.3.8 Support Vector Machine

Although the original algorithm for the construction of a linear classifier was proposed by Vapnik already in 1963, SVM is a relatively new machine learning technique of pattern recognition. SVM is capable of classifying all kinds of data sets with automated mathematical methods based on statistical learning theory, that is able to achieve a high degree of generalization and is suitable for dealing with problems with low samples and high input features (Zhiwei, Cecati and Ding, 2015). SVM technique for non-linear classification was introduced in 1992 by Boser et al. (Boser, Guyon and Vapnik, 1992). Although the development of SVM for fault detection techniques started at the end of the 1990s, the first application can be found only in 2006 (Widodo and Yang, 2007). Both, the work on SVM based fault detection presented by Yin et al. (2014) that compares the fault detection performance of the SVM classifier to the performance of the PLS algorithm and the approach of Sahri and Yusof (2014) on the replacement of missing data using k-NN analysis prior to SVM learning applied to the data of a power transformer showed encouraging results.

2.2.3.9 Decision Trees and Ensemble Classifiers

Decision trees or classification trees, whose algorithm was introduced by Morgan and Sonquist in 1963, is also rather new in use as a machine learning classification technique. Decision Trees predict a response, i.e. 'true' or 'false', following the decisions in the tree from the root down to a leaf node. Ensemble Classifiers construct more than one decision tree by applying either bootstrapping (bagging) or boosting algorithms. Whilst bootstrapping generates replicas of the data set and grows decision trees on the replicas to let them vote for the most popular class, boosting generates iteratively new weak prediction rules and combines these rules into a single strong prediction rule. Prominent and powerful bootstrapped classifiers are, Bagged Trees (Breiman, 1999) and Random Forest (RT) (Breiman, 2001). AdaBoost (Freund and Schapire, 1997) in conjunction with decision trees as weak learners is one of the most popular boosting techniques.

2.2.4 Hybrid Model-based Methods

Each of the aforementioned fault detection methodologies feature different properties, depending on process knowledge and scale they require, and have their specific advantages or disadvantages. Frequently, specifications for a certain fault detection technique are not satisfactory served with any particular strategy. In these cases, an amalgamation of various techniques is reasonable, since certain methods can complement one another and integrating or combining their advantageous features. The development of a 'hybrid detection system' is an appropriate strategy to compensate the limitations of one particular system (Venkatasubramanian et al., 2003c).

Hybrid methods for fault detection integrate more than one method within a single category (see Figure 2-2) or combine model-based and process history-based methods. Qualitative Trend Analysis (QTA) can be considered as most established hybrid fault detection methodology. QTA is performed in two stages: (i) interval-halving based trend extraction (Dash et al., 2004) and (ii) fuzzy trend matching (Dash, Rengaswamy and Venkatasubramanian, 2003). QTA techniques have been applied for fault detection, primarily in complex industrial and chemical processes. Several frameworks based on QTA technique combined with other methodologies were presented, e.g. QTA combined with SDG (Maurya, Rengaswamy and Venkatasubramanian, 2007). Reducing computational complexity using QTA in combination with a PCA model was proposed by Maurya, Rengaswamy and Venkatasubramanian (2005).

Specific applications using soft computing techniques such as fuzzy logic, neural networks and evolutionary algorithms instead of complex model-based approaches have attracted increasing interest. For example, a fuzzy model instead of physical model has been used for the generation of the parity equations (Puig and Quevedo, 2002). A hybrid data-driven and model-based fault detection method using SVM for fault detection was presented (Sheibat-Othman et al., 2014) for chemical reactors with highly nonlinear and dynamic processes and proofed its ability in isolating the faults. Several methods combining quantitative and qualitative system information for fault detection applying ANNs to minimize the probability of false-alarms and missed-alarms were discussed recently (Srivastava, Srivastava and Vashishtha, 2014). Although above applications display only a small section of the developments in the wide area of fault detection, the trend leads clearly towards hybrid models and AI approaches.

2.3 Fault Detection and Diagnosis Approaches in Water Sector

Detection of failure events at WTWs is gaining increasing importance in the eyes of water supply companies due to a raising concern about water contamination by pollutants, water security issues, legal requirements, and not least because of the needs to operate processes, in particular water treatment processes more effectively. As a result, great attention from both research and industry is focused on the development of new technologies for failure detection at WTWs.

In the water sector event recognition techniques have found their applications for the detection of faulty equipment, sensors and processes at treatment works (water and wastewater) and for the detection of faults in infrastructure, particularly in distribution systems and here mainly for the detection of leakages. Whilst the focus of the work presented in this thesis is on near real-time fault detection for WTWs, the review extends also to fault detection at water infrastructure, due to common features, e.g. detection of faulty sensors. Therefore, literature was reviewed in the field of fault detection at treatment works and infrastructure, including overlapping approaches, presented in the following sections.

2.3.1 Fault Detection Methods for Water and Wastewater Treatment Works

At this point, it has to be noted that, due to common principles of fault detection techniques for WTWs and Wastewater Treatment Works (WWTW), also fault detection methodologies applied to WWTWs are presented in this section.

Failures at water treatment works can be classified into two types of faults: sensor faults and process faults. Sensor faults refer to measurement errors or incorrect raw data resulting from faults like drift, bias, precision degradation or even from a complete breakdown of the device (Alferes et al., 2013). In contrast to sensor faults, process faults relate to either failures of WTW’s equipment, e.g. pump breakdowns or operational faults, e.g. dosing faults. Process faults usually cause problems in water quality that is described by water properties in terms of physical, chemical, thermal, and/or biological characteristics (Ritchie, Zimba and Everitt, 2003) and depending on the concentrations of chemical and biological parameters present in water. Process faults are usually amenable to detection by laboratory-based analytical methods and rectified at the treatment stage (Spellman, 2003). But laboratory based detection methods have two major drawbacks: they are costly and too slow to develop quick operational response, that is necessary for the protection of public health and/or environment if a present failure at WTW's process reaches a critical point. There is a clear need for improved online monitoring systems at all treatment stages of the WTWs (Storey, van der Gaag and Burns, 2011; Rosen and Bertrand, 2013). The use of appropriate sensor technologies for near real-time monitoring of critical water quality parameters is required to comply with this need.

Nowadays, a wide range of sensors able to monitor various water quality parameters in near real-time is commercially offered, but by far not all microbiological and chemical parameters can be measured directly by technically and economically feasible sensor products. Sensors deployed at WTWs aiming to observe the water quality usually measure "supporting" or "indicating" parameters including, among others, pH, turbidity, chlorine, electrical conductivity, nitrate, manganese and temperature. The responding behaviour of this surrogate parameters to different contamination events was investigated in the work of Hall and Szabo (2010) that demonstrated the ability of surrogate parameters to indicate rapid changes in water quality, caused, for example, by a deliberate or accidental contamination event, which in turn cannot be reliable detected in an adequate timely manner by conventional sampling and analytical methods (Banna et al., 2014).

However, performing online monitoring in an optimal manner requires a huge number of measurement values that have continuously to be refreshed, plausible and of good quality (Edthofer et al., 2010). Low data quality will limit the meaningfulness of predictions and erroneous data will lead in the worst case to faulty conclusions (Rieger et al., 2010). The water quality monitoring report published in 2013 by the Government of Newfoundland and Labrador (Government of Newfoundland & Labrador, Department of Environment and Conservation, Water Resources Management Division, 2011; Pugh, 2013) has shown, that sensor data quality is regularly affected by sensor faults like fouling and calibration drifts and the utilised sensor technology which makes it difficult to reliable assess the quality of measured sensor data. This was also shown by the work of Praus (2005) on the assessment of water quality using SVD-based PCA. This study demonstrated that hydro chemical data is mostly not normally distributed and contains not only important information for treatment technology, but also confusing noise, outliers and erroneous or nonsense values.

Since the detection of faulty processes at WTWs in near real-time is heavily dependent from reliable and accurate sensor data, it is beneficial to monitor sensor's data quality aiming to detect malfunctions or recalibration and maintenance intervals on the one hand and to raise the quality level of measured sensor data by validation and pre-processing procedures on the other. Several methods for validation of the data coming from sensors observing water quality parameters at WTW's can be found in literature. In their work Edthofer et al. (2010) presented a number of statistical methods using average, average deviations, moving average functions and Holt-Winters algorithm (Holt, 1957; Winters, 1960) to detect outliers, discontinuous measurements, noise and drifts in measured sensor data. The cleaned data is then used for event detection by pattern recognition and spectral analysis techniques (see Section 2.4.2: Developer-specific Solutions – GuardianBlue and s::can). When applied to a time series of 18 days' real data measured from sensors deployed at WTWs the sensor data validation methods demonstrated their abilities to detect outlier, discontinuous measurements, noise and drifts.

Whilst most of the techniques presented in the literature focus only on the detection of water quality events (McKenna et al., 2007; George, Chen and Shaw, 2009; Garcia-Alvarez et al., 2009; Chen and Huang, 2011) only a few, as above described method, apply sensor data validation procedures before carrying out event detection. In contrast to above, Talagala et al., (2019) developed a framework for the automated detection of outliers in water-quality data from in situ sensors caused by technical errors that make data unreliable and untrustworthy focussing solely on the identification of errors in the data due to issues unrelated to water events. Talagala et al. applied simple rules to filter out 'out-of-range', negative and missing values first, followed by statistical one sided derivative transformation to identify outlying instances from typical behaviours. The outlier were then classified by the application of several outlier scoring techniques as, among others, HDoutliers algorithm (Wilkinson, 2018), KNN-SUM algorithms (Madsen, 2018) and the calculated scores compared to a outlier threshold determined by the use of Extreme Value Theory (EVT). When demonstrated on real water-quality data the methodology improved the performance of outlier detection algorithms while maintaining low false detection rates. As an initial sensor data validation technology, this appears to perform well, demonstrated on different outlier detection algorithms.

In the field of fault detection at WTWs history-based methods SPC methods have been investigated most frequently starting in the early 2000s. Although the research focussed on multivariate methods, predominantly based on PCA very few data driven AI and univariate control chart methods were presented.

Schraa, Tole and Copp (2006) discussed the practical aspects of univariate Shewhart, Cumulative Sum (CUSUM), Exponentially Weighted Moving Average (EWMA) control charts for advanced fault detection at WWTWs. Due to autocorrelation, seasonality and non-constant variance of treatment plant measurements, EWMA control charts were evaluated as suitable control schemes. Shewhart and CUSUM control charts were assessed as difficult to apply. Although this assessment may well be correct for the application of standard univariate control charts, it seems to be highly appropriate to perform further experiments on these methods, in particular on CUSUM charts to investigate their possible adaptability to above features.

Corominas et al. (2010) developed an objective index used for performance evaluation of fault detection methods for water treatment processes. This index makes use of penalisation points, which are allotted for false acceptance, intermittent fault detection within event duration and false alarms each time the fault detection method failed, i.e. as more penalty points are allotted as worse the method performs. The usefulness of the performance index was demonstrated by experiments utilising Shewhart and four EWMA control chart variants for fault detection at wastewater treatment processes. All tested models showed moderate and partly poor detection performances explained by their lack of adapting to the time-varying and non-linear behaviour of these processes. Although the tested models show moderate detection performances, further work should be conducted to investigate modifications on control charting methods and/or explore their feasibilities for possible combinations with other detection methods to possibly adopt desired features in a hybrid system.

Unlike above univariate SPC techniques, Liu et al. (2015) developed a multivariate SPC classification method similar to CANARY (see Section 2.4.1.1) using a distance based classifier for detection of contamination events. Within a moving window Pearson correlation coefficients were calculated to determine the relationship between signals from multiple sensors followed by the calculation of corresponding correlation indicators and Euclidean distance of the formed correlation indicator vector. In case that the Euclidean distance violates a default detection threshold an alarm is triggered. The method was tested on a dataset from a contaminant injection experiment and showed higher true detection and lower false alarm rates compared to Linear Prediction Coefficient filter (LPCF) method (Hart et al., 2007) and Multivariate Euclidean Distance (MED) method (Klise and McKenna, 2006) used as baseline.

Rosen and Lennox (2001) presented a multivariate SPC method using PCA for fault detection at WWTW's processes. In their work Rosen and Lennox discussed also extensions to basic PCA and developed five models from static PCA to multiscale PCA (MSPCA). The limitation of static PCA to stationary data, because of its inability to capture changing conditions of water treatment processes was demonstrated. It was also shown in the study that PCA only using adaptive scaling parameters is not able to fully capture these changes neither. It was shown that adaptive PCA that makes use of adaptive scaling parameters and adaptive co-variance structure as well as multiscale PCA models can solve the problem of non-stationary process data. Compared to adaptive PCA the MSPCA approaches provide information on one scale at which disturbances in the system occur but have the drawback to be more complex than adaptive PCAs, this makes the methodology impractical for application in industry.

George, Chen and Shaw (2009) produced a very similar analytical approach to Rosen and Lennox (2001). George, Chen and Shaw applied PCA before using Hotelling's T^2 charts for fault detection analysis. However, in contrast to Rosen and Lennox, this method was applied for fault detection of multistage drinking water treatment processes. When applied to real time series data of 23 measured parameters collected from sensors deployed at the test site the method demonstrated feasibility to detect abnormal conditions within water treatment processes and was able to identify the parameters which contributed to disturbances in the process. Although a reasonable high number of parameters was used, only a short time period of 14 days was analysed. Due to this small-scale demonstration, it is hard to tell if the method would adapt well to the seasonality and non-constant variance of WTW's processes across long time periods.

Alferes et al. (2013) continued the work on PCA techniques and presented a PCA based method for real time monitoring of water systems and detection of sensor faults inspired by the mon*EAU* vision (Rieger and Vanrolleghem, 2008) aiming to achieve an advanced monitoring system with automatic data collection, evaluation, correction and alarm triggering. In their study Alferes et al. used PCA in combination with T^2 and Q-statistics for sensor data validation. Unlike prior PCA models, in this work sensor data pre-processing was utilised to remove outliers and perform auto-scaling (mean centring and variance scaling) prior to the application of the PCA model. Outlier detection is carried out using univariate autoregressive models. The pre-processed sensor data is projected to the PCA model and faults detected by comparing T^2 and Q values against their thresholds. When demonstrated on a dataset using eight water quality variables from a WWTW's inlet over a 3 days training period the method identified less than 1% of the samples to be abnormal and showed that the methodology enables the detection of different kinds of faults at individual sensors. Although this study show promise for the technology of sensor data validation, unfortunately, the work was tested on a relatively small test dataset with eight variables measured only at inlet stage of the WWTW.

Based on the previous work done on PCA for fault detection Aguado and Rosen (2007) produced a new approach for diagnosis of abnormal events at WWTWs. Aguado and Rosen applied adaptive PCA combined with two complementary control charts (T^2 and SPE) and introduced a static PCA model and fuzzy c-means clustering for fault diagnosis. The tested adaptive PCA approaches have demonstrated valid process monitoring most of the time. The study also showed that faster adaption results in higher detection speeds but causes a higher false alarm rate. Monitoring the process simultaneously by several adaptive models was recommended as possible solution. However, this would increase the complexity and makes the methodology impractical for application in industry.

Garcia-Alvarez et al. (2009) presented a very similar approach by combining PCA and Fisher discriminant analyses for fault detection and diagnosis at WWTWs. Garcia-Alvarez et al. used PCA and applied T^2 and Q charts for fault detection and utilised unlike the previous methodology Fisher discriminant analysis for fault diagnosis. Although the methodology applied to data of a simulated WWTW showed good results in detection and diagnosis of simulated events, the analysis of multiple charts necessary for fault diagnosis procedure makes the application impractical for daily use.

Apart from above methods based on statistical techniques several approaches using AI were presented in literature. Ruiz, Colomer and Melendez (2007) were among the first to combine multivariate statistical process control (MSPC) with an AI method. In their work Ruiz, Colomer and Melendez (2007) present Multiway PCA (MPCA) and case-based reasoning (CBR) for the assessment of the actual state of the WWTW. In contrast to the previous methodology, this technique applies PCA to batch data (MPCA) and makes use of artificial intelligence, i.e. CBR to diagnose WWTW's condition. Applied to a pilot plant the methodology demonstrated satisfactory diagnostic results and its feasibility of detecting abnormal process behaviour and classifying the state of WWTW's processes in near real-time. A notable drawback could be the expert knowledge about historical faults necessary for the case reasoning application.

A similar approach to the work of Ruiz, Colomer and Melendez (2007) for the assessment of WWTW's state was presented by Padhee, Gupta and Kaur (2012). In their study Padhee, Gupta and Kaur combined PCA for fault detection in WWTW's processes with a neural network based on backpropagation algorithm as classification technique, ascertaining healthy or faulty conditions of a multistage WWTW. Sensor data pre-processing by means of bicubic interpolation technique for reconstruction of missing data was conducted prior to the application of the PCA model. Applied to a reasonable large dataset, this methodology demonstrated to be efficient in detecting different faults in WWTW's processes, but further studies are recommended by using ICA or hybrid PCA/ICA and comparing different algorithms in backpropagation. Due to a lack of information about the number of faults applied in this demonstration, it is hard to tell if the method would generalise for the application on WTW's processes.

Lennox et al. (2001) presented the first approach found in literature not utilising SPC as basis fault detection methodology, like all above techniques. Lennox et al. applied ANN to a rapid gravity filtration process in a WTW to optimise WTW's filtration process in terms of minimising coagulant and ozone dosing as well as maximising the output water quality and filter run length. An advisory system capable to determine the optimal coagulant and ozone dosing levels was constructed. When demonstrated on two years' time period of data split into training, testing and validation periods the method showed to be capable of improving process monitoring and control. Although this promising methodology was not specifically used for fault detection it could possibly adapted for the detection of failure events at WTWs.

Punal, Roca and Lema (2002) developed an expert system for monitoring and diagnosis of anaerobic WWTWs. The methodology applied pre-processed data of a pilot scale reactor for diagnosis of the rector's state and classifies WWTW's state into labels corresponding to the commonest situations in operation. In case of abnormal process condition the model provides recommendations for the operator. Demonstrated on a small-sized dataset, the methodology showed its capability for determining the current conditions of pilot plant's processes and for classifying WWTW's state into the defined labels. A major drawback of this methodology is the deep knowledge about historical process conditions required for the application of expert systems.

Immune Feedforward Neural Network (IFNN) method using an ANN for fault detection in water quality monitoring was developed by Chen and Huang (2011). The IFNN was constructed with an eight dimensional vector as input (four measured water quality parameters at two time steps) and a 5 dimensional vector (five types of failure) as output of the network. The neural network's structure and parameters were optimised by training the IFNN with Levenberg-Marquardt method followed by the application of an immune algorithm to improve fault detection accuracy of the system. This technique was demonstrated on a small-sized dataset containing eight input and six output parameters and has shown feasibility for fault detection by detecting faults faster and more accurately than a feedforward NN. Due to the small-sized dataset used for the demonstration, it is difficult to say if the method would be capable for the application on WTWs processes.

Lamrini, Lakhal and Le Lann (2014) developed an ANN combined with fuzzy logic as classification technique. Fault detection, data validation-reconstruction and predictive control methods were utilised by the decision support tool to predict the optimal coagulant dosage to be used for coagulation process in WTWs. The decision support tool is built from a framework of four models, (i) classification model to identify the functional states of the treatment plant utilising learning algorithm for multivariate data analysis (fuzzy methodology) to detect this states; (ii) data validation model (NN model) developed by using self-organising maps of Kohonen (Kohonen, 1995) to validate invalid data; (iii) data reconstruction model (NN model) built by means of self-organising maps algorithm for reconstruction of missing data; (iv) coagulant dose prediction (NN model) using multilayer perceptron to predict the optimal coagulant dose. When demonstrated on a relatively small dataset (2311 samples) containing five water quality parameters collected at WTW's inlet the method proofed its effectiveness and reliability with a prediction rate close to 93.6%. Apart from the small dataset used for testing, the only notable drawback could be the a priori knowledge necessary of raw water descriptors for flexible and reliable prediction.

Page, Waldmann and Gahr (2017) introduced an adaptive technique for monitoring changes in water quality based on multivariate pattern analysis using multivariate analysis and artificial neural networks (ANN). The method applies self-organizing maps to calculate system states based on six online parameters and analyses identified patterns in complex system data, particularly to (i) reduce the dimension of the data set, (ii) calculate actual system states based on online time-series and (iii) lay emphasis on the similarities and dissimilarities between system states. Although the method captured actual changes in system's state and makes them visible, it was shown only on small-sized data of a small-scale treatment plant (only one treatment stage) and seemed not have been tested on a validation dataset.

A comprehensive event detection and event management solution was presented by (Bernard et al., 2017). The main developments presented in this approach are CBRN sensors, Event Detection System (EDS) and Event Management System (EMS). The EDS is based on several algorithms: (i) single sensor algorithms, which learn the statistical limits, trends and rates of change of each variable separately (threshold based), (ii) multi sensor algorithms using density functions of past combinations of sensor values, (iii) rule-violation algorithms for testing sensor values with a set of predefined engineering rules, and (iv) hazard similarity detection to identify the critical state when water quality reaches a hazardous combination. The methodology was tested and evaluated in “real life” at three water utilities and has demonstrated to detect successfully single contamination events. Since alarms are triggered by each above algorithm separately this methodology possibly generates a high false positive rate.

Piciaccia et al. (2018) developed a data-driven approach for learning the optimal control parameters using Support Vector Machine (SVM) algorithm to predict WWTWs’ process behaviour in terms of future plant states, estimation of optimal chemicals dosage and identification of the most influential parameters. A range of parameters, mainly water quality parameters served the input of the classifier in order to predict the final turbidity value at WWTWs’ end process that represents the water quality as well as WWTWs’ status. Additionally, the framework was adopted to address the prediction of chemicals dosage followed by ablative analysis for feature elimination (backward elimination) aiming to remove redundant parameters. When demonstrated on a reasonable large-sized dataset the model predicted 85% of times a correct plant status as well as at any given time the amount of chemicals to obtain a satisfactory status. This model showed very promising results and appears to be the most applicable and best demonstrated classification technique.

A probabilistic outlier detector implemented by a Deep Neural Network (DNN) for anomaly detection at water treatment systems was proposed by Inoue et al. (2018). In their work Inoue et al. applied a DNN consisting of a Long Short-Term Memory (LSTM) layer followed by feedforward layers of multiple inputs and outputs to time series data of a testbed treatment plant for the prediction of engineered contamination events. The performance of the DNN was evaluated by the calculation of the F-measure and compared to the F-measure (F_1 score) of a one-class SVM classifier used as baseline method applied to the same dataset. Both methods, DNN and SVM demonstrated good detection capabilities by generating similar F_1 scores of 0.8 and 0.79 respectively. Whilst the SVM detects slightly more anomalies, the DNN showed fewer false positives than the one-class SVM.

Although by far not all fault detection approaches were presented at this point, the methods shown above should provide a comprehensive overall view of the current state of art in the field of fault detection for water treatment processes. In addition, several methods for fault detection at water infrastructure were presented in the following section with regard to their possible adaption for fault detection at WTWs processes.

2.3.2 Fault Detection and Diagnosis Approaches for Water Infrastructure

FDD at water infrastructure refers in this context mostly to approaches for drinking water distribution systems (WDS). Faults in WDS arise either from abnormal hydraulic events, caused by leakage or pipe burst or from water quality events caused by deliberate or accidental contamination, or related to water discolouration mainly caused by increased levels of manganese and iron (Tumula and Danso-Amoako, 2014). In the presence of these events a fast detection and localisation of faults within the WDS is most important to prevent significant water loss due to hydraulic failures or protect human health in case of contamination. In order to fulfil this task in an adequate reliable and timely manner the use of real-time sensors, their optimal local placement as well as the utilisation of reliable and high-performing event recognition systems is crucial.

Numerous approaches have been presented in the past referred to sensor placement for fault detection in WDS e.g. to solve the problem of optimal sensor location for effective contamination detection (Weickgenannt et al., 2010) or to maximise the Isolability with a reasonable number of sensors for leakage detection (Sarrate, Nejjari and Rosich, 2012), but also to the detection of hydraulic sensor faults (Bouzid and Ramdani, 2013), pipe bursts and leakage in near real-time (Mounce, Boxall and Machell, 2010; Ye and Fenner, 2011; Romano, Kapelan and Savic, 2014). The mentioned approaches only exemplarily reflect a small part of the various methods developed in this field.

However, the focus here is set on fault detection techniques of water quality events in WDS. To detect water quality events in WDS following three principal tasks have to be carried out: (i) selection of water quality parameters as indicators for contamination, (ii) determination of the optimal number and locations of sensors within the WDS, and, (iii) performing temporal data analysis for possible fault identification (Perelman et al., 2012), although in the following review of techniques the determination of sensor locations was not taken into consideration.

Arad et al. (2013) developed an ANN model for analysing multivariate water quality time series combined with Bayesian sequential analysis for estimating the probability of a potential contamination event. The model was tested on real data from a water utility and demonstrated to be a powerful tool, although still a high number of false alarms was generated for certain scenarios. Further improvements on the model including the application of dynamic thresholds were proposed by the authors. This could either be achieved by using unequal water quality weights or by the integration of multiple sensor information into event detection methodology, which latter only can be achieved by multivariate analysis.

With regard to the issue raised in Arad et al. (2013) about integrating multiple sensor information, Oliker and Ostfeld (2014) proposed a contamination event detection approach with an autonomic decision support system using a weighted SVM classifier that provides, in contrast to the model of Arad et al. (2013), a multivariate analysis for outlier detection followed by sequence analysis for the classification of events. This method was demonstrated on a small-sized dataset containing six water quality parameters analysed over a four weeks' time period and has shown increased accuracy and detection ratio compared to the aforementioned model.

Oliker and Ostfeld, 2015 expanded their previous method by combining multiple sensor with network hydraulic data. The method applies, in contrast to the previous methodology, a minimum volume ellipsoid classifier for outlier detection followed by sequence analysis and the extended local and spatial decision rules integrating hydraulic data for event classification. The model proposed by Oliker and Ostfeld (2015) contains an integrated single spatial warning system by parallel event classification incorporating data analysis of all sensors together with the hydraulic model of the network and therefore can be classified as hybrid model. Although the model demonstrated good true detection ability and clear advantages over single sensor approaches, the false positive rate is still very high.

Housh and Ostfeld (2015) expanded the model of Arad et al. through the optimal integration of detection from all different water quality parameters into the event detection framework (ILD method). When applied on the same dataset as used by Arad et al enabling a valid comparison the method demonstrated a significant improvement in terms of detecting events with higher event probabilities in case of true events, but also training for the method is more complex, since the best threshold control variables have to be found for all water parameters simultaneously.

Meyers, Kapelan and Keedwell (2017) developed a classification based data driven methodology for short-term forecasting of turbidity levels for early detection of discoloration events. The threshold based classification method applies Random Forest (RF) classifiers to forecast turbidity and the Extra Trees variant to reduce the chance of overfitting on used calibration data. When tested and verified on a reasonable large-sized dataset containing turbidity measurements of a real UK trunk main network, the methodology accurately forecasts up to 5 hours turbidity events and hence corresponding discoloration events. This model appears to perform well, in particular regarding the low false detection rate of around 25% within the 5 hours forecast.

Zheng, Yekun and Qiao (2018) developed a new methodology for the detection of abnormal events in WDSs using a Deep Belief Neural network (DBN) integrated with Extended Kalman Filter (DBN-EKF). DBN-EKF makes use of data pre-processing by (i) missing data reconstruction, (ii) elimination of duplicated values, (iii) interpolation of irregular data, and (iv) removal of sensor failure stamps (extreme values and flat line faults) followed by Seasonal-Trend Decomposition Procedure based on Loess (STL) to remove the influence of the trend component and seasonality. For anomaly event detection the method applies a Deep Belief Network (DBN), which involves a two-step training including pre-training and fine-tuning by the use of Extended Kalman filter (EKF). When demonstrated on a reasonable large dataset of two years raw data and compared to several SPC techniques, the method showed ability to detect outliers and advantage over compared X-bar, EWMA, CUSUM, Seasonal Hybrid-Extreme Student Deviate SPC control chart methods. Since the method was only proofed on a ‘few obvious events’ these results are difficult to interpret.

Main findings of above reviewed event detection methodologies, which could be used favourably for the development of a reliable high performing ERS at WTWs are further discussed in section 2.5. Only a small number of fault detection methods have found their implementation in specialised software applications already used as ERS at WTWs by water utilities. In the following sections software applications structured by developer independent and developer dependent software solutions as well as promising ongoing ERS software developments are presented, starting with the developer independent products.

2.4 Fault Detection and Diagnosis Software in Water Sector

As presented in the preceding sections, a wide range of methodologies for the detection of failure events at water processes have been developed and are applicable to the specific requirements in water sector. Although several of these methods have shown their capability for reliable and high performing detection approaches and for the great need for real-time ERS applications, only a few applications have found their implementation in specialised software systems for practical use in industry and various of them have not proven their potential to reliably detect measurement or equipment failures (Rieger and Vanrolleghem, 2008) or rather to distinguish between sensor faults and fault processes. In the following sections software systems are presented that are already used by industry or bearing the potential to be implemented by water utilities in the near future. The diverse software approaches presented in the following sections are subdivided into developer-independent solutions, which are compatible to different sensor types and manufacturers or developer-specific solutions, which only can be used with certain sensor brands and finally into ERS software under development.

2.4.1 Developer-independent Solutions – CANARY, optiEDS and BlueBox

2.4.1.1 CANARY

CANARY (Hart et al., 2007; Hart and McKenna, 2009) developed by Ensemble Vulnerability Assessment Research Team (TEVA), which is composed of researchers from EPA, Sandia National Laboratories, the University of Cincinnati, and Argonne National Laboratory (EPA, 2010) is an open source computer software, written in MATLAB® (MathWorks 2008), that is able to read water quality sensor data in real time and can be used by anyone free of charge. The software requires input data from water quality sensors coming usually from water utility's SCADA[1] system. CANARY is able to include and handle additional hydraulic data such as tank levels, flow rates and valve settings. CANARY includes several event detection algorithms and enables to identify abnormalities in the water quality offline (data supplied offline by the user) or online where the data is provided online by the SCADA system.

The algorithms used in CANARY are (i) time series increments, which is an implicit estimation model to predict the value of a water quality parameter at the next time step based on the value measured at the previous time step and calculates then the difference between the estimated and the actual measured value, (ii) a linear filter, denoted as linear prediction-correction filter (LPCF) uses a linear predictor to estimate the current value of a time series based on a weighted sum of past values, and (iii) the multivariate nearest neighbour algorithm to define the background state of the water quality. To aggregate the event detection approaches over multiple time steps and to calculate the probability for an event a further algorithm is used, denoted as binomial event discriminator (BED). Although the algorithms are considered to be very effective at detecting water quality changes (Szabo and Hall, 2014) the CANARY platform allows to implement and use other fault detection algorithms.

Training data sets can be used either online or offline before implementation, to identify optimal parameter settings (window length and threshold values) to be used in the event detection algorithms determining a pattern for normal operation

[1] SCADA (Supervisory Control and Data Acquisition): System for remote monitoring and control

on typical background water quality. Past applications of fault detection tools have shown that changes in water quality due to routine hydraulic operations of a utility causes a high false alarm rate (Allgeier and Umberg, 2008). A case study application of CANARY as part of the UK multidisciplinary “Pipe Dreams” to historic data from a UK distribution system demonstrated that the correlation of events to relevant network information, e.g. pipe bursts reduces the number of unexplained faults (ghosts) and decreases the false alarm rate (Mounce, Machel and Boxall, 2012). One way to overcome this problem is to incorporate improved algorithms, which are able to recognize normal changes in water quality due to hydraulic operations. The implementation of clustering techniques for identifying patterns within time series data and use of water quality template libraries was proposed (EPA, 2010). In the meantime, the software is extended by trajectory clustering pattern matching algorithms and composite signals capacity, but their application did not significantly improve the performance (EPA, 2013). As mentioned before online application of CANARY requires input data of online sensors and therefore the integration in water utility’s SCADA system, which is generally feasible, because CANARY’s implementation is not depending on specific sensor manufacturers.

2.4.1.2 OptiEDS

The optimal Event Detection System (optiEDS) designed by Elad Salomons (OptiWater, 2018) enables monitoring of measured and computed water quality parameters such as chlorine, TOC and pH as well as operational data to detect anomalous water quality conditions. The system is capable to define a normal dynamic baseline of parameters and to monitor large set of data in real-time. Furthermore, it allows customized adjustments to utility’s water network. In case of detected abnormal process conditions the system triggers an alarm and reports the “suspicious” parameters. The software is based on trend analysis to detect the deviations from water quality baseline and is able to incorporate the unique water network operation logic to provide engineers and operators of water utilities with additional knowledge to the conditions of the system. More detailed information about the used methodologies and algorithms has not been provided by the developer. OptiEDS was one of the systems tested in the EPA EDS challenge. EPA constituted OptiEDS a good performance in the detection of basic events (EPA, 2013).

2.4.1.3 BlueBox

Since the participation in the EPA EDS challenge new features of the event detection system BlueBox (Whitewater Security, 2018) have been developed resulting in continuous improvements on the system, which has led to implementations and its utilisation by several water utilities in practice. The system is able to define and incorporate operational data, such as indication of pumps or changes in measurement of operational values, e.g. pressure or water flow and can therefore distinguish between suspected abnormal quality changes and deviations due to normal operations. Providing additional information about correlation between water network operation and changes of water quality to the system resulting in an increased accuracy in fault detection. BlueBox is able to differentiate whether a change in water quality is caused by an equipment fault, e.g. pump breakdown or by a water quality event. Furthermore, the system features a self-learning event classification approach enabling the user to categorise unknown events as “true” or “false” and establish an event classification library whose utilisation probably will result in an increased detection accuracy and decreased false alarm rate.

The feasibility of BlueBox to incorporate time parameters (division into time periods) similarly enables the reduction of false alarm rate due to seasonal effects. An integrated reporting module allows utility’s operators to generate data analysis reports, e.g. alarm statistics or events history. More detailed information about the used methodologies and algorithms has not been provided by the developer. Further features and improvements including an auto calibration approach to configure the EDS automatically for each monitoring station, a planning tool for calculation the optimal sensor locations within the WDS and a spatial detection module for fault detection in WDS sub regions (EPA, 2013) are scheduled in early future.

2.4.2 Developer-specific Solutions – Guardian^{Blue} and s::can

In this section two commercially available sensor developer-specific event detection systems, namely Guardian^{Blue} and s::can are presented.

2.4.2.1 Guardian^{Blue}

Guardian^{Blue} (Hach Homeland Security Technologies, 2007) developed by Hach Homeland Security Technologies monitors continuously total chlorine, conductivity, pH, turbidity, temperature and pressure by the water panel module and total organic carbon by the TOC analyser. Guardian^{Blue} integrates the sensor data provided by the water panel and TOC analyser and analyses the data by the application of a proprietary algorithm to calculate the water quality baseline.

The system analyses collected sensor data every 60 seconds and calculates the trigger signal indicating deviations from water quality baseline. If trigger signals violate the user defined thresholds the system raises alarms followed by an automatically triggered capture of a real-time sample at designated sensor locations. The system analyses then plant and agent libraries with containing event fingerprints of previous water quality changes to classify the abnormal condition. After event classification, the system reports a probability of a certain event to the utility's operator. If no match is found the operator can define the event as "unknown event". Guardian^{Blue} is designed to work with water sensors developed by Hatch, which limits the choice of equipment that may be used and could be a hindrance to its implementation from water utility perspective (Szabo and Hall, 2014).

2.4.2.2 s::can

The software approach **s::can** (s:can, 2013) developed by s::can integrates three software modules: (i) sensor- and station management module (moni::tool), (ii) data validation module (vali::tool), and (iii) event detection module (ana::tool). All necessary operational information about measurement devices, such as maintenance, calibration etc. is provided to the user by the integrated sensor- and station management software. The real-time validation module ensures that only "clean" data are used for further analysis, training and alarms. Several statistical methods were applied to remove outliers, correct discontinuous measurements, and reduce noise from instruments or measured parameters using state estimation and residual classification techniques. Sensor drifts in the measurement of the observed parameters were identified by modelling sensor readings utilising the Holt-Winters method (Holt, 1957; Winters, 1960).

The software module for event detection establish the state of the system and triggers an alarm in case a significant deviation from normal state is detected. The methods used for calculating individual alarms are static thresholds, multiple values outside of tolerance band limits, pattern recognition using specific correlations between observed values and changes in the light absorbance spectrum of water, assumed that at least one spectrophotometer probe is installed. The results of individual alarms are combined to a cumulative alarm, which triggers automatically a final alarm if significant abnormal state conditions were detected, caused by any alarm individually or a combination of all individual alarm algorithms (Edthofer et al., 2010).

Avoiding non-event-related data entering the detection module results in a decrease of the false alarm rate, which has been confirmed by the results in the EPA EDS challenge. With 0.9 false alarms per week ana::tool exhibit the lowest average false alarm rate of all tested tools (EPA, 2013). The software trains itself on the incoming data and after the training period it monitors conventional water quality parameters with the feasibility to detect exiguous deviation in water quality by tracking changes in the spectral fingerprint, i.e. simultaneous changes in the light absorbance spectrum of the water. The software accepts any type of developer-specific s::can sensors. A notable drawback of the system could be the use of spectrophotometers, since UV/Vis sensors are expensive and still not widely used for data analysis at WTWs.

2.4.3 ERS Software under Development – SAFEWATER and H_2O Sentinel

2.4.3.1 SAFEWATER

The SAFEWATER (SAFEWATER, 2015; Bernard et al., 2015) project funded by the EU is investigating a global water management model to detect and mitigate chemical, biological, radiological and nuclear drinking water contamination events. The aim of the project is the development of a comprehensive water management solution including spatial models for detection of abnormal water quality events within drinking water system and simulators for the determination of contamination sources as well as new water quality sensor technologies.

SAFWATER's solution will be tested by integration into the SCADA systems of three water utilities participating in this project.

The new water management system contains an event detection model based on an unsupervised machine learning technology to detect changes of water quality parameters and on a decision support tool to provide advice about the best mitigation measures in case of a water quality event. An analysis algorithm is applied on the data provided by the sensors respective by the SCADA system to estimate the location of the contamination source.

The event detection module learns the normal behaviour of the system as well as arising abnormal occurrences and generates indications about its condition by using different methods, such as violation of limits, rare combinations of abnormal occurrences, similarities to past event situations and violation of rules. Each indication is based on one or multiple detectors, which are algorithms to detect specific abnormalities in data, e.g. exceeding a statistical limit of a variable. In this case the system generates an alert. An event is triggered if several detectors are alerting simultaneously. The classification of an event needs to be processed by the system's user, whereat the system is capable to learn from the classification and improves its alerting policy.

Provision is made to incorporate advanced simulations of the hydraulic behaviour in the system's network and the dynamics of water quality parameters to support the decision making process of water utilities by utilising existing platforms, such as EPANET-MSX and SIR-3S after their enhancement through SAFEWATER. The project furthermore includes the development of two innovative water quality sensor technologies. For the detection of chemical contaminants a compact bacteria-based chemical online sensor based on measuring rapid light changes emitted by natural marine luminescent bacteria and for the detection of E.coli bacteria an antibody-based sensor will be provided.

2.4.3.2 H2O Sentinel

Regarding the event detection system H2O Sentinel under development by Frontier Technology, Inc., no valid information was provided as well as for several other ongoing approaches, reported by EPA (EPA, 2013) and for this reason it is only been given a brief mention here as a notable fault detection software system under development. In addition to the above, the reviewed FDD methodologies and approaches in water sector are discussed in the following section.

2.5 Discussion

This literature review covers general fault detection methods and previous work in the field of event recognition in the water sector as well as specific software applications for the use in practice. A various number of methodologies were developed in the past years. In this section key findings are discussed and suggestions on the further work presented.

In general terms, it can be noted that only a limited set of already available ERS technologies appear to be applicable and practical to be used by the water industry. As major drawback, the lack of reliability could be identified, since these state-of-art systems still generate a high number of false alarms. This insight was confirmed by the published EPA Water quality event detection system challenge report (EPA, 2013) whose results of testing five ERS facilities have outlined, that the event detection performance of the participant systems varies greatly and the number of invalid alerts (false alarm rate) is generally high. None of these systems is convincing and hence there is a clear need for the development of new technologies enabling the detection of failure events at WTWs in a reliable and timely manner. The recent research in this field is focussing to overcome this issue and therefore several new fault detection methodologies have been proposed in recent past.

However, the vast majority of methodologies proposed in literature are history-based fault detection techniques. This is not surprising, since no deep process knowledge is required for the application of these data-driven methods. Pure quantitative or qualitative model based models are very extensive to develop, since modelling of systems with permanently changing conditions such as water treatment processes is complex and adapting a model for modifications would be to too laborious and time-consuming. Given this context, it is an obvious requirement for a successful technique not to be dependent on deep a priori knowledge. Only data-driven methodologies meet this condition, but also some data-driven methodologies, such as Expert Systems (Punal, Roca and Lema, 2002) and case-based reasoning techniques (Ruiz, Colomer and Melendez,

2007) require knowledge about historical process behaviour. Although it may not as extensive as for model-based systems, it is still a reasonable effort and frequently not easy to acquire all necessary process knowledge for the successful application of these methods, what makes Expert Systems and CBR applications impractical for use in industry.

Most extensively investigated data-driven methodologies applied SPC techniques from which the overwhelming part combine multivariate PCA based on dimension reduction with diverse classification methods. Details of reviewed methods are all summarised in Table 2.2. at the end of this section. Although the presented PCA methods have shown promising results, most of them have been tested either on short-sized or simulated data, which makes it difficult to compare their performance. Furthermore, PCA or extended ICA applications are not easy to handle, since first the construction of the model is difficile regarding their feature selection and second the analysis of the classification results is impractical, because multiple chats have to be reviewed and evaluated by the user. The application of univariate SPC control charting methods would enable to offset these disadvantages.

Unfortunately, univariate control charts have not been extensively investigated for their fault detection capabilities at WTWs. The moderate detection performances demonstrated by the experiments utilising Shewhart, EWMA and CUSUM control chart variants (Corominas et al., 2010; Zheng, Yekun and Qiao, 2018) are difficult to interpret, because either no baseline method for a valid comparison was tested or methods were applied by using standard parameter settings. Although, or even because these methods have not been further investigated, the potential for combining them with appropriate other methods aiming to overcome the possible drawback of moderate detection performance and take advantage of their easy handling should be explored intensively, which was done and will be presented in the further course of this thesis.

The most successful methodologies found in literature applied AI techniques such as artificial neural networks, but also other machine learning classifiers as support vector machines. The approach introduced by Piciaccia et al. (2018) has proven to be promising for event classification. Most machine learning techniques were applied for detection of contamination events and achieved accuracies between 80-90%. Although the majority of these results have been achieved by the application to synthesized datasets and engineered contamination events, which are assumed to be detected more easily than minor events, machine learning classifiers in combination with SPC techniques for fault detection are found to be most promising for the development of the new ERS for WTW.

A successful application of detection techniques depends also on the quality of data used. Only a few approaches were presented in the literature that provide separate data validation methodologies. Improving the quality of measured data from sensors deployed at WTWs by applying procedures based on automated data validation and correction has promise to be beneficial to increase detection performance and thus the reliability of event detection methods (Talagala et al., 2019). The application of simple rules and statistical transformations, as it was done by Talagala et al., has demonstrated to be efficient in the detection of technical sensor faults and outlier removal from real-world data collected by in situ sensors monitoring water-quality in a natural river system. Although these improvements have shown not to be tremendous, an appropriate sensor data validation methodology should be the first step in the development of a new ERS.

Table 2-2 *Reference* summary *of fault detection approaches in water sector.*

No	Model Reference Author (Year)	Model Class[1]	Model Type[2]	Optimization	Variant[3]	Application[4]	Fault Detection (FD) Method[5]	Parameters analysed	Test Method	Notes[6]
1	Lennox et al. (2001)	PHb	ANN		MV	WTW	ANN	NTU, flow rate, ozone/coagulant dosage, pressure	onsite	applied to water filtation process
2	Rosen and Lennox (2001)	PHb	PCA	adaptive scaling paramerters; adaptive PCA; Multiscale PCA (MSPCA)	MV	WWTW	T^2 / SPE Chart (FD)	Flow rate, pH, NH_3, T	simulation	PCA not capable; adaptive PCA not fully capable
3	Punal, Roca and Lema (2002)	PHb	Expert System		MV	WWTW	rule-based (FD/FDi)	Flow rate, CH_4, T, CO, pH	pilot plant	experimental events applied
4	Praus (2005)	PHb	SVD		UV	WTW	FA	Colour, Ca, Cl_2, NO_3^-, NO_2^-, Fe, pH, SO_4^{2-}, EC, TU, MnO_4^--index	onsite	No=253 drinking water samples
5	Aguado and Rosen (2007)	PHb	PCA	PCA; adaptive scaling; adaptive Covariance Matrix	MV	WWTW	T^2 / SPE Chart (FD) Fuzzy c-means (FDi)	Flow rate, NH_4, T, K_L, NO_3^-,NH_4	simulation	BSM1 data-replacement
6	Ruiz, Colomer and Melendez (2007)	PHb	MPCA		MV	WWTW	MPCA (FD) CBR-AI-tool (FDi)	DO, T, pH, ORP	pilot plant	batch process hybrid model
7	Garcia-Alvarez et al. (2009)	PHb	PCA		MV	WWTW	T^2 / Q Chart (FD) FDA (FDi)	physical, chemical and biological parameters	simulation	synthetic model simulated for 14 days
8	George, Chen and Shaw (2009)	PHb	PCA		MV	WTW	T^2 / Q Chart (FD)	DOC, TU, T, pH, Cl_2, CT, Al, Lime Unit, Colour, Flow Rate	onsite	applied to real data (14 d), not been validated on unseen data
9	Chen and Huang (2011)	PHb	ANN		MV	WDS	NN / immune algoriyhm (FD)	pH, EC, T, DO	simulation	Sensor fault detection
10	Padhee, Gupta and Kaur (2012)	PHb	PCA		MV	WWTW	backpropagation based neural classifier (FD)	physical, chemical and biological parameters	simulation	data pre-treated
11	Perelman et al. (2012)	PHb	ANN		UV/MV	WDS	sequential Bayesian analysis; multivariate fuse information (FD)	TC, EC, pH, T, TOC, TU	simulation	applied to real data (4 month), split into calibration and validation period
12	Alferes et al. (2013)	PHb	PCA		MV	WWTW	T^2 / Q Chart (FD)	pH, EC, T, TU, DO, TOC, DOC, K, NH_3, NO_3^-	onsite	data pre-treated, auto-scaling, fault diagnosis: DB setup
13	Lamrini, Lakhal and Le Lann (2013)	PHb	ANN		MV	WTW	fuzzy-neural model (FD)	T, pH, TU	simulation	data pre-treated

14	Arad et al. (2013)	PHb	ANN	extended No. (12): dynamic thresholds	UV/MV	WDS	sequential Bayesian analysis; multivariate fuse information (FD)	TC, EC, pH, T, TOC, TU	simulation	offline analysis of dynamic tresholds
15	Oliker and Ostfeld (2014)	PHb	SVM		MV	WDS	weighted SVM; sequence analysis (FD)	TC, EC, pH, T, TOC, TU	simulation	data pre-treated
16	Liu, Smith and Che (2015)	PHb	SC		MV	WDS	Pearson correlation coefficient (FD)	T, pH, TU, EC, ORP, DOC, NO_3^-, PO_4^{3-}	pilot plant	pilot scale laboratory CIE
17	Liu et al. (2015a)	PHb	SC		MV	WDS	Pearson correlation coefficient; euclidian distance (FD)	T, pH, TU, EC, ORP, DOC, NO_3^-, PO_4^{3-}	simulation	data from pilot scale laboratory CIE
18	Oliker and Ostfeld (2015)	Hybrid	Phb	extended No. (13)	MV	WDS	Minimum volume ellipsoid; sequence analysis; local+spatial analysis (FD)	TC, EC, pH, T, TOC, TU	simulation	multiple sensors: artificial database local+spatial data hydraulic integrated
19	Housh and Ostfeld, (2015)	PHb	ANN	extended No. (15): integrated logit detection (ILD)	UV/MV	WDS	ILD; sequential Bayesian analysis (FD)	TC, EC, pH, T, TOC, TU	simulation	offline analysis of dynamic tresholds
20	Meyers, Kapelan and Keedwell (2017)	PHb	RF		UV	WDS	Random Forest Classifier	Flow rate, TU	simulation	applied to real data (1 year), split into calibration and validation period
21	Page, Waldmann and Gahr (2017)	PHb	ANN		MV	WTW	self-organizing maps, Sammon projection	EC, SAC254, pH, T	simulation	applied to real data (14 d), not been validated on unseen data
22	Bernard et al. (2017)	PHb	N/K		MV	WTW/WDS	several N/K algorithms	pH, ORP, EC, DO, T, TCC	onside	simulation of seven contamination szenarios
23	Piciaccia et al. (2018)	PHb	SVM		MV	WTW	SVM classifier, ablative analysis	Flow rate, T, Polymer Dosage, among others	simulation	applied to real data (6 month)
24	Inoue et al. (2018)	PHb	ANN SVM		MV	WTW	(a) Deep NN (DNN) (b) SVM classifier	water quality variables of 25 sensors and 26 actuators	pilot plant	pilot scale, engineered contamination events
25	Zheng, Yekun and Qiao (2018)	PHb	ANN		UV	WDS	Deep Belief NN (DBN) with Extended Kalman Filter (DBN-EKF)	Flow rate	simulation	applied to real data of (4 years), split into calibration and validation period

Abbrevisions:
1) PHb-Process History-based,
2) ANN-Artificial Neural Network, Mb-Model-based, MPCA-Multiway PCA, PCA-Principal Component Analysis, SC-Statistical Classifier, SVD-Singular Value Decomposition, SC-Statistical Classifier, SVM-Support Vector Machine, RF-Random Forest
3) UV-Univariate, MV-Multivariate
4) WTW-Water Treatment Work, WWTW-Waste Water Treatment Work, WDS-Water Distribution System
5) SPE-Squared Prediction Error, FA-Factor Analysis, CBR-Case-Based Reasoning, AI-Artificial Intelligence, FDA-Fisher Discriminant Analysis, VDA-Vecor Distance Algorithm, SVM-Support Vector Machine, ILD-Integrated Logit Detection
6) BSM1-Benchmark Simulation Model no.1, CIE-Contaminant Injection Experiment

2.6 Summary and Conclusions

This chapter has discussed previous work in the wide field of fault detection in general and for the water sector. In section 2.2 an overview of general FDD methods was presented which includes quantitative and qualitative model-based mythologies as well as process history-based, i.e. data-driven and hybrid model-based techniques. Section 2.3 provided a survey of relevant work done so far in the water sector, including (i) fault detection methods for WTWs and WWTWs, (ii) fault detection methods for water infrastructure and, (iv) near real-time ERS software applications. Finally, main findings of the survey have been discussed.

The key chapter conclusions are as follows:

- At the present ERSs used in practice do not achieve great performances in the detection of failure events at WTWs and generate a high number of false alarms. The results of the Water Event Challenge (EPA, 2013) have not reported of ERSs achieving precision values (see Section 4.5) greater than 60%.
- Proposed methodologies for the detection of failure events at WTWs found in literature have been applied either to data generated by pilot plants and/or engineered events, i.e. usually contamination events. Demonstration of the methods on real data of a comprehensive number of sensors deployed at demonstration sites and on real events identified by the analysis of signals used has not been shown in literature.
- The application of hybrid models that combine SPC methods and AI machine learning techniques for the event detection at WTWs has not been studied in literature.
- Sensor data validation and pre-processing methods proposed in literature have not elucidated their potential to reliably detect sensor failures or their capabilities to distinguish between faulty WTW sensor data and faulty processes at WTW in near real-time.
- None of the presented sensor data validation and pre-processing methodologies has proven its effectivity on the detection of failure events at WTW’s processes by demonstrating resulting detection performance on real sensor data and real events (particularly minor events).

Chapter 3: Case Study Description and Data

3 CASE STUDY DESCRIPTION AND DATA

The methodologies presented in this thesis are established for general application to different water quality sensor signals of various water treatment processes. All development and validation have been performed using real sensor data taken from a selected WTW located in the north-west region of the UK. The application of data-driven statistical process control and machine learning techniques utilized for event detection procedures require the use of historical sensor data. The effective application of used supervised learning techniques requires a significant amount of data streams containing labelled events for the mapping of new events.

This chapter presents the WTW and the utilised sensor data and failure events used throughout this thesis. The WTW description starts by briefly presenting the WTW's water treatment processes followed by the description of the process for collection and identification of sensor signals (critical alarm points) utilised for the generation of the final dataset used throughout the thesis. The chapter continues with the description of the techniques used for the identification and labelling of both the major and minor failure events within the explored dataset. Finally, a brief summary of the chapter is given.

3.1 Real-life WTW

UK water companies supply water to domestic and industrial customers usually similar in principle. Water is fully treated before being supplied to a distribution system and delivered to customers. Water treatment processes are designed to remove microbiological organisms, physical and chemical substances, e.g. algae, suspended solids (turbidity), nitrate that affect health aspects and/or aesthetic acceptability.

Adequate water treatment to provide the required drinking water quality is achieved by the physical removal of contaminants and usually consists of a number of stages. These stages typically include sedimentation frequently combined with coagulation processes by the addition of chemical agents, filtration and disinfection processes. Design and effectiveness of a treatment is heavily dependent not only on site conditions, but also on the chemical and microbiological consistence of the water to be treated which determines the chemical dosing requirements (DWI, 2016).

A real-life WTW operated by United Utilities is selected as a study site throughout this thesis. This WTW is situated in the North West of England and supplies water to around 200,000 domestic and industrial customers with 73.5 Ml/d flow capacity. The process flow scheme is shown in Figure 3-1.

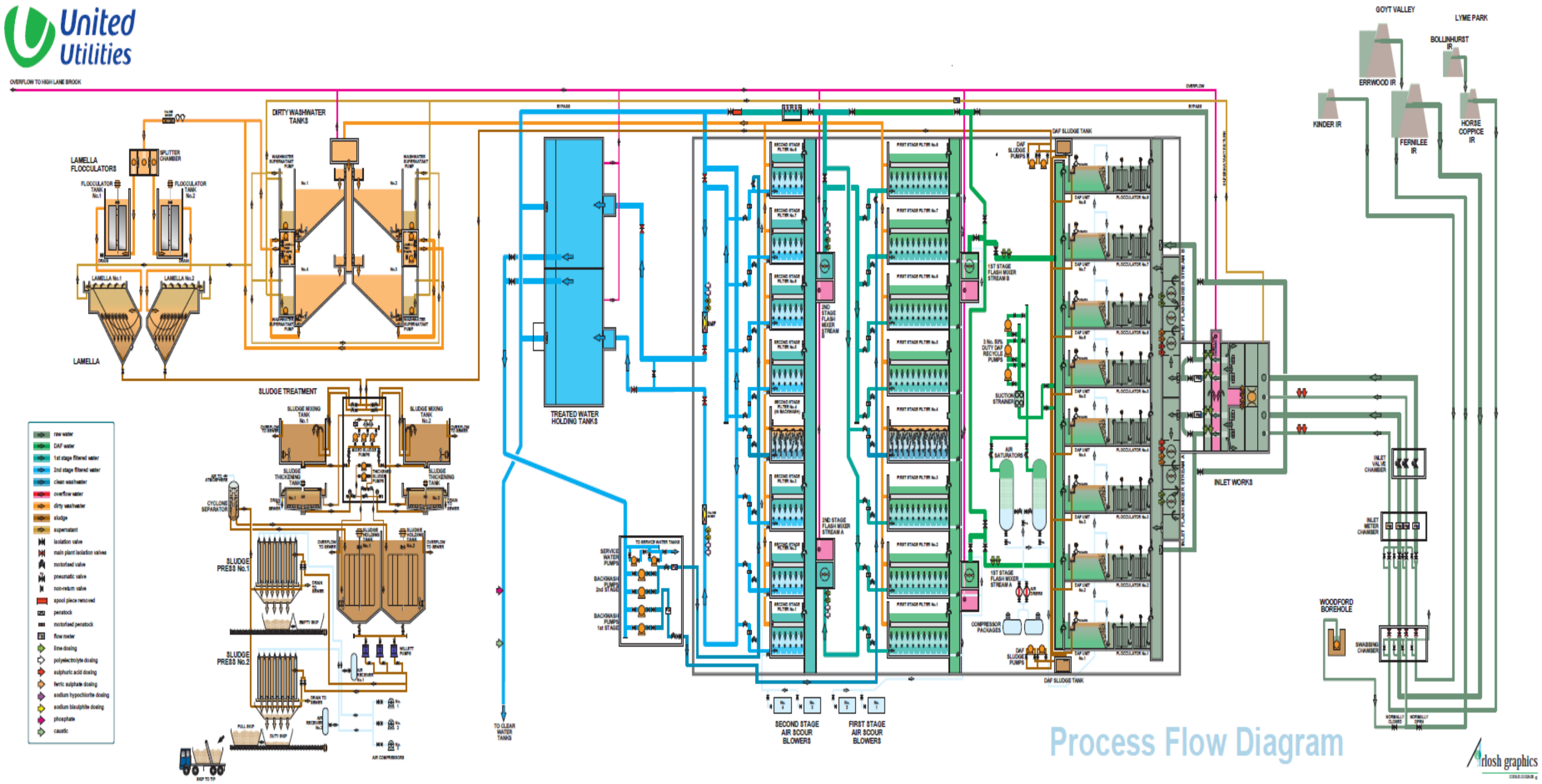

Figure 3-1 *Process flow diagram of a typical drinking water treatment works (Courtesy: United Utilities).*

As it can be seen from Figure 3-1, raw water is abstracted from different water sources and enters the WTW at the inlet chamber, where it is mixed with supernatant recycled flow from dirty backwash water and afterwards split into two separate streams (stream A and B). After dosing for coagulation and pH adjustment, water of each stream is treated by Dissolved Air Flotation (DAF), first stage filtration and second stage filtration processes. After filtration, treated water enters the water holding tanks at the outlet works where both streams are combined and presented for the final disinfection procedure. Sludge produced by the filtering processes is thickened, pressed and discharged. At larger scale WTWs, water is frequently split into two (as pictured here) or more streams for separate treatment using similar treatment processes.

To ensure the required drinking water quantity and quality the WTW is heavily automated and controlled. Usually, Supervisory Control and Data Acquisition Systems (SCADA) are used to control the water treatment processes by near real-time monitoring of water quality and flow parameters by sensors deployed at the WTW.

3.2 WTW Sensor Data

Historical data for 56 sensors over four and a half calendar years from 01/01/2012 to 30/06/2015 and at a 5-minute resolution was collected as continuous time series data of the individual signals. For event detection relevant water quality signals of both streams (stream A and B) were identified within each single treatment stage and mapped to their corresponding sensors/tags and locations. Initial data screening resulted in 28 signals selected for further analysis. A basic schematic showing the sensor locations (continuously numbered from WTW’s inlet to the outlet stage) is presented in Figure 3-2.

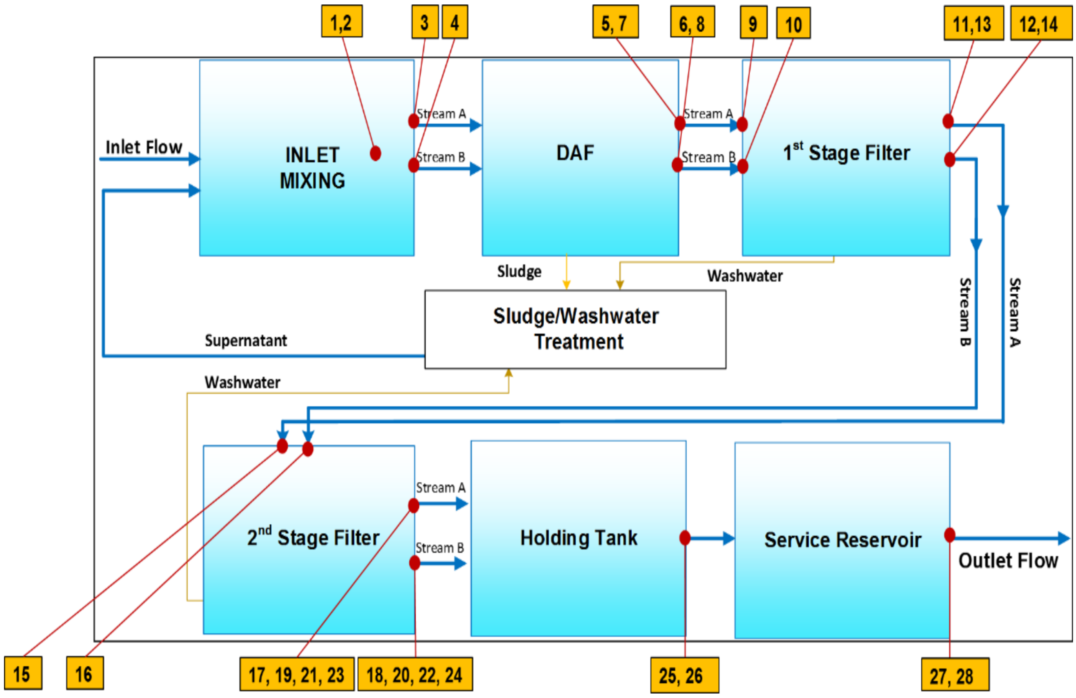

Figure 3-2 *Basic schematic of mapped sensor locations.*

Corresponding to their mapped locations shown in Figure 3-2, an overview of the collected signals is provided in the Table 3-1. The table presents both stream A and stream B signals and indicates their deployed locations and treatment stages, respectively.

Table 3-1 *Sensor signals and corresponding treatment stages.*

Sensor Signal	Sensor (No.)	Stream	Treatment Stage (Location)
Raw Water Turbidity	1	Combined	Inlet Works
Raw Water pH	2	Combined	Inlet Works
Pre Flocculation pH	3,4	A,B	Flocculation & Flotation
Post Flotation Turbidity	5,6	A,B	Flocculation & Flotation
DAF Iron	7,8	A,B	Flocculation & Flotation
Pre 1st Stage pH	9,10	A,B	1st Stage Filtration
Post 1st Stage Turbidity	11,12	A,B	1st Stage Filtration
Post 1st Stage Iron	13,14	A,B	1st Stage Filtration
Pre 2nd StagepH	15,16	A,B	2nd Stage Filtration
Post 2nd Stage Turbidity	17,18	A,B	2nd Stage Filtration
Post 2nd Stage Chlorine	19,20	A,B	2nd Stage Filtration
Post 2nd Stage Colour	21,22	A,B	2nd Stage Filtration
Treated Water pH	23,24	A,B	2nd Stage Filtration
Outlet Contact Tank Chlorine	25	Combined	Outlet Works
Outlet Contact Tank pH	26	Combined	Outlet Works
Final Water pH	27	Combined	Outlet Works
Final Water Chlorine Residual	28	Combined	Outlet Works

The quality of sensor data utilised is an important factor that affects the performance of any detection system. Low data quality will limit the meaningfulness of predictions and erroneous data will lead in the worst case to faulty conclusions (Rieger et al., 2010). For this reason, the quality of the collected data streams has been assessed aiming to extract final datasets only with signals of sufficient high data quality that are crucial for a robust event detection and therefore suitable to be utilised for further analysis. The valuation of data quality was geared into two directions. Single signal streams of poor data quality over considerable long time periods, i.e. more than one month were explored first, followed by the investigation of certain time periods in which multiple signals show low data quality aiming to exclude both unreliable signals and time periods from the datasets used for the following tasks.

Availability and consistency of data were considered as major criteria for the data quality assessment. The valuation against the availability criterion was conducted by a missing data analysis applied to every single signal. If data was missing for more than one month continuously in some signal that signal was considered unreliable. This way, six signals, i.e. 1st stage iron (stream A and B), post 2nd stage colour (stream A and B), outlet contact tank chlorine and outlet contact tank pH signals were identified as unreliable and hence omitted from further analysis. The remaining 22 signals, hereinafter also referred to as critical alarm points, remained for the generation of the final datasets.

Data consistency was used as criterion to evaluate time periods containing multiple signals of poor data quality. Statistical analysis was conducted to determine basic indicators, such as minimum, maximum and mean values on the one hand as well as additional parameters including range, variance and frequency measures to establish baseline values for the following data consistency's assessment of the remaining 22 water quality signals on the other. Additional information of missing data (days/month) and flat line duration (days/month) on corresponding signals supplemented the analysis which was conducted on each signal individually. The parameters were derived for the whole time period (4.5 years) first and then for each month of that period separately. This way single months in which multiple signals showed striking abnormalities (inconsistencies) were explored and then judged as "unsuitable", i.e. contains too many inconsistent signals to be used within the final datasets if these inconsistencies apply to more than half of the analysed signals. Figure 3-3 shows as an exemplary the graphs of the selected pH value and turbidity in the time period from 20.03.2015 until 30.06.2015. The figure illustrates that all of the 12 signals displayed, i.e. more than half of the total of 22 signals, show only flat lines from 27.05.2015 to 30.06.2015. Therefore, the time period from 01.05.20015 until 30.06.2015 has been excluded from the final dataset used for further analysis.

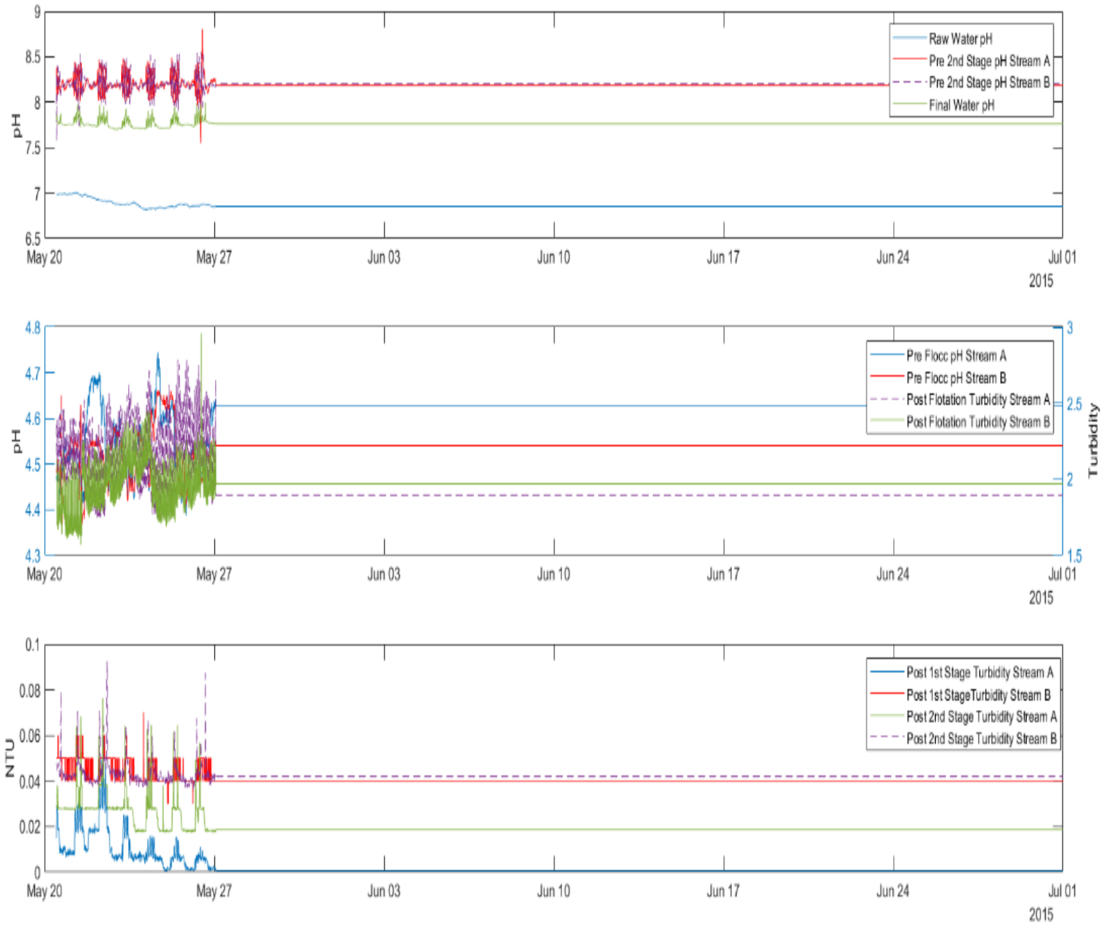

Figure 3-3 *Example pH and turbidity signals showing striking anomalies during the time period from 27/05/2015 until 30/06/2015*

Figure 3-4 shows a selected range of pH and turbidity signals in the period from 02/2015 to 06/2015 where the data quality of the pictured signals continues to decrease with progressing time from 09/03/2015 on (i.e. increasing number of missing data, unusual spikes and flat line faults marked with grey bars) until all signals show frozen values on 27/05/2015.

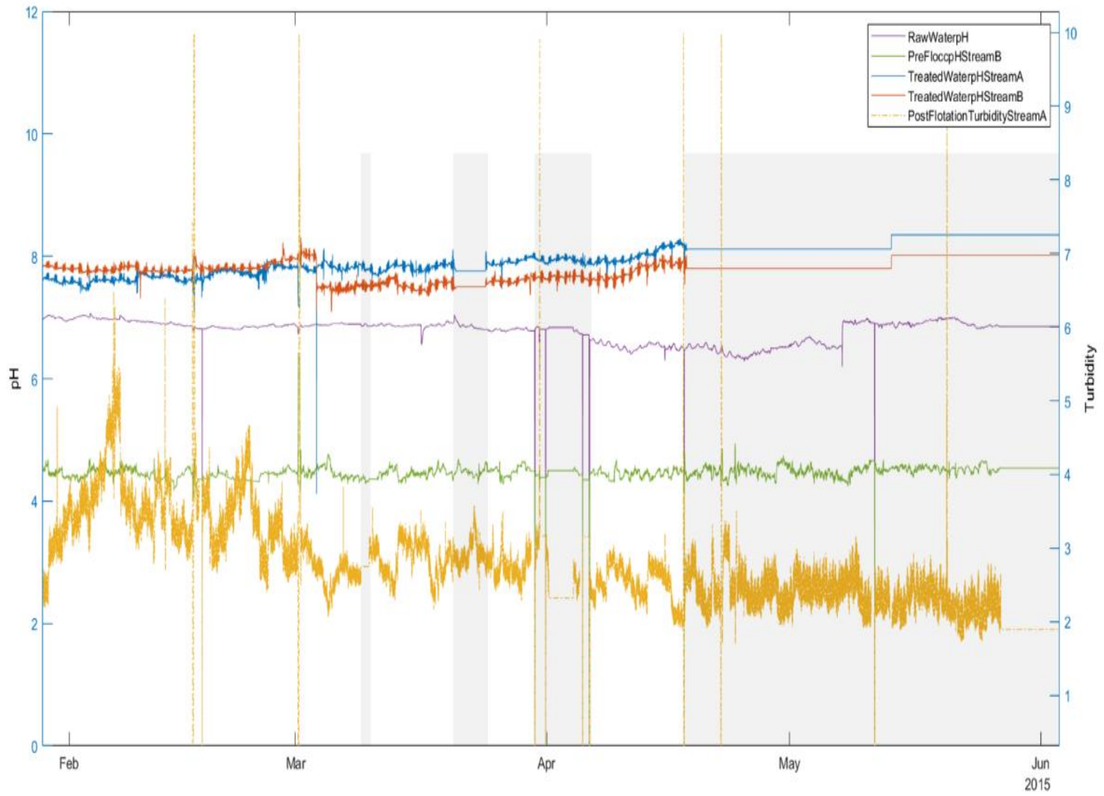

Figure 3-4 *Example pH and turbidity signals showing striking anomalies during the time period from 09/03/2015 until 30/06/2015*

This way the time period between March 2015 until June 2015 was identified rather to be excluded from further analysis, since the data quality of the vast majority of individual signals within this period were judged to be poor.

Additionally, to confirm the results obtained from the above described statistical analysis a spectral analysis was conducted on each individual signal across the whole time period of collected data. Example spectrogram plots of a selected range of signals is shown in Figure 3-5.

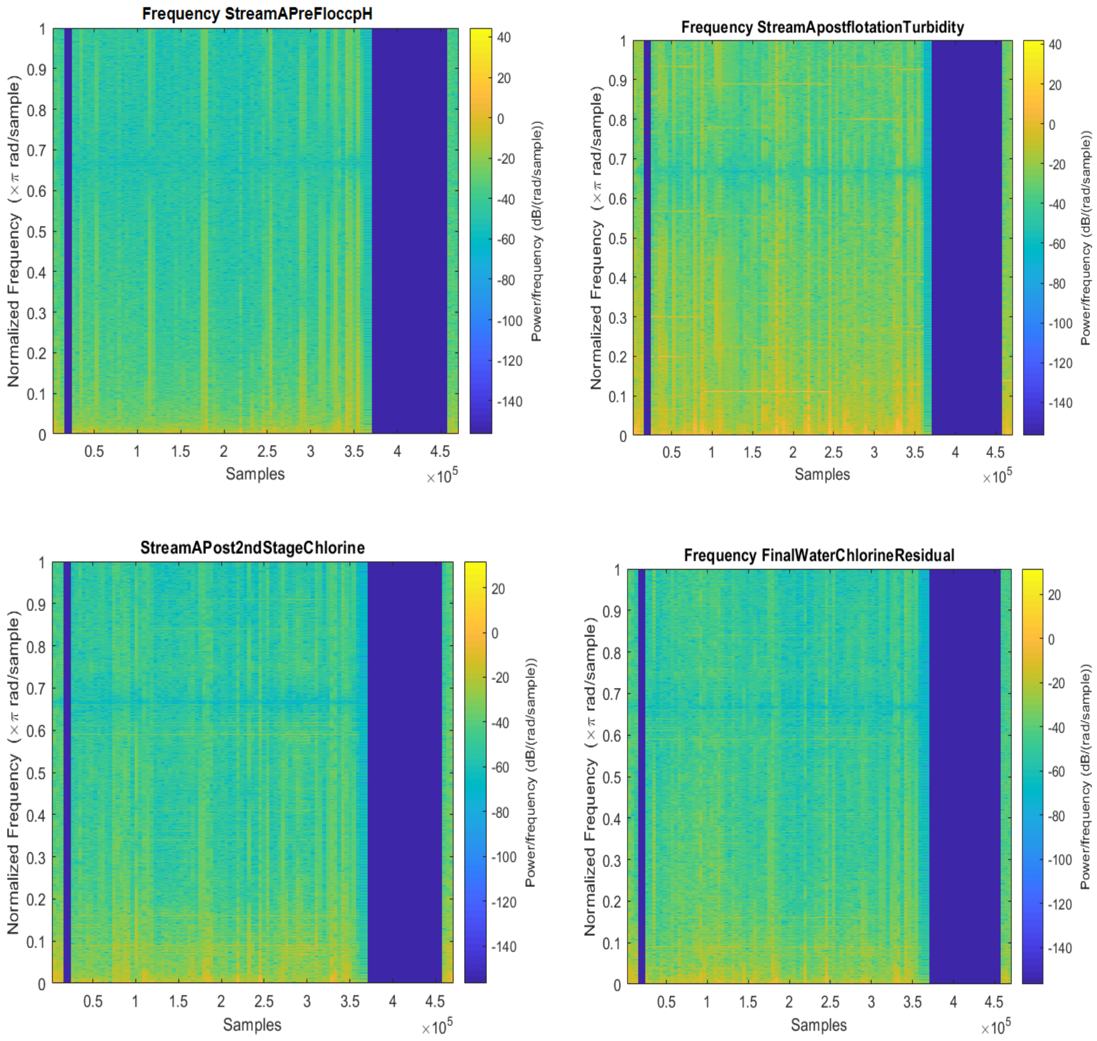

Figure 3-5 *Example spectrogram plots.*

From these plots it can be seen that the presented signals have, in contrast to preceding periods, almost no frequency (dark blue colour) at the end of sample period. Similar observations were made on the vast majority of the remaining signals (not shown here to save space). These low frequency periods indicate significant data inconsistencies of multiple signals towards the end of the time period for which the data were collected.

For this reason, only the time period from 01/01/2012 until 01/03/2015 including the 22 critical alarm points described above was used for further analyses. The data was then split into datasets for calibration of detection models (time period from 01/01/2012 until 28/02/2014, i.e. ~ 70% of total time period) and follow-on validation on unseen data (time period from 01/03/2014 until 01/03/2015, i.e. ~30% of total time period).

Due to the reasonable large number of signals utilised, i.e. 22 water quality parameters measured throughout all five treatment stages and its high number of samples (i.e. 7,324,416 data points over all 22 signals) the new generated dataset seems to be well-suited for the development of a generic data-driven event detection method. This ensures that the methodologies described in the following can be transferred to other WTWs using different treatment processes and/or different water quality parameters. Additionally, within the reasonable large time period from 01/01/2012 until 01/03/2015 used in the dataset, there should be a sufficient number of faults (i.e. 5 major events and 158 possible minor events, see Section 3.3) to comprehensively train some machine learning algorithms. Details concerning the evaluation of these failure events can be found in the following section.

3.3 Minor and Major Events

Given that only a small number of mostly major events were confirmed or reported from the water company, failure events that have occurred at the WTW during the used time period had to be explored before an assessment of the deployed detection system could take place. To address this issue, a number of historical events were identified first followed by their classification either as major, minor or sensor fault events.

The identification of events was carried out by visual inspection of the 22 water quality signals across the given time period of the used dataset. Here, major (or "zero-flow") events were defined as events that caused an interruption of the production flow and led to an unplanned shutdown of the whole WTW. This way, 5 zero-flow events were identified. Figure 3-6 shows a typical picture of a major event causing a shutdown of WTW's stream A at 12:40 on 14/09/2013 as a result of an alarm triggered by the stream A post flotation turbidity signal. A partial shutdown was followed by a corresponding drop of the inlet flow from around 55 ML/d to approx. 35 ML/d and its recovering to normal state after the restart of stream A at 16:45 on 14/09/2013.

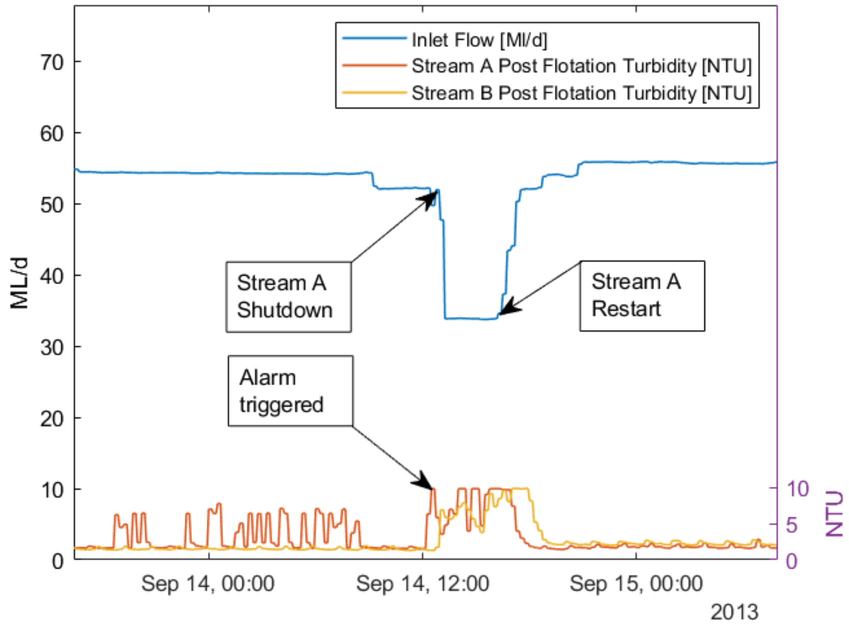

Figure 3-6 *Example major event - shutdown and restart of WTW's stream.*

Minor events were identified by looking at simultaneous deviations of more than one signal from normal operating process conditions without causing any WTW's shutdowns. Figure 3-7 shows a typical example of a minor event, where stream A post 2nd stage chlorine and stream B post 2nd stage chlorine signals together with the final water chlorine residual signal have dropped to zero almost simultaneously at 08:15 on 28/01/2014.

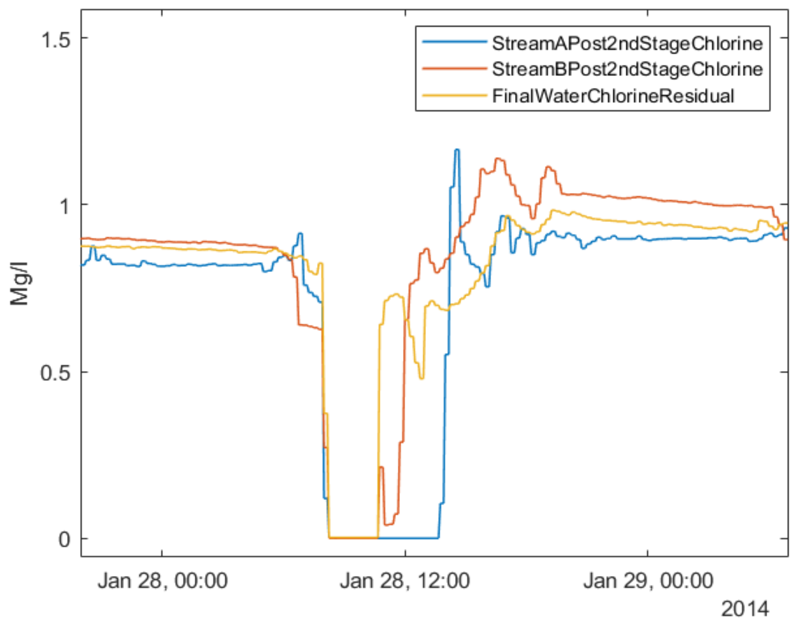

Figure 3-7 *Example of minor event.*

To identify these kind of events, normal WTW's operating conditions were analysed based on common statistical indicators for minimum, maximum, mean and range of the 22 selected signals. Bivariate correlations between parameters were then calculated using the Spearman's correlation coefficient to derive possible related deviations of multiple signals from the corresponding normal values. Table 3-2 shows several examples of parameter combinations including the corresponding Spearman's correlation coefficients.

Table 3-2 *Bivariate correlations between parameters and corresponding Spearman correlation coefficients.*

Signal 1	Signal 2	Spearman's Correlation Coefficient (r_S)
Raw Water Turbidity	Raw Water pH	0.75
Raw Water Turbidity	Stream A Treated Water pH	-0.71
Raw Water pH	Stream B Treated Water pH	-0.70
Stream B post flotation Turbidity	Raw Water pH	0.64
Stream A Post 1st Stage Turbidity	Stream A Post 2nd Stage Turbidity	0.76
Stream A Post 1st Stage Turbidity	Stream A Pre 2nd Stage pH	0.61
Stream B Post 1st Stage Turbidity	Stream B Pre 2nd Stage pH	0.61
Stream A Treated Water pH	Final Water pH	0.75
Stream A Treated Water pH	Raw Water pH	-0.67
Stream B Treated Water pH	Final Water pH	0.83
Final Water pH	Raw Water pH	-0.77

Abnormal conditions were then identified by visual inspection of the displayed deviations. All analysed signals were plotted below each other for the full time period analysed (01/01/2012 to 01/03/2015). Figure 3-8 shows exemplary a small section of the plot illustrating the method used for the identification and labelling of minor events (pictured by grey bars within the plot). Significant deviations from normal process condition were marked for each individual signal and compared to the behaviour of the remaining signals. In case of simultaneous deviations of two or more signals the presence of a minor event was assumed, as it was shown in the example pictured in Figure 3-7. Deviations of single signal values from normal process conditions were classified as sensor faults. Using this methodology 158 possible minor events were identified during the analysed time period.

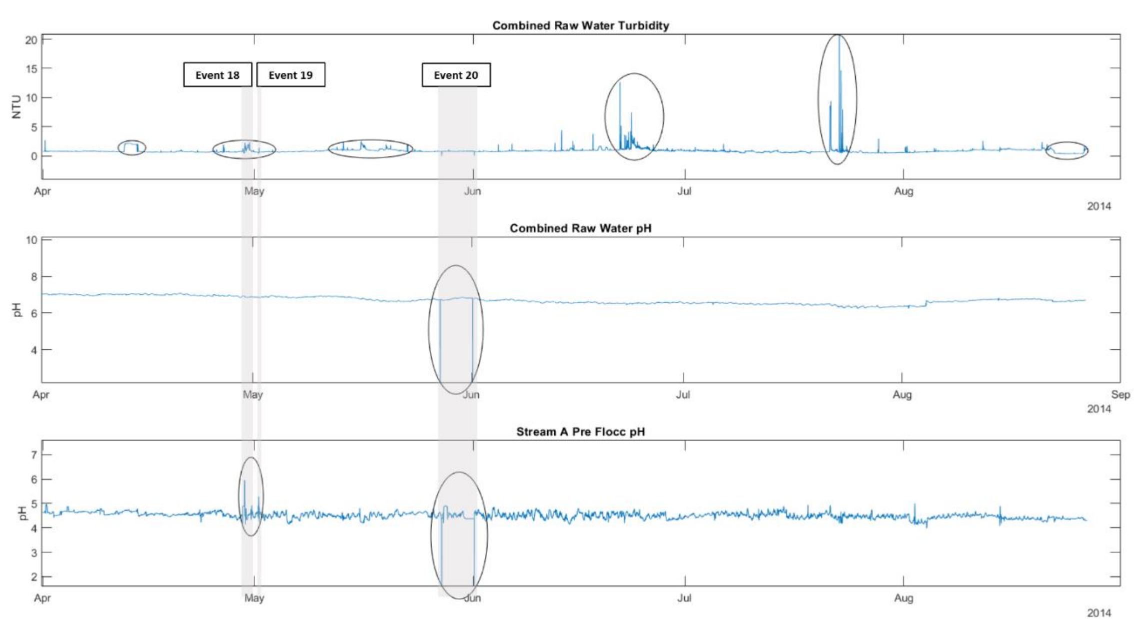

Figure 3-8 *Cut-out of the time period analysed and signals used illustrating the technique applied for the identification and labelling of minor events shown as grey bars.*

A limited number of selected minor events identified this way were reviewed by an expert from United Utilities to confirm the approach for the determination of minor events. Once the events were identified, major and minor events were labelled within the final dataset. The identification of events complemented all necessary information for the establishment of baseline by evaluating the performance of the currently used detection system described in the later course of the thesis.

3.4 Summary

This chapter has briefly detailed the WTW study site and the datasets used to develop and evaluate the work presented by this thesis, besides their methods of data collection and labelling of example events. After a brief overview of WTW's treatment processes, this section has outlined the methods of data collection and validation applied for the identification of 22 water quality parameters (critical alarm points) and the exploration of suitable time periods that contain only signals of assessed data quality for the use within the final dataset. Finally, the procedure applied for the identification of different types of events and their classification into major and minor events including their labelling within the generated datasets has been described.

Chapter 4: Event Recognition Methodology

4 EVENT RECOGNITION METHODOLOGY

4.1 Introduction

WTWs operated by UK water companies are nowadays usually observed by threshold-based event detection systems utilising online sensors for the monitoring of water quality parameters in near-real time. The data streams provided by these sensors supply the required information for the identification of failure events and/or abnormal conditions within WTW's processes. As discussed in Section 2.3.1, online water quality and quantity monitoring technologies for WTW's operation have made significant progress in recent years, but still there is a clear need for improved online monitoring systems (Storey et al., 2011).

Nevertheless, although several methods for event detection at WTWs have been recently developed, only a few, such as Canary (Hart et al., 2007; Hart and McKenna, 2009) or GuardianBlue from Hach Lange (Hach Homeland Security Technologies, 2007) have found their implementations in practice and were utilised by the water industry. Most of them still suffer from a range of shortcomings, such as insufficient true detection capability or too many false alarms (Bernard et al., 2015). Moreover, the results of the EPA Water quality event detection system challenge published by EPA (2013) have shown that event detection performances of the five tested event detection systems vary greatly and the number of invalid alerts (false alarms) generated by these systems is generally high. New and more efficient technologies need to be developed to address these issues. The focus of further research is set on innovative, cost-effective and wherever possible predictive near real-time event recognition systems. Therefore, it is no surprise that the development of new technologies for near-real time sensor data validation and recognition of failure events at WTWs has become an increasing priority for water companies.

The work presented in this thesis shows the methodologies utilised to investigate possible improvements to existing, typically threshold-based event detection systems used to date by UK's water companies to control their WTW and outlines a novel methodology for improved, near-real time recognition of failure events. The following Section 4.2 details the assessment of the currently used detection system at the selected WTW to evaluate its performance. The result of the performance assessment established this way serves as a baseline for possible improvements and the development of the novel event recognition technology. Section 4.3 provides a description of the methods applied to achieve improvements on threshold-based detection systems utilising optimised threshold and persistence values and continues with the presentation of the methods utilised for sensor data validation and pre-processing. The development of the novel Hybrid CUSUM event recognition methodology is presented in Section 4.4 followed by Section 4.5 outlining the methods applied for the assessment of ERS' performance. Finally, a concluding summary of this chapter is given in Section 4.7.

4.2 Existing Event Recognition System (E-ERS)

Water quality data from the sensors assembled at the WTW to monitor its treatment processes can be utilised for the analysis of the deployed event detection system aiming to evaluate its detection performances by quantifying detection statistics on the basis of observed data streams. For this assessment, the historical sensor data of the real-life WTW operated by United Utilities were used. This section outlines the threshold based event detection system currently in use - hereinafter referred to as E-ERS - by providing an overview of E-ERS, aiming to determine the basics for its assessment and performance evaluation, as described later in section 4.5 of the thesis.

As mentioned in Section 3.1, water treatment processes are usually controlled by monitoring of crucial process parameters with sensors deployed at the different treatment stages of the WTWs in near real-time. Typical parameters monitored at WTWs are, e.g. turbidity, pH, temperature, dissolved organic carbon (DOC), conductivity and flow rates. Turbidity and pH are normally controlled continuously throughout all process stages (frequently multiple validated, i.e. two or more sensors measure the same signal at the same location). Other parameters such as chlorine and chlorine residual are usually monitored at the outlet of the WTW. Most WTWs in the UK make use of deployed event detection systems that automatically generate alarms after the detection of abnormal behaviour on observed signals to ensure an early detection of abnormal process conditions.

Commonly used event detection systems - as in the case of the WTW used as demonstration site - apply thresholds to the monitored signals for the identification of faulty processes, respectively failure events at their WTWs. In the following, the E-ERS is outlined and the exploration of how it is working in terms of rules applied by the system for triggering alarms is presented.

The E-ERS applies pre-defined thresholds to the monitored signals and carries out default actions (alarm/no alarm) in case of limit violations. Every 5 minutes, each sensor signal is checked against the default low and/or high thresholds. In addition to the limits a “time dead-band”, i.e. persistence is used by the system. Persistence defines the time a signal has to be continuously above/below a threshold before triggering an alarm. In case of two threshold values are set on a single signal, i.e. low and high limit, which allow the signal to vary between those values without triggering an alarm the same persistence value is used for both default limits. Using an example, the operating principle of the E-ERS can be illustrated as follows: for pre 1st stage pH signals the E-ERS applies low and high limits of 5.8 pH and 7.5 pH respectively, both with a default persistence value of 10 min. An alarm is raised (after exceeding of the default persistence given by 10 min) if the pH value of the pre 1st stage pH signal goes below 5.8 pH or above 7.5 pH and all subsequent measurements from this sensor remain below or above these limits within the next 10 min. Once an alarm is triggered it has to be checked and cleared by a human expert. In case of alarms triggered by auto-shutdown water quality set points (i.e. critical alarm points, see Section 3.2) cannot be cleared within a default set time period (e.g. 30 min for post flotation turbidity signal) an automatic partial or a complete WTW’s shutdown is initiated by the system.

All relevant information concerning the described WTW’s detection system including monitored parameters and configuration of the applied high/low thresholds and persistence was provided by the UU experts. Once the architecture of the detection system was identified, the E-ERS was simulated over the entire time period analysed using the data of the final dataset to assess its detection performance and establish the baseline for further improvements and developments.

4.3 Modified Event Detection System (M-ERS)

After the performance of the E-ERS was evaluated and the baseline was established (see Section 5.2), it was clear that the existing ERS needs to be improved. This was done initially by developing the modified ERS (denoted as M-ERS here). The M-ERS was developed by optimising the thresholds and persistence values of all water quality signals used in the E-ERS (Section 4.3.1). In addition, a methodology was developed to pre-process sensor signals with the aim to validate and if applicable to correct the sensor signal values in near real-time (section 4.3.2). A brief summary of the work done is given in the section 4.3.3.

4.3.1 Optimised Thresholds

A sensitivity analysis (Saltelli et al., 2004) was performed to investigate possible improvements to the E-ERS by changing the detection thresholds and persistence values. This was done on the calibration data set only (see Section 3.2). Plausible ranges of high/low detection thresholds for the analysed 22 signals were identified first. This was done by analysing the extremes of historical values under different WTW operating conditions. Within these ranges, new detection thresholds were created by applying a gradual increase of the threshold values in increments of 0.05 (e.g. 2nd stage chlorine thresholds we allowed to vary between 0.5 and 1.6 in 0.05 mg/l increments). The gradual change of low and high thresholds in increments of 0.05 was applied for all signals, except for post flotation turbidity signals where increments of 0.01 NTU have been used for increasing the low limits, i.e. post flotation turbidity was allowed to vary between 0.01 and 7 by gradually increasing low and high thresholds in 0.01 NTU and 0.05 NTU increments, respectively. The value of 0.05 was considered to be suitable since for various signals the E-ERS uses thresholds with a 0.05 resolution, e.g. the high limit of final water chlorine residual which is set on 1.35 mg/l. The value of 0.01 NTU used for the gradual increase of low threshold values for post flotation turbidity signals has been selected to capture the threshold value of 0.01 NTU set by E-ERS for these signals and to allow a finer tuning of this specific threshold value. Persistence values were changed from 0 to 12 time steps (i.e. from 0 to 60 minutes). The value of 60 min as maximum persistence was selected since the same value was utilised by the E-ERS as highest persistence applied to individual signals, e.g. to post 1st stage iron signals. Therefore, the persistence value of 60 min was desired to be explored for the other signals as well. The threshold and persistence value ranges applied by the sensitivity analysis in comparison to current used limits are shown for all water quality parameters in Table 4-1.

Table 4-1 *Threshold settings and threshold ranges used by sensitivity analysis.*

		Current Detection System		Modified Detection System	
Signal	Unit	Low Limit	High Limit	Low Limit Range	High Limit Range
Raw Water Turbidity	NTU	-	10.00	0.05 - 0.50	5.00 - 15.00
Raw Water pH	pH	5.50	7.90	4.00 - 6.00	7.00 - 9.00
Pre Flocculation pH Stream A	pH	4.0	4.80	4.00 - 4.40	4.50 - 4.80
Pre Flocculation pH Stream B	pH	4.0	4.80	4.00 - 4.40	4.50 - 4.80
Post Flotation Turbidity Stream A	NTU	0.01	6.50	0.01 - 0.50	5.50 - 7.00
Post Flotation Turbidity Stream B	NTU	0.01	6.50	0.01 - 0.50	5.50 - 7.00
DAF Iron Stream A	mg/l	-	2.50	0.00 - 0.05	2.00 - 4.00
DAF Iron Stream B	mg/l	-	2.50	0.00 - 0.05	2.00 - 4.00
Pre 1st Stage pH Stream A	pH	5.80	7.50	5.00 - 6.95	7.00 - 9.00
Pre 1st Stage pH Stream B	pH	5.80	7.50	5.00 - 6.95	7.00 - 9.00
Post 1st Stage Turbidity Stream A	NTU	-	0.50	0.00 - 0.10	0.15 - 0.60
Post 1st Stage Turbidity Stream B	NTU	-	0.50	0.00 - 0.10	0.15 - 0.60
Pre 2nd Stage pH Stream A	pH	6.80	8.60	6.00 - 7.75	8.00 - 9.00
Pre 2nd Stage pH Stream B	pH	6.80	8.60	6.00 - 7.75	8.00 - 9.00
Post 2nd Stage Turbidity Stream A	NTU	-	0.40	0.00 - 0.05	0.10 - 0.60
Post 2nd Stage Turbidity Stream B	NTU	-	0.25	0.00 - 0.05	0.10 - 0.60
Post 2nd Stage Chlorine Stream A	mg/l	0.60	1.40	0.50 - 0.70	1.00 - 1.60
Post 2nd Stage Chlorine Stream B	mg/l	0.60	1.40	0.50 - 0.70	1.00 - 1.60
Treated Water pH Stream A	pH	6.80	8.60	6.50 - 7.20	8.00 - 9.00
Treated Water pH Stream B	pH	6.80	8.60	6.50 - 7.20	8.00 - 9.00
Final Water pH	pH	7.00	9.00	6.50 - 7.00	9.00 - 9.50
Final Water Chlorine Residual	mg/l	0.60	1.35	0.60 - 0.80	1.00 - 1.50

This way a total of up to 7,540 sensitivity tests were conducted for each of the 22 signals resulting in estimated corresponding true and false positive detection rates. The optimised new thresholds and persistence value combinations were then derived for each sensor signal by selecting the combination with the maximum value of the ratio of true positives to false positives (see Figure 4-6 in

Section 4.5). The flowchart shown in Figure 4.1 illustrates the workflow of above presented optimisation procedure.

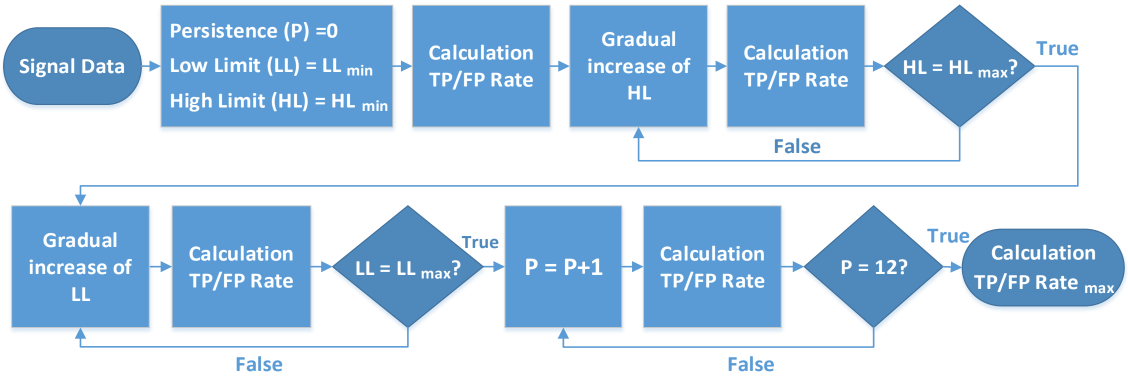

Figure 4-1 *Flowchart outlining the procedure applied for the optimisation of threshold and persistence values.*

Once the optimised threshold and persistence values were identified, the M-ERS that makes use of optimised thresholds was simulated over the entire time period analysed using the same historical time series data (5 min intervals) of the 22 critical alarm points with the same historical events as used before for the evaluation of the E-ERS. The results obtained this way were then compared to evaluate detection performance improvements.

4.3.2 Sensor Data Validation and Pre-processing

To carry out monitoring and control in an adequate manner requires a huge number of measured values that needs to be continuously refreshed (Edthofer et al., 2010). Due to various factors, such as frequently varying water demand, changing influent conditions, dynamics in water treatment processes and unreliable or missing sensor data WTWs monitoring and controlling is a challenging task for water supply companies.

Low data quality will limit the meaningfulness of predictions and erroneous data will lead in the worst case to faulty conclusions (Rieger et al., 2010). On the other hand, sensor data validation methods used in water systems are often inefficient and frequently a systematic analysis is missing (Branisavljević, Prodanović and Pavlović, 2010) For this reason, it is beneficial for the performance of any detection system to raise the quality level of sensor data measured in such systems. To achieve this higher level of data quality desired for more robust and reliable events detection, it is feasible to validate and pre-process the collected sensor data before its possible use for event detection.

This section describes the methodologies established for near real-time validation and pre-processing of WTW's measured sensor data with the aim to investigate whether this data can be trusted and used for further fault detection at the WTW. The methods presented in this section enable the identification of faulty sensor data including the detection of different kind of WTW sensor faults on the one hand and near real-time data pre-processing to lift the data quality used for event detection on the other. First, the philosophy of sensor data validation is outlined followed by the description of methodologies developed for sensor data validation and pre-processing.

As mentioned before, a reliable, accurate and rapid detection of failure events (i.e. process faults and sensor faults) is of immense importance for the efficient and effective WTW's operation. In contrast to process faults, that possibly adversely affect the quality of WTWs' processes, sensor faults cause a decrease of accuracy and reliability in the measurements that may lead to erroneous control action and false perception (Yoo et al., 2008). For this reason, the capability to distinguish rapidly between sensor faults and genuine process failures is beneficial for any new ERS. This only can be achieved by the validation and if appropriate correction (pre-processing) of the measured data from sensors deployed at WTWs in near real-time.

The philosophy of the data validation method developed is focussing on the validation of near real-time data coming from sensors that measure "indicating" parameters which are most relevant for event detection, i.e. critical alarm points. The method aims to verify water quality data observed by the WTWs sensors and to check whether the data can be trusted and reliably utilised for the detection of faulty WTWs processes. Furthermore, the method was developed to identify certain sensor faults. The focus was set on detection of the following sensor fault types: (i) erroneous data (i.e. negative values), (ii) missing data, (iii) unusual spikes (i.e. sharp change in measured value in a small number of successive samples), also referred to as an outlier and, (iv) flat line faults (i.e. constant values across a larger number of successive data points). Examples of respective sensor faults are shown in Figure 4-2.

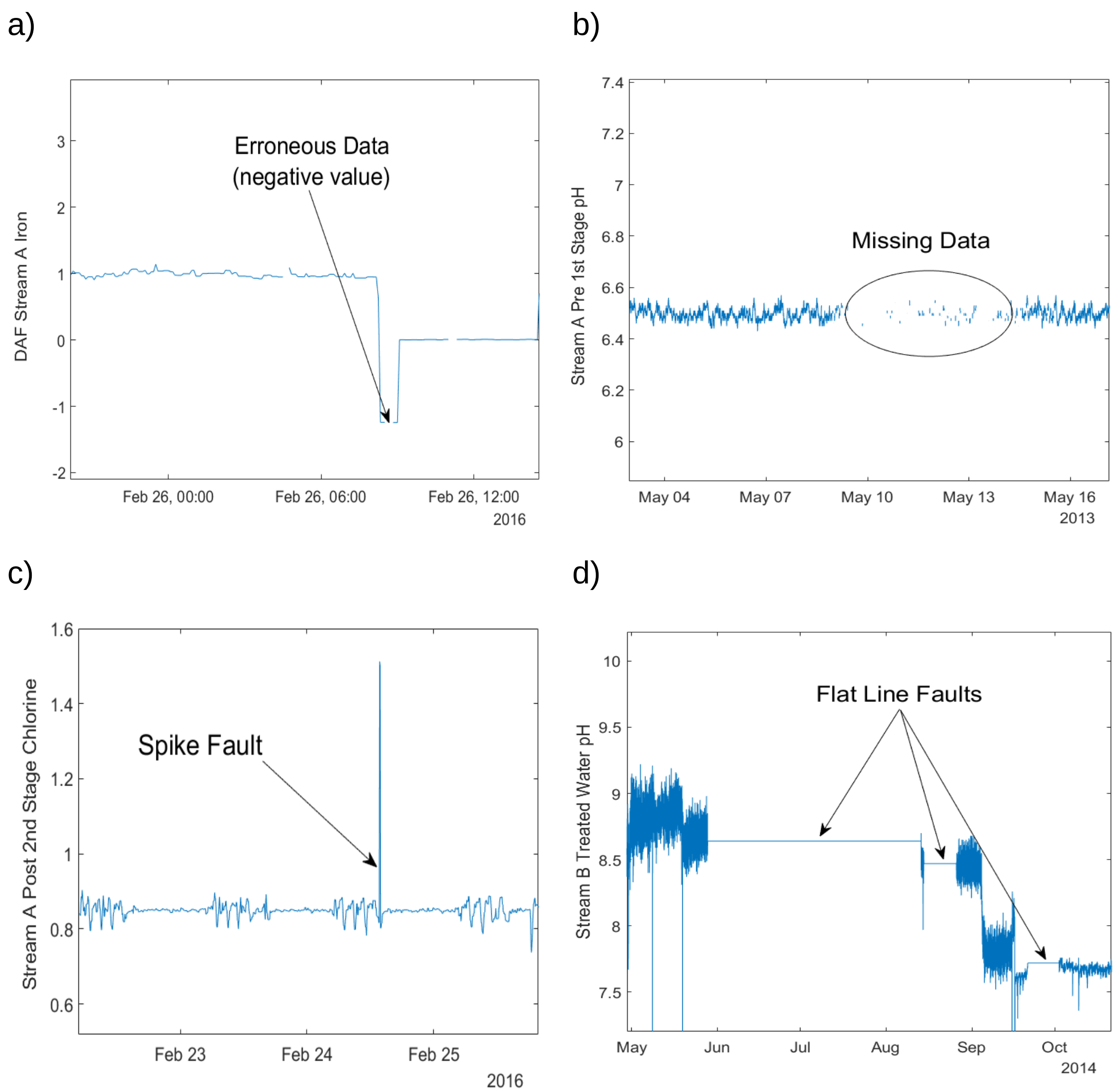

Figure 4-2 *Example sensor faults: (a) Erroneous data, (b) Missing data, (c) Spike fault, (d) Flat line faults.*

In addition to the presented sensor faults, there are other types of sensor faults such as noise, bias, drifts, but it was assumed that most of them will not have this great impact on detection system's performance. Therefore, those types of sensor faults have not been considered in the development of the below described sensor data validation and pre-processing methodology.

Therefore, the methodology developed here is focusing on the identification, marking and replacement of the following faults: (i) erroneous data points, (ii) missing data, (iii) spike faults and, (iv) flat line faults. The data validation and pre-processing procedure adopted consists of four statistical tests and can be described as follows.

Data points of each individual signal coming from WTWs sensor are online checked at every 5 min time step against its validity by four subsequently applied statistical tests.

In the first test, erroneous data, i.e. negative sensor measurements are identified, and in case of their presence the data point is marked as invalid data. Each negative value is then replaced by the preceding validated sensor measurement.

The second test comprises the identification of missing data. Blank sensor readings were detected and subsequently marked as missing data point. The identified data point is then replaced by its preceding validated sensor measurement.

In the third test, unusual spikes were detected by the identification of suddenly appearing sharp changes in the measured values. The recognition of these spikes was done by calculating the corresponding gradient values for every new data point and comparing its value to pre-defined thresholds. Suitable threshold values for both positive and negative slopes were determined by analysing the probability distribution of gradients over the calibration time period for each signal separately. This was done by means of histograms grouping the data into bins of equal size corresponding to the value of 0.1 for the gradient change per TS (5 min). In the histogram a rectangle is erected over the bin whose height represents the proportion of data points in the bin. With normalization, the height of each bar is equal to the probability of selecting an observation within that bin interval, and the height of all bars sums to 1. Analysis of the histogram plots for each signal has shown that the shapes of the gradient distribution nearly follow gaussian distributions, except at the histograms' edges where outliers have been pictured. The bins at these edges are representing the data points with extreme low (negative) or high (positive) gradient values considered as unusual spikes (outliers). The visual inspection has further shown that only a small fraction of the total amount of data points represent those outliers at each edge, i.e. usually less than 0.05%. Therefore, the value of 0.05% of the total number of analysed data points was considered as most suitable for the differentiation between gradient values representing signal's normal deviations and unusual spikes. This value was then utilised to calculate for each signal corresponding positive and negative gradient values used as thresholds for smoothing the unusual spikes. These thresholds were derived by sequentially grouping all bins at the edges of the histogram whose cumulated probability distribution do not exceed the anticipated threshold of 0.05% outliers. In each case the edge of the resulting bin represents the respective gradient threshold. The evaluation of the positive gradient threshold for the raw water pH signal is exemplary illustrated in Figure 4-3. The probability distribution plot of signal's gradients pictures that the overwhelming number of gradients range between -0.2 and 0.2, while only for a small proportion, i.e. ~0.05% gradients between 0.2 to 7.1 are shown (illustrated by the data label of the right bin). Since the probability of the bin on the right edge of the plot representing the cumulated probability of gradient values between 0.2 to 7.1 is just below the outlier threshold of 0.05% all data points within this bin are considered to be outliers. The value of 0.2 on the left edge of the bin represents the positive gradient threshold value derived this way.

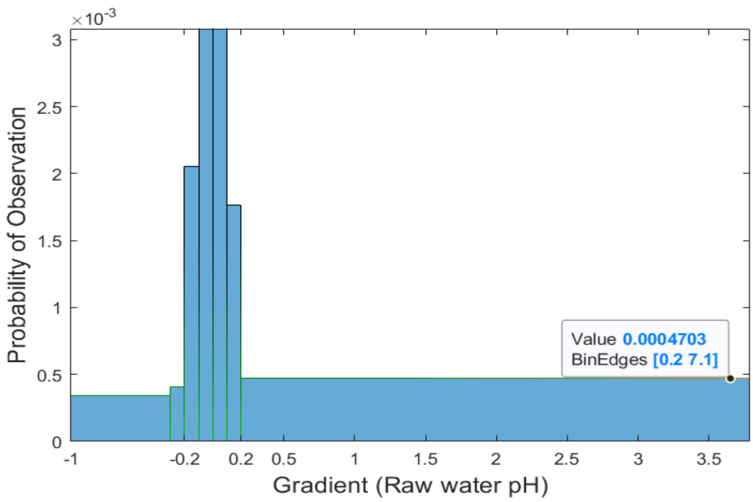

Figure 4-3 *Example distribution of gradients.*

Only in cases where the calculated gradient value of a new measured data point violates the default gradient threshold values an outlier is considered and the corresponding data point is marked as spike fault. The identified data point is then replaced by its preceding validated sensor measurement. Using the above generic calculation of gradient thresholds allows a high degree of automation regarding the process of spike removal.

The final, fourth validation test comprises the identification of flat line faults. At every time step, the actual sampled data value is compared to its last measured value. The number of consecutive duplicated values is counted and compared to a user set threshold. A threshold value of 2 hrs (i.e. 24 consecutive time stamps) was selected for this. By choosing this value it was aimed to use a single threshold value commonly applicable to all signals. The analysis of signal's behaviour at normal WTWs operating conditions has revealed that several signals, in particular signals with small measured values such as 2nd stage turbidity (usually range between 0.005 and 0.1 NTU at normal process condition) frequently show constant values over a duration of one hour or more, but usually not longer than 2 hrs. Those cases were considered to be more related to sensor's precision rather than sensor faults. For this reason, 2 hrs were selected as most suitable threshold value. Only if the counted number of consecutive duplicated values exceeds this threshold value the presence of a flat line fault is considered, and all following duplicated sensor readings marked as flat line faults. In case of sensor flat line faults, no further data pre-processing procedure is applied.

All data readings from sensors that monitor the analysed 22 critical alarm points were treated by this procedure in the same way. The sensor data verified and/or pre-processed was then used for a further simulation of the system followed by the performance assessment of the modified event detection system that makes now use of validated and pre-processed sensor data.

4.3.3 Section Summary

In this section the methods applied for the development of the M-ERS have been presented. The section has outlined two strategies to achieve possible improvements on the performance of the E-ERS, first the use of optimised thresholds and persistence values and second the application of high quality sensor data by using sensor data validation and pre-processing. The procedure for the investigation of optimal threshold and persistence values using sensitivity analysis has been described first. After that, different types of sensor faults have been briefly described and the novel methods for the validation and pre-processing of sensor data collected by the sensors deployed at the WTW outlined by discussing the four simple statistical tests applied for the identification of erroneous data, missing data and unusual spikes in near real-time. All described methodologies have been developed to lift the detection performance of E-ERS.

The following section 4.4 “Hybrid CUSUM Event Recognition System (HC-ERS)”, provides an overview of different sensor faults and describes the methodologies developed for sensor data validation and pre-processing followed by the evaluation of possible improvements on detection performance of the modified system after the utilisation of pre-processed sensor data.

4.4 Hybrid CUSUM Event Recognition System (HC-ERS)

4.4.1 Overview

Threshold-based detection systems often lack the sensitivity and specificity needed for accurate classification whereas sensitivity is defined by the true positive rate (TPR) and specificity by the complement of the false positive rate (FPR) for a test (Pichumani, 1997). It is this very effect that was observed by the assessment of the threshold-based detection system’s performance, where both, E-ERS and M-ERS show moderate true positive rates and suffer from a high number of false positives (see Sections 5.2 and 5.3). The drawback of lacking specificity and sensitivity was only overcome by moving away from threshold based to more sophisticated event detection methods.

This section describes the new methodology behind the Hybrid CUSUM Event Recognition System (HC-ERS). This detection system combines the classic SPC methodologies for fault detection with modern machine learning techniques for event classification. The first fault detection method identifies abnormal deviations of individual signals from their process means and generates labelled faults on each signal as binary output. This output is then used as input for the second event detection method (RF classifier) which estimates the probability of the presence of an event as ultimate output of the HC-ERS.

The new hybrid CUSUM event recognition method that makes use of the near real sensor data validation and pre-processing technique already described in Section 4.3.2 comprises two principle stages: (1) CUSUM fault detection and (2) RF event detection. The process scheme of the event detection procedure is pictured in Figure 4-4 highlighting the two integral event detection stages of the HC-ERS.

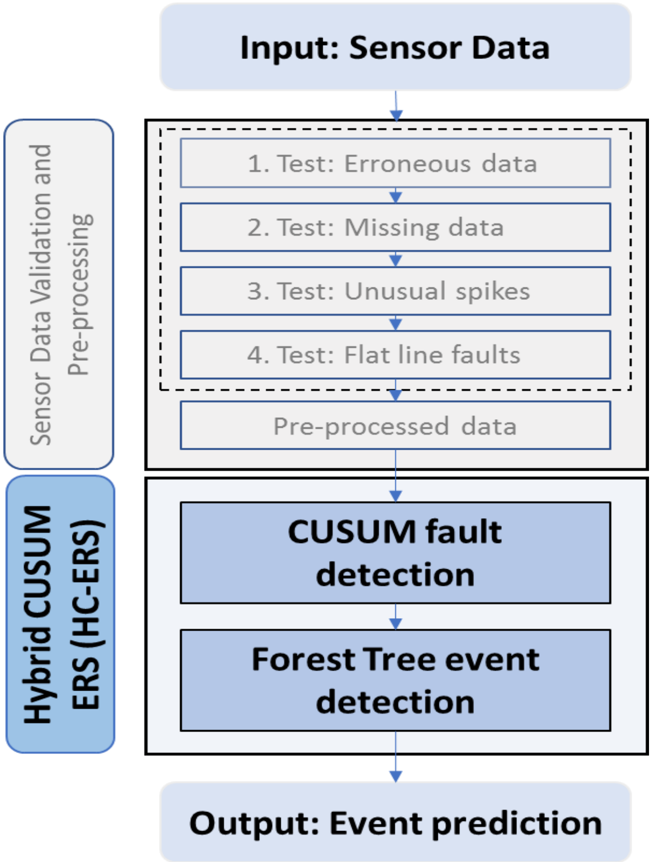

Figure 4-4 *Process scheme of Hybrid CUSUM ERS.*

This section is organised as follows. In section 4.4.2 the development of the fault detection method as an integral part of the new HC-ERS is described first. The development evolves the evaluation of several established detection methods including their testing and performance assessment. Section 4.4.3 presents the event classification methodology as second component of the HC-ERS as well as their optimisation methodologies carried out by an improved feature selection for the classifier input. Finally, a brief summary of the section is given in the Section 4.4.4.

4.4.2 Fault Detection Method

The first fault detection stage of the HC-ERS is described in this section. The fault detection methodology aims to detect the presence of process faults at WTWs by identifying relevant deviations of individual water quality parameters from normal process conditions and is applied to continuous data of 22 observed signals (critical alarm points). The method was developed with respect to the objectives defined in Section 1.3. These premises involve the development of a generic methodology capable of fault detection for water treatment processes in a reliable and timely (near real-time) manner. As such, the methodology is stipulated to quickly detect small shifts from normal process conditions in near real-time, i.e. at best detecting faults on observed sensor data at once the data is measured.

The detection method itself, utilises the data-driven Statistical Process Control (SPC) technique presented in Section 2.2.3. To establish the specific approach, that performs best for detection of faulty processes at WTWs, selected SPC control chart methods that have proven their ability to perform well for the use in fault detection of small shifts in the process mean (Montgomery, 2009) were chosen for further analysis. Control charts, in general, utilise sampled data of process variables taken over a period of time (in the following also referred to as window size) to represent the statistic of their quality measure (e.g., mean, range) aiming to investigate process changes over time. Control charts are constructed on the same principle as follows: the average of the statistic over all samples is calculated and represent the centre-line of the chart. The standard deviation (σ) of the statistic is calculated over all samples to define Upper Control Limit (UCL) and Lower Control Limit (LCL) which are used as thresholds at which the process is considered ‘out of control’. LCL and UCL are specified as number (n) of standard deviations, where n is an integer equal to or greater than 1 (n=3 is typically used) above and below the centre-line. Any observations outside LCL or ULC indicate that the process is out of control. In this work, particularly $\bar{x}$ -, r- , s-charts (Shewhart, 1931) and EWMA charts (Roberts, 1959), but also CUSUM control charts (Page, 1954) have been adapted and tested. Whilst for the construction of $\bar{x}$ -, r- , s-charts that utilise averages, ranges and standard deviation of the samples as statistical quality measure only sample size and number of standard deviations utilised for the calculation of LCL and UCL have to be selected. EWMA and CUSUM control charts require additional parameters to be defined. EWMA charts that utilise exponentially weighted moving averages as statistical quality measure making use of information from observations collected prior to the most recent data point. To address this task, EWMA charts apply the lambda (λ) value, where λ is greater than zero and less or equal than one (typically λ= 0.4 is used). This smoothing (weighting) parameter that controls how much the current prediction is influenced from past observations weights most recent samples more highly than older samples. Higher lambda values give less weight to past observations in favour of the current observation.

Rather than examining the mean of the statistic independently, CUSUM charts represent the accumulation of information of current and previous samples. CUSUM control charts involve the cumulated sums of allowed deviations from a target value (average of the sampled variable) as statistical measure. The number of standard deviations the statistical measure is allowed to deviate from the target is represented by the reference value k (k value), where k is greater than zero standard deviations (typically used value k= σ). As greater the k value as larger the deviations the process is allowed to shift from its mean.

Since all these methods are data driven, the same validated and pre-processed historical time series data (5 min intervals) of the 22 critical alarm points with labelled historical events already used for the assessment of the E-ERS were utilised to carry out the experiments on the SPS methods. The mentioned SPC techniques make use of standard parameter settings for window size, LCL, UCL, λ and k value, i.e. same values for those parameters were applied on each individual signal. A sensitivity analysis, described later in this section, was then performed to fine-tune these parameters for each signal individually.

Using a sliding window technique, statistically significant abnormalities relative to the “normal pattern” were detected on each individual signal. Once a deviation from normal condition was detected by the applied SPC method, the corresponding time step is marked with the binary value ‘1’. In case normal condition of the signal the respective time step is labelled with ‘0’. This way, for each observed signal a vector containing ones or zeros at each observed data point was generated as output of the applied SPC fault detection methodology. All SPC methods were tested on different window sizes, i.e. 1 day (288 time steps of 5 min) and 1 week (2016 time steps). A range of standard parameters, particularly Upper/Lower Control Limits (ULC/LCL) were also applied for the testing. For EWMA charts a lambda (λ) value of 0.4 that is commonly used in literature was utilised as default value for the experiments. The reference value required for the CUSUM charts, was set by with k=σ for the conducted tests. Utilised parameters applied for the assessment of different SPC methods are detailed in Table 4-2.

Table 4-2 *SPC parameters used for standard test.*

Parameter	$\bar{x}$ chart	s chart	R chart	EMWA	CUSUM
window size	1d/1week	1d/1week	1d/1week	1d/1week	1d/1week
k	-	-	-	-	σ
λ	-	-	-	0.4	-
ULC/LCL	±3σ/6σ/12σ	±3σ/6σ/12σ	±3σ/6σ/12σ	±3σ/6σ/12σ	±3σ/6σ/12σ

All above described SPC methods were tested by the use of the same dataset (22 critical alarm points) and historical events (see Section 3.3). Once the testing was done, the evaluation of detection performances was carried out for each SPC method separately in the same way as it was done in the case of E-ERS and M-ERS separately for calibration time period and follow-on validation on unseen data. Among the tested SPC methods CUSUM has proven to perform best in fault detection (see section 5.4.2) and for this reason this method was chosen to be used in further analysis for investigating possible improvements of the event recognition method. CUSUM charts monitor the cumulative sums of deviations of observed values from a target value against time and uses the out of control signals to locate anomalous points or sequences (Farkas, 2016). CUSUM control charts display the cumulation of information of current and previous observations and therefore, they are generally able to detect small shifts in the mean of a process (Montgomery, 2009).

Even though CUSUM charts are mostly automated, some parameter can be fine-tuned for their optimised adaption to the specific fault detection application. In particular, CUSUM control charts require a precise definition of the mean shift parameter (k value). By changing the k value, the sensitivity of CUSUM method can be adjusted. As higher the k value as less sensitive the CUSUM charting method gets. Since results of the performance assessment of CUSUM fault detection using standard parameters presented in Section 5.4.2 show a high number of false positives, i.e. a high level of sensitivity the focus was set on the reduce of false alarms by the testing of higher k values. Therefore, further fine tuning of the system was conducted by adjusting the CUSUM parameters on each individual signal to investigate possible improvements on detection system's performance.

To achieve this, a type of sensitivity analysis was performed by changing k values for different control limits and time windows aiming to define best possible CUSUM parameter settings for each individual signal. This was done on the calibration data set only. New mean shift arguments and control limits were created by applying a gradual increase of the k values from 1σ to 9σ in steps of 1σ and control limit values were changed from 1σ to 3σ, 6σ and 12σ for the windows sizes of 1d (288 time steps) and 1 week (2016 time steps). This way a total of 1.584 sensitivity tests were conducted, i.e. 72 sensitivity tests for each of the 22 signals resulting in the estimated corresponding true and false positives. Parameters used for the CUSUM fine-tuning are shown in Table 4-3.

Table 4-3 *Parameters used for CUSUM finetuning,*

CUSUM fine-tuned Parameter	window size 1d	window size 1 week
k	1σ,2σ…9σ	1σ,2σ…9σ
ULC/LCL	±1σ/3σ/6σ/12σ	±1σ/3σ/6σ/12σ

The optimal fine-tuned control limits and k value combinations were then derived for each individual signal and time window by selecting the combination with the maximum Performance Indicator (PI), introduced for the assessment of the results as ratio of true positives to false positives (see Figure 4-6 in Section 4.5). The performance indicator is calculated by PI = TP/FP. PIs of the single signals were then averaged to identify an overall PI rate for each window size. By comparing the PIs of each window size, that window size showing the highest PI was considered as representing the best performing window size and chosen for the further analysis. Once the optimal window size was derived, optimal parameters, i.e. k value and upper/lower control limits for each individual signal were derived in the same way as it was done for the investigation of optimal window size and the parameter combination showing highest PI was chosen as optimal finetuned parameter setting for the respective signal. Further fine-tuning of CUSUM parameters was conducted in the course of the classifier input data optimisation described in the following section.

The CUSUM fault detection system that makes use of the optimised window size and individual finetuned parameter settings was then tested in the same way as it was done for the system using the standard parameter settings for each signal.

4.4.3 Event Detection Method

The objective of methodology described in this section is to investigate possible improvements on the fault CUSUM detection performance by moving away from the application of detection rules to individual sensor signals only. With the move away from treating these signals independently (i.e. univariate detection methods) towards a detection system that considers relevant relationships between multiple signals (i.e. multivariate detection method) a reduction in false alarms is expected. The second event classification stage of Hybrid CUSUM ERS method that aims to address this issue is described below. The second event classification stage of Hybrid CUSUM ERS method that aims to address this issue is described below.

The event classification process makes use of the predictions on individual signals received by the preceding CUSUM fault detection procedure. Once the binary output for each signal is generated as result of CUSUM fault detection process, a prediction about the output's contents can be made by a trained machine learning classifier. This classifier estimates from input data (CUSUM fault detection output) the probability of the presence of an event. The methodologies described below utilise specified thresholds to raise alarms if the probability of the event prediction is above the default threshold. Even though, the thresholds were set at a fixed value, those thresholds can be changed in order to generate Receiver Operating Characteristic (ROC) curves to evaluate the performance of the classification process at all threshold values (see Section 5.4.3). The work presented in this section uses Artificial Neural Network (ANN), Support Vector Machine (SVM) and Bagged Trees (BT) classifiers trained on the dataset of previously labelled CUSUM outputs.

ANN was chosen for its prevalence in the literature, successfully using functional relationships between data patterns and fault classes without modelling the internal process states or structure (Sin et al., 2012) to classify processes into healthy or faulty conditions (Padhee, Gupta and Kaur, 2012) by minimising the probability of false and missed alarms (Srivastava, Srivastava and Vashishtha, 2014). The SVM was selected for its recognition in the literature, to be applied in some of the most recent and successful methodologies (Yin et al., 2014; Sahri and Yusof, 2014). Furthermore, SVM is well suitable in treating problems with low samples and high input features (Zhiwei, Cecati and Ding, 2015). This ability can be beneficially used for rapid prediction of events in the near-real time. Moreover, SVM classifiers have already shown high accuracy and detection ratio in the classification of contamination events (Oliker and Ostfeld, 2014). Finally, a selected range of ensembled decision tree classifiers, using boosting and bagging algorithms were investigated, because both methods, bagging and boosting have demonstrated very successfully improvements on the accuracy of conventional classifiers (Bauer and Kohavi, 1999). The focus here was set on AdaBoost, short for Adaptive Boosting (Freund and Schapire, 1997) and Bootstrap Aggregation (bagging) utilised by Bagged Trees (BT) classifiers (Breiman, 1999). While AdaBoost (in conjunction with decision trees as weak learners) is often referred to one of the most successful techniques in large-scale classification (Chapelle, Chang and Liu, 2011) (Kégl, 2013). BT classifiers were also chosen, because the bagged decision trees as class probability estimation model often has shown to outperform other one-class classifier variants by a significant margin (Hempstalk, Frank and Witten, 2008).

Similar to CUSUM fault detection, all above mentioned classification methods are data driven, learning relevant relations from a dataset of observed signals that contains pre-labelled events aiming to classify the condition of WTWs processes as normal or faulty, respectively to predict the presence of an event. For reliable predictions of the process states, suitable relations across candidate signals, i.e. multiple signals needed to be analysed by the classifier. To achieve this, CUSUM's binary output (normal or faulty signal condition) of the analysed signals served as input dataset for the training of the respective classification procedure. This classification process results in the triggering of alarms if an abnormal condition at WTW's processes respectively the presence of an event is predicted.

All classifiers described in this section were trained on data of the calibration time period using the CUSUM detection output of the 22 critical alarm point signals as predictor variables (input) with the labelled historical events as response variable (output). It was assumed that the training database contains a sufficient number of identified process faults and historical events. These requirements were met documented by the results of the case study (see Section 5.4.2), which have shown that 1.293 faults have been detected by CUSUM fault detection method across the used 22 critical alarm points and 102 historical events identified during calibration time period (i.e. training period for classifiers). After training, the classifier models were tested on the unseen data of validation time period and their respective performance assessed by quantifying detection statistics on observed, historical data and events in the same way as it was done in the previous experiments.

First, the capability of an Artificial Neural Network (ANN) for the desired classification procedure was investigated. ANNs can be trained to solve problems that are too difficult to address with a simple algorithm and have been already successfully applied to classify data for fault detection and other fields (Lennox et al., 2001) (Dong, Cheng and Chan, 2009) (Chen and Huang, 2011). Particularly feedforward neural networks (the information moves only forward from input to output nodes) can be used to construct several types of classifiers (Huang, Chen and Babri, 2000). Therefore, a feedforward neural network was implemented to perform the mentioned classification task. The performance of any ANN is heavily dependent on its architecture that means it depends on the way of how the used computing elements are connected and on the strengths of these connections (weights). The weights are automatically adjusted by training the network according to a defined learning rule. To perform the classification task a two-layer ANN was built that is fed by the predictor variables (CUSUM output) as input which is transferred via a hidden layer by a sigmoid activation function into the output layer, that makes use of a softmax function to represent its prediction as probability distribution over the different classes (Goodfellow, Bengio and Courville, 2016).

A two-layer feedforward ANN utilising above described transfer functions and 100 neurons (see Appendix A) between input and output layer was trained to identify particular patterns and classify them into one of the two classes: (i) normal condition (i.e. no event), and (ii) event taking place. To adjust weights properly, the Scaled Conjugate Gradient (SCG) backpropagation algorithm for non-linear optimization, developed by Moller was applied by training the ANN. Training the ANN on pattern recognition, the SCG is more effective and faster than most other learning algorithms, such as standard back propagation (Rumelhart, Hinton and Williams, 1986) or conjugate gradient (Johansson, Dowla and Goodman, 1991) algorithm (Moller, 1993). The trained ANN was then applied on the validation data set to predict for each observation the probability of an event. An alarm was raised if the predicted probability exceeded a threshold value of 0.5.

In the second experiment the SVM classifier was explored. SVM is a well-established classification method, popular for its high accuracy and ability to deal with high-dimensional data (Oliker and Ostfeld, 2014). The training data set is used to construct hyperplanes which separates a higher dimensional space into two classes. (Boser, Guyon and Vapnik, 1992). In this work several standard kernels were tested including the radial basis function (RBF) kernel, but also nonlinear polynomial and sigmoidal kernels (Schölkopf, Smola and Bach, 2002). Internal parameters such as kernel coefficient (γ) and regularisation constant (C) were automatically learnt during training using cross validation (Kohavi, 1995) that additionally prevents overfitting. SVM classifiers were trained on the above described dataset by maximising the separating area between the two defined classes and minimising the error of misclassified vector using kernel-based learning algorithms (Press et al., 2007). The SVM using a radial basis function (RBF) kernel with a γ of 3.1 and C of 1 was tested on the unseen data of validation time period.

The third experiment was investigating the classification capabilities of the mentioned ensemble classifiers. Random Forest (RF) classifiers that makes use of ‘bagging’ in tandem with random feature selection growing a combined ensemble of decision trees to let them vote for the most popular class (Breiman, 2001) were tested. Each tree utilised in an ensemble of 100 decision trees (see Appendix A) was trained on the calibration data set individually to generate the decision rules, after which for each observation of the validation time period each tree has generated its vote for the estimated class (event or no event). The proportion (non-weighted average) of votes from all trees in the ensemble in favour of a class represents the estimated probability of the class membership. Finally, an alarm is risen if the estimated probability of an event is above threshold value of 0.5. The same procedure was followed for exploring the event detection capabilities of AdaBoost classifiers. 10 decision trees in ensemble (see Appendix A) have been utilised by the adaptive boosting technique. Unlike RF, the AdaBoost algorithm apply weights to represent the estimated class membership for each observation as weighted sum of the predictions made by each individual tree in the forest. An alarm was raised if the predicted classification value exceeded the threshold of zero. When evaluating each approach, speed and detection performance were considered. Among all tested classification methods RF classifier has shown best performance (see Section 5.4.3).

Additional experiments were conducted on the classifier models to investigate possible improvements on detection performance by (i) optimisation of the classifier input data and (ii) feature selection to explore the importance of single signals aiming to remove redundant signals. The optimisation of the classifier input was conducted by a further refinement of the CUSUM detection. The output of CUSUM detection system was revised by utilising same windows sizes (1d and 1 week) and control limits (1σ, 3σ, 6σ, 12σ) as described in Section 4.4.2, but by the application of modified target and k values to the CUSUM detection system. Instead of using mean values and standard deviation of the mean, median and Median Absolute Deviation (MAD) was used as target and k value. Parameters used for the CUSUM refinement are shown in Table 4-4.

Table 4-4 *Parameters used for CUSUM refinement.*

Parameter	window size (1d)	window size (1 week)
k	1MAD,2MAD…9MAD	1MAD,2MAD…9MAD
ULC/LCL	±1σ/3σ/6σ/12σ	±1σ/3σ/6σ/12σ

A type of sensitivity analysis that makes use of the modified parameters was conducted in the same way as described in section 4.4.2. Optimised control limit and k value combinations for each sensor signal as well as the optimal window size were derived by evaluating the performance of the HC-ERS for each criterion and each signal separately. The following 3 criteria were applied the performance evaluation:

1. Maximum value of the ratio of true positives rates to false positives (see Figure 4-4 in Section 4.5)
2. Maximum regression value (linear regression between CUSUM output and historical events)
3. Maximum number of true detections (sum of true positives and true negatives)

Best performing CUSUM parameter settings of each individual signal were then derived for each criterion by the analysis of their CUSUM detection output data against the above criteria. After recalculation of the number of neurons for the ANN and decision tress for RF and AdaBoost classifiers (see Appendix B) the optimised output of each individual signal derived this way was then feed as input into above classifier models followed by the calculation of corresponding detection statistics. The results of these experiments have shown that best performance was achieved by the application of the CUSUM parameters derived by criterion 3 (see Appendix B) and therefore, these parameters were used as default values in the CUSUM fault detection system for each signal. The results have also demonstrated HC-ERS in combination of refined CUSUM with RF-based classification (criterion 3) as best performing method as yet (see section 5.4.3). For this reason, RF classifiers has been selected to set on further analysis.

On basis of the system developed so far, additional improvements on the classification process were investigated. It is well known that RF performs well in presence of a small number of informative predictors among a greater number of noisy variables, it is impossible to distinguish the contributions of single predictors to the outcome of the RF classification process (Sandri and Zuccolotto, 2009). Therefore, further analysis was conducted by the optimisation of feature selection for the classification procedure aiming to remove redundant signals or signals that possibly adversely affect the performance of the classification process. Stepwise backward elimination using the wrapper method similar to the approach of Kohavi and John (Kohavi and John, 1997) was utilised to identify these redundant signals. Stepwise backward elimination comprises the step-by-step rejection of insignificant predictors from a model until all variables are significant or eliminated. Once all input variables have been eliminated, they can be ranked according to their order of importance, i.e. significance for the model.

This approach was conducted starting by utilising all 22 CUSUM output vectors as input for the RF classifier model. Sequentially one of each of the 22 signals was rejected followed by the evaluation of corresponding model's performance on unseen data of validation time period and the re-training of the classifier. The performance of the model was assessed by comparing the stepwise calculated ratio of true positives to false positives (see Figure 4-6 in Section 4.5). The signal that showed the least influence on the performance was considered as insignificant for the model and therefore permanently eliminated for the following test. The testing was then repeated with the remaining 21 signals in the same way as it was done with 22 signals. This procedure was repeated until all significant signals were eliminated.

The number of significant signals was identified by visual inspection of the graphed results. After each permanent elimination of a signal the performance of the model (TP/FP ratio) was graphed against the remaining number of signals (see Figure 5-11 in Section 5.4.3). By analysis of the graph the optimal number of signals was derived, respectively these signals itself were identified, which contribute significantly to model's performance by selecting the model that has pictured the best overall detection performance, i.e. the best TP/FP ratio. Finally, the definitive model was built with the identified signals and its detection performance was evaluated in the same way as it was done for the previous tests.

4.4.4 Section Summary

This section has focussed on the presentation of the methodologies used for development of the novel Hybrid CUSUM ERS for WTWs processes based on their two process stages, fault detection and event classification.

For the fault detection approach, a number of SPC strategies has been presented and CUSUM based fault detection identified as best performing and most suitable method. The CUSUM based fault detection utilises control charts applied to each observed signal individually and detects their out of control conditions. The output of CUSUM fault detection serves as input for the second event classification stage. For the evaluation of the most appropriate event classification methodology, a collection of machine learning techniques has been applied to point out Random Forest classifier as best performing classification method for predicting the two response classes event or no event. In the following section methodologies and metrics applied for performance assessment of ERSs are outlined.

4.5 ERS Performance Assessment

Performance assessment was conducted by simulating the ERSs using the historical time series data (5 min intervals) of the 22 water quality signals with labelled events contained in the datasets (see Section 3.2). The performance of developed ERS methodologies was demonstrated on unseen data of the validation time period (see Chapter 5). To avoid double or multiple counting of sequenced alarms in the course of ERSs' performance evaluation follow-on alarms raised by individual signals in rapid succession were supressed over a defined number of subsequent time steps.

The time period used for this suppression, in the following referred as to suppression time, had to be investigated first. For this reason, a distribution analysis was conducted investigating the distribution of event durations and timely intervals between single events (see Figure 4-5). Figure 4-5(a) demonstrated that the duration for the largest fraction of events is less than 24 hrs (first bar on the left hand side of the plot). Figure 4-5(b) illustrates that only for a small fraction of events (i.e., 13 minor events out of a total of 163 events) is shorter than 1d. Therefore, the period of one day (i.e. 288 time steps) was selected as most suitable suppression time to (a) minimise the risk that real events will not be detected (suppressed) and (b) reduce the number of multiple counts of subsequent alarms in rapid succession. The suppression time of 1d has been uniformly applied throughout the experiments conducted in this work.

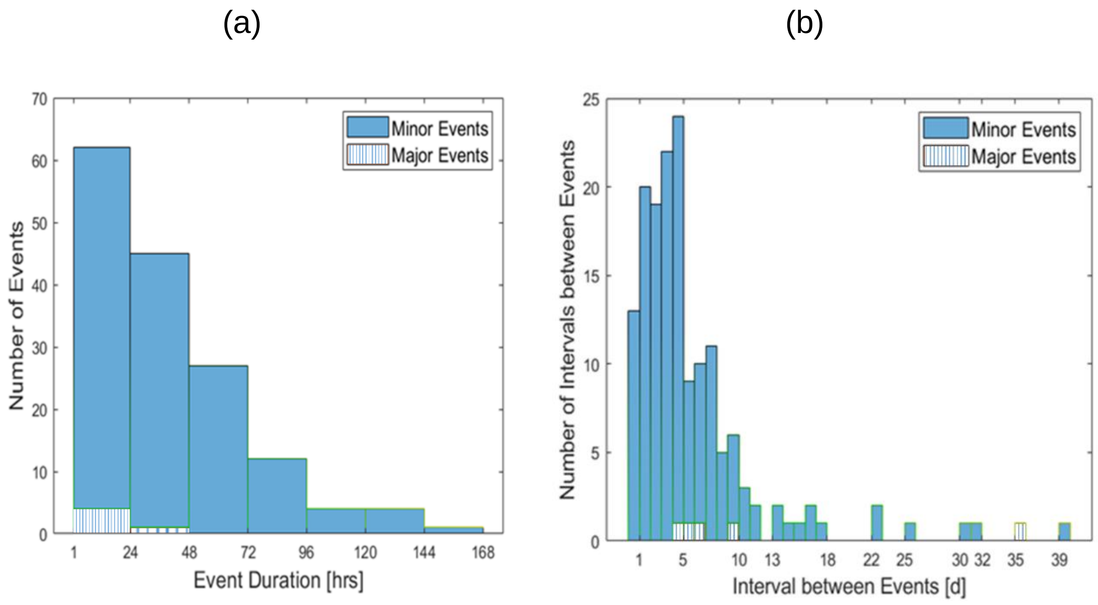

Figure 4-5 *Analysis investigating the distribution of (a) Duration of events and (b) timely interval between events.*

Once the ERS' were simulated, two-by-two confusion matrices with true/false positives/negatives, representing the distribution of possible outcomes for each signal, were generated and the corresponding true detection and false alarm rates calculated. The utilised confusion matrices scheme is shown in Figure 4-6 followed by Figure 4-7 presenting the key metrics for the performance assessment and the formulas used for calculation of the corresponding parameters.

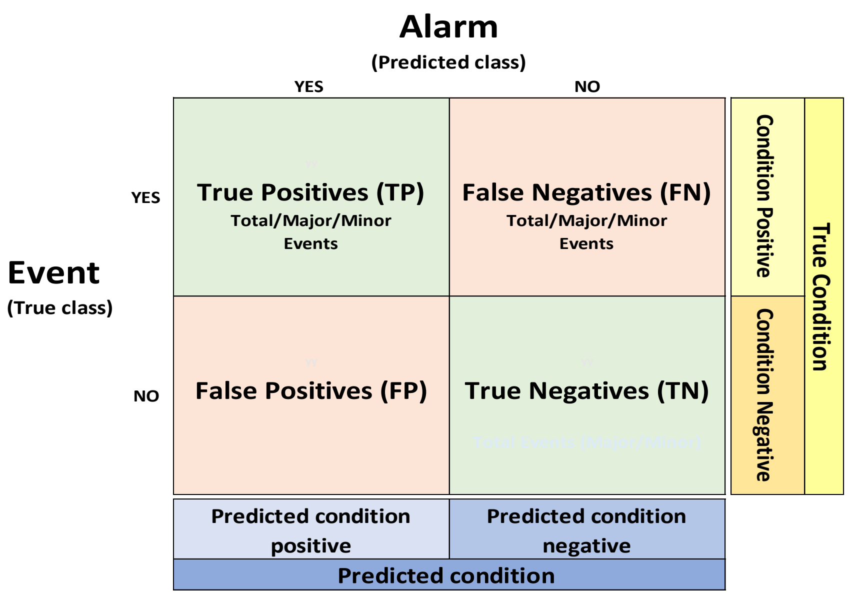

Figure 4-6 *Confusion Matrix*.

Performance Metrics

True Detections				False Detections		
True Positive Rates (TPR)			Positive predictive value (PPV)	False discovery rate (FDR)	False Positives (FP)	False Negative Rate (FNR)
Total	Major	Minor				

Figure 4-7 *Performance Metrics.*

It has to be noted that instead of the more common False Positive Rate (FPR) the False Discovery Rate (FDR) was used to display the rate of incorrect alarms raised by the detection system. The false discovery rate (hereafter also referred to as false alarm rate) is a method of conceptualising the rate of type I errors in null hypothesis testing when conducting multiple comparisons (Benjamini and Hochberg, 1995).

The derived detection statistics contain the True Positive Rate (TPR), also referred to as recall or sensitivity calculated for total events (sum of major and minor) on the one hand and for major and minor events separately on the other. Additionally, the Positive Predictive Value (PPV) also called precision is shown in the assessment of the detection performance. Both, TPR and PPV describe the true detection capabilities of the system. Above described false discovery rate, the absolute number of false positives and finally the False Negative Rate (FNR) all shown in the detection performance tables (see Figure 4-7) refer to false detections and therefore they are a suitable measure of performance against faulty predictions of the system. Formulas used for the calculation of the detection metrics are as follows:

$$TPR = \frac{\sum True\ positive}{\sum Condition\ positive} = \frac{TP}{TP+FN} \tag{1}$$

$$PPV = \frac{\sum True\ positive}{\sum Predicted\ condition\ positive} = \frac{TP}{TP+FP} \tag{2}$$

$$FDR = \frac{\sum False\ positive}{\sum Predicted\ condition\ positive} = \frac{FP}{TP+FP} \tag{3}$$

$$FP = \Sigma\ False\ positive \tag{4}$$

$$FNR = \frac{\sum False\ negative}{\sum Condition\ positive} = \frac{FN}{TP+FN} \qquad (5)$$

In addition to above detection metrics for the comprehensive ERS's performance evaluation the harmonic mean of precision and recall, i.e. the F measure also referred as to F_1 score used in literature for the comparison of event detection methods (Inoue et al., 2017) and false alarms per week are calculated as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \qquad (6)$$

$$False\ alarms\ per\ week = \frac{\sum False\ positive}{Number\ of\ weeks} \qquad (7)$$

Finally, the above described detection rates of individual signals corresponding to a single treatment stage were then averaged to display detection statistics of the respective treatment stage. In the same way detection statistics were calculated for the overall system as averaged detection rates of all analysed signals. All this was done for calibration and validation time periods separately.

Evaluating the performance of E-ERS this way, the baseline used for the evaluation of improvements was established by quantifying its detection statistics on observed, historical data and events. The same assessment procedure was applied for the evaluation of M-ERS's and HC-ERS's detection performances. The results obtained this way were then compared to assess the improvements of M-ERSs and HC-ERS against the E-ERS.

It has to be noted, that in contrast to the E-ERS and M-ERS, for the evaluation of HC-ERS's performance only a single confusion matrix with true/false positives/negatives is generated followed by the calculation of corresponding true detection and false alarm rates, since the output of its classification process based on the multivariate CUSUM predictor input serves for each time step only a single prediction for the condition (normal state or event) of the overall system.

For the further assessment of the new Hybrid CUSUM ERS methodology its detection results were additionally compared to the detection results of the well-established CANARY event detection software from U.S. Environmental

Protection Agency (EPA). CANARY was tested to serve as benchmark for the Hybrid CUSUM ERS. The simulation of CANARY was carried out by the use of default, respectively recommended parameter settings suggested by the software documentation (detailed results are shown in section 5.4.5).

4.6 Implementation of HC-ERS Methodologies

For the implementation of the methodologies used by the HC-ERS, coding was performed using the MATLAB environment and functions provided by MATLAB. In order to be able to integrate the sensor data validation and pre-processing method (see Section 4.3.2) with the fault and event detection methodologies of the HC-ERS (see Section 4.4.2 and 4.4.3) into a single software code, a modular approach was chosen for this purpose. Therefore, the final MATLAB code developed consists of three main sections for: (i) sensor data validation and pre-processing, (ii) CUSUM fault detection, and (iii) Random Forest event detection procedures.

In the course of this work, several types of codes have been developed. On the one hand as offline applications for calibration, training and testing the methods, and on the other hand for validation of the applied methodologies on unseen data. The latter is applicable for "online" implementation of HC-ERS's methodologies and allows event recognition at WTWs in near real time. The “offline” codes are used to determine the parameters required to perform the CUSUM fault detection algorithm, i.e. size of the sliding widow, k and upper/lower control limits and the number of decision trees used by the RF classifier for event detection. For this purpose, multiple sensitivity tests were performed by the code to investigate the best possible parameter or parameter combinations for the respective application.

The parameters obtained this way and the decision rules achieved from the trained classifier were used as default in the online code, which works as follows.

After the initialisation of the required data (sensor readings from the observed signals) the online code performs the sensor data validation and pre-processing routine. The statistical tests described in Section 4.3.2 were conducted for each signal individually and each observation according to the following scheme.

1) Checking for erroneous data and if applicable replacement with the preceding validated sensor measurement.

2) Identification of missing data and if applicable replacement with the preceding validated sensor measurement.

3) Identification of unusual spikes if applicable replacement with the preceding validated sensor measurement.

4) Identification of flat signals and if applicable labelling as “flat line fault”.

The purpose of the subsequent fault detection section is to identify abnormal (“out of control”) conditions in the analysed signals. The code uses the default values in combination with the CUSUM function, which is applied to each observed water quality signal individually to detect their possible out of control conditions. This code section executes the following procedure.

1) Within the applied historical window, the upper and lower cumulative sums of deviations are calculated for each water quality signal separately.

2) For each signal and observation, a vector is created, filled with ones if upper/lower or both cumulative sums of deviations violated the applied control limit at the present data point (sample) and otherwise with zeros if the signal is “in control”.

3) In the last step the code generates the CUSUM output matrix using for each observation the CUSUM output of individual signals. The CUSUM output matrix serves as input for the following event classification procedure.

The task of the code in the final event detection section is to provide the prediction of an event occurring at WTW’s processes. The code uses the default number of trees and default decision rules provided from the “offline” code to calculate the probability of an event for each observation from the unseen data of the validation period. Finally, if the probability exceeds a default threshold value of 0.5 an alarm is risen.

4.7 Summary

Online data of water quality parameters from sensors deployed in a sufficient number to monitor various processes at WTWs provide the necessary information for a reliable and robust event recognition in near real-time. The novel Hybrid CUSUM event recognition system that makes use of these data has been presented in this chapter.

In this concluding section the methodologies and developments outlined in this chapter can be summarised as follows. After a brief introduction, the assessment of the E-ERS has been described, detailing its architecture as currently used in the real-live WTW. The chapter continues with the presentation of strategies used to achieve improvements on the performance of the E-ERS discussing the sensitivity analysis applied to investigate optimised thresholds and persistence values for each individual signal and the developed sensor data validation and pre-processing methods that make use of four simple statistical tests to improve the data quality in near real-time used for M-ERS. Furthermore, details have been given about the fundamentals of the HC-ERS and the implementation of methodologies used by the system. Procedures applied for the optimisation of the Hybrid CUSUM ERS have been described as well as the methods of how its event detection performance has been evaluated and assessed against E-ERS, M-ERS and CANARY detection systems.

Chapter 5: Case Study Results and Discussion

5 CASE STUDY RESULTS AND DISCUSSION

5.1 Introduction

This chapter presents the results of application of the E-ERS's, M-ERS's and HC-ERS's detection methods on the WTW case study. For all cases presented in this chapter the dataset including water quality signals from sensors deployed at the above test site (see Section 3.1) was used with the labelled events as described in Section 3.3.

The first case performed on the E-ERS is aimed to investigate, test and illustrate the capabilities of the existing detection system to establish a baseline for the performance of ERSs. The subsequent studies on M-ERS and HC-ERS were conducted aiming to test and demonstrate the event detection capabilities of new methodologies applied to M-ERS and HC-ERS and outline improvements achieved against the currently used E-ERS.

This chapter is organised as follows. After this introduction, Section 5.2 details the results of E-ERS's detection performance evaluation. The results achieved for the E-ERS serve as baseline for the assessment of possible improvements by the application of new methodologies to M-ERS and HC-ERS. Section 5.3 provides the results of M-ERS's performance evaluation and improvements achieved utilising the type of sensitivity analysis presented in Section 4.3.1 aiming to optimise the threshold and persistence values applied by M-ERS and continues with the illustration of the outcome and benefits accomplished using validated and Pre-processed Sensor Data (PSD) applied to the M-ERS - denoted as M-ERS (PSD) here - by utilising the methods described in Section 4.3.2. The results of the developed CUSUM based fault detection (see Section 4.4.2) and event detection (see Section 4.4.3) methodologies using Random Forest (RF) classifiers applied by the novel HC-ERS (see Section 4.4) including its performance evaluation and improvements achieved are presented in Section 5.4. These results are outlined for the different development stages shown in Sections 5.4.2, 5.4.3. and 5.4.4. Finally, the detection performance of the HC-ERS is compared to E-ERS, M-ERS, M-ERS (PSD) and the well know CANARY method shown in Section 5.4.5 followed by a concluding summary in Section 5.5.

5.2 E-ERS Results and Discussion

This first case investigates the fault detection capabilities demonstrated on data from the WTW described in Section 3.1 and presents performance evaluation of E-ERS. The case demonstrates the application of the dataset described in Section 3.2 with labelled major and minor events (see Section 3.3) to the simulated detection system (E-ERS) of the test site. The analyses aimed to evaluate the performance of E-ERS according to the methods described at Section 4.5 focussing on the capabilities to correctly detect the labelled minor and major events in conjunction with corresponding false alarm rates generated by the system.

The results of the analysis serve as baseline for the assessment of possible improvements achieved by M-ERS and HC-ERS. Once the E-ERS was calibrated using the same threshold and persistence values as currently utilised at the test site. The default parameters used for this case study are shown in Table 5.1.

Table 5-1 *E-ERS: Default parameters.*

Signal	Unit	Low Limit	High Limit	Persistence [Time Step = 5min]
Raw Water Turbidity	NTU	-	10.00	0
Raw Water pH	pH	5.50	7.90	1
Pre Flocculation pH Stream A	pH	4.0	4.80	0
Pre Flocculation pH Stream B	pH	4.0	4.80	0
Post Flotation Turbidity Stream A	NTU	0.01	6.50	1
Post Flotation Turbidity Stream B	NTU	0.01	6.50	1
DAF Iron Stream A	mg/l	-	2.50	6
DAF Iron Stream B	mg/l	-	2.50	6
Pre 1st Stage pH Stream A	pH	5.80	7.50	2
Pre 1st Stage pH Stream B	pH	5.80	7.50	2
Post 1st Stage Turbidity Stream A	NTU	-	0.50	2
Post 1st Stage Turbidity Stream B	NTU	-	0.50	1
Pre 2nd Stage pH Stream A	pH	6.80	8.60	1
Pre 2nd Stage pH Stream B	pH	6.80	8.60	2
Post 2nd Stage Turbidity Stream A	NTU	-	0.40	3
Post 2nd Stage Turbidity Stream B	NTU	-	0.25	3
Post 2nd Stage Chlorine Stream A	mg/l	0.60	1.40	1
Post 2nd Stage Chlorine Stream B	mg/l	0.60	1.40	1
Treated Water pH Stream A	pH	6.80	8.60	0
Treated Water pH Stream B	pH	6.80	8.60	0
Final Water pH	pH	7.00	9.00	1
Final Water Chlorine Residual	mg/l	0.60	1.35	0

For this analysis the observed data of the entire time period was split into datasets for re-calibration of existing detection thresholds (time period from 01/01/2012 until 28/02/2014) and follow-on validation on unseen data (time period from 01/03/2014 until 01/03/2015). The results obtained by testing the E-ERS on unseen data (i.e. validation dataset) are provided in this section.

According to the formulas shown in Section 4.5 detection rates corresponding to the generated confusion matrices containing True Positive Rates (TPR), Positive Predictive Value (PPV), False Positive Rate (FPR) and False Discovery Rate (FDR) were calculated for each signal individually. These metrics are summarised in Table 5-2. It has to be noted that multiple detections of events are possible if an event last longer than the default suppression time (1 day here). Whilst multiple counts of TPs for these events have been neglected for the calculation of true positives, i.e. for the detection of a single event only a single TP was counted, multiple detections have been considered for the calculation of PPVs and FDRs. This way an adequate calculation of the detection metrics was achieved according to which (a) the number of possible TPs corresponds to the number of real events and (b) the number of alarms used in the metrics calculation corresponds to the number of total alarms raised by the ERS.

Table 5-2 *E-ERS: Detection statistics for individual signals.*

E-ERS (22 critical alarm points)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Raw Water Turbidity	3%	0%	3%	50%	50%	2	97%
Raw Water pH	5%	0%	5%	57%	43%	3	95%
Pre Flocculation pH Stream A	34%	100%	33%	58%	42%	18	66%
Pre Flocculation pH Stream B	33%	100%	32%	68%	32%	12	67%
Post Flotation Turbidity Stream A	21%	100%	20%	89%	11%	2	79%
Post Flotation Turbidity Stream B	8%	0%	8%	64%	36%	4	92%
DAF Iron Stream A	25%	100%	23%	49%	51%	19	75%
DAF Iron Stream B	21%	100%	20%	54%	46%	16	79%
Pre 1st Stage pH Stream A	28%	100%	27%	78%	22%	5	72%
Pre 1st Stage pH Stream B	13%	100%	12%	90%	10%	1	87%
Post 1st Stage Turbidity Stream A	25%	100%	23%	56%	44%	16	75%
Post 1st Stage Turbidity Stream B	34%	0%	35%	53%	47%	31	66%
Pre 2nd Stage pH Stream A	67%	100%	67%	53%	47%	51	33%
Pre 2nd Stage pH Stream B	44%	0%	45%	56%	44%	28	56%
Post 2nd Stage Turbidity Stream A	3%	100%	2%	40%	60%	3	97%
Post 2nd Stage Turbidity Stream B	8%	100%	7%	100%	0%	0	92%
Post 2nd Stage Chlorine Stream A	16%	100%	15%	80%	20%	3	84%
Post 2nd Stage Chlorine Stream B	20%	100%	18%	56%	44%	11	80%
Treated Water pH Stream A	8%	0%	8%	43%	57%	8	92%
Treated Water pH Stream B	44%	0%	45%	31%	69%	115	56%
Final Water pH	7%	0%	7%	63%	38%	3	93%
Final Water Chlorine Residual	10%	100%	8%	75%	25%	3	90%

The following can be observed from Table 5-2: (a) the E-ERS has generated low TPRs between 3% and 45% for almost all signals, except for stream A pre 2nd stage pH signal where a significantly higher TPR of 67% was produced, which indicates that signals seem to react differently to process changes based on the pre-determined thresholds and persistence values used by the E-ERS, (b) the same effect applies to the FDRs illustrated by widely varying values between 0% to 69% which confirms the different sensitivity of signals to process changes, (c) the number of 115 false positives shown in the table for the treated water pH stream B signal is by far the largest value among the signals indicating that threshold and persistence values on this signal are not optimally set.

The detection statistics for the overall E-ERS calculated by averaging the detection rates and summation of false positives over all observed signals is shown in Table 5-3.

Table 5-3 *E-ERS: Averaged overall detection statistics.*

E-ERS	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System	**22%**	**64%**	**21%**	**62%**	**38%**	**354**	**78%**

As it can be seen from Table 5-3, the E-ERS is able to detect only 22% of total events, split into TPRs of 64% for major and 21% for minor events respectively. The significant higher true detection rate for major events was expected since these events are easier to detect than the minor ones. Although a large number of events were not detected as pictured in the table by the FNR of 78%, the E-ERS generates a considerable high number of false alarms demonstrated by the FDR of 38% and the high number of 354 FPs produced within the tested one year validation time period.

The assessment of E-ERS's performance has resulted in estimated 6.8 false alarms per week (derived as ratio of 354 FP alarms and the 52 weeks). This additional measure is an important information regarding practical WTW's operation, since an ERS that generate seven or more false alarms (invalid alerts) per week is considered as system of limited practical relevance (EPA, 2013) (s::can, 2013). With 6.8 false alarms per week the E-ERS is only just below this critical value. Given this and the low true detection rate of 22% it can be concluded that the E-ERS in not suited well for the detection of water quality events at WTWs.

Whilst the results displayed in Table 5-3 already give a good overview about the detection capabilities of the E-ERS, the F_1 score (see Section 4.5) was estimated here as well. This score is the harmonic mean of precision (PPV) and recall (TPR) and is widely considered as a suitable measure for the ERS performance assessment and the comparison of different detection systems. The F_1 score ranges between 0 and 1 (as higher the F_1 score as better the system performs), an ideal F_1 score would achieve a value of 1. The calculated F_1 score for the E-ERS is only 0.31 further confirming a rather poor detection performance of this method.

The evaluation of the E-ERS's detection performance was complemented by calculating the detection matrices for each treatment stage by averaging the detection statistics of single signals across the corresponding treatment stage. The corresponding metrics of the detection statistics for the single treatment stages are presented in Table 5-4.

Table 5-4 *E-ERS: Detection statistics corresponding to single treatment stages.*

E-ERS (WTW's Treatment Stages)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Inlet	19%	50%	18%	58%	42%	9	81%
Flotation/Flocculation	19%	75%	18%	64%	36%	10	81%
1st Stage Filtering	25%	75%	24%	69%	31%	13	75%
2nd Stage Filtering	27%	83%	26%	64%	36%	10	73%
Outlet	17%	25%	17%	53%	47%	32	83%

The results presented in Table 5-4 show that the E-ERS enables the true detection of 19% to 25% of the events already at WTW's early treatment stages (WTW's inlet to 1st stage filtering). The TPR increases only slightly (by 1%) at the 2nd stage filtration, which indicates that most of the detected events seems to be manifesting at the early treatment stages and then propagating throughout the subsequent treatment stages.

This observation was confirmed by further analysis of signal's behaviour for minor events. Selected examples presented in Figure 5-1 illustrate deviations of signals from normal condition in the presence of minor events and shows the propagation of these faults to signals of subsequent treatment stages.

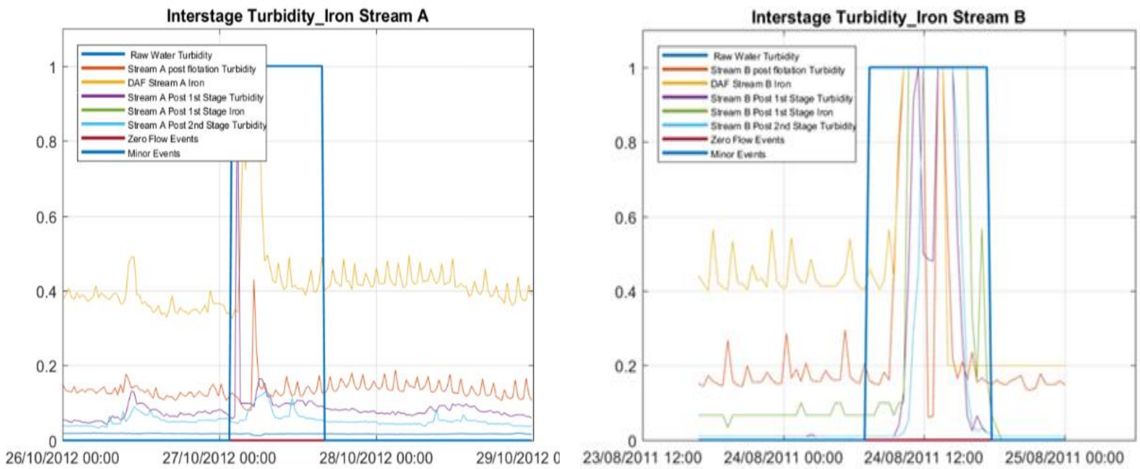

Figure 5-1 *Example minor events – Propagation of turbidity faults to subsequent WTWs stages.*

From Figure 5-1 it can be seen that the water quality signals shown are starting to deviate from the normal values at the WTW's inlet stage (see raw water turbidity signal) and this is then propagated further downstream, to the following treatment stages (see other signals). This propagation effect correlates with the observation that TP rates generated by the E-ERS from early treatment stages to downstream 2nd stage filtering do not differ much (i.e. 19% for the inlet stage, 25% for flocculation/flotation stage and 27% for the 1st and 2nd filtration stages).

The fact that a significant high number of events influence water quality signals' behaviour already at the early treatment stages is beneficial for an early detection of these events. At outlet stage the lowest TPR (17%) combined with the highest number of 32 false positives over all treatment stages indicate that default limits set on signals at the outlet stage have not been selected in an optimal manner.

This observation is confirmed by analysing the confusion matrices of single signals at this stage shown in Figure 5-2 which presents the confusion matrices (explained in Section 4.5, see Figure 4-6) generated for each of the 22 analysed signals. In addition to the numbers of true positives and false negatives corresponding to all events (major and minor), the resulting confusion matrices show true positives and false negatives for major and minor events separately (values displayed in brackets, whereas first numbers shows the true/false detections for major and second the numbers for minor events).

Event (True Class)	Alarm (Predicted class) YES	NO
YES	2 (0/2)	59 (1/58)
NO	2	63

Raw Water TU **(a)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	3 (0/3)	58 (1/57)
NO	3	64

Raw Water pH **(b)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	21 (1/20)	40 (0/40)
NO	18	79

Pre Flocculation pH (A) **(c)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	20 (1/19)	41 (0/41)
NO	12	73

Pre Flocculation pH (B) **(d)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	13 (1/12)	48 (0/48)
NO	2	63

Post Flotation TU (A) **(e)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	5 (0/5)	56 (1/55)
NO	4	65

Post Flotation TU (B) **(f)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	15 (1/14)	46 (0/46)
NO	19	80

DAF Iron (A) **(g)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	13 (1/12)	48 (0/48)
NO	16	77

DAF Iron (B) **(h)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	17 (1/16)	44 (0/44)
NO	5	66

Pre 1st Stage pH (A) **(i)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	8 (1/7)	53 (0/53)
NO	1	62

Pre 1st Stage pH (B) **(j)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	15 (1/14)	46 (0/46)
NO	16	77

Post 1st Stage TU (A) **(k)**

Event (True Class)	Alarm (Predicted class) YES	NO
YES	21 (0/21)	40 (1/39)
NO	31	92

Post 1st Stage TU (B) **(l)**

Alarm (Predicted class)

Event (True Class)	YES	NO
YES	41 (1/40)	20 (0/20)
NO	51	112

Pre 2nd Stage pH (A) **(m)**

Event (True Class)	YES	NO
YES	27 (0/27)	34 (1/33)
NO	28	89

Pre 2nd Stage pH (B) **(n)**

Event (True Class)	YES	NO
YES	2 (1/1)	59 (0/59)
NO	3	64

Post 2nd Stage TU (A) **(o)**

Event (True Class)	YES	NO
YES	5 (1/4)	56 (0/56)
NO	0	61

Post 2nd Stage TU (B) **(p)**

Event (True Class)	YES	NO
YES	10 (1/9)	51 (0/51)
NO	3	64

Post 2nd Stage Cl (A) **(q)**

Event (True Class)	YES	NO
YES	12 (1/11)	49 (0/49)
NO	11	72

Post 2nd Stage Cl (B) **(r)**

Event (True Class)	YES	NO
YES	5 (0/5)	56 (1/55)
NO	8	69

Treated Water pH (A) **(s)**

Event (True Class)	YES	NO
YES	27 (0/27)	34 (1/33)
NO	115	176

Treated Water pH (B) **(t)**

Event (True Class)	YES	NO
YES	4 (0/4)	57 (1/56)
NO	3	64

Final Water pH **(u)**

Event (True Class)	YES	NO
YES	6 (1/5)	55 (0/55)
NO	3	64

Final Water Cl **(v)**

Figure 5-2 *Confusion matrices for E-ERS generated by the application to data of the observed signals.*

From this figure it can be seen (e.g. Figure 5-2t) that the number of 115 false positives generated by stream B treated water pH signal is significantly higher than the 3 to 8 false positives of the other signals at this treatment stage (e.g. Figure 5-2s, 5-2u and 5-2v). The single signals' confusion matrices illustrate also that some signals are more sensitive in detection of true events than others. Whilst stream A and B pre 2nd stage pH signals (see Figure 5-2m and 5-2n) are able to detect 41 and 27 events respectively, only two and five events are truly identified by post 2nd stage turbidity stream A and B signals (see Figure 5-2o and 5-2p), respectively. The confusion matrices also demonstrate that some of the events are harder to detect than others. Figure 5-3 shows the distribution of events detected simultaneously by a number of signals.

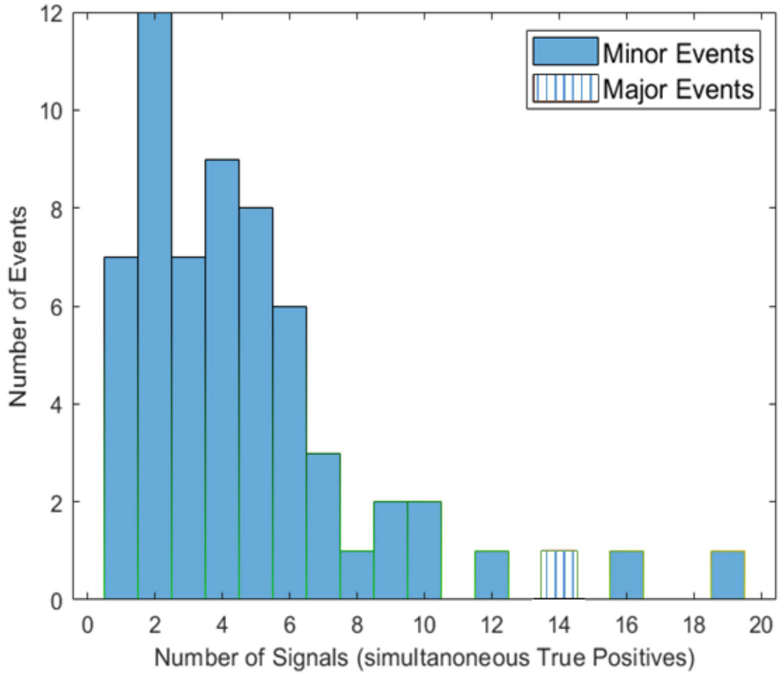

Figure 5-3 *Distribution of events simultaneously detected by multiple signals.*

From this figure it can be seen that at least 7 minor events out of a total of 61 events within the validation time period are harder to detect than the other ones since only one signal is triggering an alarm for these events.

In order to detect more of those events that are difficult to spot, the thresholds used to define the presence of an event could be reduced, resulting in identifying more events at the cost of increasing the number of false alarms. Figure 5-3 Distribution of events simultaneously detected by multiple signals.

also shows that the overwhelming number of events, i.e. 54 of the total 61 events have been identified by two or more (up to 19) signals simultaneously. This high number of simultaneous detections illustrates that investigating signals simultaneously rather than analysing signals individually can be favourably used for the development of an event detection methodology that is more sophisticated and robust against false alarms caused by individual signals.

As already mentioned in Section 2.3.1 the detection success relies on the quality of data processed and adequate validation of input data can be used to improve the performance of any ERS. The E-ERS is lacking such a data validation procedure resulting in a certain amount of false predictions (mainly false alarms) related to sensor faults. These ‘outliers’ caused by erroneous sensor measurements or transmission failures will be counted and treated by the E-ERS as real events. Having said this, the E-ERS makes no distinction between faulty processes and faulty sensors/telemetry measurements. This will possibly influence the detection ability in this way that alarms are triggered by faulty sensor data which is one reason for the high number of false alarms generated by the E-ERS. Figure 5-4 pictures two examples of alarms, both triggered by possible sensor faults that have led to false interpretations by the E-ERS. Whilst the alarm (false positive) shown on the left hand side of the plot illustrates a typical false alarm, i.e. E-ERS predicted an event although no event is present, the alarm (true positive) pictured at the right hand side shows an ‘incidentally’ detection of a true event caused by alarm triggering on faulty sensor data.

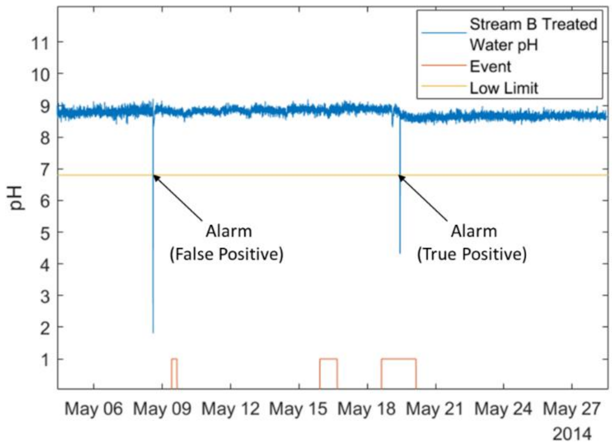

Figure 5-4 *Examples for E-ERS’ false interpretation of triggered alarms.*

Sensor data validation and data pre-processing procedures can be beneficially used for lowering the false alarm rate which is confirmed by the results achieved with the implementation of corresponding methods shown in the following section.

5.3 M-ERS Results and Discussion

This second case explores the event detection capabilities of M-ERS resulting from testing and validating of the methodologies described Sections 4.3.1 and 4.3.2. First, the performance of M-ERS that makes use of optimised threshold and persistence values and second the performance of M-ERS (PSD) applied to validated and pre-processed sensor data was evaluated. Improvements were investigated by the comparison of M-ERS' and M-ERS' (PSD) detection statistics against the statistics obtained by the E-ERS method.

Optimised threshold and persistence values were identified by performing the sensitivity type analysis described in Section 4.3.1. For each signal new thresholds and persistence values were created by gradually changing low and high thresholds for a defined range of persistence values (see Section 4.3.1 for further details). The optimised new thresholds and persistence value combinations were then derived for each sensor signal by selecting the combination with the maximum value of the ratio of true positives to false positives. The new optimised threshold and persistence values obtained this way are presented in Table 5-5.

Table 5-5 *Optimal thresholds/persistence values derived by sensitivity analysis.*

Signal	Unit	Low Limit	High Limit	Persistence [Time Step = 5min]
Raw Water Turbidity	NTU	0.05	14.05	0
Raw Water pH	pH	5.10	7.70	0
Pre Flocculation pH Stream A	pH	4.05	4.80	0
Pre Flocculation pH Stream B	pH	4.00	4.80	0
Post Flotation Turbidity Stream A	NTU	0.01	6.50	1
Post Flotation Turbidity Stream B	NTU	-	5.70	0
DAF Iron Stream A	mg/l	-	3.00	4
DAF Iron Stream B	mg/l	-	2.80	3
Pre 1st Stage pH Stream A	pH	5.80	7.50	2
Pre 1st Stage pH Stream B	pH	5.80	7.50	2
Post 1st Stage Turbidity Stream A	NTU	-	0.45	2
Post 1st Stage Turbidity Stream B	NTU	-	0.50	1
Pre 2nd Stage pH Stream A	pH	6.80	8.60	1
Pre 2nd Stage pH Stream B	pH	7.60	8.60	1
Post 2nd Stage Turbidity Stream A	NTU	-	0.25	3
Post 2nd Stage Turbidity Stream B	NTU	-	0.25	3
Post 2nd Stage Chlorine Stream A	mg/l	0.50	1.30	0
Post 2nd Stage Chlorine Stream B	mg/l	0.60	1.35	0
Treated Water pH Stream A	pH	7.10	8.80	0
Treated Water pH Stream B	pH	7.00	8.65	0
Final Water pH	pH	6.90	9.00	0
Final Water Chlorine Residual	mg/l	0.60	1.35	0

M-ERS testing was carried out in the same way as it was done for the E-ERS but using the optimised thresholds and persistence values given in Table 5-5 instead of the default values utilised by the E-ERS.

The detection metrics of each individual signal shown in Table 5-6 were calculated in the same way as it was done for the E-ERS.

Table 5-6 *M-ERS: Detection statistics of individual signals.*

M-ERS (22 critical alarm points)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Raw Water Turbidity	7%	0%	7%	56%	44%	4	93%
Raw Water pH	5%	0%	5%	57%	43%	3	95%
Pre Flocculation pH Stream A	34%	100%	33%	58%	42%	18	66%
Pre Flocculation pH Stream B	33%	100%	32%	68%	32%	12	67%
Post Flotation Turbidity Stream A	21%	100%	20%	89%	11%	2	79%
Post Flotation Turbidity Stream B	8%	0%	8%	64%	36%	4	92%
DAF Iron Stream A	28%	100%	27%	50%	50%	20	72%
DAF Iron Stream B	33%	100%	32%	56%	44%	21	67%
Pre 1st Stage pH Stream A	28%	100%	27%	78%	22%	5	72%
Pre 1st Stage pH Stream B	13%	100%	12%	90%	10%	1	87%
Post 1st Stage Turbidity Stream A	26%	100%	25%	58%	43%	17	74%
Post 1st Stage Turbidity Stream B	34%	0%	35%	53%	47%	31	66%
Pre 2nd Stage pH Stream A	67%	100%	67%	53%	47%	51	33%
Pre 2nd Stage pH Stream B	53%	100%	52%	59%	41%	31	48%
Post 2nd Stage Turbidity Stream A	5%	100%	3%	50%	50%	4	95%
Post 2nd Stage Turbidity Stream B	8%	100%	7%	100%	0%	0	92%
Post 2nd Stage Chlorine Stream A	18%	100%	17%	81%	19%	3	82%
Post 2nd Stage Chlorine Stream B	21%	100%	20%	60%	40%	10	79%
Treated Water pH Stream A	12%	0%	12%	57%	43%	6	89%
Treated Water pH Stream B	25%	0%	25%	30%	70%	59	75%
Final Water pH	7%	0%	7%	63%	38%	3	93%
Final Water Chlorine Residual	10%	100%	8%	75%	25%	3	90%

In the same way as it was done for the E-ERS, detection statistics for the overall M-ERS presented in Table 5-7 was calculated by averaging the detection rates and summation of false positives over all observed signals.

Table 5-7 *M-ERS: Averaged overall detection statistics.*

M-ERS (22 critical alarm points)	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System	23%	68%	22%	64%	36%	308	77%

As it can be seen from Table 5-7 the M-ERS is able to detect 23% of total events with 68% major and 22% minor events detected by the system. When compared to the E-ERS method only minor improvements on TPRs are achieved, i.e. 4% for major and 1% of minor events, respectively. When comparing the M-ERS's PPV of 64%, FDR of 36% and FNR of 77% the table also shows minor improvemets with increased rates for PPV of 2% and decreased values for FDR and FNR of 2% and 1% respectivey. Although the number of false alarms generated by the M-ERS is reduced by 46 FPs compared to 354 FPs produced by the E-ERS, the total number (308 FPs) is still very high. Similar to the E-ERS, the M-ERS' FNR of 77% (non-detected events) is too still high. The calculated F_1 score of 0.32 for the M-ERS only increased by 0.1 towards the value of 0.31 for the E-ERS. Compared to 6.8 false alarms per week produced by the E-ERS, the value of 5.9 calculated for the M-ERS shows an improvement by 0.9 less false alarms per week.

Altough the M-ERS shows slight improvements on all detection metrics, these improvements started from a low base. In particular the poor true detection rate of 22% for minor events and the high number of 5.9 false alarms per week illustrate the still moderate detection performance of M-ERS. The sigificant higher true detection rate of 68% for major events seems acceptable, but this can be expected for the same reason as already described for the E-ERS (see Section 5.2).

The same picture can be seen when looking at the detection statistics across the individual treatment stages that are presented in Table 5-8.

Table 5-8 *M-ERS: Averaged detection statistics of single treatment stages.*

M-ERS (WTW's Treatment Stages)	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Inlet	20%	50%	19%	60%	40%	9	80%
Flotation/Flocculation	23%	75%	22%	65%	35%	12	77%
1st Stage Filtering	25%	75%	25%	70%	30%	14	75%
2nd Stage Filtering	29%	100%	28%	67%	33%	17	71%
Outlet	13%	25%	13%	56%	44%	18	87%

As it can be seen from this table, when compared to the E- ERS method, an increase of TPRs between 1% to 4% combined with a decrease of FDRs between 1% to 3% is achieved by the M-ERS at inlet, flotation/flocculation and 2nd stage filtration processes. Only at WTW's outlet stage the M-ERS displays a lower TPR (i.e. -4%) than the E-ERS at the same stage, but also the FDR is by 3% lower compared to the FDR generated by the E-ERS at this stage.

Further analysis of single signals' confusion matrices at outlet stage (see Figure 5.2 and 5.5) explains the root cause for this observation.

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	4 (0/4)	57 (1/56)
NO	4	65

Raw Water TU **(a)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	3 (0/3)	58 (1/57)
NO	3	64

Raw Water pH **(b)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	21 (1/20)	40 (0/40)
NO	18	79

Pre Flocculation pH (A) **(c)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	20 (1/19)	41 (0/41)
NO	12	73

Pre Flocculation pH (B) **(d)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	13 (1/12	48 (0/48)
NO	2	63

Post Flotation TU (A) **(e)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	5 (0/5)	56 (1/55)
NO	4	65

Post Flotation TU (B) **(f)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	17 (1/16)	44 (0/44)
NO	20	81

DAF Iron (A) **(g)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	20 (1/19)	41 (0/41)
NO	21	82

DAF Iron (B) **(h)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	17 (1/16)	44 (0/44)
NO	5	66

Pre 1st Stage pH (A) **(i)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	8 (1/7)	53 (0/53)
NO	1	62

Pre 1st Stage pH (B) **(j)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	16 (1/15)	45 (0/45)
NO	17	78

Post 1st Stage TU (A) **(k)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	21 (0/21	40 (1/39)
NO	31	92

Post 1st Stage TU (B) **(l)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	41 (1/40)	20 (0/20)
NO	51	112

Pre 2^{nd} Stage pH (A) **(m)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	32 (1/31)	29 (0/29)
NO	31	92

Pre 2^{nd} Stage pH (B) **(n)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	3 (1/2)	58 (0/58)
NO	4	65

Post 2^{nd} Stage TU (A) **(o)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	5 (1/4)	56 (0/56)
NO	0	61

Post 2^{nd} Stage TU (B) **(p)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	11 (1/10)	50 (0/50)
NO	3	64

Post 2^{nd} Stage Cl (A) **(q)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	13 (1/12)	48 (0/48)
NO	10	71

Post 2^{nd} Stage Cl (B) **(r)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	7 (0/7)	54 (1/53)
NO	6	67

Treated Water pH (A) **(s)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	15 (0/15)	46 (1/45)
NO	59	120

Treated Water pH (B) **(t)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	4 (0/4)	57 (1/56)
NO	3	64

Final Water pH **(u)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	6 (1/5)	55 (0/55)
NO	3	64

Final Water Cl **(v)**

Figure 5-5 *Confusion matrices for M-ERS generated by the application to data of the observed signals.*

Whilst stream A treated water pH, final water and final chlorine residual signals (see Figure 5-5s, 5-5u, 5-5v and Figure 5-2s, 5-2u, 5-2v) show identical or similar numbers between 4 and 7 TPs and between 3 and 8 FPs for E-ERS and M-ERS respectively. In particular, stream B treated water signal (see Figure 5-2t) applied to the E-ERS generates exceptionally high 27 TPs and 115 FPs at this treatment stage. Regarding the significantly higher values for both TPs and FPs it can be concluded that the thresholds applied on stream B treated water signal are set too thigh resulting in a high number of alarms triggered by the E-ERS. This observation is confirmed by analysis of Figure 5-6 that pictures stream B treated water signal with the corresponding upper alarm thresholds applied by E-ERS.

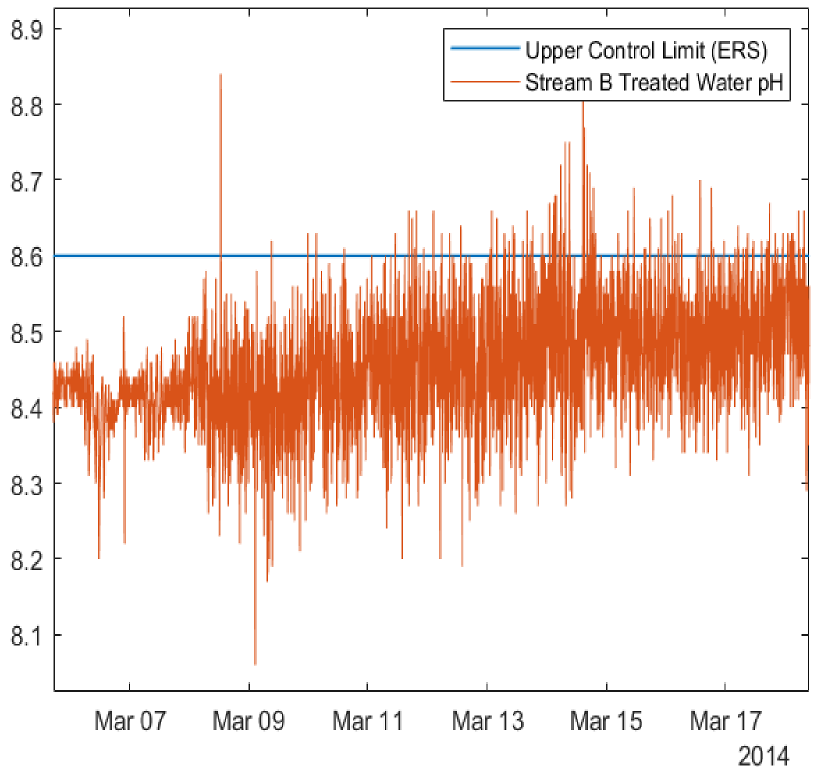

Figure 5-6 *Stream B treated water signal including upper control limit applied by E-ERS.*

From this figure it can easily be seen that stream B treated water signal is generally very noisy and shows a high number of unusual spikes that causes frequently violations of the default thresholds used by E-ERS. Applying a wider upper threshold is beneficial to the reduce the number of the false alarms generated in this case. Although the number of alarms raised by the M-ERS is still very high compared to other signals, the application of the higher upper threshold utilised by the M-ERS significantly improves the number of false alarms produced by the stream B treated water signal (i.e. reduction from 115 to 59) but at the expense of a lower TPR (i.e. reduction from 44% to 25%). This example captures well the general limitation of threshold-based detection systems. The use of wider thresholds to make the system less sensitive results in lower level of false alarms but also to the decrease of true detection rates.

One approach that might at least partially help further reduce the high false alarm rate is an application of sensor data validation and pre-processing. This is done with the aim to avoid alarm triggering on faulty sensor data. For this reason, the M-ERS was tested on validated and pre-processed sensor data. Following the sensor data validation and pre-processing methodologies described in Section 4.3.2, technical faults in sensors or sensor readings, i.e. (i) erroneous data points, (ii) missing data, (iii) spike faults and, (iv) flat line faults were identified and replaced (except flat line faults) before the M-ERS is applied to the data. The M-ERS that makes use of Validated and Pre-processed Sensor Data (PSD) was then tested in the same way as it was done for the E-ERS and M-ERS and the resulting detection statistics calculated. The confusion matrices obtained for each individual signal are shown in Figure 5-7.

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	2 (0/2)	59 (1/58)
NO	3	64

Raw Water TU **(a)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	0 (0/0)	61 (1/60)
NO	0	61

Raw Water pH **(b)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	19 (1/18)	42 (0/42)
NO	15	76

Pre Flocculation pH (A) **(c)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	18 (1/17)	43 (0/43)
NO	9	70

Pre Flocculation pH (B) **(d)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	13 (1/12)	48 (0/48)
NO	2	63

Post Flotation TU (A) **(e)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	5 (0/5)	56 (1/55)
NO	3	64

Post Flotation TU (B) **(f)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	16 (1/15)	45 (0/45)
NO	19	80

DAF Iron (A) **(g)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	18 (1/17)	43 (0/43)
NO	20	81

DAF Iron (B) **(h)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	18 (1/17)	43 (0/43)
NO	4	65

Pre 1st Stage pH (A) **(i)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	9 (1/8)	52 (0/52)
NO	0	61

Pre 1^{st} Stage pH (B) **(j)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	14 (1/13)	47 (0/47)
NO	11	72

Post 1^{st} Stage TU (A) **(k)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	21 (0/21)	40 (1/39)
NO	30	91

Post 1^{st} Stage TU (B) **(l)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	37 (0/37)	24 (1/23)
NO	50	111

Pre 2^{nd} Stage pH (A) **(m)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	32 (1/31)	29 (0/29)
NO	28	89

Pre 2^{nd} Stage pH (B) **(n)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	3 (1/2)	58 (0/58)
NO	3	64

Post 2^{nd} Stage TU (A) **(o)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	5 (1/4)	56 (0/56)
NO	0	61

Post 2^{nd} Stage TU (B) **(p)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	8 (1/7)	53 (0/53)
NO	0	61

Post 2^{nd} Stage Cl (A) **(q)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	10 (1/9)	51 (0/51)
NO	9	70

Post 2^{nd} Stage Cl (B) **(r)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	6 (0/6)	55 (1/54)
NO	6	67

Treated Water pH (A) **(s)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	15 (0/15)	46 (1/45)
NO	59	120

Treated Water pH (B) **(t)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	1 (0/1)	60 (1/59)
NO	0	61

Final Water pH **(u)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	4 (1/3)	57 (0/57)
NO	0	61

Final Water Cl **(v)**

Figure 5-7 *Confusion matrices for M-ERS (PSD) generated by the application to data of the observed signals.*

The corresponding detection metrics shown in Table 5-9 are calculated for each individual signal in the same way as it was done for the E-ERS and M-ERS.

Table 5-9 *M-ERS (PSD): Detection statistics of individual signals.*

M-ERS (PSD) (22 critical alarm points)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Raw Water Turbidity	3%	0%	3%	40%	60%	3	97%
Raw Water pH	0%	0%	0%	0%	0%	0	100%
Pre Flocculation pH Stream A	31%	100%	30%	58%	42%	15	69%
Pre Flocculation pH Stream B	30%	100%	28%	70%	30%	9	71%
Post Flotation Turbidity Stream A	21%	100%	20%	88%	12%	2	79%
Post Flotation Turbidity Stream B	8%	0%	8%	67%	33%	3	92%
DAF Iron Stream A	26%	100%	25%	49%	51%	19	74%
DAF Iron Stream B	30%	100%	28%	54%	47%	20	71%
Pre 1st Stage pH Stream A	30%	100%	28%	82%	18%	4	71%
Pre 1st Stage pH Stream B	15%	100%	13%	100%	0%	0	85%
Post 1st Stage Turbidity Stream A	23%	100%	22%	66%	34%	11	77%
Post 1st Stage Turbidity Stream B	34%	0%	35%	54%	46%	30	66%
Pre 2nd Stage pH Stream A	61%	0%	62%	49%	52%	50	39%
Pre 2nd Stage pH Stream B	53%	100%	52%	61%	39%	28	48%
Post 2nd Stage Turbidity Stream A	5%	100%	3%	50%	50%	3	95%
Post 2nd Stage Turbidity Stream B	8%	100%	7%	100%	0%	0	92%
Post 2nd Stage Chlorine Stream A	13%	100%	12%	100%	0%	0	87%
Post 2nd Stage Chlorine Stream B	16%	100%	15%	55%	45%	9	84%
Treated Water pH Stream A	10%	0%	10%	50%	50%	6	90%
Treated Water pH Stream B	25%	0%	25%	30%	70%	59	75%
Final Water pH	2%	0%	2%	100%	0%	0	98%
Final Water Chlorine Residual	7%	100%	5%	100%	0%	0	93%

The results obtained for the performance evaluation of the overall M-ERS (PSD) by averaging the detection rates and accumulation of false positives across all individual signals are shown in Table 5-10.

Table 5-10 *M-ERS (PSD): Averaged detection statistics of across all signals.*

M-ERS (PSD)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System	20%	64%	20%	68%	32%	271	80%

From Table 5-10 can be seen that the TPR for minor events generated by the M-ERS (PSD) method has slightly fallen to 20% in comparison to the 22% and 23% values obtained for the E-ERS and M-ERS methods, respectively. Whilst M-ERS (PSD) generates the same TPR of 64% for major events as the E-ERS, the M-

ERS has with a TPR of 68% a 4% higher value of true detections for major events. The PPV of 68% for M-ERS (PSD) also shown in the table has been raised against the PPVs of E-ERS and M-ERS by 6% and 4% respectively.

When compared to the E-ERS and M-ERS methods the most notable improvement of the M-ERS (PSD) was achieved on the absolute number of false positives. With 271 false positives against the 354 and 308 generated by the E-ERS and M-ERS methods respectively, the M-ERS (PSD) method was able to reduce the number of false positives by 83 and 37 respectively. The lower number of M-ERS's false positives brought down the false alarms per week from 6.8 for E-ERS and 5.9 for M-ERS to 5.2 false alarms per week generated by the M-ERS (PSD). At the same time the F_1 score of 0.31 calculated for the M-ERS (PSD) decreased by 0.1 only towards the value of M-ERS but is still equal to the value for E-ERS that indicates the M-ERS performs better than M-ERS (PSD).

Digging deeper to explore possible root causes for the lower performance of M-ERS (PDS) against M-ERS, the analysis of the TPRs across the different treatment stages are presented in Table 5-11.

Table 5-11 *M-ERS (PSD): Averaged detection statistics of single treatment stages.*

M-ERS (PSD) (WTW's Treatment Stages)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Inlet	16%	50%	15%	56%	44%	7	84%
Flotation/Flocculation	21%	75%	20%	64%	36%	11	79%
1st Stage Filtering	25%	75%	25%	75%	25%	11	75%
2nd Stage Filtering	26%	83%	25%	69%	31%	6	74%
Outlet	11%	25%	10%	70%	30%	16	89%

As it can be seen from this table, the M-ERS (PSD) has TPRs of 11%-26% of total events at different treatment stages, which represents a reduction in TPRs for inlet, 1st stage filtration, 2nd stage filtration and outlet stages between 2% to 4% compared to the corresponding TPRs of the M-ERS shown in Table 5-8. Only at the 1st filtration stage the M-ERS (PSD) produced the same TPR as M-ERS. The greatest difference in TPRs of 4% is generated at the WTWs inlet stage.

Comparison of the detection statistics of individual signals for M-ERS and M-ERS (PSD) presented in Table 5-6 and Table 5-9 has identified raw water pH signal that is primary responsible for the fallen TPR produced by the M-ERS across this treatment stage. The corresponding confusion matrices of the raw water pH signal shown in Figure 5-5 and Figure 5-7 display values of three and zero true positives for the M-ERS and M-ERS (PSD), respectively. This indicates that the 3 TPs generated by the M-ERS were (presumably) caused by unusual spikes or erroneous sensor data that are coinciding with the genuine events taking place. The identification of the faulty sensor data performed by the M-ERS (PSD) method ensures that no alarms were triggered on the raw water pH signal resulting in no events detected by this method on this signal. This is confirmed by visual inspection of the graphed signal and labelled events, as shown in Figure 5-8. Part (a) of this figure shows the original sensor data for raw water pH for the October to December 2014 time period (which is part of the validation time period). From this figure it can be seen that the measured pH value of the analysed signal drops three times to zero or nearly zero causing alarms by violating its default low limit of 5.1 at same time when minor events are taking place. It is obvious that these alarms are caused by faulty sensor measurements since sudden drops of pH from 7.5 to 0 are unrealistic in real-life WTW's operation. Therefore, the true positives generated by the M-ERS are most likely caused by the sensor faults. Figure 5-8(b) shows the pre-processed raw water pH sensor data resulting in no true positives generated be the M-ERS (PSD).

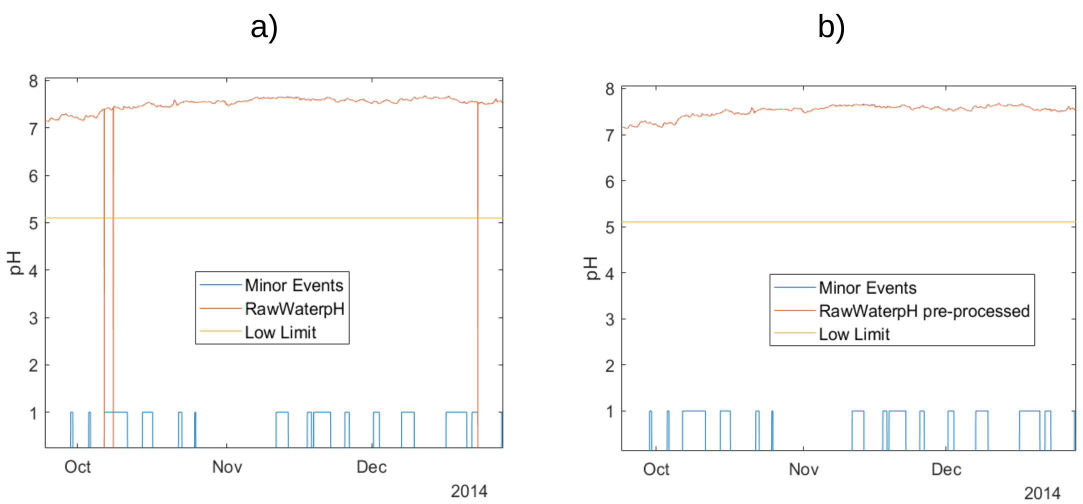

Figure 5-8 *Example of alarms raised by raw water pH signal with (a) origin sensor measurements and (b) pre-processed sensor data.*

It is very likely that the lower TPRs produced by the M-ERS (PSD) compared to the higher TPRs generated by E-ERS and M-ERS is mainly caused by the above described effect that is also occurring at other signals.

The above example indicates that the M-ERS (PSD) improves the quality of alarms raised. This is also confirmed by the higher PPV of 68% achieved by the M-ERS compared to the PPVs of 62% and 64% generated by the E-ERS and M-ERS, respectively. Although the improvements achieved by the M-ERS and M-ERS (PSD) are only of minor nature, the presented results have shown the beneficial use of optimised limit and persistence values for threshold-based event detection systems, the useful application of sensor data validation and pre-processing procedures but also the limitations of threshold-based ERSs.

5.4 HC-ERS Results and Discussion

5.4.1 Introduction

The final case focuses on the investigation of the event detection capabilities of the newly developed HC-ERS method described in Sections 4.4.2 and 4.4.3. The HC-ERS is applied to the same pre-processed dataset used for the M-ERS case (see preceding section) and using the same set of major and minor events (as used for all methods). The aim is to evaluate the performance of the HC-ERS method and assesses its capabilities for improvements in true detections and corresponding false alarm rates. This is done by comparing the HC-ERS detection statistics to those of E-ERS, M-ERS, M-ERS (PSD) and the well-established CANARY event detection system. The results of the analysis demonstrate that major improvements can be achieved by the application of HC-ERS' methodologies to the real data of the described demonstration site.

As mentioned in methodology section 4.4.1, the HC-ERS method consists of two principal stages:

1. At the first stage the fault detection methodology is applied resulting in individual signal alarms, based on the identification of deviations of water quality signal values from their 'normal' values.

2. At the second stage the methodology for the classification of identified individual signal alarms is applied to identify actual events and raise overall alarms by analysing multiple signals simultaneously.

5.4.2 First Stage Results

This section starts with the analysis to evaluate the most suitable fault detection methodology to be used at the first stage. This was done by exploring different SPC methods (see Section 4.4.2). Each SPC method was tested using the pre-processed data of the 22 observed sensor signals for the validation time period. After that the corresponding detection metrics were calculated in the same way as it was done for E-ERS and M-ERS methods. The detection capability of each method was then assessed by averaging the ratio between TPRs and FDRs for each signal and across all signals. The Performance Indicator (PI) derived this way is used to compare different SPC methods. The larger the value of PI the better.

The resulting PI values obtained are summarised in Table 5-12. Confusion matrices and detection metrics of all tests conducted in this context are not shown here to save space.

Table 5-12 *Performance Indicator (PI) of the selected SPC methods tested for window sizes (a) 1d and (b) 1 week and different control limits.*

(a)

SPC	PI		
	CL = ±3σ	CL = ±6σ	CL = ±12σ
CUSUM	1.656	1.661	1.665
EMWA	1.651	1.626	1.505
X-bar	1.647	1.640	1.577
r-chart	1.618	1.470	1.260
s-chart	1.580	1.541	1.489

(b)

SPC	PI		
	CL = ±3σ	CL = ±6σ	CL = ±12σ
CUSUM	1.654	1.646	1.620
s-chart	1.616	1.523	1.192
X-bar	1.541	1.358	1.123
r-chart	1.570	1.125	0.690
EMWA	1.480	1.278	1.008

As it can be seen from Table 5-12, the CUSUM method outperformed all other tested SPC methods by generating the highest PIs (range between 1.665 and 1.620) for both time windows and all thresholds used for upper and lower Control Limits (CL). Therefore, the CUSUM method has proven to be most promising stage one method. The performance of this methods was further optimised by moving away from standard (literature) parameters for this method by optimising the mean shift and thresholds values for each signal. This was done using the sensitivity type analysis described in Section 4.4.2. The resulting fine-tuned mean shift arguments and threshold limits are summarised in Table 5-13.

Table 5-13 *CUSUM fine-tuned parameters and PI values for 1d/1week window size.*

Signal	window size = 1d			window size = 1week		
	k	ULC/LCL	TPR/FDR	k	ULC/LCL	TPR/FDR
Raw Water Turbidity	1	+/- 6	1.52	1	+/- 1	1.52
Raw Water pH	2	+/- 1	1.53	1	+/- 12	1.51
Pre Flocculation pH Stream A	3	+/- 1	1.54	1	+/- 6	1.54
Pre Flocculation pH Stream B	1	+/- 6	1.52	8	+/- 12	3.32
Post Flotation Turbidity Stream A	1	+/- 6	1.54	1	+/- 1	1.51
Post Flotation Turbidity Stream B	4	+/- 1	1.55	5	+/- 12	1.55
DAF Iron Stream A	1	+/- 12	1.51	1	+/- 12	1.49
DAF Iron Stream B	1	+/- 3	1.43	1	+/- 6	1.48
Pre 1st Stage pH Stream A	2	+/- 1	1.53	1	+/- 3	1.45
Pre 1st Stage pH Stream B	1	+/- 6	1.53	1	+/- 1	1.43
Post 1st Stage Turbidity Stream A	2	+/- 6	1.56	1	+/- 12	1.49
Post 1st Stage Turbidity Stream B	2	+/- 1	1.55	1	+/- 1	1.51
Pre 2nd Stage pH Stream A	1	+/- 12	1.58	2	+/- 1	1.56
Pre 2nd Stage pH Stream B	1	+/- 1	1.55	2	+/- 1	1.54
Post 2nd Stage Turbidity Stream A	4	+/- 3	1.58	1	+/- 3	1.51
Post 2nd Stage Turbidity Stream B	3	+/- 12	1.61	1	+/- 1	1.44
Post 2nd Stage Chlorine Stream A	2	+/- 1	1.54	1	+/- 6	1.51
Post 2nd Stage Chlorine Stream B	3	+/- 1	1.56	9	+/- 12	1.80
Treated Water pH Stream A	1	+/- 6	1.56	1	+/-1	1.48
Treated Water pH Stream B	5	+/- 6	1.69	1	+/- 1	1.51
Final Water pH	3	+/- 6	1.57	1	+/- 1	1.53
Final Water Chlorine Residual	3	+/- 6	1.58	9	+/- 3	2.11
PI			1.55			1.63

As it can be seen from Table 5-13, the PI of 1.63 for the window size of 1 week is higher than the corresponding value of 1.55 for the 1d time window. The higher PI value demonstrates that higher detection performances can be achieved using a window size of 1 week rather than a 1d time window. Therefore, the CUSUM fault detection method performance was further evaluated by using the fine-tuned parameters of each signal for the 1 week sliding time window. The resulting confusion matrices obtained for the validation time period are shown in Figure 5-9.

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	59 (1/58)	2 (0/2)
NO	190	251

Raw Water TU **(a)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	60 (1/59)	1 (0/1)
NO	218	279

Raw Water pH **(b)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	206	267

Pre Flocculation pH (A **(c)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	8 (1/7)	53 (0/53)
NO	1	62

Pre Flocculation pH (B) **(d)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	207	268

Post Flotation TU (A) **(e)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	25 (1/24)	36 (0/36)
NO	7	68

Post Flotation TU (B) **(f)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	60 (1/59)	1 (0/1)
NO	179	240

DAF Iron (A) **(g)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	203	264

DAF Iron (B) **(h)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	58 (1/57)	3 (0/3)
NO	164	225

Pre 1st Stage pH (A **(i)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	55 (1/54)	6 (0/6)
NO	181	242

Pre 1st Stage pH (B) **(j)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	55 (1/54)	6 (0/6)
NO	167	228

Post 1st Stage TU (A) **(k)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	59 (1/58)	2 (0/2)
NO	159	220

Post 1st Stage TU (B) **(l)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	183	244

Pre 2^{nd} Stage pH (A) **(m)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	60 (1/59)	1 (0/1)
NO	189	250

Pre 2^{nd} Stage pH (B) **(n)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	59 (1/58)	2 (0/2)
NO	191	252

Post 2^{nd} Stage TU (A) **(o)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	196	257

Post 2^{nd} Stage TU (B) **(p)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	59 (1/58)	2 (0/2)
NO	193	254

Post 2^{nd} Stage Cl (A) **(q)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	9 (1/8)	52 (0/52)
NO	3	64

Post 2^{nd} Stage Cl (B) **(r)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	50 (1/49)	11 (0/11)
NO	160	221

Treated Water pH (A) **(s)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	54 (1/53)	7 (0/7)
NO	166	227

Treated Water pH (B) **(t)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	61 (1/60)	0 (0/0)
NO	207	268

Final Water pH **(u)**

Event (True Class)	Alarm (Predicted class) YES	Alarm (Predicted class) NO
YES	7 (1/6)	54 (0/54)
NO	1	62

Final Water Cl **(v)**

Figure 5-9 *Confusion matrices for CUSUM fault detection using finetuned parameters generated by the application to data of the observed signals.*

Using Figure 5-9 results, the corresponding detection rates including True Positive Rates (TPR), Positive Predictive Value (PPV), False Discovery Rate (FDR), False Positives (FP) and False Negative Rate (FNR) were calculated for each water quality signal. The resulting detection metrics are summarised in Table 5-14.

Table 5-14 *CUSUM (fine-tuned): Fault detection statistics of individual signals.*

CUSUM (finetuned) Fault Detection (22 critical alarm points)	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Raw Water Turbidity	97%	100%	97%	42%	58%	190	3%
Raw Water pH	98%	100%	98%	38%	62%	218	2%
Pre Flocculation pH Stream A	100%	100%	100%	40%	60%	206	0%
Pre Flocculation pH Stream B	13%	100%	12%	89%	11%	1	87%
Post Flotation Turbidity Stream A	100%	100%	100%	41%	59%	207	0%
Post Flotation Turbidity Stream B	41%	100%	40%	81%	19%	7	59%
DAF Iron Stream A	98%	100%	98%	42%	58%	179	2%
DAF Iron Stream B	100%	100%	100%	39%	61%	203	0%
Pre 1st Stage pH Stream A	95%	100%	95%	40%	60%	164	5%
Pre 1st Stage pH Stream B	90%	100%	90%	40%	60%	181	10%
Post 1st Stage Turbidity Stream A	90%	100%	90%	40%	60%	167	10%
Post 1st Stage Turbidity Stream B	97%	100%	97%	46%	55%	159	3%
Pre 2nd Stage pH Stream A	100%	100%	100%	45%	55%	183	0%
Pre 2nd Stage pH Stream B	98%	100%	98%	43%	57%	189	2%
Post 2nd Stage Turbidity Stream A	97%	100%	97%	42%	58%	191	3%
Post 2nd Stage Turbidity Stream B	100%	100%	100%	43%	57%	196	0%
Post 2nd Stage Chlorine Stream A	97%	100%	97%	41%	59%	193	3%
Post 2nd Stage Chlorine Stream B	15%	100%	13%	75%	25%	3	85%
Treated Water pH Stream A	82%	100%	82%	43%	57%	160	18%
Treated Water pH Stream B	89%	100%	88%	44%	56%	166	12%
Final Water pH	100%	100%	100%	42%	58%	207	0%
Final Water Chlorine Residual	12%	100%	10%	88%	13%	1	89%

The corresponding detection metrics are shown in Table 5-15.

Table 5-15 *CUSUM (fine-tuned): Averaged fault detection statistics.*

CUSUM (finetuned) (22 critical alarm points)	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System	82%	100%	82%	49%	51%	3371	18%

The analysis of the detection statistics presented in the Table 5-15 has shown that fine-tuned CUSUM fault detection method is far more sensitive to signals' deviations from normal process condition than threshold-based ERSs presented in Section 5.2 and 5.3. CUSUM fault detection methodology enables the true detection of a great number of events displayed by a TPR of 82% but the system also shows a high FDR of 51% which is reflected by the vast number of 3371 false positives. Compared to E-ERS and M-ERS the TPR and FDR increased by around 60% and by 13% to 19% respectively. The high number of false positives was expected since the CUSUM control charts are well known to be highly sensitive by detecting already small shift in the process mean.

5.4.3 Second Stage Results

Given the high number of false alarms generated by the CUSUM method, classification method was developed and used at the second stage. The aim of the classification method is to learn possible relationships across multiple water quality signals thus enable the multivariate based event detection. It is anticipated that this way the large number of false alarms generated by the CUSUM can be reduced.

As mentioned in Section 4.4.3 several classification methods were tested. These include the NN, SVM, bagged (RF) and boosted (AdaBoost) decision trees. The aim was to evaluate and assess (a) event detection capabilities of different classification methods and (b) further fine tune the CUSUM method parameters.

The resulting detection metrics of the experiments with different classifiers are presented in Table 5-16.

Table 5-16 *Comparison of tested classifiers' detection statistics on CUSUM finetuned output data.*

Classifier	True Detections				False Detections			PI
	TPR			PPV	FDR	FP	FNR	
	Total	Major	Minor					
Random Forest	90%	100%	90%	48%	52%	89	10%	1.73
AdaBoost	98%	100%	98%	43%	57%	158	2%	1.72
Artificial Neural Network	98%	100%	98%	42%	58%	151	2%	1.69
SVM (RBF)	90%	100%	90%	44%	56%	117	10%	1.61

From the table can be seen that all classifiers except SVM with a lower PI of 1.61 display similar performance indices between 1.69 and 1.73, whereat RF classifiers have shown the highest value. Even though the table shows no great differences between the classifiers in performance, it gives an indication that RF seems to be promising for the event detection task.

However, further analysis is necessary to evaluate performance of the classifiers and the whole system, since the classifier models (a) are highly dependent on reliable input data and (b) estimate probabilities of event's presence. The occurrence of an event in WTW's processes is considered if the probability exceeded a threshold value of 0.5. Changing the threshold value will influence the detection results. Therefore, the performance of the classifier models was investigated over the full range of thresholds utilising Receiver Operating Characteristic (ROC) curves (Fawcett, 2006) combined with the calculated Area Under Curve (AUC) that is widely used for assessing and comparing classifier's performances. The higher the AUC value, the better the classification model performs. Figure 5-10 shows the ROC curves picturing each classification technique applied to outputs generated by the CUSUM fine-tuned fault detection method.

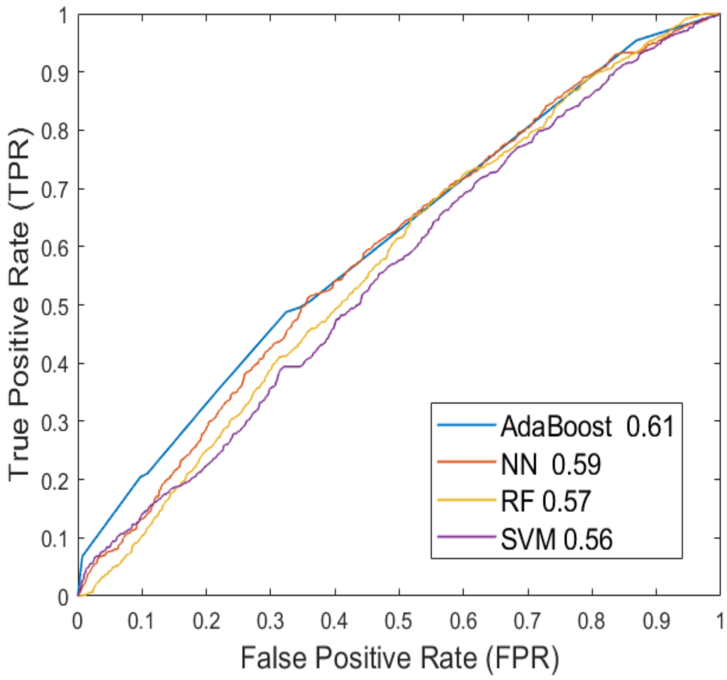

Figure 5-10 *Receiver Operating Characteristic (ROC) curves for AdaBoost, feedforward Neural Network (NN), Random Forest (RF) and Support Vector Machine (SVM) classifiers generated by their application to CUSUM finetuned output data.*

When analysing the detection metrics shown in Table 5-16 all systems demonstrated TPRs between 90% to 98%, i.e., high true detection rates. The number of false positives generated (between 89 to 158) are much lower compared to 354 and 271 false positives produced by the E-ERS and M-ERS (PSD) respectively. Although improvements of TPR and false positives were achieved the system shows FDRs between 52% and 58% which indicate that the models are not working satisfactory well.

This is confirmed by the analysis of the ROC curves shown in Figure 5-10. All of the demonstrated curves show similar behaviour with only a flat increase from south west to north east. A ROC curve close to ideal would rapidly increase from origin to nearly 1 and become flat at the following. The AUC values between 0.56 and 0.61 also shown for each technique separately in Figure 5-10 confirm the moderate performance of the models. The numbers demonstrated in the plot reflect a weak performance of the tested systems. When looking deeper into the CUSUM fine-tuned detection metrics of single signals (see Table 5-14) to explore the root cause for this issue it is obvious that the system produces exceptional high TPRs and FDRs for the vast majority of signals (e.g., TPR of 100% and FDR of 59% for post flotation turbidity stream A), whereas in contrast comparable signals show significant lower TPRs and FDRs, e.g. 41% and 19% for post flotation turbidity stream B signal respectively. This observation leads to the conclusion that the parameters utilised by the fine-tuned CUSUM fault detection method were not configured in an optimal manner to be used as suitable input for event classification. Therefore, further refinements on the CUSUM fault detection and parameter selection methodologies were necessary aiming to better capture the margins between signals' normal background deviations and anomalous deviations from that background caused by an event. Parameters achieved by the CUSUM refinement utilising criteria 2 and 3 (see Section 4.4.3) are summarised in Table 5-17.

Table 5-17 *CUSUM refined parameters for the respective criterion applied.*

CUSUM (criterion 2)	k [MAD]	ULC / LCL	Signal	k [MAD]	ULC / LCL
Raw Water Turbidity	5	+/- 3σ	Post 1st Stage Turbidity Stream B	4	+/- 12σ
Raw Water pH	6	+/- 12σ	Pre 2nd Stage pH Stream A	4	+/- 12σ
Pre Flocculation pH Stream A	9	+/- 6σ	Pre 2nd Stage pH Stream B	4	+/- 12σ
Pre Flocculation pH Stream B	5	+/- 6σ	Post 2nd Stage Turbidity Stream A	6	+/- 12σ
Post Flotation Turbidity Stream A	5	+/- 6σ	Post 2nd Stage Turbidity Stream B	6	+/- 12σ
Post Flotation Turbidity Stream B	5	+/- 6σ	Post 2nd Stage Chlorine Stream A	3	+/- 12σ
DAF Iron Stream A	5	+/- 1σ	Post 2nd Stage Chlorine Stream B	8	+/- 12σ
DAF Iron Stream B	6	+/- 6σ	Treated Water pH Stream A	7	+/- 3σ
Pre 1st Stage pH Stream A	7	+/- 1σ	Treated Water pH Stream B	3	+/- 12σ
Pre 1st Stage pH Stream B	5	+/- 1σ	Final Water pH	9	+/- 1σ
Post 1st Stage Turbidity Stream A	5	+/- 12σ	Final Water Chlorine Residual	7	+/- 12σ
CUSUM (criterion 3)	**k [MAD]**	**ULC / LCL**	**Signal**	**k [MAD]**	**ULC / LCL**
Raw Water Turbidity	9	+/- 12σ	Post 1st Stage Turbidity Stream B	9	+/- 12σ
Raw Water pH	9	+/- 12σ	Pre 2nd Stage pH Stream A	8	+/- 12σ
Pre Flocculation pH Stream A	9	+/- 12σ	Pre 2nd Stage pH Stream B	9	+/- 12σ
Pre Flocculation pH Stream B	8	+/- 12σ	Post 2nd Stage Turbidity Stream A	9	+/- 12σ
Post Flotation Turbidity Stream A	6	+/- 6σ	Post 2nd Stage Turbidity Stream B	9	+/- 12σ
Post Flotation Turbidity Stream B	7	+/- 6σ	Post 2nd Stage Chlorine Stream A	9	+/- 12σ
DAF Iron Stream A	9	+/- 12σ	Post 2nd Stage Chlorine Stream B	9	+/- 12σ
DAF Iron Stream B	9	+/- 12σ	Treated Water pH Stream A	9	+/- 12σ
Pre 1st Stage pH Stream A	9	+/- 12σ	Treated Water pH Stream B	9	+/- 12σ
Pre 1st Stage pH Stream B	9	+/- 12σ	Final Water pH	9	+/- 12σ
Post 1st Stage Turbidity Stream A	9	+/- 12σ	Final Water Chlorine Residual	7	+/- 12σ

After refinement and testing CUSUM fault detection methodology on pre-processed data of the 22 signals for validation time period corresponding detection rates for both applied criteria were calculated and summarised in Table 5-18.

Table 5-18 *CUSUM (refined): Fault detection statistics of individual signals according to the respective criterion applied.*

	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
CUSUM (criterion 2)	Total	Major	Minor				
Raw Water Turbidity	62%	100%	62%	46%	54%	81	38%
Raw Water pH	30%	100%	28%	49%	51%	25	71%
Pre Flocculation pH Stream A	21%	100%	20%	81%	19%	3	79%
Pre Flocculation pH Stream B	64%	100%	63%	50%	51%	51	36%
Post Flotation Turbidity Stream A	61%	100%	60%	67%	33%	29	39%
Post Flotation Turbidity Stream B	72%	100%	72%	70%	31%	32	28%
DAF Iron Stream A	80%	100%	80%	46%	54%	109	20%
DAF Iron Stream B	64%	100%	63%	49%	51%	61	36%
Pre 1st Stage pH Stream A	64%	100%	63%	59%	41%	43	36%
Pre 1st Stage pH Stream B	64%	100%	63%	45%	55%	74	36%
Post 1st Stage Turbidity Stream A	67%	100%	67%	48%	52%	81	33%
Post 1st Stage Turbidity Stream B	77%	100%	77%	50%	50%	93	23%
Pre 2nd Stage pH Stream A	74%	100%	73%	50%	50%	62	26%
Pre 2nd Stage pH Stream B	72%	100%	72%	49%	51%	74	28%
Post 2nd Stage Turbidity Stream A	69%	100%	68%	54%	47%	73	31%
Post 2nd Stage Turbidity Stream B	74%	100%	73%	56%	44%	75	26%
Post 2nd Stage Chlorine Stream A	85%	100%	85%	45%	56%	137	15%
Post 2nd Stage Chlorine Stream B	38%	100%	37%	63%	37%	16	62%
Treated Water pH Stream A	49%	100%	48%	63%	37%	27	51%
Treated Water pH Stream B	57%	100%	57%	59%	41%	46	43%
Final Water pH	41%	0%	42%	67%	33%	17	59%
Final Water Chlorine Residual	28%	100%	27%	68%	32%	10	72%
CUSUM (criterion 3)							
Raw Water Turbidity	39%	100%	38%	52%	49%	33	61%
Raw Water pH	7%	100%	5%	56%	44%	4	93%
Pre Flocculation pH Stream A	18%	100%	17%	85%	15%	2	82%
Pre Flocculation pH Stream B	20%	100%	18%	60%	40%	8	80%
Post Flotation Turbidity Stream A	57%	100%	57%	75%	25%	17	43%
Post Flotation Turbidity Stream B	54%	100%	53%	79%	21%	12	46%
DAF Iron Stream A	51%	100%	50%	46%	54%	49	49%
DAF Iron Stream B	38%	100%	37%	54%	46%	22	62%
Pre 1st Stage pH Stream A	53%	100%	52%	68%	32%	18	48%
Pre 1st Stage pH Stream B	44%	100%	43%	71%	29%	13	56%
Post 1st Stage Turbidity Stream A	56%	100%	55%	56%	44%	40	44%
Post 1st Stage Turbidity Stream B	62%	100%	62%	58%	42%	49	38%
Pre 2nd Stage pH Stream A	53%	100%	52%	70%	30%	16	48%
Pre 2nd Stage pH Stream B	36%	100%	35%	71%	29%	12	64%
Post 2nd Stage Turbidity Stream A	51%	100%	50%	56%	44%	40	49%
Post 2nd Stage Turbidity Stream B	64%	100%	63%	57%	43%	48	36%
Post 2nd Stage Chlorine Stream A	23%	100%	22%	54%	46%	16	77%
Post 2nd Stage Chlorine Stream B	34%	100%	33%	69%	31%	11	66%
Treated Water pH Stream A	28%	100%	27%	85%	15%	4	72%
Treated Water pH Stream B	30%	0%	30%	75%	25%	8	71%
Final Water pH	25%	0%	25%	76%	24%	6	75%
Final Water Chlorine Residual	28%	100%	27%	68%	32%	10	72%

Finally, the detection metrics for the overall systems shown in Table 5-19 were calculated as averaged detection rates and summation of false positives over all signals.

Table 5-19 *CUSUM (refined): Averaged fault detection statistics according to the respective criterion applied.*

CUSUM (refined) (22 critical alarm points)	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System (Criterion 2)	60%	95%	59%	56%	44%	1219	40%
Overall System (Criterion 3)	40%	91%	39%	65%	35%	438	61%

Comparing the detection results presented in Table 5-19 to the CUSUM fine-tuned detection metrics (shown in Table 5-15) it can easily be seen that the refined CUSUM model is by far less sensitive in fault detection demonstrated by the TPRs that have fallen by 22% and 42% combined with decreased FDRs of 7% and 16% for criterion 2 and criterion 3 respectively. The detection metrics of single signals presented in Table 5-18 show for comparable signals such as post flotation turbidity stream A and post flotation turbidity stream B now also similar TPRs, i.e. 57% and 54% and FDRs, i.e. 25% and 21% respectively (see criterion 3). This demonstrates improvements that have been made in the way of parameter choice and selection utilised by the CUSUM refined fault detection methodology.

TPRs of 40% and 60% as averaged rates across all signals generated by the CUSUM for criterion 2 and criterion 3 respectively should make a sufficient number of true detections available for an effective use of the classifier techniques.

Based on above, the best performing HR-ERS method so far is the combination of fine-tuned CUSUM methods and RF-based classification (criterion 3). The performance of this method is shown in the following table:

Table 5-20 *HC-ERS: Event detection statistics.*

HC-ERS (22 critical alarm points)	TPR			PPV	FDR	FP	FNR
	Total	Major	Minor				
Overall System	87%	100%	87%	77%	23%	27	13%

5.4.4 Final HC-ERS Method Optimisation

Even though, great improvements in detection performance were already achieved by HC-ERS a final optimisation of the system was conducted by investigating those sensor signals that are redundant or negatively affecting the performance of the event detection aiming to remove them from the system. A stepwise backward elimination analysis as described in Section 4.4.3 was conducted for the investigation of those signals. Figure 5-11 pictures the results of stepwise elimination of sensor signals by the plot comparing TPR vs FDR for each signal that has been eliminated.

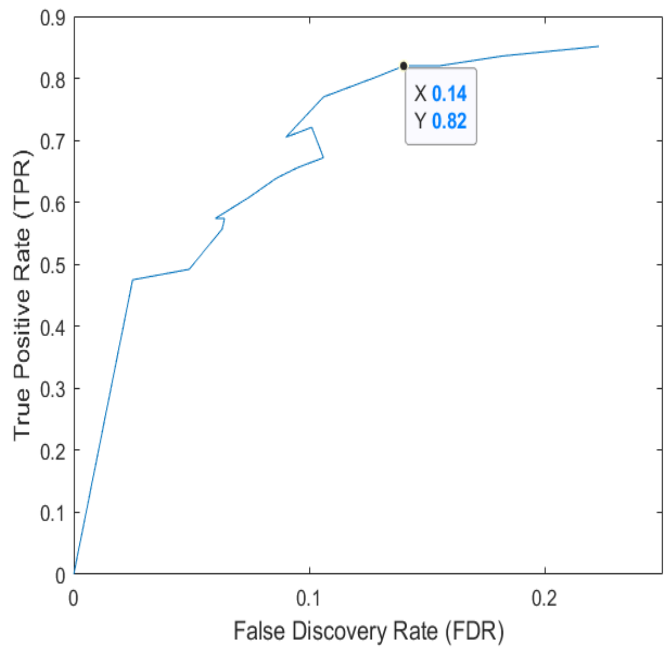

Figure 5-11 *Plot comparing TPR vs FDR for the stepwise elimination of sensor signals.*

When removing stepwise unfavourable signals starting initially with 22 signals (right hand side of the plot) the plot demonstrates a flat decrease of TPRs combined with a significant decrease of FDRs towards the labelled data point caused by the stepwise decreased number of signals used. From this point on more severe drops of TPRs are shown as more signals have been removed. The labelled point has been identified as optimal trade-off between TPR and FDR, where the system provides the most favourable event detection prediction. At this point the HC-ERS generates a TPR of 82% and FDR of 14%. These detection rates were achieved with 16 sensor signals that have been identified as most important signals that perform the HC-ERS in an effective manner. Removed signals as well as the signals applied to the HC-ERS are presented in Table 5-21.

Table 5-21 *Signals identified as most important for the detection performance of HC-ERS.*

Signals used by HC-ERS after optmisation	
Raw Water Turbidity	Pre 1st Stage pH Stream A
Raw Water pH	Post 1st Stage Turbidity Stream A
Pre Flocculation pH Stream A	Pre 2nd Stage pH Stream A
Pre Flocculation pH Stream B	Pre 2nd Stage pH Stream B
Post Flotation Turbidity Stream A	Post 2nd Stage Turbidity Stream B
Post Flotation Turbidity Stream B	Treated Water pH Stream A
DAF Iron Stream A	Final Water pH
DAF Iron Stream B	Final Water Chlorine Residual
Signals removed by HC-ERS after optmisation	
Pre 1st Stage pH Stream B	Post 2nd Stage Turbidity Stream A
Post 2nd Stage Turbidity Stream B	Post 2nd Stage Chlorine Stream A
Post 2nd Stage Chlorine Stream B	Treated Water pH Stream B
Pre Flocculation pH Stream A	Pre 2nd Stage pH Stream A

The final HC-ERS was tested using the remaining sensor signals shown in Table 5-21 on the pre-processed data for validation time period. The resulting detection metrics calculated in the same way as it was done for E-ERS and M-ERS is presented in Table 5-22.

Table 5-22 *HC-ERS: Event detection statistics after stepwise elimination of unfavourable signals.*

HC-ERS (16 signals)	TPR			PPV	FDR	FP	FNR	False alarms per wk	F_1 score
	Total	Major	Minor						
Overall System	82%	100%	82%	86%	14%	13	18%	0.3	0.84

5.4.5 Comparison of HC-ERS to Other Detection Methods

To further assess the performance of HC-ERS, its detection performance was compared to the well know CANARY method (Hart et al., 2007) event detection system. CANARY is providing three event detection algorithms: the time series increment (INC), linear prediction coefficient filter (LPCF) and multivariate nearest-neighbour (MVNN) (Klise and McKenna, 2006). Since LPCF and MVNN algorithms have proven to be the most effective (USEPA, 2014), the INC algorithm was only preliminary tested and not further used in this work. Following the parameter optimisation procedure described in USEPA Canary was calibrated and tested conducting a kind of sensitivity analysis to explore the most suitable key parameters for the application of Canary's LPCF and MVNN event detection algorithms. The evaluation of detection performances for both methods was done in the same way as it was done for HC-ERS by applying the same pre-processed sensor data for validation time period.

Canary event detection algorithms require five key parameters to be defined, in particular the following: (a) history window length in time steps used to calculate the baseline variability of water quality signals, (b) outlier threshold measured in units of standard deviations applied for the detection of outliers, (c) Binominal Event Discriminator (BED) window size in time steps used to provide the event probability for comparison against (d) the user defined number of required outliers (NRO) to determine an event, (e) the event threshold as value of probability used to declare a group of outliers as ultimate event. Both Canary algorithms were tested using the USEPA recommended configuration parameter values shown in Table 5-23.

Table 5-23 *Configuration parameter values used for the sensitivity analysis.*

Parameter	Configuration Values
History window	**2016 data points**
Outlier threshold	**0.5 - 3.0 standard deviations**
BED window	**12 data points**
Number of outliers (N_{RO})	**8, 9, 10**
Event threshold$_{1,2,3}$	**0.927, 0.981, 0.997**

BED, binomial event discriminator

Since the performance of CANARY methods is particularly sensitive to the outlier threshold the sensitivity analysis is aimed to explore suitable threshold values for LPCF and MVNN algorithms. For this analysis only a large history window of 2016 TS (7d) was selected because it has the same size as the window used for HC-ERS and it was proven that increasing the history window results in fewer alarms, while lower values (lower than 1.5 days) will increase the number of alarms (EPA, 2010). Corresponding to the experiments conducted by USEPA a window size of 12 TS (1hr) was selected for the BED window because similar to above shorter BED sizes will raise the number of alarms, while with larger windows events of short duration (shorter than the BED) will not be detected. Since a limited number of historical events contained in the validation dataset having a duration of ~1hr only a BED window of 12 TS was used for the analysis. The number of outliers (N_{RO}) required to define the event thresholds must be chosen as a number less than the BED window. The numbers of outliers used for the analyses were calculated as follows:

$$N_{RO} = \sum_{i=0}^{2} \left(\frac{2}{3} BED + i \right) \qquad (8)$$

N_{RO} can then be used to calculate the event thresholds. The event thresholds utilised for the sensitivity analysis (see Table 5-23) were defined as follows:

$$\text{Event Threshold} = \sum_{i=0}^{N_{RO}} \left(\frac{\text{BED!}}{i!\,(\text{BED} - i)!} \right) \left(\frac{1}{2} \right)^{BED} \qquad (9)$$

$$In\ Excel = BINOMDIST(N_{RO}, BED, 0.5, TRUE)$$

Following the recommendations of USEPA additional parameters were set on each individual signal. Setting these parameters lowers alarms caused by invalid data. Therefore, valid ranges for the sensor data were introduced to the analysis of both LPCF and MVNN Canary event detection algorithms. Any data outside this range is treated as having originated from a sensor fault (which presumably has not a great impact to methodologies' performance since the data used for the analysis has already been pre-processed). The range of valid sensor values provided from the water company is shown for each signal separately in Table 5-24.

Table 5-24 *Engineering ranges of sensors.*

Signal	Units	Engineering Range From	Engineering Range to
Raw Water Turbidity	NTU	-1.25	51.25
Raw Water pH	pH	1.75	12.25
Pre Flocculation pH Stream A	pH	1.75	12.25
Pre Flocculation pH Stream B	pH	1.75	12.25
Post Flotation Turbidity Stream A	NTU	-0.25	10.25
Post Flotation Turbidity Stream B	NTU	-0.25	10.25
DAF Iron Stream A	mg/l	-0.13	5.13
DAF Iron Stream B	mg/l	-0.13	5.13
Pre 1st Stage pH Stream A	pH	1.75	12.25
Pre 1st Stage pH Stream B	pH	1.75	12.25
Post 1st Stage Turbidity Stream A	NTU	-0.05	2.05
Post 1st Stage Turbidity Stream B	NTU	-0.05	2.05
Pre 2nd Stage pH Stream A	pH	1.75	12.25
Pre 2nd Stage pH Stream B	pH	1.75	12.25
Post 2nd Stage Turbidity Stream A	NTU	-0.05	2.05
Post 2nd Stage Turbidity Stream B	NTU	-0.05	2.05
Post 2nd Stage Chlorine Stream A	mg/l	-0.05	2.05
Post 2nd Stage Chlorine Stream B	mg/l	-0.05	2.05
Treated Water pH Stream A	pH	1.75	12.25
Treated Water pH Stream B	pH	1.75	12.25
Final Water pH	pH	1.75	12.25
Final Water Chlorine Residual	mg/l	-0.05	2.05

Once the configuration parameters were defined the sensitivity analysis was conducted by gradually increasing the outlier threshold in increments of 0.25 standard deviations from 1 to 3 standard deviations and evaluating the test results for each event threshold value (see Table 5-23). This way the sensitivity tests were conducted for both detection algorithms resulting in corresponding detection metrics and F_1 scores. The optimised outlier and event threshold combination for LPCF and MVNN algorithms was derived by selecting the combination with the maximum F_1 score. Detection metrics and F_1 score of both methods using optimised outlier and event thresholds are shown in Table 5-25.

Table 5-25 *Detection metrics of optimised LPCF and MVNN Canary event detection methods tested on validation time period.*

Detetion Method	True Detections				False Detections			False alarms per wk	F_1 score
	TPR			PPV	FDR	FP	FNR		
	Total	Major	Minor						
Canary (LPCF)	79%	100%	78%	69%	31%	33	21%	0.6	0.73
Canary (MVNN)	100%	100%	100%	48%	52%	145	0%	2.8	0.65

Canary's LPCF algorithm using an outlier threshold of 2.75 standard deviations combined with an event threshold of 0.981 has demonstrated best detection performance among both algorithms and all tested configurations. The test results are in line with the studies conducted by USEPA testing both algorithms on WTW real-life data, whose results have shown that Canary's LPCF usually outperformed the MVNN method. Therefore, the detection results of Canary's system utilising LPCF detection algorithm have been used for comparison with E-ERS', M-ERS' and HC-ERS' detection performances.

For better comparison of all tested systems a summary of the detection metrics of each method supplemented by the number of false alarms per week and F_1 score is sown in Table 5-26 (sorted in order of highest to lowest F_1 scores).

Table 5-26 *Detection metrics of analysed event detection systems tested on validation time period.*

ERS	True Detections				False Detections			False alarms per wk	F_1 score
	TPR			PPV	FDR	FP	FNR		
	Total	Major	Minor						
HC-ERS	82%	100%	82%	86%	14%	13	18%	0.3	0.84
Canary (LPCF)	79%	100%	78%	69%	31%	33	21%	0.6	0.73
M-ERS	23%	68%	22%	64%	36%	308	77%	5.9	0.32
M-ERS (PSD)	20%	64%	20%	68%	32%	271	80%	5.2	0.31
E-ERS	22%	64%	21%	62%	38%	354	78%	6.8	0.31

From Table 5-26 it can be seen that the HC-ERS method outperforms the other ERSs in all key figures. The good performance of HC-ERS is illustrated by a 3% higher TPR for total events and more than halved FDR in contrast to the second best CANARY system. The table also demonstrates the limitations of threshold-based fault detection systems displayed by the far lower F_1 scores generated by E-ERS and M-ERSs.

Moreover, considering the computational efficiency of HC-ERS including sensor data validation and pre-processing procedure the system is able to process approximately 300 observations per second, whereas CANARY processed around 100 observations per second. These results were achieved on a commercial laptop with i5 2.2 GHz processor having 12GB RAM, i.e. both, HC-ERS and CANARY enable event detection in near-real time.

Even though the removal of six redundant signals lead to a drop of the TPR by 5% it also causes a decrease in FDR by 9% in contrast to the system using the initial 22 sensor signals. The lower TPR is the trade-off for the benefit of reduced false alarms resulting in only 13 false positives generated by the HC-ERS. The system shows the highest F1 score of 0.84 and produces with a value of 0.3 by far the least false alarms per week among all tested event detection systems.

5.5 Summary

In this section the results of the case studies described in this chapter can be summarised as follows. After a brief introduction, the results of the performance evaluation of E-ERS have been presented and discussed in Section 5.2. This first case study has demonstrated that the threshold-based E-ERS shows moderate event detection capabilities achieving a TPR of 22% combined with an FDR of 38% and generates a high number of false alarms, i.e. 6.8 false alarms per week.

In the Section 5.3 the outputs of the second case study have been presented and discussed. The evaluation of M-ERS' detection performance has shown that improvements of only minor degree can be achieved by the application of optimised threshold and persistence values to the E-ERS. The results of M-ERS' performance evaluation have demonstrated an increased TPR by 1% combined with 1% drop of the FDR in contrast to the E-ERS. This case study has also demonstrated the beneficial use of validated and pre-processed sensor data. The evaluation of M-ERS' (PSD) performance showed a decrease in FDR by 6% and 4% at the expense of 2% and 3% lower TPRs compared to E-ERS and M-ERS respectively. But similar to the application of optimised thresholds, the improvements achieved by the use of pre-processed sensor data were not dramatically high.

The final Section in this chapter presents the results of the performance evaluation of the new developed HC-ERS. Overall the event detection method has been successfully applied to the real-wold water quality sensor data of the demonstration WTW. CUSUM fault and RF event detection in combination have proven to perform well in the detection of failure events at WTWs' processes. The evaluation of HC-ERS' performance with a TPR of 82% and an FDR of 14% respectively has demonstrated major improvements against threshold-based E-ERS and M-ERS and also clearly better detection capabilities compared to its benchmark CANARY. Moreover, HC-ERS has proven to be sufficiently fast for near real-time event detection. These results indicate the potential of the system to be effectively used for event detection at WTW.

Chapter 6: Conclusions

6 CONCLUSIONS

This thesis has presented a range of methodologies for near real-time detection of failure events at WTW's processes and demonstrated their application on real WTW's sensor data. The work comprised the evaluation of the: (a) existing, threshold-based event detection system (E-ERS) currently used by the United Utilities water company at the selected demonstration site, (b) the optimisation of the existing threshold-based event detection system resulting in the modified ERS (M-ERS) and (c) the development of a novel, hybrid event detection system (HC-ERS) that makes use of the CUSUM-based fault detection and RF event detection. Section 6.1 provides a summary of the work done in the thesis followed by key conclusions and contributions made to the research area presented in Section 6.2. Finally, recommendations for future developments in this work and the field of near real-time recognition of failure events at WTW's processes are given in Section 6.3.

6.1 Thesis Summary

The first chapter of this thesis provided an introduction (see Section 1.1) into the field of event detection at WTW's processes and highlighted in Section 1.2 the shortcomings of event detection systems currently used in water industry. In this way the scope and objectives of the thesis were defined in Section 1.3 as part of the overall aim to develop and validate a new methodology and a near real-time event recognition system for the detection of faulty sensor data and faulty processes at Water Treatment Works (WTWs). The methodology must be able to detect and distinguish fault sensor data and fault WTW's processes in near real-time. Additionally, the methodology should be cheap in implementation and practical in real world operation. Finally, the structure of the thesis was outlined in Section 1.4.

Starting with a brief introduction in Section 2.1, in the second chapter previous work relevant to the development of this thesis was reviewed. Section 2.2 included a brief overview about general fault detection methods and focused in Section 2.3 on methodologies developed for the water sector as well as in Section 2.4 on software applications already used in water industry for event detection at WTW's processes. The review provided a summary of current methods including widely spread Statistical Process Control (SPC) techniques such as control charts and Principle Components Analysis (PCA), but also - most evident to the development of the novel near real-time event recognition system presented in this thesis - recently increasing machine learning techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM).

The third chapter starts by the description of the real-life WTW in Section 3.1 and the related data used for the development of the near real-time event detection methodologies in Section 3.2. Real world data from the demonstration site provided by the water company was used for the developments within this thesis. The dataset used for the case studies included data of 22 water quality signals observed by sensors deployed at the WTW over a time period of more than 3 years and contained a large variety of failure events affecting WTW's treatment processes identified by visual inspection of single sensor signals. The 163 events identified this way were classified into 5 major events and 158 minor events (see Section 3.3). All case studies applied this comprehensive dataset that was split into ~70% of total time period of data used for calibration and ~30% of total time period for validation on unseen data. Finally, a brief summary of the chapter was given in Section 3.4.

In Chapter 4 methodologies of existing event recognition system (E-ERS), modified event recognition system M-ERS and new developed CUSUSM Hybrid event recognition system (HC-ERS) were discussed. After a brief introduction in Section 4.1, an overview of E-ERS' architecture was provided in Section 4.2 aiming to determine the fundamentals for its assessment and performance evaluation. Section 4.3 detailed the methods applied for M-ERS improving E-ERS's detection capabilities. The section has outlined two strategies to achieve enhanced detection performance of the E-ERS that comprised first the use of optimised thresholds and persistence values and second the application of validated and pre-processed sensor data. Whilst the enumeration method used for determining optimised of threshold and persistence values was detailed in Section 4.3.1, sensor data validation and pre-processing methods applied to the sensor raw data were outlined in Section 4.3.2. Detection methodologies used for development of the novel Hybrid CUSUM Event Recognition System (HC-ERS) were presented in Section 4.4. HC-ERS that combines the CUSUM control chart techniques described in Section 4.4.2 for fault detection on individual water quality signals with Random Forest (RF) machine learning classifiers detailed in Section 4.4.3 for the prediction of event occurrence probability and hence ultimate overall event detection. Methodologies applied and metrics used for the performance evaluation of the different ERSs were outlined in Section 4.5 followed by a concluding summary of the chapter in Section 4.6.

Finally, Chapter 5 provided the results of the case studies conducted on E-ERS, M-ERS and HC-ERS. After a brief introduction in Section 5.1, the E-ERS that makes use of thresholds and persistence values applied to singe signals was assessed by evaluating its capabilities of detecting failure events at WTW's processes in Section 5.2. To demonstrate E-ERS's performance and for comparison to M-ERS and HC-ERS, it was applied to the dataset of validation time period and resulting detection metrics evaluated. The E-ERS performed moderate, achieving a F_1 value of 0.31 and generating 6.8 false alarms per week. Section 5.3 discussed the results of performance evaluation of M-ERS. To demonstrate M-ERS's performance that makes use of optimised threshold and persistence values, it was first applied to the same dataset used for the performance evaluation of E-ERS and resulting detection metrics evaluated. M-ERS achieved only minor improvements compared to E-ERS, generating a F_1 value of 0.32 and 5.9 false alarms per week. The beneficial use of validated and Pre-processed Sensor Data (PSD) were demonstrated by the application of M-ERS (PSD) on pre-processed sensor data achieving a fall from 5.9 to 5.2 false alarms per week compared to M-ERS applied to sensor raw data.

In Section 5.4 the results of the case studies conducted on the HC-ERS's were discussed. After a final refinement of CUSUM fault detection parameters and selection of RF as best performing classifiers HC-ERS was applied to pre-processed sensor data and the resulting detection metrics compared to above ERSs and the well-established CANARY (Hart and McKenna, 2009) event detection systems evaluated to demonstrate the detection performance of HC-ERS. Compared to all other ERSs the novel HC-ERS performed well achieving a F_1 value of 0.82 and 0.5 false alarms per week in contrast to CANARY's F_1 value of 0.70 and 0.9 false alarms per week. After final optimisation by removing redundant signals the evaluation of HC-ERS's detection performance conducted by its application on pre-processed sensor data of 16 remaining signals demonstrated further improvements of the system achieving a F_1 value of 0.84 and generating 0.3 false alarms per week.

6.2 Conclusions and Contributions

The work conducted on the development of the novel HC-ERS forms meaningful contribution to the research area. The key conclusions and contributions presented in this thesis are as follows:

- **A novel event recognition methodology (HC-ERS) was developed.** The new methodology is capable of identifying the presence of failure events at WTW's processes in near-real-time by processing water quality signals coming from sensors deployed at WTW. Unlike other ERSs found in the literature, which usually deploy a single method for event detection, the HC-ERS utilises a hybrid of two data-driven methods. The new HC-ERS was tested and validated on real life data and has proved to be effective. The HC-ERS has achieved a true positive detection rate of 82%, an F_1 value of 0.84 and a false alarm rate of 14% (equivalent to only 0.3 false alarms per week) when applied on the validation (i.e. unseen) data set (see Section 5.4 for details). For comparison, the equivalent values obtained by using the CANARY method were 79% true detection rate, 0.73 F_1 score, 31% false alarm rate and 0.6 false alarms per week. Therefore, the HC-ERS method shows promise for practical application in the water industry.
- **A new automated method for sensor data validation and pre-processing in near real-time was developed**. This method is used for the identification and correction of different types of sensor faults in near real-time. The method applies four statistical tests to the collected raw sensor data (see Section 4.3.2) for the identification and correction of faulty sensor data. The new sensor data validation and pre-processing method was tested, validated and demonstrated by comparing the performance of the M-ERS method with and without Pre-processed Sensor Data (PSD) method. When compared to the M-ERS method, the detection results achieved by the M-ERS (PSD) method demonstrated to be effective by

reducing false alarms from 5.9 to 5.2 false alarms per week (see Section 5.3 for details). The PSD was integrated into the new HC-ERS method.

- **Detailed testing and optimisation of the existing, threshold-based event detection system (E-ERS) was conducted**. This method is currently used by the water company and has demonstrated fairly moderate detection performance. The system achieved a modest F_1 value of 0.31 with barely acceptable 6.8 false alarms generated per week. An attempt was made to overcome the limitations of the E-ERS method (see Section 5.2 for details) by optimising its thresholds and related persistence time values. However, the resulting M-ERS method obtained this way demonstrated only minor improvements. The use of PSD method applied to M-ERS demonstrated no substantial advancements in detection performance either (see Section 5.3 for details). All this demonstrates the clear limitations of threshold based detection methods which, unfortunately, continue to dominate in engineering practice.

6.3 Future Work Recommendations

Future work should involve further validation of the new HC-ERS method on additional real world data collected at different WTW sites. In this thesis, the testing and validation was done on a single WTW due to limitations in availability of real world data. Tests at additional WTWs with potentially different sensors and failure events would not only enable a more thorough validation and demonstration of the proposed HC-ERS detection method, this would, more importantly, provide an opportunity to gain additional insights and hence further generalise the observations made in this thesis.

Future work could also consider novel machine learning and other methods as this is a constantly developing field. The latest rapid progress in development of machine learning and artificial intelligence techniques could be explored to further improve the detection methodology. This applies especially to the emerging field of deep learning techniques such as deep learning artificial neural networks. The application of improved event detection techniques would enable further improvements in the detection performance and lowering the false positive rate.

Future work could also look into the classification of identified events into different subtypes of respective root causes with the aim to further improve detection performance. At this development stage the classification of events into subtypes was not possible due to a lack of knowledge about the root causes of labelled events within the used dataset. Analysis of failure events detected by the HC-ERS and investigation of their root causes would enable to set up a continuously growing database containing different fault types labelled by their root causes. This specific fault types should be first identified and classified by experts from the water company up to this stage the database is filled with of a sufficient high number of each fault types. From this stage on the classification procedure could be conducted automatically by system's recognition of respective fault types and incorporated into the HC-ERS and other methods to enable the system to be additionally used to identify fault types and thus will help to determine the possible root cause of a problem. This further development of HC-ERS towards fault type classification in turn, would allow to locate faults and map these to corresponding WTW's processes and single sensors. However, diagnosis of events done this way would be extremely advantageous for practical WTW's operation, because an unusually high or frequent occurrence of failure events at certain locations of the treatment processes would indicate weak spots, e.g. dosages, retention times, etc. that could then be corrected by WTW's operators.

Further research should be undertaken to overcome the specific issue described in Appendix B (see Figure B-5). Due to the time delay of downstream signal's deviations in the presence of an event, the ERS is not continuously identifying the event over its entire duration. It was frequently observed that multiple alarms were raised in the presence of an event caused by intermediate interruptions of the event status. To overcome this issue, once a fault is identified by CUSUM fault detection a persistence value could be introduced aiming to prolongate the duration of labelling this fault for a number of specified time steps. The use of such kind of persistence value presumably would be beneficial for avoiding above described interruptions. The more continuous event predictions achieved this way should further increase the detection performance and also lead to higher AUC values when generating ROC curves for the performance evaluation of the classifier. Further testing should be conducted to investigate the optimal number of time steps to be used for these persistence values and to explore to what extend this measure improves the detection performance of the system.

The use of enhanced sensors that can provide their or nearby asset's 'health status' should be investigated to examine possible options of integrating this additional metadata into the detection process. The use of this additional information could be beneficial for the more reliable detection of process faults and would likely improve the system's overall detection performance. For example, using sensors with self-diagnostic feasibility and integrating additional sensor information about the status/condition of assets, e.g. of pump speed, vibration, etc. will certainly help to improve the reliability of fault predictions. The additional information, however, could possibly reduce the value of separate sensor data validation and pre-processing procedure used in this thesis due to the use of more specific information coming directly from the WTW sensors.

The HC-ERS could also be integrated into a decision support tool with a suitable user-friendly interface enabling improved visibility of the WTW's process states in near real-time. The decision support tool should ideally display an overall view of all WTW's treatment processes by visually indicating the condition of single process states, e.g. similar to traffic lights systems (green for healthy, orange for early warning of processes in danger of getting out of control and red for faulty processes). A great visibility of WTW's process conditions at a glance is a key feature and crucial for a successful and practical implementation of HC-ERS in the water industry.

APPENDIX A SIDE CALCULATION OF PRELIMINARY CLASSIFIER PARAMETER SETTINGS

Appendix A presents the preliminary analysis to explore the number of neurons and decision trees that represents the most suitable choice to be used by the ANN and RF or AdaBoost classifiers for the detection of failure events at WTW's processes.

First, the analysis was done using the ANN classification method with the aim to identify the adequate number of ANN's neurons. This was done by growing the number of neurons and comparing this number against the ratio of TPR and FDR generated by the neural network. Figure A-1 shows that the number of 100 neurons was the most suitable choice. The ANN with 100 neurons was tested using the CUSUM (finetuned) output data for validation time period. After that, the detection performance of the method was evaluated by calculating the detection metrics in the same way as it was done for E-ERS and M-ERS (see Section 5.4.3).

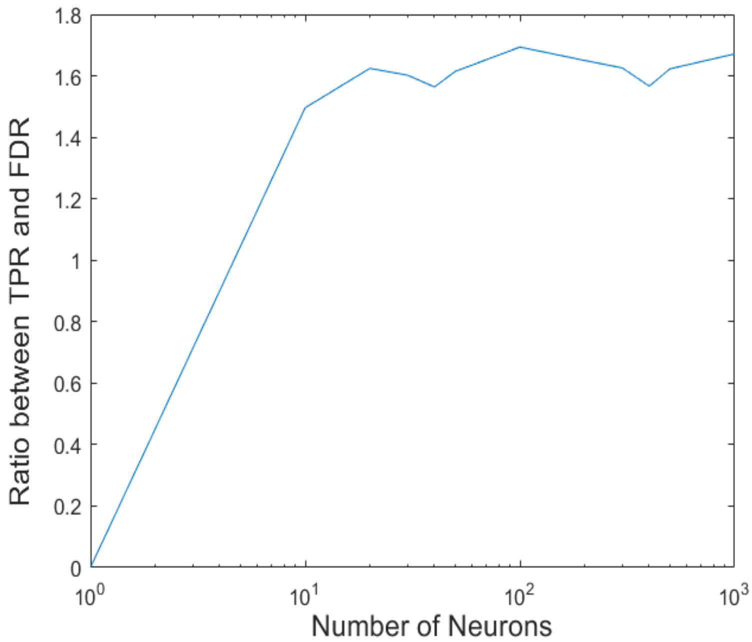

Figure A-1 *Comparison of the number of neurons for the Neural Network against the ration of True Positive Rate (TPR) to False Discovery Rate (FDR) generated by its application to CUSUM finetuned output data.*

Similar to above, the number of trees for RF and AdaBoost classifiers were explored by growing template trees and comparing the number of trees against the ratio of TPR of FDR. Numbers of 10 and 100 decision trees were derived as most suitable for the application of AdaBoost and RF classifiers to the CUSUM finetuned output data of validation time period (see Figure A-2).

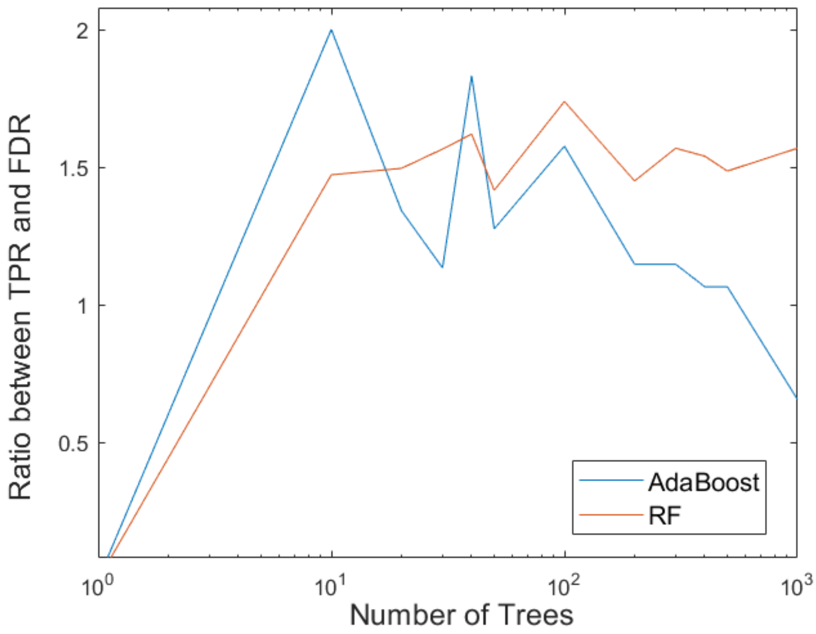

Figure A-2 *Comparison of the number of decision trees for RF and AdaBoost classifiers against the ration of True Positive Rate (TPR) to False Discovery Rate (FDR) generated by their application to CUSUM finetuned output data.*

AdaBoost and RF classifiers that make use of 10 and 100 decision trees respectively were then tested followed by the evaluation of their detection performances in the same way as it was done before (see Section 5.4.3).

APPENDIX B SIDE CALCULATION AND ANALYSIS OF REFINED CUSUM AND CLASSIFIER METHODS

In Appendix B side calculations to evaluate the most suitable parameter settings for the refinement of the event classification methods are presented first. Additionally, the analysis undertaken to evaluate the CUSUM criteria whose output serves the most adequate input for the event classifier model is outlined. Finally, in Appendix B the outcome of ROC curves and AUCs used for the performance evaluation of the classification procedures is discussed.

Prior to the final tests of the classifier models (see Section 5.4.3) again the adequate number of neurons for the NN and trees for AdaBoost and RF classifiers for the new CUSUM data were explored first. For the NN the numbers of neurons were derived by 10 and 400 Neurons for the application to CUSUM output data according to criterion 2 and criterion 3 respectively (see Figure B-1).

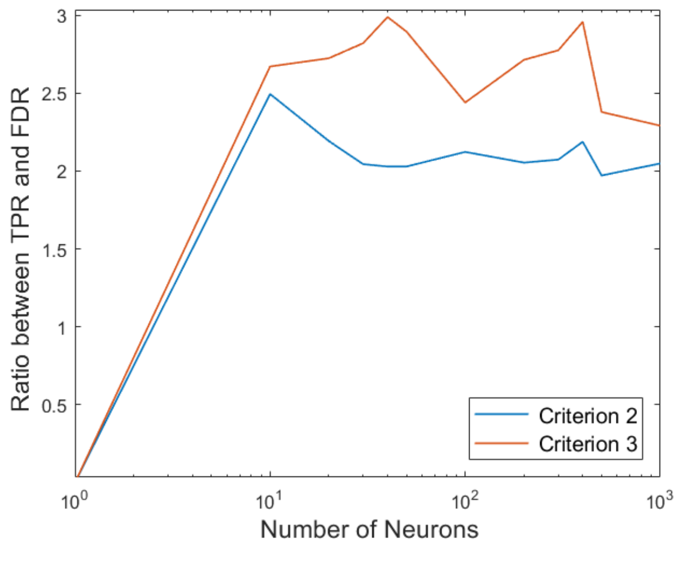

Figure B-1 *Comparison of the number of neurons for the Neural Network (NN) against the ratio of True Positive Rate (TPR) to False Discovery Rate (FDR) when applied to CUSUM refined criterion 2 and 3 output data.*

Figure B-2 shows the plot that displays the numbers of 50 and 100 decision trees derived as best choice for the application of AdaBoost and the RF classifier to CUSUM criterion 2 and 3 output data.

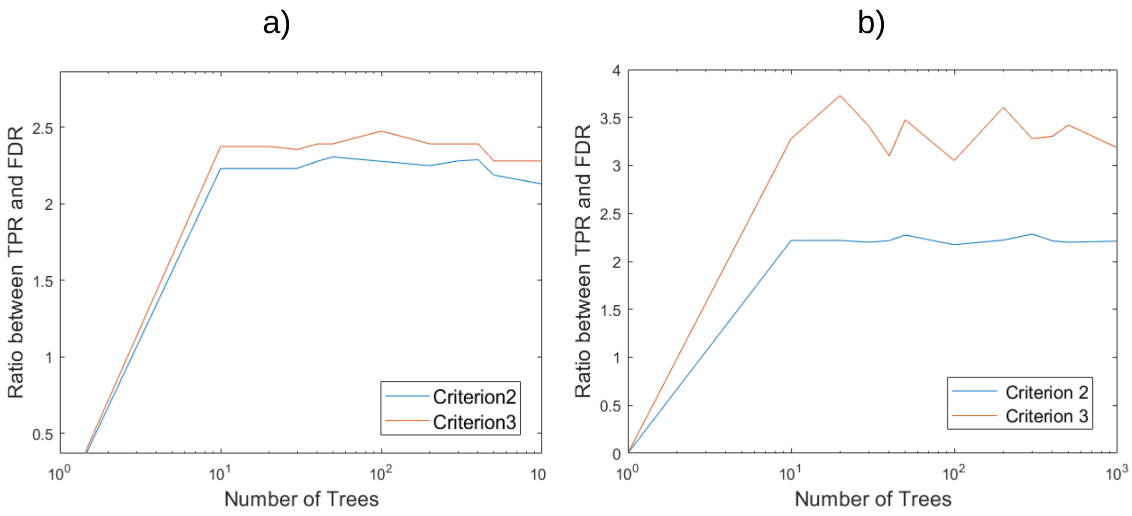

Figure B-2 *Comparison of the number of trees for AdaBoost (a) and Random Forest (b) ensemble classifier against the ratio of True Positive Rate (TPR) to False Discovery Rate (FDR) when applied to CUSUM criterion 2 and criterion 3 output data.*

Performance of each classifier technique was evaluated in the same way as it was done for the initial experiment by testing the models on the pre-processed sensor data of the 22 sensor signals for validation time period. The resulting detection metrics are shown in Table B-1.

Table B-1 *Comparison of tested classifiers' detection statistics on CUSUM refined criterion 2 and 3 output data.*

Classifier	True Detections				False Detections		
	TPR			PPV	FDR	FP	FNR
Criterion 2	Total	Major	Minor				
AdaBoost	71%	100%	70%	69%	31%	30	30%
Artificial Neural Network	80%	100%	80%	68%	32%	39	20%
Random Forest	87%	100%	87%	62%	38%	55	13%
SVM (RBF)	61%	100%	60%	75%	25%	19	39%
Classifier	**True Detections**				**False Detections**		
	TPR			**PPV**	**FDR**	**FP**	**FNR**
Criterion 3	**Total**	**Major**	**Minor**				
AdaBoost	61%	100%	60%	75%	25%	19	39%
Artificial Neural Network	75%	100%	75%	75%	26%	26	25%
Random Forest	87%	100%	87%	77%	23%	27	13%
SVM (RBF)	80%	100%	80%	67%	33%	42	20%

From the detection metrics presented in Table B-1 can be seen that the classifiers show improved event detection capabilities compared to preliminary tests with CUSUM finetuned parameters (see Table 5-16). The systems now generate reasonable high TPRs between 61% to 87% combined with significant lower FDRs of 23% to 38% compared to FDRs of 52% to 58% produced by the initial tests with CUSUM finetuned parameters. In general, the models using CUSUM output data with parameters refined according to criterion 3 (parameters have been optimised to maximise the number of true detections, i.e. true positives and true negatives) perform better than the models with CUSUM parameters refined after criterion 2 (parameters optimised by using regression techniques). This is demonstrated by usually lower FDRs of 23% to 33% at the same level of TPRs generated with CUSUM parameters selected by criterion 3 compared to the FDRs of 25% to 38% produced by the models using CUSUM criterion 2.

Regarding performance of the classification techniques RF classifiers were identified as best performing by achieving highest TPR of 87% and lowest FDR of 23% among all tested classifiers.

As it was done for the preliminary classifier tests ROC curves and AUCs were calculated to evaluate the performance of the classification procedures. Figure B-3 shows the ROC curves and AUCs of each technique for both CUSUM parameter selection criteria.

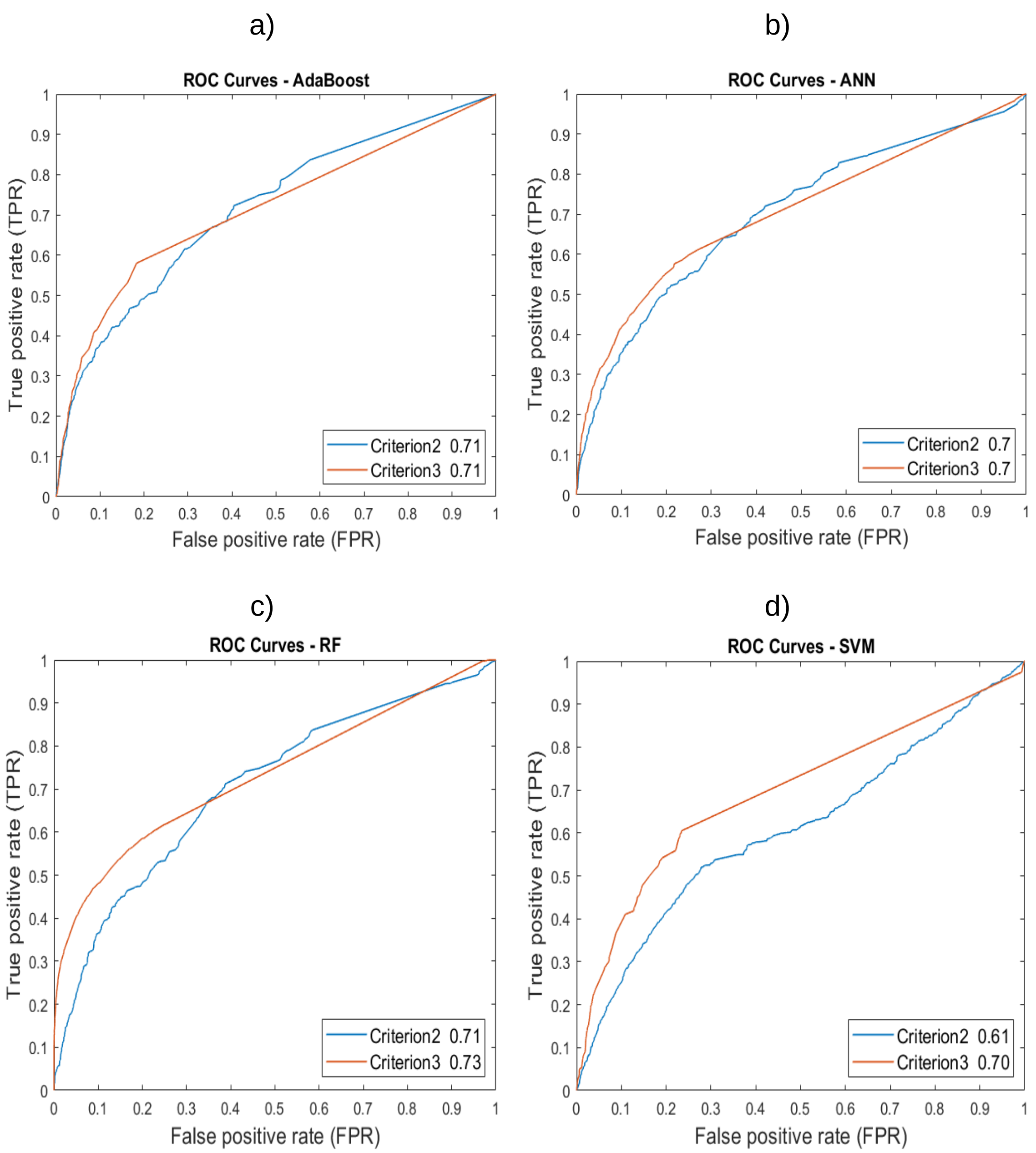

Figure B-3 *Receiver Operating Characteristic (ROC) curves for (a) AdaBoost, (b) Neural Network (NN), (c) Random Forest (RF), and (d) Support Vector Machine (SVM) classifiers generated by their application to refined CUSUM criterion 2 and criterion 3 output data.*

From the ROC curves as shown in Figure B-3 it can be seen that the performance of the models has greatly improved according to the results achieved by the initial tests. This is also confirmed by increased AUC values of around 0.7 (except the AUC of 0.61 generated by SVM for criterion 2) compared to the AUC values (mostly below 0.6) achieved by the initial testing of the classifiers. The ROC curves also confirm CUSUM criterion 3 as better choice than CUSUM criterion 2 since the curves produced by the classifiers using CUSUM output data generated with parameters optimised according to criterion 3 dominate in all cases the criterion 2 curves at the most important portion of the ROC curves (i.e. FPRs from 0 to 0.3). Figure B-4 shows the above mentioned portion of the ROC curves of the classifiers for CUSUM criterion 3 only.

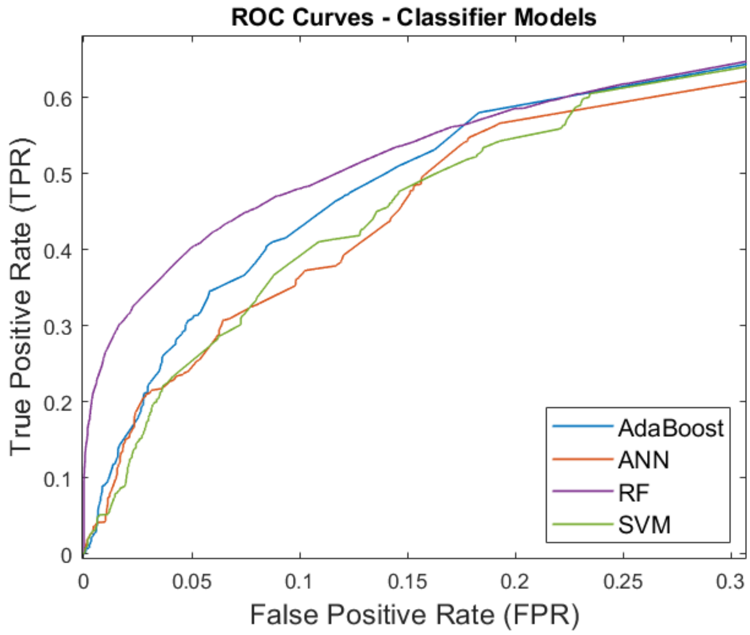

Figure B-4 *Receiver Operating Characteristic (ROC) curves for (a) AdaBoost, (b) Neural Network (NN), (c) Random Forest (RF), and (d) Support Vector Machine (SVM) classifiers generated by their application to refined CUSUM criterion 2 and criterion 3 output data.*

The figure shows that the ROC curve for RF dominate all other curves and thus confirms aside the highest AUC value of 0.73 and promising detection metrics (see Table B-1) achieved by RF classifiers this technique as most suitable event detection method.

Even though the event classification methodology appears to perform well some difficulties with ROC curves and AUC as measure for performance evaluation of ERS arise (EPA, 2013). When creating the ROC curves and calculating the AUC value each time step during an event is classified as a true positive or false negative. Due to inconsistencies and time lags in signal’s deviations the classifier is unable to clearly define continuously periods of event predictions during the whole time of the presence of an event. This issue is demonstrated by Figure B-

5, which shows the predictions of the RF classifier (red) compared with the labelled events (blue) over a cut-out of validation time period (from 17th December to 8th February). White spaces between event predictions during the presence of an event demonstrate above described effect. Although this effect does not seem to influence the performance of the event detection technique, the achieved AUC values do not fully reflect the event detection capabilities of the classifier.

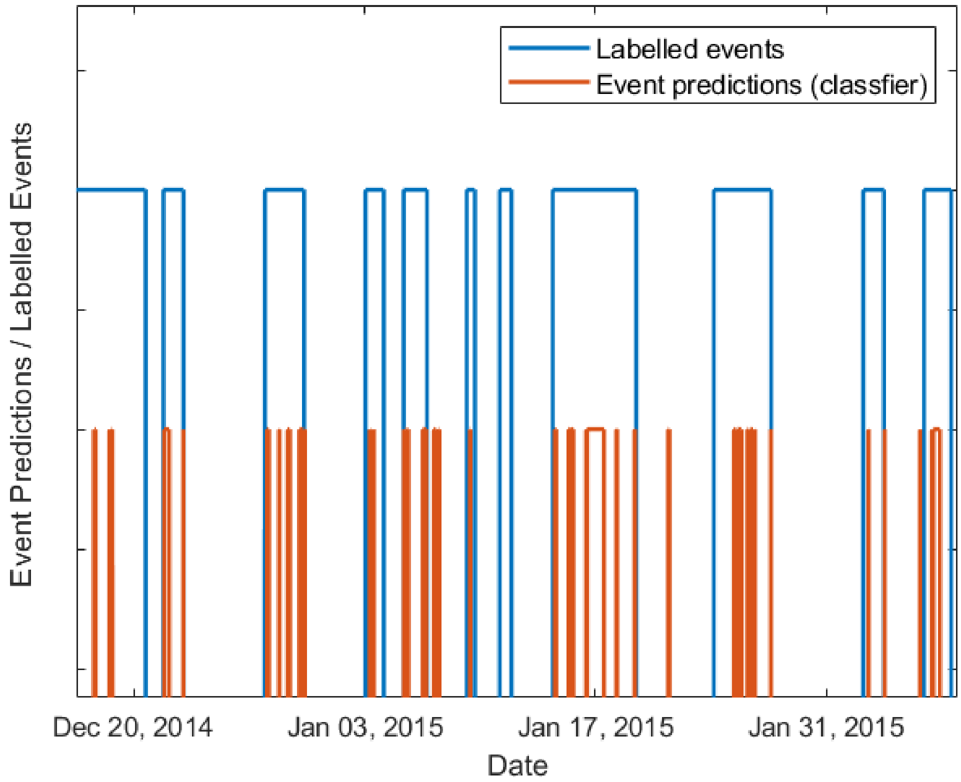

Figure B-5 *Event predictions of the RF classifier (red) compared with the labelled events (blue).*

Not only for the reason that RF classifiers have demonstrated best detection performance among the tested methods, but also for their easy implementation combined with a high computational efficiency this classification technique were chosen as best suitable event detection method for the new HC-ERS. RF required the least time to process compared to the other tested classifiers.

BIBLIOGRAPHY

Papers from the Candidate

Riss, G., Romano M., Woodward K., Memon F. A., Kapelan Z. (2018) 'Improving Detection of Events at Water Treatment Works: A UK Case Study', 13th International Conference on Hydroinformatics, HIC 2018, Palermo, Italy.

Riss, G., Romano M., Memon F. A., Kapelan Z. (2020, in preparation) 'Detection of Water Quality Failure Events at Treatment Works using Hybrid CUSUM Method'.

References

Abdi, H. and Williams, L.J. (2010) 'Principal component analysis', *WIREs Computational Statistics*, vol. 2, pp. 433-459.

Agneli, C. (2010) 'Diagnostic expert systems: from expert's knowledge to real-time systems', in Sajja, P.S., Patel, S. and Akerkar, R. (ed.) *Advanced knowledge based systems: model, applications & research*, Sajja & Akerkar edition, e-Book.

Aguado, D. and Rosen, C. (2007) 'Multivariate statistical monitoring of continuous wastewater treatment plants', *Engineering Applications of Artificial Intelligence*, vol. 21, pp. 1080-1091.

Alcala, C.F. and Qin, J.S. (2011) 'Analysis and generalization of fault diagnosis methods for process monitoring.', *Journal of Process Control*, vol. 21, no. 3, pp. 322–330.

Alferes, J., Tik, S., Copp, J. and Vanrolleghem, P.A. (2013) 'Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection', *Water Science & Technology*, vol. 68, no. 5, pp. 1022-1030.

Allgeier, S.C. and Umberg, K. (2008) 'Systematic evaluation of contaminant detection through water quality monitoring', AWWA Water Security Congress, Denver, CO.

An, L. and Sepehri, N. (2005) 'Hydraulic actuator leakage fault detection using extended Kalman filter', *International Journal of Fluid Power*, vol. 6, no. 1, pp. 41-51.

Arad, J., Housh, M., Perelman, L. and Ostfeld, A. (2013) 'A dynamic thresholds scheme for contaminant event detection in water distribution systems', *Water Research*, vol. 47, pp. 1899-1908.

Banna, M.H., Imran, S., Francisque, A., Najjaran, H., Sadiq, R., Rodriguez, M. and Hoorfar, M. (2014) 'Online drinking water quality monitoring: review on available and emerging technologies', *Critical Reviews in Environmental Science and Technology*, vol. 44, no. 12, pp. 1370-1421.

Barnard, G.A. (1959) 'Control Charts and Stochastic Proeesses', *Journal o/ the Royal. Statisticai Society*, vol. B 21, pp. 239-271.

Bauer, E. and Kohavi, R. (1999) 'An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants', *Machine Learning*, vol. 36, pp. 105-139.

Bellotti, T., Nouretdinov, I., Yang, M. and Gammerman, A. (2014) 'Feature Selection', in *Conformal Prediction for Reliable Machine - Theory, Adaptations and Applications*.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society*, vol. 57, no. 1, pp. 289-300.

Bernard, T., Mossgraber, J., Madar, A.E., Rosenberg, A., Deuerlein, J., Lucas, H., Boudergui, K., Ilver, D., Brill, E. and Ulitzur, N. (2015) 'SAFEWATER - innovative tools for the detection and mitigation of CBRN related contamination events of drinking water', 13st International Conference on Computing and Control for the Water Industry, CCWI 2015, Leicester, UK.

Bernard, T., Rosenberg, A., Lucas, H., Rieder, A., Mossgraber, J., Deuerlein, J., Brill, E., Boudergui, K., Ilver, D., Ulitzur, N. and Madar, A.E. (2017) 'SAFEWATER – Application and Results of Innovative Tools for the Detection and Mitigation of CBRN- related Contamination Events in Drinking Water Supply Systems', CCWI 2017 – Computing and Control for the Water Industry, Sheffield, U.K.

Boser, B.E., Guyon, I.M. and Vapnik, V.N. (1992) 'A training algorithm for optimal margin classifiers', Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, 144-152.

Bouzid, S. and Ramdani, M. (2013) 'Sensor fault detection and diagnosis in drinking water distribution networks', *8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), Mazafran-Algiers, Algeria*.

Branisavljević, N., Prodanović, D. and Pavlović, D. (2010) 'Automatic, semi-automatic and manual validation of urban drainage data', *Water Science & Technology*, vol. 62, no. 5, pp. 1013-1021.

Breiman, L. (1999) 'Bagging predictors', *Machine Learning*, vol. 24, no. 2, pp. 123–140.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, vol. 45, pp. 5-32.

Chapelle, O., Chang, Y. and Liu, T.Y. (2011) 'Yahoo! Learning-to-Rank Challenge', *JMLR W&CP*.

Cheng, X.B. and Wang, F.K. (2018) 'The performance of EWMA median and CUSUM median control charts for a normal process with measurement errors', *Quality and Reliability Engineering International*, vol. 34, no. 2, pp. 203-213.

Chen, and Huang, H. (2011) 'Immune feedforward neural network for fault detection', *Tsinghua Sience and Technology*, vol. 16, no. 3, pp. 272-277.

Chester, D., Lamb, D. and Dhurjati, P. (1984) 'Rule-based computer alarm analysis in chemical process plants', 7th Micro-Delcon, Newark, Delaware, 22-29.

Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosen, C. and Peter, A.V. (2010) 'Performance evaluation of fault detection methods for wastewater treatment processes', *Biotechnology and Bioengineering*, vol. 108, pp. 333-344.

Cover, T.M. and Hart, P.E. (1967) 'Nearest neighbor pattern classification', *IEEE Trans Inf Theory*, vol. 13, no. 1, pp. 21–27.

Crosier, R.B. (1988) 'Multivariate Generalizations of Cumulative Sum Quality Control Schemes', *Technometrics*, vol. 30, pp. 291- 303.

Dash, S., Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V. (2004) 'A novel interval-halving framework for automated identification of process trends', *AIChE Journal*, vol. 50, no. 1, pp. 149–162.

Dash, S., Rengaswamy, R. and Venkatasubramanian, V. (2003) 'Fuzzy-logic based trend classification for fault diagnosis of chemical processes', *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 347–362.

Dong, M., Cheng, T. and Chan, S. (2009) 'On-line fast motor fault diagnosis based on fuzzy neural networks', *Tsinghua Science and Technology*, vol. 14, no. 2, pp. 225-233.

Dror, G., Boulle, M., Guyon, I., Lemaire, V. and Vogel, D. (2009) 'Proceedings of KDD-Cup 2009 competition', JMLR Workshop and Conference Proceedings.

Dunia, R. and Qin, S.J. (1998) 'Subspace approach to multidimensional fault identification and reconstruction', *American Institute of Chemical Engineers Journal*, vol. 44, no. 8, pp. 1813-1831.

DWI (2016) *Water treatment processes*, 31 Aug, [Online], Available: http://dwi.defra.gov.uk/private-water-supply/installations/Treatment-processes.pdf.

Edthofer, F., van den Broeke, J., Ettl, J. and Weingartner, A. (2010) 'Reliable online water quality monitoring as basis for fault tolerant control', Nice, France.

Edwards, C., Spurgeon, S.K. and Patton, R.J. (2000) 'Sliding mode observers for fault detection and isolation', *Automatica*, vol. 36, pp. 541-553.

EIP Water Action Group (2015) 'Innovative monitoring technologies and EU Water legislation', *A White paper*.

EPA (2010) *Water quality event detection systems for drinking water contamination warning systems*, Cincinnati, OH: EPA.

EPA (2013) *Water quality event detection system challenge: methodology and findings*, Cincinnati, OH: EPA.

Ewan, W.D. (1963) 'When and How to Use CUSUM Charts', *Technometrics*, vol. 5, pp. 1-22.

Farkas, K. (2016) *CUSUM Anomaly Detection*, 28 Jun, [Online], Available: https://www.measurementlab.net/publications/CUSUMAnomalyDetection.pdf.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874.

Feigenbaum, E.A. (1982) *Knowledge engineering in the 1980s*, Stanford: Department of Computer Science, Stanford University.

Filbert, D. and Metzger, K. (1982) 'Quality test of systems by parameter estimation', 9th IMEKO Congress, Berlin.

Fisher, R.A. (1938) 'The statistical utilization of multiple measurements', *Annals of Eugenics*, vol. 8, pp. 376–386.

Frank, P.M. (1996) 'Analytical and qualitative model-based fault diagnosis - a survey and some new results', *European Journal of Control*, vol. 2, pp. 6-28.

Frank, P.M. and Ding, X. (1997) 'Survey of robust residual generation and evaluation methods in observer-based fault detection systems', *Journal of Process Control*, vol. 7, no. 6, pp. 403–424.

Frank, P.M. and Wuennenberg, J. (1989) 'Robust fault diagnosis using unknown input observer schemes', in Patton, R.J., Frank, P.M. and Clark, R.N. *Fault diagnosis in dynamic systems: theory and applications*, NY: Prentice Hall.

Freund, Y. and Schapire, R. (1997) 'A decision-theoretic generalization of on-line learning and an application to boosting', *Journal of Computer and System Sciences*, vol. 55, pp. 119–139.

Fukunaga, K. (1972) *Introduction to statistical pattern recognition*, New York: Academic press.

Fussell, J.B. (1974) 'Fault tree analysis - state of the art', *IEEE Transactions on Reliability*, vol. 23, no. 1, pp. 51-53.

Garcia-Alvarez, D., Fuente, M.J., Vega, P. and Sainz, G. (2009) 'Fault detection and diagnosis using multivariate statistical techniques in a wastewater treatment plant', 7th IFAC Symposium on Advanced Control of Chemical Processes, Istanbul, Turkey, 952-957.

George, J.P., Chen, Z. and Shaw, P. (2009) 'Fault detection of drinking water treatment process using PCA and Hotelling's T2 chart', *World Academy of Science, Engineering and Technology*, vol. 26, pp. 970-975.

Gertler, J. (1991) 'Analytical redundancy methods in fault detection and isolation', IFAC SAFERPROCESS Symposium, Baden-Baden, p. 91.

Gertler, J. and Singer, D. (1990) 'A new structural framework for parity equation-based failure detection and isolation', *Automatica*, vol. 26, pp. 381-388.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) 'Deep Feedforward Networks', in *Deep Learning*, MIT Press.

Government of Newfoundland & Labrador, Department of Environment and Conservation, Water Resources Management Division (2011) *Comparison of three turbitity instruments*, Government of Newfoundland & Labrador, Department of Environment and Conservation, Water Resources Management Division.

Grantham, S.D. and Ungar, L.H. (1990) 'A first principles approach to automated troubleshooting of chemical plants', *Computers and Chemical Engineering*, vol. 14, no. 7, pp. 783-798.

Grantham, S.D. and Ungar, L.H. (1991) 'Comparative analysis of qualitative models when the model changes', *American Institute of Chemical Engineers Journal*, vol. 37, no. 6, pp. 931-943.

Hach Homeland Security Technologies (2007) *GuardianBlue - early warning system*, Loveland, CO: Hach Homeland Security Technologies.

Hall, J. and Szabo, J. (2010) 'On-line water quality monitoring in drinking water distribution systems: a summary report of USEPA research and best practices', *Journal AWWA*, vol. 102, no. 8, pp. 20-22.

Hart, D.B. and McKenna, S.A. (2009) *CANARY user’s manual, version 4.1*, Cincinnati, OH: U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center.

Hart, D., McKenna, S.A., Klise, K., Cruz, V. and Wilson, M. (2007) 'CANARY: a water quality event detection algorithm development tool', World Environmental and Water Resources Congress, Reston, VA.

Haykin, S. (2009) *Neural Networks and Learning Machines*, 3rd edition, Upper Saddle River, New Jersey: Pearson Education, Inc.

Hempstalk, K., Frank, E. and Witten, I.H. (2008) 'One-Class Classification by Combining Density and Class Probability Estimation', Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Antwerp, pp. 505-519.

Henley, E.J. (1984) 'Application of expert systems to fault diagnosis', AIChE Annual Meeting, San Francisco.

Hérault, J. and Ans, B. (1994) 'Réseau de neurones à synapses modifiables : Décodage de messages sensoriels composites par apprentissage non supervisé et permanent', *Comptes Rendus de l'Académie des Sciences*, vol. III, no. 13, pp. 525–528.

Himmelblau, D.M. (1978) 'Fault detection and diagnosis in chemical and petrochemical processes', *Elsevier, Amsterdam*.

Holt, C.C. (1957) 'Forecasting, seasonals and trends by exponentially weighted moving averages', *Office of Naval Research (ONR 52)*.

Hotelling, H. (1947) 'Multivariate quality control', in Eisenhart, C., Hastay, M.W. and Wallis, W.A. *Selected techniques of statistical analysis for scientific and industrial research, and production and management engineering*, New York: McGraw-Hill.

Housh, M. and Ostfeld, A. (2015) 'An integrated logit model for contamination event detection in water distribution systems', *Water Research*, vol. 75, pp. 210-223.

Huang, G., Chen, Y. and Babri, H.A. (2000) 'Classification Ability of Single Hidden Layer Feedforward Neural Networks', *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 11, no. 3, pp. 799-801.

Hwang, I., Kim, S., Kim, Y. and Seah, C.E. (2010) 'A survey of fault detection, isolation, and reconfiguration methods', *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636-653.

Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M. and Sun, J. (2017) *Anomaly Detection for a Water Treatment System Using Unsupervised Machine Learning*, [Online], Available: https://arxiv.org/pdf/1709.05342.pdf.

Iri, M., Aoki, K., O'Shima, E. and Matsuyama, H. (1979) 'An algorithm for diagnosis of system failures in the chemical process', *Computers and Chemical Engineering*, vol. 3, no. 1-4, pp. 489-493.

Isermann, R. (2005) 'Model-based fault-detection and diagnosis – status and applications', *Annual Reviews in Control*, vol. 29, pp. 71-85.

Isermann, R. and Balle, P. (1997) 'Trends in the application of model-based fault detection and diagnosis of technical processes', *Control Eng. Practice*, vol. 5, no. 5, pp. 709-719.

Johansson, E.M., Dowla, E.U. and Goodman, D.M. (1991) 'Back-propagation learning for multi-layer feed-forward neural networks using the conjugate gradient method', *International Journal of Neural Systems*, vol. 2, no. 4, pp. 291-302.

Johnson, N.L. and Leone, F.C. (1962) 'Cumulatıve Sum Control Charts, Parts I, II & III', *Industruu Qııaiıtll Control*, vol. 18;19, pp. 15-21;29-36.

Kégl, B. (2013) 'The return of AdaBoost. MH: multi-class Hamming trees', *arXiv preprint arXiv:1312.6086*.

Khoukhi, A., Khalid, H., Doraiswami, R. and Cheded, L. (2012) 'Fault detection and classification using Kalman filter and hybrid neuro-fuzzy systems', *International Journal of Computer Applications*, vol. 45, no. 22, pp. 7-14.

Klise, K.A. and McKenna, S.A. (2006) 'Water quality change detection: multivariate algorithms.', SPIE, (International Society for Optical Engineering), Defense and Security Symposium 2006, Orlando. Florida, 9pp.

Kohavi, R. (1995) 'A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection', Montreal, Canada.

Kohavi, R. and John, G.H. (1997) 'Wrappers for feature subset selection', *Artificial Intelligence* , vol. 97, pp. 273-324.

Kohonen, T. (1995) *Self-organizing Maps*, 3rd edition, Berlin: Springer-Verlag Berlin Heidelberg.

Kramer, M.A. and Palowitch, B.L. (1987) 'A rule based approach to fault diagnosis using the signed directed graph', *American Institute of Chemical Engineers Journal*, vol. 33, no. 7, pp. 1067-1078.

Kresta, J.V., MacGregor, J.F. and Marlin, T.E. (1991) 'Multivariate statistical monitoring of process operating performance', *The Canadian Journal of Chemical Engineering*, vol. 69, pp. 35-47.

Lamrini, B., Lakhal, E.K. and Le Lann, M.V. (2013) 'A decision support tool for technical processes optimization in drinking water treatment', *Desalination and Water Treatment*, vol. 52, pp. 4079–4088.

Lennox, B., Montague, G.A., Frith, A.M., Gent, C. and Bevan, V. (2001) 'Industrial application of neural networks - an investigation', *Journal of Process Control*, vol. 11, pp. 497-507.

Liu, S., Che, H., Smith, K. and Chen, C. (2015) 'A method of detecting contamination events using multiple conventional water quality sensors', *Environ Monit Assess*, vol. 187, p. 4189.

Liu, S., Smith, K. and Che, H. (2015a) 'A multivariate based event detection method and performance comparison with two baseline methods', *Water Research*, vol. 80, pp. 109-118.

Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) 'A multivariate exponentially weighted moving average control chart', *Technometrics*, vol. 34, no. 1, pp. 46-53.

Lucas, J.M. and Crosier, R.B. (1982) 'Fast Ininal Response for Cusum Quality-Control Schemes: Give Your Cusum a Head Start', *Technometrics*, vol. 24, pp. 199-206.

Madsen, J.H. (2018) *DDoutlier: Distance and Density-Based Outlier Detection*, [Online], Available: https://CRAN.R-project.org/package=DDoutlier.

Maiti, J. and Banerjee, R.N. (2012) 'Process monitoring and fault detection strategies: a review', *International Journal of Quality & Reliability Management*, vol. 29, no. 7, pp. 720-752.

Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V. (2005) 'Fault diagnosis by qualitative trend analysis of the principle components', *Chemical Engineering Research and Design*, vol. 83, no. A9, pp. 1122–1132.

Maurya, M., Rengaswamy, R. and Venkatasubramanian, V. (2007) 'A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis', *Chemical Engineering Research and Design*, vol. 85, no. 10, pp. 1407-1422.

McKenna, S..A., Hart, D., Klise, K., V., C. and Wilson, M. (2007) 'Event Detection from Water Quality Time Series', World Environmental and Water Resources Congress 2007, Tampa, Florida, 1-12.

Meyers, G., Kapelan, Z. and Keedwell, E. (2017) 'Data-driven Approach to Short-Term Forecasting of Turbidity in a Trunk Main Network', CCWI 2017 – Computing and Control for the Water Industry, Sheffield, U.K.

Miljkovic, D. (2011) 'Fault detection methods: A literature survey', Proceedings of the 34th International Convention MIPRO 2011, Opatija, Croatia, 750 - 755.

Moller, M.E. (1993) 'A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning', *Neural Networks*, vol. 6, pp. 525-533.

Montgomery, D.C. (2009) *Introduction to statistical quality control*, 6th edition, John Wiley and Sons.

Morgan, J.N. and Sonquist, J.A. (1963) 'Problems in the analysis of survey data, and a proposal', *J. Amer. Statist. Assoc.*, vol. 58, pp. 415–434.

Mounce, S.R., Boxall, J.B. and Machell, J. (2010) 'Development and verification of an online artificial intelligence system for burst detection in water distribution systems', *Journal of Water Resources Planning and Management*, vol. 136, no. 3, pp. 309-318.

Mounce, S.R., Machel, J.M. and Boxall, J.B. (2012) 'Water quality event detection and customer complaint clustering analysis in distribution systems', *Water Science and Technology: Water Supply*, vol. 12, no. 5, pp. 580-587.

Murray, R., Haxton, T., McKenna, S.A., Hart, D.B., Klise, K., Koch, M., Vugrin, E.D., Martin, S., Wilson, M., Cruz, V. and Cutler, L. (2010) *Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems: Development, Testing, and Application of CANARY*, 60010036th edition, Cincinnati, OH: U.S. Environmental Protection Agency.

Nidsunkid, S., Borkowski, J.J. and Budsaba, K. (2017) 'The effects of violations of the multivariate normality assumption in multivariate Shewhart and MEWMA control charts', *Thailand Statistician*, vol. 33, no. 8, pp. 2563-2576.

Oliker, N. and Ostfeld, A. (2014) 'A coupled classification - evolutionary optimazation model for contamination event detection in water distribution systems', vol. 51, pp. 234-245.

Oliker, N. and Ostfeld, A. (2015) 'Networks hydraulics inclusion in water quality event detection using multiple sensor stations data', *Water Research*, vol. 80, pp. 47-58.

OptiWater (2018), [Online], Available: http://www.optiwater.com/optieds/.

Padhee, S., Gupta, N. and Kaur, G. (2012) 'Data driven multivariate technique for fault detection of waste water treatment plant', *International Journal of Engineering and Advanced Technology*, vol. 1, no. 4, pp. 45-50.

Page, E.S. (1954) 'Continuous inspection schemes', *Biometrika*, vol. 41, pp. 100-115.

Page, R.M., Waldmann, D. and Gahr, A. (2017) 'Online Water-Quality Monitoring based on Pattern Analysis', CCWI 2017 – Computing and Control for the Water Industry, Sheffield, U.K.

Parsons, S.A. and Jefferson, B. (2006) 'Introduction to potable water treatment processes' Oxford, UK.

Pearson, K. (1901) 'On lines and planes of closest fit to systems of points in space', *Philosophical Magazine*, vol. 2, pp. 559-572, Available: http://pbil.univ-lyon1.fr/R/pearson1901.pdf.

Perelman, L., Arad, J., Housh, M. and Ostfeld, A. (2012) 'Event detection in water distribution systems from multivariate water quality time series', *Environmental Science & Technology*, vol. 46, pp. 8212-8219.

Pichumani, R. (1997) *Survey of Current Techniques - Limitations of thresholding*, 07 Jul, [Online], Available: http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/RAMANI1/node21.html.

Piciaccia, L., Croce, D., Basili, R., Pettersen, J. and Ryfors, P. (2018) 'A Data-driven Approach for Optimal Control parameters in WWTP: the VEAS Experience in Scandinavia', HIC 2018. 13th International Conference on Hydroinformatics, Palermo, Italy.

Pignatiello, J.J. and Runger, G.C. (1990) 'Comparisons of multivariate cusum charts', *Journal of Quality Technology*, vol. 22, no. 3, pp. 173–186.

Praus, P. (2005) 'Water quality assessment using SVD-based principal component analysis of hydrological data', *Water SA*, vol. 31, no. 4, pp. 417-422.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2007), in *Numerical Recipes: The Art of Scientific Computing*, Third Edition edition, New York: Cambridge University Press.

Pugh, R. (2013) *Real-time water quality monitoring program - pilot project report*, Government of Newfoundland & Labrador; Department of Environmental and Conservation; Water Resources Management Division.

Puig, V. and Quevedo, J. (2002) 'Passive robust fault detection using fuzzy parity equations', *Mathematics and Computers in Simulation,* vol. 60, pp. 193-207.

Punal, A., Roca, E. and Lema, J.M. (2002) 'An expert system for monitoring and diagnosis of anaerobic wastewater treatment plants', *Water Research*, vol. 36, pp. 2656-2666.

Quin, S.J. (2003) 'Statistical process monitoring: basics and beyond', *Journal of Chemometrics*, vol. 17, pp. 480-502.

Raghuraj, R., Bhushan, M. and Rengaswamy, R. (1999) ' Location of sensors in complex chemical plants based on fault diagnostic observability criteria', *AIChE Journal*, vol. 45, pp. 310-322.

Rahman, A.A., Yahaya, S.S.S. and Atta, A.M.A. (2018) 'The eﬀect of median based estimators on CUSUM chart', *Journal of Telecommunication*, vol. 10, no. 1-10, pp. 49–52.

Rasmussen, N.C. (1975) *Reactor safety study. An assessment of accident risks in U. S. commercial nuclear power plants. Executive summary: main report*, Washington, D.C., USA: Nuclear Regulatory Commission.

Rasmussen, J. (1986) *Information processing and human-machine interaction: an approach to cognitive engineering*, New York: North Holland.

Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P.A. and Comeau, Y. (2010) 'Data reconciliation for wastewater treatment plant simulation studies - planning for high-quality data and typical sources of errors', *Water Environment Research*, vol. 82, no. 5, pp. 426-433.

Rieger, L. and Vanrolleghem, P.A. (2008) 'monEAU: a platform for water quality monitoring networks', *Water Science & Technology - WST*, vol. 57, no. 7, pp. 1079-1086.

Ritchie, J.C., Zimba, P.V. and Everitt, J.H. (2003) 'Remote sensing techniques to assess water quality', *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 6, pp. 695-704.

Roberts, S.W. (1959) 'Control chart tests based on geometric moving averages', *Technometrics*, vol. 1, no. 3, pp. 239-250.

Romano, M., Kapelan, Z. and Savic, D. (2014) 'Automated detection of pipe bursts and other events in water distribution systems', *Journal of Water Resources Planning and Management*, vol. 140, no. 4, pp. 457-467.

Rosen, J.S. and Bertrand, T. (2013) 'Using online water quality data to detect events in a distribution system', *Journal AWWA*, vol. 105, no. 7, pp. 22, 24-26.

Rosen, C. and Lennox, J.A. (2001) 'Multivariate and multiscale monitoring of wastewater treatment operation', *Water Research*, vol. 35, pp. 3402-3410.

Ruiz, M., Colomer, J. and Melendez, J. (2007) 'Multiway principle component analysis and case-based reasoning approach to situation assessement in a waste water treatment plant', European Control Conference, Kos, Greece, 5354-5360.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) 'Learning internal representations by error propagation', in Rumelhart, D.E. and McClelland, J.L. *Parallel distributed processing. Exploration in the microstructure of cognition*, Cambridge: MIT Press.

Russell, E.L. and Braatz, R.D. (2000b) *Data-driven techniques for fault detection and diagnosis in chemical processes*, Springer, New York.

s:can (2013) *Event detection from a practical view*, Vienna, Austria: s:can.

SAFEWATER (2015) 'SAFEWATER - innovative tools for the detection and mitigation of CBRN related contamination events of drinking water', Antibes - Juan-les-Pins, France.

Sahri, Z. and Yusof, R. (2014) 'Support vector machine-based fault diagnosis of power transformer using k-nearest-neighbor imputed DGA dataset', *Journal of Computer and Communications*, vol. 2, no. 9, pp. 22-31.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004) *Sensitivity analysis in practice*, Chichester: John Wiley & Sons.

Sandri, M. and Zuccolotto, P. (2009) 'Variable Selection Using Random Forests', pp. 1-8.

Sarrate, R., Nejjari, F. and Rosich, A. (2012) 'Sensor placement for fault diagnosis performance maximization in distribution networks', *Mediterranean Conference on Control & Automation (MED), Barcelona, Spain*.

Schölkopf, B., Smola, A.J. and Bach, F. (2002) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Cambridge: MIT Press.

Schraa, O., Tole, B. and Copp, J.B. (2006) 'Fault detection for control of wastewater treatment plants', *Water Science & Technology*, vol. 53, no. 4, pp. 375–382.

Severson, K., Chaiwatanodom, P. and Braatz, R.D. (2015) 'Perspectives on process monitoring of industrial systems', *IFAC-PapersOnLine 48-21*, pp. 931–939.

Shaocheng, T., Bin, C. and Yongfu, W. (2005) 'Fuzzy adaptive output feedback control for MIMO nonlinear systems', *Fuzzy Sets and Systems*, vol. 156, pp. 285-299.

Sheibat-Othman, N., Laouti, N., Valour, J. and Othman, S. (2014) 'Support vector machines combined to observers for fault diagnosis in chemical reactors', *The Canadian Journal of Chemical Engineering*, vol. 92, pp. 685-694.

Shewhart, W.A. (1931) *Economic control of quality of manufactured product*, New York: Van Nostrand.

Sin, G., Ghosh, K., Natarajan, S., Srinivasan, R., Adhitya, A., Karimi, I.A., Papadokonstantakis, S., Hungerbühler, K. and Angelo, P. (2012) '7. Abnormal events management and process safety', Process Systems Engineering, 1-41.

Sobhani-Tehrani, E. and Khorasani, K. (2009) *Fault detection and diagnosis*, LLC: Springer Science+Business Media.

Sokoowski, A. (2004) 'On some aspects of fuzzy logic application in machine monitoring and diagnostics', *Engineering Applications of Artificial Intelligence*, vol. 17, pp. 429-437.

Spellman, F. (2003) *Handbook of water and wastewater treatment plant operations*, Boca Raton, Florida: CRC Press LLC.

Srivastava, N.P., Srivastava, R.K. and Vashishtha, P.K. (2014) 'Fault detection and isolation (Fdi) via neural networks', *Journal of Engineering Research and Applications*, vol. 4, no. 1, pp. 81-86.

Storey, M.V., van der Gaag, B. and Burns, B.P. (2011) 'Advances in on-line drinking water quality monitoring and early warning systems', *Water Research* , vol. 45, no. 2, pp. 741–747.

Szabo, J. and Hall, J. (2014) 'On-line water quality monitoring for drinking water contamination', *Comprehensive Water Quality and Purification*, vol. 2, pp. 266-282.

Talagala, P.D., Hyndman, R.J., Leigh, C., Mengersen, K. and Smith-Miles, K. (2019) 'A feature-based framework for detecting technical outliers in water-quality data from in situ sensors', *JEL classification*, p. Submitted on 17 Feb 2019.

Tumula, D. and Danso-Amoako, E. (2014) 'Influence of chemical and biological parameters on iron and manganese accumulation in water distribution networks', 12th International Conference on Computing and Control for the Water Industry, Perugia, Italy, 1353 – 1361.

Ulerich, N.H. and Powers, G.A. (1988) 'Online hazard aversion and fault diagnosis in chemical processes: the digraph+fault tree method', *IEEE Transactions on Reliability*, vol. 37, no. 2, pp. 171-177.

United Nations (2015) 'World water development report 2015'.

USEPA (2014) 'Configuring Online Monitoring Event Detection Systems', in *Report EPA/600/R-14/254*, Washington, D.C: Agency, U.S. Environmental Protection.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N. and Yin, K. (2003a) 'A review of process fault detection and diagnosis Part I: Quantitative model-based methods', *Computers and Chemical Engineering*, vol. 27, pp. 293-311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N. and Yin, K. (2003b) 'A review of process fault detection and diagnosis Part II: Qualitative models and search strategies', *Computers and Chemical Engineering*, vol. 27, pp. 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N. and Yin, K. (2003c) 'A review of process fault detection and diagnosis Part III: Process history based methods', *Computers and Chemical Engineering*, vol. 27, pp. 327-346.

Verron, S., Tiplica, T. and Kobi, A. (2008) 'Fault detection and identification with a new feature selection based on mutual information', *Journal of Process Control*, vol. 18, no. 5, pp. 479-490.

Weickgenannt, M., Kapelan, Z., Blokker, M. and Savic, D.A. (2010) 'Risk-based sensor placement for contaminant detection in water distribution systems', *Journal of Water Resources Planning and Management*, pp. 629-636.

WhitewaterSecurity (2018), [Online], Available: http://w-water.com.

WHO (2012) 'Water quality and health strategy 2013-2020'.

WHO/UNICEF (2015) 'Progress on sanitation and drinking water - update and MDG assessement'.

Widodo, A. and Yang, B. (2007) 'Support vector machine in machine condition monitoring and fault diagnosis', *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574.

Wilkinson, L. (2018) 'Visualizing Big Data Outliers through Distributed Aggregation', *IEEE transactions on visualization and computer graphics*, vol. 24, no. 1, pp. 256–266.

Winters, P.R. (1960) 'Forecasting sales by exponentially weighted moving averages', *Management Science*, vol. 6, pp. 324-342.

Wold, H. (1966) 'Estimation of principal components and related models by iterative least squares', in Krishnaiah, P.R. *Multivariate Analysis*, New York: Academic Press.

Woodall, W.H. and Ncube, M.M. (1985) 'Multivariate cusum quality-control procedures', *Technometrics*, vol. 27, no. 3, pp. 285–292.

Yang, L., Pai, S. and Wang, Y. (2010) 'A Novel CUSUM Median Control Chart', International MultiConference of Engineers and Computer Scientists (IMECS), Hongkong, 1707-1710.

Ye, G. and Fenner, R.A. (2011) 'Kalman filtering of hydraulic measurements for burst detection in water distribution systems', *Journal of Pipeline Systems Engineering and Practice*, vol. 2, no. 1, pp. 14-22.

Yin, .S., Ding, S.X., Haghani, A., Hao, H. and Zhang, P. (2012) 'A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process', *Journal of Process Control*, vol. 22, pp. 1567-1581.

Yin, S., Gao, X., Karimi, H. and Zhu, X. (2014) 'Study on support vector machine-based fault detection in Tennessee Eastman process', *Abstract and Applied Analysis*, vol. 2014, pp. 1-8, Available: Article ID 836895.

Yoo, C.K., Villez, K., Van Hulle, S.W. and Vanrolleghem, P.A. (2008) 'Enhanced process monitoring for wastewater treatment systems', *Envirometrics*, vol. 19, no. 6, pp. 602–617.

Zheng, Y.W., Yekun, H. and Qiao, L. (2018) 'Comparing deep learning with statistical control methods for anomaly detection', 1st International WDSA / CCWI 2018 Joint Conference, Ontario, Canada.

Zhiwei, G., Cecati, C. and Ding, S.X. (2015) 'A survey of fault diagnosis and fault-tolerant techniques Part II: fault diagnosis with knowledge-based and hybrid/active approaches', *IEEE Transactions on Industrial Electronics*.