Université de Lille

&

Université Libanaise

Thèse présentée par :

Christine SAAB

Pour obtenir le grade de **Docteur en Génie Civil**

SMART SYSTEM FOR WATER QUALITY CONTROL: FEEDBACK FROM LARGE-SCALE EXPERIMENTATION

Soutenue le 2 Juillet 2018 devant le jury composé de:

Isam SHAHROUR	Professeur, Université de Lille	Directeur
Fadi HAGE CHEHADE	Professeur, Université Libanaise	Directeur
Fadi COMAIR	Professeur, DG, Ministère de l'Energie et de l'Eau, Liban	Rapporteur
Azzedine HANI	Professeur, Université d'Annaba	Rapporteur
Sawsan SADEK	Professeur, Université Libanaise	Examinateur
Aziz SOULHI	Professeur, Ecole Nationale Supérieure des Mines, Rabat	Examinateur

Acknowledgments

To deserve the Doctor degree, hard work, great motivation as well as support and encouragement are needed. In these few lines, I would like to express my sincere appreciations to all my surroundings who contribute to the achievement of this PhD thesis.

First, I want to express my high gratitude to my supervisor in France, Pr. Isam Shahrour for all the helps, advices and significant guidance. It was a great honor to make this work under his supervision and to benefit from his large experience. Also, I want to thank my supervisor in Lebanon, Pr. Fadi Hage Chehade who gave me this opportunity to do a co-tutorial thesis between France and Lebanon.

I would like also to thank all my committee members, Pr. Fadi Comair, Pr. Azzedine Hani, Pr. Sawsan Sadek and Pr. Aziz Soulhi, for accepting to review and evaluate my thesis work.

I am thankful also to all my friends in the Laboratory of Civil Engineering and geo-Environmental (LGCgE) for their encouragements and all the beautiful moments we have shared together.

A special thanks to my dearest friend Elias Farah for his continuous helping and encouraging. Your support was very meaningful to me.

Last but not least, I would like to express my biggest acknowledgments to my beloved family. Thank you my wonderful sisters, Juliana, Joelle and Aline to all your encouragements that motivate me indefinitely. My deepest thanks goes to my father Melhem and my mother Wadad for whom this work is dedicated. Thank you my lovely parents for all the sacrifices you made. I will always be grateful for your unconditional love that blessed me in my whole life.

Abstract

This works presents the real-time control of drinking water quality using the smart technology. The deployment of water quality sensors in the distribution networks provides indication of contamination risks. However, the use of these innovative devices is recent and yet requires field experimentations. This thesis concerns this issue and enhances the feedback in this domain. It presents a field study of online supervision of water quality, which was conducted at the Scientific Campus of Lille University, within SunRise project. This work is also a part of the European project "SmartWater4Europe" which consists in developing 4 demonstrator sites for the smart water management.

This thesis includes five chapters. The literature review highlights the impact of water contamination on human health as well as the drawbacks of conventional water supervision methods. A large-scale experimentation is conducted at Lille University, where two types of sensors (S::CAN and EventLab) are implemented at two different locations. The detailed analysis of recorded water quality signals showed the occurrence of some events, generally correlated with the variation of hydraulic parameters or the network interventions.

Different methodologies for the detection of water anomaly are presented and applied to S::CAN data. Statistical and Artificial Intelligence (Support Vector Machine) methods discriminate between normal and unexpected measurements. An Event Detection System (EDS) has been developed within Canary software. It showed a good performance in the identification of water abnormalities recorded by S::CAN. The last part proposes a combination between the risk assessment approach and the smart monitoring. The improved risk assessment methodology allows a real-time detection and classification of water anomaly risk as well as an identification of the priority attention required.

Keywords: drinking water, distribution network, quality, smart technology, sensors, online-monitoring, contamination, field study.

Résumé

Ce travail présente le contrôle en temps réel de la qualité de l'eau potable par la technologie intelligente. Le déploiement des capteurs de qualité de l'eau dans les réseaux de distribution fournit une indication des risques de contamination. Cependant, l'utilisation de ces dispositifs innovants est récente et nécessite encore des expérimentations. Cette thèse concerne cette problématique; elle vise à améliorer le retour d'expérience dans ce domaine. Elle présente la supervision en ligne de la qualité de l'eau sur le Campus « Cité Scientifique » de l'Université de Lille, qui est réalisée dans le cadre du projet SunRise. Ce travail fait également partie du projet Européen "SmartWater4Europe" qui vise à développer 4 sites démonstrateurs pour la gestion intelligente de l'eau.

Cette thèse comporte cinq chapitres. L'étude bibliographique met en évidence l'impact de la contamination de l'eau sur la santé humaine ainsi que les inconvénients des méthodes conventionnelles de la surveillance de la qualité de l'eau. Une expérimentation à grande échelle est menée à l'Université de Lille, où deux types de capteurs (S::CAN and EventLab) sont implémentés dans deux endroits différents. L'analyse des signaux enregistrés a montré l'occurrence de certains évènements, généralement corrélés avec la variation des paramètres hydrauliques ou des interventions sur le réseau.

Différentes méthodologies pour la détection d'anomalie de l'eau sont présentées et appliquées aux données S::CAN. Les méthodes Statistiques et de l'Intelligence Artificielle (Machine à Vecteurs de Support) distinguent entre les mesures normales et celles inattendues. Un Système de Détection des Evènements (SDE) a été développé en utilisant le logiciel Canary. Il a montré une bonne performance dans l'identification des anomalies de l'eau enregistrées par S::CAN. La dernière partie propose une combinaison entre l'approche « Évaluation de risques » et « surveillance intelligente ». La méthode d'évaluation des risques développée permet une détection et une classification, en temps réel, du risque d'anomalie de l'eau, ainsi qu'une identification de la priorité d'attention requise.

Mots-clés: eau potable, réseau de distribution, qualité, technologie intelligente, capteurs, surveillance, contamination, in situ.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAS	Atomic Absorption Spectroscopy
ACC	Accuracy
ADWG	Australian Drinking Water Guidelines
AF	Asset Framework
AFSSA	Agence Francaise de Sécurité Sanitaire des Aliments
AI	Artificial Intelligence
AMR	Automatic Meters Reader
ANN	Artificial Neural Network
AR	Autocorrelation
AT	Averaging Time
ATP	Adenosine Triphosphate
BED	Binomial Event Discriminator
BHAA	Aerobic and Anaerobic Heterotrophic Bacteria
BW	Body Weight
CIRE	Centre Interrégional d’Epidémiologie
CSHPF	Conseil Supérieur d'Hygiène Publique de France
CSV	Comma-Separated Value
CW	Concentration in Water of chemical of concern
CWS	Contamination Warning System
DDASS	Directions Départementales des Affaires Sanitaires et Sociales
DMA	Dose Maximale Admissible
DOC	Dissolved Organic Carbon
E.Coli	Escherichia Coli
ECL	Ecole Centrale de Lille
ED	Exposure Duration
EDS	Event Detection System
ELISA	Enzyme-Linked Immunosorbent Assays
ENSCL	Ecole Nationale Supérieure de Chimie de Lille
EPA	Environmental Protection Agency
EU	European Union
EWS	Early Warning System
EF	Exposure Frequency
FISH	Fluorescent In Situ Hybridization
FN	False Negative
FNU	Formazin Nephelometric Units
FP	False Positive
FRR	False Positive Rate
FT-IR	Fourier Transform Infrared Spectroscopy
FTU	Formazin Turbidity Units
GEA	Acute Gastroenteritis
GPRS	General Packet Radio Service
H.P	High Priority
HRV	Human Rotavirus
I.P	Intermediate Priority
ICT	Information and Communication Technology
IFSTAR	Institut Français des Sciences et Technologies des Transports, de l’Aménagement et des Réseaux
IM-ECL	Immunomagnetic Electrochemiluminiscence
IMS	Immunomagnetic Separation
IMS-FCM	Immunomagnetic Separation and Flow Cytometry Method

INSPQ	Institut National de santé publique de Québec
IR	Ingestion Rate
ISO	International Standards Organization
IT	Information Technology
IP	Internet Protocol
L.P	Low Priority
LGCgE	Laboratory of Civil Engineering and geo Environmental
Lp	Legionella pneumophila
LPC	Linear Predictive Coding
LPCF	Linear Prediction Correction Filter
MF	Membrane Filtration
MLB	Minimal Lactose Enrichment Broth
MPN	Most Probable Number
MVNN	Multivariate Nearest-Neighbor
MZI	Mach-Zehnder Interferometer
NGS	Next-Generation Sequencing
NHSRC	National Homeland Security Research Center
NIR-CI	Near-Infrared and Chemical Imaging
NIRS	Near-Infrared Spectroscopy
NTU	Nephelometric Turbidity Units
P	Priority attention Score
PCR	polymerasechainreaction
ppm	parts per million
PPV	Positive Predictive Value
PVC	Polyvinyl Chloride
QCRA	Quantitative Chemical Risk Assessment
QMRA	Quantitative Microbial Risk Assessment
qPCR	quantitaitve polymerasechainreaction
qRT-PCR	quantitative reverse-transcriptase polymerasechainreaction
R	Risk Score
RI	Refractive Index
ROC	Receiver Operating Characteristics
RTD	Resistance Temperature Detector
S	Severity Score
SCADA	Supervisory Control and Data Acquisition
SMaRT-Online^{WDN}	Online Security Management and Reliability Toolkit for Water distribution Networks
SNT	Serum Neutralization Test
SUNRISE	Smart Urban Networks for Resilient Infrastructure and Sustainable Ecosystems
SV	Support Vectors
SVM	Support Vector Machine
SW4EU	Smart Water for Europe
THM	Trihalomethanes
TN	True Negative
TNR	True Negative Rate
TOC	Total Organic Carbon
TP	True Positive
TPR	True Positive Rate
U.P	Urgent Priority Attention
UFC	Units Forming Colony
UNICEF	United Nations International Children’s Emergency Fund
VBA	Visual Basic for Applications
VPN	Virtual Private Network
W	Weight
WDN	Water Distribution Network
WDS	Water Distribution System
WHO	World Health Organization

Introduction

Drinking water is a crucial resource for health and well-being of human. According to the *World Health Organization (WHO)*[1], approximately 844 million people lack access to safe water. Due to the pressure of increasing population, aging infrastructure and limited resources, the challenge of sustainable water-quality management constitutes a great concern for water utilities.

Although water quality is well controlled in treatment plant, this is not the case in water distribution networks (WDNs). WDNs are not inert transport systems since various physico-chemical and biological interactions could occur. In addition to malicious attacks, accidental contamination from incidents (backflow, network interventions, contaminant penetration, etc.) could degrade the water quality in the distribution networks. The consumption of contaminated drinking water can transmit dangerous agents and can threat the human health. Based on the *WHO*[2] guidelines, the most predominant waterborne disease, diarrhea, has an estimated annual incidence of 4.6 billion episodes and causes 2.2 million deaths every year. Thus, providing safe drinking water to consumers constitutes the main purpose of water supply.

The water quality supervision is generally done using conventional methods, such as Membrane Filtration, Immunological detection, chemical and aesthetic analyses, etc. These methods are based on taking manual samples in a periodic basis from different locations in the water system, followed by laboratory analyses. Results from laboratory are compared with Standards to determine the organoleptic characteristics of the water as well as the presence of microorganisms. Although laboratory-based methods allow a detailed analysis of water characteristics, they present some limitations such as: i) long time to obtain results and thus long delay to take corrective measures and ii) economic issues (equipment, intensive labor, operations, etc.). Therefore, there is an urgent need for a real-time supervision of the water quality in distribution system to protect early the public health from harmful impacts of contamination. On the other hand, a global annual reduction of more than $120 million can be ensured by moving from manual sampling to online monitoring according to *Sensus*, *Water 20/20*[3].

Nowadays, the Information and Communication Technology (ICT) is used in several urban systems within the concept of sustainable and smart city. The integration of ICT in water distribution system enhances the security of water network. Therefore, WDN should be turned into smart water system through the use of smart technology, in order to ensure the online control of the water quality.

The deployment of water quality sensors is an essential requisite in smart water networks. However, compound sensors measuring specific agent are not efficient in real water networks

[1] WHO/UNICEF, Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines, *World Health Organization*, Geneva, 2017.

[2] WHO, author. Water for health, WHO Guidelines for Drinking-water Quality. Geneva, Switzerland: WHO, 2010.

[3] Sensus, Water 20/20: Bringing Smart Water Networks into Focus, *Report by Sensus*, 2012.

since the contaminant type is not known in advance. The use of conventional sensors that measure continuously surrogate parameters (such as pH, Turbidity, Temperature, etc.) is more reliable. Studies have demonstrated the effect of the presence of contaminant in water on the change in the water quality parameters.

The real-time access to sensors' data allows a continuous analysis of the drinking water quality. The variation of the water quality signals provides indication of possible abnormalities in the water system. It will help in the decision-making of water utilities to reduce the contamination risk. However, it is still challenged to differ between normal variations and those due to water contamination.

The use of the smart technology in the online monitoring of water quality is recent; the feedback is very restricted in this domain. In general, the reliability of sensors is tested in pilot-scale systems where contaminant injection can be controlled. In laboratory station, reference lines can be established for normal drinking water. The deviation of signals can be then analyzed after contaminant injection. However, real water networks are considered as extensive and complex systems usually affected by several factors (hydraulic conditions, corrosion, stagnation, etc.). The performance of sensors should be evaluated in real condition to test their efficiency in water anomaly detection. Various reasons (contamination, connection issues, sensors faults, etc.) can induce a variation in the water quality signals in real site. There is a clear need for a field study for sensors implementation at large-scale.

The aim of this work is to analyze the online water quality control using the smart technology. This study is based on a large scale experimentation for the real-time monitoring of water quality. Within SunRise project, the water network of the Scientific Campus of Lille University is used as a field study for sensors instrumentation. The campus stands for a small town hosting around 25000 users in 150 buildings. Two types of water quality sensors: S::CAN and EventLab are installed at two locations in the campus.

This work is also a part of the European project SmartWater4Europe "SW4EU" which consists in developing and demonstrating 12 innovative solutions in smart water management (water quality management, leak management, energy optimization and customer interaction) at 4 demonstration sites in Europe (France, Netherlands, Spain, United-Kingdom).

This thesis includes five chapters and it is divided in two main sections. In the first part, the identification of abnormalities is based on the analysis of deviations observed in the water quality signals, while the second part (fourth and fifth chapters) proposes different methodologies for the early detection of water anomaly.

The first chapter presents a literature review. It describes a list of possible contaminants in the water system, their guidelines, and their health impacts as well as conventional techniques (laboratory-based methods) used for water quality control. This bibliography also details the use of indicators for contamination detection. It introduces the concept of the smart water network for the early identification of water abnormalities.

The second chapter presents the demonstration site as well as the water distribution network of the University of Lille. It describes two types of water quality sensors (S::CAN and EventLab) used in this work with their installation in the campus. S::CAN measures several water quality parameters (Turbidity, Conductivity, Chlorine, etc.) each minute, while EventLab controls continuously the variation of refractive index in water.

The third chapter presents a detailed analysis of the water quality at Lille demo site. It presents the continuous monitoring of the water quality signals. Different abnormalities occurring at critical periods, are also detailed with the possible reason of each event. A comparison between sensor's responses is also presented in this chapter.

The fourth chapter details the use of Statistical (linear predictive coding) and Artificial Intelligence (Support Vector Machine) methods for the identification of unexpected sensor data. This chapter also presents the use of Event Detection System (EDS) approach for the early detection of water quality anomaly. The efficiency of these methods is evaluated through their application to S::CAN measurements.

The fifth chapter describes the water quality control using the qualitative risk assessment method. Based on a combination between the online monitoring and the risk assessment concept, two main approaches are developed to detect in near real-time abnormal events in water and to rank their risk level. A detailed description of each approach is presented. This chapter includes also a comparison between the proposed methodologies for the water anomaly detection.

Chapter 1. State of the Art – Control of Drinking Water Quality

1.1 Introduction

This chapter presents a state of the art of the water quality control in distribution networks. Standards for water quality parameters will be described. Different cases of water contamination with their health impacts will be exposed. A literature review shows the conventional techniques (laboratory-based methods) used in the water quality control. The problem of the long delay of these methods demonstrates the need for innovation in this domain. To meet this objective, the role of the Smart Technology in real demonstration site will be detailed.

1.2 Water Distribution Network

Water Distribution Networks (WDNs) are designed to ensure water demands such as domestic, industrial, fire-fighting, etc. The main objective of Water Distribution System (WDS) is to provide safe water in terms of quality and quantity. Different hydraulic elements constitute the distribution system, such as pipes, hydrants, pumping stations, meters, etc. The system can be divided in three main classes: i) branching with dead ends, ii) grid where any point can be supplied from at least two directions and iii) combination of the two preceding systems (branching & grid) in case of sharp variation in topography.

Distribution systems should provide adequate and reliable water to the customer; adequate means providing all the water the customer needs for quality, at a pressure no less than 20 psi (1.4 bars); reliable means that customers can expect to obtain all the water they need, anytime they need it [1].

1.3 Drinking water quality

The main purpose of WDS is to deliver a safe drinking water to users. However, WDS can be subjected to accidental or malicious attacks which can degrade the quality of water. High quality water leaving treatment facilities could deteriorate as it travels through extensive, often convoluted, distribution networks, via a number of mechanisms associated with distribution network materials, hydraulic conditions, chemical and biological reactions, or ingress of polluting materials [2]. The impact of drinking water contamination on human health can be very dangerous. Anomalous in water quality can lead to serious diseases, or to death in very critical cases. The water quality supervision is of high importance for water utilities to protect the human health. Figure 1.1 shows the WDS as a complex reactor where physico-chemical and biological interactions occur.

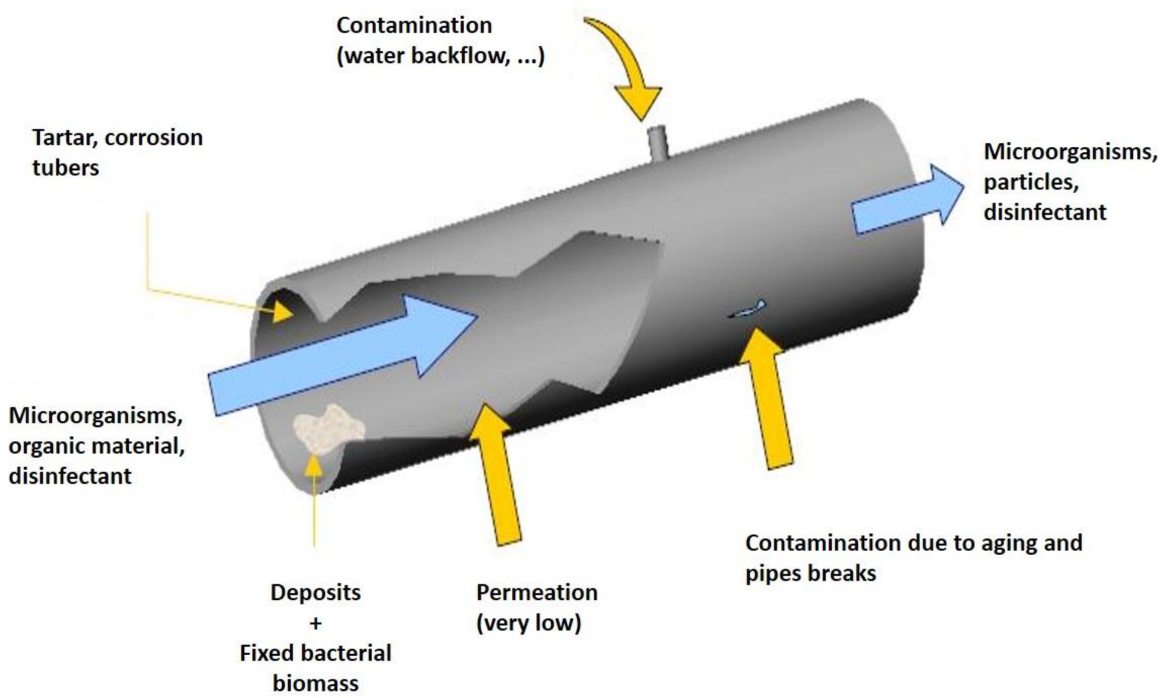


Figure 1.1. Schema of reactor water network [3].

To prevent any anomaly in the distribution network, many parameters have to be monitored. They should be compared with Standard thresholds. These parameters should constitute a good indicator of possible intrusion in the water system. The ideal indicator should be [4]:

- Non-pathogenic.
- Always present when pathogenic viruses are present.
- More abundant than pathogenic viruses.
- More resistant to disinfection treatments and environmental conditions than pathogenic viruses.
- Easily and quickly quantifiable at low cost.
- Identifiable, without ambiguity, in all sample types.
- Distributed randomly in the analyzed sample.
- Do not multiply in the environment.

1.3.1 Contaminants

According to the *Environmental Protection Agency (EPA)*, a contaminant can be defined as a physical, chemical, biological or radiological substance that exists in water. The impact of contaminant in drinking water depends on many factors: the type, the concentration and the health effects. Contaminant leading to a big damage for the population could have some specific characteristics [5]:

- Infectious at low dosage.
- Resistant to Chlorine.
- Stable in water.
- Difficult to detect by consumers by the appearance, smell or taste of water.
- Causing serious illness or death.

The analysis of different type of contaminants, their risks and their Standards is essential. Firstly, it is necessary to differ between limits and references of water quality. Limits are threshold values for hazardous and harmful substances and therefore their application is imperative. For quality limits, the health risk is assessed by "*Directions Départementales des Affaires Sanitaires et Sociales* "(*DDASS*) based on quality requirements of the Public Health Code. Results are then controlled by two types of parameters: i) microbiological parameters (e.g. Escherichia Coli, Enterococci) with a direct and short-term risk and ii) physico-chemical parameters (e.g. nitrates, pesticides) with medium- and long-term effects.

For quality references, the main objective is to show the operation and the efficiency of facilities. Some parameters such as Turbidity, Chlorite, etc. evaluate the operation of treatment stations, others such as Temperature, Sulfates, pH, Sodium, etc. give information on the natural structure of the water and certain parameters define the organoleptic characteristics of water (color, smell, and flavor).

1.3.1.1 Quality limits

Water potability is defined by two types of indicators:

- Bacteriological: microbiological analyses are based on the research of the indicator of bacteria of fecal contamination. These bacteria have been chosen based on different factors: i) their presence in large numbers in the stools of warm-blooded animals which are frequent sources of serious contamination, ii) easily detectable and iii) they do not develop in pure water [6].
- Physical and chemical: these substances can be divided in two parts: i) undesirable whose quantity is allowed up to a certain limit such as Fluorine and ii) those with toxic effects which have very low thresholds such as Lead, Arsenic and Cadmium.

1.3.1.1.1 Standards

At the beginning of the 20th century, the number of water quality parameters was limited to the number of 5. In the 1950s, this number increased to about 20 parameters describing the water quality. In the 1980s, the various factors of progression contributed to increase this number. Today, 54 parameters control the quality of water.

In France, requirements of water quality are determined in the *Public Health Code*, which is based on the *European Directive 98/93/EC* and is completed by the request of the "*Agence Francaise de Sécurité Sanitaire des Aliments" (AFSSA)* and "*Conseil Supérieur d'Hygiène Publique de France" (CSHPF)*. The European Directive threshold values are based on *WHO* guidelines values except those for pesticides, whose values are almost inferior to those proposed by *WHO*, for more safety.

A Standard value will be represented by a lower limit to be respected or a maximum value not to be exceeded. We define Maximum Allowable Dose *"Dose Maximale Admissible"* (DMA) as a maximum amount of a substance that can be consumed daily without being harmful to human health. The threshold values for some parameters are shown in Table 1.1.

Table 1.1. Standards for parameters indicators of contamination.

Indicative Parameters	European Union (EU)	French Standards	WHO
Copper	2 (mg/l)	2 (mg/l)	2 (mg/l)
Chloride	250 (mg/l)	200 (mg/l)	250 (mg/l)
clostridium perfringens (including spores)	0/100ml	spores 0/100ml	
Iron	0,2 (mg/l)	0,2 (mg/l)	<0,3 (mg/l)
Manganese	0,05 (mg/l)	0,05 (mg/l)	0,1 (mg/l)
Odor	acceptable for consumers without abnormal odor	acceptable	
Oxidant power	5mg/l O_2	5mg/l O_2	≥5 (mg/l)
Sulfate	250 (mg/l)	250 (mg/l)	500 (mg/l)
Sodium	200 (mg/l)	200 (mg/l)	200 (mg/l)
Taste	acceptable for consumers without particular taste	acceptable	
Number of colonies at 22 °C	No abnormal concentration (100/ml)	below 100/ml	
Coliform bacteria	0/100ml	0/100ml	
Total Aluminum	0,2 (mg/l)	0,2 (mg/l)	0,2 (mg/l)
Ammonium	0,5 (mg/l)	0,1 (mg/l)	0,5 (mg/l)
Tritium		100Bq/l	10000Bq/l

1.3.1.1.2 Microbiological parameters

A contaminant of microbial origin constitutes a serious threat to public health. The research of microorganisms, potentially dangerous, is unrealistic for technical and economic reasons. The strategy of control is currently based on the search of bacteria known as "*germes témoins de contamination fécale* ", easy to detect, not directly pathogenic, and whose presence suggests the existence of pathogenic germs for humans. Limits of quality are then fixed for indicator microbiological parameters such as Escherichia Coli (E.Coli) and Enterococci [7].

- **Coliforms**

According to the *French Association of Standardization* [8], coliforms are bacillus not sporulated, grams-negative, oxidase-negative, aerobic or anaerobic. They are able to grow in the presence of bile salts or any other surface agent having equivalent properties and capable of fermenting lactose, with production of gas and acid in 48 hours at 37 ± 1 °C. They are rod-shaped bacteria.

Although total coliforms are not harmful to health, their presence indicates the possibility of dangerous contamination in a source of water supply. Their detection is based on laboratory tests since these bacteria have no color, odor or taste. Measurement of total coliforms can also verify the effectiveness of treatment phase. Their presence may explain the bacteria growth due to a low concentration of Free Chlorine or an intrusion in the water quality. It is important to mention that the presence of total coliforms in water doesn't necessarily mean a serious risk to human health. The use of total coliforms may reflect the reliability of the treatment phase in water systems.

In addition of total coliforms, fecal or thermo tolerant coliforms may be used as indicators of human or animal fecal contamination. E.Coli is the most commonly used coliform for water quality assessment. It is a bacteria in the digestive tract of warm-blooded animals and humans or in the intestines of mammals, in particular humans. The detection of this type of coliforms in water can involve the existence of virus or dangerous bacteria. The presence of E.Coli, of fecal origin, in water requires corrective measures to prevent health risks. This type belongs to Enterobacteriaceae family and can be found in naturel water or soils subjected to fecal contamination. The majority of E.Coli is not pathogenic except certain types such as E. Coli: H7 which may be harmful.

These coliforms are generally used in contamination detection because they have an ability to survive equivalent to that of dangerous bacteria. They constitute a reliable indicator of the intensity of fecal contamination.

According to *World Health Organization (WHO)*, guideline values are defined as 0/100 ml for thermo tolerant coliforms and 0/100 in 95 % of treated water samples. The same thresholds are set by *French Standards* and the *European Union* (*EU*) except that for E.Coli. According to *Council Directive 98/93/EC* adopted on 3 November 1998 by the EU, the value is set at 0/250 ml.

The absence of E.Coli in water with the existence of total coliforms can be analyzed according to three possibilities:

- Presence of bacteria (organic films) in the pipes.
- Penetration of surface water.
- Water comes from aquifers that contain bacteria.

- **Enterococci**

These are Gram-positive lactobacilli found in the environment, in the digestive system of humans and animals and in certain foods. They belong to the type Enterococcus of fecal streptococci family. Enterococci are characterized by their growth in unfavorable conditions, especially species of *E. avium, E. casseliflavus, E. cecorum, E. aurons, E. faecalis, etc.* Most of these species are of fecal origin (human or animal). They are generally considered specific indicators of fecal pollution.

The Enterobacteriaceae are of high importance since many of these bacteria are pathogenic (Salmonella, E.coli, Yersinia pestis, etc.). Some of them are used as indicators of fecal contamination of drinking water [9].

- **Aerobic and Anaerobic Heterotrophic Bacteria (BHAA)**

According to drinking water and public health summaries, the presence of BHAA gives an idea of the quality of water. It can indicate a problem in the treatment. Most types of BHAA do not constitute a hazard to human health. However, some types can cause diseases especially for people with low immunity, or those aged or for very young children.

- **Coliphages**

Bacteriophages or coliphages can be considered as indicators of the treatment of drinking water. They are viruses that infect bacteria, such as coliform bacteria. They are divided in two types: somatic coliphages and male specific coliphages. In order to have efficient disinfection, water must not contain any somatic or male specific coliphage. If the water is not disinfected and there are no other water quality indicators (such as E.Coli), coliphages may indicate the presence of human enteric viruses. If water is disinfected and there are no other quality indicators (such as E.Coli), coliphages indicate deficiency in treatment. For distributed water samples, the presence of coliphages does not provide more information than that given by BHAA.

1.3.1.1.3 Chemical parameters

Several chemical parameters can be used as indicators of water quality anomaly, especially Arsenic, Cadmium, Cyanide, Mercury, Lead, Chromium, Nickel, Antimony and Selenium, some hydrocarbons, and also pesticides and Nitrates.

- **Nitrates**

For a high precaution, the Standard for Nitrates is set at 50mg/l according to the *European Directive and the French Standards*. This threshold is based on the risks that can affect pregnant women and infants which are most sensitive. Nitrates can come from human activities (agricultural for example), or from natural origin (Nitrates resulting from the transformation of Nitrogen in the water and soils: cycle of the Nitrogen).

The *WHO* set also the limit for Nitrate as 50 mg/l. An exceeding of this limit is noticed in small drinking water production facilities. According to *"Conseil Supérieur d'Hygiène Publique de France" (CSHPF)*: if the concentration is between 50 & 100 mg/l: cessation of consumption for pregnant and newborn; if the concentration exceeds 100 mg/l: cessation of consumption for the whole population.

- **Sulfates**

Water containing a significant amount of Sulfate can induce problems of dehydration and diarrhea. These problems affect especially children unaccustomed to high concentrations. A concentration, that exceeds the limit, induces unpleasant taste of water. A very high concentration can produce corrosion of pipes.

- **Pesticides**

This category contains several hundred substances, including: herbicides, insecticides, acaricides, nematicides (against worms), fungicides, rodenticides, etc. The *French Standard* for pesticides complies with the *European directive* and defines the limit as [10]:

- 0.1 μg/l maximum concentration for each substance; Except for Aldrin, Dieldrin, Heptachlor and Heptachlorepoxide whose maximum acceptable dose is 0.03 μg / l.
- 0.5 μg/l total concentration of pesticides.

For certain pesticides, *WHO* defines a specific guideline value, for example Atrazine: 2μg / L, Terbuthylazine: 7μg / L and Isoproturon: 9μg / L. In 2003, Atrazine and Atrazine-Desethyl were the most detected pesticides. Based on health risk control, the bacteria are divided into three main groups: Triazines, substituted ureas and organochlorines.

The *CSHPF* classifies distribution units and populations into three categories: A (permanent Conformity), B1 (presence of pesticides without restriction of water use), B2 (frequent or important presence of pesticides with restriction of water use). In 2003, 91% are in situation A, 9% are in situation B1 and B2.

- **Arsenic**

It is a very toxic substance, very dangerous to human health and can lead to death. Based on its harmful effects, the *WHO* acceptable value was reduced from 50 to 10 µg/l. This guideline value of *WHO* has been incorporated into European law (Council Directive 98/83/EC of November 3, 1998) and French law (Decree 2001-1220 of December 20, 2001) in the form of a "maximum acceptable concentration" and a "quality limit'", set to 10 µg/l instead of the 50 ug/l fixed in 1989 [11].

- **Iron**

Iron in water may be derived from: i) iron salts which are used in some cases instead of Aluminum salts in water flocculation, ii) industrial waste and iii) corrosion of metal pipes. An exceeding of Iron Standards increases the risk of cardiovascular disease and cancer. Very high amounts can contribute to several neurodegenerative diseases such as Alzheimer and Huntington's chorea.

- **Manganese**

According to the *Health Agency and socials services of "chaudière-Appalaches" Québec*, a high concentration of Manganese gives a bad taste. Its presence in drinking water could have an impact on the neurological development of the child, such as a decrease in the intellectual quotient.

- **Ammonium**

It results from a reaction between the minerals comprising Iron and Nitrates. Its presence in water indicates incomplete degradation of organic matter. It is a good indicator of the pollution of water by organic discharges of agricultural, domestic or industrial origin. It is not very harmful but can have some health effects. According to *LENNTECH* [12], Ammonium is not very toxic, but it can cause several problems such as: corrosion of the pipes, bacterial revivification within them, decreased effectiveness of Chlorine treatment and development of microorganisms responsible of unpleasant flavors and odors.

- **Aluminum**

According to the *Water Information Center* [13], some studies have shown the possibility of increasing the risk of Alzheimer's when drinking water is too rich in Aluminum. Aluminum is toxic to nerve cells and researchers are trying to find out if it can promote the development of several pathologies, such as multiple sclerosis and especially Alzheimer's disease. This disease affects in France more than 400 000 people, with about 100,000 new cases each year. Several studies conclude that the risk of developing the disease increases when the concentration exceeds 0.1 mg/l of water [14].

- **Sodium**

Saline intrusion, mineral deposits, seawater spray, sewage effluents, and salt used in road de-icing can all contribute significant quantities of Sodium to water [15]. Sodium is not acutely toxic, but it can have some human effects. This effect differs between infants and adults. Taste of water could be affected by an amount of Sodium that exceeds the limit of 200 mg/l.

- **Copper**

Copper concentration in drinking-water varies widely as a result of variations in water characteristics, such as pH, hardness and Copper availability in the distribution system [16].

The gastrointestinal effects of Copper depend on the temporal aspects of exposure (Acute or longer-term) and on the consumed concentration.

- **Tritium**

According to "*Institut National de santé publique de Québec" (INSPQ)*, Tritium, as well as all radioactive elements, is considered by International health organizations as a carcinogenic agent.

1.3.1.2 Quality References

In addition to the quality limits, several parameters, defined as quality references, are used to detect a disturbance in water. The principal indicators of water quality are: Turbidity, Total Organic Carbon (TOC), Chlorine, Conductivity, pH and Temperature. Table 1.2 illustrates different threshold Standards for quality references.

1.3.1.2.1 Turbidity

This parameter indicates the existence of perturbation in the system. It can be a result of pipes break or a problem in the amount of residual Chlorine. A removal of deposits from wall surface of the pipe, corrosion, or abnormalities in the treatment, are all considered as source of troubled water. The Turbidity is an indicator of the presence of suspended matters, which could be due to microorganisms (virus, bacteria, and protozoa) in drinking water. The most widely used technique for measuring Turbidity is the Nephelometry. It is based on the measurement of scattered light at an angle of 90° relative to the incident. This factor is correlated with the amount of suspended matters that disturb the water.

1.3.1.2.2 Organic materials

The measurement of organic matter indicates the presence of heterotrophic bacteria in the reservoirs and in the distribution system. A high level of TOC can be explained by the release of biofilms in the distribution system or by a "breakthrough" in the treatment station. On the other hand, Dissolved Organic Carbon (DOC) describes the remaining Organic Carbon in a sample after filtering, generally using 0.45 µm filter. High DOC in chlorinated water produces the Trihalomethane (THM) which is harmful for human health. Many organics materials absorb in the ultraviolet. The measurement of the absorbance UV is an efficient indicator of water quality. It is correlated with TOC, color, THM and its measurement is more simple and economical than TOC [17].

1.3.1.2.3 Residual Chlorine

Chlorine is a principal disinfectant agent, used to minimize bacterial growth in the distribution networks. A main benefit of Chlorine over other disinfectants, is that it leaves a residual disinfectant. This residual assists in preventing recontamination during distribution. Its absence may indicate the possibility of post-treatment contamination. There are three types of residual Chlorine [18]:

- Free Chlorine: most reactive species, such as hypochlorite ion.
- Combined Chlorine: less reactive but more persistent species, resultant from the reaction of Free Chlorine species with organic materials and Ammonia.
- Total Chlorine: sum of Free Chlorine with Combined Chlorine.

According to the *WHO*, a concentration of Free Chlorine in treated water between 0.2 and 0.5 mg/l must be maintained. A rapid decrease in Chlorine levels indicates a water quality variation induced generally by biofilm growth. The amount of residual Chlorine in distribution network in France is between 0.05-0.1 mg/l, it can be increased to 0.2-0.3 mg/l during a period of risk [5].

1.3.1.2.4 Conductivity

The measurement of the Conductivity is defined as the capacity of the water to conduct the "current" between two electrodes. It indicates the quantity of dissolved salts in the water. Conductivity values can indicate pollution or infiltration and can identify a mixture of different water sources.

1.3.1.2.5 pH (Hydrogen potential)

It is important to measure pH at the same time as residual Chlorine since the efficiency of disinfection with Chlorine is highly pH-dependent: where the pH exceeds 8.0, disinfection is less effective [18]. pH value (concentration in Hydrogen ion) must be measured in the field using either comparators or a pH meter or by colorimetric technique [5]. The pH measurement is important for two main reasons: i) some disinfection products depend on the pH value, ii) for corrosion control.

1.3.1.2.6 Temperature

The measurement of temperature is of high importance in water quality monitoring. A temperature contrasts indicates the presence of bacteria which can cause a bad flavor or odor and even corrosion problems.

Table 1.2. Quality references according to different Standards.

Parameter	EU	French Standards	WHO
Turbidity	acceptable for consumers without abnormal variation	0,5 NFU, 2NFU (2003) at tap,1NFU (2008)	< 5NTU
TOC	No abnormal variation	2 (mg/l)	
Residual Chlorine		0.2 (mg/l)	
Conductivity	2500 µS/cm at 20°C	180, 1000 µS/cm at 20°C & 200, 1100 µS/cm at 25 °C	250 µS/cm
pH	≥6,5 and ≤9,5	≥6,5 and ≤9	6,5-8,5
Temperature		25 ° C	

1.3.2 Factors leading to water quality degradation

The quality of tap water can differ from that produced by treatment plants. To monitor the quality of drinking water, it is important to identify the factors responsible of water quality degradation. These factors constitute the origin of drinking water contamination.

First of all, the "hydraulic residence time" of the water in the pipe can directly affect the quality of water and cause its degradation. This is derived from water stagnation characterized by bacterial growth, and some corrosion forms. A low flow velocity can induce such phenomenon of stagnation. A renewal of water in the distribution networks is required. The maximum "hydraulic residence time", before the need of pipe's purge, is one month according to the *Standard EN 1717*. Although the increase in pipe diameters will be effective for fire protection reasons, it may be a source of stagnation, and therefore a pollution source in case of low consumption. A long "hydraulic residence time" of the water in the distribution networks leads to a high microbial density. It induces a high risk of bacterial contamination. The influence of the "hydraulic residence time" on the microbiological quality of water is illustrated in Figure 1.2. Other factors such as temperature may be at the origin of the re-growth of certain bacteria.

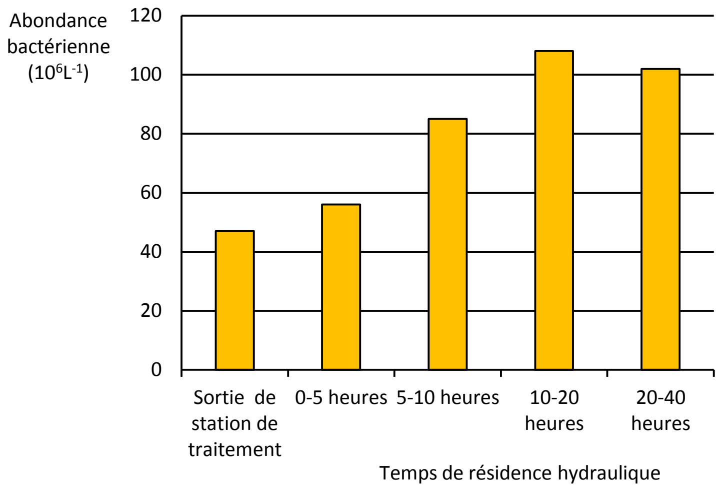


Figure 1.2. Example of the alteration of microbiological quality along a drinking WDS [19].

On the other hand, some points, in a water network, provide pathways to a contamination by microorganisms. This is the case of tanks where water is in contact with air, and where the unprotected orifices can allow the passage of dust or insects [20].

Microorganisms, in distribution networks, can be protected from disinfectants by the formation of biofilms on the pipes walls. Biofilms are defined by a set of micro colonies joining together several species such as bacteria and protozoa. The presence of these microorganisms with some amount of biodegradable organic materials leads to microbial contamination. These microorganisms can multiply within the network, despite the low concentration of nutrients and the presence of disinfectants. The presence of microorganisms, particularly bacteria, in drinking WDSs comes from several sources: repair actions, connections, network interventions, leakage or breakage problems, poorly protected areas, water backflow, in addition of all the bacteria that resist to treatment steps in the stations. The growth and multiplication of these bacteria, developed in the presence of biodegradable organic carbon, lead to the formation of biofilms in the pipes.

The presence of microorganisms in water networks induces several problems. The results of bacterial growth do not concern only the distribution systems but the consequences can reach the consumers. For example, the risk of gastroenteritis can be a direct consequence of biological instabilities in distribution networks. This proves the importance of limiting bacterial reviviscence within drinking water pipelines.

In addition, one of the main problems of drinking water systems is the phenomenon of corrosion. The corrosion involves the presence of several materials such as Iron, Zinc, etc. It induces a decrease in the amount of residual Chlorine. However, the quantity of residual Chlorine should be maintained to limit the presence of pathogenic microorganisms. Some bacteria can accelerate also the corrosion phenomenon. The distribution capacities of a network can be then decreased by the increase in resistance forces, induced by the biofilms [21]. An efficient maintenance of the networks is required to avoid corrosion phenomena. Corrosion is divided in two main types: i) external corrosion, often galvanic, produced in the presence of dissimilar or electrolytic metals and ii) internal corrosion responsible of water quality degradation in the distribution pipes. Internal corrosion occurs in the case of uncoated metallic pipes. The negative effect of this type of corrosion on water quality can be manifested in different forms: i) corrosion derivatives such as rusty or red water and ii) tubers arising from the metal oxides produced from corrosion.

The constituent elements of the pipes can affect also the quality of water in distribution networks. They can be of two types: i) materials of organic types such as Polyvinyl Chloride (PVC) and ii) metallic materials such as Cast Iron and Steel. Among these materials, those which lead to the presence of microorganisms depending on their roughness, or induced the formation of biofilms. Others involve corrosion tubers.

Materials used in production or distribution systems and in contact with water, intended for human consumption, shall not be capable of altering the quality of water. They must respond to the conditions laid down in a decree, on the advice of "*Agence française de sécurité sanitaire des aliments*", by the ministers responsible for Health, Industry, Consumer Affairs and Construction according to *Article 7 of the Decree of January 3, 1989.*

Article 3 of the departmental-sanitary health regulation, diffused by the circular of August 9, 1978 of the Minister for Health, recalls that:

- "Bituminous coatings, Oil-based coatings and all similar products as well as plastic coatings, may be used in the only conditions that, where they are in contact with water, they are in danger of disintegrating or communicating to the latter unpleasant favors or odors".
- "The substances used in the composition of piping materials, devices and part of them, and accessories made of plastic materials must comply with the rules, concerning materials and articles placed in contact with foodstuffs".

Although water produced by treatment plants is not sterile water, the main objective is to deliver water tap without health risk. It is essential to avoid bacterial reviviscence within the distribution networks. Distribution networks are affected by biological and physicochemical interactions. Water utilities should ensure that tap water is suitable for human consumption. Table 1.3 summarizes the main origin of water quality degradation in distribution network, with their consequences.

Table 1.3. Main origin of water quality degradation [20].

Origin	Reason	Consequences	Hazards
Tanks	ventilation holes or poorly protected access routes	Penetration of insects or other animalcules	Alteration of water organoleptically or microbiologically
	oversizing	water stagnation	Alteration of initial qualities for various parameters
water backflow	Depression (intensive pumping on the network, pipes break …) or backpressure (pressurizing in private installation)	siphoning or suppression of undesirables or polluting substances	toxic, microbiological or organoleptic pollution
Exterior environment of the pipe	leaks, permeation	Introduction of pollutants	Microbiological pollution, especially organoleptic or toxic pollution
Network faults	Inadequate material for drinking water supply	Excessive adhesion of germs, corrosion	Microbiological contamination, Alteration of metal parameters (Fe, Zn, Pb, Cu, Cd, ...)
Interventions on the network	Inadequate disinfection following a repair or renewal,	Germs development	Microbiological contamination
Intrusion of pollutants at the level of suction cup (rare)	problem in suction cup location	contaminated drinking water	Microbiological contamination
Interior installation	Inadequate material, over-dimensioning (low consumption…)	corrosion, water stagnation	Organoleptic, chemical (NH_3), microbiological contamination (Legionella)

1.3.3 Waterborne diseases (Health risks)

The consumption of contaminated water threats the human health. It may induce many risks, which can be very dangerous in some cases. Their harmful effects on health should be evaluated.

As already mentioned, the presence of E.Coli in drinking water indicates a contamination that may contain bacteria, viruses or parasites. All of them can induce severe illness. The health risk varies according to the age and to the body immunity of the person. The most common symptoms are: nausea, diarrhea, vomiting and in the most intense cases, infections of the eyes, lungs, skin, nervous system, kidneys, liver and in very serious chronic or even fatal cases.

Similarly, Enterococci can have dangerous impacts. They can induce intra-abdominal infections, blood infections (septicemia) of the cardiac wall (endocarditis), or urinary tract. Enterobacteriaceae have been found for about 20 years in half of nosocomial infections. Responsible species are mainly E.Coli but also *Klebsiella*, *Enterobacter*, *Serratia*, *Proteus*, *Providencia* [9]. According to bacteria’s type, pathogenic effect of Enterobacteriaceae is summarized in Table 1.4. Heterotrophic bacteria also have a detrimental consequence on human health, which is indicated in Table 1.5.

Table 1.4. Pathogenicity of several types of bacteria.

	Type of bacteria	**Health risk**
Enterobacteriaceae	Yersinia pestis	plague
	Salmonella serovarTyphi	typhoid fever
	E.coli Salmonella enterica	Diarrheal syndromes and other intestinal infections
	Shigellasp	Diarrheal syndromes and other intestinal infections
	Yersinia enterolitica	Diarrheal syndromes and other intestinal infections
	Enterococci	Intra-abdominal infections, blood infections (septicemia), cardiac wall (endocarditis) or urinary tract.

Table 1.5. Diseases induced by heterotrophic bacteria.

BHHA (Aerobic and Anaerobic Heterotrophic Bacteria)	Pseudomonas aeruginosa	Nosocomial infections
	Legionellapneumophila	legionellosis, Pontiac fever
	Aeromonassp	diarrhea

Chemical material can also induce several health risk, if consumed at high quantity. A high amount of Nitrates cause risks of methemoglobinemia for the infants: Nitrates transformed into Nitrites cause variations in the properties of hemoglobin in the blood; they hinders the correct oxygen transportation by red blood cells. Also, Nitrite that results from Nitrate transformation can induce some types of cancer for adults.

For pesticides, the consumption of very low but repetitive doses can lead to cancer diseases (especially leukemia), disorders of nervous system and reproductive disorders due to chronic exposure. Arsenic may also have harmful effects such as severe digestive disorders and vascular risks.

Other substances may have hazardous health effects and therefore their thresholds are set according to the Standards. The health risks of these parameters are summarized in Table 1.6.

Table 1.6. Health risks induced by substances that may exist in water.

Parameter	Health risk
Arsenic	Skin and internal cancers, severe digestive disorders, vascular risk and risk on carotid atherosclerosis
Fluorides	Dental or bone fluorosis
Lead	saturnism
Copper	Fever, irritation of nose, mouth and eye, headache and stomach pain, dizziness, vomiting, diarrhea. In case of high dose, kidney and liver damage and even death
Chloride	A relationship is demonstrated between the high concentrations of chlorine by-products and the increased incidence of humans cancer
Manganese	Neurotoxic, disorders suggestive of Parkinson's disease: memory deficit, signs of depression, hallucinations, lack of memory, and problems with nerves, pulmonary embolism and bronchitis, schizophrenia, boredom, muscle weakness, Headaches and insomnia.
Sulfate	Dehydration and diarrhea (especially in children), unpleasant taste
Sodium	Damage the kidneys and increase the risk of high blood pressure. Contact with sodium, including sweat, causes the formation of sodium hydroxide vapors, which are highly irritating to the skin, eyes, nose and throat. This can cause sneezing and coughing. Very severe exposures can cause difficult breathing, coughing or bronchitis
Total Aluminum	Increased risk of Alzheimer's
Ammonium	Redness, sore throat, vomiting, nausea, inhalation, cough
Tritium	Carcinogen, nausea, hair loss

1.4 Water quality monitoring using conventional methods

In general, water quality monitoring is done using traditional methods. These methods are generally based on taking manual samples from different locations in the distribution network. Laboratory analyses are done to evaluate the quality of water. Laboratory analyses can be divided in two main categories: i) physico-chemical, which take more than two hours, to determine organoleptic characteristics of tested water and ii) microbiological to identify the presence of pathogenic microorganisms and this type requires more than one day to obtain the result. In addition of the problem of high time consuming, these conventional methods have to be applied with specified conditions. They have to be repeated continuously (7 days between samples) to verify the accuracy of results. Some parameters are directly analyzed on site with certain instruments. The work of labors, that take samples, should be very precise and accurate.

Several conventional methods are already used for water quality control. Among these methods those based on culture, DNA and immunology. Although these techniques require a long time to identify the presence of pathogens, they are not very expensive in general.

1.4.1 Traditional techniques (culture method)

Conventional methods, used for the detection and identification of pathogens, are based mainly on microbiological and biochemical identification. These methods require pre-enrichment of the samples before proceeding to the analysis stage [22]. One of these conventional methods is that based on the cultivation and counting of colonies.

The two traditional methods often used are Most Probable Number (MPN) and Membrane Filtration (MF):

- MPN: a method of replicating the growth of the liquid, that allows the enumeration of bacteria present in a given sample. The most probable number is calculated by a statistical table giving the number of positive tubes (producing a gas that indicates the possible presence of coliforms) among the inoculated tubes.
- MF: this method is based on the enumeration of bacteria in colony units in 100 ml after filtration of water samples on membranes with small pore size. Different media and incubation conditions (time and temperature) may be used depending on the type of fecal bacteria researched [23].

Several studies have also highlighted the importance of bio indicators such as fecal coliforms in the detection of water quality anomalies. Several criteria define a good indicator [24]:

- Present in fecal samples at high concentrations.
- Absent in uncontaminated water samples.
- Measurable.
- Having the same resistance to environmental conditions and to disinfectants as the pathogens they represent.

According to *WHO*, indicators are divided in three groups:

- General Microbial Indicators (Heterotopic bacteria or total coliforms remaining after disinfection).
- Fecal specific indicators representing a group of organisms directly indicating fecal contamination (Thermo tolerant coliforms such as E.Coli).
- Index or organism's model of pathogen (E.Coli index of Salmonella & F-RNA coliphages as model of human enteric viruses) [25], [26].

The accuracy of these methods, based on enumeration and laboratory culture, can be improved by combining the use of viral indicators with bacterial indicators [27].

These conventional methods have many limitations, where the major one is time-consuming. Also, several infectious and dangerous agents cannot be cultivated, in addition to the absence of specific indicators for viruses.

1.4.2 Molecular detection method (Nucleic acid method)

To improve the accuracy of detection, methods based on the use of nucleic acid will be more efficient and rapid. Among these methods: *PCR (polymerasechainreaction), quantitative PCR (qPCR), quantitative reverse-transcriptase PCR (qRT-PCR), sequencing of PCR amplicons, & next-generationsequencing.* Molecular biology analyses are based on the recognition of specific DNA sequences of microorganisms [27], [5]. These methods can detect waterborne pathogens such as Camylobacter, E.Coli, and Salmonella. For example, the PCR method allows bacterium identification by the demonstration of one or more fragments of its genome. In the case of pathogenic bacteria for which the conventional detection techniques, where they exist, are complex, long and unreliable, PCR constitutes a very good alternative [22]. Although the

PCR technique has the advantage over traditional techniques of being able to detect viable as well as dead bacteria, the problem remains in the discrimination between these bacteria. Another limitation of PCR is the high detection threshold in some cases. To reduce this problem, immunomagnetic separation based on DNA concentration of specific sequences may be an effective solution. Other technique called Multiplex PCR allows detection of several organisms by a single reaction and then a reduction of the time detection. Real time PCR and Real time qPCR quantify, detect and differ between different contaminants (such as E.Coli O157: H7) in real time. A technique called 16 SrDNA allows the identification of microbes present in the sample. But, it cannot be used for virus detection. FISH method (Fluorescent In Situ Hybridization) and confocalmicroscopy are methods used mainly for the detection of microbial pathogens in sewage treatment plants. Microarray method based on the hybridization of a sequence of target nucleic acid (DNA or RNA) to a complementary sequence is intended to identify microbes in the water samples. The Next-Generation Sequencing (NGS) technique is a method for the direct detection of a microbial sequence without prior knowledge of the type of contaminant [27].

Although the latter techniques are rapid, effective and accurate, they do not reflect the real health risk. Other limitations are the presence of inhibitory substances which may influence on the specificity of these methods.

1.4.3 Immunological detection

One of the conventional methods is the technique of immunological detection. The principle of detection is the interaction antigen-antibody. It is based on the ability of antibodies to recognize specific three-dimensional structures (e.g. parts of proteins or polysaccharides) of biological macromolecules [28]. Immunological detection includes, for example, Serum Neutralization Test (SNT), immunofluorescence, and Enzyme-Linked Immunosorbent Assays (ELISA) [29]. Although they are faster than methods based on nucleic acid, they are less specific and cannot provide real-time detection.

These immunological methods are now coupled with other methods for more efficient detection of pathogens. For example, the combination of immunomagnetic separation with flow cytometry for the detection of *L. monocytogenes* [30]. Also, an estimation of E.Coli O157 was done, using a Minimal Lactose Enrichment Broth (MLB) with Immunomagnetic Electrochemiluminiscence detection (IM-ECL) [31].

1.4.4 Immunomagnetic Separation with flow cytometry

Immunomagnetic Separation (IMS) was used to detect enteric viruses. Flow cytometry method is rapid and quantitative method [5]. Although this method is fast, it does not differ between viable and non-viable cells. IMS and flow cytometry can be combined to improve the detection techniques. For example, a waterborne pathogen, such as Legionella pneumophila (Lp), is detected by filtration, Immunomagnetic Separation, double fluorescent staining, and Flow Cytometry Method (IMS-FCM) [32].

1.4.5 Biosensor

The biosensor is formed from two main parts: i) the molecular recognition element and ii) the transducer which detects physical changes and converts them into a signal. The biosensor is not effective for the detection of human pathogens. These pathogens often correspond to species or specific tissue requiring incubation time (a few days to weeks) before the symptoms of the disease are perceptible [33]. Figure 1.3 illustrates the configuration of a biosensor.

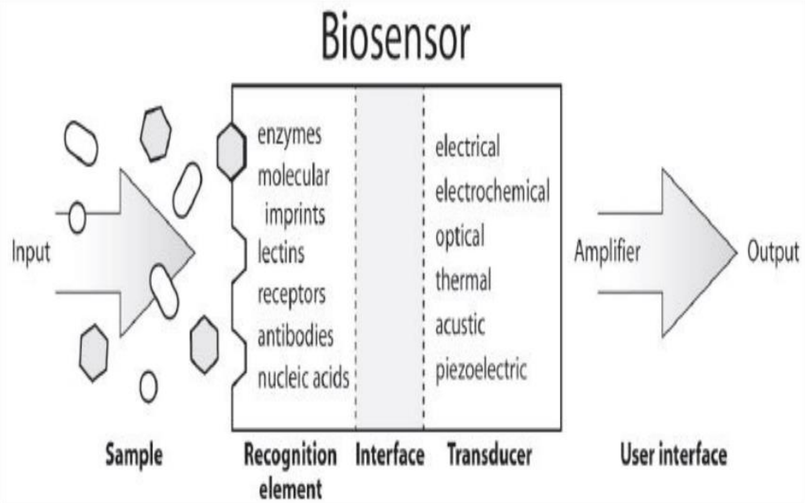


Figure 1.3. Configuration of a biosensor showing bio recognition, interface, and transduction elements [34].

1.4.6 Adenosine Triphosphate (ATP) Test

Adenosine Triphosphate (ATP) is a molecule present in all bacteria. There is a good correlation between cellular ATP measurement and the number of viable bacteria present [35]. The identification of ATP, based on bioluminescence, requires the presence of Luciferase, Luciferin, Magnesium and Oxygen [36]. A main problem that reduces the use of ATP technique is the high cost.

1.4.7 Vibrational spectroscopy

The spectroscopy is based on the interaction between radiation and substance. It is used for the identification of substances through the spectrum emitted from or absorbed by them [37]. The vibrational spectroscopy is noninvasive method, used in the detection of bacterial cells. It can be divided in two main techniques:

- Fourier Transform Infrared Spectroscopy (FT-IR): based on interferometer principle. It is used for bacterial pathogen identification and differentiation [38]. (FT-IR) combined with multivariate analysis allows the identification of the bacterial contaminants Pseudomonas aeruginossa and E.Coli in drinking water [39].
- Raman spectroscopy: used for a rapid identification of molecules. The Raman effect can be defined as inelastic scatter of light. A micro-Raman-spectroscopy with excitation in the near infrared or visible range has the potential for the analysis of single bacterial and yeast cells [40].

1.4.8 Turbidimetry

Turbidity indicates the relative clarity of water. The analysis of Turbidity is based on the optical properties that cause the scattering of light through water [41]. In general, Turbidity is measured in Nephelometric Turbidity Units (NTU). The measured value illustrates the presence of suspended materials in water. It indicates the loss in the intensity of transmitted light.

When disinfection is applied, value should be ideally less than 1 NTU. To prevent Turbidity variation during sample transportation, it is measured usually on site using electronic meter

(Turbidimeter). Although electronic meter is very useful for low Turbidity (less than 5 NTU), it can be easily damaged and requires a power supply (mains or battery). When sensitivity is not essential in small community, extinction methods (using turbidity tube) can be efficient for value higher than 5 NTU. Although turbidity tube is cheaper than electronic meter, the latter is more precise.

1.4.9 Chemical analyses

Although the majority of water quality problems are related to fecal contamination, a significant number of serious problems may occur due to chemical contamination [18]. It would be costly to measure a wide range of chemical parameters on a periodic basis. It is essential to determine the most health-related parameters (Fluoride, Pesticides, Arsenic, Nitrate, etc.). The determination of chemical contaminant can be done on two ways: i) field test methods and ii) laboratory test methods.

1.4.9.1 Field test methods

These methods could be conducted using a portable test kit, which provides rapid results on site. The portable test kit includes comparators, photometers, etc.[42]:

- Test (Reagent) Strip: Test strips typically have a plastic handle with a reagent area at one end. The reagent area is immersed in the water sample. The color of the reagent is then compared with color chart. The main limitation of this method is the low accuracy since it requires a visual interpretation.
- Color Disc Comparator: based on the comparison between the standard test colors provided on the disc and the color produced by each chemical test. Color comparators can be more accurate than test strips, but they are more expensive.
- Colorimeter and Photometer: Colorimeters and photometers use a light source to measure the chemical concentration in a water sample. Although they are more accurate than test strips, they require training to ensure they are used properly.
- Digital meters: Some portable field kids include various digital meters to measure parameters like pH. Their use is simple, but they require calibration.
- Arsenic test kits: New kits have been developed making field testing easier and more accurate. The Arsenic concentration is determined by comparing the intensity of the stain with a color chart.

1.4.9.2 Laboratory test methods

Some chemical parameters require laboratory analyses such as Atomic Absorption Spectroscopy (AAS), Chromatography, Colorimetric method, etc.:

- AAS: analyzes the presence of metals. It is based on the principle that free atoms (gas) can absorb radiation at specific frequency. The analyte concentration is quantified from the amount of absorption.
- Chromatography: used to determine metallic, organic or inorganic substance. There are many type of chromatography: liquid chromatography, gas chromatography, ion-exchange chromatography. All of them have the same principle of the mixture separation. It is based on the affinity difference between two phases: stationary and mobile.

- Colorimetric method: measures the color intensity of a reaction product or a colored target chemical. The concentration is determined by means of a calibration curve obtained using known concentrations of the determinant [42].

1.4.10 Aesthetic analyses

Aesthetic parameters, such as color, odor and taste, indicate qualitatively the status of water. They are simple to detect by sensation and inexpensive to be controlled on site:

- Taste and odor: many factors could induce odors in water: organics substances, industrial contamination, biological activities, etc. The problems of taste are identified by consumer's complaints. It can detect the presence of inorganic compounds of metals (Calcium, Sodium, etc.). In case of taste and odor problems in water, sampling analyses are required.
- Color: it is a useful indicator of the need of laboratory analyses. The variation of water color could indicate the presence of metals (Iron, Manganese), or organic substance.

1.5 Cases of water contamination

1.5.1 Global contamination frequency

Several studies have highlighted many cases of water contamination, inducing disease situations as well as a number of deaths. The following is a summary of the various percentages of contamination affecting public health worldwide.

In 2008, about 884 million people lacked access to an improved drinking water source [43]. In 2015, 2.1 billion people (or some 3 in 10 people worldwide) lacked safely managed drinking water services; among them 844 million do not have even a basic drinking water service [44]. In particular, about 25% of the global population (1.8 billion people in 2012) consumes water contaminated with fecal materials [45]. More than 50 countries continue to report active cholera cases to *WHO*; millions of people are exposed to hazardous levels of Arsenic and Fluoride and about 260 million suffer from waterborne diseases [46].

Many water-borne viruses can produce health problems, including: Norovirus, Adenovirus, Hepatitis A, Hepatitis E, Bacteriophages, etc. Children under the age of 5 are the most sensitive; about a hundred thousand deaths a year are induced by viral diseases.

- Hepatitis A: The virus can reach the liver where it can cause severe damage. The fatal situation is less than 1% and is greater for those over the age of 50 [47].
- Hepatitis E: The virus causes inflammation of the liver which is not fatal but induces pain, vomiting and fever [48]. Hepatitis E can lead to a mortality rate that could reach 25% among pregnant women [49].
- Rotavirus: Human Rotaviruses (HRVs) are the most important cause of children mortality worldwide. Typically, 50-60% of cases of acute gastroenteritis for hospitalized children around the world are caused by HRVs.
- Norovirus: Symptoms, starting after one to two days, can vary between nausea, vomiting, diarrhea, etc. In the United Kingdom, each year, Norovirus affects between 600,000 and 1,000,000 people [50].

According to *United Nations International Children's Emergency Fund* (*UNICEF)*, contaminated water, inadequate sanitation, and poor hygiene are responsible for about 90% of deaths. Diarrhea is the second leading cause of death among children under five worldwide (1.2 million deaths in 2012) [51].

1.5.2 Water contamination in France

In France, water quality monitoring includes two principal phases:

- Sanitary control of water supply. This type of regulatory control is executed by the *DDASS*, which ensures, under the authority of the prefects, the implementation of health policies, the protection against epidemics and the control of hygiene rules.
- Quality monitoring by either on-site & laboratory analyzes on a regular basis or by the use of sensors allowing continuous monitoring. This is done by the responsible of water supply.

The water control is of two types: i) routine control which evaluates the reliability of treatment and gives information on organoleptic and microbiological quality and ii) complete control which assesses the compliance with other quality requirements according to the *Public Health Code*.

Microorganisms existing in water can cause gastroenteritis and can lead to an epidemic situation. The epidemic cases recorded in France are of one annual situation, among them 800 cases of gastroenteritis in Ain in 2003 and 400 cases of gastroenteritis in Saone-et-Loire in 2001.

A study shows that during 10 years (between 1998 and 2008), ten epidemics due to drinking water networks are recorded in France. A large population have been exposed (estimated number between 1000 and 60 000). On average, more than 1 000 people infected in each epidemic situation. In total, the number is 9000 affected, of whom 1000 are hospitalized. The most implicated contaminants: Norovirus, Cryptosporidium, Campylobacter and Rotavirus indicating fecal contamination. It should be noted that 3 to 5 contamination resources are due to a dysfunction of the chlorination phase [52].

In 2002, the percentage of people receiving permanently water that doesn't comply with microbiological quality limits is estimated at 5.8%. This is often the case of small distribution units in rural areas. A total of 380,000 people, living in 2,087 communes, are potentially exposed to an exceeding of bacterial standards [53].

Many cases of water contamination are presented to evaluate the frequency of pollution in WDN in France. Critical situations were occurred in different regions.

1.5.2.1 Strasbourg, May 2000.

A bacteriological contamination was taking place on May 26, 2000 in a part of a drinking water system in Strasbourg. Nearly 60 000 inhabitants are deprived of drinking water for about 15 days. The consequences of this pollution were manifested by cases of gastroenteritis (vomiting and/or diarrhea).

To find a relationship between reported cases of gastroenteritis and pollution of the water network, an epidemiological investigation was carried out. It showed a dependence between the possibility of exposure to gastroenteritis and the age of the consumer. Three age intervals have been studied: 1-14 years, 15-60 years, over 60 years. This research estimated that there were almost 53 cases of vomiting, fever and diarrhea in the contaminated network area.

To analyze also the water quality, samples are taken daily from the polluted part in the first stage and then on a wider region of the network. The type of analysis carried out are bacteriological, making it possible the identification of coliforms, fecal streptococci, etc. These analyses give the following results: maximum levels of microbiological contamination were found on May 26th (16 to 36 Units Forming Colony (UFC)/20 ml sulfito-reducing spores, 4 to 22 UFC /100 ml fecal coliforms, 2 to 25 UFC/100 ml total coliforms). Then, positive results were found until June 5th at different points (1 to 4 UFC /20 ml sulfur-reducing spores) [54].

This contamination situation in Strasbourg highlights the high probability of pollution that could affect drinking water systems. The contamination has led to several risks (cases of gastroenteritis) on public health. The verification of this disruption required several surveys. Either the surveys that were executed to correlate situations of gastroenteritis with water pollution, or the bacteriological analysis. The latter requires daily sampling, especially at the contaminated network. It determines the presence of the microorganisms that have caused the pollution.

1.5.2.2 Rennes, January 2011.

A situation of accidental contamination, in drinking water network, was reported on January 11, 2011, in Rennes. The contaminant was Ethylene Glycol, generally used as an antifreeze. Intoxication caused by this contamination results in several symptoms, such as digestive disorders.

Several recommendations were taken into account on January 13, 2011, concerning people who could use this contaminated water. The number is estimated to be about 400 people. Among these recommendations, the instructions of non-consumption of water in the affected network in addition to the execution of purges.

The results of the contamination recognition were: 96.7% aware of the contamination: 10.8% prevented on Wednesday and 86.4% on Thursday, of which 25.8% before 10:30. While for clinical signs: 27.9% reported having at least one symptom, 14.2% at least water between headaches, abdominal pain, nausea/vomiting and drinking [55].

These various symptoms of poisoning justify the importance of real-time monitoring, since traditional techniques may take a long time to justify water pollution.

1.5.2.3 Drôme Department

The water backflow could be an essential factor of accidental contamination in drinking water systems. It is the result of water pressure variation, either an overpressure in case of private connection (with an irrigation network as an example), or a depression, for example due to pipes break, and which can be avoided by a non-return systems.

A recent epidemic of gastroenteritis has been highlighted in a district of a commune in the department of Drôme [56]. The source of the pollution was the backflow from an irrigation network to the drinking water network. This contamination in Drôme illustrates another source of accidental contamination. Irrigation systems or private wells should not be connected to drinking water systems.

1.5.2.4 Divonne-les-Bains (Ain), August-September, 2003

Between the end of August and the beginning of September 2003, a situation of accidental contamination was reported in the commune of Divonne-les-Bains. Several cases of gastroenteritis were reported among the affected population.

A survey was conducted in order to: i) confirm water pollution in the affected distribution network, ii) determine the source of contamination and iii) assess the health risk associated to this quality degradation. Several health authorities have been involved in this survey, such as the *DDASS* and “*Centre Interrégional d’Epidémiologie ”* (*CIRE)*.

A water backflow from the purification plant was the origin of the microbiological contamination. For a population of about 10,000 people, the estimated impact of the epidemic episode was close to 800 excess cases of Acute Gastroenteritis (GEA); the analysis executed for the distributed water and the resource showed frequent contamination with Cryptosporidium and Giardia [57].

The executed investigation was carried out at several levels:

- A survey of doctors to evaluate the rate of consultation for gastroenteritis cases during the period of contamination (between the end of August and the beginning of September 2003).
- A research in pharmacies to determine the sale of medicament used for treatment of gastroenteritis diseases.
- Microbiological analyzes of the stools of patients residing at Divonne-les-Bains for the research of gastroenteritis agents: bacteria, protozoa (Cryptosporidium and Giardia), and viruses (Rotavirus, Enterovirus, Adenovirus, Calicivirus, Astrovirus) and Hepatitis A virus [57].
- Environmental analysis, specifically water analysis to identify the cause of accidental pollution and the pathogens responsible for contamination.

These various surveys face several difficulties, such as: i) data transmission between the responsible authorities, ii) lack of complete information needed to assess the percentage of the affected population and iii) financial problems for executing certain analysis for parasites identification, for example, etc.

This example of accidental contamination highlights several problems in case of drinking water pollution. Among these problems, the difficulties of investigations, the assessment of health risks, the determination of the origin of pollution and responsible pathogens, etc.

1.6 Smart Network

A smart water network is an integrated set of products, solutions and systems that enables utilities to remotely and continuously monitor and diagnose problems, prioritize and manage maintenance issues, and use data to optimize all aspects of WDN performance [58]. Smart water grid will also be able to isolate contaminated section. For example, valves can be shut off remotely to avoid further displacement of contaminated water within WDN [59].

The transformation from a passive distribution network to a smart grid has many benefits: water loss reduction, energy savings, detection efficiency, water quality monitoring, etc. Figure 1.4 shows different advantages of using Information Technology (IT) in the WDN.

Three principal elements constitute the smart water network [60]:

- Information: making full use of all data generated by water utility.
- Integration: utilizing current IT systems to maximize previous investments.
- Innovation: designing a system flexible enough to meet future challenges.

Category	Savings as Percentage of Baseline Cost		Description
Leakage and Pressure Management	2.3 - 4.6	(3.5%)	Reduction in leakage levels by precise detection of leaks; Predictive modeling to estimate potential future leaks and pressure management
Strategic Capital Expenditure Prioritization	3.5 - 5.2	(12.5%)	Improved dynamic assessment, maintenance, replacement, planning and designing of network to optimize spending on infrastructure needs
Water Quality Monitoring	0.3 - 0.6	(0.4%)	Automatic water sampling, testing and quality monitoring; Reduction in costs from labor and truck rolls for manual sample collection
Network operations And Maintenance	1.0 - 2.1	(1.6%)	Real-time, automated valve/pump shutoff to facilitate flow redirection and shutoffs; more efficient and effective workflow planning
Total Smart Water Savings Opportunity	**7.1 – 12.5**	**(7.4%)**	

$U.S.billion

Figure 1.4. Summary of global savings by smart water solution [58] .

1.6.1 Role of Smart Technology

Nowadays, the smart technology is used in different urban networks (drinking water, electrical, heating and sewage). In WDN, the implementation of smart technology becomes an essential requisite to maintain a safe quality of water. Although water coming from treatment plan is well controlled, its quality can degrade through WDS. In this regard, there is a need for better on-line water monitoring systems given that existing laboratory-based methods are too slow to develop operational response and do not provide a level of public health protection in real time [61]. Water quality sensors can be used to ensure a real-time control. The main objective of sensors is to monitor continuously water quality parameters (pH, Conductivity, Turbidity, etc.) significantly affected by the presence of contaminants. Generally, there are two types of water quality sensors [62]:

- Non-compound specific or conventional sensors used normally for routine water quality parameters (Chlorine, Temperature, etc.).
- Compound specific water quality sensors or advanced sensors, which are capable of confirmative detection at low concentration for a specific component. Due to their high cost, the use of specialized sensors is not very effective, especially that the potential contaminant is unknown in real network.

Water utilities are also interested by remote sensing for economic reason. Twenty percent of average water quality monitoring cost is attributed to sample collection. Those cost could be reduced from 30 to 70 percent by moving from manual sampling to online monitoring, which lead to a global annual reduction of $120 to $270 million [58].

1.6.1.1 Need of field study

Although the use of smart technology in the field of water resources is desirable, its application faces many challenges. The use of sensors in the domain of water quality control is recent. Their efficiency is tested firstly in laboratory station. Different pilot-scale systems were studied in previous works [62], [63] where contaminant injection can be controlled. A case study is the station installed at the Laboratory of Civil Engineering and geo Environmental (LGCgE) [5]. The system tests the performance of two types of sensors for chemical and microbiological contaminants. It verifies their capacity in the contaminant detection. The feedback of the real-time monitoring of water quality remains very limited. Most of research consists of pilot station and few works are conducted in real-site. A field study is required to validate their reliability in online monitoring. The yields of smart water technologies are not well understood. Their instrumentation in real-network is needed to prove their benefits.

1.6.2 Early Warning System (EWS)

An Early Warning System (EWS) is an integrated system for monitoring, analyzing, interpreting, and communicating monitoring data, which can then be used to make decisions that are protective of public health and minimize unnecessary concern and inconvenience to the public [64]. To enhance the public health protection from water contamination events, a combination between online monitoring with a EWS is required. The EWS have many objectives to accomplish [65]:

- Identify low-probability/high-impact contamination events in sufficient time.
- Provide a fast and accurate means to distinguish between normal variations, contamination events and differences in quality due to biochemical and physical interactions.
- Detect both accidental and intentional contamination events with minimum of false positives and negatives.

An ideal EWS should be able to detect a wide range of contaminants, allows source identification and localization of the contaminant, and ensures alert management, remote operation and decision response. Four main elements, of same importance, constitute the base of the EWS [66]:

- Risk assessment: provides essential information for identifying strategy priorities for reducing and preventing disasters.
- Monitoring and forecast: need of devices of monitoring and forecast that can provide prediction about the potential risks affecting the community, economy and the environment.
- Forwarding information: need of communication systems that are capable of forwarding warning messages to potentially affected places and alarming the local and regional government organizations.
- Response: Key points of effective forecast: coordination, responsible governance, appropriate action plans.

1.6.3 Contamination Warning System (CWS)

In the context of drinking water control, Contamination Warning System (CWS) constitutes an essential component of the EWS. The main purpose of CWS is to protect earlier public health from the impacts of drinking water pollution. A CWS is a combination of monitors, institutional arrangements, analysis tools, emergency protocols, and response mechanisms designed to provide early warning of contaminants to minimize customer exposure [67]. It includes various approaches such as: water quality sensors deployment in the water grid, spot sampling and laboratory analysis, public health surveillance systems and customers complaint. It should have the capacity to quickly detect any deviation of water quality from accepted Standards.

The basic process for CWS operation (Figure 1.5) is described as follows, according to [68]:

- Monitoring and Surveillance: five basic components are required at this stage: online water quality monitoring, sampling and analysis, enhanced security, consumer complaint and public health surveillance. Monitoring and surveillance of these components and information streams occurs on a routine basis, until an anomaly from the baseline is detected.
- Event Detection and Possible Determination: Event detection is the process by which a deviation from the baseline is detected. If possible contamination is validated, credibility determination step is initiated, otherwise the CWS components return to routine monitoring and surveillance.
- Credibility determination: these procedures are performed using information from all CWS components as well as external resources when available and relevant. If contamination is determined to be credible, additional confirmatory and response actions are initiated.
- Confirmed determination: In this stage of consequence management, additional information is gathered and assessed to confirm drinking water contamination. Response actions initiated during credible determination are expanded and additional response activities may be implemented.
- Remediation and Recovery: Once contamination has been confirmed, and the immediate crisis has been addressed through response remediation and recovery actions defined in the consequence management plan are performed to restore the system to normal operations.

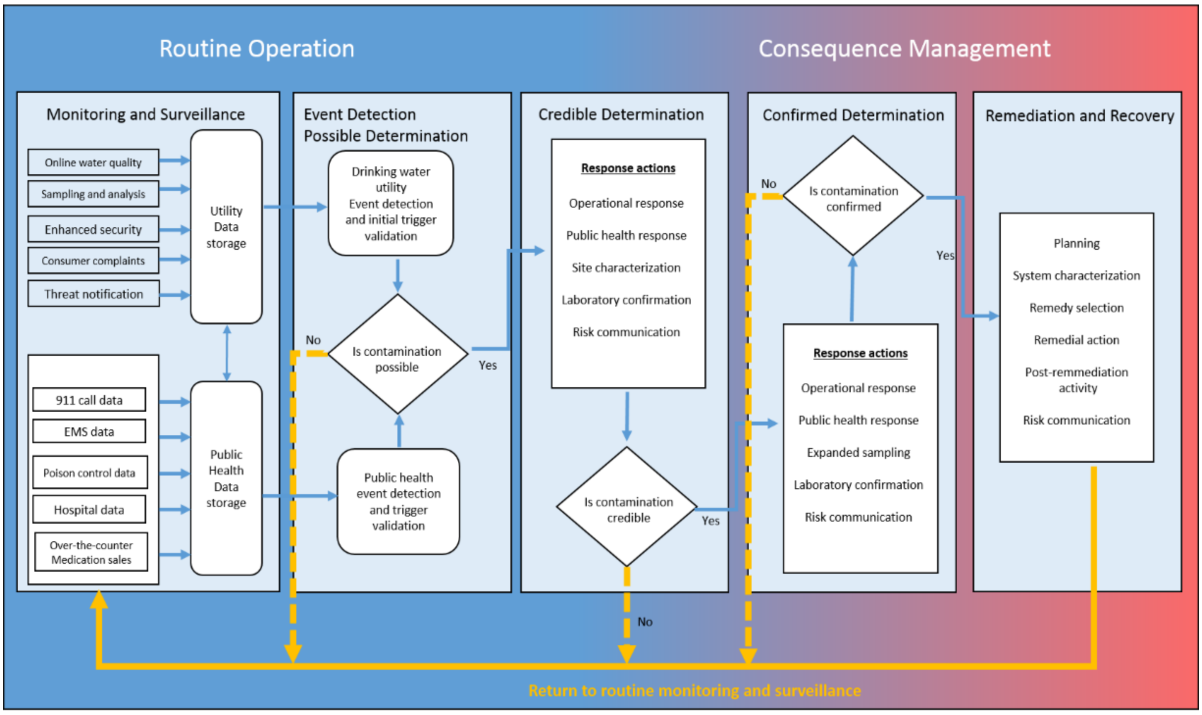


Figure 1.5. Architecture of the water security CWS [68].

The effectiveness of a CWS is based on the protection of public health, by reducing the contamination impacts. This objective highly depends on two factors: i) the time delay between the contamination event and the response actions, ii) the optimal placement of minimum number of sensors.

In a study in Ann Arbor, Michigan [69], it was demonstrated that the effectiveness of the CWS is reduced significantly for a total delay of time greater than 8 hours between the entry of contaminant in the distribution system and the cessation of water use by customers.

As mentioned above, data collected from online sensors should be analyzed to determine the probability of event. In order to detect a contamination, event detection method should be used. The existing methods for anomaly detection of water-contamination events based on online measurements of water-quality parameters are mainly divided into three categories, namely, statistical, artificial intelligence and data mining methods [70].

1.6.3.1 Statistical methods

These methods are based on time-series prediction. Based on historical measured data, the analysis of time-series allows the estimation of the future value. Statistical methods analyze one single parameter. This algorithm operates generally on normalized data with zero as mean and one as standard deviation. This normalization is done over the data contained within a moving window of previously measured value [71]. Different statistical models applied to the previously observed data can provide predictions of future water quality values [72]. Two principal time-series methods are generally used:

- Time Series Increments: It represents an implicit prediction model where the prediction of the water quality parameter value at the current time step, $\hat{x}$(n), is simply the value measured at the previous time step [71]:

$$\hat{x}\,(t) = x\,(t\text{-}1) \tag{1.1}$$

And the time series increments, δ (t) are calculated as follows:

$$\delta\,(t)=\hat{x}\,(t)\text{-}x\,(t)=x\,(t\text{-}1)\text{-}x\,(t) \tag{1.2}$$

This dependence on only the single previous measurement is the definition of a Markov model [73]. δ, expressed in units of standard deviation, is compared with a threshold δ_c to identify events. The length of moving window is p equal to the number of previous data.

- Linear filter: the Linear Prediction Correction Filter (LPCF) estimates the future value as a linear function of previous samples. It is based on a weighted sum of historical values. The most known criterion used in this method is the Autocorrelation (AR). The main objective is to minimize the squared prediction error. The prediction model is represented by the following equation:

$$\hat{x}\,(t) = -\sum_{i=1}^{p} a_i\, x\,(p\text{-}i) \tag{1.3}$$

Where $\hat{x}$ (t): predicted signal value,

x (p-i): previous observed value,

a_i: predictor coefficients.

p: order of the prediction filter polynomial and the size of the standardization window.

The error is then calculated by the following equation:

$$e\,(t)=\hat{x}\,(t)\text{-}x\,(t) \tag{1.4}$$

These two types of statistical methods have been tested in [71] to predict the variation in background water quality and to detect events of anomalous water quality superimposed upon the background models. Results shows that more developed algorithm, such as Multivariate Nearest-Neighbor (MVNN) (detailed latter), has a better performance than the traditional statistical methods.

1.6.3.2 Artificial Intelligence

These methods consist on the classification based techniques to extract models of normal and contaminated data samples automatically [74]. The Artificial Intelligence (AI) has many applications, in Finance, Medicine, Industries, Robotics, etc. There are two main types of AI:

- Artificial Neural Network (ANN): it is a form of computing inspired by the functioning of the brain and nervous system [75]. As well as the statistical method, ANN can be used as forecasting tool to predict future values based on historical set of data. However, with ANN methods, the statistical distribution of data does not have to be known [76]. Also, nonstationarities in data, such as seasonal variations, are accounted for by the hidden layer nodes [77]. One of the most used neural networks is the Perceptron. The linear Perceptron with threshold takes n values x_1, x_2, …x_n as input and calculates the output o as illustrated in Figure 1.6 [78], where w_i: Synaptic coefficients and θ: threshold. w_i are updated until the convergence, when the calculated output reaches the desired output.

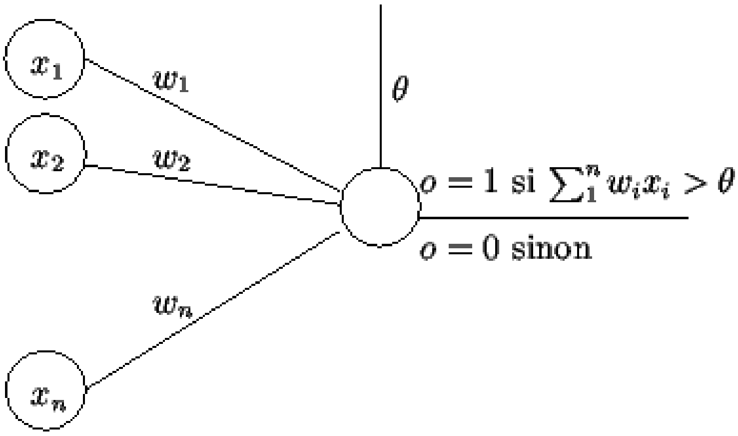


Figure 1.6. Perceptron with threshold [78].

The ANN method applied in [79] for water quality supervision shows a recognition rate up to 90 % with high capacity and rapidity in training phase. The main disadvantage remains in the choice of neurons numbers and hidden layers.

- Support Vector Machine (SVM): The main objective is to divide data according to two classes: normal and anomalous. This pattern recognizers is developed from the statistical learning theory. It is one of the Key area in machine learning [80]. It is based on the decision boundary that should be maximized for both classes. Among different types of SVM (monoclass, multiclass, etc.), the binary SVM is the simplest one. The base idea is to classify data in two main groups (+1 or -1), by separating them in a hyper plane. The optimal separator hyper plane, which maximizes the margin between the two classes, is obtained by solving the following problem [81]:

  Minimizing $\frac{1}{2}\|w\|^2$ such as: $y_i\ (w^T x_i + b) \geq 1 \forall\ i = 1, \ldots n$. Support vectors are those verifying the equality.

  Where w: weight vector of dimension m.
  b: term.
  x: example to be classified (input).
  y: output $\in\ \{+1; -1\}$

  As an application of SVM method, [82] use this technique to monitor the potability of a water tank. This method has shown a good performance in water quality classification with a low rate of false alarms. The time of training phase is of high importance in SVM technique.

1.6.3.3 Data-mining

In comparison with statistics methods, data mining techniques tend to be more robust to both messier real world data and also more robust to being used by less expert users [83]. The main purpose of data mining is to transform the huge amounts of data in useful information. The two essential type of these methods:

- Multivariate Nearest-Neighbor (MVNN): The MVNN approach compares the Euclidean distance between the current measured water quality and any water quality measurement within the recent past in the multivariate space [70]. All water quality signals provided from all available sensors are used simultaneously at each time step. The MVNN method indicates the similarity between the quality of tested water and a p previous historical samples. The minimum distance between the points, in J-dimensional space, is retained as the distance, Δ, which is compared to the threshold [72]:

$$\Delta = Min_{\,i=1\ldots p}\left|\sqrt[J]{\sum_{j=1}^{J} z^{j}(t+1) - Z^{j}(p-i+1)}\right| \tag{1.5}$$

Where J: number of water quality signals
p: number of water quality samples in the history window
$z^{j}(t+1)$: new water quality signal

- K-means: The main objective of k-means algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized [84]. The importance challenge in this method is that the number of cluster has to be specified first [85]. This is an iterative algorithm. It is a point-based clustering method that starts with the cluster centers initially placed at arbitrary positions and proceeds by moving at each step the cluster centers in order to minimize the clustering error [86].

A case study of data mining technique, especially the MVNN algorithm within Canary software, is applied to historical data from a UK drinking WDS. The performance evaluated with artificial events gives approximately 25 % of false negatives [2].

1.7 Smart Water Projects

- **Vigires' eau**

This is a system for drinking water network operation which proposes an original solution based on a network of sensors, existing tools and statistical detection algorithms [87]. It is a technical monitoring platform whose primary objective is the detection of any anomaly situation linked to the abnormal drop in the level of residual Chlorine. The Chlorine monitoring is based on its two main advantages: i) it constitutes a powerful oxidant, which have the capacity to destroy viral or bacterial agents, even at low doses and ii) it is a good tracer of water quality and network integrity since any contaminant in the network immediately gives abnormal Chlorine levels [88]. The aims of this prototype is to determine the optimal location of sensors. Figure 1.7 illustrates the principal components of Vigires'eau project. [87].

Figure 1.7. Main components of Vigires'eau Platform [87].

The detection is executed in the basis of two approaches (parametric and nonparametric). In the parametric approach, two models of Free Chlorine variation are established: i) the first expresses the normal reduction in the amount of Chlorine due to its reaction with microorganisms present in the water or with the walls of the pipes, and ii) the second studies the case of a rapid abnormal decrease resulting from a contamination phenomenon. An algorithm allows the detection of any anomaly affecting a given node of the network. It is based on the reduction of the probability of non-detection while avoiding as much as possible situations of false alarms. Due to uncertainties of drinking water systems, a parametric approach, based on an analytical model, is not sufficient for effective anomaly detection. A non-parametric approach based on statistical learning theory is needed. These methods of recognition consist in determining the distribution of a set of measurements. However, a nonparametric approach alone is not enough to understand the physical mechanism of contamination. So, Vigires'eau suggests a semi-parametric approach to combine the flexibility of nonparametric model with the precision of parametric one [89]. In the semi-parametric approach, the main purpose is to consider any observation as a sum of parametric and non-parametric model to prove any real anomaly with reducing the generation of false alarms.

- **SMaRT-Online^{WDN}**

(Online Security Management and Reliability Toolkit for Water distribution Networks)

This is a French-German cooperative project research whose main purpose is the development of an online security management toolkit for WDN that is based on sensor measurements of water quality as well as water quantity [90]. Concept and architecture of the project is illustrated in the Figure 1.8.

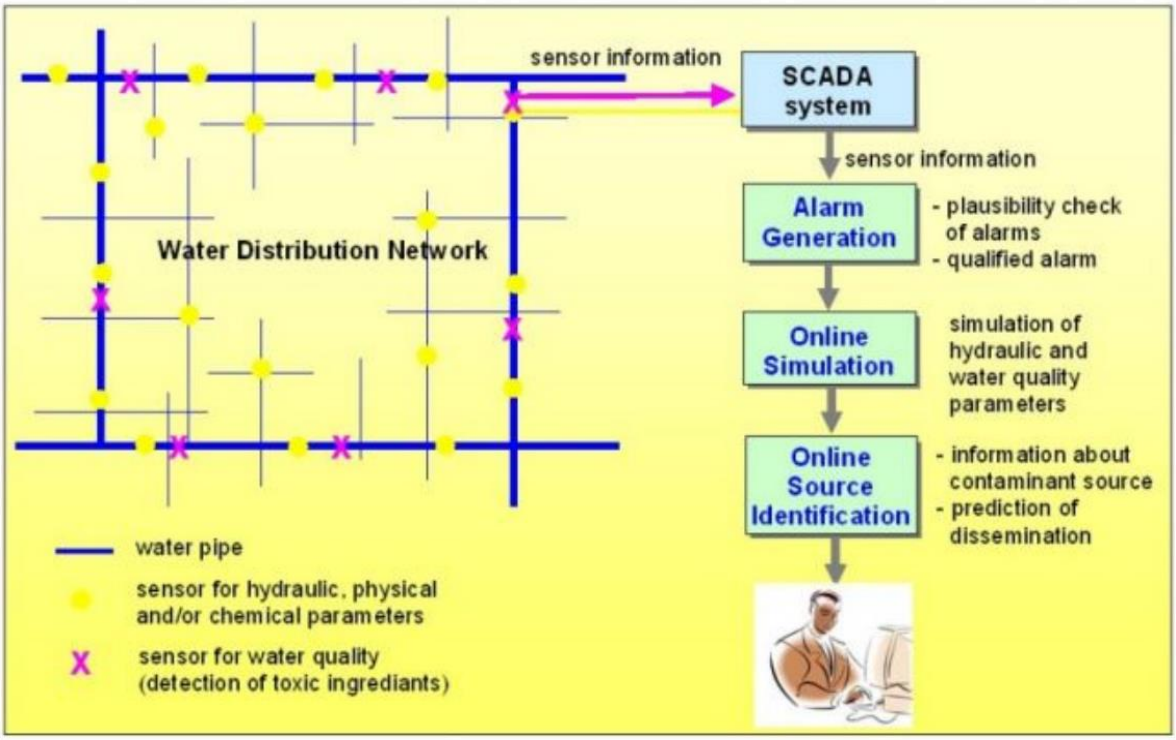


Figure 1.8. System architecture of SMaRT-Online^{WDN} [90].

The principal objectives of this project are divided into five main stages [91]:

- Smart quality sensors and alarm generation: Monitoring of water quality by an analytical model remains complicated due to the large number of parameters and then a high time consuming. One solution to this problem is the modeling technique based on the data learning method. A model based on the data history of the sensors can lead to the implementation of an effective monitoring system.
- Online simulation model for reliable predictions of water quantity and quality: To show efficiently the hydraulic state of the system, the use of intelligent sensors must be combined with a simulation model.
- Optimal sensor placement: Several research was conducted to evaluate the optimal location of the sensors in a distribution network. However, most of these studies, based on offline outputs, do not reflect the actual case of contaminant propagation. To solve this problem, an optimal location based on online measurements will be more appropriate. This requires a regular updating of the data.
- Online contaminant source identification: Deterministic and probabilistic methods for identifying the source of contamination are integrated into the simulation model to determine online the responsible sources.
- Risk analysis and impact assessment: Risk assessment focuses on three main areas: environmental, social and economic. Consumer perceptions can define and modify the management of water distribution.

- **AquaSense**

Before detailing the objective of this project, it is important to define the following terms:

- Aquaphotomics: a new technique for rapid water analysis based on the interaction between water and light.

- Near-Infrared Spectroscopy (NIRS): an imaging technique characterized by its non-invasive and portable approach, which allows activity evaluation where under some conditions the standard methods would not be applicable [92].

AquaSense project, funded by the European Union, aims to combine Aquaphotomics with NIRS and Chemical Imaging (NIR-CI) to ensure real-time water analysis. This technique can detect and identify contaminations affecting the water quality. However, the problem lies in the low concentrations of contaminants where detection will be more complex.

1.8 Conclusion

Ensuring safe and potable drinking water is the great challenge for water utilities. Accidental or intentional contamination can deteriorate the quality of water and threat the public health. This chapter presents a state of the art of the water quality control in WDS. Contaminants that could present in the WDS have been detailed: i) microbiological parameters (E.Coli, Enterococci, etc.) and ii) chemical parameters (Sulfates, Pesticides, etc.). To control the amount of these substance, Standards set threshold limits. In France, Standards, indicated in the Public Health Code, are fixed by the European Directive, based on the *WHO* guidelines. Over years, the water quality supervision is developed to monitor indicators parameters such as Turbidity, TOC, Conductivity, etc. Many factors can be the source of contamination, among them the corrosion, water backflow, network interventions and pipes break. These factors induce the introduction of pollutant and/or the bacteria growth through the WDS. This can lead to serious risks on human health. The impacts depend on many criteria: the nature of contaminant, its amount, the age of consumer and his immunity system.

In general, the water quality supervision consists of taking manually samples from different points of the network. Laboratory analyses are then executed. Different conventional methods have been used in the literature to analyze the water quality. Among them, the culture method, the chemical and the aesthetics analyses. The advantage and the limitation of each method have been presented. **The major inconvenience of all these conventional methods is the long delay (from several hours to many days) to obtain results.** Much research has demonstrated the importance of the WDS security since the global frequency of contamination is significant. The percentage of waterborne disease is critically high. Especially in France, many cases of contamination have been occurred in different regions, leading to multiple diseases.

This literature review shows both the complexity of the early detection of water contamination and the great potential of the Smart Technology to meet this challenge. However, this technology is recent, consequently we need more research in this area with mainly demonstration projects in real field condition. This research work constitutes a contribution to this objective. It concerns the implementation and use of the Smart Technology for water quality control at the Scientific Campus of Lille University. It is conducted within the European Project “SmartWater4Europe”. In the following chapters, we present the demonstration site, the smart monitoring, data analysis and try to propose some recommendations for the use of this technology.

Chapter 2. Presentation of the Demonstration site – Water Quality Devices Installation

2.1 Introduction

This chapter presents the demonstration site, which is used in this research. It is a part of SunRise project, which concerns the construction of a large-scale demonstrator of the Smart City. The Scientific Campus of the University of Lille is used for this large-scale demonstrator. This chapter presents successively the Scientific Campus, the water distribution network, the water quality sensors used in this work, results obtained in a Lab pilot and finally the implementation of the water quality sensors in the demonstration site.

2.2 Demonstration Site Description

The demonstrator site used in this research is the Scientific Campus of the University of Lille. The campus, inaugurated in 1967, is located in Villeneuve d'Ascq, Lille in the north of France. It stands for a small town, covering an area of around 110 hectares. It includes 150 buildings for different uses:

- The University of Lille, Sciences and Technology: with 2 Engineering school (Polytech'Lille and Telecom Lille1) and 2 Institutes (IUT A and CUEEP).
- "Ecole Centrale de Lille" (ECL).
- "Ecole Nationale Supérieure de Chimie de Lille" (ENSCL).
- "Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux " (IFSTAR).
- Residences, Restaurants, and sport equipment.

It hosts around 25000 users, among them about 4000 students live in the campus. The campus includes about 100 km of urban networks: drinking water, electrical network, heating and sewage. Figure 2.1 illustrates the plan of the Scientific Campus.

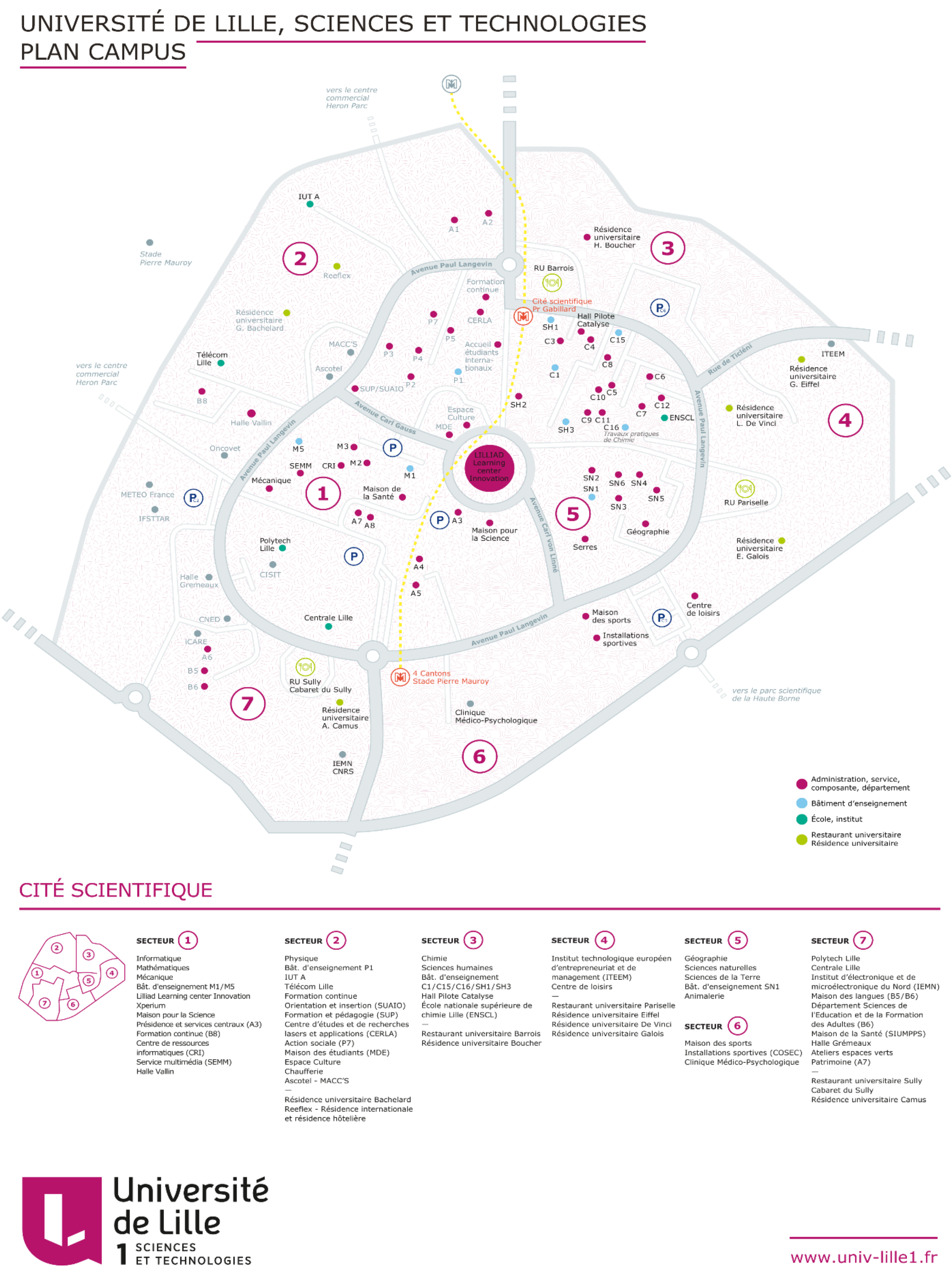


Figure 2.1. Scientific Campus of the University of Lille.

2.3 Water Distribution Network

The water distribution network length is about 15 km: 13.5 km for private sector (Campus) and 1.5 km for public sector (Eaux du Nord). The pipes are made, in majority, of grey cast iron. The diameter of pipes varies between 20 and 300 mm, as indicated in Figure 2.2. It includes 49 hydrants, 250 isolations valves, purges and stabilizers. The water is supplied to the campus at five sections located in the North, West and South of the campus:

- Cité Scientifique.
- 4 Cantons.
- ECL
- Bachelard.
- Building M5.

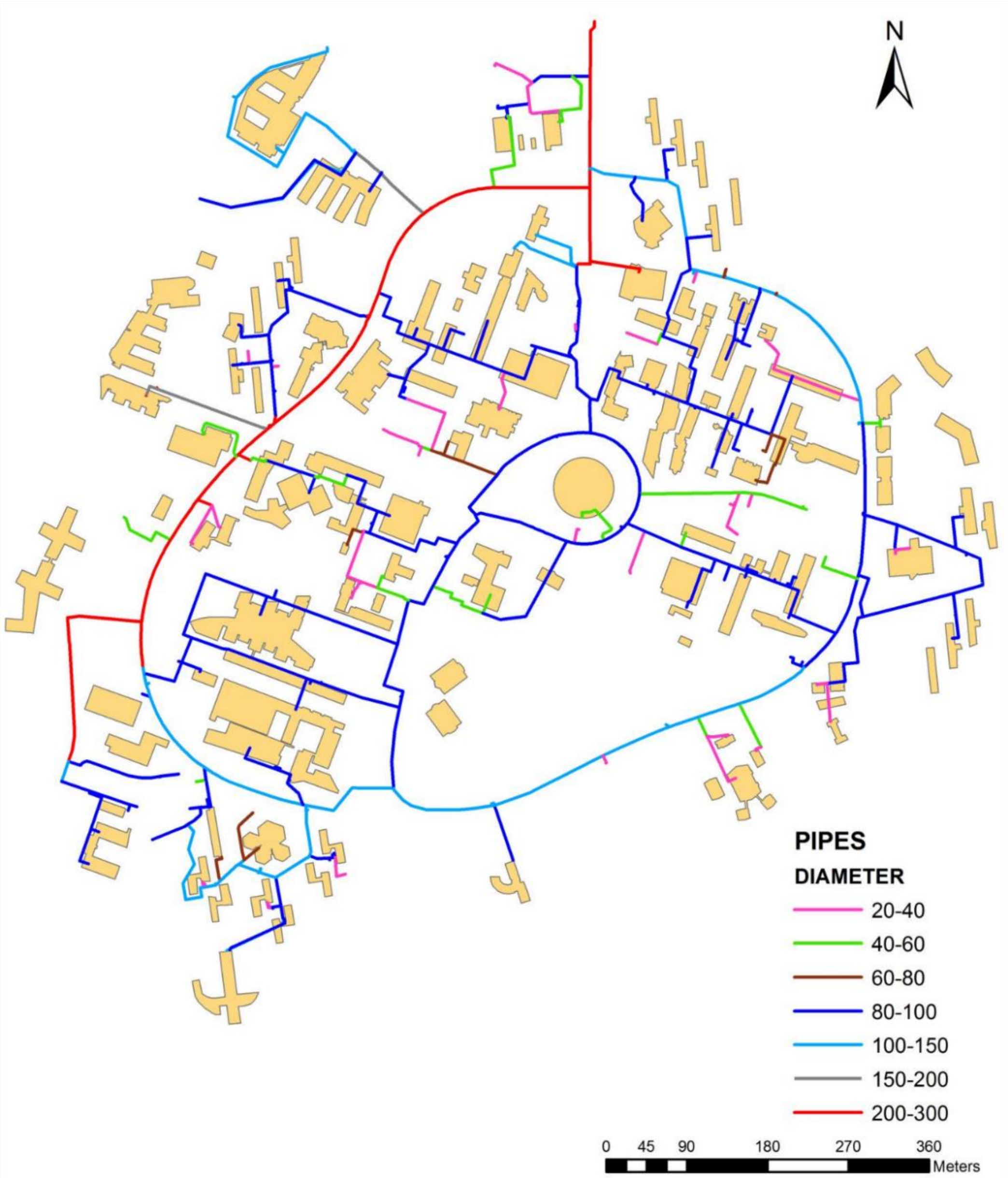


Figure 2.2. Pipes distribution according to the diameter [93].

2.3.1 Water network instrumentation (Farah thesis, 2016, [93])

The water network is equipped with 93 Automatic Meters Readers (AMR). AMR readings are recorded every one hour and transmitted every 24 hours, to a central server via a radio frequency of 169 MHz. The water supply in the campus is calculated according to 13 general meters regrouped as follows (Figure 2.3) [93]:

- 4 AMRs entitled: 4CANTONS, ECL, BACHELARD, M5 and 5 AMRs CITE SCIENTIFIQUE: looped network.
- 4 AMRs entitled: CUEEP, DELTEC_ICARE, HALLE-VALIN and LML: branched network.

The consumption of the main buildings is measured by 80 AMRs distributed as follows [93]:

- University of Lille (55).
- Reeflex (2).
- CROUS (Centre Régional des Œuvres Universitaires et Scolaires) (14).
- ENSCL (1).
- ECL (6).
- The company Bonduelle and the Clinic 4Cantons (2).

The pressure is also monitored by a set of 5 piezoresistive pressure sensors. Data loggers, attached to the sensors, send pressure values measured each 15 minute via SMS [93]. The pressure sensors are installed in different locations in the campus, covering the majority of zones (Figure 2.3):

- Barroi Restaurant.
- Building C1 for chemical research.
- SN5 for biology research.
- Polytech'Lille University.
- Bachelard university residence.

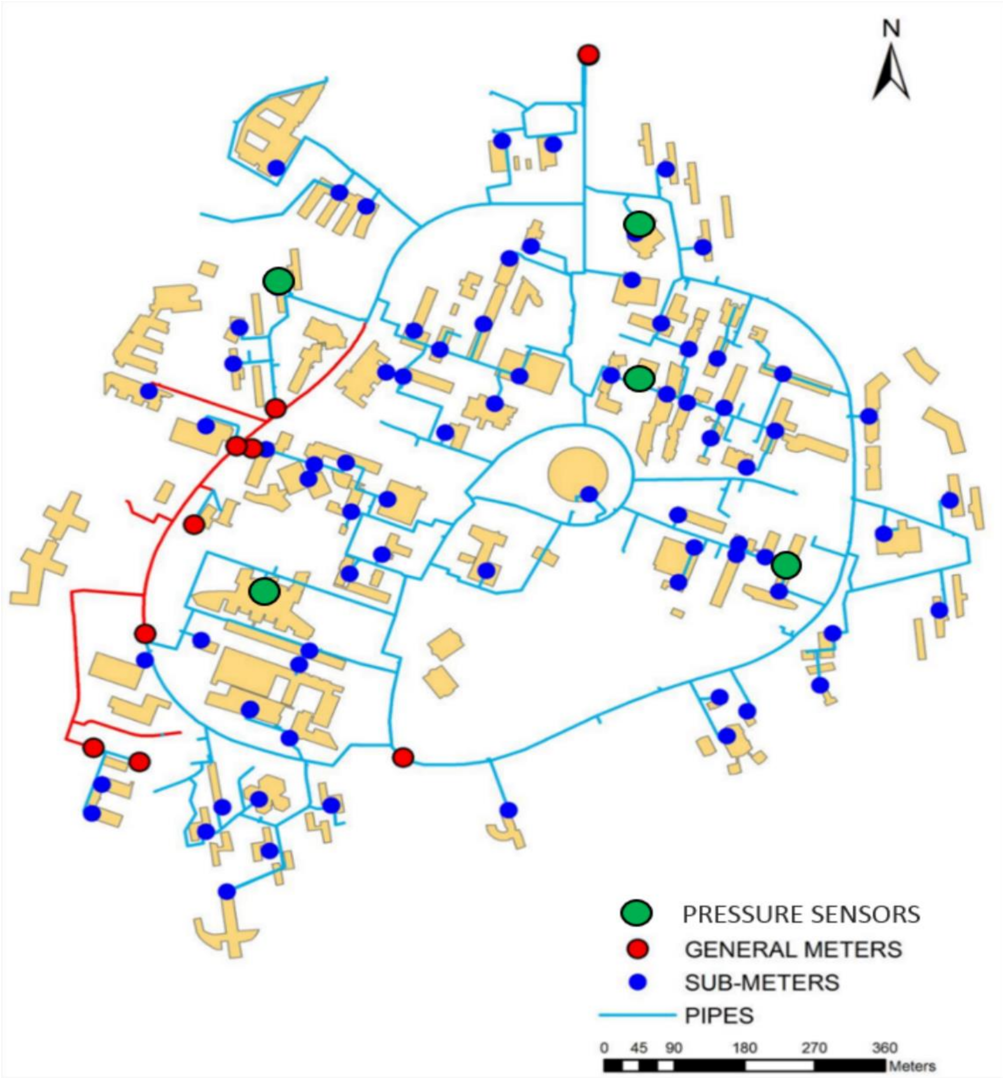


Figure 2.3. Distribution of the smart sensors in the water network [93].

2.4 Water quality sensors

2.4.1 S::CAN

S::CAN micro::station allows online monitoring of various water quality parameters. It is used for real time control of drinking water quality. Figure 2.4 illustrates the S::CAN sensor. The main components-spectro::lyser, sensors and controller- are assembled with required flow cells, mounting fittings and pipework on a compact panel [94]. A con::cube terminal with moni::tool software ensures data acquisition, data display and station control [95].

Different elements constitutes the micro::station S::CAN [95]:

- Con::cube controller with moni::tool software.
- Flow cell with auto brush cleaning device to provide cleaning of the optical measuring windows.
- System tubing included in the panel assembly.
- Flow detector that gives alarm if the flow decreases below 0.25 L/min. The recommended flow is about 0.5 L/min.
- Inlet strainer that avoids the entrance of coarse material in the station.
- Pressure transmitter that supplies the pressure signal to the con::cube.

- Main panel that assembles all components.
- Flow restrictor for automatic flow restriction and backflow prevention in by-pass.
- Probes (i::scan, pH::lyser, Chlori::lyser and Condu::lyser) ensuring the continuous measurements of water quality parameters.

Each parameter, measured every one minute by S::CAN, is represented by a signal. For a normal drinking water, a reference line is established for every parameter. The main purpose of S::CAN probes is to detect significant deviation from the stable line. An exceeding of the reference will be analyzed to identify the origin of the abnormality. Many reasons can explain the perturbation in the signals, such as normal variations, instrument faults, contaminations, connections errors, etc. A change in water quality signal can be divided in three main groups:

- Outlier: a significant deviation, generally due to noise in Supervisory Control and Data Acquisition (SCADA) system, at a single time step. A sudden increase or decrease in single time steps do not generate an alarm. The unexpected values return to the reference line at the next time step.
- Event: deviation from expected signal in a specified period of time steps. It can be defined as a group of outliers which values are unacceptable for a period of time. This event will generate an alarm to analyze the source of the abnormalities. The number of time steps and the amplitude of deviation determines the importance of the event.
- Baseline change: This change happens suddenly but induces a persistent variation in the mean value of the water quality signals. Generally, it results from operational variations, such as turning on or off a pump delivering water with different quality characteristics.

Figure 2.4. S::CAN sensor [96].

2.4.1.1 i::scan

i::scan probes are multiwavelength photometer probes, capable of online measurements of absorption spectra (UV, UV-Vis, UV-Vis-Nir, or derived parameters) [97]. Figure 2.5 illustrates the i::scan probe. It allows the monitoring of various reliable indicator of water quality parameters:

- Absorbance UV, for a wavelength of 254 nm, in Abs/m.
- Turbidity according to *International Standards Organization ISO* 7021 in FNU (Formazin Nephelometric Units) or FTU (Formazin Turbidity Units) and also to *EPA* 180.1 in NTU.
- Organic substances: TOC and DOC in mg/l.
- Temperature in °C.
- Color in Hazen.

The probe can be used immediately after delivery. Global calibration is available for typical applications, such as drinking water. However, a local calibration can be performed on site, without demounting the probe, to get more adjusted parameters [97].

Figure 2.5. i::scan probe [97].

2.4.1.2 pH::lyser

pH::lyser is an ion-selective device designed for continuous monitoring of the logarithmic concentration of dissolved hydrogen ions (H^+) [98]. It measures also the temperature and then corrects the measured concentration. Figure 2.6 illustrates a pH::lyser. As for i::scan, the probe is precalibrated in the factory. However, adjustments (offset or linear calibration) for measurements can be performed for more accurate data.

Figure 2.6. pH::lyser probe [98].

2.4.1.3 Condu::lyser

Condu::lyser is a probe designed for continuous monitoring of the Conductivity in water, expressed in µS/cm [99]. The temperature is also measured and used to correct the corresponding measured Conductivity. Figure 2.7 shows the Condu::lyser probe. Such others probes, local calibration, especially slope calibration (SPAN), can adapt the global calibration to the actual monitored parameters.

Figure 2.7. Condu::lyser probe [99].

2.4.1.4 Chlori::lyser

Chlori::lyser is an electrochemical based sensor designed for the continuous monitoring of Free Chlorine (concentration of residual Chlorine) in water, expressed in mg/l [100]. A slope calibration (SPAN) adjusts the monitored values. Two main elements are essential in this probe: i) electrolyte, ii) membrane cap. The electrolyte needs to be replaced every 3-6 months, while the membrane cap needs to be replaced once per year or if the local calibration failed [100]. Figure 2.8 illustrates a Chlori::lyser probe.

Figure 2.8. Chlori::lyser probe [100].

The specifications for each parameter, including the measuring principle, units, measuring range, resolution and accuracy are presented in Table 2.1.

Table 2.1. Specifications for each parameter of S::CAN sensor ([97]-[100]).

Parameter	Measuring principle	Units	Measuring Range	Resolution	Accuracy
UV	Absorption	Abs/m	[0-60]	0,015	±10 %
TurbidityEPA	90 degree scattered light	NTU	[0-800]	0,001	±7 %
TurbidityISO	90 degree scattered light	FNU	[0-800]	0,001	±2,5 %
TOC	Absorption	mg/l	[0,1-25]	0,01	±3 %
DOC	Absorption	mg/l	[0-450]	0,035	±2,5 %
Temperature	Semiconductor	°C	[-20°C-70°C]	0,0625	0,5 (0°C-65°C) 1 (-20°C-70°C)
Color	Absorption	Hazen	[1-70] mg/l	0,01 mg/l	±2,5 %
pH	Potentiometric		[2-12]	0,01	±0,01
Conductivity	4 electrode contacting	µS/cm	[0-500,000]	1	±0,1% of current reading
Free Chlorine	Amperometric 3 electrode sensor	mg/l	[0-2]	0,001	0,23

2.4.1.5 Data Transmission

Using a 3G SIM card, the transmission of data is ensured via a web server that can be connected to a Supervisory Control and Data Acquisition (SCADA) system.

In order to access to the nano::station, a Virtual Private Network (VPN) server with specific username and password is used. Once the OpenVPN is connected, the con::cube access is provided via a web browser (Chrome, Mozilla, etc.) by using a specific Internet Protocol (IP) address. The same VPN server is used for both locations, however, the IP address differs between the two stations. When the IP address is reached, measurements can be monitored in real-time. Historical data are also displayed in time series form. By login to the service mode (with username and password), outputs can be downloaded in Comma-Separated Value (CSV) files. All others options of the con::cube can be controlled remotely through the VPN connection. An example of data display is illustrated in Figure 2.9.

Figure 2.9. Example of data display from S::CAN sensor.

2.4.1.6 Data archiving

Data transmitted from sensors are stored and analyzed using PI system, which collects, stores and manages data. Figure 2.10 illustrates a typical diagram for PI system components. Through PI interface nodes, information collected from data source, are stored in PI tags on data archive. The access to data is done either directly from data archive or from Asset Framework (AF) server [101].

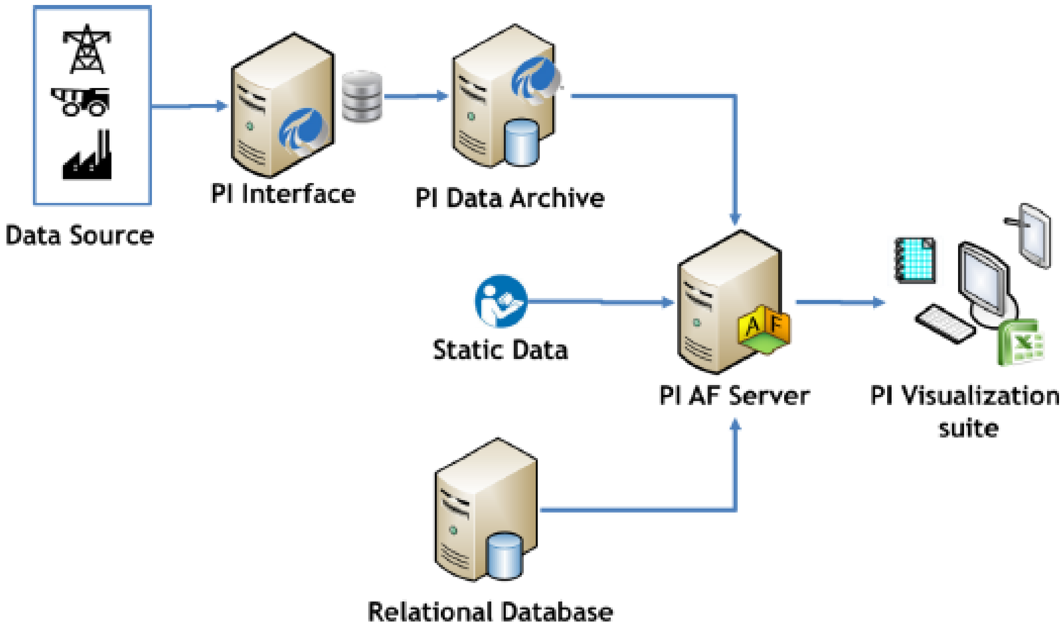


Figure 2.10. Diagram of the components of a typical PI system [101].

According to the measured parameters, 13 tags have been created, for S::CAN installed at Polytech'Lille: UV254, Turbidity ISO, Turbidity EPA, TOC (Total Organic Carbon), Temperature1 (probe i::scan), Temperature2 (probe Condu::lyser), T10, R alarm, pH, Free Chlorine, Flow, Conductivity and Color. For S::CAN installed at Barroi, 9 tags are stored in

PI system: UV254, Turbidity ISO, Turbidity EPA, TOC, Temperature, pH, Free Chlorine, DOC (Dissolved Organic Carbon) and Conductivity.

Data collected continuously from S::CAN are installed in the PI system. A PI Datalink, is used, as Microsoft Excel add-in, to import data of historical period in a spreadsheet. Archive, Compressed, Sampled data and other tools are used for data gathering and monitoring.

2.4.2 Optiqua EventLab

EventLab sensor uses an optical probe that ensures an Early Warning System (EWS) for water distribution grid, without the need of reagents or any consumables (Figure 2.11). It can be installed and accessed for servicing without interrupting the main flow [102]. This sensor is controlled remotely via a web server. EventLab meets four key requirements of an EWS [103]:

- Continuous real-time detection.
- Generic, one sensor covering full spectrum of possible chemical contaminants.
- High sensitivity.
- Low cost and low maintenance (no consumables).

Figure 2.11. Optiqua EventLab sensor.

EventLab measures each minute the change in Refractive Index (RI). The RI is an effective indicator of water quality. Any substance dissolved in water affects the RI of the water matrix [103]. RI has many benefits that ensure an early alert of water contamination. The use of RI has a number of advantages for water quality monitoring and the detection of water quality incidents [104]:

- The only generic parameter available: detect all chemical changes, irrespective of their nature, while others sensors are sensitive only to a part of the spectrum of contaminants.
- Consistency in response: the consistence sensitivity of EventLab allows an estimation of the order of magnitude of a contamination event. The sensor operates at a sensitivity level equivalent to parts per million (ppm) levels for any chemical contaminant [105].

- Response linear with concentration: a linear relationship is maintained between RI and the concentration of contaminant [106].
- High resistance to matrix interference: with RI, there is no dependence on matrix effects. The only factor that can affect the RI is the temperature. Its effect is fully accounted for in the compensation mechanism of EventLab system.

The variation of RI is illustrated by the measurement of variation of phase to which it is directly proportional. In a normal drinking water, the RI is quasi-stable as well as the variation of phase. A perturbation in the baseline will be analyzed to identify the possible existence of contaminant.

The system is based on the Mach-Zehnder Interferometer (MZI) principle as illustrated in Figure 2.12. The basic layout of the MZI consists of an input channel wave-guide that splits into two identical branches, which are then combined again to form the output wave-guide [107].

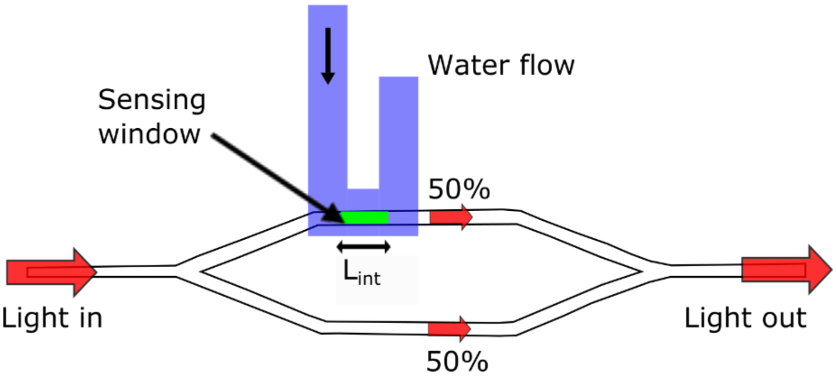


Figure 2.12. Basic Layout of the Optiqua MZI sensor [103].

The main output signal of EventLab is the variation of phase $\Delta\Phi_{m\ (measured)}$ of light propagating over the interaction window, given by the following equation [106]:

$$\Delta\Phi_m = (2\pi/\lambda)\, L_{int}\, (\partial n_{eff}/\partial n_{water})\, \Delta n_{water} \quad [radians] \tag{2.1}$$

where λ is the wavelength of the light in vacuum, L_{int} is the interaction light of the sensing window, $\partial n_{eff}/\partial n_{water}$=0.21 and Δn_{water} is the refractive index change in the water. With λ=850 nm and L_{int}=10 mm and using (eqn 2.1), the refractive index change Δn_{water} is given by:

$$\Delta n_{water} = 4x10^{-4}\, (\Delta\Phi_m/2\pi) \tag{2.2}$$

At each time step i, the system measures the phase Φ (i), then its variation is calculated as follows:

$$\Delta\Phi\,(i) = \Phi\,(i+1) - \Phi\,(i) \tag{2.3}$$

In addition to the phase, EventLab measures other parameters that should be monitored continuously:

- Temperature based on Resistance Temperature Detector (RTD).
- Signal health and signal level as indicators of probe status and the need of maintenance.
- Response which is the phase corrected after taken into consideration the effect of temperature.
- F24 Response which indicates how much a response value is above or below the average of the preceding 24 hours.

To ensure correct monitoring, some operating conditions should be verified; a water flow of 0.1-0.5 L/min is controlled by flow regulator and generally a pressure of 2 bar adjusted by the pressure reduction valve [108].

2.4.2.1 Data transmission

Using a 3G SIM card, a Wireless communication allows a continuous data transmission. A GPRS (General Packet Radio Service) modem allows data integration into existing data infrastructures [102]. Using a web browser, the system can be controlled remotely through the website of Optiqua (*https://optiqua.eventlabonline.com*). A username and password are required to reach the website. Data are displayed in real-time and can be collected in csv file. An example of data transmission is illustrated in Figure 2.13.

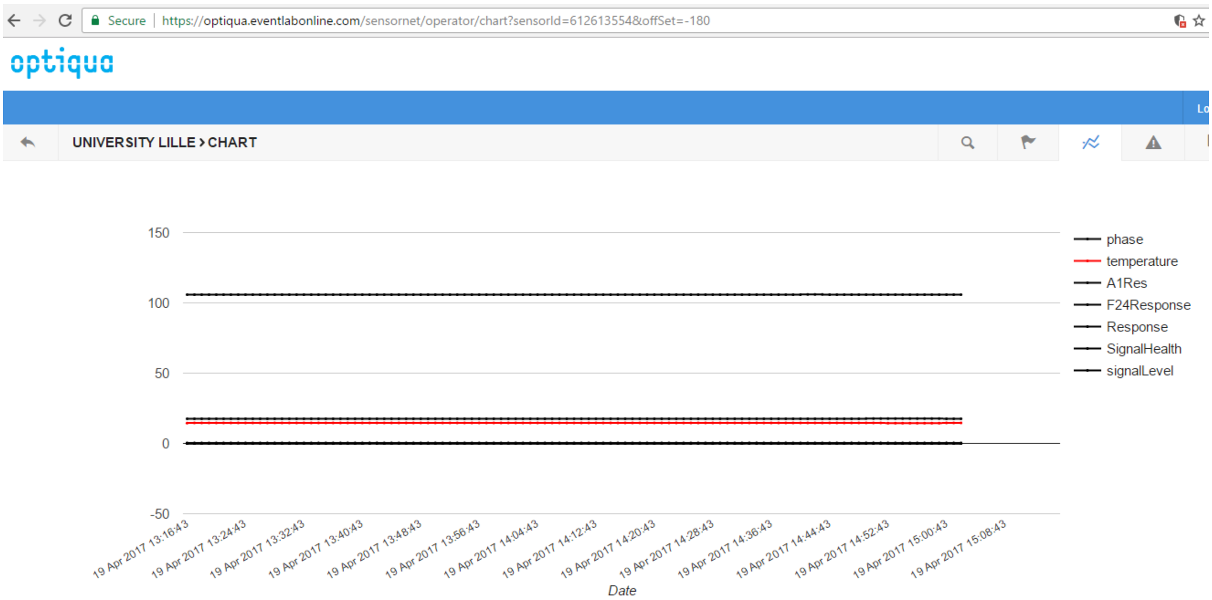


Figure 2.13. Example of data transmission from EventLab sensor.

2.4.2.2 Data archiving

For EventLab sensor, data collecting and archiving were done in the same way as S::CAN, using PI system. The tags created are equal to the number of parameters measured. We archive 12 parameters: the algorithms A1Res, A2Res, A3Res, DetectorA1, DetectorA2 and DetectorA3, then F24 Response, phase, Response, signal health, signal level and temperature. PI Datalink is used to report data in excel and then analyze the results.

2.4.3 Data Management

Data can be visualized using PI System Explorer, as a configuration and management tool for AF. In PI System Explorer, elements represent either physical or logical entities which can be organized in several hierarchies [101]. In Lille demo site, smart water network comprising pressure sensors, water meters [93] and water quality sensors are assembled in the group of SUNRISE Smart Water. Water quality sensors are divided into two main groups: i) EventLab sensor and ii) SCAN sensor, and then each one contains a sub-group referring to the installation location. Data are assembled according to their category (real time data or static data) and the last added values are displayed (Figure 2.14).

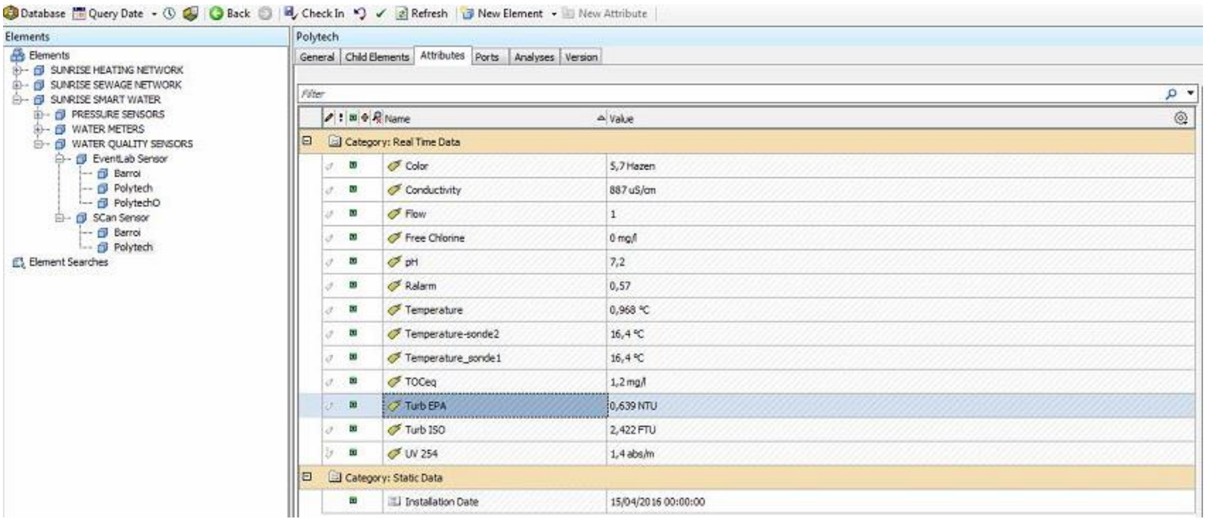


Figure 2.14. Data in PI System Explorer.

2.5 Lab pilot (Abdallah thesis, 2015, [5])

The feedback of water quality sensors in the online supervision is limited. Since the WDN is complex, the performance of sensors should be tested before the installation in real site. For this objective, a Lab pilot has been installed previously in the Laboratory of Civil Engineering and geo-Environmental (LGCgE). The pilot station reproduces the same conditions of WDN in terms of pressure, velocity, materials, etc. [5]. Figure 2.15 illustrates the Lab pilot with the different hydraulic components: pipes, tanks, discharge system, injection system of contaminants, pumps and other hydraulic equipment.

The performance of S::CAN and EventLab, installed at 41 m from the injection system, has been tested. Two types of contaminants have been injected at different concentrations: i) biological (E.Coli and Enterococcus faecalis), ii) chemical (Sodium Hypochlorite NaClO, Glyphosate (N-phosphométhyl) glycine $C_3H_8NO_5P$, Cadmium Chloride $CdCl_2$ and Chloride of Mercury $HgCl_2$) [5].

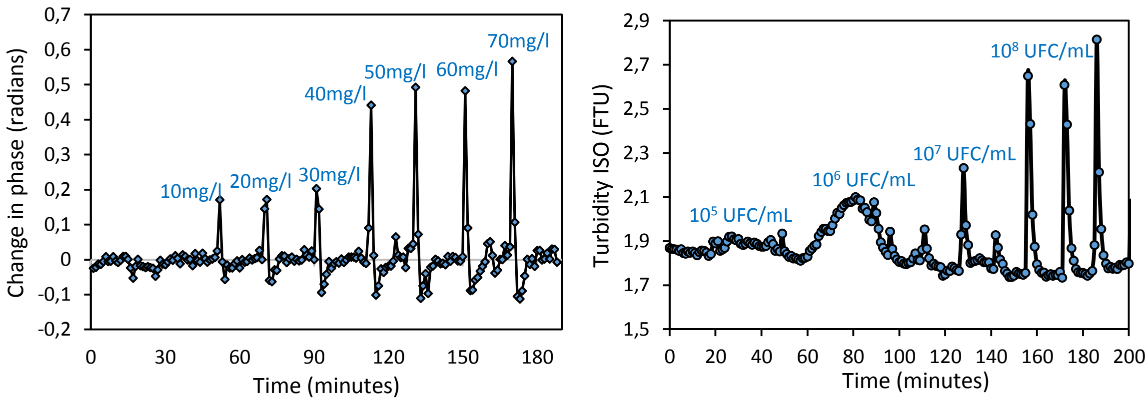
Figure 2.15. Lab pilot for water quality control [5].

2.5.1 EventLab response

The response of EventLab is illustrated by the variation of phase, after taken into consideration the effect of temperature. EventLab proves a high efficiency in the detection of chemical contaminants, even at low concentrations [5]. Figure 2.16 (a) shows an example of EventLab response after the injection of Chloride of Mercury ($HgCl_2$). Peak values have been observed when the contaminant reaches the sensor. However, EventLab did not detect microbiological contaminants.

2.5.2 S::CAN response

For chemical contaminants, S::CAN has proved a good capacity of detection. This is illustrated by the augmentation of signals after the injection of contaminations. For microbiological contaminants, S::CAN could detect this type from a bacterial concentration of 10^6 UFC/ml [5]. The amplitude of the peak is proportional to the injected concentration, as showed in Figure 2.16 (b) for E.Coli example.



(a) (b)

Figure 2.16. Sensors response. (a) Variation of phase after the injection of Glyphosate; (b) Variation of Turbidity after the injection of E.Coli [5].

2.6 Installation of water quality devices at the demonstration site (Scientific Campus)

To control the water quality in the Scientific Campus, S::CAN and EventLab have been installed in the water network. The optimal placement of sensors depends on many factors:

- Cover the majority of the campus.
- Compare the water quality in different types of building's usage.
- Ease of installation.

According to these factors, five locations have been prepared for sensors installation (Figure 2.17):

- The University residence Bachelard, which is the larger residence in the campus.
- The engineering school Polytech'Lille.
- The building SN5, research and teaching building in the biology sector.
- The building C1, teaching building in the chemical sector.
- The University Restaurant Barroi.

Because of budget restrictions, devices were only installed in two locations:

- The engineering school Polytech'Lille.
- The University Restaurant Barroi.

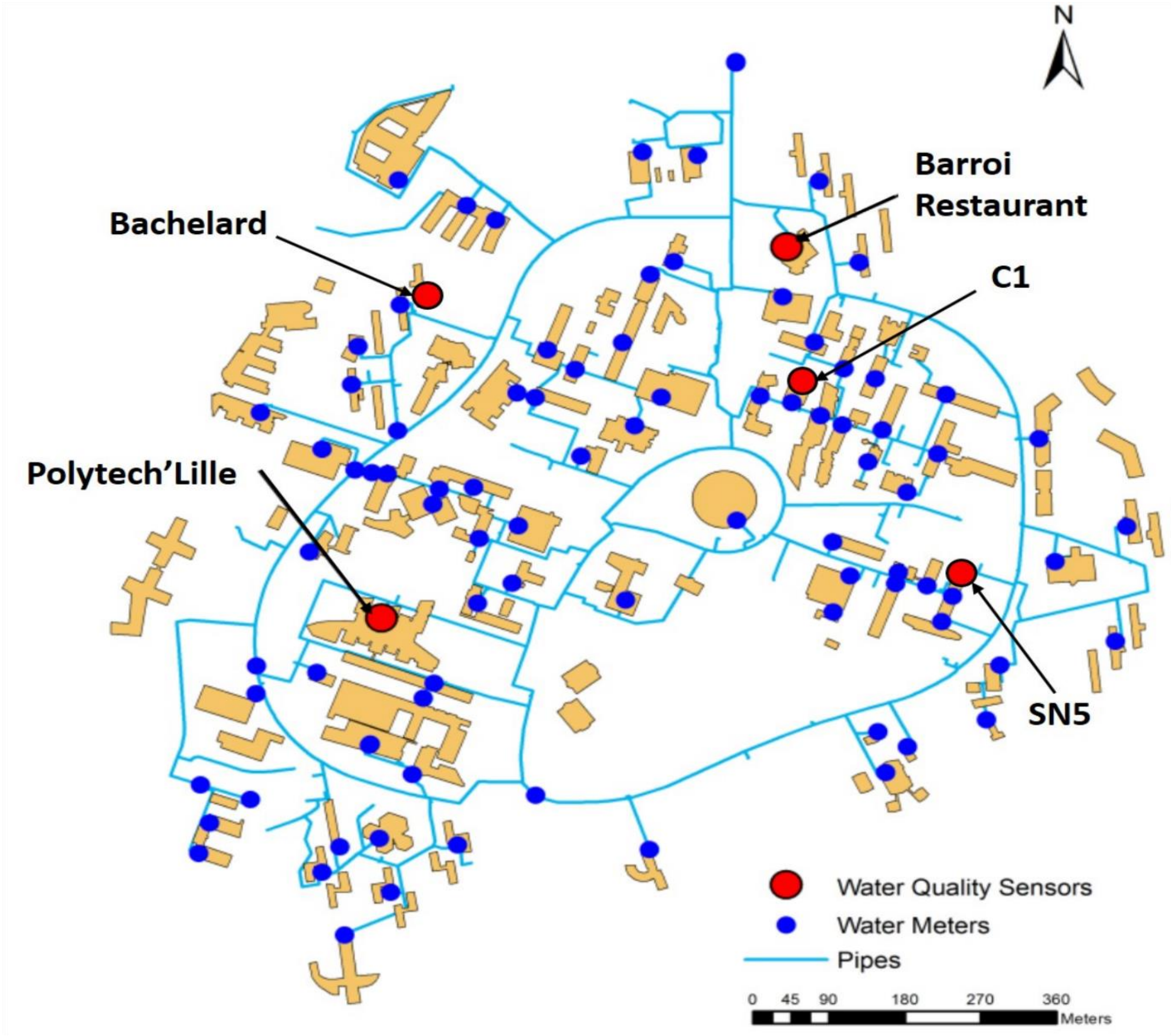


Figure 2.17. Locations of water quality sensors.

2.6.1 Installation at Polytech'Lille

Both devices (S::CAN and EventLab) were fixed on one panel and installed in the technical room of Polytech'Lille in April 2016. A derivation from the water pipe was used for sensors water supply. A check valve is used to prevent a backflow in the main system. Water samples used by the devices passes through an evacuation system. The technical room electricity was used for the powers supply. Figure 2.18 and Figure 2.19 display some details of the installations.

(a) (b)

Figure 2.18. Sensors at Polytech'Lille. (a) S::CAN; (b) EventLab.

(a) (b)

Figure 2.19. Installations details at Polytech'Lille. (a) Water supply; (b) Discharge systems.

2.6.2 Installation at Barroi restaurant

EventLab was installed in October 2016, while S::CAN was installed in November 2016. The installation was conducted following the method used in Polytech'Lille. A main pipeline passes near the target location. A connection is taken from this pipe and sensors are placed on it. Valves system, ensuring water samples to the sensors, are also fixed on a panel system. Water is then evacuated in a discharge. The installation details are showed in Figure 2.20.

Figure 2.20. Installations details at Barroi.

2.6.3 Maintenance and cleaning

During the exploitation of sensors, it is important to make sure that they are clean. The maintenance and cleaning tasks depend on the type of sensors.

For S::CAN, each probe has a specific procedure to be cleaned according to the manufacturer. Cleaning can be done in a periodic basis or in case of error indicated by the function check. This procedure requires the use of distilled water and sometimes other liquids such as Ethanol or Isopropanol. In particular, the maintenance of Chlori::lyser should be done very carefully. It can be divided into two steps [100]:

- Replacement of Electrolyte every 3-6 months or if the local calibration failed.
- Replacement of the membrane cap once a year or if the local calibration failed although the electrolyte has been replaced.

The cleaning of Chlori::lyser requires the use of specific materials (Gel-Electrolyte E-507) and plastic sheet. Moreover, once it is installed, Chlori::lyser requires a continuous flow without any interruption. Figure 2.21 illustrates a part of the cleaning procedure of Chlori::lyser.

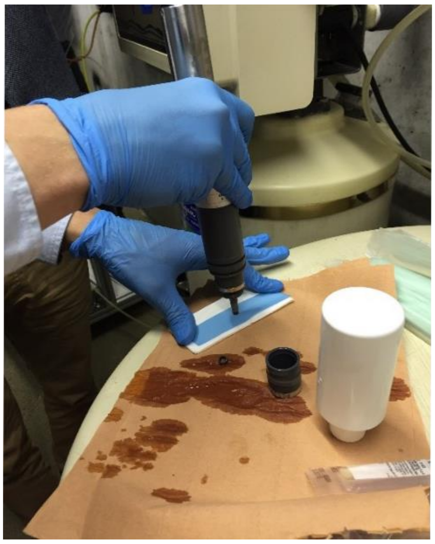


Figure 2.21. Cleaning of Chlori::lyser probe.

The EventLab probe requires regular maintenance. The water quality condition and the nature of installation determine the frequency of maintenance. Its maintenance can be divided in two parts:

- Cleaning of sensors.
- Need for filter.

The value of signal strength, monitored continuously, indicates the need of maintenance. If the signal health parameter drops below 0.15, the probe needs to be cleaned. If, after many cleaning, the signal level parameter remains below 0.15, the probe needs to be replaced.

The cleaning of EventLab sensor (sensing window) requires the use of many materials: cleaning agent, Polyurethane swab and drinking water. Firstly, we open the top cap of the flow cell to remove the EventLab sensor. Then, cleaning agent is used to wet the measurement surface. The next step consists of rubbing the surface using Polyurethane swab. Finally, it should be rinsed using drinking water. After cleaning, we verify that the signal health is 0.5 or higher. Figure 2.22 illustrates these different steps of cleaning [108].

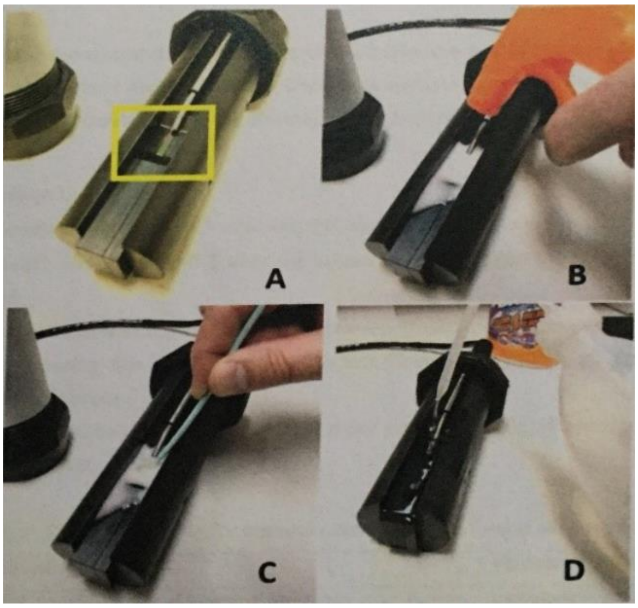


Figure 2.22. Cleaning of EventLab sensor [108].

Depending on the local water quality, the signal strength can be deteriorate quickly within few days. In such case, the use of filter is required [109]. To prevent the accumulation and the clopping up of the fluidic system, a filter unit can be used to capture large particles. The pore size of the default installed filter cartridge is 1µm, ensuring dissolved components are not removed from the sample stream [108].

2.6.4 Example of S::CAN measurement

An example of S::CAN measurement is displayed in Figure 2.23. It illustrates the variation of parameters: UV, Turbidity EPA, Turbidity ISO, and TOC in function of time, at Barroi during April 2017. The signals are quasi-stable and below the Standards limits.

Figure 2.23. Example of S::CAN measurement at Barroi.

2.6.5 Example of EventLab measurement

Figure 2.24 illustrates an example of EventLab measurement on May 15, 2017 at Polytech'Lille. It indicates the variation of two parameters: F24 Response and the signal level in function of time. During this period, F24 Response is quasi-constant which indicates normal quality of water. Also, the signal level is above the limit value of 0.15. Others parameters can be monitored and visualized each minute.

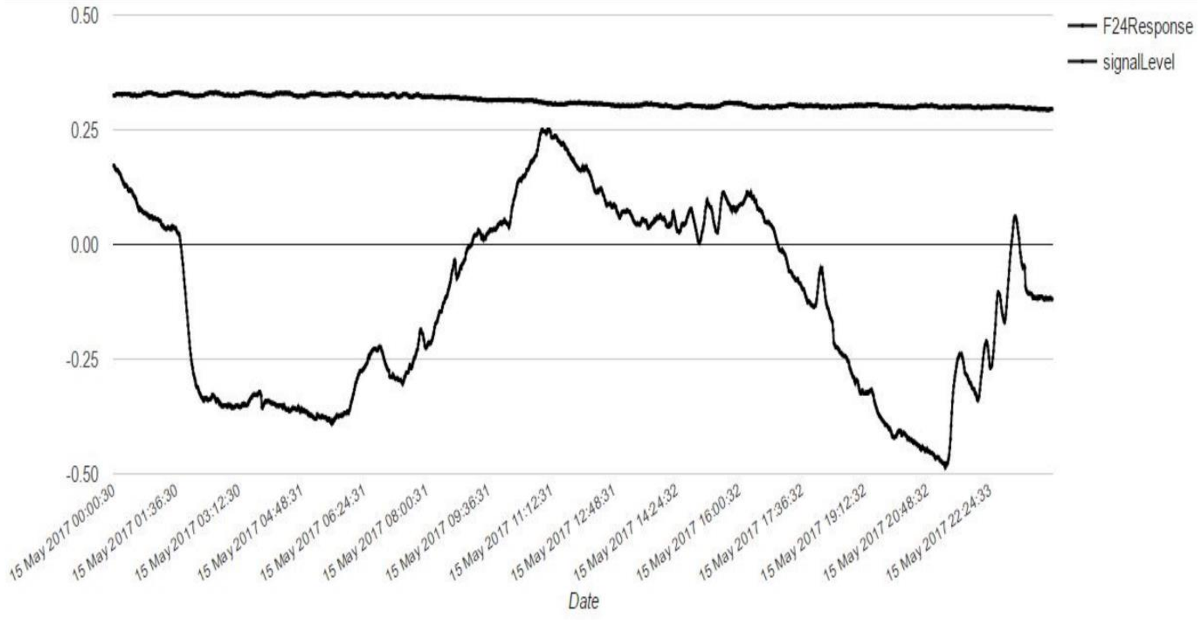


Figure 2.24. Example of EventLab measurement at Polytech.

2.7 Conclusion

This chapter presented the demonstration site used in this research for the real-time control of the water quality. The drinking water network of the Scientific Campus of the University of Lille is used as a support for the demonstration activity. This network was already monitored by sensors for water leakage (AMR, pressure sensors).

For the water quality control, S::CAN and EventLab were installed at two locations in the campus. We provided a detailed description of these devices: components, measured parameters, principle of measurements, calibration procedure, installations, etc.

The chapter presented the integrated solution used for the water quality control: data acquisition, data transmission, data storage, processing and visualization. The chapter presented also the maintenance of the installed devices. Examples of recorded results with S::CAN and EventLab were presented. They show that the monitoring system operates well.

The analysis of sensor's data for early detection of abnormalities is presented in the following chapters. Different event detection methods will be applied to these data to prevent rapidly any contamination.

Chapter 3. Analysis of Water Quality Signals in Lille Demo Site

3.1 Introduction

This chapter presents a detailed analysis of water quality signals at the Scientific Campus of the University of Lille. It is divided in two main parts: i) S::CAN signals and ii) EventLab data.

In the first part, the methodology used for signals analysis will be described. S::CAN data will be firstly compared with laboratory analyses results to validate the measurements. The relationship between the different S::CAN parameters will be evaluated using the correlation matrix. Different critical periods are chosen to describe the variation of S::CAN data at both locations (Polytech'Lille and Barroi). The continuous analysis allows the identification of some abnormalities that affect the water quality signals. The detected deviations should be interpreted to define the reason of abnormality and the potential occurrence of event in water. Data measured are compared between the two installation sites to verify the importance of sensor placement.

The second part will describe data monitored by EventLab sensor. The variation of phase as well as F24 Response will be analyzed during different periods. Signal health and signal level are controlled to test the probe performance. An example of an event recorded by EventLab will be presented.

A comparison between the measurements of S::CAN and EventLab will allow to explore their reliability as well as their limitation.

3.2 Analysis of S::CAN records

3.2.1 Methodology of analysis

Several research have proved the feasibility of using sensor data for the early detection of water contamination. A correlative response between the change in water quality parameters and the presence of contaminant in water has been verified in various studies [110].
S::CAN sensor measures, each minute, multiple water quality parameters. Each parameter is represented by a signal. For a safe drinking water, signals should be below the Standards thresholds.

Figure 3.1 illustrates an example of S::CAN measurements at Barroi. It indicates different water signals below the acceptable limits; each stable line refers to an indicator parameter. Therefore, any intrusion in the water system could induce deviation from the baseline. An abnormality in water quality is generally illustrated by an increase in S::CAN parameters (UV, Turbidity, etc.) and a decrease in Chlorine concentration. Figure 3.2 shows an example of data collected at Polytech, where peaks in UV, Turbidity and TOC are coupled with a decrease in the Chlorine level.

However, it is important to distinguish normal variation from that due to water quality degradation. Many factors could cause perturbations in water signals: contaminations, connections issues, etc. The analysis of deviation should identify the type of unexpected values according to three classes: outlier, event, baseline change (detailed in Chapter 2).

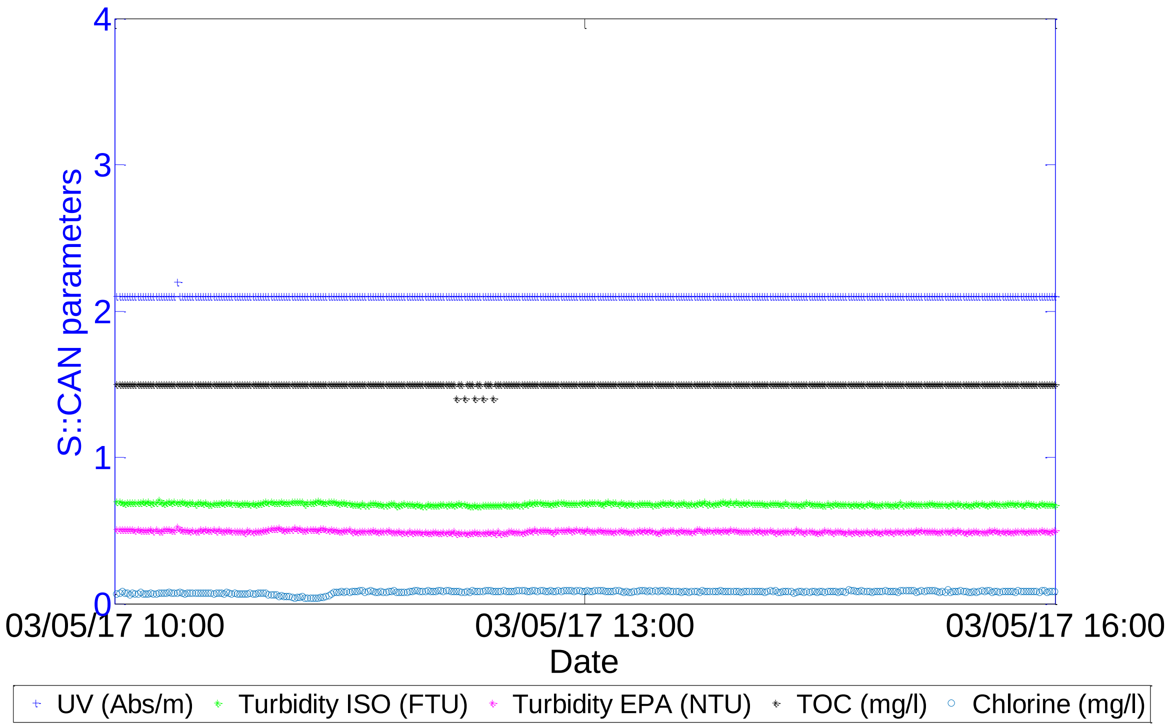


Figure 3.1. Example of stable S::CAN signals.

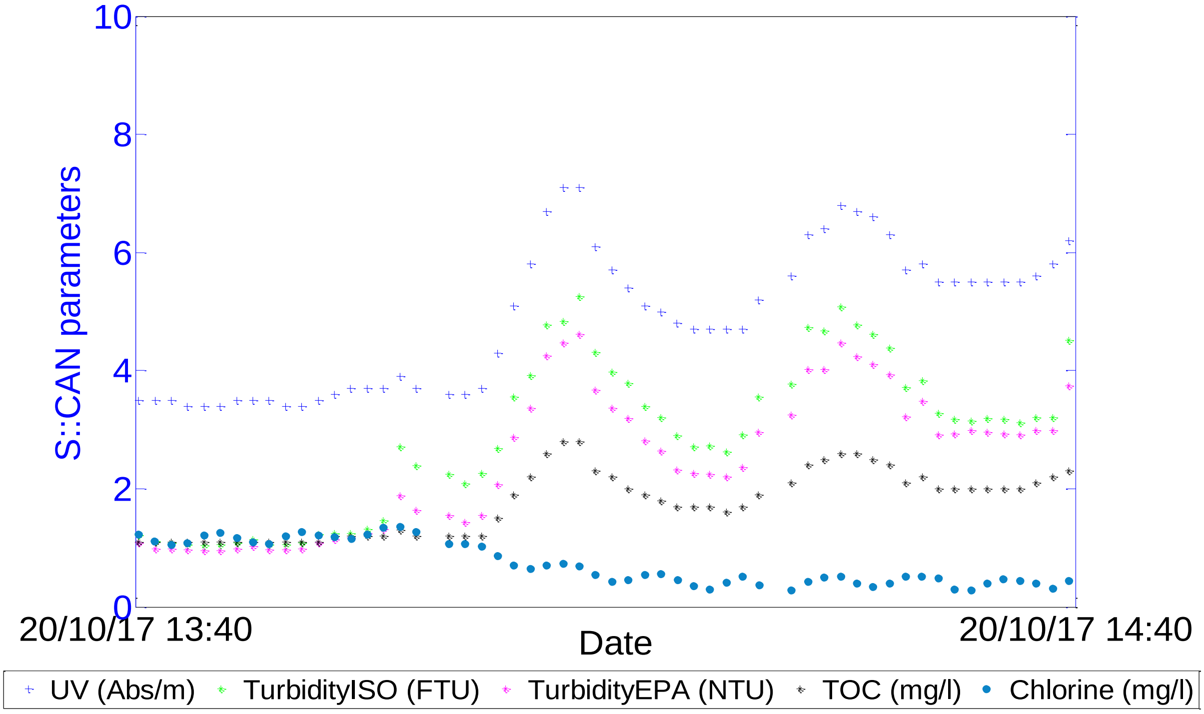


Figure 3.2. Example of deviations of S::CAN signals.

The water quality system could suffer from faults due to the sensor or to data transmission problems. When data is missed from a time step, an “Error” status is indicated by S::CAN. The absence of measurement will generate an alarm to solve the problem of data loss. A case of data missing is illustrated in Figure 3.3. It indicates a measurement interruption on May 29, 2017 between 4:17 and 4:59 pm.

Figure 3.3. Loss of S::CAN data at Polytech'Lille on May 29, 2017.

3.2.2 Data Validation: Laboratory Analyses and Calibration

In order to validate the measurements of S::CAN sensor, laboratory analyses were conducted on water samples collected from the two locations: Polytech'Lille and Barroi. Samples are taken on February 7, 2017 and on February 27, 2017. Laboratory tests included:

- Microbiological analysis concerning microorganisms such as Escherichia Coli and Intestinal Enterococci. The presence of microorganisms is directly correlated with a low amount of Free Chlorine induced generally by biofilm growth.
- Physico-chemical analyses concerning TOC, DOC, pH, Turbidity, absorbance UV 254, and Conductivity.
- Total and Free Chlorine
- Anions tests concerning Nitrates and Sulfates.
- Cations tests concerning Sodium and Ammonium.
- Metals tests concerning Aluminum, Iron, Arsenic, Copper and Manganese.

Table 3.1 gives the results of laboratory tests at both locations. The comparison with Standards values indicates satisfactory microbiological and physico-chemical results [111]. Results obtained from laboratory analyses were compared to S::CAN data.

Table 3.1. Laboratory analyses results at Polytech'Lille and Barroi [111].

TYPE OF TEST	Laboratory Results				Standards
	07/02/2017		27/02/2017		
OBSERVATION-IN-SITU	Polytech	Barroi	Polytech	Barroi	
Water temperature (°C)	11,9	11,2	13,1	11,4	29
MICROBIOLOGICAL					
Intestinal enterococci (UFC/100ml)	<1	<1	<1	<1	0
Coliform bacteria (UFC/100ml)	<1	<1	<1	<1	0
Escherichia Coli (UFC/100ml)	<1	<1	<1	<1	0
Germs revivable at 36 ° C, 44h (without dilution) (UFC/ml)	<1	<1	<1	2	
Germs revivable at 22 °C, 68h	<1	<1	<1	<1	

(without dilution) (UFC/ml)					
Spores of anaerobic bacteria sulfite-reducing (UFC/ ml)	<1	<1	<1	<1	0
RESIDUAL DISINFECTANTS					
Total Chlorine (in situ) (mg/l)	0,26	0,21	0,27	0,28	
Free Chlorine (in situ) (mg/l)	0,22	0,13	0,16	0,117	
ORGANOLEPTIC TESTS					
Appearance / Color	limpid	limpid	limpid	limpid	
Color (visual examination)	absence	absence	absence	absence	
Odor / flavor at 25 ° C	1	1	1	1	3
PHYSICO-CHEMICAL					
Total Organic Carbon (TOC) (mg/l)	1,5	1,5	1,3	1,4	2
Dissolved Organic Carbon (DOC) (mg/l)	1,5	1,5	1,3	1,3	
pH	7,5	7,5	7,4	7,5	6.5-9
Temperature of pH measurement (° C)	17,3	17	16,7	16,4	
Turbidity (NFU)	0,9	0,61	0,33	0,45	2
UV (at 254 mm) (u.abs)	<0.001	<0.001	<0.001	<0.001	
Conductivity (at 25 ° C) (µS/cm)	857	852	883	885	200-1100
Anion					
Nitrates (mg/l)	15,8	15,9	11,8	12,2	50
Sulfates (mg/l)	150	144	130	132	250
Cation					
Sodium (mg/l)	44,04	43,91	42,47	42,81	200
Ammonium (mg/l)	<0.05	<0.05	0,08	0,06	0,1
METALS					
Aluminum (µg/l)	<5	8	10	10	200
Iron (µg/l)	57	79	98	166	1mg/l
Arsenic (µg/l)	<1	<1	<1	<1	10
Copper (µg/l)	1,9	1,7	2,5	2,4	1mg/l
Manganese (µg/l)	1,64	2,87	2,67	3,1	50

Figure 3.4 shows results of comparison between S::CAN records and laboratory analyses. Since S::CAN is controlled to give data each minute, an average value (during one hour) is used to make the comparison with laboratory data. The relative error is calculated as follows:

$$Error\ (\%) = \frac{|Sensor\ data - Laboratory\ data|}{Laboratory\ data}\ x\ 100 \qquad (3.1)$$

Figure 3.4 indicates:

- A good agreement between S::CAN data and laboratory results for some parameters such as pH (error less than 3 %).
- A small error (less than 10 %) for Conductivity and Temperature
- Significant error for parameters such as Chlorine and TOC (error between 20 and 40%).

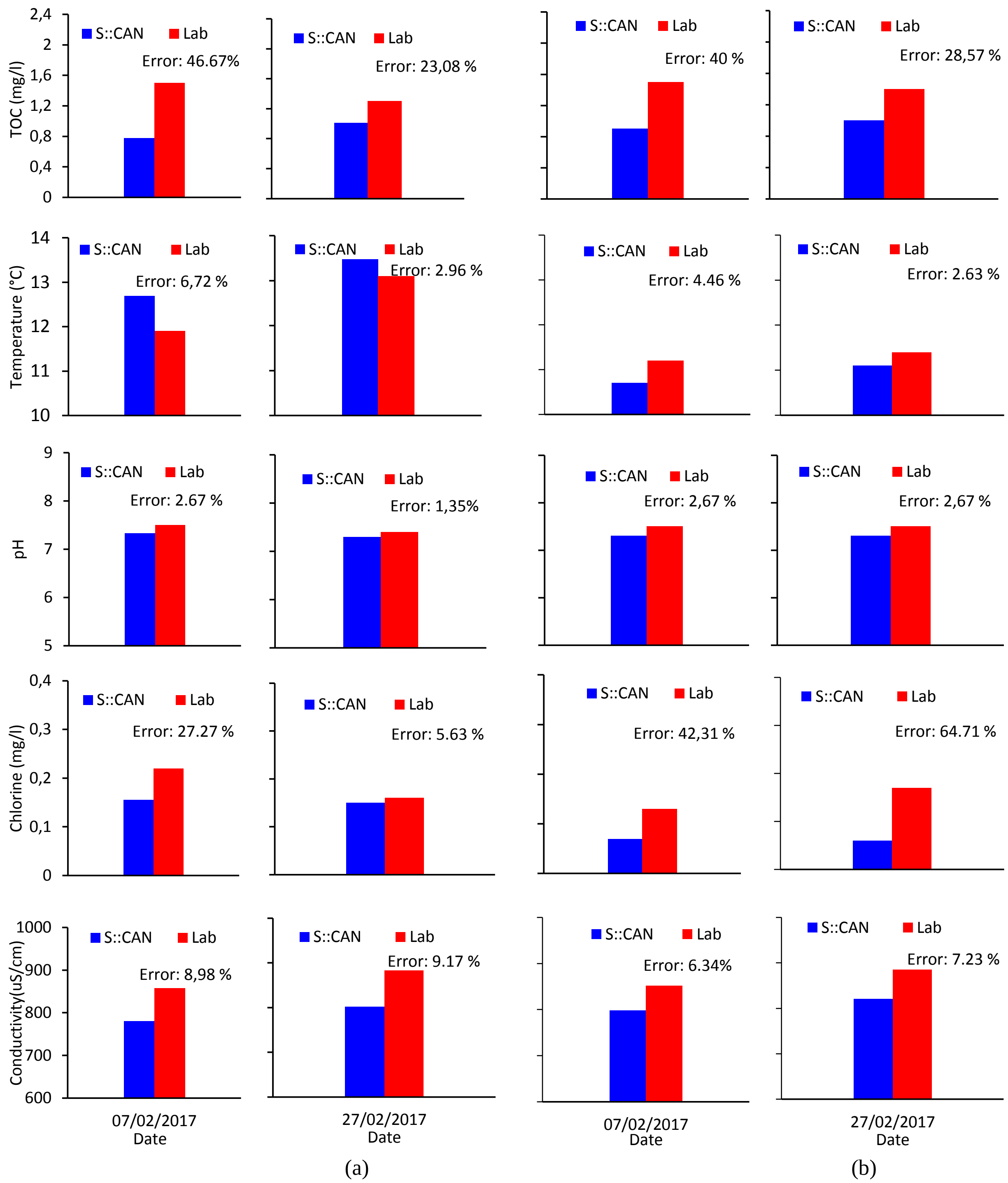


Figure 3.4. Laboratory comparison. (a) Polytech'Lille; (b) Barroi.

For Turbidity, the difference between laboratory results and S::CAN measurement is major. Table 3.2 gives the Turbidity recorded by S::CAN in different units (ISO and EPA) as well as the laboratory value. Since laboratory results are significantly lower than those measured by S::CAN, calibration for Turbidity parameter is needed.

Table 3.2. Comparison of Turbidity values between S::CAN and Laboratory analyses.

	TurbidityISO_S::CAN (FTU)	TurbidityEPA_S::CAN (NTU)	TurbidityISO_Lab (NFU)
Polytech_Jan 7	2,10	1,10	0,90
Polytech_Jan 27	2,40	1,95	0,33
Barroi_Jan 7	0,15	0,15	0,61
Barroi_Jan 27	0,23	0,24	0,45

Calibration of S::CAN was conducted as follows:

- For Conductivity and Free Chlorine, we calculate the correction factor: quotient between laboratory and S::CAN data. Table 3.3 (a) indicates an average factor of 1.2 for Chlorine at Polytech'Lille.
- For i::scan parameters (Turbidity, TOC, and DOC), we calculate the corresponding offset: difference between laboratory and S::CAN data. Table 3.3 (b) shows an average offset of 0.5 for TOC at Barroi.

Table 3.3. Adjustments calculation. (a) Chlorine at Polytech'Lille; (b) TOC at Barroi.

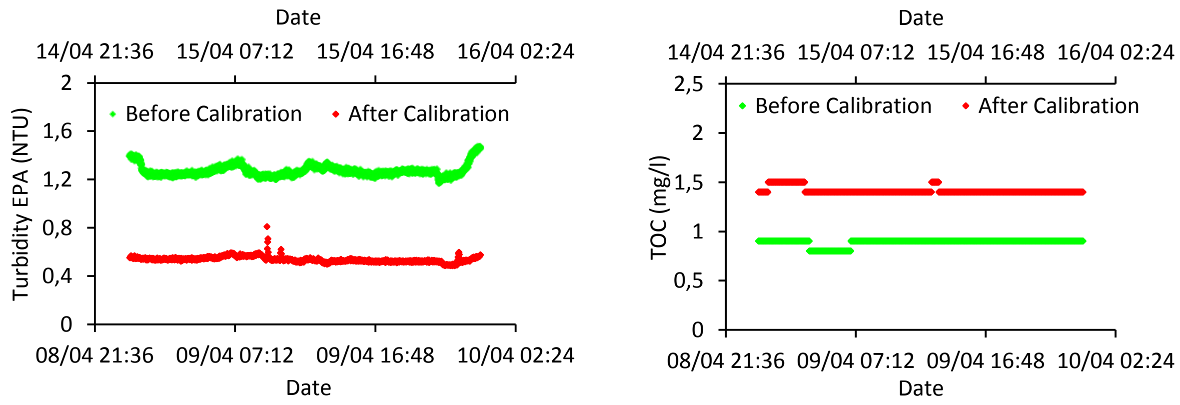
(a)

Chlorine_Polytech (mg/l)			
Date	**S::CAN**	**Laboratory**	**factor**
07/02/2017	0,16	0,22	1,4
27/02/2017	0,15	0,16	1,1
Average			**1,2**

(b)

TOC_Barroi (mg/l)			
Date	**S::CAN**	**Laboratory**	**offset**
07/02/2017	0,9	1,5	0,6
27/02/2017	1	1,4	0,4
Average			**0,5**

Based on the calculated offset and correction factor for each parameter, the calibration was conducted at both locations on April 10, 2017. After calibration, S::CAN data became closer to laboratory results. An example is illustrated in Figure 3.5. The Turbidity EPA, at Polytech'Lille (Figure 3.5 (a)), is reduced from an average about 1.4 NTU to 0.6 NTU to fit with laboratory tests. However, TOC value increased from an average of 0.9 mg/l to 1.5 mg/l, at Barroi (Figure 3.5 (b)).



(a) (b)

Figure 3.5. Comparison of signals before and after calibration. (a) Polytech'Lille; (b) Barroi.

3.2.3 Correlation between measured parameters

In order to find the relationship between measured parameters of S::CAN, the Pearson coefficient ρ is calculated. It measures the strength of linear relationship between paired values and can vary between -1 and 1. Positive coefficient indicates that if one variable increases the other has a tendency to increase and inversely. Negative correlation denotes that the two data vary in opposite direction. A coefficient ρ close to 0 indicates that there is no linear correlation. The closer the value ρ is to 1 or -1, the stronger the linear correlation [112]. Table 3.4 gives the correlation coefficients for S::CAN data at Polytech'Lille during January 2017 (data variation given in Figure 3.9). It indicates a Pearson coefficient higher than 0.8 between UV, Turbidity (ISO and EPA), TOC and Color. For these parameters, ρ is close to 0 with Temperature, Conductivity, pH and Chlorine. Negative coefficient is observed between Temperature and Conductivity (-0.15).

Table 3.4. Correlation coefficients for S::CAN parameters at Polytech'Lille.

Polytech Parameters	UV (Abs/m)	Turbidity ISO (FTU)	Turbidity EPA (NTU)	TOC (mg/l)	Color (Hazen)	Tempe-rature 1 (° C)	Conductivity (µS/cm)	Tempe-rature 2 (° C)	pH	Free Chlorine (mg/l)
UV (Abs/m)	1	0,87	0,88	0,95	0,81	-0,01	0,06	0,00	0,03	0,06
Turbidity ISO (FTU)	0,87	1,00	0,99	0,86	0,93	-0,07	0,04	-0,07	0,06	0,08
Turbidity EPA (NTU)	0,88	0,99	1,00	0,87	0,93	-0,06	0,04	-0,06	0,09	0,09
TOC (mg/l)	0,95	0,86	0,87	1,00	0,80	-0,05	0,07	-0,03	0,03	0,05
Color (Hazen)	0,81	0,93	0,93	0,80	1,00	-0,06	-0,02	-0,08	0,06	0,05
Temperature 1 (° C)	-0,01	-0,07	-0,06	-0,05	-0,06	1,00	-0,15	0,99	0,08	0,24
Conductivity (µS/cm)	0,06	0,04	0,04	0,07	-0,02	-0,15	1,00	-0,14	-0,29	0,20
Temperature 2 (° C)	0,00	-0,07	-0,06	-0,03	-0,08	0,99	-0,14	1,00	0,05	0,21
pH	0,03	0,06	0,09	0,03	0,06	0,08	-0,29	0,05	1,00	0,00
Free Chlorine (mg/l)	0,06	0,08	0,09	0,05	0,05	0,24	0,20	0,21	0,00	1,00

To analyze these results, the correlation matrix is given in function of regression coefficient R^2 (or coefficient of Determination). R^2 values are between 0 and 1. It indicates how close the data are to a fitted regression line. Figure 3.6 illustrates the correlation matrix of S::CAN parameters. It gives the correlation equation and coefficient for paired data. It shows that:

- For UV, $R^2 \geq 0.66$ with Turbidity (ISO and EPA), TOC and Color while R^2 tends to 0 with Temperature, Conductivity, pH and Chlorine.
- For Turbidity (ISO and EPA), $R^2 \geq 0.74$ with UV, TOC and Color. As for UV, R^2 is close to 0 with the remaining parameters.
- For TOC and Color, $R^2 \geq 0.63$ with UV, Turbidity (ISO and EPA) and Color.
- For Temperature 1 and 2, R^2 is close to 1, while R^2 between Temperature and all other parameters is approximately null.
- For Conductivity, pH and Free Chlorine, R^2 is very close to 0 with all parameters.

Analysis of the correlation matrix allows to divide S::CAN parameters into three main categories:

- Category 1: including UV, Turbidity (ISO and EPA), TOC and Color parameters. The regression factor between these parameters is higher than 0.66. In this category, a strong correlation is observed between UV and TOC ($R^2 = 0.9$) and between Turbidity ISO and Turbidity EPA ($R^2 \approx 1$).

- Category 2: for Temperature 1 and Temperature 2 ($R^2 \approx 1$). The only linear relation for this category is between the two measurements of Temperature (by i::scan and Condu::lyser).
- Category 3: including Conductivity, pH and Free Chlorine parameters. There is no linear relationship between each of this parameter and others S::CAN data.

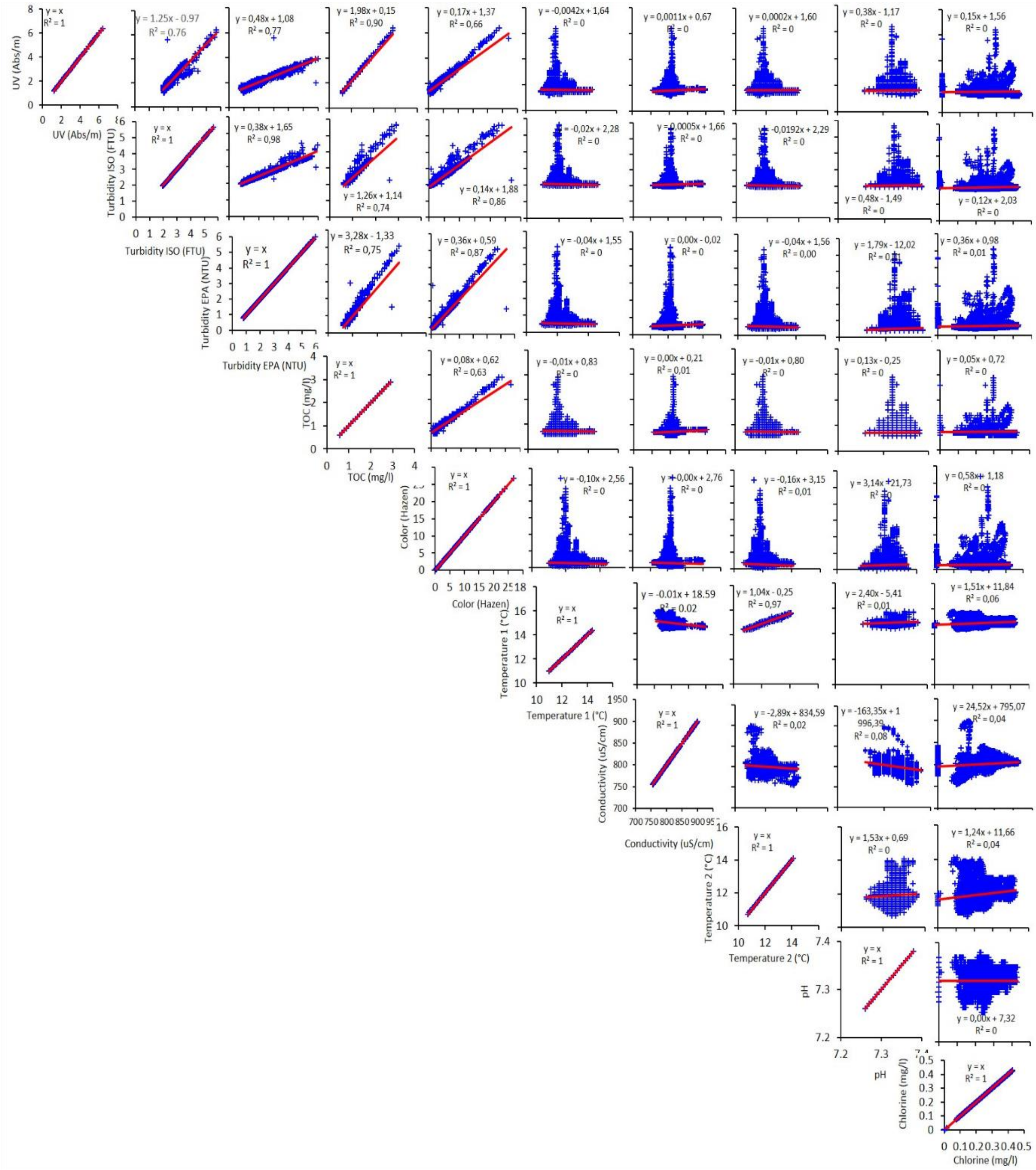


Figure 3.6. Correlation matrix for S::CAN parameters at Polytech'Lille.

3.2.4 Data Analysis

3.2.4.1 Polytech and Barroi- December 2016 till October 2017

Figure 3.7 shows the variation of different S ::CAN data at Polytech'Lille, for the period between December 2016 and October 2017. During this period, the majority of signals did not exceed the threshold limits. However, different deviations from the stable lines are observed. For i::scan parameters (UV, Turbidity and TOC), all instantaneous exceeding of 5 *units* is considered as outlier, due to measurement errors. It should be mentioned that the term "*units*" is related to the corresponding unit of measurement for each parameter (Abs/m, NTU, mg/l, etc.). More significant events, that last for several hours, have been detected during this period (for example in September 2017). Parameters, such as pH, Temperature and Conductivity, show less deviations than UV, Turbidity and TOC. For Chlorine, different problems in measurement have occured during the monitoring period.

Figure 3.7 indicates also a change of the baseline for many parameters due to calibration adjustments on April, 10. Signals become more stable between April and October, 2017. A loss of data is observed between March and April, 2017.

The variation of S::CAN parameters, between December 2016 and October 2017, at Barroi, is given in Figure 3.8. Although some deviations were recorded, signals are more stable than those measured at Polytech.

An adjustment of measured values is observed after the calibration done on April, 10, especially in Turbidity, TOC and Conductivity signals. An augmentation in Temperature is noticed after May 2017 but it remains always less than 20 ° C. As for Polytech, pH is constant during the monitoring period. The variation of Chlorine is quasi-stable (between 0 and 0.1 mg/l).

The list of events, their date of occurrence, their characteristics as well as the analysis of the reason of abnormalities detected, during this period, will be detailed below.

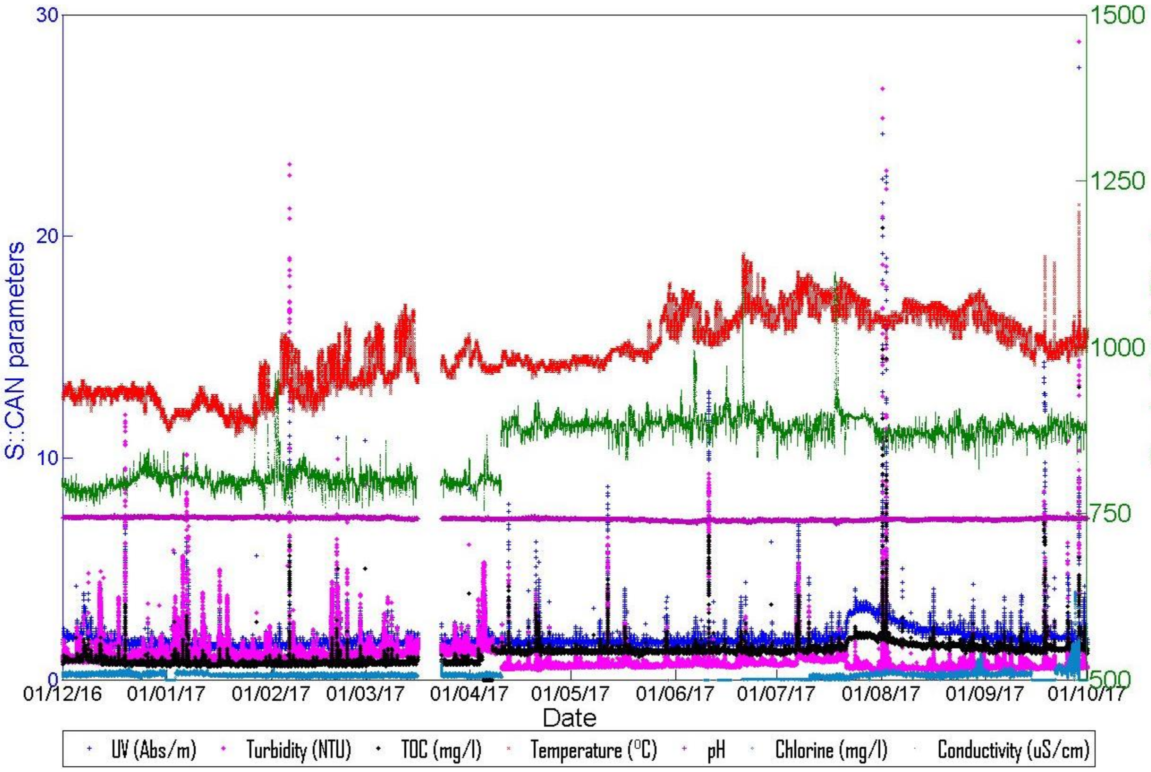


Figure 3.7. S::CAN data at Polytech'Lille.

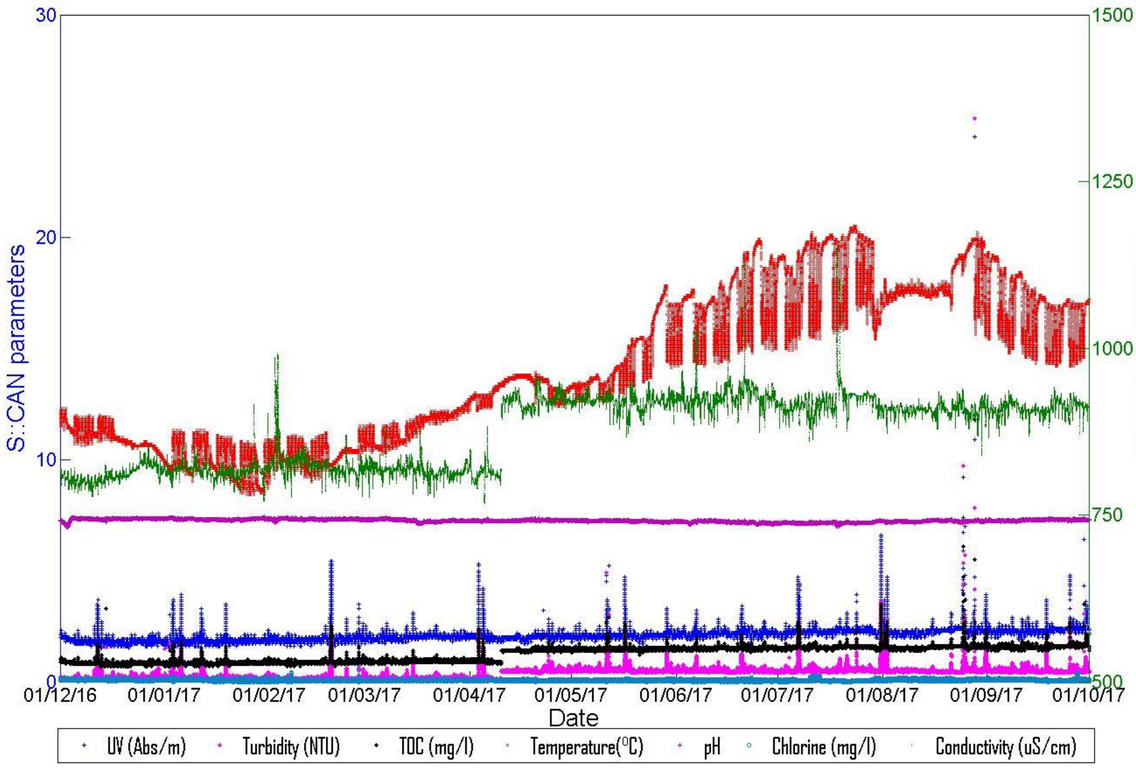


Figure 3.8. S::CAN data at Barroi restaurant.

3.2.4.2 Polytech-January 2017

An example of data collected by S::CAN at Polytech during January 2017 is given in Figure 3.9. Data missing is observed on January, 22. Connection problems induce this loss of data. The figure shows that:

- The TOC profile remains below the maximum limit of 2 mg/l, with an average of 0.7 mg/l. During the nights and weekends, the signal can be considered as totally stable. However, some perturbations have been observed at the morning when the water flow increases. The TOC profile is strongly correlated with UV variation (1.6 Abs/m as average value of UV) with 0.95 as correlation coefficient.
- The Turbidity profile is less constant in comparison with other signals. Different deviations (outliers/events) were detected, especially before January, 22. However, the average Turbidity remains acceptable, around 1 NTU.
- The average Temperature is 12.2 °C, with a standard deviation of about 0.5 °C. All values of temperature are below the limit of 15 ° C.
- The pH signal is quasi-constant (7.3) between (6.5) and (8.5).
- The Chlorine signal is very close to the limit of 0.2 mg/l. The variation of Chlorine concentration depends on the consumption value but it can be considered as acceptable.
- The Conductivity signal is around 800 µS/cm with a standard deviation of 10 µS/cm. It remains between the accepted limits of 200 and 1100 µS/cm.

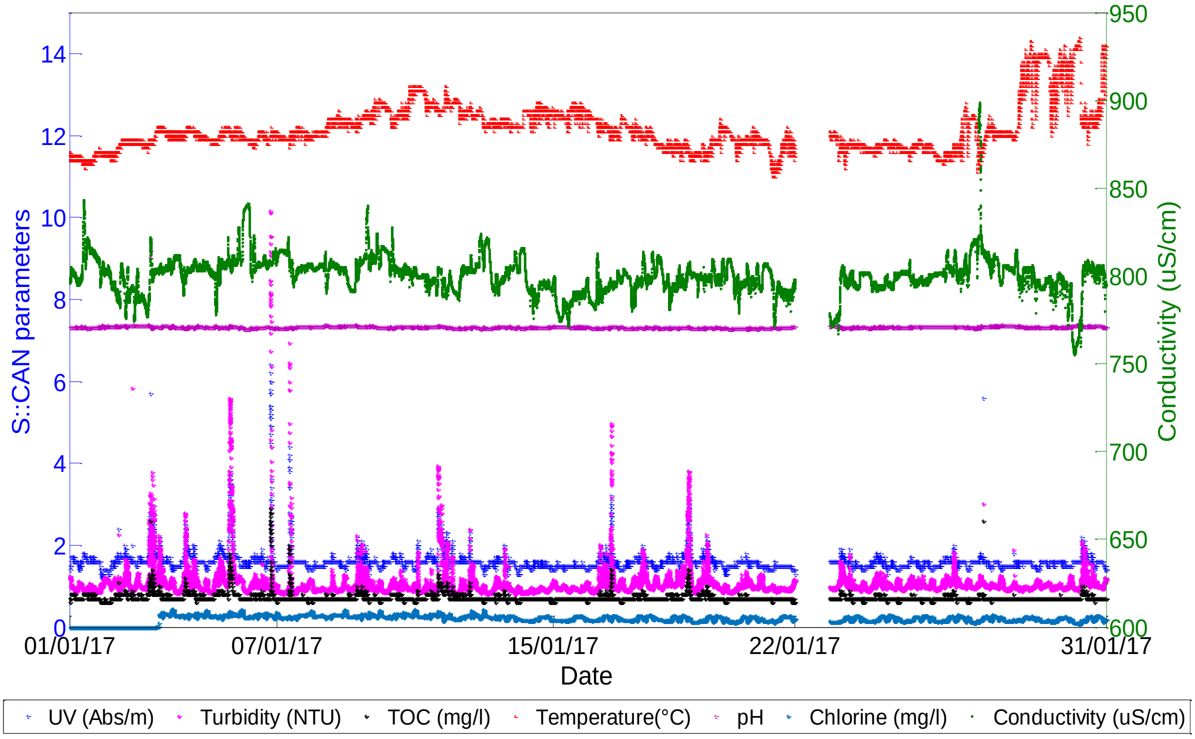


Figure 3.9. S::CAN data during January 2017 at Polytech'Lille.

3.2.4.3 Barroi-December 2016.

Figure 3.10 illustrates the variation of S::CAN parameters during December 2016 at Barroi restaurant. The figure also reports the threshold recommended by the *WHO*. It indicates that:

- The UV profile is constant with an average of 1.8 Abs/m.
- The Turbidity signals are relatively stable and below the threshold of 1 NFU. The measurements in two units ISO and EPA are similar with an average of 0.1 NFU.
- The amount of organic substances is illustrated by the measurements of TOC and DOC (average 0.8 mg/l). The two corresponding signals are similar and approximately equal. The variation of both TOC and DOC are quasi-stable, with a standard deviation around 0.070 mg/l. Although the signals of UV, TOC, DOC and Turbidity are quasi-stable, a significant deviation is observed on December, 12. This event is illustrated by a peak of parameters from the stable lines.
- The temperature is around 12 ° C. It has few regular variation till December, 17, and becomes more stable between December 17, 2016 and December 31, 2016 (during Christmas holidays).
- The pH signal is considered stable without any deviation. The average value is 7.3 between the Standards of 6.5 and 8.5.
- The amount of Chlorine is smaller than the limit of 0.2 mg/l. The average value is 0.1 mg/l.
- The reference line of Conductivity is relatively constant (average value 810 µS/cm). The measurement does not exceed the upper limit of 1100 µS/cm.

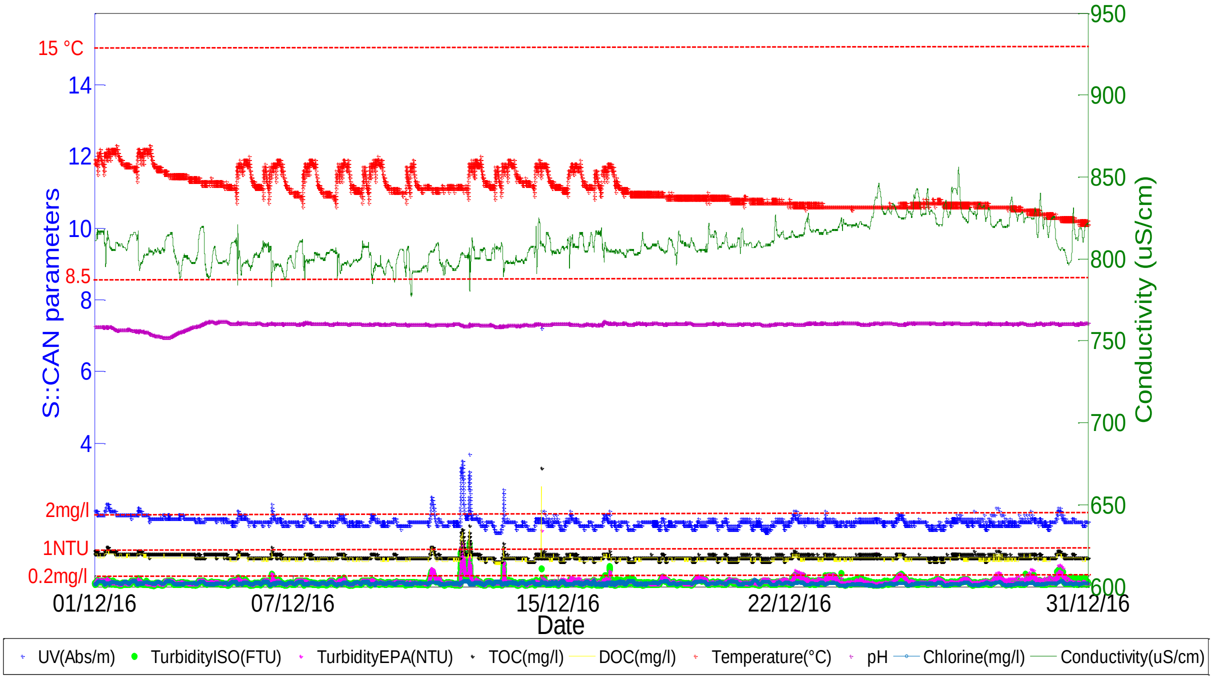


Figure 3.10. S::CAN data during December 2016 at Barroi.

3.2.4.4 Polytech and Barroi-June 2017.

The variation of multiple i::scan data during June 2017 at Polytech'Lille is given in Figure 3.11. During this period, an outlier is observed on June, 11. A very high deviation of signals is detected. However, this large variation is instantaneous; it can be due to some faults in data measurement or transmission. Since signals returned rapidly to quasi-stable values, this outlier should not generate an urgent alarm. A check up for the functionality of the sensor is required in this case.

In addition to the outlier, an event is detected between June 6 and June, 8. It is characterized by a significant increase in i::scan parameters. The amplitude of augmentation is around 1 *unit* and it lasts many hours. This event could indicate a potential variation in the water quality. Figure 3.11 shows also a clear difference in signals in the Week of June 12, 2017 (as indicated in the curly brackets). The variation of S::CAN data (UV, Turbidity and TOC), during this week, is illustrated in Figure 3.12. The shape of graphs differs between workdays and weekend. Signals are more stable during weekend and more disturbed in workdays (especially at 8 am). This difference could be attributed to the low consumption during the weekend. From Monday till Friday, the water is more consumed by students and employees, which implies more disturbance in water quality; when the flow increases, suspended matters can be released from aging water pipes. This intrusion appears in the increase of indicator parameters (UV, Turbidity and TOC). This phenomenon is not observed during weekend because the University is generally closed, and the consumption is reduced.

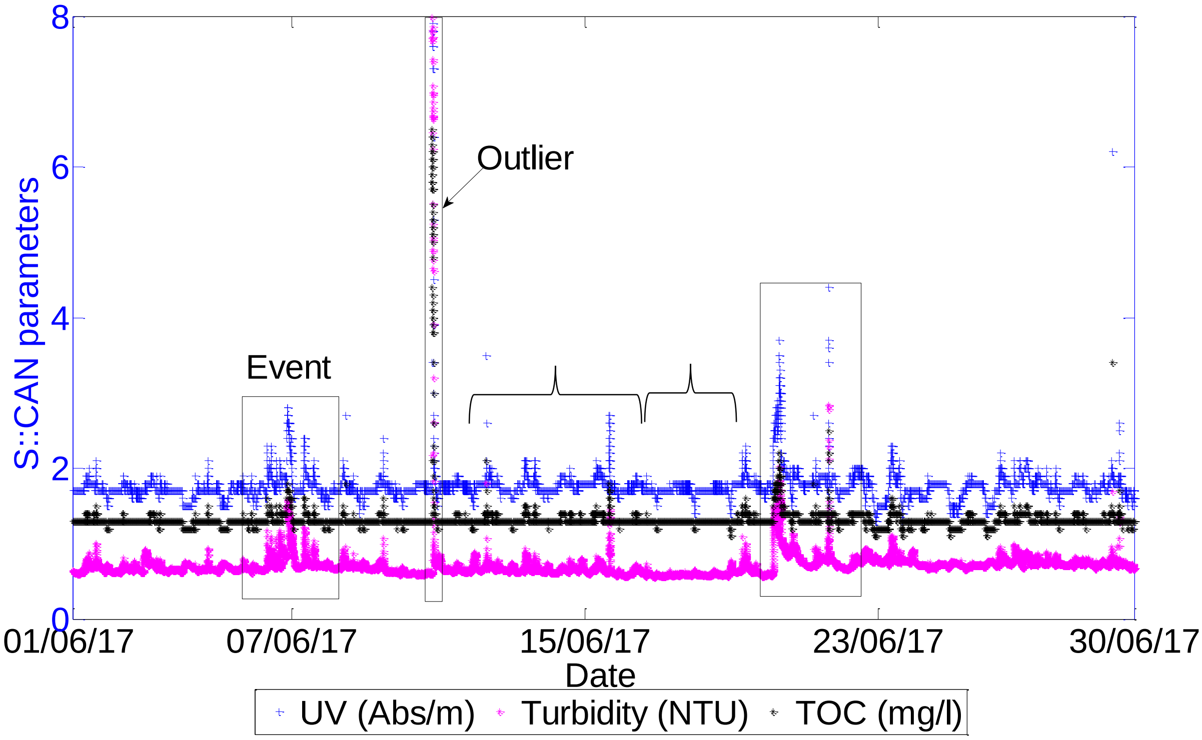


Figure 3.11. S::CAN data during June 2017 at Polytech'Lille.

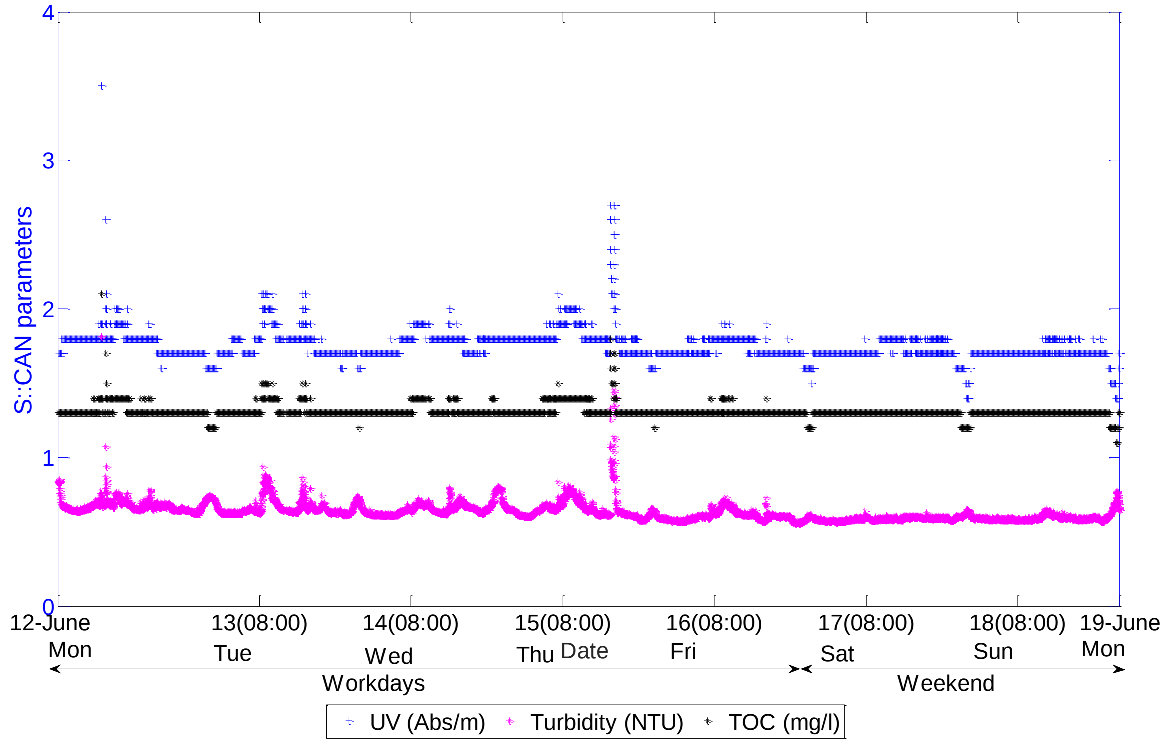


Figure 3.12. S::CAN data on the week of June 12, 2017 at Polytech'Lille.

The signals of i::scan at Barroi are given in Figure 3.13. Some events have been detected, such as the abnormality observed on June, 7. However, the duration and the amplitude of this event are more important at Polytech than at Barroi.

The variation of signals, between June, 12 and June, 19, is illustrated in Figure 3.14. Although the difference is not obviously clear as for Polytech, signals at Barroi remain more stable during the weekend in comparison with workdays. From Monday till Friday, parameters profile show low perturbations. These profiles become generally constant during weekend when the restaurant is closed.

The comparison of the two locations indicates that signals at Barroi are more stable than those at Polytech, even during workdays. This can be explained by the fact that the consumption of water in the University is usually greater than that at the restaurant. This is illustrated in water quality signals of S::CAN sensor.

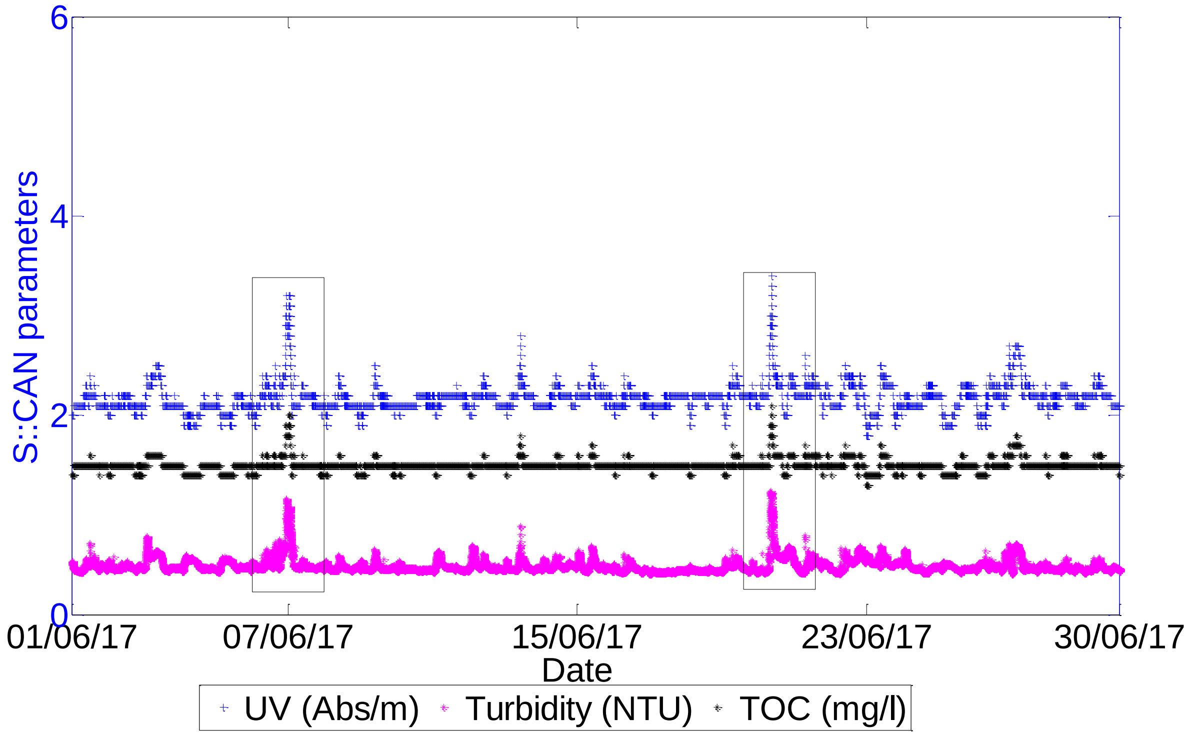


Figure 3.13. S::CAN data during June 2017 at Barroi.

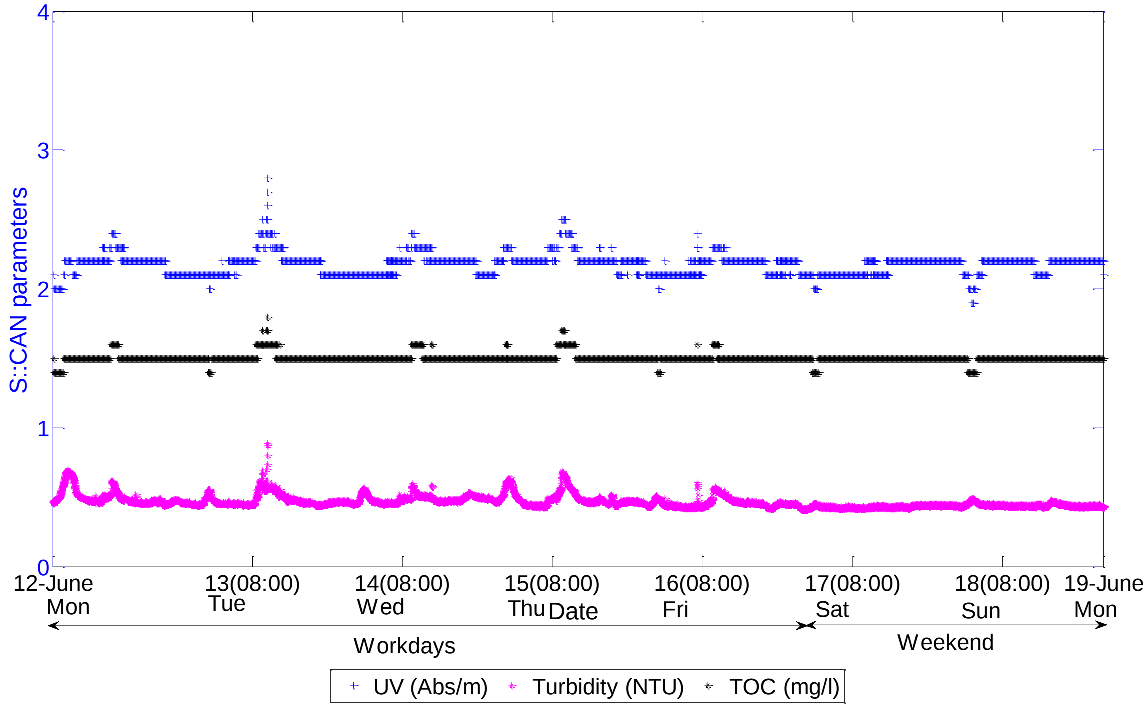


Figure 3.14. S::CAN data on the week of June 12, 2017 at Barroi.

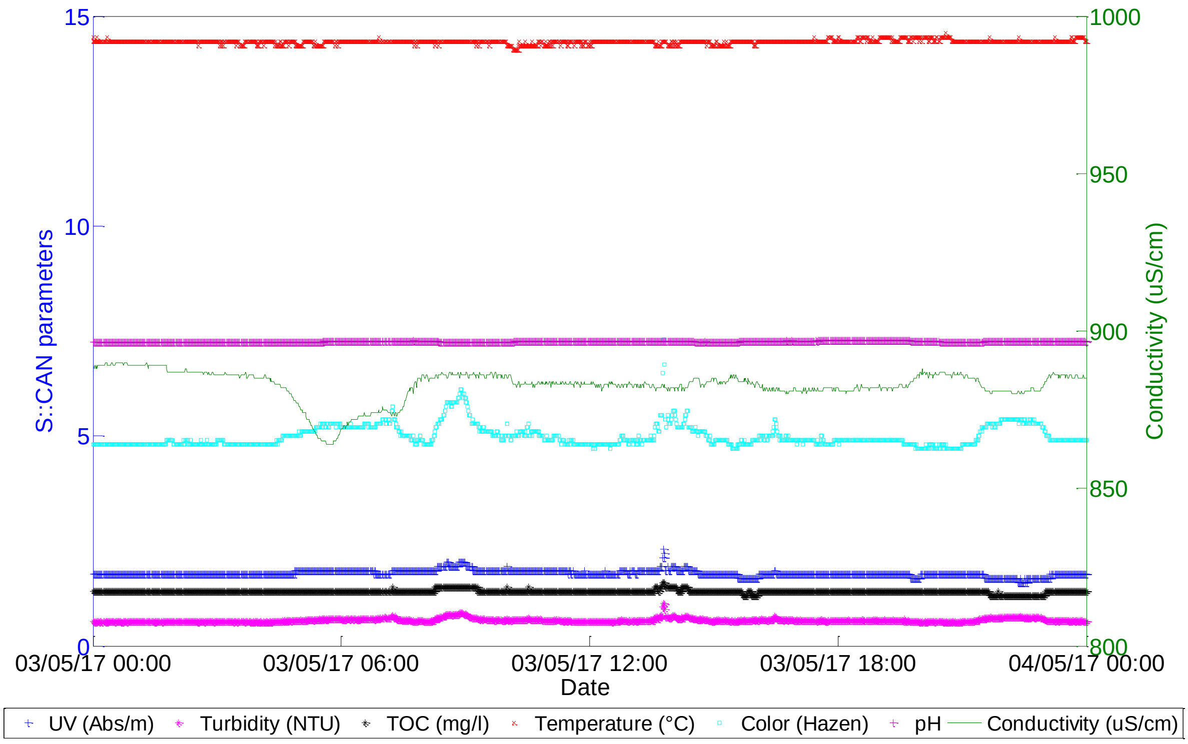
3.2.4.5 Polytech and Barroi-May 3, 2017

An example of the daily variation of S::CAN parameters is given on May 3, 2017. Figure 3.15 (a) shows the variation of S::CAN parameters on the third of May at Polytech'Lille. During this day, no major event has been detected. Most of signals are quasi-constant with slight deviation from the baseline. Among the measured parameters: UV, Turbidity, TOC and Color remain lower than 5 *units*. pH and Temperature can be considered stable with value around 7.2 and 14 ° C, respectively. For Conductivity, values vary in the interval [850-900] µS/cm.

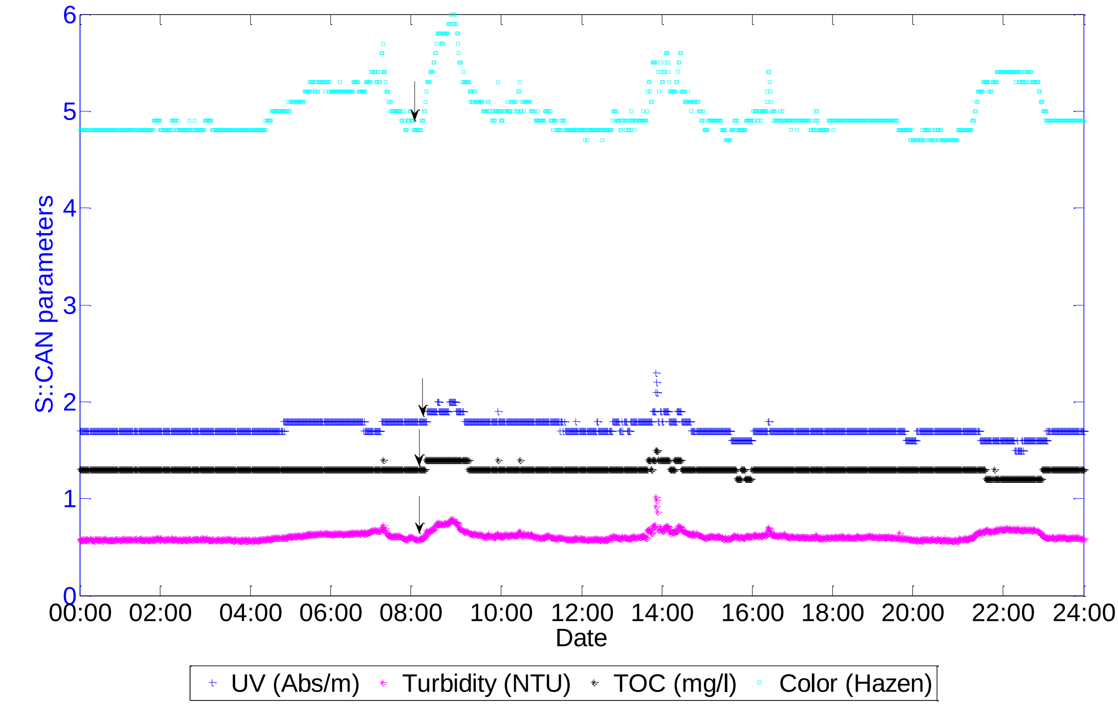
To take into account the magnitude of each parameter, analysis can be done for each group separately. Figure 3.15 (b) indicates the variation of several parameters. It shows that:

- At 8:00 am, an increase is observed in all signals during 1 hour. For UV, Turbidity and TOC, the average deviation is 0.2 *units*, while for Color, the variation is about 1 Hazen.
- At 2:00 pm, UV, Turbidity, TOC and Color deviate of 0.5 *units* from the stable lines for about 1 hour.
- At 9:30 pm, a very small increase (around 0.1 *units*) in UV and TOC signals is coupled with a decrease in Turbidity data. The augmentation in Color signal is more important (about 0.5 Hazen).

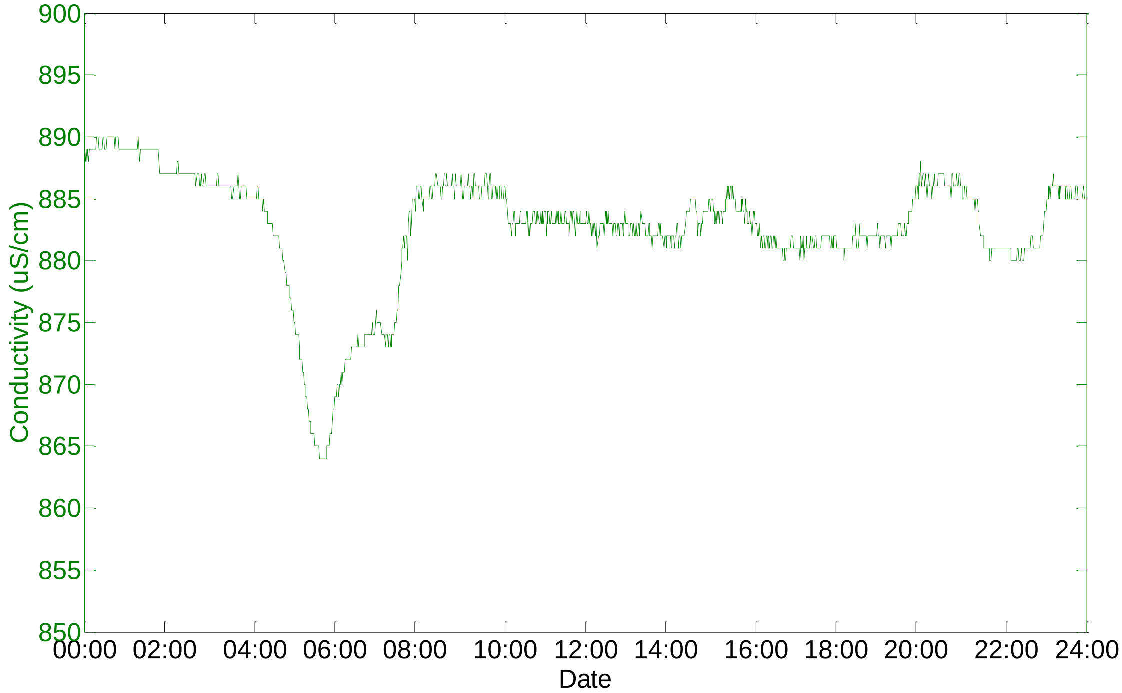
Figure 3.15 (c) illustrates the variation of Conductivity on May 3, 2017. A decrease is observed at 4:00 am with an amplitude of 20 µS/cm. This deviation takes about 4 hours. Another variation occurred at 9:30 pm for about 1hour 30 minutes, with a deviation of 5 µS/cm.



(a)



(b)



(c)

Figure 3.15. S::CAN data on May 3, 2017 at Polytech'Lille. (a) All measured parameters; (b) i::scan parameters; (c) Variation of Conductivity.

Figure 3.16 gives the signals of multiple water quality parameters on May 3, 2017 at Barroi. A very small deviation from the baseline is observed around 8:30 am. This increase is observed in UV, Turbidity and TOC signals. A decrease in Conductivity signals precedes this augmentation of i::scan parameters. The deviation in the morning, is much more important at

Polytech than at Barroi. Other signals, such as pH, Temperature and Chlorine are generally more stable.

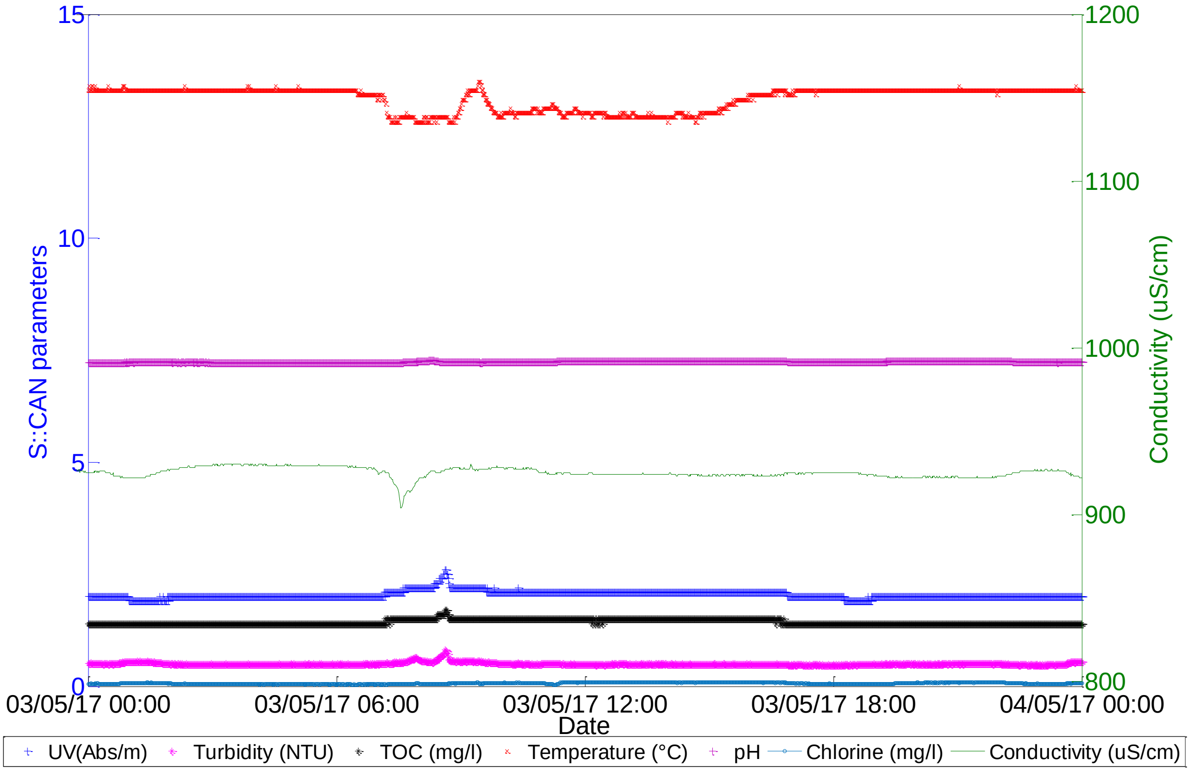


Figure 3.16. S::CAN data on May 3, 2017 at Barroi.

3.2.4.6 Statistical Analysis

A statistical analysis is conducted for different S::CAN parameters during three days in 2017: May, 3, April, 22 and June, 5 at Polytech'Lille. At each day, the average, standard deviation as well as the maximum and minimum and the normal variation ($\bar{X} \pm 3\sigma$) are calculated. The results are shown in Table 3.5. It indicates similar results during these three days. The average of most parameters is below the thresholds (detailed in Chapter 1). The Standard deviation is relatively small for most of parameters except for Conductivity which could have significant standard deviation (reaches 13 µS/cm on June, 5). The maximum and minimum values determine the upper and lower bounds for each parameter and remain generally between the limit of normal variation ($\bar{X} \pm 3\sigma$).

Table 3.5. Statistical Analysis for S::CAN data at Polytech'Lille.

	UV (Abs/m)			Turbidity ISO (FTU)			Turbidity EPA (NTU)		
	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017
Average	1,8	1,7	1,7	2,4	2,4	2,4	0,6	0,6	0,7
Standard deviation	0,094	0,08	0,100	0,072	0,042	0,026	0,071	0,044	0,031
Max	2,1	2,3	1.9	2,7	2,8	2,6	0,9	1	0,8
Min	1,6	1,5	1,5	2,3	2,3	2,4	0,5	0,6	0,6
$\bar{X}+3\sigma$	2,1	1,9	2,0	2,6	2,5	2,5	0,8	0,7	0,8
$\bar{X}-3\sigma$	1,5	1,5	1,4	2,2	2,3	2,4	0,4	0,5	0,6

	TOC (mg/l)			Temperature 1 (° C)			Temperature 2 (° C)		
	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017
Average	1,3	1,3	1,3	14,2	14,4	17,2	14,1	17,3	1,0
Standard deviation	0,048	0,039	0,046	0,095	0,048	0,166	0,065	0,064	0,186
Max	1,5	1,5	1,4	14,4	14,6	17,4	14,3	14,3	17,6
Min	1,2	1,2	1,2	14,1	14,2	16,4	13,9	14	16,5
$\bar{X}$ **+3**σ	1,4	1,4	1,4	14,5	14,5	17,7	14,3	14,4	17,9
$\bar{X}$ **-3**σ	1,2	1,2	1,1	13,9	14,3	16,7	13,9	14	16,8

	Color (Hazen)			pH			Conductivity (µS/cm)		
	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017	April 22, 2017	May 3, 2017	June 5, 2017
Average	4,2	5	6,4	7,3	7,2	7,1	882,9	882,7	885,6
Standard deviation	0,375	0,264	0,169	0,013	0,013	0,012	5,308	4,84	13,021
Max	5,6	7,3	7,1	7,3	7,3	7,2	892	890	912
Min	3,7	4,7	6,0	7,3	7,2	7,1	874	864	848
$\bar{X}$ **+3**σ	5,3	5,8	6,9	7,3	7,2	7,1	898,8	897,2	924,7
$\bar{X}$ **-3**σ	3,1	4,2	5,9	7,3	7,2	7,1	867,0	868,2	846,5

3.2.5 Analysis of detected events

3.2.5.1 Effect of hydraulic parameters

- **Consumption**

Figure 3.17 shows data collected by S::CAN at Polytech during February 2017. Although the signals are quasi-constant, some events have been detected during this period. To identify the source of these deviations, the consumption profile was plotted for the same period. A strong correlation is observed between the peaks of water quality parameters and the increase in the consumption profile. During night, the consumption is very low and signals are considered as relatively stable. In the morning, the flow increases due to the consumption (at around 8:00 am). The sudden augmentation in the flow induces the extraction of particles from the aging water pipes. The release of deposits is followed by an increase in the water quality parameters (such as UV, Turbidity, etc.). The deviation of these parameters from the stable line indicates the presence of suspended matters.

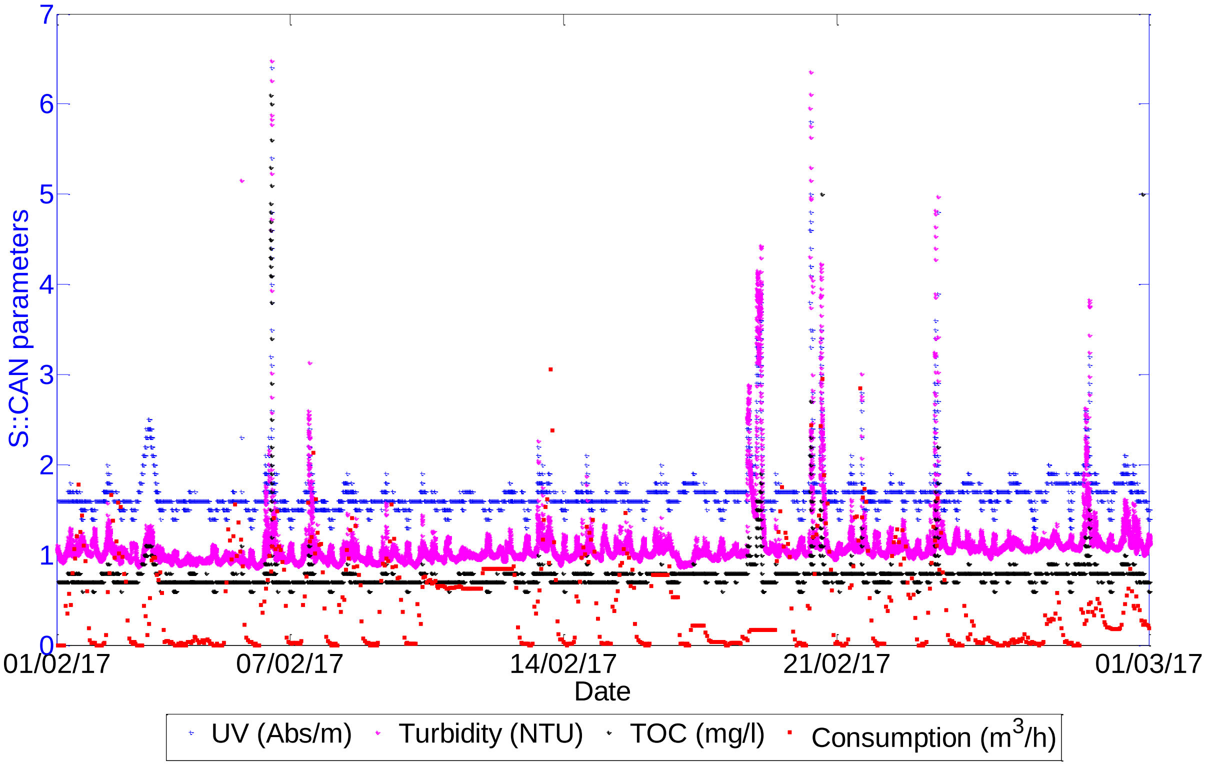


Figure 3.17. Data recorded at Polytech'Lille during February 2017.

- **Pressure**

Figure 3.18 illustrates the variation of S::CAN data, with hydraulics parameters (pressure and consumption), from May 21, 2017 till June 1, 2017 at Polytech'Lille. A significant event has been detected between May, 28 and May, 29. It is characterized by a large deviation of water quality parameters (UV, Turbidity, TOC, etc.) with an amplitude of 2 *units*. To analyze the origin of this event, the consumption and the pressure profiles are plotted during this period. As indicated in the black rectangle of Figure 3.18, the augmentation of water quality parameters is preceded by a sudden variation of the pressure. On May, 28 around 4:00 pm, a large deviation is observed in the pressure profile. The pressure drop could result from network interventions.

The important diminution of pressure is followed by a sudden extraction of suspended matters from water pipes. The peaks of water quality parameters illustrate the deposits release resulting from this pressure variation. These peaks coincided also with a flow increase. This event remains till May, 29 around 2:00 pm. After this period, the water quality parameters returned to standard variations.

Figure 3.19 gives the data recorded at Barroi for the same period (Week of May, 21). The same phenomenon is observed on May, 28. The sudden variation in pressure profile is followed by an increase in water quality parameters of around 1.2 *units*. This event appears clearly at Barroi from 4 am, while it was observed from 12 am at Polytech.

A good correlation is observed between hydraulic parameters (consumption and pressure) and water quality parameters. A variation in consumption or pressure profile can induce an increase in the values of indicators parameters. The extraction of suspended materials from water pipes explains this relation between consumption and/or pressure and water quality signals.

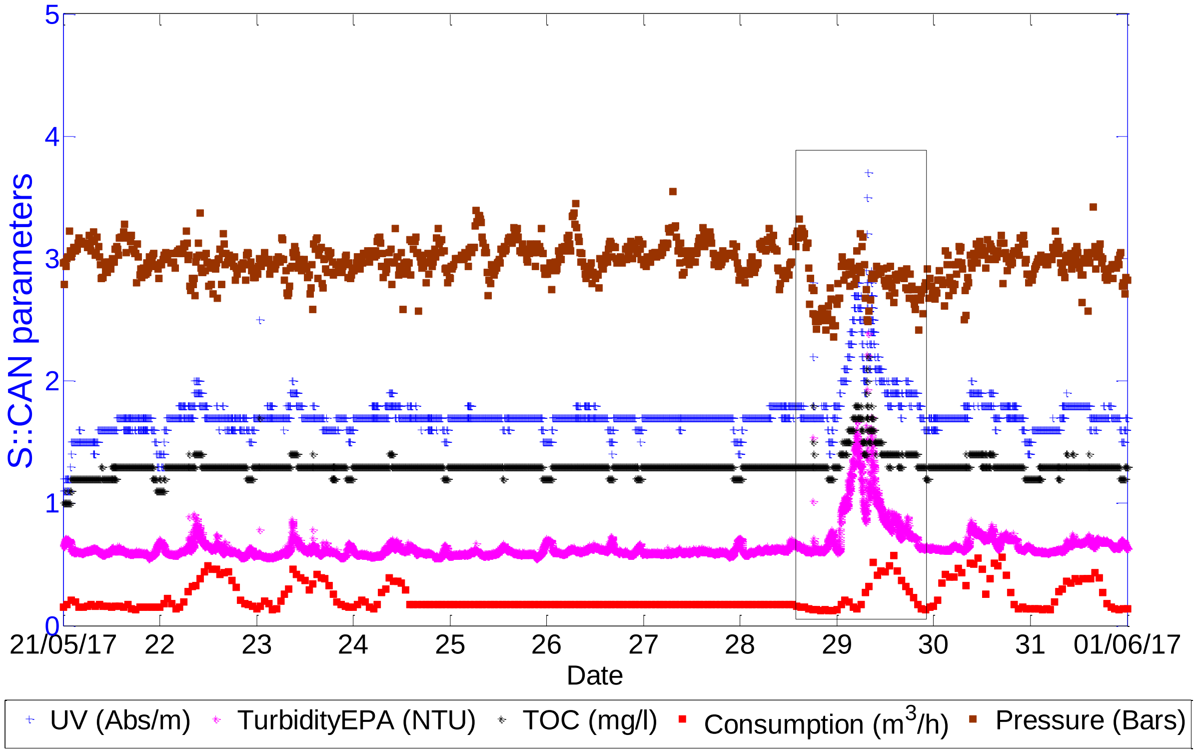


Figure 3.18. S::CAN data with hydraulic parameters on the week of May 21, 2017 at Polytech'Lille.

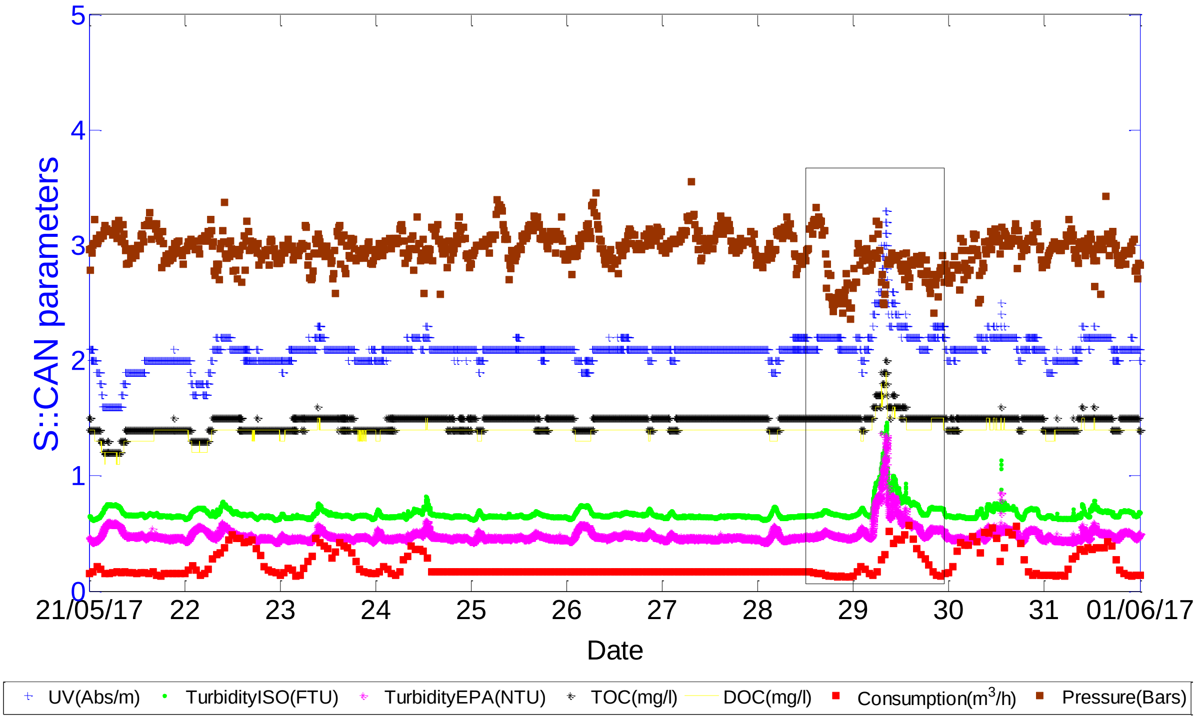


Figure 3.19. S::CAN data with hydraulic parameters on the week of May 21, 2017 at Barroi restaurant.

3.2.5.2 Network interventions

In addition to events correlated with hydraulic parameters, other factors can induce perturbations in the water quality signals. Figure 3.20 illustrates the variation of the water quality parameters on October 20, 2017 at Polytech. An event occurred observed at 1:00 pm.

It is characterized by significant deviation of UV, Turbidity and TOC. The event remains for approximately 2 hours with a mean amplitude of 3 *units*.

During renovation works at block D in Polytech'Lille, "a pipe pulling on the primary network of drinking water passing through the construction site" required the intervention on the network for repair actions on October, 20 around 1 pm.

This intervention is correlated with meaningful peaks of water quality parameters. Repair actions can be the source of anomaly in water, especially the presence of microorganisms or the germs development. UV, Turbidity and TOC are affected by such intrusion in water. The augmentation of their values can be explained by the network intervention.

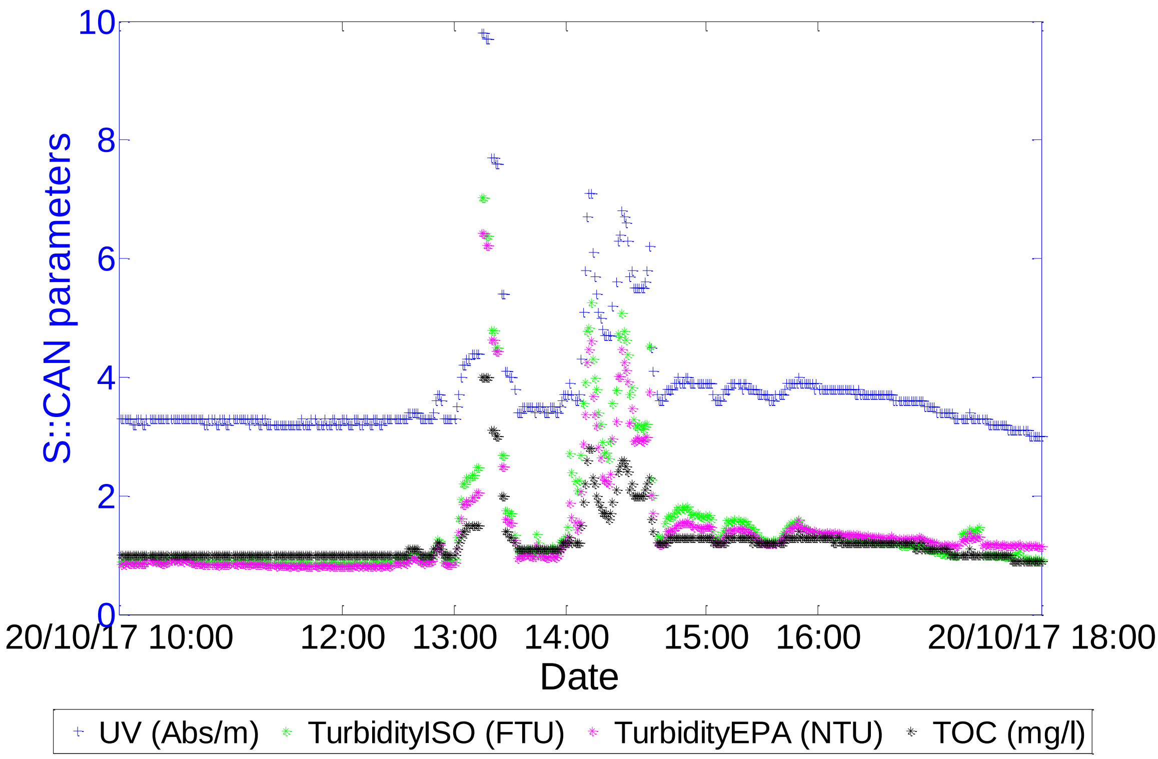


Figure 3.20. S::CAN data on October 20, 2017 at Polytech'Lille.

Figure 3.21 shows another example of the effect of network intervention on the water quality variation. It indicates three events, which occurred during the last week of September 2017. During this week, multiple "water interruption" occurred at Polytech'Lille during the construction works of block D. "Water interruption" has been illustrated by the deviation of water quality parameters. The event on September, 25 has a low amplitude and ends after few hours. The event on September, 28 can be interpreted as outlier, since the amplitude is very high, but the deviation is instantaneous. The major event occurred on September, 30. This event, of amplitude 1 *unit*, began on September, 29 at 10:00 pm and remained till September, 30 at 1:00 pm.

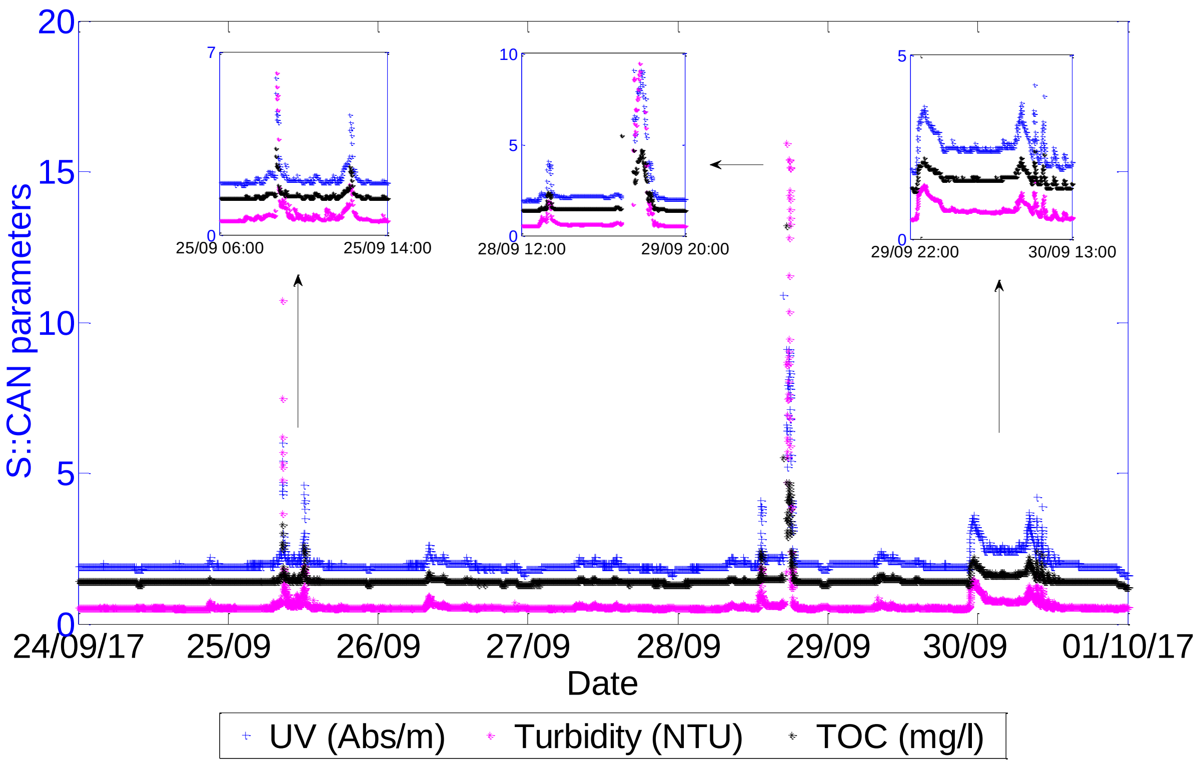


Figure 3.21. S::CAN data on the Week of September 24, 2017 at Polytech'Lille.

3.2.5.3 Effect of Temperature

Figure 3.22 gives the variation of UV, Turbidity and TOC between June 15, 2017 and June 25, 2017 at Polytech'Lille. An event occurred on June 20, 2017. A significant increase is observed in the profiles of S::CAN data. This variation in the water quality is also detected at Barroi (Figure 3.23). At Polytech'Lille, the deviation began around 3:00 am with an average amplitude of 1.5 *units*, while at Barroi, the event is observed from 7:00 am with an average of 0.7 *units*. On June 21, 2017, signals returned to their normal variations.

During this period, a "heat wave" has taken place in Villeneuve d'Ascq. An augmentation of the ambient temperature is noticed. This change in temperature could have an impact on the functionality of sensors. This event highlights the importance of the operating conditions of the sensor. The room temperature should not be very high nor too low, in order to ensure correct control of the water quality parameters.

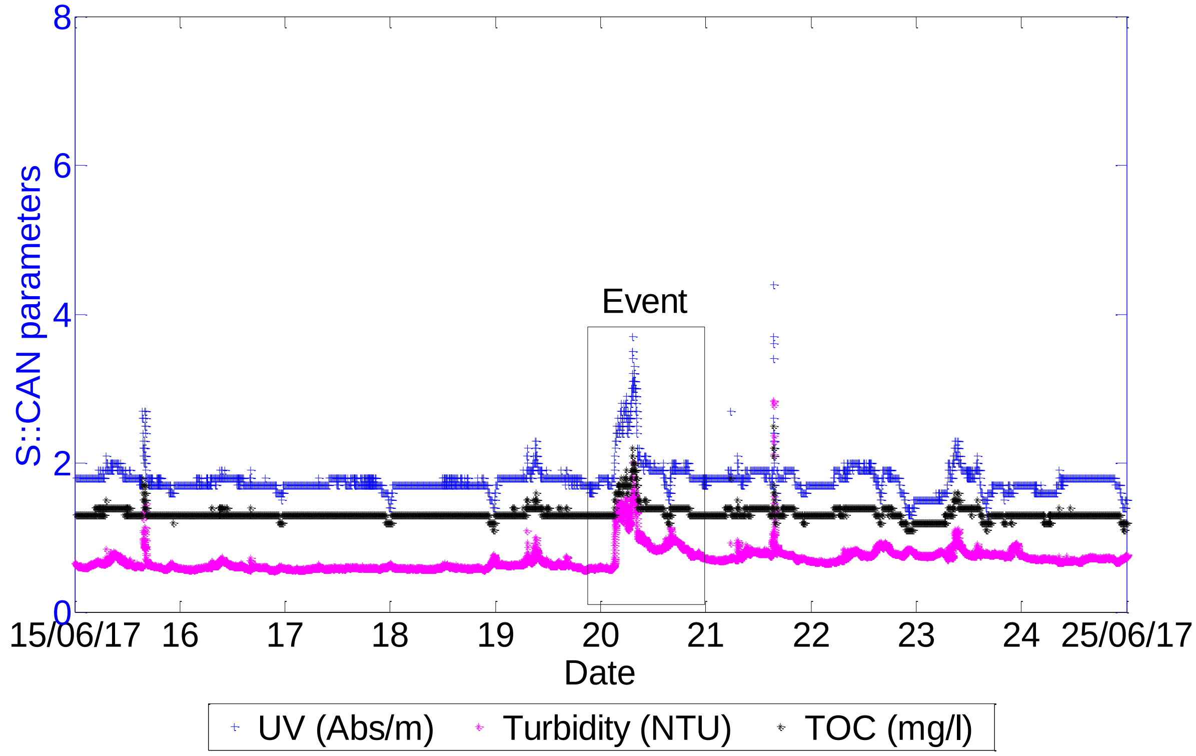


Figure 3.22. S::CAN data between June 15, 2017 and June 25, 2017 at Polytech'Lille.

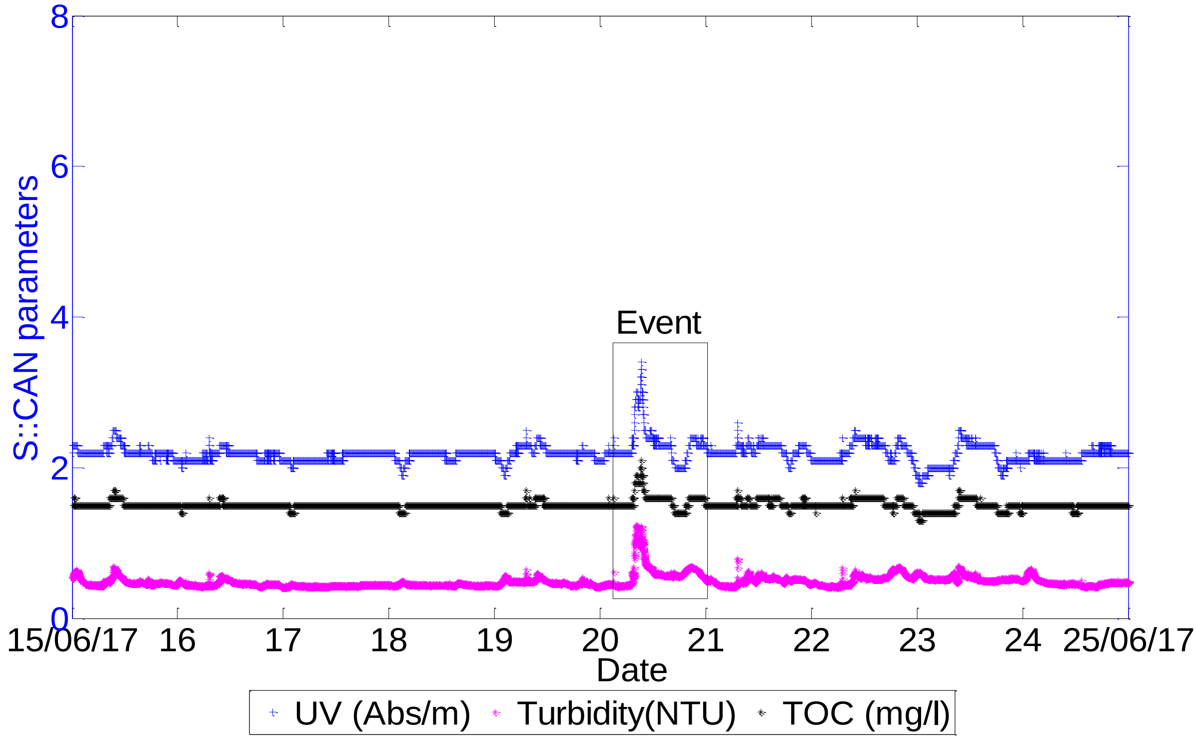


Figure 3.23. S::CAN data between June 15, 2017 and June 25, 2017 at Barroi.

3.2.5.4 Unknown anomaly

The variation of multiple S::CAN parameters between October 10, 2016 and October 30, 2016 at Polytech is given in Figure 3.24. A large deviation has been observed in October 2016. This event began on October, 14 at around 12:00 pm and remained till October, 30. During this period, the reference lines for all signals deviated significantly. For UV, the maximum deviation is about 12 Abs/m while for Turbidity and TOC, the variation is around 6 *units*. This anomaly indicates a major intrusion in water. The cause of this abnormality is unknown. After October, 30, data recovered normal values.

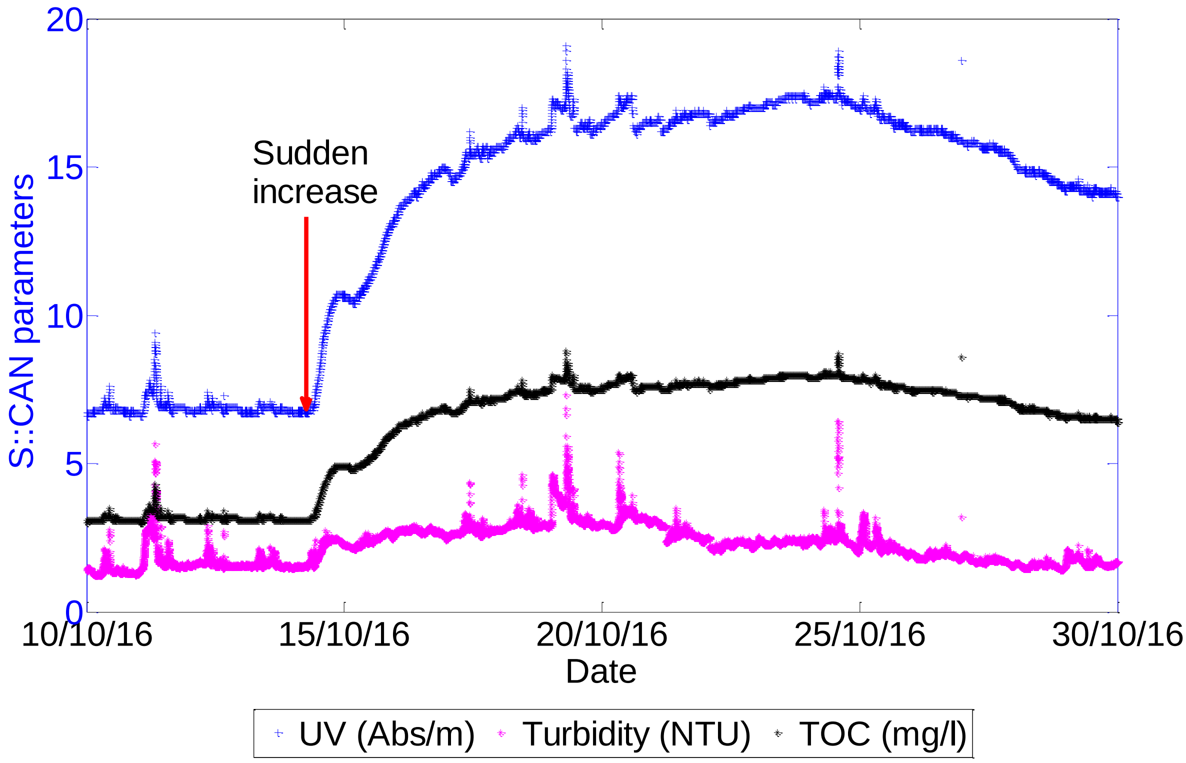


Figure 3.24. Event detected on October 14, 2016 at Polytech'Lille.

3.2.5.5 List of events

Table 3.6 reports significant events recorded during the monitoring period in 2016 and 2017. It indicates that the increase in consumption, each day at 8:00 am, is followed by an increase in the water quality parameters, especially at the University.

In addition to the variation correlated with the flow increase, deviations from the stable lines are observed. The reason of each event is indicated in Table 3.6. These variations are illustrated clearly in the deviation of some specific parameters: UV, Turbidity, TOC and Color.

However, these interpretations cannot certify that a real contamination, due to the presence of pollutant, has occurred. The perturbations in signals indicate only that an anomaly occurred in water. The online analysis contributes to understand the cause of these deviations.

Table 3.6. List of events detected during the monitoring period.

Date of event	Reason	Locations
Every day at 8 am	Consumption increase	Polytech
October 14, 2016	Unknown	Polytech
May 28 and May 29, 2017	Sudden and large deviation of pressure	Polytech and Barroi
June 20, 2017	"Heat wave"	Polytech and Barroi
Week of September 24, 2017	"Water interruption"	Polytech
October 20, 2017	Repair actions	Polytech

3.2.6 Comparison of data recorded at Polytech'Lille and Barroi

Figure 3.25 illustrates S::CAN data for the period between July 23, 2017 and July 30, 2017 at Polytech'Lille and Barroi. It indicates that:

- Figure 3.25 (a): Turbidity profiles are very similar at both locations, except some peaks detected at Polytech'Lille.
- Figure 3.25 (b): TOC data have the same order of magnitude (between 1.5 and 2 mg/l).
- Figure 3.25 (c): the shape of Conductivity variation is the same at both locations. However, a difference in the measured value is observed. Conductivity, recorded at Polytech, is higher by 50 µS/cm from that at Barroi.
- Figure 3.25 (d): Temperature is more stable at Polytech with an average value of 17 °C. At Barroi, Temperature has regular variation between 17 and 20 °C.
- Figure 3.25 (e): pH data are quasi-constant with same values at both locations.
- Figure 3.25 (f): Chlorine at Barroi is very low in comparison with that at Polytech. At Barroi, Chlorine remained under 0.1 mg/l, while it varied between 0.1 and 0.4 mg/l at Polytech.

Analysis conducted in this chapter proves that signals measured at Barroi are generally more stable, since the use of water is lower at the Restaurant. The comparison shows that certain water quality parameters could differ between two locations at the campus, due to the difference in pipes ages and quality in the campus.

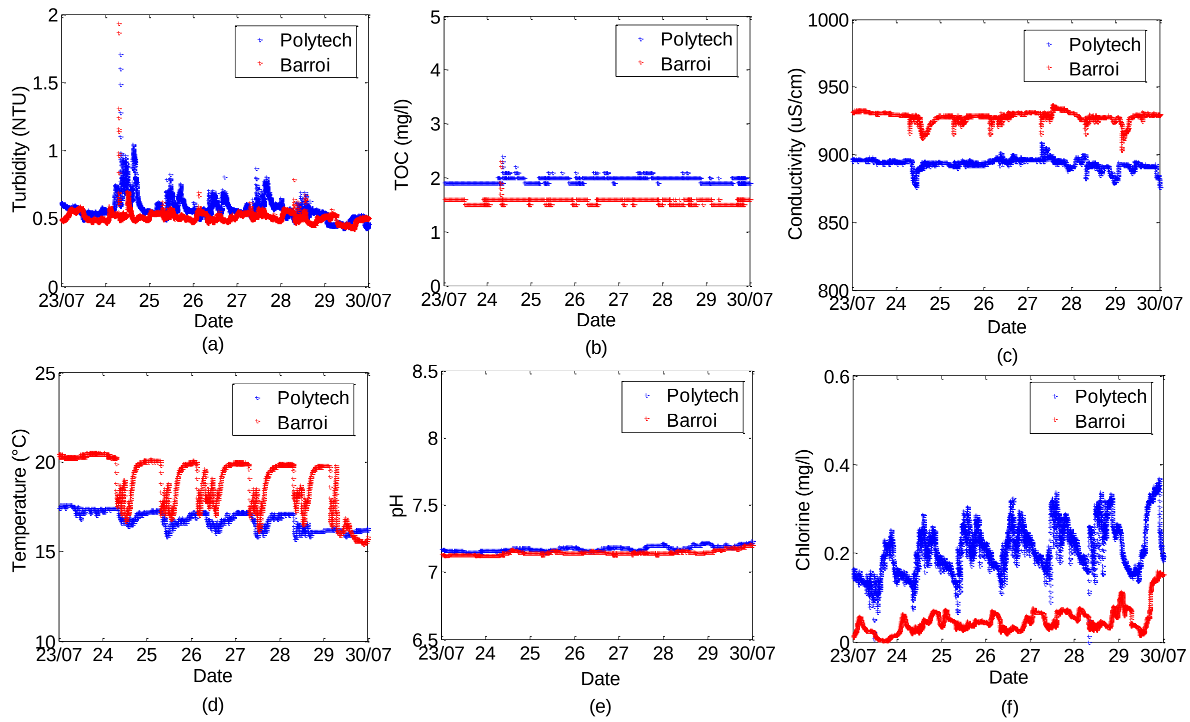


Figure 3.25. Comparison of S::CAN data between Polytech'Lille and Barroi. (a) Turbidity (NTU); (b) TOC (mg/l); (c) Conductivity (µS/cm); (d) Temperature (°C); (e) pH; (f) Chlorine (mg/l).

3.3 Analysis of EventLab data

3.3.1 Variation of phase and F24 Response

Any change in the composition of the water matrix will induce a change in the combined Refractive Index (RI). Figure 3.26 shows an example of the variation of the RI in function of the presence of solute in water. It gives the concentration of Sodium Sulfate in pure water and the resulting change in RI [113]. It shows a linear relationship between the concentration of solute in water and the corresponding variation in RI.

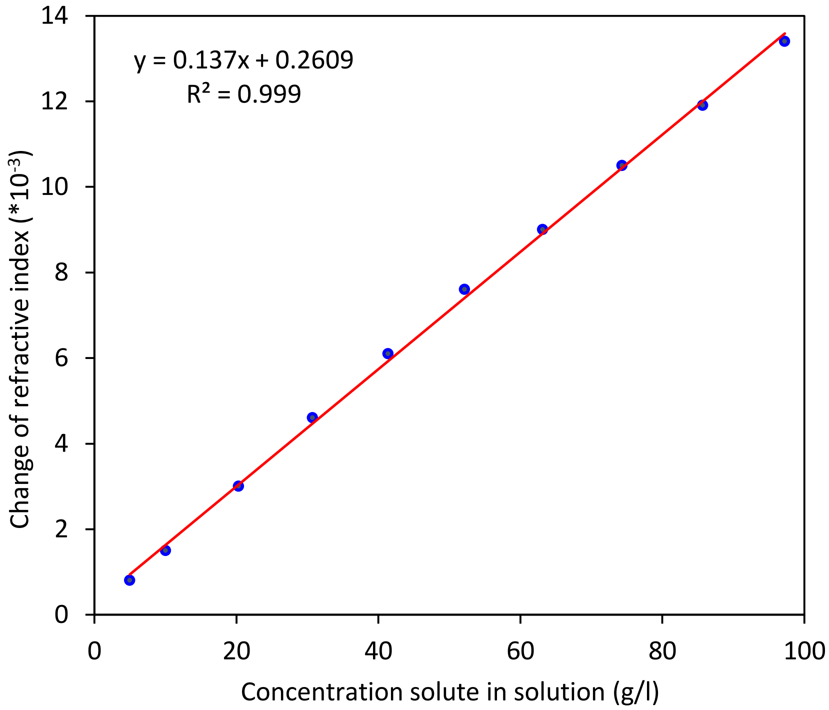


Figure 3.26. Variation of Refractive Index (RI) as function of the concentration of Sodium Sulfate [113].

The main purpose of EventLab sensor is to measure the variation of phase, which is proportional to the change in RI. These changes are controlled continuously by EventLab through the measurement of phase. At each minute i, the variation of phase is calculated (eqn 2.3 in Chapter 2) as follows:

$$\Delta\Phi = \Phi(i+1) - \Phi\ (i) \tag{3.2}$$

In a safe drinking water, the change in RI as well as the variation of phase is quasi-constant. Exceeding the normal variations ($\pm 3\sigma$) [114] is considered as abnormality, which could indicate a potential contamination of water. Data recorded outside these limits should be analyzed to identify the type of deviation (event, outlier, etc.) and its origin.

An example of the variation of phase measured by EventLab is illustrated in Figure 3.27. It shows the change in phase monitored between October 3, 2016 and October 4, 2016 at Polytech'Lille. The variation lies between the upper and lower accepted bounds, which indicates a normal drinking water. However, two abnormalities were observed (red circles) during this period (on October, 4 around 5 am and 7 am). In order to analyze the source of these deviations, the consumption profile is plotted in Figure 3.27 for the same period. It indicates that a sudden increase in the consumption is followed by a peak in EventLab signals. This correlation can be explained by the extraction of suspended matters when the flow increases suddenly. The drawn of such substances from aging pipes affects the water quality composition and therefore its RI. This impact is illustrated in the deviation of $\Delta\Phi$ from the normal variations.

Figure 3.28 shows an example of the variation of phase measured on January 10, 2017 at Barroi restaurant. The majority of events remained between the two limits ($\bar{X} \pm 3\sigma$). However, some events exceeded the lower limit (-0.02 radians). These measurements are observed in a periodic way (each hour). Since abnormality occurred for one single time step, these values can be considered as outliers and should not generate alert.

In addition to the phase, EventLab measures F24 Response which is a moving average of response, providing information on slow changes in water composition. It indicates how much a response data is above or below the average of the preceding 24 hours. Figure 3.29 provides the variation of F24 Response between May 19, 2017 and May 27, 2017 at both locations (Polytech and Barroi). It indicates small values in both sites. However, the variation is more important at Polytech. At Barroi, F24 response remains between ± 0.3 while it ranges between ± 0.7 at Polytech. This indicates that signals are less stable in Polytech.

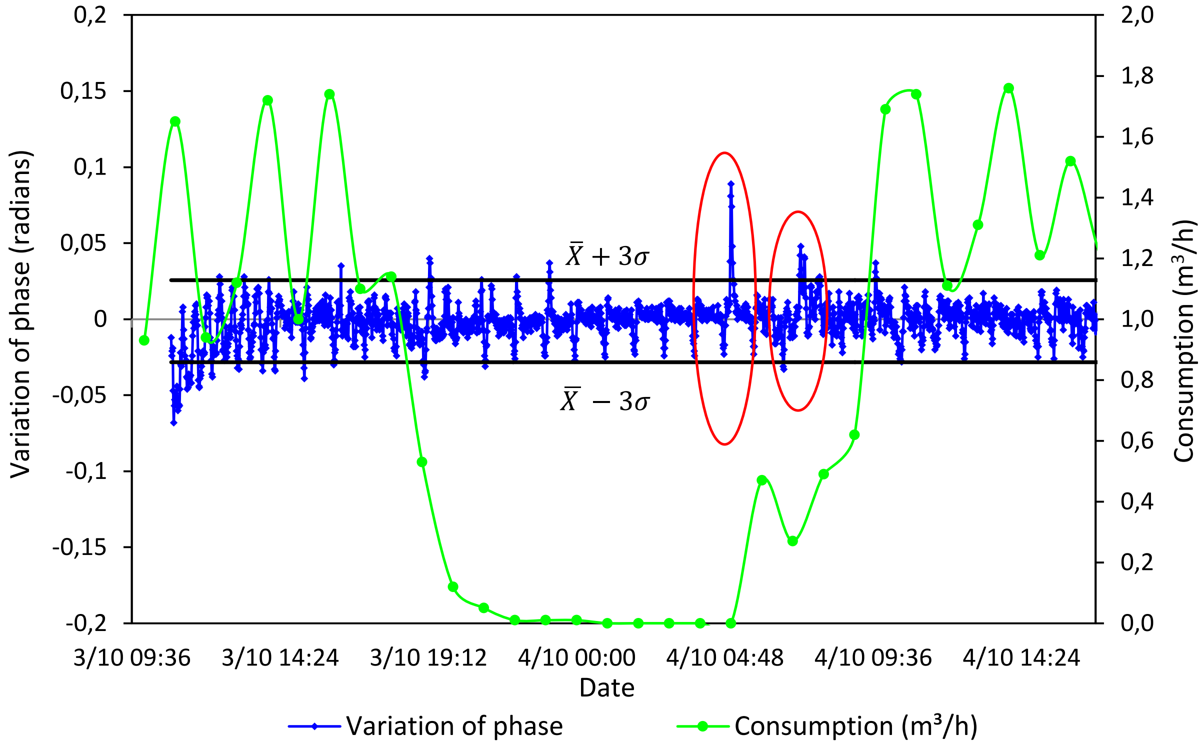


Figure 3.27. Variation of phase with consumption profile for October 3, 2016 and October 4, 2016 at Polytech'Lille.

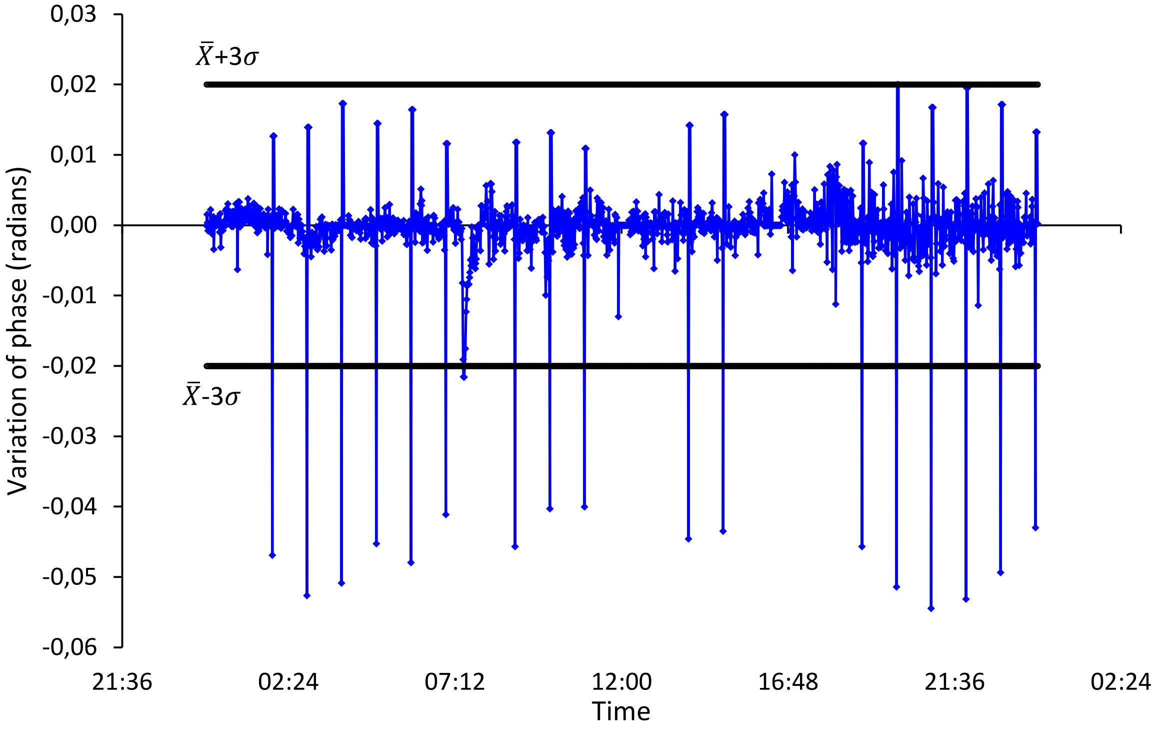


Figure 3.28. Variation of phase on January 10, 2017 at Barroi.

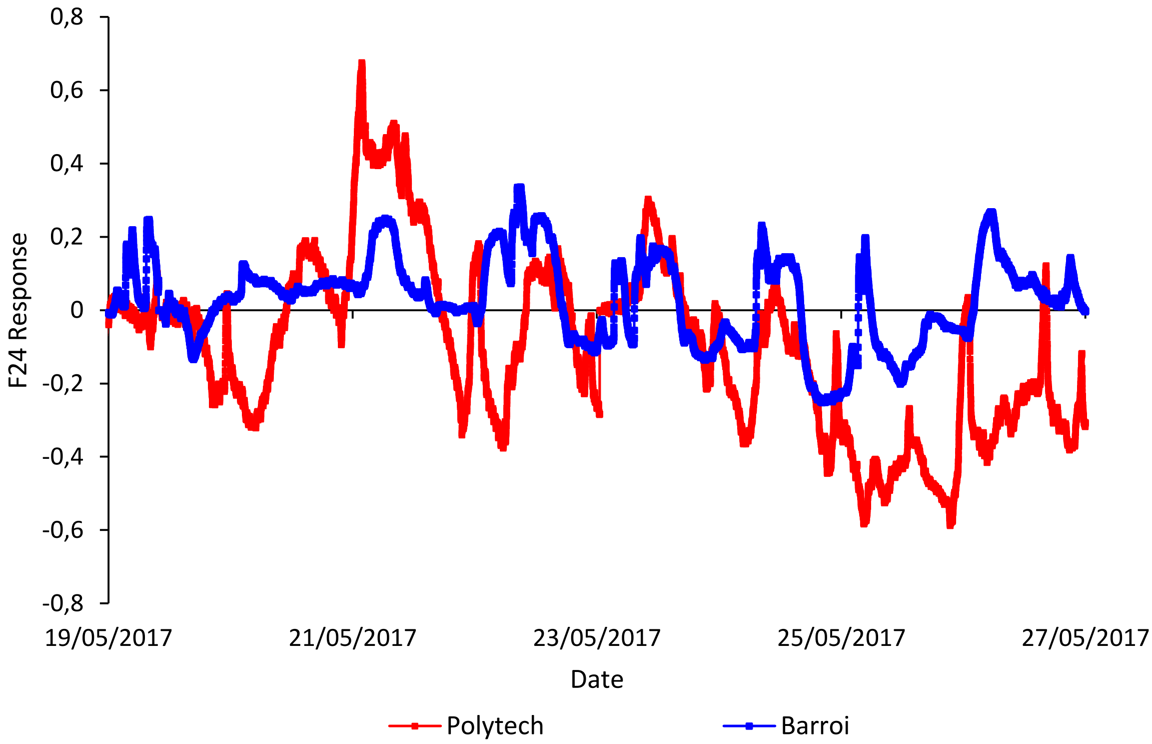


Figure 3.29. F24 Response between May 19, 2017 and May 27, 2017.

3.3.2 Variation of signal health and signal level

In order to control the probe performance, the signal health and the signal level are controlled continuously. As mentioned previously (in Chapter 2), their values should not drop below 0.15. If signal health and/or level is less than 0.15, maintenance and cleaning task are required. In some critical cases, very low values could imply the need of probe replacement, if the cleaning is not sufficient.

Figure 3.30 shows the signal health and the signal level between January 13, 2017 and January 16, 2017 at both locations (Polytech and Barroi). Their variations are relatively stable and above the accepted limit of 0.15. However, values recorded at Polytech are significantly smaller than those measured at Barroi. This result indicates that the probe is cleaner at Barroi. The EventLab probe installed at Polytech requires more maintenance and cleaning.

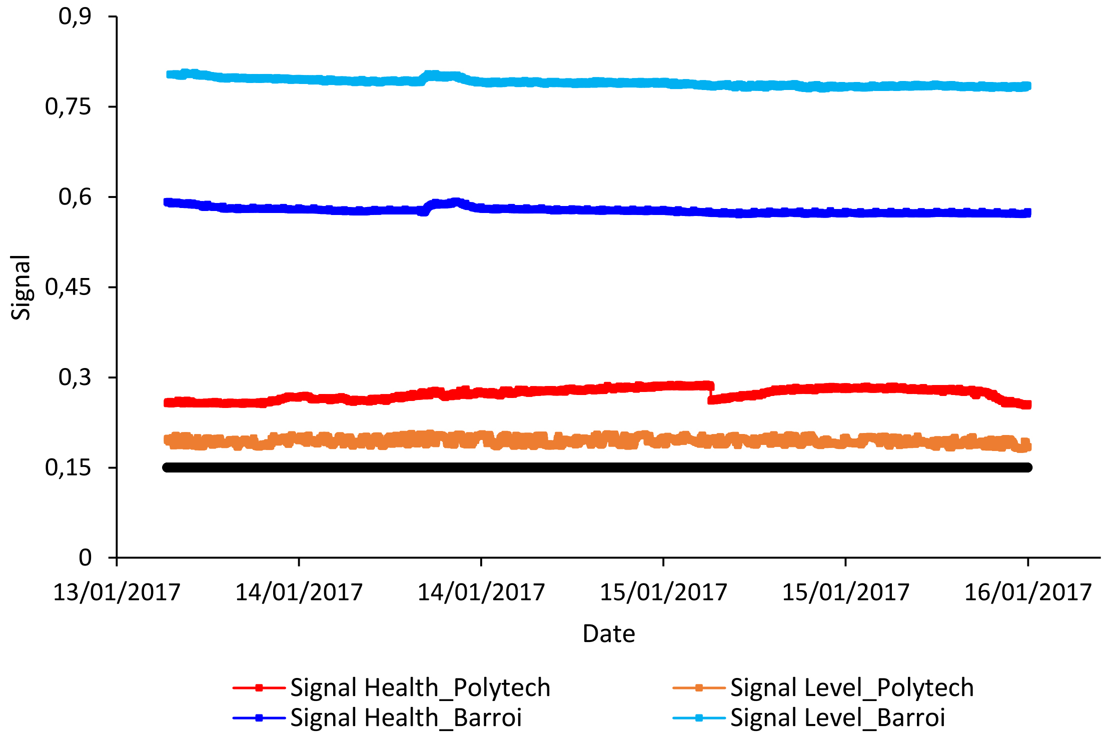
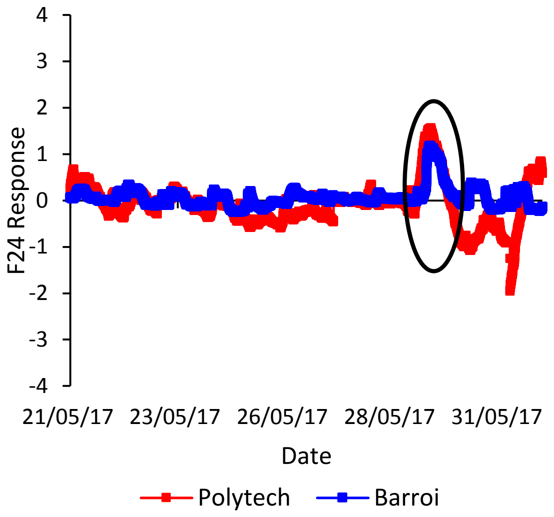


Figure 3.30. Example of signal health and signal level at Polytech'Lille and Barroi.

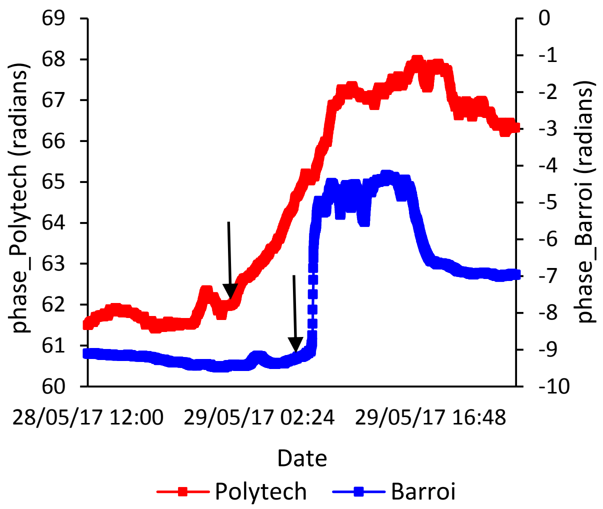
3.3.3 Event detected

The online analysis of EventLab data allows the detection of some abnormalities in water. An example of event visible in EventLab data is given in Figure 3.31. A significant event was detected on May 29, 2017. Figure 3.31 (a) gives the variation of F24 Response between May 21, 2017 and May 31, 2017 at both locations. Before May, 29, F24 Response tends to be constant (about 0). However, an important deviation (higher than 1) is observed on May, 29 in F24 Response signals (as indicated in the black circle). At Polytech, the deviation appeared from 12:00 am while it was visible from 4:00 am at Barroi.

Figure 3.31 (b) shows the signals of phase on May 28, 2017 and May 29, 2017. A variation of phase, about 5 radians, is observed on May, 29 at both locations. Since the sensor worked well during this period, it means that a real anomaly has occurred, which could involve a change in the water composition. This event has been also detected by S::CAN sensor and related to a sudden variation in pressure profile (section 3.2.5.1: Pressure).



(a)



(b)

Figure 3.31. Event detected on May 29, 2017. (a) F24 response; (b) Phase (radians).

3.4 Comparison between S::CAN and EventLab

In order to compare the performance of S::CAN and EventLab, the variation of some S::CAN parameters, the phase as well as the consumption profile are plotted for the same period. Figure 3.32 illustrates the variation of TOC, Turbidity, phase and consumption on October 4, 2016 at Polytech'Lille. A deviation from the stable lines is observed in TOC and Turbidity signals around 7:00 am (black arrows). It is followed by a significant variation of phase, detected in EventLab signal. This deviation is correlated with an increase in consumption profile between 7:00 am and 8:00 am.

However, another deviation is detected around 10:00 am by S::CAN (clearly in Turbidity signal) which is correlated by a significant increase in flow. However, this anomaly is not observed in the phase signal of EventLab sensor.

The comparison proves that both sensors have a good performance in the detection of abnormalities in water. However, the reliability of S::CAN sensor is better in the identification of event, since certain deviations are not recorded in the phase variation of EventLab sensor.

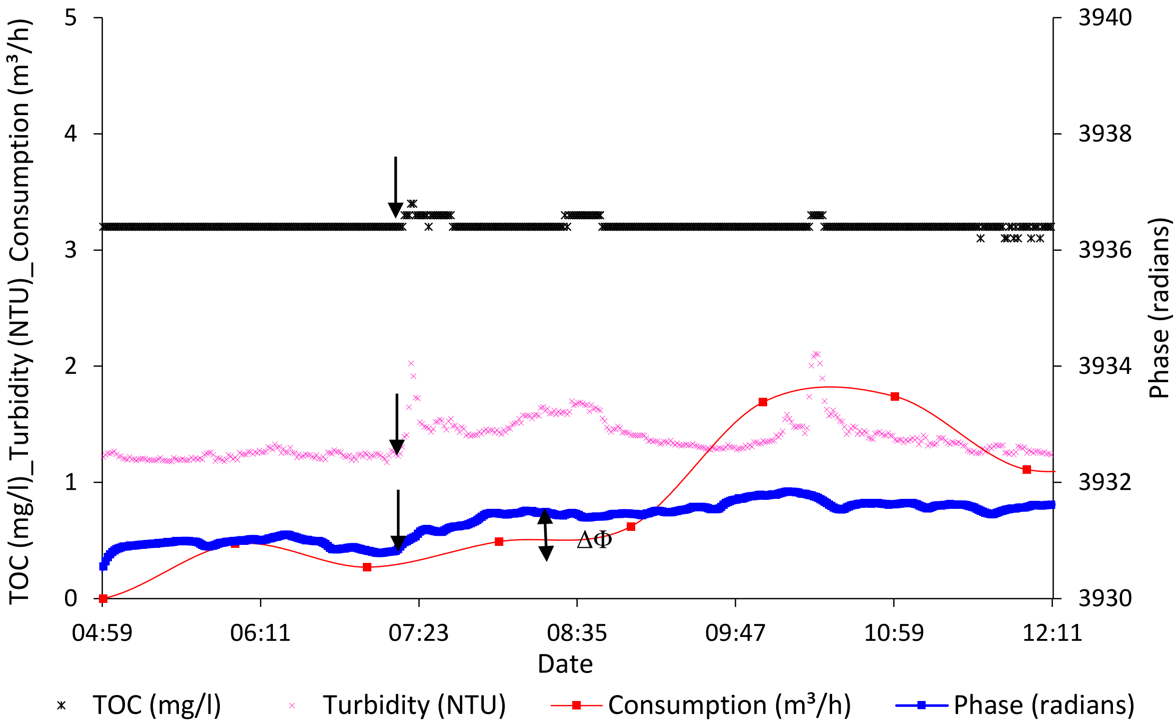


Figure 3.32. Comparison between S::CAN and EventLab response on October 4, 2016 at Polytech'Lille.

3.5 Conclusion

This chapter presented analysis of data recorded by S::CAN and EventLab for the water quality control at the Scientific Campus. Analysis focused on the deviation of recorded values from the baseline. This deviation is expected to result from a variation in the water quality.

The comparison of data recorded by S::CAN with laboratory analyses allowed the validation of the good functioning of S::CAN for some water quality parameters (pH, Temperature and Conductivity) and the necessity to calibrate regularly other parameters such as Turbidity, TOC and Chlorine. This result shows the necessity to conduct regularly a control of the on-line water quality devices.

Data recorded by S::CAN showed good correlations between UV, Turbidity, TOC and Color, which indicates a good functioning of S::CAN. They also showed the occurrence of some events (deviation from baseline values), which generally were correlated with a change in the hydraulic parameters (water flow or water pressure). We expect that this change in the water quality is due to aging pipes. An increase in the water flow or pressure could lead the liberation of particles from the pipes and then to a perturbation in the water quality.
The comparison of S::CAN and EventLab records showed good general agreement; however S::CAN detected more events than EventLab.

These results show that the on-line monitoring allows to detect anomalies in the water quality (deviation from the baseline). However, it is still difficult to understand the real cause and nature of these events, because of the complexity of the water quality control. In the following chapter, three methodologies of water anomaly detection are presented and applied to S::CAN data.

Chapter 4. Water Anomaly Detection Using Statistical, Artificial Intelligence and Event Detection System (EDS) Methods

4.1 Introduction

This chapter presents different methods for anomaly detection of water contamination. It describes three methodologies to detect abnormalities observed in sensor's responses. The first part describes the use of the linear prediction model as statistical method for data forecasting. At each time step, observed value of each signal will be compared with expected data to identify unacceptable measures. The second part concerns the Support Vector Machine (SVM) as Artificial Intelligence method. It will be applied as binary classifier to differ between normal and anomalous classes of water quality. The advantage as well as the limitation of these two methods will be presented.

Results show the need of an Event Detection System (EDS) for an early identification of water event. Therefore, an event detection model will be developed using Canary software. It will analyze data in real-time to determine the probability of events. A sensitivity analysis is conducted to determine appropriate parameters that reduce the generation of false alarms and increase the rate of detection of real events.

4.2 Sensitivity Analysis

This section presents the methodology for the sensitivity analysis used for parameters selection. A confusion matrix is used to evaluate the performance of detection methods. Figure 4.1 gives the corresponding confusion matrix for event classification in comparison with the real condition. It defines four main categories:

- True Positive (TP): An actual event is detected. The method indicates an event which is well detected by the sensor.
- False Negative (FN): A real event is not detected. An actual event, identified by the sensor, is not reported by the method.
- False Positive (FP): The method reports an event while no true event occurred; generation of false alarm.
- True Negative (TN): Both the sensor and the method do not record an event. There is no real event and the method do not identify an event.

		True condition	
		Positive condition: Event	Negative condition: Non-Event
Predicted condition	Positive prediction: Event	True Positive	False Positive
	Negative prediction: Non-Event	False Negative	True Negative

Figure 4.1. Confusion matrix for event classification.

A decision is considered correct if the real and the predicted condition are similar (green box), i.e. if an event occurs, the method reports it, while if the true condition indicates there is no event then the method also do not record any event. However, if the estimated condition differs from the true condition, we have incorrect decision. Two cases can lead to this situation. If the real condition indicates there is no true event, while the method generates a false alarm of event (yellow box), utility should verify the existence or not of a real event. However, the most dangerous case occurs if a true event is missed by the method (red box). This last situation is the most critical for water utility and should be avoided.

The following performance parameters will be used for parameters selection and are calculated as follows:

$$Sensitivity = \frac{TP}{TP+FN} = TPR\ (True\ Positive\ Rate) \tag{4.1}$$

$$Specificity = \frac{TN}{TN+FP} = TNR\ (True\ Negative\ Rate) \tag{4.2}$$

$$1 - Specificity = \frac{FP}{FP+TN} = FPR\ (False\ Positive\ Rate) \tag{4.3}$$

$$Precision = \frac{TP}{TP+FP} = PPV\ (Positive\ Predictive\ Value) \tag{4.4}$$

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = ACC \tag{4.5}$$

4.3 Linear Prediction

4.3.1 Principle

Linear prediction modelling is used for different applications such as data forecasting, speech coding, etc. Based on historical data, the method will predict future values. At time step t, the predictor model forecasts a value $\hat{x}$ (t) using a linearly weighted combination of n past samples [115]:

$$\hat{x}(t)=\sum_{k=1}^{n} a_k\, x\,(t-k) \text{ with } a_k\text{: the predictor coefficients.} \tag{4.6}$$

The prediction error is then calculated as the difference between the actual value $x\,(t)$ and the predicted one $\hat{x}$ (t):

$$e\,(t) = x(t) - \sum_{k=1}^{n} a_k\, x\,(t-k) \tag{4.7}$$

The main purpose is to find the coefficients that minimize the least mean square error defined as follows [115]:

$$E\left(e^2(t)\right) = E\left[\left(x(t) - \sum_{k=1}^{n} a_k\, x\,(t-k)\right)^2\right] \qquad (4.8)$$

4.3.2 Use of Linear Prediction for the detection of water anomaly

The Linear Predictive Coding (lpc) method can be applied in water quality monitoring for anomaly detection. The statistical method will be able to estimate future data of water quality parameters, based on the analysis of historical data. In this study, S::CAN signals, recorded at Barroi, are used to test the efficiency of lpc method in identifying anomaly. The methodology used can be summarized in the three following steps:

- Training phase: Calculation of coefficients a and error e, based on a set of historical data.
- Test phase: Prediction of future values using the calculated coefficients a.
- Residual and Class group: Calculation of residual as the absolute difference between measured and predicted data. Result is compared to a user predefined threshold, then classified in two main categories: i) Class= 1 if $|Residual|$ > Threshold (which indicates unacceptable value) and ii) Class = 0 otherwise (which indicates normal value).

This method is developed in a Matlab code and it will be applied consecutively to two different periods:

- Period 1: From April 11, 2017 till April 30, 2017. This period will be used to determine the appropriate polynomial degree n and the adequate predefined threshold.
- Period 2: From April 11, 2017 till May 11, 2017. During this period, a sensitivity analysis is conducted to select the size of history window required for training phase.

For each period, two applications will be evaluated: i) analysis of one single parameter (Turbidity) (Figure 4.2) and ii) Analysis of several S::CAN parameters (Figure 4.3).

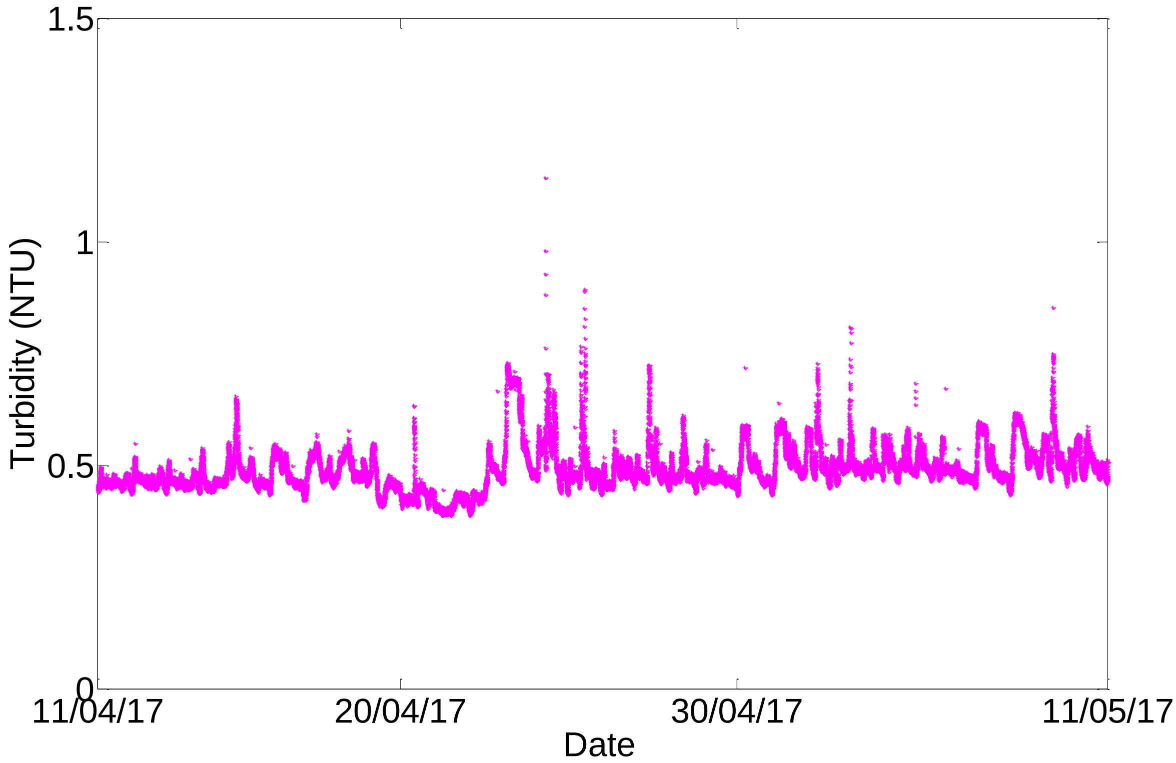


Figure 4.2. Variation of Turbidity between April 11, 2017 and May 11, 2017 at Barroi.

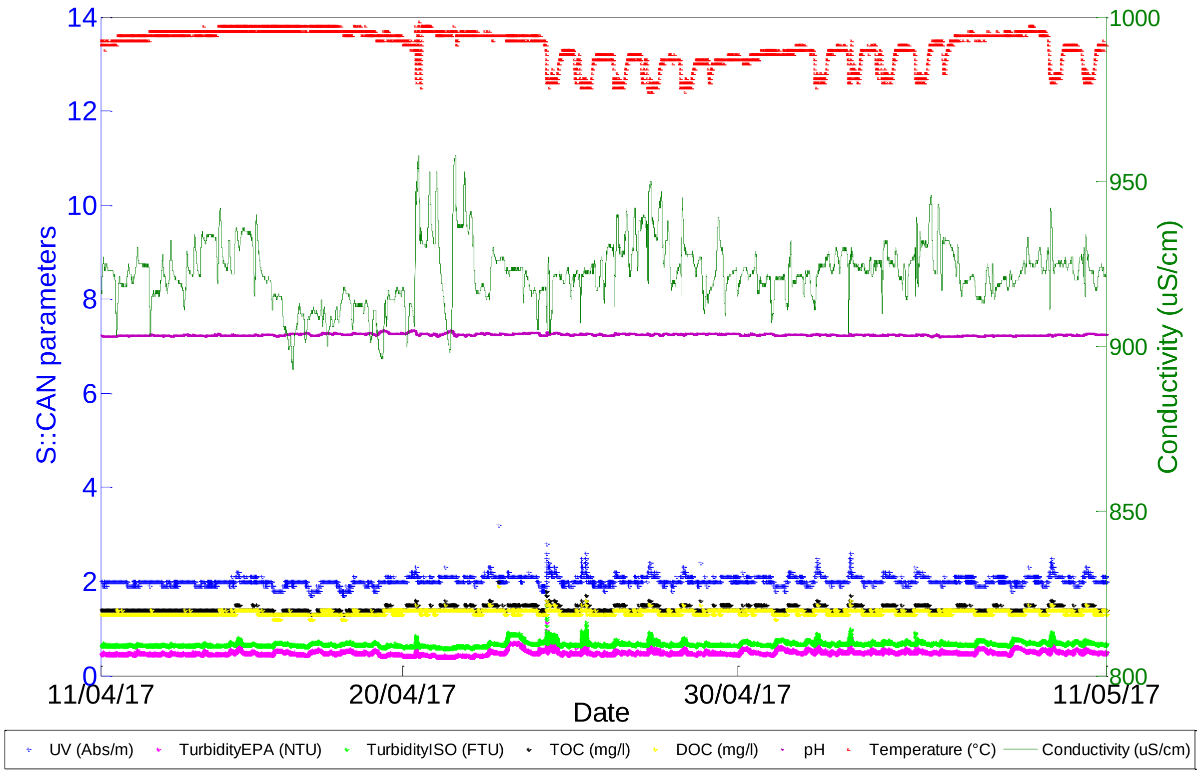


Figure 4.3. Variation of S::CAN data between April 11, 2017 and May 11, 2017 at Barroi.

4.3.2.1 Analysis of period 1

The variation of Turbidity EPA (Figure 4.2) during period 1 indicates 6 days in which significant deviations were observed: 15/04/2017; 22/04/2017; 23/04/2017; 24/04/2017;

25/04/2017 and 27/04/2017. During this period, data are divided as follows: i) 15000 time steps (about 10 days) for training phase, and ii) 12352 time steps (about 9 days) for test phase.

Different degree n of the polynomial equation used in the training phase are studied. In each case, the corresponding error is calculated. The results of the variation of the error in function of the degree n are illustrated in Figure 4.4. It indicates that the error decreases with the augmentation of the degree n. However, the error converges and reaches a minimum value of around 5.35E-05 from a degree n=3. A larger value n will increase the computational time without decreasing significantly the error. Therefore, a degree n=3 will be chosen in the following study.

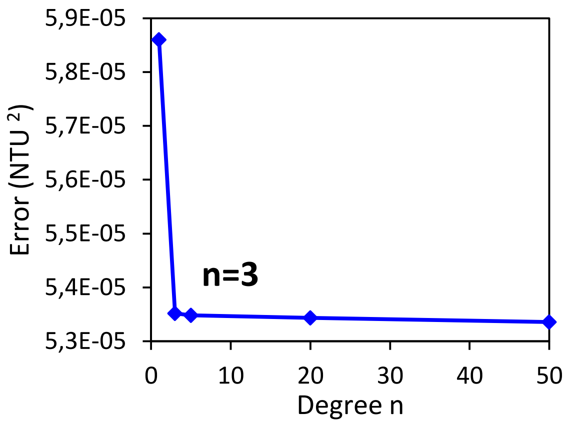


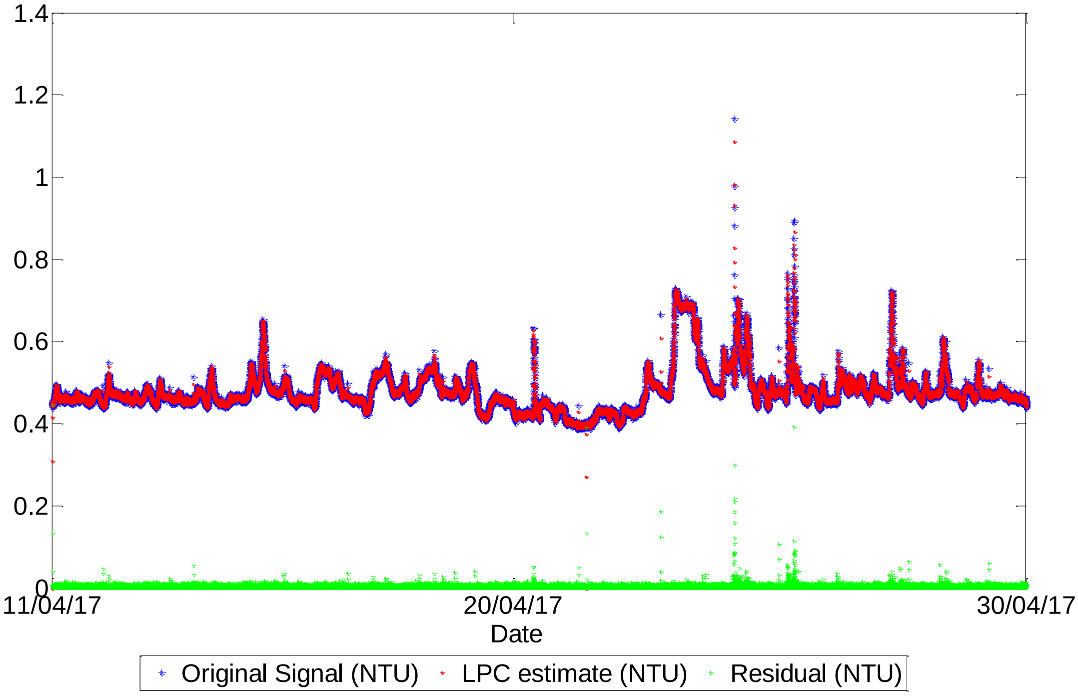
Figure 4.4. Error in function of degree n.

A sensitivity analysis is done to select the most appropriate predefined threshold. For each time step where the identified class group of the method is equal 1, an anomaly is assigned to the corresponding day. The comparison between the 6 real deviations observed and the date of anomaly detected by the method allows the calculation of the performance measures. Results of the calculation are given in Table 4.1. It should be mentioned that the threshold is expressed in function of standard deviation. From Table 4.1, a threshold = σ will be selected since it makes a tradeoff between increasing the true positive rate and reducing the generation of false alarms (FPR).

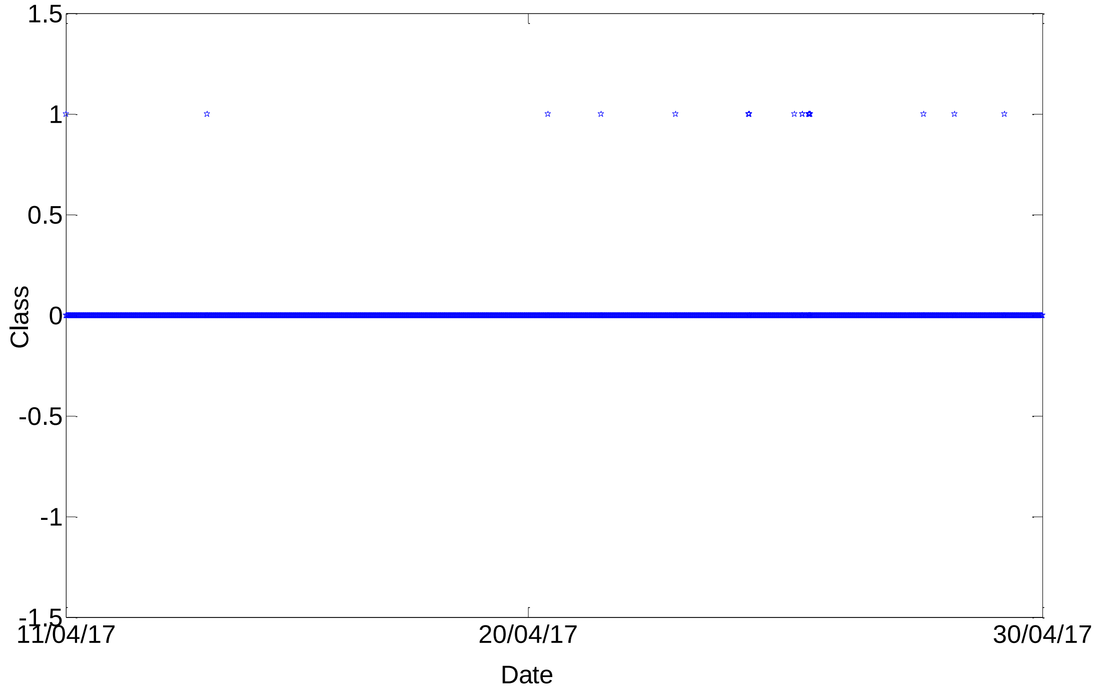
Table 4.1. Sensitivity analysis for different thresholds.

n=3, training=15000 time steps			
Predefined Threshold	TPR (%)	TNR (%)	FPR (%)
3б	50.00	92.31	7.69
2б	50.00	92.31	7.69
б	66.67	61.54	38.46
0.85б	66.67	53.85	46.15
0.75б	66.67	46.15	53.85
0.5б	100.00	15.38	84.62

Figure 4.5 (a) illustrates the results of the application of lpc method using n=3 and threshold= σ during period 1. It indicates relatively small difference between the measured signal and the estimated one. The calculated residual varies between 0 and 0.05 NTU, and reaches a maximum of 0.45 NTU. Figure 4.5 (b) shows the classes of each value of Turbidity. The majority of data belongs to the class 0 indicating normal values. Values of class 1 indicate deviation in the signal or a false alarm generated by the method.



(a)



(b)

Figure 4.5. Results of statistical method applied to Turbidity. (a) Comparison between measured and estimated Turbidity; (b) Classification of data.

During period 1, the statistical method lpc is also tested for several S::CAN parameters simultaneously: UV, Turbidity EPA, Turbidity ISO, TOC, DOC, pH, Temperature and Conductivity. Some deviations occurred in the different monitored signals (Figure 4.3). The same methodology described previously is used. However, to determine the final class group, at each time step, a new index s is defined as follows:
If Class (parameter i) =1 then s (i) =1, otherwise s (i) =0. The summation of s is calculated $\sum s\ (i)$ and two hypotheses are then tested:

- Hypothesis 1: $\sum s\ (i) > 1$ (which means at least two parameters belong to the class 1) implies that the final group class= 1, otherwise final class=0.
- Hypothesis 2: $\sum s$ (i) > 2 (which indicates that at least three parameters are of class 1) gives a final class=1, otherwise final class=0.

The sensitivity analysis applied, using hypothesis 1, shows an increase of the FPR as the TPR increases (Table 4.2) with low percentage of accuracy and precision. Each identification of real deviation is followed by the generation of false alarm. Using hypothesis 1, it is impossible to get simultaneously acceptable levels of sensitivity and specificity. This hypothesis will be excluded from the study, since it gives high rate of false positive alarm.

The results of sensitivity analysis obtained using hypothesis 2, are given in Table 4.3. It indicates very similar FPR for the different thresholds. In order to get the highest rate of true positive (100 %) with high accuracy (84.21 %), a threshold= σ (as the case for Turbidity) should be selected. However, it should be mentioned that the precision of this method is low. Even for the selected threshold, a maximum precision of 50 % can be obtained. This can be explained by a large number of negative examples (positive class is the minority). The FP will overwhelm the TP even at low FPR. This method is not very precise even it can detect true unexpected values.

Table 4.2. Sensitivity analysis for different S::CAN signals during period 1 according to hypothesis 1.

n=3, $\sum s\ (i) > 1$					
Threshold	TPR (%)	TNR (%)	FPR (%)	PPV (%)	ACC (%)
3б	33.3	76.92	23.08	40	63.16
2б	33.3	76.92	23.08	40	63.16
б	100	30.77	69.23	40	52.63
0.5б	100	30.77	69.23	40	52.63

Table 4.3. Sensitivity analysis for different S::CAN signals during period 1 according to hypothesis 2.

n=3, $\sum s\ (i) > 2$					
Threshold	TPR (%)	TNR (%)	FPR (%)	PPV (%)	ACC (%)
3б	66.67	81.25	18.75	40	78.95
2б	66.67	81.25	18.75	40	78.95
б	100	81.25	18.75	50	84.21
0.5б	100	62.5	37.5	33.33	68.42

4.3.2.2 Analysis of period 2

After April 30, 2017, several deviations have occurred in Turbidity signal (Figure 4.2): 02/05/2017; 03/05/2017; 05/05/2017; 09/05/2017. To select the adequate size of history window, four different cases have been evaluated: 5, 10, 15 and 20 days. For each value, the performance measures are calculated as well as the error and the correlation between the measured and the estimated parameter. Results, given in Table 4.4, indicate an improvement in the performance measures with the increasing of the training time. From a history window of 15 days, the rate of true positive reaches its maximum (85.71 %). However, the augmentation of the value till 20 days decreases the generation of false alarm, as well as the error with high correlation factor (0.9898). It should be noticed that data are firstly normalized. Normalization is done according to the maximum and minimum of each parameter, as shown in the following equation:

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin} \quad (4.9)$$

X_{norm}: normalized value.
X: input parameter.
X_{max}, X_{min}: maximum and minimum of each parameter.

The effect of normalization is showed in the last row of Table 4.4. It indicates lower value of TPR with higher value of FPR. This certifies the importance of normalization of data to get more accurate results.

Table 4.4. Sensitivity analysis for different history windows.

n=3, threshold=σ					
Training time (days)	TPR (%)	TNR (%)	FPR (%)	Error	Correlation
5	71.43	52.17	47.83	4.9947E-04	0.9863
10	57.14	47.83	52.17	3.9261E-04	0.9912
15	85.71	60.87	39.13	1.0334E-04	0.9893
20	85.71	65.22	34.78	9.3864E-05	0.9898
20 (without normalization)	71.43	52.17	47.83	6.828E-05	0.9881

The linear predictor model is tested for other S::CAN parameters separately such as UV, TOC, etc. Table 4.5 shows the performance measures calculated for each parameter. Results depend strongly on the nature of monitored parameter. For UV and Turbidity, TPR and FPR values are acceptable (respectively higher than 75 %, and lower than 45 %), while for TOC a high rate of true positive is combined with high generation of false alarm (85 %). The low precision obtained is explained by the large number of negative classes in comparison with the positive examples.

On the other hand, using hypothesis 2 (detailed previously), the application of the linear predictor model (with 20 days as training phase and σ as threshold) to S::CAN matrix (UV Turbidity EPA, Turbidity ISO, TOC, DOC, etc.) gives 83.33 % as TPR with the generation of 25 % as false alarm (FPR). However, the precision obtained is low (41.67 %).

Table 4.5. Sensitivity analysis for each of S::CAN parameters during period 2.

n=3 , threshold=6, Training time=20 days							
Parameter	TPR (%)	TNR (%)	FPR (%)	Error	Correlation	PPV (%)	ACC (%)
Turbidity EPA (NTU)	85.71	65.22	34.78	9.3864E-05	0.990	42.86	70
UV (Abs/m)	83.33	54.17	45.83	2.4373E-04	0.977	31.25	60
Turbidity ISO (FTU)	75.00	86.36	13.64	6.2840E-05	0.992	66.67	83.33
TOC (mg/l)	100	14.81	85.19	4.7701E-04	0.979	11.54	23.33

4.3.3 Discussion

At each time step, the linear predictor model compares water quality measurement of S::CAN to predicted value. If the difference is large, the corresponding data is identified as unacceptable. However, this method will allow to determine outlier but not real anomaly that lasts several minutes or hours. Also, the large number of negative examples leads to low precision, even at low FPR (and high TPR).

Another challenge in linear prediction is the determination of the predefined threshold. In addition, the identification of the final class group at a given time step is complicated. The determination of the number of signals that must exceed the threshold to identify a positive class (unexpected data) will be based on the user judgment. This can lead to the generation of high rate of false alarm.

4.4 Support Vector Machine (SVM) method

4.4.1 Classification Principle

The learning method Support Vector Machine (SVM) is generally used for data classification. It has several applications in different research and engineering domains such as medical diagnosis, marketing, biology, recognition of handwritten characters and human faces [81]. SVM method can be applied through different types: binary, multiclass, monoclass and regression. However, the simplest method is the binary classification usually applied to distinguish between positive and negative examples.

The main purpose of binary SVM is to divide data in two main classes: {+1; -1}. The objective is to define a hyper plan that separates the two classes. Although there is a multitude of valid hyper plan, the aim is to find the one that passes in the "middle" of the points of the two classes [116]. Therefore, the distance (namely "margin") between the hyper plan and the example should be maximized for better classification. Figure 4.6 illustrates a schematic representation of binary SVM. It shows the optimal hyper plan as well as the support vectors defined by the nearest points used for determining this hyper plan.

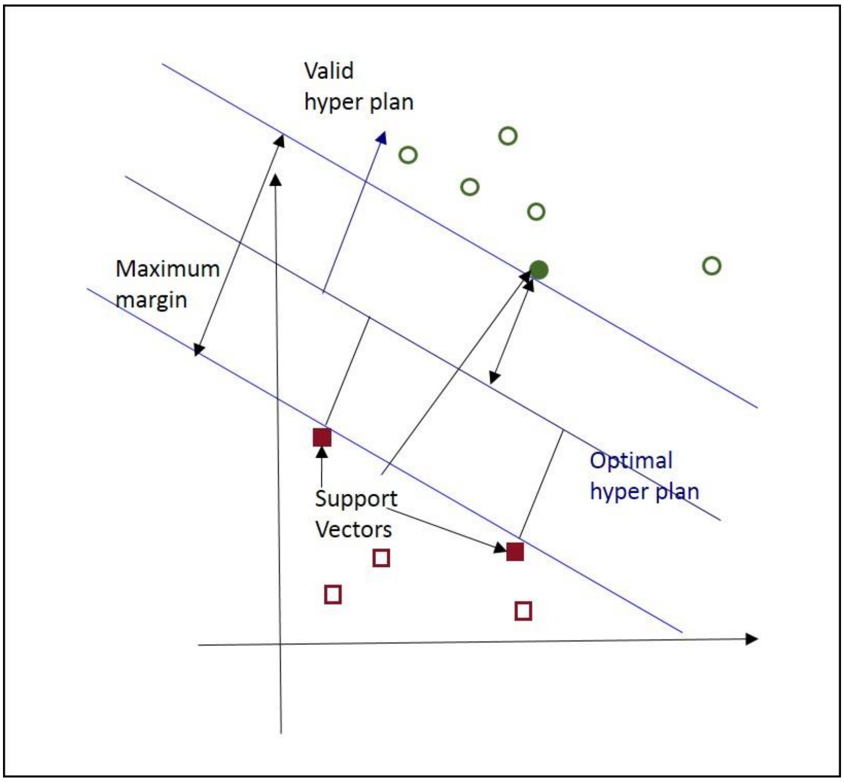


Figure 4.6. Binary SVM [116].

The hyper plan is described by the following equation [81]:

$$H(x) = w^T x + b \tag{4.10}$$

With: w: weight vector of dimension m.
b: term.
x: example to be classified (input).

Since the two classes are linearly separable, the decision function to be used for classification:

$$\begin{cases} Class = 1 \; if \; H(x) > \; +1 \\ Class = -1 \; if \; H(x) < -1 \end{cases} \tag{4.11}$$

To find the hyper plan that maximizes the margin, the quadratic optimization problem can be transformed to a dual problem using Lagrange multipliers α_i [117], [118]:

$$Maximize \; L\,(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \, \alpha_j \, y_i \, y_j \, (x_i, x_j) \tag{4.12}$$

$$with \; \sum_{i=1}^{n} \alpha_i \, y_i = 0 \text{ , } 0 \leq \alpha_i \leq C, \text{ n: number of observations, y } \epsilon \; \{+1; -1\}$$

C indicates the error of the classification. The Support Vectors are defined as [82]:

$$SV = \{xi \text{ such as } \alpha_i > 0\} \tag{4.13}$$

The decision function is then calculated as follows [117], [119]:

$$f\,(x) = \sum \alpha_i \, y_i \, K(x_i, x) + b \text{ with K: Kernel function} \tag{4.14}$$

Figure 4.7 illustrates an example of the process used in SVM method. X is the input vector to be classified. According to the sign of the decision function f, the output of the method will be either the class 1 or -1.

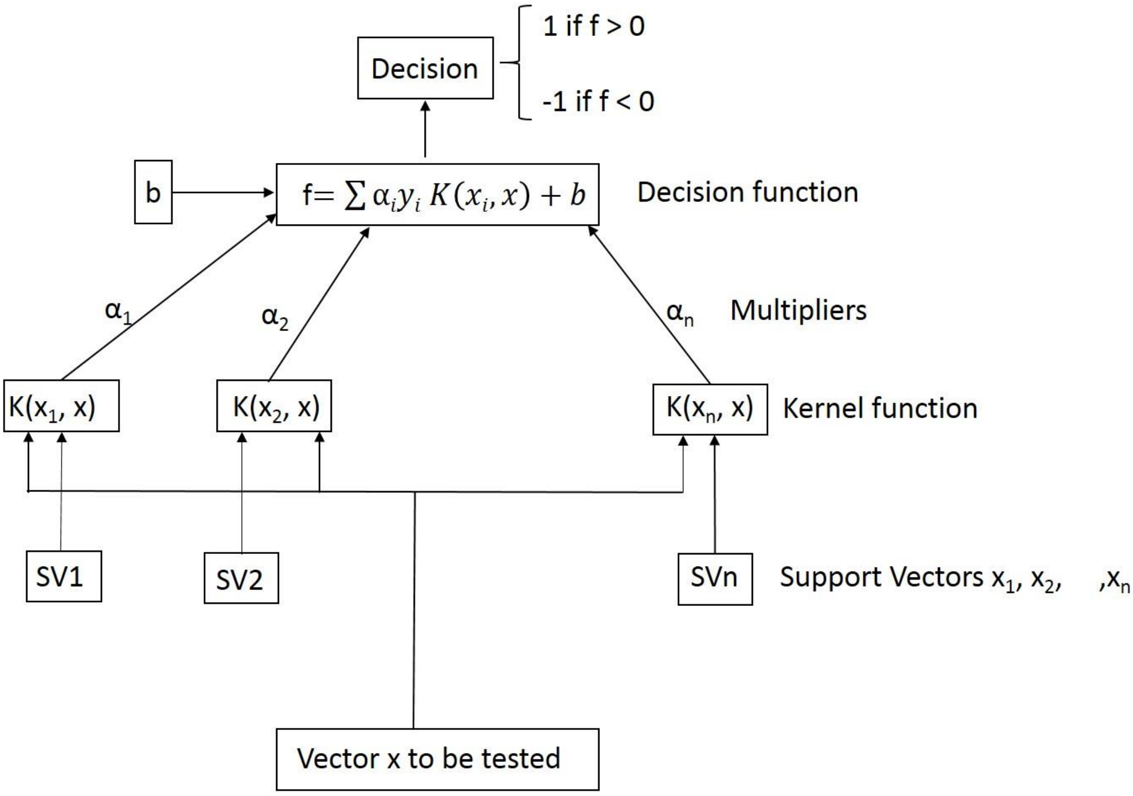


Figure 4.7. Architecture of Support Vector Machine (according to [81]).

SVM method requires a training phase to find the appropriate properties of the optimal separator. During this phase, a set of predictor data (numeric matrix where rows are the observations and columns are the features) with their known classes labels (columns vectors) should be used. The analysis of this data will define the hyper plan that will be used later to classify each new example. After training phase, a corresponding class is assigned to each new data using information from the trained classifier.

4.4.2 Use of SVM method for the detection of water anomaly

Since binary SVM classifies data into main classes, the method can be applied for the detection of water anomaly. The main objective is to analyze water quality parameters in order to test the potability of water. In this chapter, the two main classes of SVM method are defined as follows: i) Class +1 which indicates unsafe drinking water due to anomaly in quality and ii) Class -1 which indicates normal drinking water.

The methodology applied can be summarized in five main steps:

- Definition of the input data: i) matrix of S::CAN data (columns contain water quality parameters and rows include measured values at each time step), and ii) Y: class label.
  The identification of Y is divided to two parts. Firstly, we compare each observation with a predefined threshold. If the value is bigger than the threshold so its corresponding class is identified as 1, otherwise the class is -1. After this classification, we obtain a matrix of classes {+1; -1}. Then, for each row of the matrix, if one parameter belongs to the class 1 so Y is equal to 1, otherwise Y is -1.

- Normalization of data: all input parameters should be normalized before classification, according to their maximum and minimum.
- Training phase: a set of historical data with known class label are used (in *svmtrain* function of Matlab software) to determine different properties of an appropriate classifier.
- Test phase: new input should be tested (with *svmclassify* function of Matlab software) using the classifier obtained from training phase.
- Calculation of the method accuracy: comparison between the predefined class labels (for test data) and the SVM output.

4.4.3 Threshold definition

As detailed previously, the identification of the class label Y depends on a predefined threshold. In order to test different applications of the method, four thresholds have been evaluated:

- Case 1: threshold for each parameter is fixed as the corresponding Standard limit (described in Table 4.6).

Table 4.6. Standard limits for S::CAN parameters.

Turbidity EPA (NTU)	Turbidity ISO (FTU)	TOC (mg/l)	Conductivity (µS/cm)	Temperature (°C)	pH
1	1	2	1000	15	8.5

- Case 2: Each measurement is considered out of limit when it exceeds the normal variation defined by $\bar{X} \pm 3\sigma$ [114], with $\bar{X}$ the average and σ the standard deviation for each parameter.
- Case 3: In order to reduce the acceptable limits, each parameter is compared with $\bar{X} \pm 2\sigma$.
- Case 4: the lowest predefined threshold in this study is taken $\bar{X} \pm \sigma$.

4.4.4 Application

SVM method has been tested to analyze and classify S::CAN data (Turbidity EPA, Turbidity ISO, TOC, Conductivity, Temperature and pH) at Barroi for the period between April 11, 2017 and June 11, 2017. The corresponding variation of water quality parameters is illustrated in Figure 4.8. Some deviations of signals were observed during this period.

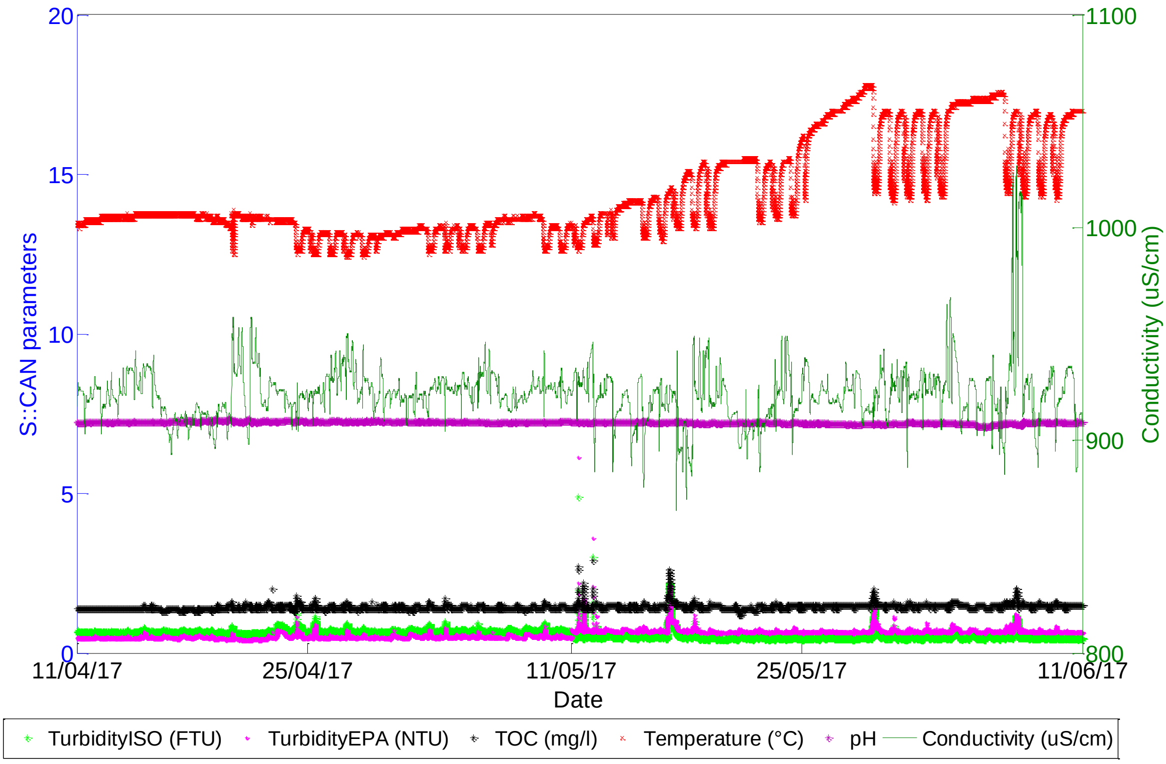


Figure 4.8. Variation of S::CAN data between April 11, 2017 and June 11, 2017 at Barroi.

4.4.4.1 Selection of predefined threshold

To choose the adequate threshold, the accuracy should be compared for the different cases (detailed in section 4.4.3). To make the comparison, the history window used in training phase should be fixed. The set of data trained should contain the two classes {+1; -1} in order to define classifier's properties. To achieve this objective, different trials were tested in each case and the results of the minimum window size needed is given in Table 4.7. The number of iterations used to solve the quadratic optimization problem is taken 15000 by default. Results show that a history window of 14 days is required to make the comparison between the thresholds.

Table 4.7. History window for the different cases.

Case	History window for training (days)
1	14
2	10
3	6
4	2

For a training phase of 14 days, a test is conducted for the rest of data till June 11, 2017. In each case, the comparison between the predefined classes and the SVM classification allows to calculate the corresponding accuracy. The histogram of Figure 4.9 illustrates the results of accuracy. It should be noticed that, for Case 3 and 4, the number of iterations should be increased to 60000 to ensure the convergence. The accuracy obtained from Case 3 is very low (14.29%) while it is acceptable for Case 4 (50.75 %) and Case 1 (60.25 %). The best accuracy is obtained for Case 2 (96.39%). This comparison proves that a predefined threshold of ($\bar{X} \pm 3\sigma$) should be selected to ensure the best performance of the SVM classifier.

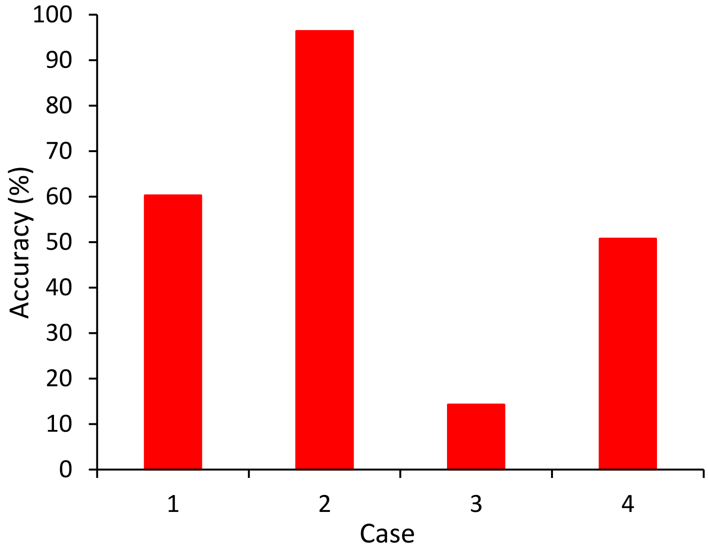


Figure 4.9. Accuracy for different cases of predefined thresholds.

4.4.4.2 Selection of history window

As the threshold is selected to be $\bar{X} \pm 3\sigma$, the history window can take a minimum value of 10 days (Table 4.7). To test the impact of the history window size used for training, different cases are studied: 10, 15 and 20 days. For each case, two sizes of data are used for the test phase: 5 and 15 days. The results of the accuracy is illustrated in Table 4.8. It indicates an improvement of the SVM performance with the increase of the history size. However, to reduce the computational time, a training phase of 15 days will be selected.

Figure 4.10 gives an example of the SVM classification tested between May 26, 2017 and June 11, 2017, using a training phase of 15 days. It indicates close results between the predefined classes and SVM results for the majority of time steps. However, some unexpected data (for example on June, 4 and June, 5) are not reported by SVM classifier.

Table 4.8. Accuracy for different training and test phases.

Test (days)	Training (days)	Accuracy of Test (%)
5	10	86,41
	15	86,9
	20	87
15	10	91,64
	15	92,03
	20	92,23

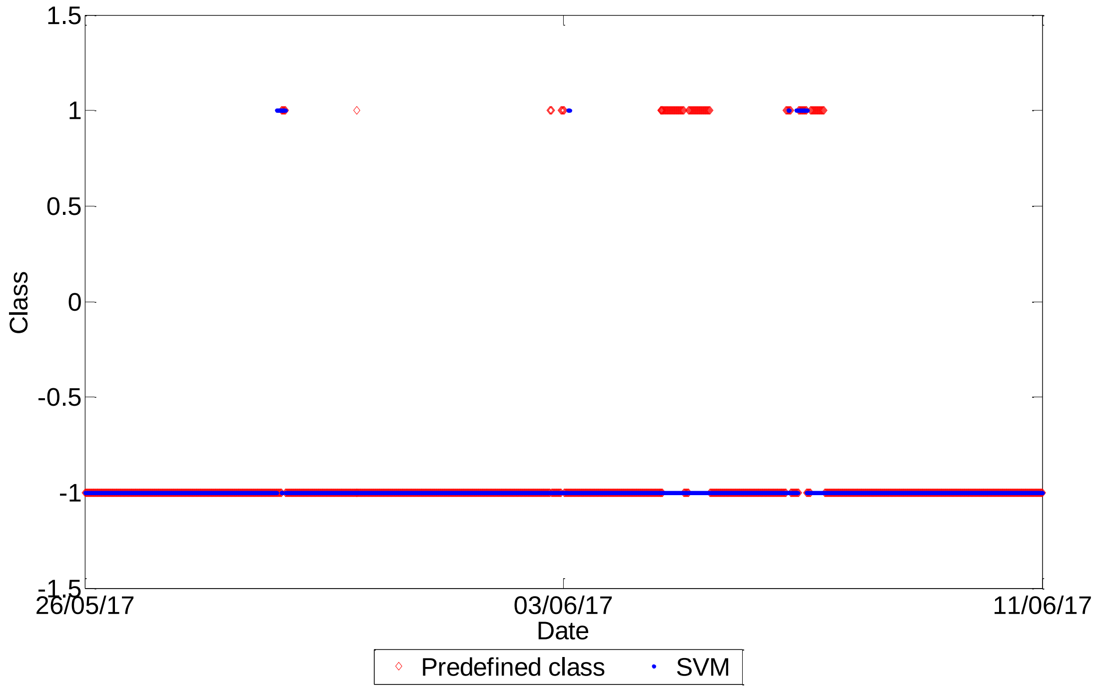


Figure 4.10. SVM classification for a training phase of 15 days.

4.4.5 Discussion

The use of SVM method within the online monitoring of water quality allowed a discrimination of S::CAN data in two classes at each time step: i) accepted values indicating normal drinking water and ii) unexpected data implying unacceptable water quality.

As for linear prediction, SVM requires the identification of predefined threshold to classify data. In addition, the method faces limitations concerning problems in the convergence (number of iterations) and training phase which should contain examples from the two classes. On the other hand, the methodology used within SVM is based on the hypothesis that a deviation in one parameter induces the identification of "anomaly" class for all the input parameters.

Therefore, the application of SVM method can be efficient as learning method to distinguish between normal and unexpected values at each time step. However, the detection of real abnormality through SVM remains difficult. An early detection system is required to identify the occurrence of true anomaly in water.

4.5 Event Detection System (EDS) - Canary

4.5.1 Role of Event Detection System (EDS)

The two previous methods (lpc and SVM) have been used to distinguish between normal and unacceptable data at each time step. Results identify unexpected values without distinction between outliers and true anomaly. This indicates the need of an Event Detection System (EDS) that reports significant abnormalities that could involve change in water composition.

Contamination Warning Systems (CWSs) have been proposed as a promising approach for the early detection and management of contamination incidents in drinking water distribution systems [120]. The main purpose of CWS is to protect the public health by reducing the time between contamination detection and an effective decision-making. A robust CWS includes different components: Online monitoring, customer complaint monitoring programs, routine sampling and laboratory analysis. The main part on a CWS is the online monitoring of water quality based on the deployment of sensors through the distribution networks. Recently, surrogates sensors are more used than direct sensors that detect specific contaminants. Surrogates sensors measures continuously indicators water quality parameters (Turbidity, pH, Conductivity, etc.). The deviation of such parameters from the stable lines could indicate the potential presence of contaminants. Therefore, EDS is needed to differ between periods of normal and anomalous water quality variability from measures made with surrogate sensors [121]. An EDS is an automated system that ensures an interface for analyzing real-time data and detecting unexpected variations.

In general, water utilities detect the change in water quality parameters by comparing them with thresholds. This method will trigger an alarm for each event exceeding of the set point value. However, some contamination incidents might not cause water quality parameters to move outside of set points boundaries [121]. In such cases, an EDS is relevant. It will detect all variation in water quality data that differ significantly from the background values whether or not they exceed the set point limits. Another advantage of the EDS is the complexity to detect variation in water quality signals for different parameters monitored simultaneously by using only the set point approach. This is demonstrated in Figure 4.11. Only Signal 1 indicates an exceeding of the upper threshold limit. However, the relative deviation is Signal 2 and Signal 3 is only identified if an EDS is used.

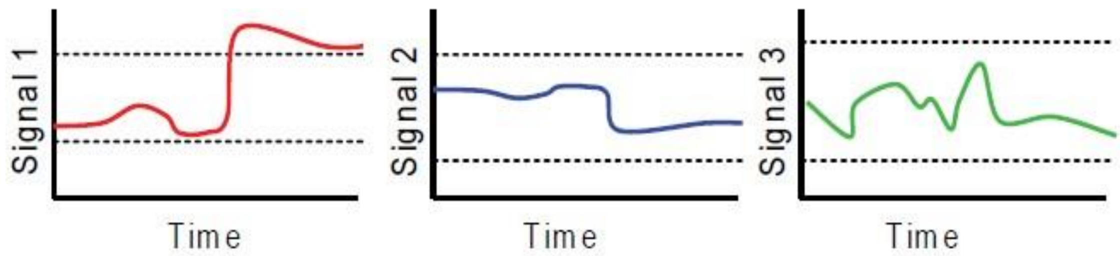


Figure 4.11. Schematic diagram of changes in three different water quality signals over time (The dashed lines represent the set points of signals) [121].

4.5.2 Detection methodology

The open source CANARY EDS software has been developed by Sandia National Laboratories in collaboration with EPA's National Homeland Security Research Center (NHSRC). The EDS developed within Canary increases the likelihood and speed of anomaly detection [122]. The main goal of Canary is to analyze data from a Supervisory Control and Data Acquisition (SCADA) system in near real-time. Therefore, it will evaluate the probability of a water quality event. It uses statistical and mathematical algorithms to identify the onset of periods of anomalous water quality data, while at the same time limiting the generation of false alarms [123]. It should be noted that Canary works in both offline (historical data but simulated online by processing in time series order one step at a time) and online (via SCADA connection) modes.

The main purpose of Canary is to detect automatically significant deviations from expected water quality data collected from sensors at each time step. In the basic mode of operation, the event detection algorithms deployed in Canary software involve four consecutive steps [124]: i) Estimation of the future water quality, ii) Comparison and Residual calculation, iii) Residual classification and vi) Probability calculation. Figure 4.12 illustrates the process used for event detection.

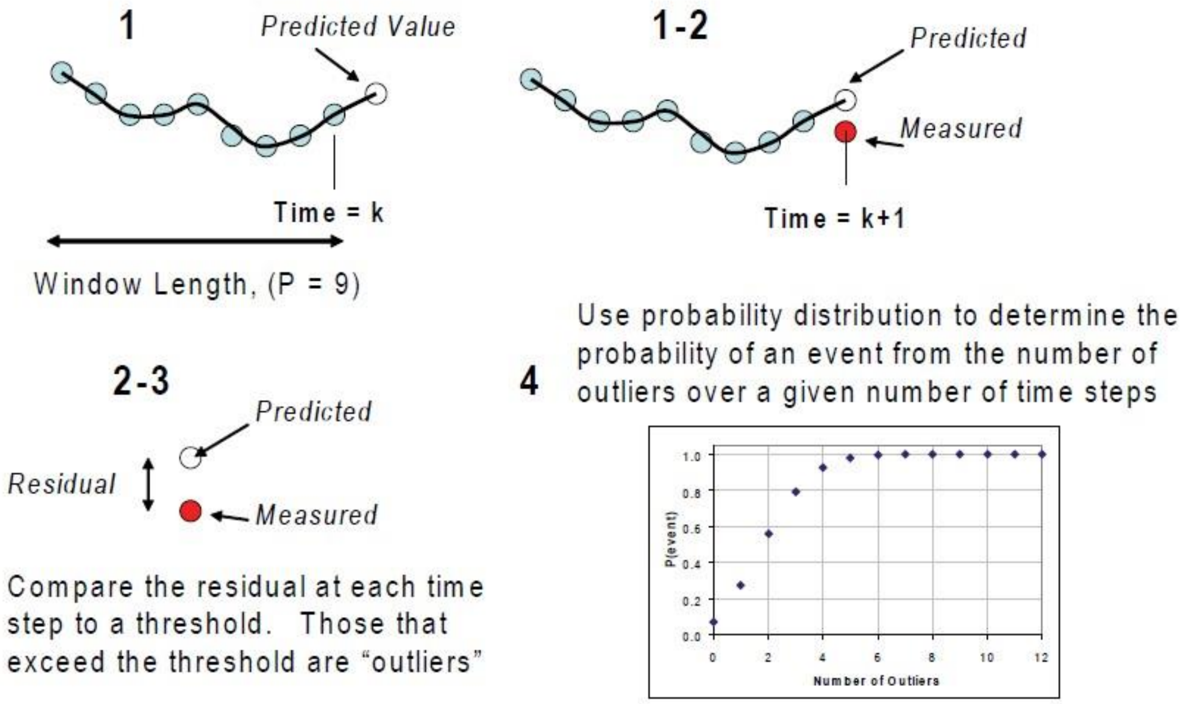


Figure 4.12. Steps in the Canary event detection process: 1) Estimation, 2) Comparison, 3) residual classification, 4) probability calculation [121].

4.5.2.1 Estimation

For each time series of data for every single parameter, Canary predicts the expected data in the next time step. It looks backward in a user predefined moving window of previous time steps and uses data in this window to make the estimation. Data are firstly normalized with zero as mean and one as standard deviation. The normalization removes the units of measurements so signals can be easily combined later. For the estimation, two approaches can be used within Canary [124]:

- Linear filtering (LPCF): An auto-covariance function, computed independently for every signal, is used to calculate a set of weights. At each time step, an optimal set of weights is applied to each of the previously measured standardized observations. The calculated weights indicate the importance of previous data in the prediction of the next value no matter how far in the past that value has occurred. Weights, calculated automatically within Canary, are updated at each time step. The weighted average of the predefined set of previous values serves as the estimation of the water quality value at the next time step.
- Multivariate Nearest Neighbor (MVNN): At each time step, the set of values provided from n different water quality signals can be considered as a point in n-dimensional space. All data in previous time steps can be grouped as points and distance between them can be calculated. At each new time step, a new point in n-dimensional space is created and its 'nearest neighbor' or the closet point in the set of previous values serves as the estimated value for this time step.

4.5.2.2 Comparison and residual calculation

At the current time step, when the observation is available in the SCADA system, it is normalized and compared with the predicted value. A residual value is then calculated as the difference between the measured and the estimated value. For each water quality signal, a residual value is expressed in unit of standard deviations.

4.5.2.3 Residual Classification

Across all water quality signals, the maximum residual is determined. It is compared with a user-defined threshold, given in unit of standard deviations. If the maximum exceeds the threshold, the water quality at this time step is classified as 'outliers' and is not be used for the estimation of the future value.

4.5.2.4 Probability calculation

At the current time step, the probability of a water quality event should be analyzed in function of the number of outliers. A Binomial Event Discriminator (BED) was developed for CANARY to create a time-integrated probability of an event *P (event)* [124]. The BED employs the properties of the binomial distribution to define *P(event)* as a function of the number of outliers within the BED integration window, the length of the BED integration window, and the probability of an outlier occurring at any given time step [121].

4.5.3 Configuration file

Canary operates using a configuration file written in YML markup language. The configuration file describes all details concerning data source (Comma-Separated Value (CSV) file or link to SCADA system), time period and interval, water quality signals, and algorithms to be used in the EDS.

The main purpose is to configure the EDS in a way to decrease the number of missed detection, with the minimum number of false alarms [125]. In addition to the type of the Event detection algorithm used in the estimation of the future values, four associated parameters define the performance of the EDS: i) History window, ii) Outlier threshold, iii) Event threshold and iv) BED window.

4.5.3.1 History window

The history window is the number of previous data (in units of time steps) used to predict the value of the water quality signal at the next time step. It is a fixed length applied equally to all signal but it moves forward in time. The selection of the history window size will influence the accuracy of the EDS. A window size of 1.5 to 2.0 days has proven to be the most accurate; smaller or larger values result in decreased accuracy [126].

4.5.3.2 Outlier threshold

The outlier threshold defines the limit, expressed in unit of standard deviations that should be met or exceeded to identify an outlier. In general, outlier threshold will be near 1. Increasing its value will reduce the number of time steps classified as outliers and, thus, the number of events, which makes the event detection algorithm less sensitive in terms of detecting significant changes in the water quality signal [126].

4.5.3.3 Event threshold

The event threshold determines the maximum probability of an event that should be exceeded before a group of outliers generates an alarm. As the threshold is increased, Canary is less sensitive, detecting fewer events [126]. The event threshold should be selected in a way to increase the sensitivity of the EDS.

4.5.3.4 BED window

In general, water quality data has significant, short-lived spikes in the values that last for a few time steps and could be due to SCADA communication failures or sensor malfunction [121]. To ignore these spikes, the BED function within Canary aggregates evidence of an event (i.e., outliers) over multiple consecutive time steps before identifying an event [126]. In the binomial model, the number of outliers represents NFAILURES within a number of time steps indicating NTRIALS-window (BED window). The BED window specifies the size of the window, in unit of time steps, for the binomial distribution. A short BED window allows to determine an event faster because fewer outliers are sufficient to define an event. However, false positives alarms (induced by noisy or bad data) can be generated in this case.

The aim is to make a tradeoff between the faster determination of event and the generation of false alarm. To provide the high detection of events, all these parameters should be adjusted in a training phase done for historical data.

4.5.4 Sensitivity Analysis at Polytech’Lille

4.5.4.1 Case study

S::CAN sensor, installed at Polytech’Lille, is selected to determine the appropriate parameters that adjust the configuration file. Data, used for this training phase, are collected between August 1, 2016 and October 10, 2016. Ten parameters, measured by S::CAN, are evaluated within Canary: UV, Turbidity ISO, Turbidity EPA, TOC, Color, Temperature1 (from i::scan), Conductivity, Temperature2 (from Condu::lyser), pH and Free Chlorine. The variation of some of these parameters is given in Figure 4.13. It indicates different deviations, especially in Turbidity signal. During this period, 29 anomalies have been detected. The date of each abnormality is listed in Table 4.9. Since S::CAN measures water quality parameters each one minute, the time step for running Canary is taken the same (1 minute) as the sensor.

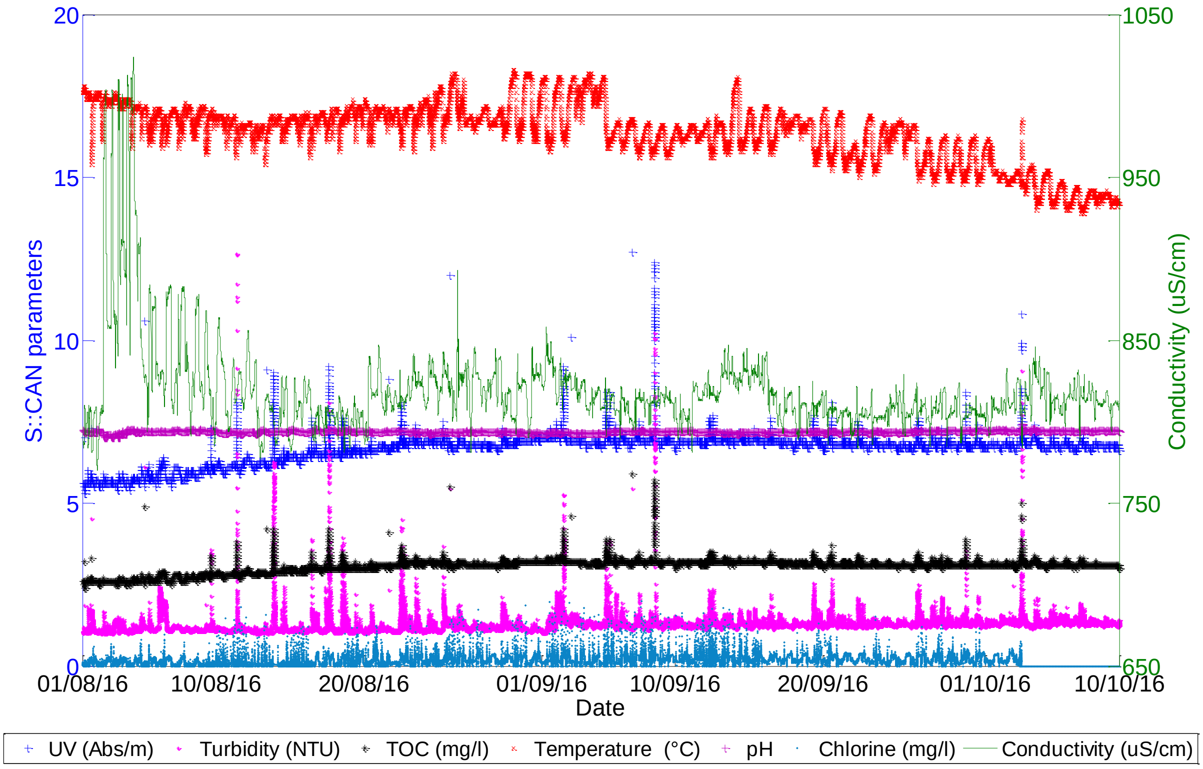


Figure 4.13. Variation of S::CAN data between August 1, 2016 and October 10, 2016 at Polytech'Lille.

Table 4.9. List of abnormalities. (a) August, 2016; (b) September, 2016; (c) October, 2016.

Anomaly on August, 2016
02/08/2016
04/08/2016
06/08/2016
09/08/2016
11/08/2016
13/08/2016
16/08/2016
17/08/2016
18/08/2016
20/08/2016
22/08/2016
23/08/2016
25/08/2016
29/08/2016

Anomaly on September, 2016
02/09/2016
05/09/2016
08/09/2016
12/09/2016
16/09/2016
19/09/2016
20/09/2016
22/09/2016
26/09/2016
29/09/2016

Anomaly on October, 2016
03/10/2016
04/10/2016
05/10/2016
06/10/2016
07/10/2016

4.5.4.2 Parametric study

A parametric study is conducted to select adequate values for the five parameters cited above (Event detection algorithm, History window, Outlier threshold, Event threshold and BED window). The confusion matrix and the performance measures (described in section 4.2) will be used for this analysis.

- **Selection of the event detection algorithm**

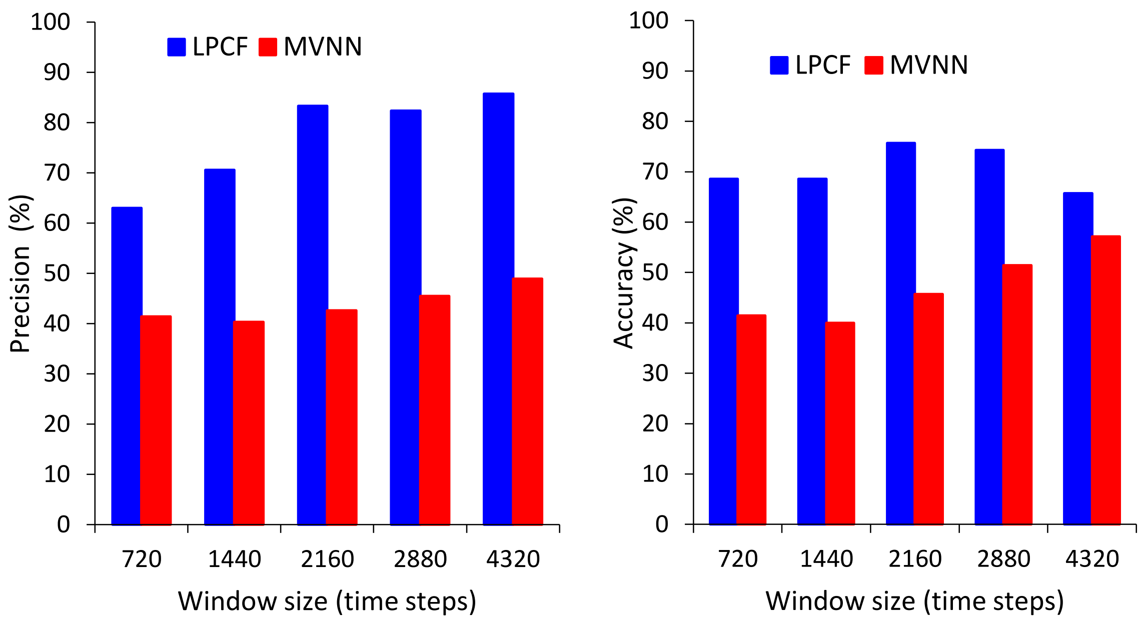
In order to select the appropriate algorithm type (LPCF or MVNN) for S::CAN signals, we compare the precision as well as the accuracy for each type using different history window values (cited in Table 4.10), while other parameters are fixed as follows: i) Outlier threshold: 0.85, ii) Event threshold: 0.99 and iii) BED window: 200 time steps.

Table 4.10. List of history windows.

History window (time steps)	720	1440	2160	2880	4320
Duration (days)	0.5	1	1.5	2	3

Figure 4.14 gives the results of the calculation of both the precision and the accuracy. Figure 4.14 (a) indicates a large difference in the precision of LPCF and MVNN algorithms. The precision obtained with LPCF varies between 62% and 85 %, while for MVNN, the precision is lower (do not exceed 49 %). This shows that LPCF is more precise. In the same way, Figure 4.14 (b) gives the accuracy comparison between the two algorithms. The accuracy obtained from MVNN type is low (maximum value 57 %). However, LPCF accuracy is significantly better (can reach 75 %).

This comparison proves that a highest precision and accuracy are obtained with LPCF. From this result, we select "LPCF" as an appropriate type of algorithm to be used in the configuration file of S::CAN signal analysis.



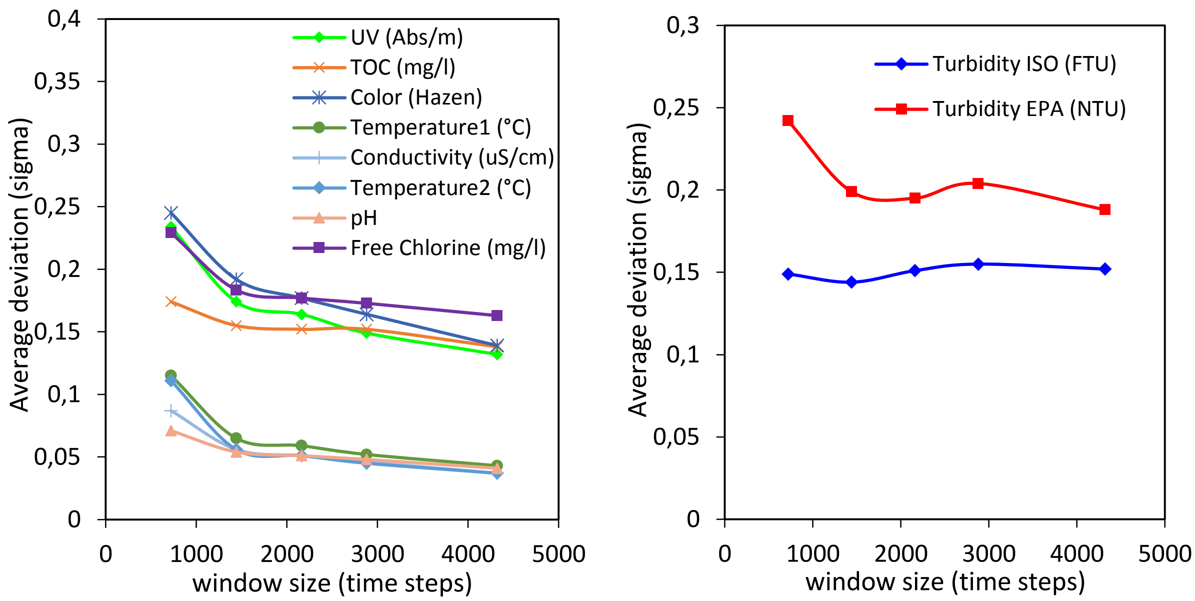
(a) (b)

Figure 4.14. Comparison of detection algorithm. (a) Precision; (b) Accuracy.

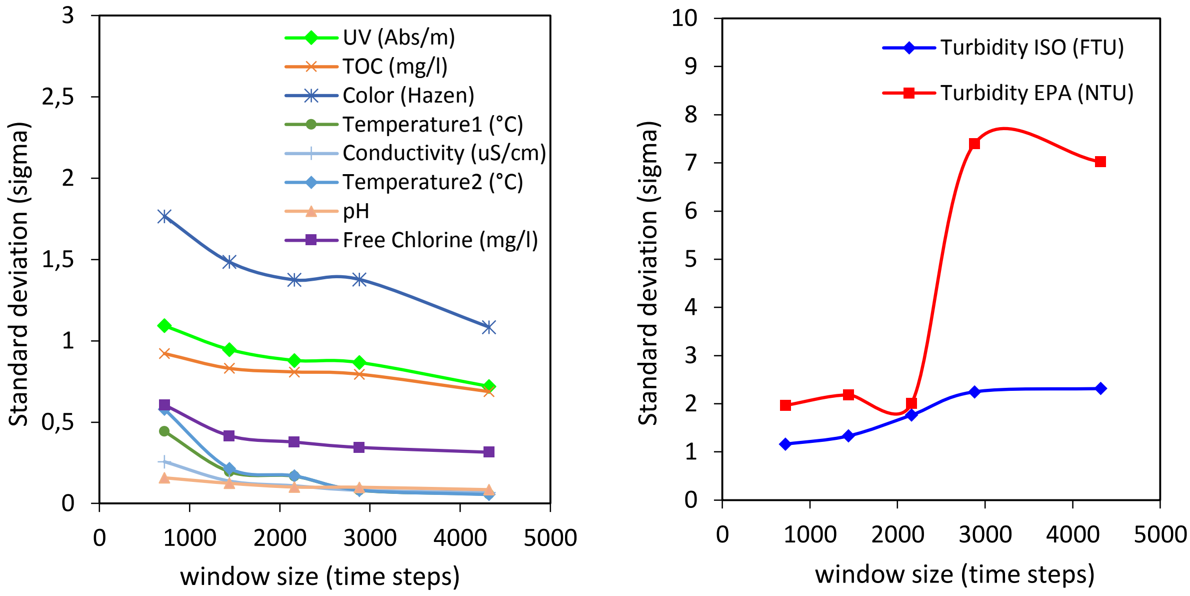
- **Determination of history window**

The selection of the appropriate window size (history window) is the basic step required for configuration. Using a training data set, performance measures should determine the accuracy of water quality prediction made by Canary. The quality of the estimation is evaluated in function of the average absolute value of the residuals (difference between measured and predicted value) and the corresponding standard deviations of residuals.

For the selected LPCF algorithm, these two performance measures (average deviation and standard deviation) are calculated for the list of history window of Table 4.10. However, outlier threshold, event threshold and BED window are fixed as indicated in the previous section. The results of calculation are given in Figure 4.15. The main objective is to find the history window that increases the precision of the estimation i.e. that induces the lowest average residual and standard deviation. Figure 4.15 shows that both the absolute residual (Figure 4.15 (a)) and the standard deviation (Figure 4.15 (b)) decrease as the window size increases. However, it is important to find a tradeoff between reducing the average deviation and standard deviation and the long computational time induced by large window size. Figure 4.15 indicates that the average residual and standard deviation reach approximately a minimum and start to deviate from a window size of 2160 time steps (1.5 days). From this value of history window, the average deviation ranges between 0.05 and 0.17, while the standard deviation remains between 0.1 and 1.37 for all monitored parameters (except Turbidity data that have few larger values). The exception observed in Turbidity, after 1.5 days, can be induced by some fluctuations in the signal during the training phase. Despite this exception, the choice of 1.5 days as history window provides useful and precise estimations of the future water quality parameters.



(a)



(b)

Figure 4.15. Comparison of history window for LPCF algorithm. (a) Average deviation; (b) Standard deviation.

- **Selection of event threshold**

It is obvious that a significant deviation from the expected value at a single time step should not trigger an alert. The generation of alarm depends on both the BED window and the event threshold values. These two parameters represent: i) the time period in which to look for the onset of an event and ii) the number of outliers that need to occur within that time period to indicate an event [122].

The selection of the appropriate event threshold determines the probability that should be exceeded to consider a group of outliers as an event and to trigger an alarm. A parametric study is conducted using four values of event threshold: 0.75; 0.85; 0.9 and 0.99. An outlier threshold of 0.85 and a BED window of 200 time steps are fixed in this section, while the use of LPCF algorithm and a history window of 1.5 days is verified previously. For each case, performance measures are calculated and illustrated in Table 4.11. It indicates very close sensitivity results for the different event thresholds. A value of 0.99 as event threshold will be selected. It gives high precision and accuracy with small FPR.

Table 4.11. Sensitivity analysis for different event thresholds.

Event threshold	TPR (%)	TNR (%)	FPR (%)	PPV (%)	ACC (%)
0.99	51.72	92.68	7.32	83.33	75.71
0.90	48.28	92.68	7.32	82.35	74.29
0.85	48.28	92.68	7.32	82.35	74.29
0.75	51.72	92.68	7.32	83.33	75.71

- **Parametric study for outlier threshold and BED window**

At each time step, a signal point is labelled as outlier if the difference between the measured value and the estimated value (i.e. the residual) exceeds the outlier threshold times the standard deviation. Very low values of the outlier threshold result in more outliers, because less of the data is treated as good relative to the predicted behavior while very high values are less sensitive since almost all data might be considered good and real events might not be detected [122].

Once the outlier threshold is determined, an adequate BED window should be selected to define the size of historical data window used for testing the presence of an event. The BED window should be greater than one to avoid that a single outlier generates an alarm.

The selection of these two parameters should be done in a way to ensure the rapid detection of event while reducing false positive alarms. Firstly, the average prediction residual is calculated using 25 different combinations between the outlier threshold (0.6; 0.65; 0.75; 0.85; 1.1) and the BED window (50; 100; 200; 250; 300) time steps. Other parameters are selected previously as follows: i) Event detection algorithm: LPCF, ii) History window: 1.5 days and iii) Event threshold: 0.99. Figure 4.16 illustrates the results of calculation. It indicates that a minimum value of the average prediction residual (0.108) is obtained from the combination of 50 time steps as BED window and 0.75 as outlier threshold.

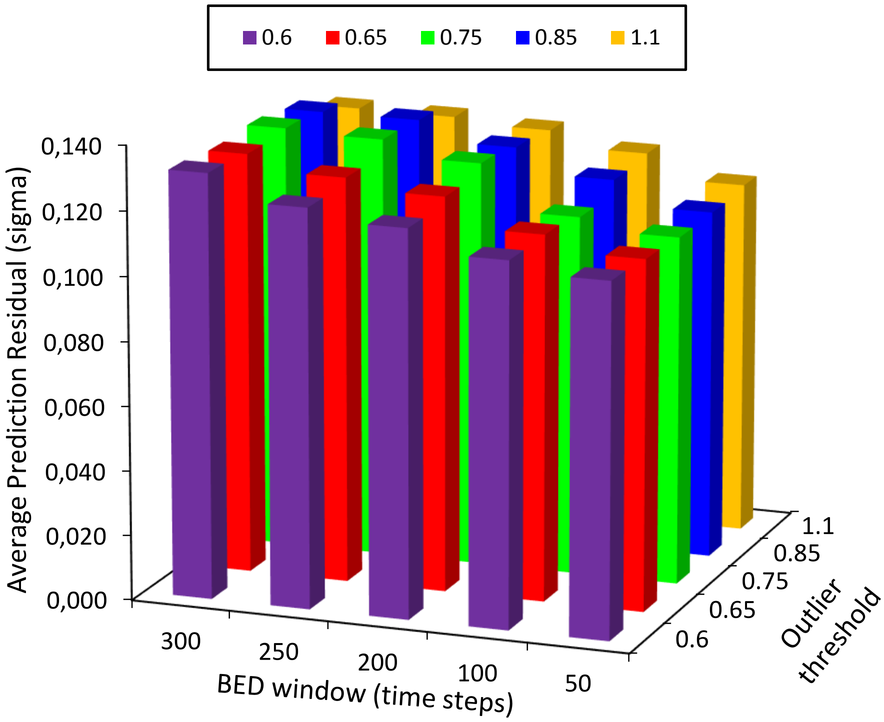


Figure 4.16. Average prediction residual in function of outlier threshold and BED window.

To validate the choice of these two parameters, a sensitivity analysis is conducted using Receiver Operating Characteristics (ROC) curve analysis. A ROC curve is a good way of visualizing a classifier's performance to select a suitable operating point, or decision threshold [127]. ROC curve graph is a two-dimensional graph in which TPR (Sensitivity) is plotted in function of FPR (1-Specificity). The main purpose of this graph is to make a tradeoff between high sensitivity and low specificity. The aim is to make a correct decision with the minimum generation of false alarms. This can be illustrated in the point (0, 1) in the graph where the sensitivity and the specificity reach 100 % (no false negatives nor false positives). More the result is close to this point (0, 1), the detection algorithm has a better performance.

Firstly, a ROC curve analysis is done using an outlier threshold of 0.75 and different value of BED window. Results are illustrated in Figure 4.17. A BED window of 100 time steps is the most appropriate to increase the probability of detection (TPR) with low value of false alarm (FPR).

In order to select adequately the outlier threshold value, a ROC analysis is also conducted for both critical cases of BED window (50 and 100 time steps) and for different outlier thresholds. Figure 4.18 gives the results of this analysis (values on the graph indicates the corresponding outlier threshold for each point). It indicates that a high sensitivity (with acceptable rate of false alarm) is obtained for 0.75 as outlier threshold with 100 time steps as BED window.

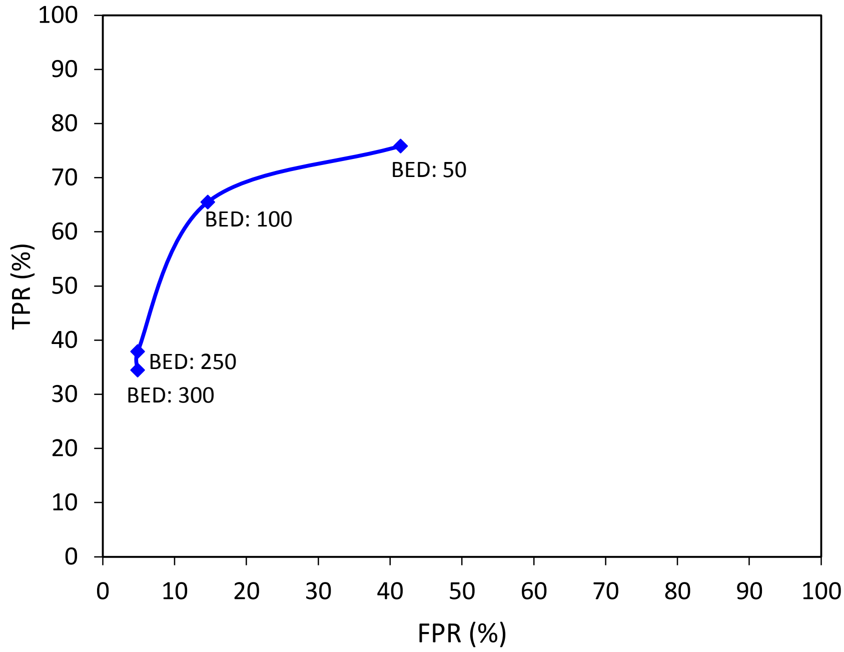


Figure 4.17. ROC curve for different BED window.

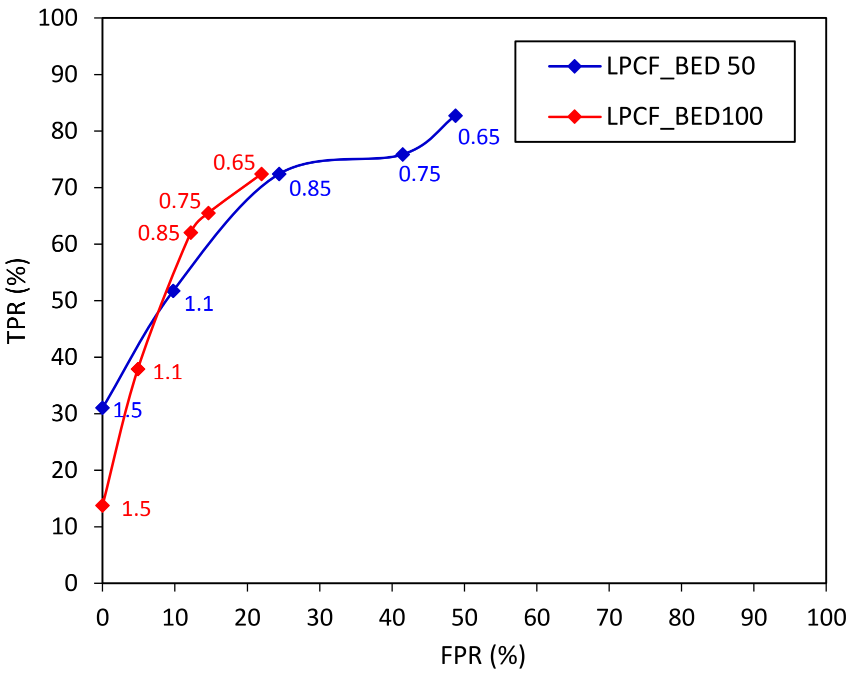


Figure 4.18. ROC curve for multiple outlier thresholds for a BED window of 50 and 100 time steps.

4.5.4.3 Application of selected parameters

The parametric study allows to determine the best appropriate parameters (as indicated in Table 4.12) for the EDS at Polytech'Lille. To test the ability of Canary to detect earlier water quality anomaly, the adjusted configuration file is applied to S::CAN data from October 10, 2016 till November 1, 2016.

Figure 4.19 illustrates the variation of several S::CAN parameters during this period. Ten significant deviations are observed: 10/10/2016; 11/10/2016; 12/10/2016; 14/10/2016; 17/10/2016; 18/10/2016; 19/10/2016; 20/10/2016; 24/10/2016 and 31/10/2016. The application of the event detection algorithm within Canary, using the adequate parameters (Table 4.12), allows the detection of 7 anomalies with 30 % as false alarm rate. Figure 4.20 illustrates an example of results. It shows the probability of event calculated between October 17, 2016 and October 24, 2016.

On the other hand, a strong deviation is observed on October 14, 2016 (Figure 4.19). Figure 4.21 illustrates the results given within Canary for the week of October 10, 2016 (blue points indicate deviation from expected values). The large variation of signals on October 14, 2016 is well detected with a probability of event equal to 1.

Table 4.12. Parameters selected to the EDS at Polytech.

Type of detection algorithm	LPCF
History window	1.5 days (2160 time steps)
Event threshold	0.99
Outlier threshold	0.75
BED window	100 time steps

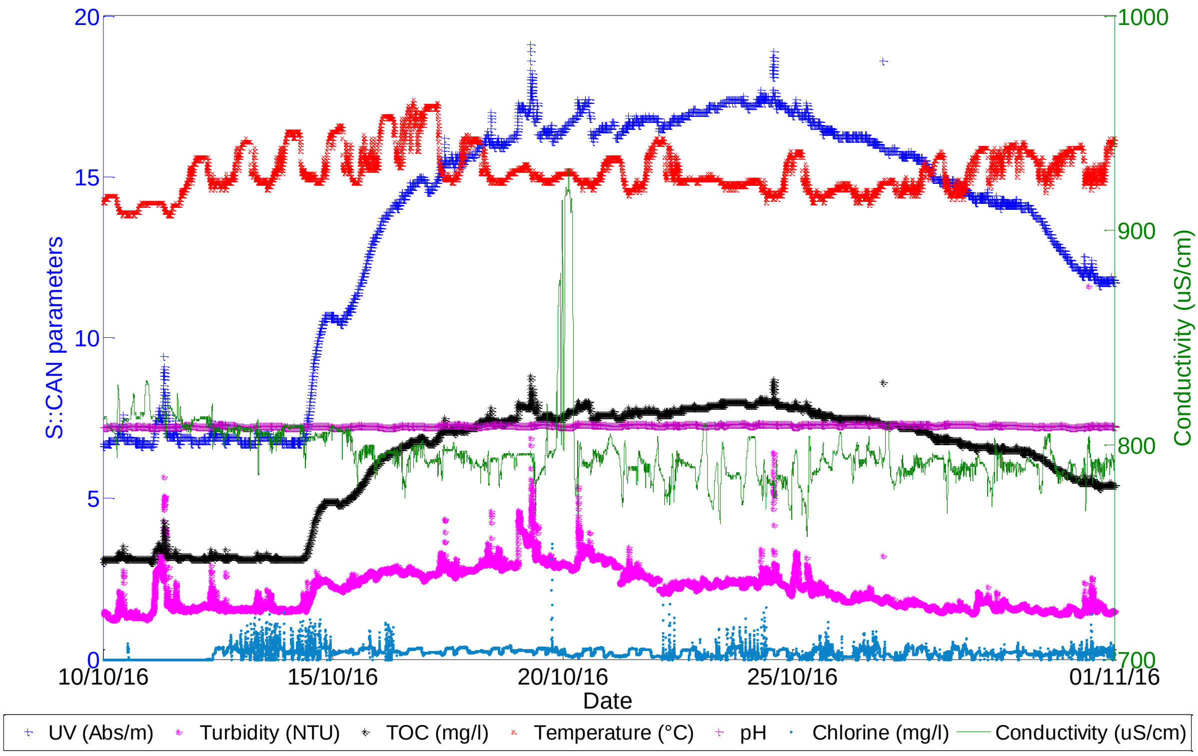


Figure 4.19. S::CAN data between October 10, 2016 and November 1, 2016 at Polytech'Lille.

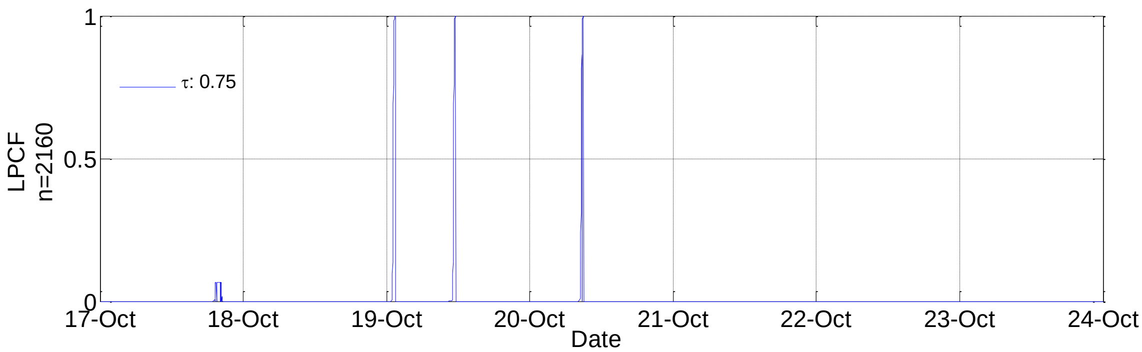


Figure 4.20. Probability of event for the week of October 17, 2016 at Polytech’Lille.

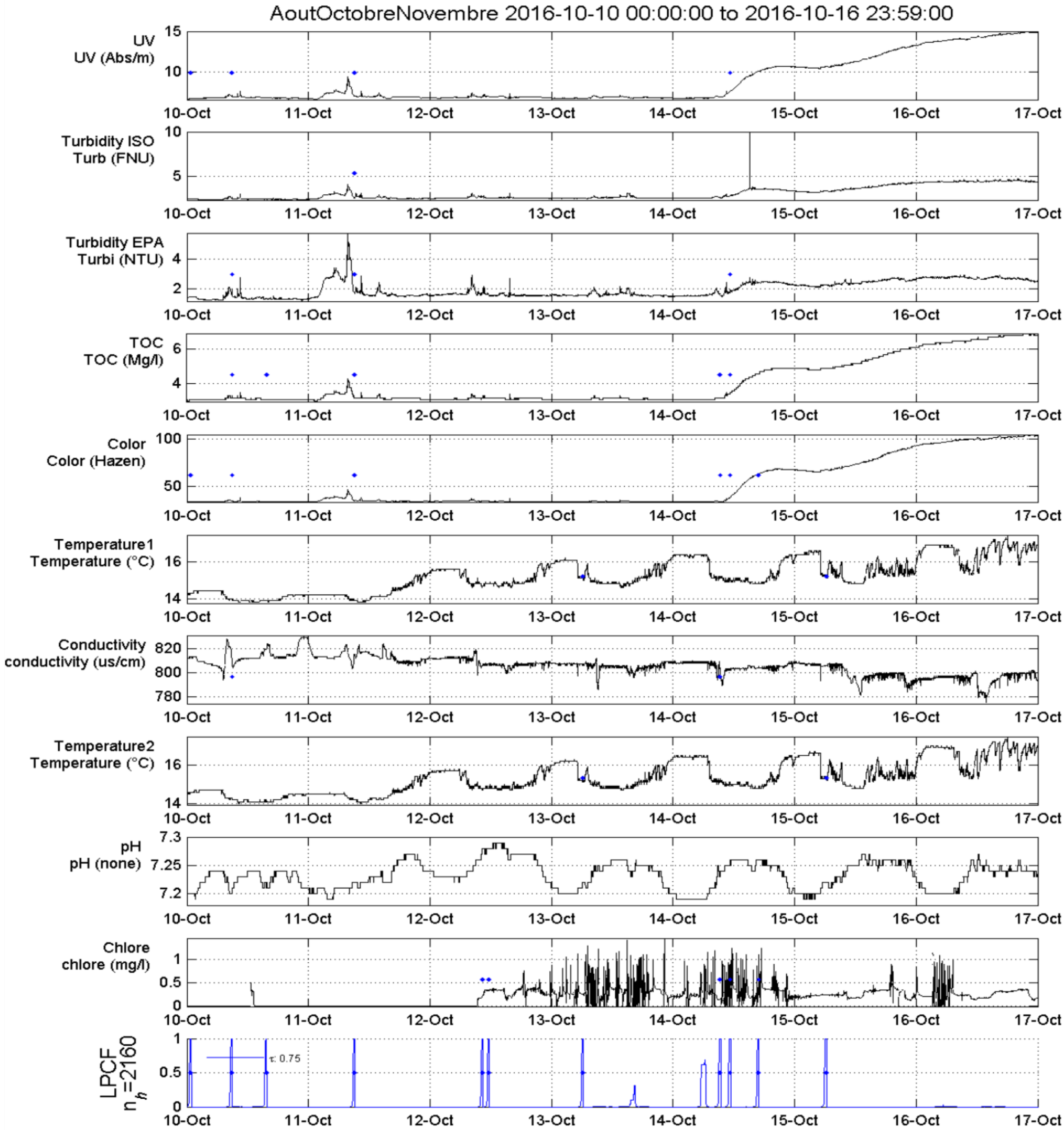


Figure 4.21. Results of the event detection algorithm for the week of October 10, 2016.

A second application of the configured model is done between May 21, 2017 and June 1, 2017. As mentioned previously (Chapter 3), a significant event has occurred between May, 28 and May, 29. Figure 4.22 gives the probability of event during this period. The event on May 28, 2017 has been successfully detected. During this period, no false alarm has been generated. It should be mentioned also that only one alarm (with probability equal to 1) has trigger on May 28, 2017. For detection, Canary is concerned with the significance of the change, not the total length of time that a signal remains abnormal [122]. During this period, the EDS can be represented by the perfect point (0, 1) of the ROC curve, where both sensitivity and specificity reach 100 %.

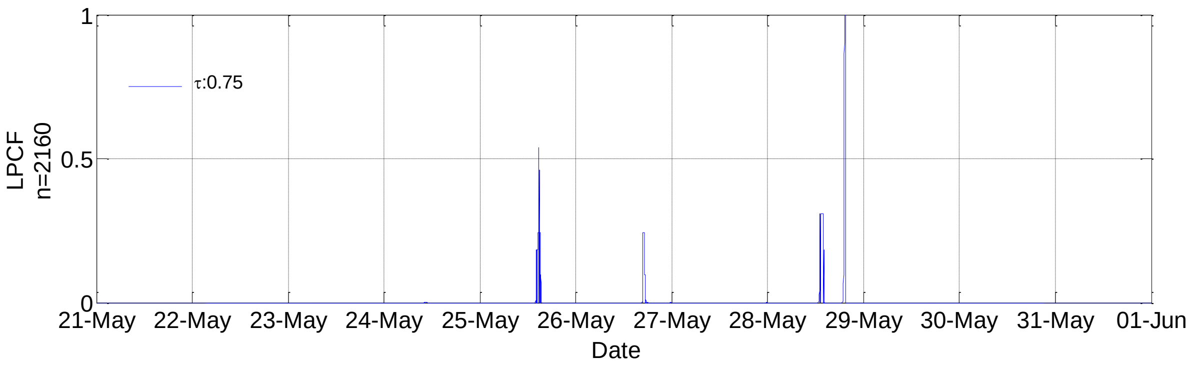


Figure 4.22. Probability of event from May 21, 2017 till June 1, 2017 at Polytech'Lille.

The different applications conducted at Polytech'Lille, proves that the EDS developed within Canary has a good sensibility in identifying significant abnormalities. False alarm (especially false positive) can be generated in some situations. However, their rate is acceptable, while the rate of detection of real event is generally high.

4.5.5 Sensitivity Analysis at Barroi

4.5.5.1 Case study

Using the selected parameters at Polytech (Table 4.12), the application of the adjusted model to S::CAN data measured at Barroi between December 1, 2016 and January 15, 2017 has led to high rate of false alarms. Since the water quality signals can differ between the two locations, a sensitivity analysis is done to S::CAN measurements at Barroi to find the adequate model.

A training phase is considered between December 1, 2016 and January 15, 2017. Nine measured parameters are analyzed within Canary: UV, Turbidity ISO, Turbidity EPA, TOC, DOC, Temperature, Conductivity, pH and Chlorine. The variation of different S::CAN data is illustrated in Figure 4.23. During this period, six events have been occurred: 11/12/2016; 12/12/2016; 13/12/2016; 03/01/2017; 05/01/2017 and 11/11/2017. These anomalies are characterized by a significant deviation of signals, especially UV, Turbidity and TOC. As for Polytech, the time step for Canary running is one minute.

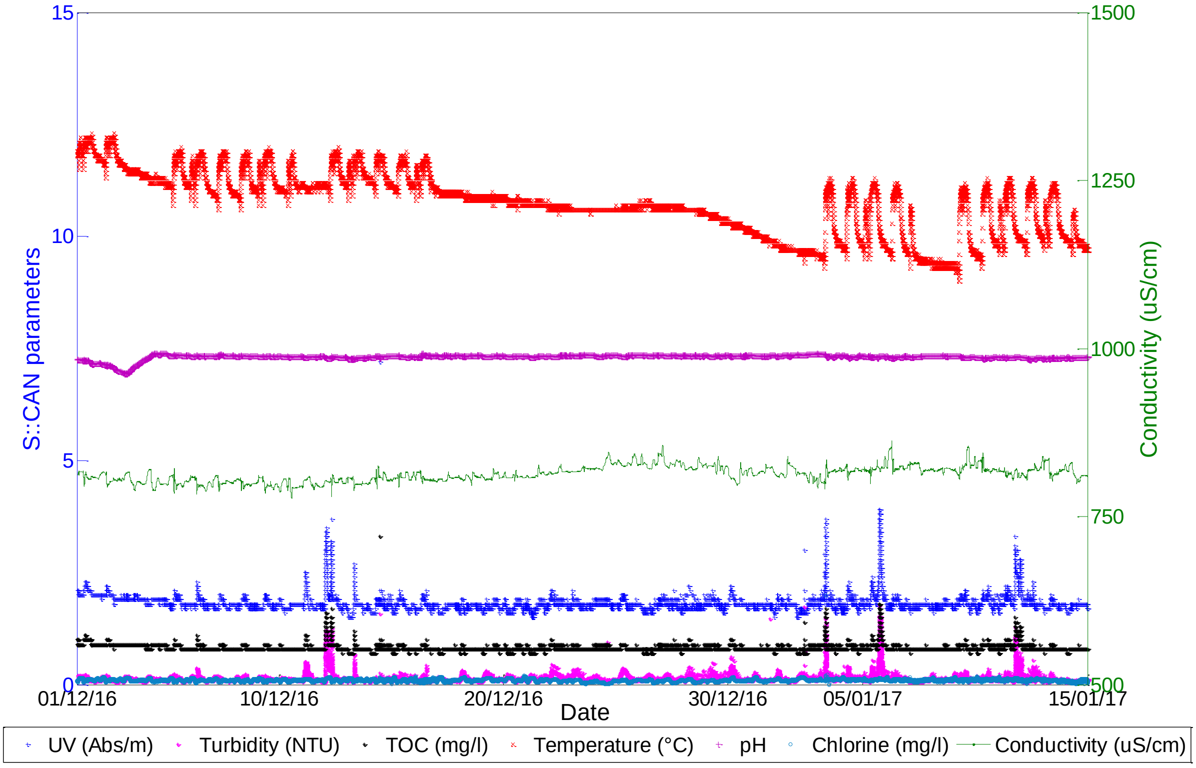


Figure 4.23. Variation of S::CAN data from December 12, 2016 till January 15, 2017 at Barroi.

4.5.5.2 Selection of parameters

During the training phase at Barroi, the use of MVNN algorithm has led to very high FPR (around 92 %) which indicates that the choice of MVNN is not appropriate for S::CAN signals measured at Barroi. As for Polytech, LPCF detection algorithm is more accurate.

As a general rule, the value of history window should be long enough to include 1.5 to 2 days of previous data [128]. Therefore, the following parametric study will take into consideration this two cases of window size: 2160 time steps (1.5 days) and 2880 time steps (2 days).

As demonstrated previously, the different values tested for the event threshold give similar sensitivity results. A value of 0.99 can be considered as acceptable in terms of precision and accuracy.

The two remaining parameters that affect the sensitivity performance of the EDS are: the outlier threshold and the BED window. Multiple combinations of these two parameters are evaluated. For each case, performance measures (TPR, TNR and FPR) are calculated and results are given in Table 4.13.

For a history window of 1.5 days, the analysis of results shows high FPR (> 50%) for an outlier 0.75 and 0.9. These two values of outliers should be excluded from the study. For an outlier of 0.99, a BED window of 100 time steps gives also a relatively high rate of false alarm (around 44%) with an acceptable TPR (66.67%). For an outlier of 1.5, the value of the BED window do not affect the results. However, the value of the TPR is very low (< 50%). This outlier should be also excluded from the selection. For an outlier of 1.1, we obtain the lowest generation of false alarm. However, a BED window of 50 time steps is better than 100 time steps, in terms of the TPR. For 1.5 days as history window, an outlier of 1.1 gives the better sensitivity results.

If the history window is increased to 2 days, the values of TPR remain the same while the FPR is well reduced. A history window of 2 days is then selected to reduce the generation of false alarm. A BED window of 50 time steps is also selected, since it increases the sensitivity (TPR).

Table 4.13. Sensitivity analysis at Barroi restaurant.

History window (time steps)	Outlier	BED window (time steps)	TPR (%)	TNR (%)	FPR (%)
2160	0,75	100	83,33	20,51	79,49
2160	0,9	100	66,67	41,03	58,97
2160	0,99	100	66,67	56,41	43,59
2160	1,1	100	50,00	66,67	33,33
2160	1,1	50	66,67	64,10	35,90
2160	1,5	100	33,33	89,74	10,26
2160	1,5	50	33,33	89,74	10,26
2880	1,1	100	50,00	76,92	23,08
2880	1,1	50	66,67	74,36	25,64

4.5.5.3 Application of selected parameters

Table 4.14 illustrates the list of selected parameters, obtained during the training phase at Barroi. The configured detection system has been applied in two different periods, to test its ability in identifying abnormalities.

The first application is applied from January 15, 2017 till January 31, 2017. The variation of S::CAN data during this period is illustrated in Figure 4.24. It indicates a significant deviation on January 19, 2017. This single anomaly is well detected with the event detection model, which gives a TPR of 100 %. However, Canary reports five other events (on 16/01/2017; 22/01/2017; 23/01/2017; 27/01/2017 and 31/01/2017) which are not true. This induces a false alarm rate about 33 %.

The adjusted model is also applied to the period between May 21, 2017 and June 1, 2017. During this period, the event between May, 28 and May, 29 is successfully detected within Canary (which gives a TPR of 100 %). Figure 4.25 illustrated the probability of event calculated with a history window of 2 days and an outlier of 1.1. It indicates that a false alarm is also generated on May, 26 which induces 11, 11 % as FPR and 66, 67 % as precision.

Table 4.14. Parameters selected to the EDS at Barroi.

Type of detection algorithm	LPCF
History window	2 days (2880 time steps)
Event threshold	0.99
Outlier threshold	1.1
BED window	50 time steps

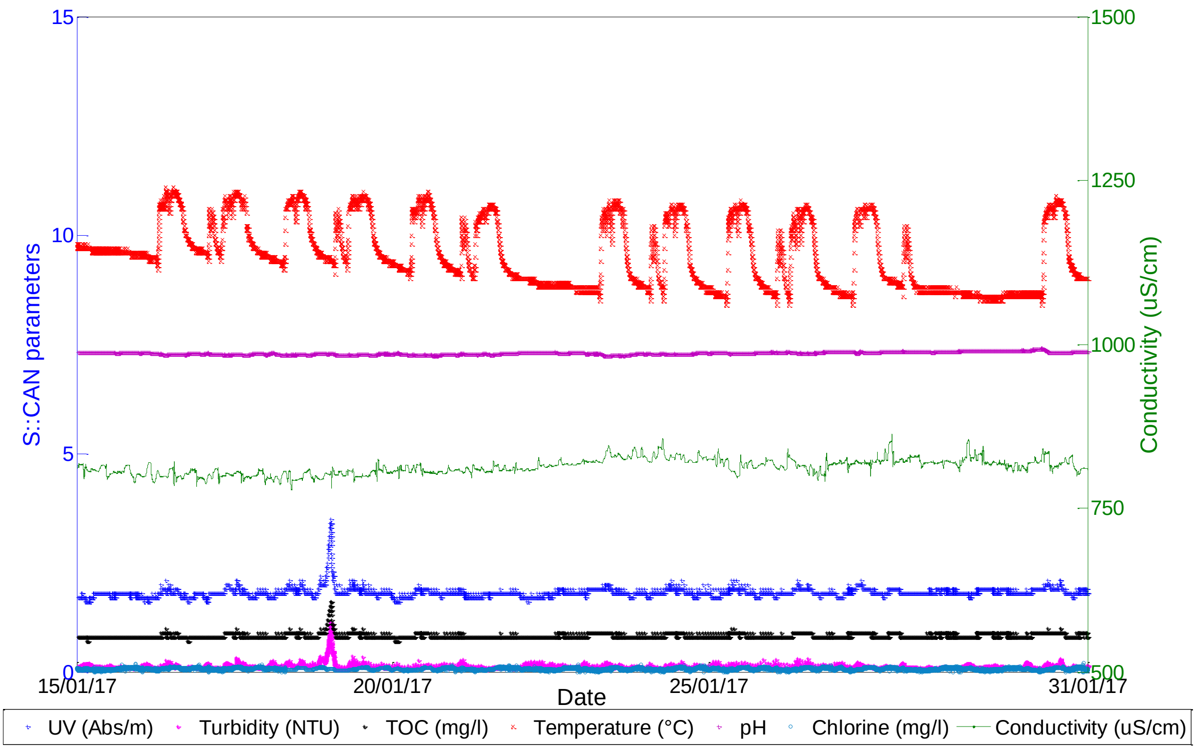


Figure 4.24. Variation of S::CAN data from January 15, 2017 till January 31, 2017 at Barroi.

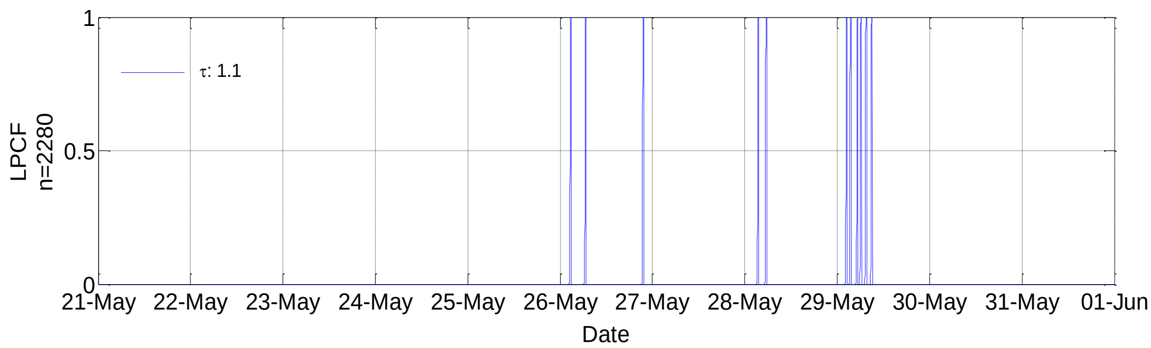


Figure 4.25. Probability of event from May 21, 2017 till June 1, 2017 at Barroi.

The EDS developed within Canary to analyze S::CAN signals at Barroi shows a high performance. The results indicates high sensitivity where real events are well detected with low rate of false alarm.

4.5.6 Discussion

The EDS developed within Canary analyzes S::CAN data in near-real time and calculates the probability of possible events. The event detection algorithm generates an alarm when a group of measurement exceeds a threshold probability during a certain period. Therefore, it differs between outliers occurring at a single time step and event that remains for a significant duration. The application of the adjusted configuration file to S::CAN signals allows the detection of several abnormalities, at both locations, with acceptable levels of false alarms.

4.6 Conclusion

This chapter presented the early detection of water quality anomaly using three methods: Statistical method (LPC), Artificial Intelligence (SVM) and Event Detection System (EDS).

The LPC and SVM methods have identified some ‘anomalies’ in recorded data by S::CAN sensor. These anomalous data usually refer to outlier at a single time step.

In order to avoid that a single outlier generates an alert, EDS approach has been developed within Canary to detect water quality abnormalities. Using a Binomial Discriminator, the probability of event is calculated in function of the number of outliers occurring in a specific duration. The application of this method to S::CAN data shows the ability of the system to detect true events with minimum false alarms. However, a sensitivity study is required for parameters selection (at each site location of S::CAN) in order to obtain the best performance of detection.

The following chapter will describe the use of risk assessment approach in combination with the online monitoring to help in the decision-making of water utilities in the case of occurrence of abnormal events in the water system.

Chapter 5. Risk Assessment of Drinking Water Contamination Using Smart Sensors

5.1 Introduction

This chapter presents the use of the qualitative risk assessment for the control of the drinking water quality. This use is illustrated through its application to the drinking water system of the Scientific Campus of the University of Lille. Based on sensor data, the qualitative assessment is conducted using two approaches based on the smart water system. These approaches analyze, in real-time, the variation of the water quality parameters (Turbidity and Chlorine), detect abnormal events and rank their risk level in function of their severity.

The first method analyses the combination Turbidity and Chlorine simultaneously. Turbidity classes are ranked in function of their probability. Chlorine concentration indicates the severity of the event. This approach is applied at two levels: i) risk assessment at each time step and ii) priority attention based on the duration of each risk category.

In the second method, the assessment is based on two criteria: i) the magnitude of measured Turbidity as indication of the event's impact, and ii) the duration of the risk. This approach is then developed. It takes into account the effect of Chlorine concentration on the severity of the event.

The detailed methodology of each approach will be described. Applications are conducted with S::CAN data. These approaches show that the real-time analysis of Turbidity and Chlorine provides an early detection and classification of risks. A comparison of results will allow to choose the appropriate method for the estimation of risk level.

A third method is developed by analyzing the variation of Turbidity and TOC simultaneously. The benefit as well as the limitation of this approach will be discussed.

5.2 Risk Assessment in drinking water quality

Risk assessment is the process of analysis and evaluation of risk. It provides information for identifying strategy for reducing and preventing disasters and designing the early warning system [66]. Analysis of sensors data can indicate different perturbations in the water quality. The associated level for potential risk should be identified. Although there are numerous contaminants that can compromise drinking water quality, not every potential hazard requires the same degree of attention [129]. The main objective is to detect possible events and classify their risk. A high risk event requires an emergency response, while a low risk needs a lower priority of attention.

The identification of risk consists of determining what, why, and how a risk can occur [130]. The following terms, defined by the *Australian Drinking Water Guidelines (ADWG)* [131], are used in risk assessment approach:

- Hazard: a biological, chemical, physical or radiological agent that has the potential to cause harm.
- Hazardous event: an incident that can lead to the presence of a hazard.
- Risk: likelihood of identified hazards causing harm in exposed populations.

The definition of risk is based on two main concepts: i) the probability that an adverse event could occur and ii) the severity of this event. Risk assessment is of high importance in the water industry for the following reasons [132]:

- To predict the burden of waterborne disease in the community, under outbreak and non-outbreak conditions.
- To help set microbial standards for drinking water supplies that will give tolerable levels of illness within the populations drinking that water.
- To determine the most cost-effective option to reduce microbiological health risks to drinking water consumers.
- To help identify the optimum treatment of water to balance microbial risks against chemical risks from disinfection by-products.
- To provide a conceptual framework to help individuals and organizations understand the nature and risk to, and from, their water and how those risks can be minimized.

Three types of risk assessment are generally used, namely: i) Epidemiological methods, ii) Quantitative Risk Assessment, and iii) Qualitative Risk Assessment.

5.2.1 Epidemiological methods

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health issues [133]. The use of the epidemiological approach in risk assessment is based on outbreaks investigations. This method consists of determining the relation between specific water quality parameters (such as Faecal Streptococci) and the level of disease in the population.

It is unlikely that epidemiological study relies on passive health surveillance, as the complexity of interpreting the results would be difficult, particularly if assessing the risks related to diarrhoeal disease [134]. Different types of epidemiological studies can be used in risk assessment of waterborne disease, especially: i) Descriptive (Ecological and Time series study), ii) Analytical (Case-control and cohort study) and iii) Intervention (Experimental study). According to the epidemiological approach, risk can be divided into four main types [135]:

- Absolute risk: generally named as incidence, it indicates the number of new cases occurring within a certain sized population during a specified time period.
- Attributable risk: the proportion of cases of a disease due to a particular risk factor.
- Relative risk: the ratio between the incidence of disease in those members of the population exposed to a possible risk factor and those not exposed.
- Odds ratio: the ratio between the probability that someone with a disease has experience of the potential environmental factor and the probability that a control has experience of the same factor.

5.2.2 Quantitative Risk Assessment

The Quantitative Risk Assessment allows the calculation of health risk and its comparison with another one agreed to be acceptable. A Quantitative Risk Assessment, that studies each component in water separately, is based on [136]:

- The presence of harmful substances and microorganisms in the water.
- Acceptable and infective doses.

- Estimations of the exposure of the water users.

Approaches based on Quantitative Risk Assessment quantify the potential risks arising from: i) hazards in source waters, ii) impact of the system in reducing the threat posed by source water through source protection and treatment, iii) residual risk from the production phase and iv) risks from recontamination during distribution [134].

To take into consideration the difference between microbial and chemical contaminants, the Quantitative Risk Assessment can be divided in two main categories: i) Quantitative Microbial Risk Assessment (QMRA) and ii) Quantitative Chemical Risk Assessment (QCRA).

5.2.2.1 Quantitative Microbial Risk Assessment (QMRA)

It can be defined as a scientific tool to evaluate the microbial safety of drinking water supplies. QMRA is used to estimate the risk of infection of a consumer, by combining the probability of occurrence of a pathogen at the tap with the consumption pattern and dose-response relationships [137]. To determine the public health risk, the main purpose of QMRA is to analyze the presence of pathogens with information about the consumed concentration and their infectivity.

As adapted by the *National Research Council* [138] , the QMRA approach involves four main steps [135]:

- Hazard identification: consists of determining pathogenic microorganisms that may induce acute or chronic effects to human health. The pathogen responsible for outbreaks must be identified with the severity of the infection, the patterns of transmission in the population and the control measures. The potential presence of "Emerging" pathogens, newly discovered and contributing to waterborne disease, should also be evaluated.
- Dose-response analysis: during an outbreak, individuals are exposed to different levels of the pathogens. To determine the relationship between pathogen exposure and infection, the volume of water ingested may be combined with the level of contamination of water. Two models of the infection process have been proposed: i) the exponential model: $Probability\ (infection) = 1 - exp(-rD)$ where D: pathogen dose, r: fraction of pathogens that survive to produce an infection and ii) the Beta-Poisson model: $Probability\ (infection) = 1 - \left(1 + \left(\frac{D}{ID50}\right)\right)^{-\alpha}$ where D: pathogen dose, α and ID_{50} are parameters of the beta-distribution used to describe variability in survival.

  Based on the study conducted in [139], an example of Dose-response analysis is done for E.Coli O157:H7. The probability of infection is calculated using β-Poisson model, with α=0.2099, N_{50}=1120 (parameters used in the literature [140]) and $ID_{50}=N_{50}(2^{\frac{1}{\alpha}} - 1)$. Figure 5.1 illustrates the dose-response relationship: probability of infection function of the different pathogen dose. It is evident that the probability of infection increases with E.Coli concentration.

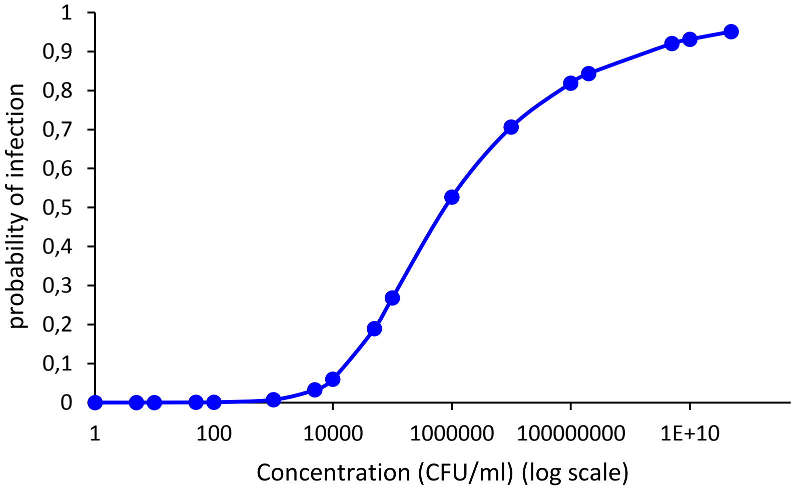


Figure 5.1. Example of Dose-response relationship for E.Coli (according to [139] and [140]).

- Exposure assessment: based on determining the size and nature of population exposed to pathogens with the duration of exposure. The level of pathogens for untreated water, combined with the proportion of surrogate removed by treatment, provides an estimation of the dose level of pathogens in water after treatment.
- Risk Characterization: studies the complete transmission system of pathogens to estimate the risk on public health. Variability and uncertainty of the estimation are taken into consideration.

5.2.2.2 Quantitative Chemical Risk Assessment (QCRA)

QCRA is developed to assess health risks posed to humans from exposure to chemicals in water. As for QMRA, it is based on four main steps [142]:

- Hazard identification: consists of determining the identity of the chemical of concern and if it is potentially hazardous to humans. For example, Cadmium can be identified as a hazardous chemical. It is classified as very toxic and carcinogenic to humans.
- Hazard characterization: based on defining the chemical's properties that induce the adverse health effects. The guideline values, with assumptions about exposure and dose, from international organizations should be determined (determination of guideline value). For Cadmium, the drinking water guideline value is 0.003 mg/l according to the *World Health Organization* (*WHO*).
- Exposure assessment: the main purpose is to obtain an exposure rate that can be compared with the guideline value. The exposure route, the frequency and the duration of exposure (short, medium or long term) should be evaluated. The exposure rate is given by the following equation [143]:

$$Intake\ (mg/kg\text{-}day) = \frac{CW\ x\ IR\ x\ EF\ x\ ED}{BW\ x\ AT} \tag{5.1}$$

Where: CW: Concentration in water of chemical of concern (mg/l).
IR: Ingestion Rate (l/day).
EF: Exposure Frequency (days/yr).
ED: Exposure Duration (yr).
BW: Body Weight (kg).

AT: Averaging Time (period over which exposure is averaged) (days)
For noncarcinogens: AT= ED*365 days per year.
For carcinogens: AT=Lifetime (70 years) *365 days per year.

For drinking water, *WHO guidelines* assume a water consumption rate of 2 liters per day and a body weight of 60 kg.

- Risk characterization: the hazard quotient is calculated as result of comparison between chemical exposure estimation and guideline value. This determines the priority level to take actions and corrective measures.

5.2.3 Qualitative Risk Assessment

Qualitative (or semi-quantitative) risk assessment approach is based on the classification of risks. It has the advantage of being relatively simple and the results are easy to present [144].

The Qualitative approach focuses on five steps [135]:

- Hazard scenario: identification of hazardous scenarios such as filter breakthrough or loss/breakdown of chemical disinfection system.
- Likelihood: ranking of how likely the event is.
- Consequence: ranking of the consequence (e.g. short-term injury).
- Risk score (Frequency x Effect): Weight attributed to likelihood and Severity of Consequences and multiplied to give value for hazardous event.
- Rank (level of risk): ranking of hazard scenario in order to provide the priority level for risk management.

Epidemiological and Quantitative approaches face some limitations and uncertainties in predicting relationship between the indicators parameters and the health risk. Epidemiological studies need a large sample sizes to uncover very small increases in risk, in addition of the costs incurred and expertise needed to mount a good study [26]. Risk quantification requires vast amount of data, could be time consuming and can give a false sense of precision. Qualitative method remains more accurate for a rapid detection of accidental contamination. It ranks the level of risk using classes, without resorting laboratory analyses. It does not seek to identify the actual levels of disease but to give a qualitative indication of the abnormalities in water. Risk ranking is the preferred qualitative method used by *WHO* in the execution of Water Safety Plan.

5.2.3.1 Methodology of Qualitative risk assessment

To identify the risk level, two main scales should be defined: i) the likelihood of event and ii) the severity of its consequences. The likelihood can be classified into four main categories: Likely, Moderate, Unlikely and Rare. A weight is assigned to each category, according to its probability of occurrence (Once per week, month, year, etc. [145]). In the same way, four classes define the severity of consequences: Major, Moderate, Minor and Insignificant. Based on the impact of the consequence (harmful or lethal to small or large population [145]), a corresponding weight is attributed. Table 5.1 provides the two scales (likelihood and consequences) used in the Qualitative Risk Assessment.

A Risk Score R is defined as follows:

$$R = Weight\ (Likelihood)\ x\ Weight\ (Severity\ of\ Consequences) \tag{5.2}$$

As indicated in Table 5.2, a Risk Level category is identified (Low, Moderate, High or Very High), according to the risk score class. To determine the magnitude of the risks and to prioritize the risks, a risk matrix is recommended [146]. A risk matrix combines the likelihood and the severity of consequences of an event in a risk level category (Table 5.3).

Table 5.1. Qualitative Risk Assessment Scale. (a) Likelihood Scale; (b) Severity of Consequences Scale.

(a)

Category	Likely	Moderate	Unlikely	Rare
Weight	4	3	2	1

(b)

Category	Major	Moderate	Minor	Insignificant
Weight	4	3	2	1

Table 5.2. Risk Score and Risk Level classification.

Risk Score	<4	4≤R≤7	7≤R≤13	>13
Risk Level	Low	Moderate	High	Very High

Table 5.3. Risk Matrix.

	Severity of Consequences			
Likelihood	Insignificant	Minor	Moderate	Major
Likely	Moderate	High	High	Very High
Moderate	Low	Moderate	High	High
Unlikely	Low	Moderate	Moderate	High
Rare	Low	Low	Low	Moderate

The major issue in Qualitative assessment concerns the descriptive definition of both scales (Likelihood and Severity of Consequence). Using the smart technology, the likelihood and the severity of consequences can be defined in function of monitored parameters (Turbidity and Chlorine). Then, the qualitative assessment will provide in real-time the risk of any event.

5.3 Application to the water system of the Scientific Campus

The risk assessment approach is applied to the drinking water quality at the Scientific Campus of the University of Lille. S::CAN data, especially the Turbidity and Free Chlorine measured at Polytech'Lille, for the period between July 1 and 15, 2017, are used. Turbidity and Free Chlorine are two main indicators of the water quality. The Turbidity is an indicator of the presence of suspended matters that disturb the water. Standards defines 1 NTU as a threshold limit for Turbidity. For Chlorine level, a rapid decrease indicates a water quality variation induced generally by biofilm growth. According to the *WHO*, the concentration of Free Chlorine in treated water must be maintained between 0.2 and 0.5 mg/l.

The variation of Turbidity and Chlorine at the water supply of Polytech'Lille is illustrated in Figure 5.2. Some events have been detected for Turbidity, especially on July 6 and 7. The concentration of Chlorine varies between 0 and 0.3 mg/l.

In Approach 1 and 2, an event is defined as exceeding of the threshold of Turbidity (Turbidity > 1 NTU).

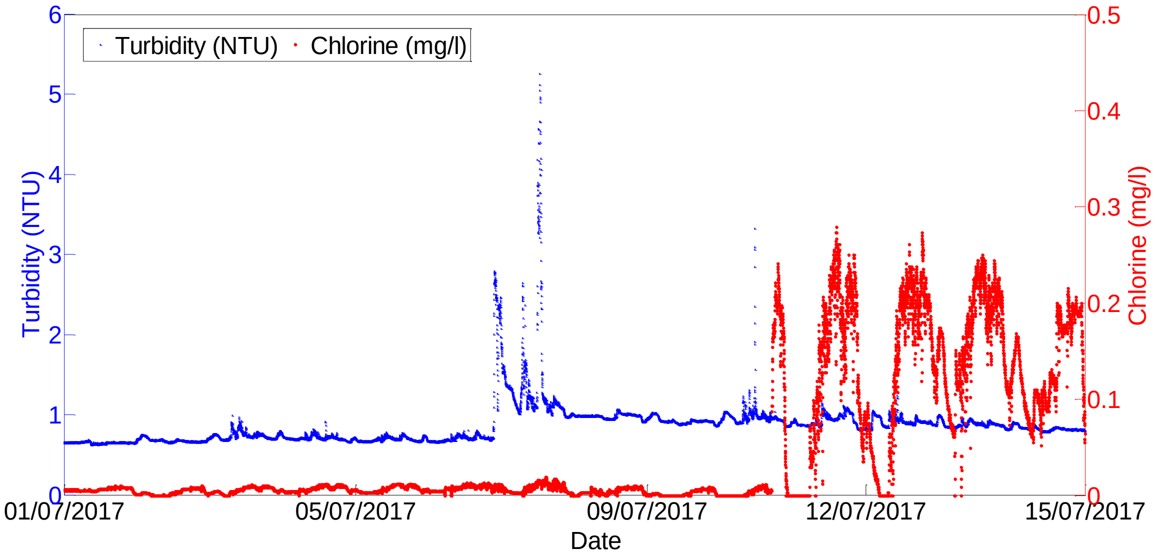


Figure 5.2. Variation of Turbidity and Chlorine between July 1 and 15, 2017.

5.4 Risk Assessment analysis

5.4.1 Approach 1-Level 1

The analysis of real-time data allows to define two major classes: i) Turbidity < 1 NTU which indicates a safe water quality, and ii) Turbidity >1 NTU which indicates a possible event. Data are firstly filtered (Event or not). The list of potential events are analyzed to estimate the level of risk induced.

In case of event, Turbidity can be divided into four main classes: [1-1.5], [1.5-3], [3-10] and > 10 NTU. The analysis of historical data shows that Turbidity values and their corresponding probability are inversely proportional. The likelihood is defined as the probability of each class of Turbidity. As indicated in Table 5.4 (a), an event with Turbidity value between 1 and 1.5 NTU occurs likely with a probability higher than 80 %. However, the probability that Turbidity reaches very high amplitudes (more than 10 NTU) is very low; it can be due to some faults in the instrument. Based on the likelihood scale of Table 5.1 (a), a weight is attributed for each category of Turbidity (Table 5.4 (a)).

The severity of consequences is defined according to Chlorine ranking in four intervals: [0-0.005] or >0.5, [0.005-0.05], [0.05-0.2] and [0.2-0.5] mg/l (Table 5.4 (b)). Since the amount of Chlorine should be maintained between 0.2 and 0.5 mg/l, the major consequence occurs when Chlorine concentration is lower than 0.005 mg/l or higher than 0.5 mg/l. The corresponding weight is assigned in function of the Severity Scale of Table 5.1 (b).

The risk level is defined as the combination of: i) the probability of Turbidity and ii) the severity of consequences based on Chlorine interval. The risk matrix, defined in Table 5.3, is used in this approach. It indicates that the risk level is highest for low Chlorine value combined with high probability of event (Turbidity between 1 and 1.5 NTU).

Analysis of the risk assessment's results will determine the priority score for risk management. The objective of prioritization is to rank hazardous events to provide a focus on the most significant hazards [134]. Risk-reduction actions will be based on the level of priority attention. High risk level identifies the need of high priority attention with emergent corrective actions. While, low risk level can be ignored or given a low priority attention. Table 5.5 gives the priority level in accordance with the risk level ranking.

Table 5.4. Description of Approach 1. (a) Likelihood (for Turbidity>1 NTU); (b) Severity of Consequences.

(a)

Descriptor	Description: Turbidity (NTU)	Weight
Likely (>80%)	[1-1,5]	4
Moderate (10-80%)	[1,5-3]	3
Unlikely (4-10%)	[3-10]	2
Rare (<4%)	>10	1

(b)

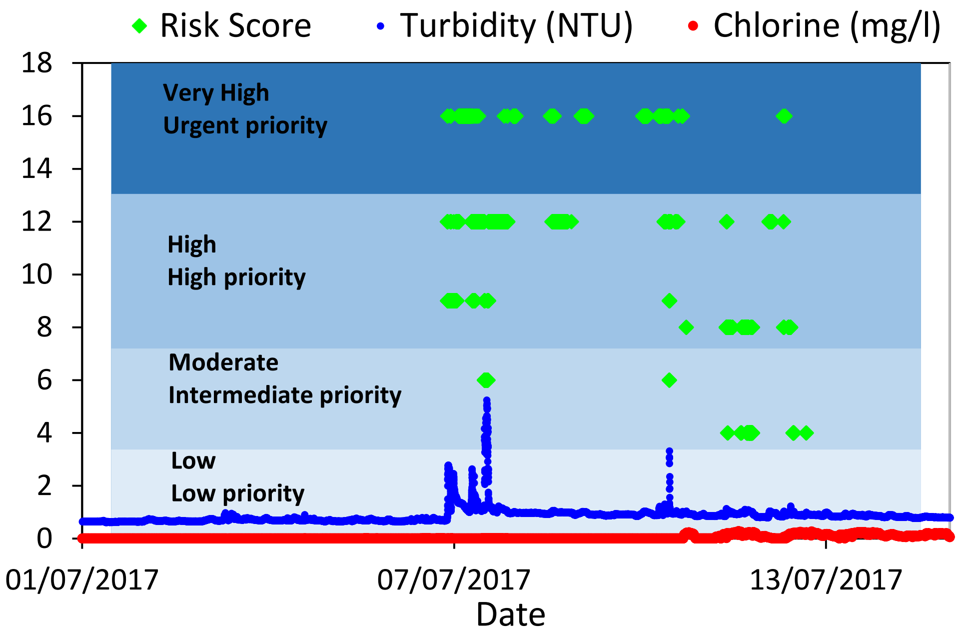
Descriptor	Description: Chlorine (mg/l)	Weight
Major	[0-0.005] or > 0.5	4
Moderate	[0.005-0.05]	3
Minor	[0.05-0.2]	2
Insignificant	[0.2-0.5]	1

Table 5.5. Priority level classification.

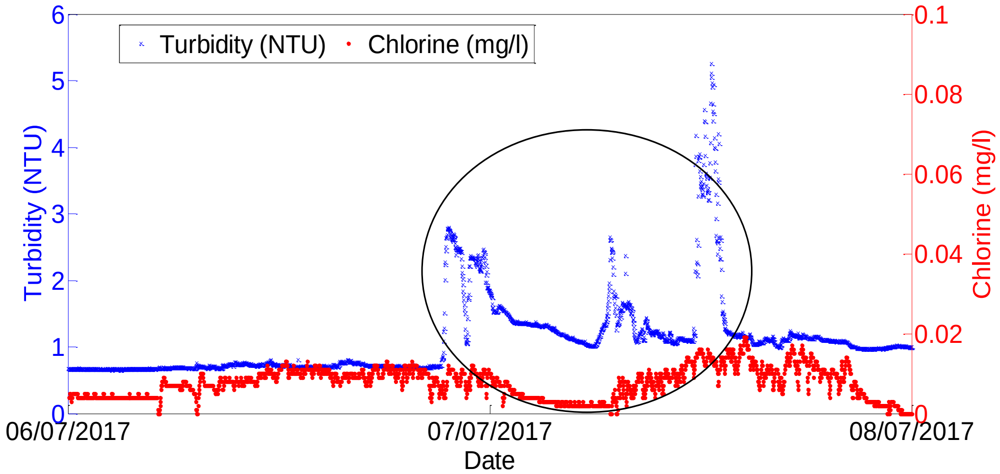
Risk Level	**Prioritization of Risk**
No Risk	No attention needed
Low	Low priority attention
Moderate	Intermediate priority attention
High	High priority attention
Very High	Urgent priority attention

The application of this approach on historical data (defined in Section 5.3) is illustrated in Figure 5.3 (a). At each time step, Turbidity and Chlorine are ranked with the specified weight. The risk score is then calculated (eqn. 5.2). The corresponding risk category is identified. Figure 5.3 (a) shows the risk score, as well as the variation of Turbidity and Chlorine between July 1 and 15, 2017. It indicates situations of very high risk, especially on July 7, 2017. This risk level is verified in Figure 5.3 (b). It shows an increase in the Turbidity for many hours coupled with sudden decrease in Chlorine concentration. This combination provides an indication of a potential event.

On the other hand, other cases of high risk are observed on July 8 and 9, 2017. They are induced by very low concentrations of Chlorine (near zero) with Turbidity values close to the threshold (1 NTU). Moderate levels are observed, in particular, on July 11 and 12, 2017. Such levels are obtained from insignificant severity of consequence (Chlorine about 0.2 mg/l) combined with a likely event of Turbidity (between 1 and 1.5 NTU).



(a)



(b)

Figure 5.3. Application of Approach 1-Level 1. (a) Risk level for S::CAN data; (b) Event detected on July 7, 2017.

The risk level could be also evaluated for real time data, with the corresponding priority attention. An example of real time assessment of risk is given in Figure 5.4 for a Turbidity of 1.6 NTU and Chlorine value of 0.02 mg/l. A weight of 3 will be given to both Turbidity and Chlorine (Table 5.4). The calculated risk score is 9 (eqn 5.2). Using Table 5.2 and Table 5.3, a High risk level is identified for this data. A High priority attention should be assigned (Table 5.5).

Turbidity (NTU):	1,6
Chlorine (mg/l):	0,02
Risk Level:	High
Prioritization of Risk:	High priority attention

Figure 5.4. Example of real time risk assessment according to Approach 1-Level 1.

In this approach, the risk level is assigned by evaluating Turbidity and Chlorine values at each single time step. The outcome of the risk assessment will determine the priority attention required. However, many factors can lead to unaccepted values of Turbidity and/or Chlorine (contamination, faults in instrument, data transmission problems, etc.). It is obvious that not all exceeding of the tolerable limit should generate an alarm. The determination of priority level, based on data analysis each minute, will not be very precise. False identification could occur in some cases.

5.4.2 Approach 1-Level 2

To determine accurately the priority attention level, the duration factor could be taken into consideration. Level 1 of Approach 1 could be developed by applying the three following steps:

- Determination of risk level at each time step (according to Approach 1-Level 1).
- Classification of risk level, function of their category (Very High, High, Moderate, and Low).
- Duration of each category of risk level during an event.

An application is done for S::CAN data at Polytech. For each event, risk levels are classified according to their category. The start and the end of each risk category are determined. The duration of each level, during an event, is calculated as follows:

$$Duration\ (min) = Date\ (End\ of\ risk\ level) - Date\ (Start\ of\ risk\ level) \qquad (5.3)$$

The histogram in Figure 5.5 (and Appendix A: Figure A.1) shows an example for the classification of risk level with the corresponding duration. It shows the importance of the duration factor. For example, an event of “High” risk level of 135 minutes (on 07/07/2017 09:25) is more dangerous than an event of “Very High” risk that occurs during 1 minute (on 06/07/2017 21:24). Such event (7/7/2017 09:25) requires more attention.

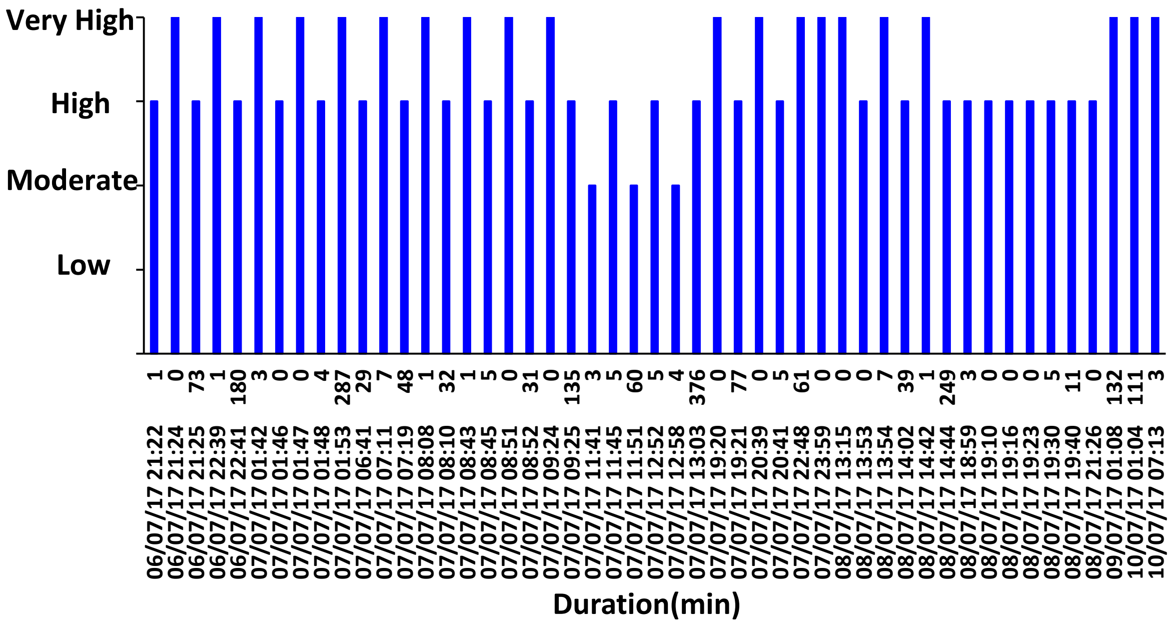


Figure 5.5. Risk level classes with the corresponding duration.

To take into account the duration of each category of risk, a new scale for priority level is given in Table 5.6. It is defined as the combination of the risk level and the duration. A corresponding weight is attributed for each category of risk level. The largest weight is assigned for "Very High" risk. In the same way, the duration is divided in four main classes (Long, Medium, Short, and Instantaneous). Weights are given in the increasing order of duration: a long duration will be the most critical. The priority attention level is defined as follows:

$$Priority\ attention\ Score\ P = Weight\ (Risk\ Level)\ x\ Weight\ (Duration) \qquad (5.4)$$

The priority score is classified in four intervals. The priority level is then identified according to the class of priority score (Table 5.7). Table 5.8 illustrates an example of results (for the period of Section 5.3) (Appendix A: Table A.1). For each new event, the risk level is classified and the corresponding duration is calculated. The priority attention for each category of risk level is then identified.

The analysis of this table illustrates the influence of the duration factor on the priority level. For example, an event of "High" risk (occurred on 06/07/2017 22:41) requires more attention higher than an event of "Very High" risk (occurred on 06/07/2017 22:39). Although the risk level is more important on 06/07/2017 22:39, the event on 06/07/2017 22:41 requires higher corrective actions, since the latter remains for more than 2 hours.

Table 5.6. Priority level scale. (a) Risk Level Scale; (b) Duration Scale.

(a)

Category	Very High	High	Moderate	Low
Weight	4	3	2	1

(b)

Category	Long (>120min)	Medium(60-120min)	Short(30-60 min)	Instantaneous(0-30min)
Weight	4	3	2	1

Table 5.7. Prioritization of risk.

Priority Score	<4	4≤P≤7	7≤P≤13	>13
Priority Level	L.P	I.P	H.P	U.P

	Duration			
Risk level	Instantaneous	Short	Medium	Long
Very High	I.P	H.P	H.P	U.P
High	L.P	I.P	H.P	H.P
Moderate	L.P	I.P	I.P	H.P
Low	L.P	L.P	L.P	I.P

L.P	Low Priority Attention
I.P	Intermediate Priority Attention
H.P	High Priority Attention
U.P	Urgent Priority Attention

Table 5.8. Risk and Priority level for S::CAN data according to Approach 1-Level 2.

Start of new level of an event	End of level	Risk Level	Weight of level	Duration (min)	Weight of duration	Priority Score	Priority level
6/7/2017 21:22	6/7/2017 21:23	High	3	1	1	3	L.P
6/7/2017 21:24	6/7/2017 21:24	Very High	4	0	1	4	I.P
6/7/2017 21:25	6/7/2017 22:38	High	3	73	3	9	H.P
6/7/2017 22:39	6/7/2017 22:40	Very High	4	1	1	4	I.P
6/7/2017 22:41	7/7/2017 1:41	High	3	180	4	12	H.P
7/7/2017 1:42	7/7/2017 1:45	Very High	4	3	1	4	I.P
7/7/2017 1:46	7/7/2017 1:46	High	3	0	1	3	L.P
7/7/2017 1:47	7/7/2017 1:47	Very High	4	0	1	4	I.P
7/7/2017 1:48	7/7/2017 1:52	High	3	4	1	3	L.P
7/7/2017 1:53	7/7/2017 6:40	Very High	4	287	4	16	U.P
7/7/2017 6:41	7/7/2017 7:10	High	3	29	1	3	L.P

All the steps of Approach 1 have been automated in a VBA script and are summarized in the flow chart of Figure 5.6. For Approach 1, Level 2 seems to be more realistic than Level 1, in terms of priority attention. For "Very high" risk (as on 06/07/2017 22:39 in Table 5.8), Level 1 will indicate the need of more attention (Table 5.5). However, Level 2 evaluates the duration of this risk category before defining the priority level as Intermediate. After detecting the risk of abnormalities, it is important to take into account the duration of the event. A short-time event should not have the same decision response as an event that remains for several minutes or even hours. The duration will help in determining the nature of the occurred event: i) instantaneous risk level could be related to sensor data (connection issues, etc.) and then ignored, ii) significant duration should be analyzed to verify the potential existence of a contamination, iii) long duration indicates the presence of anomaly in water quality and requires urgent actions.

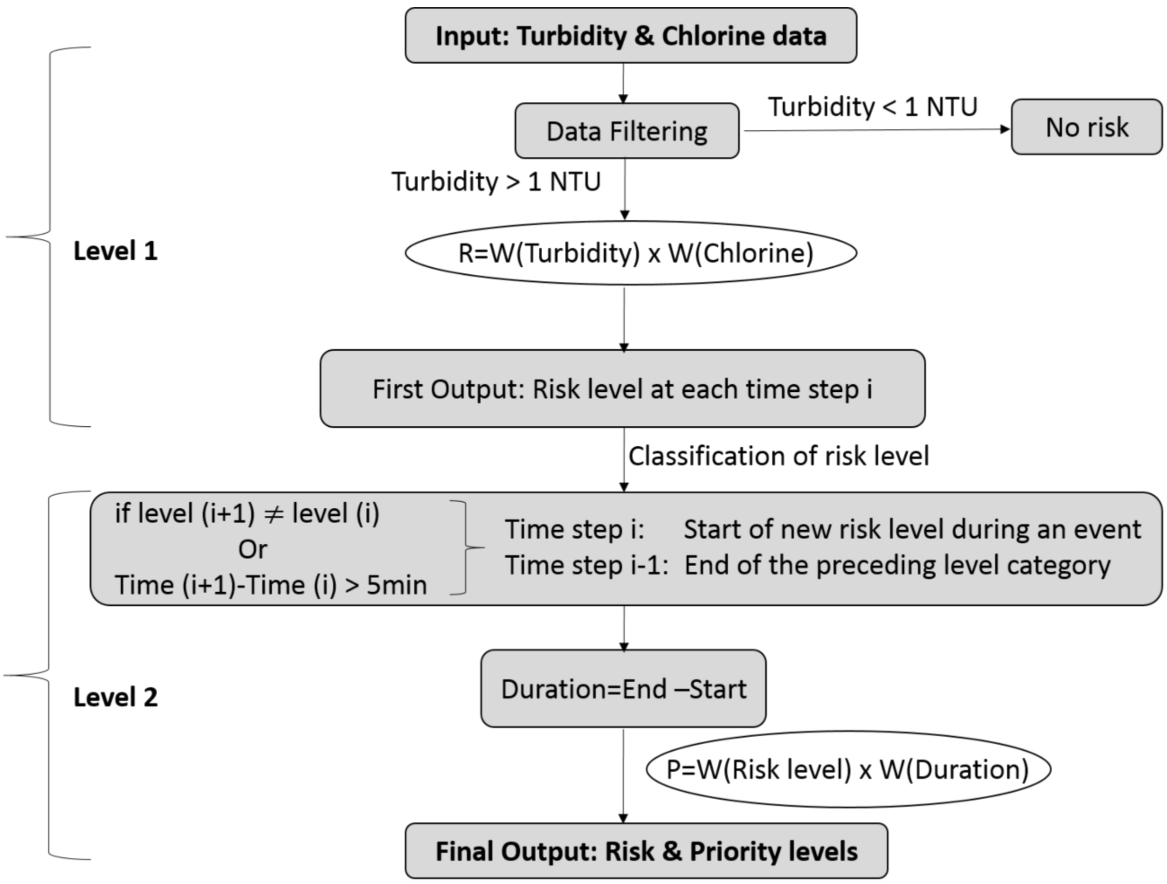


Figure 5.6. Flow chart of Approach 1.

5.4.3 Approach 2-Level 1

In Approach 1, since the likelihood scale is described as the probability of Turbidity values, highest weight is given to the lowest Turbidity that occurs frequently. However, it is obvious that the impact of an event depends on the magnitude of deviation from the Standard limit (1 NTU). The event becomes more severe when the deviation from the limit is bigger. Turbidity close to 1 NTU will not be dangerous like high values. Approach 2 will take into account the effect of parameter value (Turbidity) on the severity of consequences and will evaluate the frequency of an event in terms of the duration.

The main purpose of Approach 2 is to assess the variation of one single parameter measured by S::CAN. Approach 2 is firstly applied to Turbidity data. The main output is the list of events that occurred during a period with the corresponding risk level. Firstly, data are filtered (Event or not) by analyzing Turbidity value each minute. The start and end time of each event is identified. The corresponding duration is calculated (difference between end and start time of an event).

The severity of consequences is defined according to the average Turbidity during each event (Table 5.9 (a)). Turbidity data are classified in four main intervals. Weights are assigned in ascending order of Turbidity. An average Turbidity lower than 1.2 NTU has insignificant impact. However, a value greater than 3 NTU can induce a major impact.

The duration of event is classified in four main categories (from Instantaneous to Long) with the corresponding weight (Table 5.9 (b)).

The risk score (Table 5.10) is calculated as follows:

$$Risk\ Score\ R = Weight\ (\overline{Turbidity})\ x\ Weight\ (Duration) \qquad (5.5)$$

The ranking of the risk score identifies the category of risk level (Low, Moderate, High, and Very High) (Table 5.11). The application of this approach, to S::CAN data, is illustrated in Table 5.12 (Appendix A: Table A.2). It shows some events occurring during this period. The identification of the Start and End time of each event allows to calculate the corresponding duration. The average Turbidity is also calculated. The combination of their weight determines the risk level of each event.

A "High" risk is identified for the first event (on July 6, 2017). This event lasts for around one day with an average Turbidity about 1.5 NTU. Figure 5.7 verifies the high risk during this period. Important perturbations are observed in Turbidity signal. Turbidity data exceed significantly the limit of 1 NTU. High priority attention should be assigned in two steps: i) analysis of the water quality to identify the origin of anomaly, and ii) corrective actions if required.

Most of other events have "Low" risk level. The excess of the threshold is limited to few minutes, with values close to 1 NTU. This indicates a relatively safe drinking water with very low risk of anomaly. There is no need for urgent intervention in most cases.

Table 5.9. Description of Approach 2. (a) Severity of Consequences; (b) Duration of event.

(a)

Descriptor	**Description: $\overline{Turbidity}$(NTU)**	**Weight**
Major	>3	4
Moderate	[1,5-3]	3
Minor	[1,2-1,5]	2
Insignificant	[1-1,2]	1

(b)

Descriptor	**Description: Duration (min)**	**Weight**
Long	>120	4
Medium	60-120	3
Short	30-60	2
Instantaneous	0-30	1

Table 5.10. Calculation of risk score.

	Severity of Consequences			
Event Duration	Insignificant	Minor	Moderate	Major
Long	4	8	12	16
Medium	3	6	9	12
Short	2	4	6	8
Instantaneous	1	2	3	4
Risk Score	<4	4≤R≤7	7≤R≤13	>13
Risk Rank	Low	Moderate	High	Very High

Table 5.11. Description of risk matrix.

	Severity of Consequences			
Event Duration	Insignificant	Minor	Moderate	Major
Long	Moderate	High	High	Very High
Medium	Low	Moderate	High	High
Short	Low	Low	Moderate	High
Instantaneous	Low	Low	Low	Moderate

Table 5.12. Risk level for S::CAN data according to Approach 2-Level 1.

Start of Event	End of Event	Duration (min)	Weight of Duration	Average Turbidity (NTU)	Weight of Turbidity	Risk Score	Risk Level
06/07/2017 21:22	07/07/2017 20:46	1404	4	1,499	2	8	High
07/07/2017 22:48	07/07/2017 23:49	61	3	1,011	1	3	Low
07/07/2017 23:59	07/07/2017 23:59	0	1	1,001	1	1	Low
08/07/2017 13:15	08/07/2017 13:15	0	1	1,005	1	1	Low
08/07/2017 13:53	08/07/2017 18:53	300	4	1,017	1	4	Moderate
08/07/2017 18:59	08/07/2017 19:02	3	1	1,002	1	1	Low
08/07/2017 19:10	08/07/2017 19:10	0	1	1,004	1	1	Low
08/07/2017 19:16	08/07/2017 19:16	0	1	1,002	1	1	Low
08/07/2017 19:23	08/07/2017 19:23	0	1	1,002	1	1	Low
08/07/2017 19:30	08/07/2017 19:35	5	1	1,004	1	1	Low
08/07/2017 19:40	08/07/2017 19:51	11	1	1,003	1	1	Low

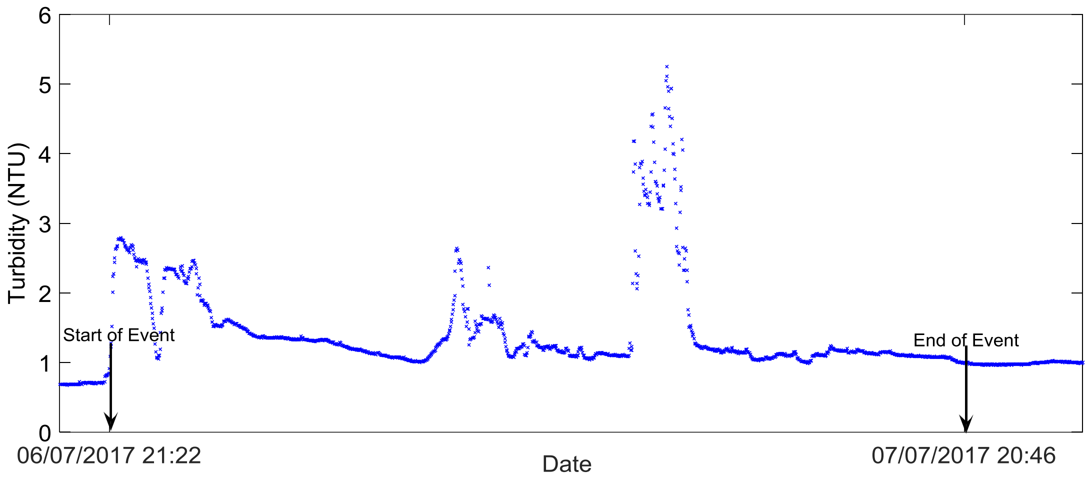


Figure 5.7. First event detected by Approach 2-Level 1 between July, 6 and 7, 2017.

5.4.4 Approach 2-Level 2

Approach 2 can be extended by analyzing the variation of two parameters: Turbidity and Chlorine simultaneously. The risk score is calculated as for Approach 2-Level 1:

$$Risk\ score = Weight\ (Severity\ of\ Consequences)\ x\ Weight\ (Duration) \quad (5.6)$$

The definition of severity of consequences depends, in this case, on the classification of both Turbidity and Chlorine (Table 5.13). For each event, the average Turbidity and the average Chlorine are calculated. Highest weight is assigned to the biggest value of Turbidity. However, for Chlorine, weights are inversely proportional to the concentration. The consequence of event is major for high Turbidity combined with small amount of Chlorine. A severity scale with assigned weight (Table 5.14) is defined as follows:

$$Severity\ Score\ S = Weight\ \left(\overline{Turbidity}\right)\ x\ Weight\ (\overline{Chlorine}) \quad (5.7)$$

Table 5.13. Description of the Severity of Consequences. (a) Turbidity classification; (b) Chlorine classification.

(a)

Descriptor	Description: $\overline{Turbidity}$(NTU)	Weight
Major	>3	4
Moderate	[1,5-3]	3
Minor	[1,2-1,5]	2
Insignificant	[1-1,2]	1

(b)

Descriptor	Description: $\overline{Chlorine}$ (mg/l)	Weight
Major	[0-0,005] or >0,5	4
Moderate	[0,005-0,05]	3
Minor	[0,05-0,2]	2
Insignificant	[0,2-0,5]	1

Table 5.14. Description of Severity Scale.

Severity Score	<4	4≤S≤7	7≤S≤13	>13
Severity Scale	Insignificant	Minor	Moderate	Major
Weight	1	2	3	4

An example of the calculation of the severity scale is illustrated in Table 5.15 (Appendix A: Table A.3). It indicates some events that occurred (between July 1 and 15, 2017) with their corresponding severity level. During this period, no Major or Moderate Consequences are observed. The level of severity for the majority of events is Insignificant or Minor. This can be verified by the small deviation of Turbidity from the limit, even when Chlorine is close to 0 mg/l.

The duration of each event is then calculated and given a corresponding weight (Table 5.9 (b)). The risk level for each event is identified (Table 5.11). A part of the results are shown in Table 5.16 (Appendix A: Table A.4). It indicates “Low” risk levels for most events. There is no major anomaly detected.

Table 5.15. Severity level for S::CAN data according to Approach 2-Level 2.

Start of Event	End of Event	Average Turbidity (NTU)	Weight of Turbidity	Average Chlorine (mg/l)	Weight of Chlorine	Severity Score	Severity Level
6/7/2017 21:22	7/7/2017 20:46	1.499	2	0.008	3	6	Minor
7/7/2017 22:48	7/7/2017 23:49	1.011	1	0.001	4	4	Minor
7/7/2017 23:59	7/7/2017 23:59	1.001	1	0	4	4	Minor
8/7/2017 13:15	8/7/2017 13:15	1.005	1	0.004	4	4	Minor
8/7/2017 13:53	8/7/2017 18:53	1.017	1	0.008	3	3	Insignificant
8/7/2017 18:59	8/7/2017 19:02	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:10	8/7/2017 19:10	1.004	1	0.009	3	3	Insignificant
8/7/2017 19:16	8/7/2017 19:16	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:23	8/7/2017 19:23	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:30	8/7/2017 19:35	1.004	1	0.007	3	3	Insignificant
8/7/2017 19:40	8/7/2017 19:51	1.003	1	0.01	3	3	Insignificant

Table 5.16. Risk level for S::CAN data according to Approach 2-Level 2.

Start of Event	End of Event	Duration (min)	Weight of Duration	Severity Level	Weight of Severity Level	Risk Score	Risk Level
6/7/2017 21:22	7/7/2017 20:46	1404	4	Minor	2	8	High
7/7/2017 22:48	7/7/2017 23:49	61	3	Minor	2	6	Moderate
7/7/2017 23:59	7/7/2017 23:59	0	1	Minor	2	2	Low
8/7/2017 13:15	8/7/2017 13:15	0	1	Minor	2	2	Low
8/7/2017 13:53	8/7/2017 18:53	300	4	Insignificant	1	4	Moderate
8/7/2017 18:59	8/7/2017 19:02	3	1	Insignificant	1	1	Low
8/7/2017 19:10	8/7/2017 19:10	0	1	Insignificant	1	1	Low
8/7/2017 19:16	8/7/2017 19:16	0	1	Insignificant	1	1	Low
8/7/2017 19:23	8/7/2017 19:23	0	1	Insignificant	1	1	Low
8/7/2017 19:30	8/7/2017 19:35	5	1	Insignificant	1	1	Low
8/7/2017 19:40	8/7/2017 19:51	11	1	Insignificant	1	1	Low

A VBA code has been developed for all the steps of Approach 2 as indicated in Figure 5.8. Some differences in risk assessment results are observed between Level 1 and 2 of Approach 2. For example, the event of 07/07/2017 22:48: Level 1 identifies a "Low" risk while Level 2 indicates "Moderate" risk (Table 5.12 and Table 5.16). During this event, Chlorine concentration (0.001 mg/l in Table 5.15) is very low which could imply major impact. The fact of ignoring this concentration of Chlorine can lead to underestimation of the risk level. The assessment's results remains more accurate by combining the variation of different parameters (Turbidity and Chlorine) simultaneously.

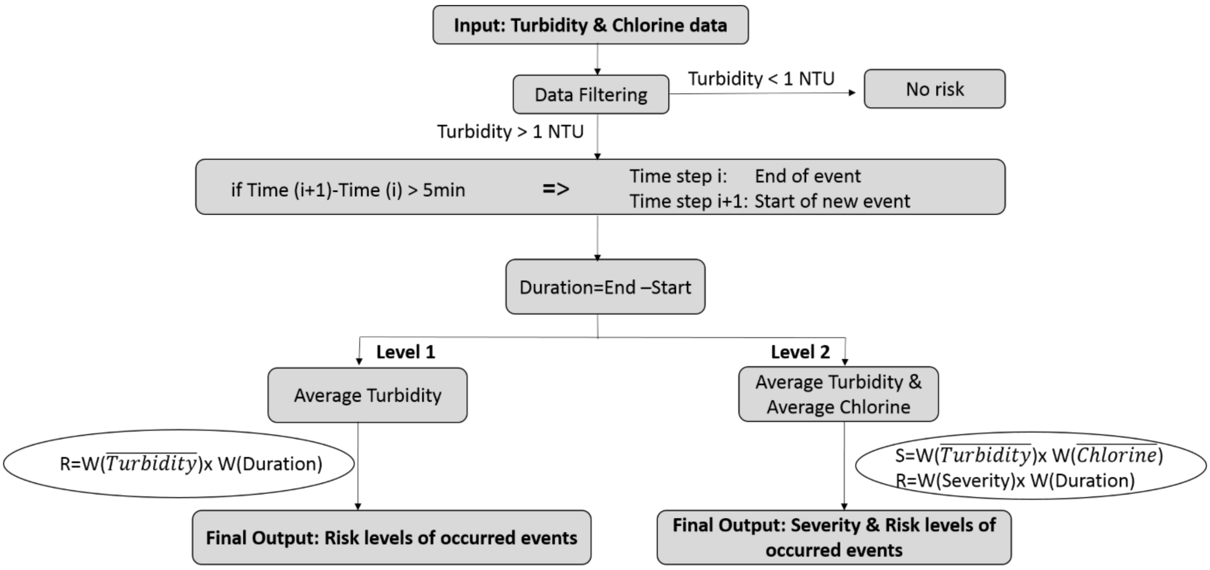


Figure 5.8. Flow chart of Approach 2.

5.4.5 Comparison between Approach 1 and Approach 2

By comparing the methodology of Approach 1 to Approach 2 (Levels 2), the main difference is the analysis of Turbidity parameter. In Approach 1, Turbidity data define the likelihood scale; Turbidity values and assigned weights are inversely proportional. However, in Approach 2, Turbidity values affect the severity of consequence; weights are attributed in the increasing order of Turbidity.

Table 5.17 illustrates an example of the classification of three events using the two approaches. Approach 1 indicates "Very High" risk for the three events. However, these events are classified, by Approach 2, as Moderate and Low risks. Figure 5.9 shows the variation of Turbidity and Chlorine during the first event. Turbidity remains around 1.01 NTU and Chlorine level around 0.001 mg/l. Although Chlorine concentration is very low, Turbidity value is very close to the limit (1 NTU). The identification of risk as "Moderate" is more precise.

For Event N° 2 and 3, a "Low" risk level seems to be more correct than a "Very High" risk. Despite the low concentration of Chlorine, Turbidity value is acceptable (near the limit). Such combinations should not generate an urgent attention, especially for an instantaneous event.

This comparison proves that Approach 1 over estimates the risk level. Approach 2 is more accurate in risk identification. The magnitude of deviation from the limit and the duration of event are the basic steps in risk assessment methods.

Table 5.17. Risk level comparison between Approach 1 and Approach 2.

Event N°	Start of Event	End of Event	$\overline{Turbidity}$ (NTU)	$\overline{Chlorine}$ (mg/l)	Risk Level	
					Approach 1	Approach 2
1	07/07/2017 22:48	07/07/2017 23:49	1,011	0,001	Very High	Moderate
2	07/07/2017 23:59	07/07/2017 23:59	1,001	0	Very High	Low
3	08/07/2017 13:15	08/07/2017 13:15	1,005	0,004	Very High	Low

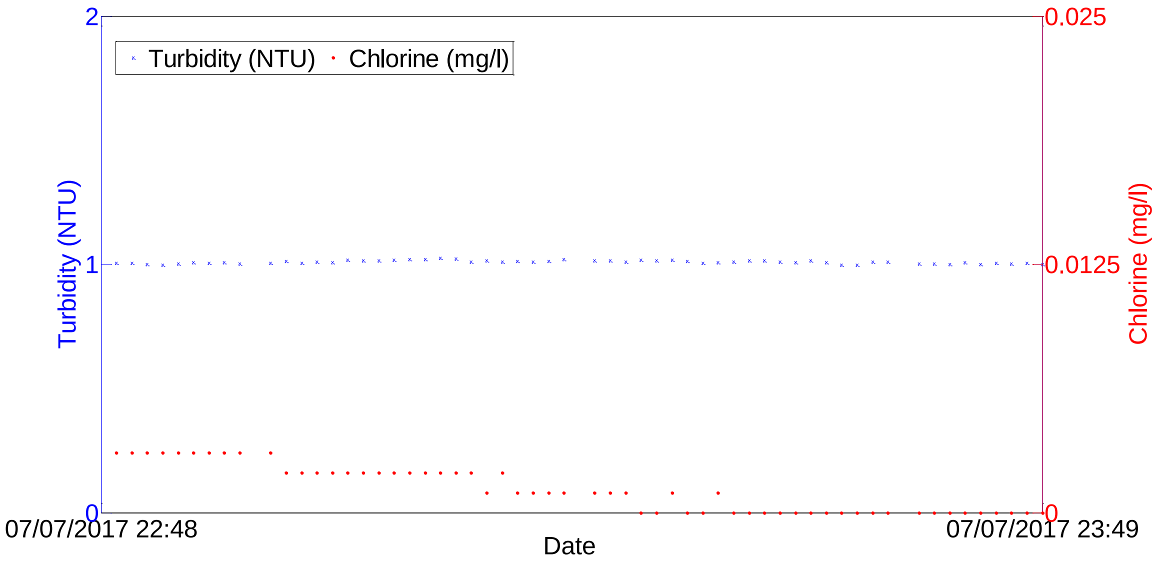


Figure 5.9. Variation of Turbidity and Chlorine on 07/07/2017 between 22:48 and 23:49.

5.4.6 Application of Approach 2 to Barroi Data

To verify the efficiency of Approach 2 over Approach 1 in risk identification, an application is done to S::CAN sensor installed at Barroi. Figure 5.10 illustrates the variation of Turbidity and Chlorine in Barroi restaurant during July, 2017. Two major events were observed during this period (On July, 7 and 31). These events are characterized by significant augmentation of Turbidity signals coupled with low amount of Chlorine.

The application of Approach 2 to these data is illustrated in Table 5.18. It shows the list of 11 events detected during this period, with the corresponding severity and risk levels. Table 5.18 indicates "High" risk on July, 7 and 31, 2017 which is verified in Figure 5.10. The detection of these two events and their classification as high risk prove the reliability of Approach 2 in the assessment of risk. Other events detected as "Low" risk should not generate an urgent alarm. Many reasons could be the source of such level of risks: connections issues, increase in the consumption, etc.

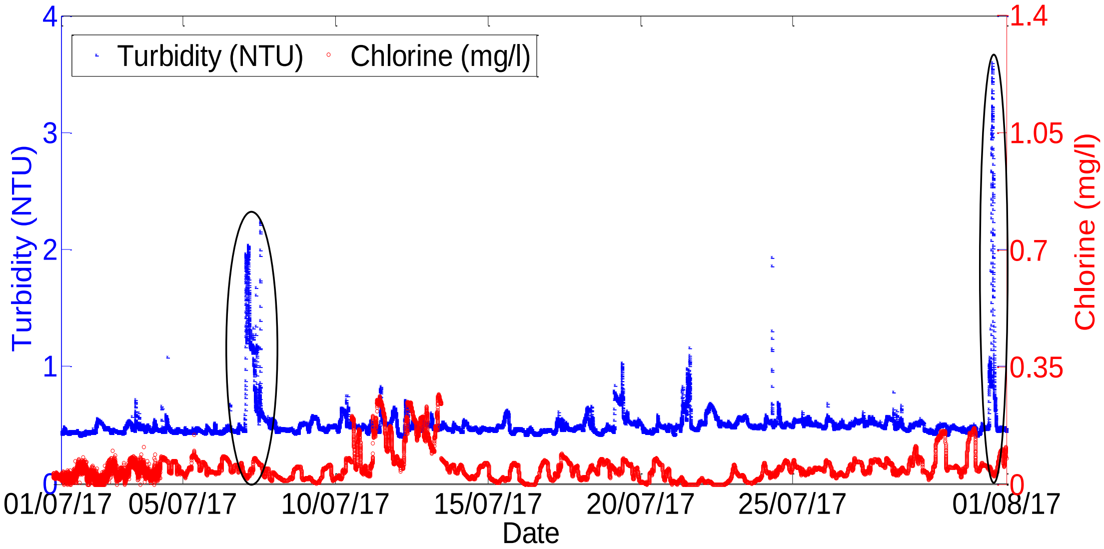


Figure 5.10. Variation of Turbidity and Chlorine during July, 2017 at Barroi restaurant.

Table 5.18. Risk level for S::CAN data at Barroi restaurant according to Approach 2.

Start of Event	End of Event	Duration (min)	Average Turbidity (NTU)	Average Chlorine (mg/l)	Severity Level	Risk Level
04/07/2017 11:21	04/07/2017 11:21	0	1,084	0,051	Insignificant	Low
07/07/2017 00:51	07/07/2017 07:32	401	1,413	0,039	Minor	High
07/07/2017 09:20	07/07/2017 09:55	35	1,185	0,039	Insignificant	Low
07/07/2017 12:23	07/07/2017 12:37	14	1,733	0,073	Minor	Low
19/07/2017 09:26	19/07/2017 09:31	5	1,022	0,029	Insignificant	Low
21/07/2017 14:22	21/07/2017 14:39	17	1,065	0	Minor	Low
24/07/2017 07:06	24/07/2017 07:15	9	1,439	0,032	Minor	Low
31/07/2017 10:08	31/07/2017 10:09	1	1,012	0,051	Insignificant	Low
31/07/2017 10:32	31/07/2017 10:47	15	1,032	0,043	Insignificant	Low
31/07/2017 11:39	31/07/2017 11:41	2	1,394	0,048	Minor	Low
31/07/2017 11:53	31/07/2017 13:57	124	2,423	0,033	Moderate	High

5.4.7 Approach 3 - Analysis of Turbidity and TOC

TOC gives an indication of the amount of organic matters in water. Also, the level of THM (Trihalomethanes), which is very harmful to human health, can be determined by the analysis of TOC. Standards define 2 mg/l as threshold limit for TOC.

In this section, the analysis of TOC and Turbidity is used in risk assessment approach. The severity of consequences is defined in function of Turbidity and TOC classification (Table 5.19). The first class, with weight 1, indicates safe drinking water. An event is detected if Turbidity exceeds 1 NTU or TOC exceeds 2 mg/l. As Turbidity and/or TOC increase, the event is more severe. Weights are assigned in the ascending order of Turbidity and TOC values.

Table 5.19. Description of the Severity of Consequences for Approach 3. (a) Turbidity classification; (b) TOC classification.

(a)

Description: Turbidity(NTU)	**Weight**
[0-1]	1
[1-1.2]	2
[1,2-3]	3
>3	4

(b)

Description: TOC (mg/l)	**Weight**
[0-2]	1
[2-2.5]	2
[2.5-4]	3
>4	4

At each time step, a severity score is calculated as follows:

$$Severity\ Score\ S = Weight\ (Turbidity)\ x\ Weight\ (TOC) \tag{5.8}$$

The severity score is then classified as indicated in Table 5.20, and given a corresponding weight. Firstly, data are filtered: Risk or not. During each identified risk, severity levels are classified according to their category (Insignificant, Minor, Moderate, and Major).

Table 5.20. Description of severity scale for Approach 3.

Severity Score	1	1<S<4	4≤S≤7	7≤S≤13	>13
Severity Scale	No risk	Insignificant	Minor	Moderate	Major
Weight	-	1	2	3	4

The duration for each category is calculated as follows:

$$Duration\ (\text{min}) = Date(End\ of\ severity\ level) - Date\ (Start\ of\ severity\ level) \tag{5.9}$$

The duration is classified as described in Table 5.9 (b). The risk score is calculated (as indicated in eqn. 5.6) and classified according to Table 5.10. Figure 5.11 illustrates the methodology of Approach 3 developed within VBA excel. An application is done for S::CAN data, at Polytech'Lille, for the period between July, 1 and 15, 2017. The variation of Turbidity and TOC is illustrated in Figure 5.12. Some events were detected during this period, especially on July, 6, 7, 8 and 10, 2017. In those days, a meaningful deviation of Turbidity and TOC from the reference line is observed.

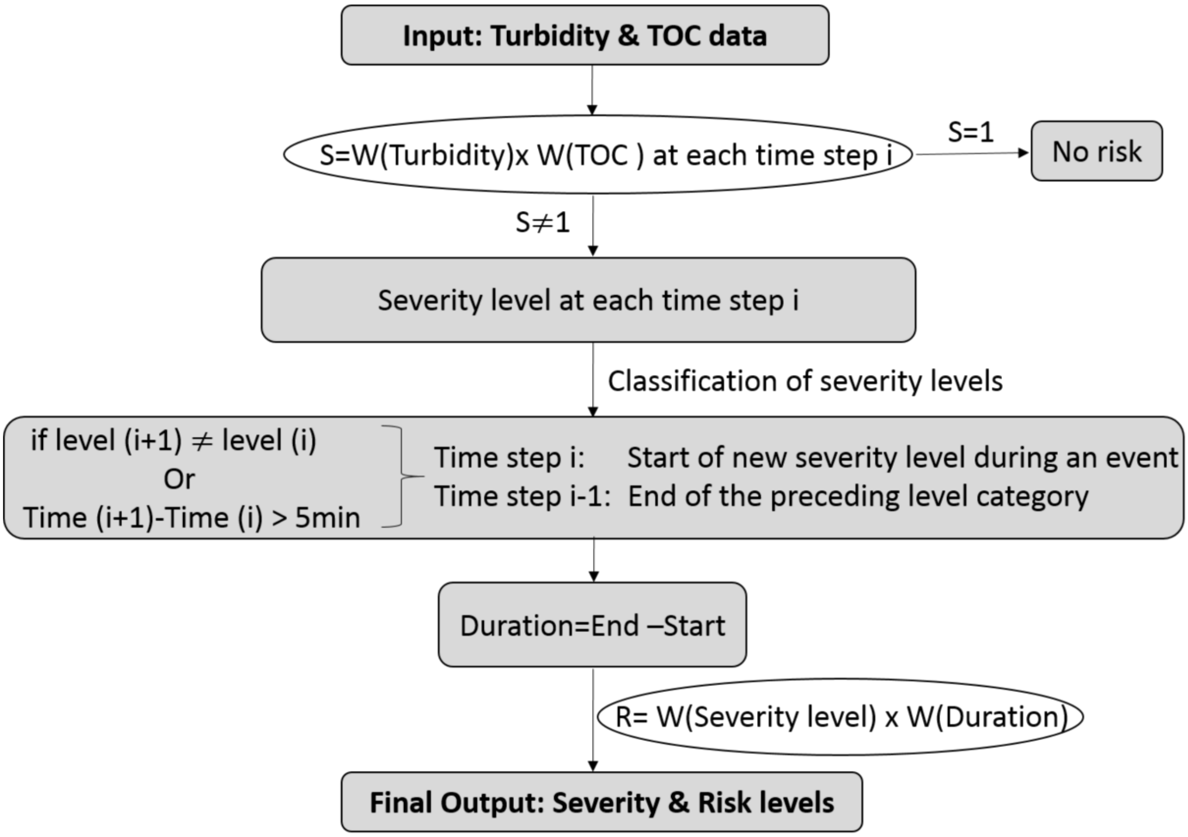


Figure 5.11. Flow chart of Approach 3.

An example of results, obtained by applying this approach, is given in Table 5.21 (Appendix A: Table A.5). The majority of events are classified as "Low" risk levels. Some "Moderate" risks are identified for example on July, 6 and 7, 2017, due to a significant increase of both signals (Turbidity and TOC).

In this approach, a weight of 2 is attributed for the second Class of Turbidity and/or TOC. Then, a small deviation from the threshold limits can induce "Minor" severity risk (S=4). However, value close to the limits can be ignored in the most cases, since it will not be very dangerous. This limitation proves that the analysis of the combination (Turbidity and Chlorine) is more accurate to estimate the magnitude of the risk and its severity. The analysis of TOC in risk assessment will be used as indication of the potential presence of organic contaminants in water.

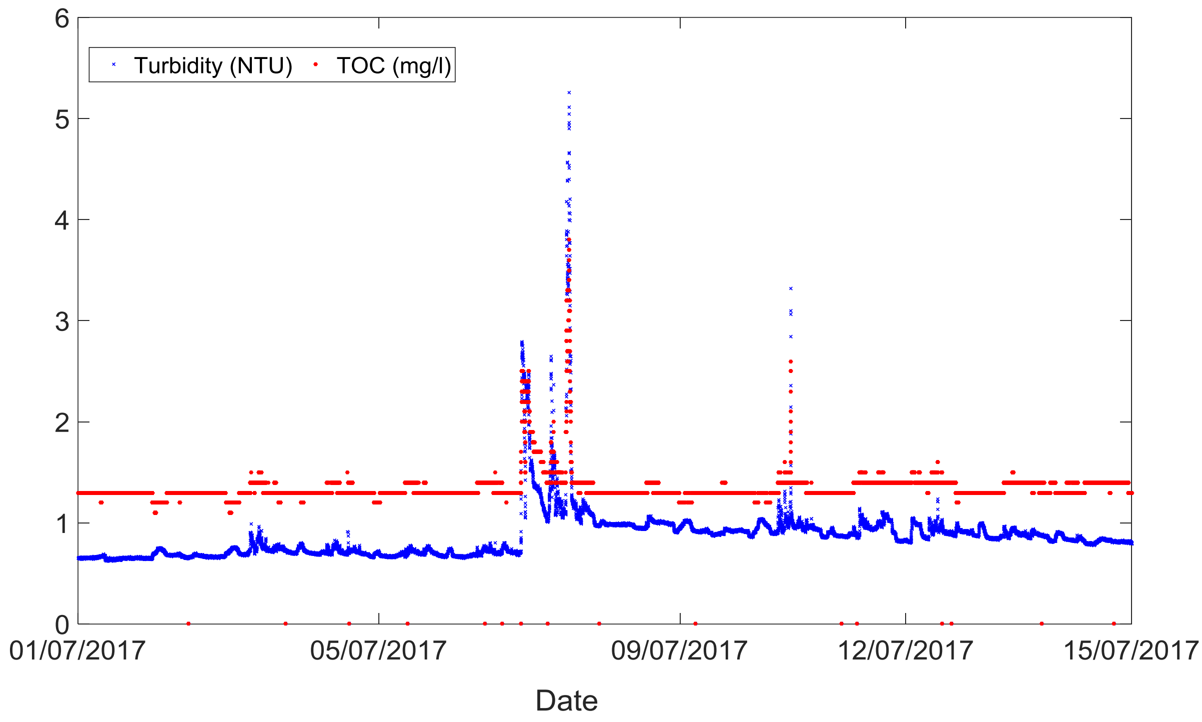


Figure 5.12. Variation of Turbidity and TOC, at Polytech'Lille, between July,1 and 15, 2017.

Table 5.21. Risk level for S::CAN data, at Polytech'Lille, according to Approach 3.

Start of new severity level	End of severity level	Severity level	Weight of severity	Duration (min)	Weight of Duration	Risk score	Risk level
06/07/2017 21:22	06/07/2017 21:25	Insignificant	1	3	1	1	Low
06/07/2017 21:26	06/07/2017 21:31	Minor	2	5	1	2	Low
06/07/2017 21:32	06/07/2017 21:47	Moderate	3	15	1	3	Low
06/07/2017 21:48	06/07/2017 21:54	Minor	2	6	1	2	Low
06/07/2017 21:55	06/07/2017 21:58	Moderate	3	3	1	3	Low
06/07/2017 21:59	06/07/2017 22:29	Minor	2	30	2	4	Moderate
06/07/2017 22:30	06/07/2017 22:46	Insignificant	1	16	1	1	Low
06/07/2017 22:47	06/07/2017 23:36	Minor	2	49	2	4	Moderate
06/07/2017 23:37	06/07/2017 23:42	Moderate	3	5	1	3	Low
06/07/2017 23:43	07/07/2017 00:08	Minor	2	25	1	2	Low
07/07/2017 00:09	07/07/2017 07:42	Insignificant	1	453	4	4	Moderate

5.5 Conclusion

The use of the smart technology for real-time control of water quality is recent. The limited feedback, in this domain, makes the decision-making difficult. Risk assessment and management constitutes a powerful tool for the analysis of water contamination risk. It allows to focus on significant risks. The combination between the smart monitoring and risk assessment approach provides an early identification of anomaly risk in real-time.

Qualitative assessment is appropriate for risk assessment of drinking water quality. It ranks potential risks into categories. This method was used for the risk assessment of the drinking water quality of the Scientific Campus of the University of Lille. Data are obtained from S::CAN sensor, which provides several water quality parameters (Turbidity, TOC, Conductivity, etc.).

Different strategies are proposed to describe the risk assessment parameters: Likelihood and Severity of Consequences in function of sensor data (Turbidity and Chlorine). Two main approaches are developed:

- Approach 1: describes the likelihood in function of Turbidity classes, while the severity of consequences is defined according to Chlorine levels.
- Approach 2: defines the likelihood in function of the duration of risk while Turbidity and Chlorine describe the severity of consequences of the risk.

Results show that the use of risk assessment approach with S::CAN data ensures the detection and classification of risks as well as the determination of priority levels required.

Analyses show that Approach 2 has a better performance in risk level estimation. It indicates that an accurate risk assessment should be based on: the magnitude of the measured parameter and the duration of the event.

A third approach is developed by analyzing the Turbidity and TOC signals. This approach estimates the possible existence of organic agents in water. However, Approach 2 remains more advantageous in risk identification.

The study conducted in this chapter shows encouraging results in the rapid detection of anomaly in water. However, research is still needed to collect more data concerning the abnormal events on water quality and their consequences on user's health.

Conclusion

Research conducted in this thesis concern the real-time control of the water quality in the Water Distribution Network (WDN) using the smart technology. The traditional laboratory-based methods taking generally long time, do not ensure the required protection of human health against accidental or malicious contamination. There is a clear need to turn this water system into a smart system that can detect rapidly any intrusion in water. The complexity and the extent dimension of the WDN makes the early detection of water contamination difficult. To meet this challenge, the deployment of the smart technology is relevant. However, the feedback is limited in this domain, because the implementation of this technology in the WDN is recent. Consequently, the water industry is interested by research works based on large-scale experimentation in this area. This works constitutes a contribution to this industrial need. It presents feedback from a large-scale experimentation of online water quality supervision.

The Scientific Campus of the University of Lille is used to investigate the real-time monitoring of the water quality within SunRise Large-Scale Demonstrator of the Smart City. This work is also a part of the European project SmartWater4Europe “SW4EU” which aims at establishing 4 demonstrator sites of the smart water networks in Europe.

The online control of the water quality implies the use of innovative systems, especially water quality sensors. They record continuously parameters generally affected by water contamination. In this work, two types of water quality devices (S::CAN and EventLab) are implemented at two locations in the campus (Polytech’Lille University and Barroi restaurant). S::CAN sensor measures continuously several water quality parameters (Turbidity, Conductivity, pH, etc.) while EventLab controls, each minute, the change in phase, which is proportional to the variation of the refractive index.

A detailed analysis of sensors’ data is presented in this work. The correlation matrix of S::CAN parameters showed the good operation of this device. Collected data are validated through the comparison with laboratory results, however some adjustments showed a need for regular control of this device. Although the majority of signals are quasi-stable, some events have been detected and characterized by deviation from baseline values. A good correlation is observed between the peaks in the water quality parameters and the variation in hydraulic parameters (pressure and/or consumption). Some abnormalities have been attributed to the intervention on the network (such as repair actions). In some cases, suspended materials are extracted from the aging water pipes, which induces a change in water quality. For EventLab sensor, the experimentation showed the need of regular maintenance. Analysis of the phase variation allowed to detect some abnormalities in the water quality. However, S::CAN showed higher performances in the identification of water anomaly.

Different detection methodologies have been adapted to the smart water system and tested on S::CAN data. The use of the Statistical and Artificial Intelligence (Support Vector Machine) methods resulted in the classification of S::CAN measurements between normal values and unexpected data. However, a single outlier (due generally to sensor mal function, connection issues, etc.) should not generate an alarm. Therefore, an Event Detection System (EDS) has been developed within Canary software to identify anomaly induced potentially by change in the water quality. The system triggers an alert if a number of outliers in a precise duration exceeds a predefined threshold. A sensitivity analysis allowed to get the highest sensibility of detection with low rate of false alarms.

To focus on significant risks and thus take effective corrective measures, a combination between the risk assessment methodology and the online monitoring has been proposed. Two main approaches have been developed where water quality parameters describe the likelihood and the severity of consequences of risk. The application of these approaches to S::CAN data showed that the magnitude of the measured parameter as well as the duration of the corresponding event should be taken into consideration for accurate risk identification. The improved method of risk assessment allows an early detection of the water anomaly risk in near real-time. It indicates also the level of risk as well as the priority attention required.

The experimentation conducted in this thesis provides a thorough feedback about the use of the smart technology (smart sensors and data analysis) for the real-time control of the drinking water quality in the distribution networks. It indicates a good reliability of these devices to detect abnormal events in the water system. Detection methods used and developed in this research, showed good performances of these methods for the early identification of water anomaly. However, this promising result requires additional effort to improve our knowledge about these innovative devices. It will be interesting to enable an automated sampling in case of abnormal events detected by sensors. Advanced methods should also be developed to enhance the response strategy of water utilities to significant change in the water quality.

References

[1] Arasmith, S., Introduction to water distribution systems, chapter 5, *Introduction to small water systems, ACR publications, American Water Works Association*, pp. 147-191, 2005. https://dec.alaska.gov/water/opcert/Docs/Chapter5.pdf.

[2] Mounce, S., et al., Water quality event detection and customer complaint clustering analysis in distribution systems. *Water Science and Technology: Water Supply*, **12**(5), pp. 580-587, 2012.

[3] Levi, Y., Les paramètres influençant le développement des germes dans les réseaux d'eau potable, *Technique Science et Méthode*, **3**, pp. 240-245, 1995.

[4] Armon, R., et al. Bacteriophage as indicators of pollution. *Critical Reviews in Environmental Sciences and Technology*, **26**(4), pp. 299-335, 1996.

[5] Abdallah, A., Réseaux d'eau intelligents : surveillance de la qualité de l'eau par des capteurs en ligne, Doctoral dissertation, Lille1, 2015.

[6] E26-Analyse et qualité de l'eau. Principales normes et indicateurs de potabilité et qualité de l'eau. http://www.wikiwater.fr/e26-analyse-et-qualite-de-l-eau.html.

[7] Dossier d'information, La qualité de l'eau potable en France: Aspects sanitaires et réglementaires, Annexes, FICHE 7: LES EXIGENCES DE QUALITE (NORMES DE QUALITE). *Direction générale de la santé, Ministère de la Santé et des Solidarités*, Septembre 2005.

[8] AFNOR (Association Française de Normalisation), Eaux-méthodes d'essais. *Recueil de Normes Françaises, 4ème édition*, La Défense, Paris, pp. 735, 1990.

[9] Joly, B., et al., Entérobactéries : systématique ct méthodes de diagnostic. *TEC & DOC (Ed.), Lavoisier*, Paris, pp. 356. 2003.

[10] Aubril-Carhon, L., Ingénieur environnement, Eau potable: Normes, 2013. http://www.wikiterritorial.cnfpt.fr/xwiki/wiki/econnaissances/view/Notions-Cles/Leaupotablelesnormes.

[11] SENAT, L'arsenic dans les eaux de boisson, Annexe 63, La qualité de l'eau et assainissement en France (annexes), 2017. https://www.senat.fr/rap/l02-215-2/l02-215-256.html.

[12] LENNTECH, Impact de l'ammonium sur la santé et l'environnement. http://www.lenntech.fr/francais/ammonium-environnement.htm.

[13] Water Information Center, De l'aluminium dans l'eau, Les thématiques de santé, September 2014. http://www.cieau.com/les-thematiques-de-sante.

[14] McLachlan, D. R. C., et al., Risk for neuropathologically confirmed Alzheimer's disease and residual aluminum in municipal drinking water employing weighted residential histories. *Neurology* **46**(2) pp.401-405, February 1996.

[15] World Health Organization, Sodium in Drinking-water, Background document for development of WHO *Guidelines for Drinking-water Quality*, *Health criteria and other supporting information*, 2nd edition, **2**, 2003.

[16] World Health Organization, Copper in Drinking-water, Background document for development of WHO *Guidelines for Drinking-water Quality*, 2004.

[17] Toutant, C., et al., Techniques=L'Absorbance à 254 nm, *média terre*, https://www.mediaterre.org/eau/actu,20130703100557,1.html.

[18] World Health Organization, Water sampling and analysis, *Guidelines for drinking-water quality*, Surveillance and community supplies, 2nd edition, **3**, Geneva, 1997.

[19] Servais, P., et al., Comparison of the bacterial dynamics in various French distribution systems, *Journal Water SRT-Aqua*, **44**(1), pp.10-17, 1995.

[20] Celerier., Jean-Luc, et al., La dégradation de la qualité de l'eau potable dans les réseaux, Document technique, Ministère de l'agriculture, de l'alimentation, de la pêche et des affaires rurales, pp. 98, 2003.

[21] Gauthier, F., Biofilms et qualité biologique de l'eau potable au cours de sa distribution, *Université d'Amiens*, 2002.

[22] Bouguelia, S., Développement de biopuces dédiées à la détection de bactéries pathogènes à faibles taux, Doctoral dissertation, Grenoble, 2012.

[23] Rompré, A., et al., Detection and enumeration of coliforms in drinking water: current methods and emerging approaches, *Journal of microbiological methods*, **49**(1), pp. 31-54, 2002.

[24] Toranzos, G. A., et al., Detection of microorganisms in environmental freshwaters and drinking waters, *Manual of Environmental Microbiology, Third Edition*, American Society of Microbiology, pp. 249-264, 2007.

[25] Lemarchand, K., et al., Molecular biology and DNA microarray technology for microbial quality monitoring of water. *Critical reviews in microbiology*, **30**(3), pp. 145-172, 2004.

[26] Ashbolt, N.J., et al., Indicators of microbial water quality. In: Fewtrell L., Bartram J. (Eds.), Water quality: Guidelines, standards and health, Assessment of risk and risk management for water-related infectious disease, *IWA Publishing*, London, UK, pp. 289-316, 2001.
[27] Alhamlan, F. S., et al., Recommended advanced techniques for waterborne pathogen detection in developing countries, *The Journal of Infection in Developing Countries*, **9**(02), pp.128-135, 2015.
[28] Schloter, M., et al., The use of immunological methods to detect and identify bacteria in the environment, *Biotechnology advances*, **13**(1), pp. 75-90, 1995.
[29] Ramírez-Castillo, F. Y., et al., (2015). Waterborne pathogens: detection methods and challenges, *Pathogens*, **4**(2), pp.307-334, 2015.
[30] Hibi, K., Abe, et al., Combination of immunomagnetic separation with flow cytometry for detection of Listeria monocytogenes, *Analytica chimica acta*, 573, pp. 158-163, 2006.
[31] Shelton, D.R., et al., Estimation of viable Escherichia coli O157 in surface waters using enrichment in conjunction with immunological detection, *Journal of Microbiological Methods*, **58** (2), pp. 223-231, August 2004.
[32] Keserue, H. A., et al., Rapid detection of total and viable Legionella pneumophila in tap water by immunomagnetic separation, double fluorescent staining and flow cytometry, *Microbial biotechnology*, **5**(6), pp. 753-763, 2012.
[33] Rodriguez-Mozaz, S., et al., Advantages and limitations of on-line solid phase extraction coupled to liquid chromatography–mass spectrometry technologies versus biosensors for monitoring of emerging contaminants in water, *Journal of Chromatography A*, **1152**(1), pp. 97-115, 2007.
[34] Chambers, J. P., et al., Biosensor recognition elements, *TEXAS UNIV AT SAN ANTONIO DEPT OF BIOLOGY*, 2008.
[35] Lee, J.Y., et al., ATP Bioluminescence for Monitoring Drinking Water Quality, *School of Public Health, University of Michigan*, August 2005.
[36] Ochromowicz, K., et al., ATP as an indicator of microbiological activity in tap water, *European Commission Directorate–General Joint Research Centre*, 2005.
[37] Spectroscopy: Principles, Theory, Techniques and Applications, May 2009. https://edurev.in/studytube/Spectroscopy-Principles,Theory,-Techniques-and-Applications/d0bb4b0e-97ec-4e41-8591-0bd850bf23af_p.
[38] Zarnowiec, P., et al., Fourier transform infrared spectroscopy (FTIR) as a tool for the identification and differentiation of pathogenic bacteria, *Current medicinal chemistry*, **22**(14), pp. 1710-1718, 2015.
[39] Al-Qadiri, H. M., et al., Fourier transform infrared spectroscopy, detection and identification of Escherichia coli O157: H7 and Alicyclobacillus strains in apple juice, *International Journal of Food Microbiology*, **111**(1), pp. 73-80, 2006.
[40] Harz, M., et al., Vibrational spectroscopy—A powerful tool for the rapid identification of microbial cells at the single-cell level, *Cytometry Part A*, **75**(2), pp. 104-113, 2009.
[41] Bin Omar, A. F., et al., Turbidimeter design and analysis: a review on optical fiber sensors for the measurement of water turbidity, *Sensors*, **9**(10), pp. 8311-8335, 2009.
[42] Introduction to drinking water quality testing, *Centre for Affordable Water and Sanitation Technology*, A Cawst training manual, June 2009.
[43] WHO/UNICEF, Drinking Water: Equity, Safety and Sustainability. *World Health Organization*, Geneva, 2011.
[44] WHO/UNICEF, Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines, *World Health Organization*, Geneva, 2017.
[45] WHO/UNICEF, Joint Water Supply, & Sanitation Monitoring Programme, *Progress on drinking water and sanitation: 2014 update*, *World Health Organization*, 2014.
[46] World Health Organization and UNICEF (2014) Progress on drinking water and sanitation. Joint Monitoring Programme update. Available: http://www.wssinfo.org/fileadmin/user_upload/resources/JMP_report_2014_webEng.pdf. Accessed on April 18, 2014.
[47] WHO, Guidelines for Drinking-water Quality, Third Edition Incorporating the First and Second Addenda – Volume 1: Recommendations, Geneva: *World Health Organization*, 2008.
[48] Carter, J., et al., Virology: Principles and Applications, second ed., *JohnWiley & Sons, UK*, pp. 395, 2013.
[49] WHO, Guidelines for drinking-water quality, Fourth Edition, Geneva, Switzerland: *WHO Press*, 2011.
[50] NHS choices, Norovirus, Queen's Printer and Controller of HMSO, 2008. http://www.nhs.uk/translationfrench/Documents/Norovirus_French_FINAL.pdf.
[51] UNICEF, Pneumonia and diarrhoea: Tackling the deadliest diseases for the world's poorest children, *New York, NY: Statistics and Monitoring Section- Division of Policy and Strategy*, 2012.
[52] French Institute for Public Health Surveillance, Assessment of waterborne outbreaks investigated in France since 1998 and key recommendations, *Infectious diseases Health and environment, Summary Report*, August, 2008.

[53] UFC-Que Choisir, La qualité de l'eau en France : à vau l'eau ?, February, 2014. http://image.quechoisir.org/var/ezflow_site/storage/original/application/9d0a442589d4d349c83a287772097b85.pdf.
[54] Deshayes, F., et al., Pollution du Réseau d'Eau Potable à Strasbourg et Survenue Concomitante de Gastro-Entérites—Mai 2000, *Bull. Épidémiologique Hebdomadaire*, **2**(1), 2001.
[55] Esvan, M., et al., Contamination d'un réseau d'eau potable par de l'éthylène glycol dans un immeuble, Rennes (France), janvier 2011, *BEH (49-50)*, pp. 526-529, 2012.
[56] Agence Régionale de Santé, projet régionale de santé, Rhône- Alpes 2012-2017, L'eau du robinet, ça se protège !, *Communiqué de presse*, May, 2014. https://www.feurs.org/images/stories/pdf/20140523_ARSRA_CP_RetoursEau.pdf
[57] Gofti-Laroche, L., et al., Epidémie de gastro-entérites liée à la pollution du réseau de distribution d'eau potable de la commune de divonne-les-Bains. *CIRE Rhône-Alpes-Auvergne*, 2003.
[58] Sensus, Water 20/20: Bringing Smart Water Networks into Focus, *Report by Sensus*, 2012.
[59] Martyusheva, O., Smart Water Grid, plan B, Technical report, Department of civil and environmental fort Collins, Colorado, 2014.
[60] Miller, J.M, et al., Why Smart Water Network Boost Efficiency, Schneider Electric White Paper. http://www.wwdmag.com/sites/wwdmag.com/files/998-2095-08-06-14AR0_EN.pdf
[61] Lambrou, T. P., et al., A low-cost sensor network for real-time monitoring and contamination detection in drinking water distribution systems. *IEEE sensors journal*, **14**(8), pp. 2765-2772, 2014.
[62] Che, H., et al., Contaminant detection using multiple conventional water quality sensors in an early warning system. *Procedia Engineering*, vol. 89, pp. 479-487, 2014.
[63] Yang, Y. J., et al., Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results. *Journal of environmental management*, **90**(8), pp. 2494-2506, 2009.
[64] Hasan, J., et al., Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: A state-of-the-art review (No. EPA/600/R-05/156), *ENVIRONMENTAL PROTECTION AGENCY WASHINGTON DC OFFICE OF WATER*, 2005.
[65] Storey, M. V., et al., Advances in on-line drinking water quality monitoring and early warning systems, *Water Research*, **45**(2), pp. 741-747, 2011.
[66] EWS Initial Study, Water quality early warning system on transboundary watercourses of the Tisza river basin, *Veszprémi Regionalis Innovacios Centrum & Environmental Institute*, 2014. http://www.danubewaterquality.eu/uploads/mod_files/WQM-EWS_part-1-4_EN_v2.0_.pdf
[67] Grayman, Walter M, Contamination of water distribution systems, In *International Seminar on Nuclear War and Planetary Emergencies—45th Session: The Role of Science in the Third Millennium*, World Scientific, pp. 389-398 June, 2013.
[68] USEPA, Water security initiative: Interim guidance on planning for contamination warning system deployment, EPA-817-R-07-002, *Office of Water, Office of Ground Water and Drinking Water*, U.S. EPA, Washington, D.C, 2007.
[69] Skadsen, J., et al,. Distribution System Monitoring For Detecting Contamination Events And Water Quality, J.AWWA. **100**(7), pp. 81-94, 2008.
[70] Hou, D., et al., Detection of water-quality contamination events based on multi-sensor fusion using an extended Dempster-Shafer method. IOP Publishing, *Measurement Science and Technology*, **24**(5), pp.18, 2013.
[71] McKenna, S. A., et al., Testing water quality change detection algorithms. In *Water Distribution Systems Analysis Symposium 2006*, pp. 1-15, 2008.
[72] Murray, R., et al., Water quality event detection systems for drinking water contamination systems, Development, Testing, and Application of Canary, *National Homeland Security Research Center*, *Office of Research and Development*, U.S. Environmental Protection Agency, CINCINNATI, OH 45268, 2010.
[73] Taylor, H.M., et al., An Introduction to Stochastic Modeling, *3rd Edition*, Academic Press, pp. 631, 1998.
[74] Raciti, M., et al., S.Anomaly detection in water management systems. In *Critical infrastructure protection*, Springer Berlin Heidelberg, pp. 98-119, 2012.
[75] Maier, H.R. et al., The use of artificial neural networks for the prediction of water quality parameters. *Water Resour Res*, **32**(4), pp.1013-1022, 1996.
[76] Burke, L. I, Introduction to artificial neural systems for pattern recognition. *Computers & Operations Research*, **18**(2), pp. 211-220, 1991.
[77] Dandy, G. C, Comparison of the Box-Jenkins procedure with artificial neural network methods for univariate time series modelling, Department of Civil and Environmental Engineering, The University of Adelaide, Research Report No.R 127,1995.
[78] Apprentissage automatique: les réseaux de neurones, Chapitre 3. http://www.grappa.univ-lille3.fr/polys/apprentissage/sortie005.html.

[79] Bouamar, M., et al., Système multicapteur utilisant les réseaux de neurones artificiels pour la surveillance des eaux potables. In *4th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, LASS, Laboratoire d'Analyse des Signaux et Systèmes, Université de M'sila, Algérie*, pp. 25-29, March, 2007.
[80] Law, M., A simple introduction to support vector machines. *Lecture for CSE*, *802*, Department of Computer Science and Engineering, Michigan State University, 2006.
[81] Djeffal, A., Utilisation des méthodes Support Vector Machine (SVM) dans l'analyse des bases de données, Doctoral dissertation, Université Mohamed Khider-Biskra, 2012.
[82] Bouamar, M., et al., La Technique SVM Appliquée à la Surveillance des Eaux Potables, *5th International Conference : Sciences of Electronic, Technologies of Information and Telecommunications, LASS, Laboratoire d'Analyse des Signaux et Systèmes, Université de M'sila, Algérie*, pp. 22-26, March, 2009.
[83] Berson, A., et al., An overview of data mining techniques. *Building Data Mining Application for CRM*, 2004.
[84] Hartigan, J. A., et al., Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**(1), pp. 100-108, 1979.
[85] Farah, E., et al., Sunrise: Large Scale Demonstrator of the Smart Water System. *International Journal of Sustainable Development and Planning*, **12**(1), pp. 112-121, 2017.
[86] Likas, A., et al., The global k-means clustering algorithm. *Pattern recognition*, **36**(2), pp. 451-461, 2003.
[87] Campagn, F., et al., Vigires' eau: système de surveillance en temps réel de la qualité de l'eau potable d'un réseau de distribution en vue de la détection d'intrusion. In *Workshop Interdisciplinaire sur la Sécurité Globale (WISG'13)*, pp. 1-8, 2013.
[88] Lerminiaux, C., Ellipse, Le magazine de l'Université de technologie de Troyes, N° 5, Printemps, 2010. http://www.utt.fr/_resources/pdf%20généraux/ellipse5.pdf.
[89] Fillatre, L., et al., Vigires' eau: Surveillance en temps réel de la qualité de l'eau potable d'un réseau de distribution en vue de la détection d'intrusions, In *Workshop Interdisciplinaire sur la Sécurité Globale (WISG'11), (ANR-CSOSG)*, 2011.
[90] Bernard, T., et al., SMaRT-Online WDN: Online Security Management and Reliability Toolkit for Water Distribution Networks, In *Proceedings of the 10th International ISCRAM Conference Baden-Baden, Germany*, May, 2013.
[91] SMaRT-Online WDN, Online Security Management and Reliability Toolkit for Water Distribution Networks, 2012.
http://www.smart-onlinewdn.eu/objectives.
[92] LESCA, La spectroscopie proche infrarouge: une technique d'imagerie trop peu connu !, Laboratoire d'Etude de la Santé Cognitive des Aines, 2015.
http://www.lesca.ca/2015/03/11/la-spectroscopie-proche-infrarouge-une-technique-dimagerie-trop-peu-connu/.
[93] Farah, E., Detection of water leakage using innovative smart water system: application to SunRise Smart City demonstrator, Doctoral dissertation, Lille 1, 2016.
[94] Manual micro::station, s::can, Intelligent. Optical. Online. Technical Description, *by s::can Messtechnik GmbH, Vienna*, June 2011. www.s-can.at
[95] Micro::station, Monitoring Stations.
http://www.s-can.at/medialibrary/datasheets/microstation_dw_hp.pdf.
[96] http://www.s-can.at.
[97] Manual i::scan V1, s::can, Intelligent. Optical. Online, *by s::can Messtechnik GmbH, Vienna*, February 2014. www.s-can.at.
[98] Manual pH::lyser/redo::lyser V2, s::can, Intelligent. Optical. Online, *by s::can Messtechnik GmbH, Vienna*, January 2010. www.s-can.at.
[99] Manual condu::lyser V2, s::can, Intelligent. Optical. Online, *by s::can Messtechnik GmbH, Vienna*, November 2010. www.s-can.at.
[100] Manual chlori::lyser V1, s::can, Intelligent. Optical. Online, *by s::can Messtechnik GmbH, Vienna*, August 2015. www.s-can.at.
[101] Building PI System Assets and Analytics with AF, *OSIsoft*, version 2016a.
[102] Optiqua EventLab, Real time on-line water quality monitoring. www.optiqua.com.
[103] Wijlen, M.A.B van. et al., Innovative sensor technology for effective online water quality monitoring. *Proceedings of the Singapore International Water Week*, 2011.
[104] Broeke, J. van den., The benefits of using refractive index for water quality monitoring in distribution network, A comparison of Optiqua EventLab and traditional water quality parameters, Optiqua, Published by *Optiqua Technologies*, www.optiqua.com.

[105] Raich, J., Review of sensors to monitor water quality, *Joint Research Centre Institute for the Protection and Security of the Citizen (IPSC)*, Commission EUR 26325 EN, Luxembourg: Publications Office of the European Union, 33pp, December 2013.
[106] Tangena, B.H., et al., A novel Approach for Early Warning of Drinking Water Contamination Events, from *Water Contamination Emergencies: Monitoring, Understanding and Acting*, pp.13-31, 2011.
[107] Optiqua, Refractive index as the new water quality monitoring standard, Mach Zehnder Interferometer. http://www.optiqua.com/mach_zehnder.html.
[108] Optiqua EventLab 2.0, User Manual, Monitoring Water Quality, *Optiqua Technologies*, June 2016. www.optiqua.com.
[109] Optiqua EventLab, User Manual, 09.2012. www.optiqua.com.
[110] Hall, J., Zaffiro, et al., On-line water quality parameters as indicators of distribution system contamination. *Journal (American Water Works Association)*, **99**(1), pp.66-77, 2007.
[111] Eurofins Hydrologie Nord SAS, https://www.eurofins.fr/env.
[112] Pearson's correlation. http://www.statstutor.ac.uk/resources/uploaded/pearsons.pdf
[113] Wolf. A.V., et al., Concentration properties of aqueous solutions, *Handbook Of Chemistry and Physics, CRS Press*, 65th edition, 1984-1985.
[114] Saw, J.G., et al., Chebyshev Inequality with Estimated Mean and Variance. *The American Statistician*, **38**(2), pp. 130-132, 1984.
[115] Vaseghi, S. V. Advanced digital signal processing and noise reduction. *John Wiley & Sons*, 2008.
[116] Hasan, M., et al., Svm: Machines à vecteurs de support ou séparateurs à vastes marges. *Rapport technique, Versailles St Quentin, France. Cité*, p. 64, 2006.
[117] Shiges, A., Analysis of support vector machines, *Neural Networks for Signal Processing, Proceedings of the 12 th IEEE Workshop*, pp. 89-98, 4-6 September, 2002.
[118] Schklopf, B., et al., Learning with kernels, Support vector machines, Regularization, Optimization and Beyond, *MIT Press*, Cambridge, MA, 2002.
[119] Reyna-Rojas, R. A. Conception et intégration VLSI d'un système de vision générique. Application à la détection et à la localisation d'objets à l'aide de" Support vector machines*" Doctoral dissertation, Toulouse, INSA*, 2002.
[120] AWWA. Contamination warning systems for water: an approach for providing actionable information to decision-makers, *American Water Works Association*, Denver, CO, 2005.
[121] Murray, R., et al., Water quality event detection systems for drinking water contamination warning systems—development, testing, and application of CANARY. *EPAI600IR-lOI036, US*, 2010.
[122] U.S. EPA. Configuring Online Monitoring Event Detection Systems. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-14/254, 2014.
[123] Mounce, S., et al., Water quality event detection and customer complaint clustering analysis in distribution systems. *Water Science and Technology: Water Supply*, **12** (5), pp. 580-587, 2012.
[124] McKenna, S.A., et al., CANARY: Event Detection Software, *Sandia National Laboratories*, R&D 100, 2010. http://www.sandia.gov/research/research_development_100_awards/_assets/documents/2010_winners/SNL_Canary_SAND2010-2228P.pdf.
[125] Saab, C., et al., Field Study of real-time water quality control. *WIT Transactions on Ecology and the Environment*, **216**, pp. 237-248, 2017.
[126] U.S. EPA. CANARY Training Tutorials. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-13/201, 2013.
[127] Bradley, A. P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, **30(**7), pp. 1145-1159, 1997.
[128] U.S. EPA. CANARY User's Manual version 4.3.2. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-08/040B, 2012.
[129] NHMRC, NRMMC, et al. Australian drinking water guidelines paper 6 national water quality management strategy. National Health and Medical Research Council, National Resource Management Ministerial Council, Commonwealth of Australia, Canberra, 2011, Version 3.4 Updated October, 2017.
[130] Apsari, M.N. et al., Analisis risiko dan optimasi kualitas air produksi Instalasi Pengolahan Air Ngagel I. Surabaya*:* Department of Environmental Engineering ITS, 2014.
[131] Water Quality Protection Note. Risk Assessment of Public Drinking Water Source Areas. *WQ 77, Government of Australia Western*, Department of Environment. Risk Assessment of PDWSAs (Public Drinking Water Source Areas), June, 2005. https://www.water.wa.gov.au/__data/assets/pdf_file/0011/4115/59829.pdf.
[132] Percival, S. L., et al., Microbiological aspects of biofilms and drinking water. *CRC Press*, 2000.
[133] Last, J. M., et al., (Eds.). A dictionary of epidemiology, Vol. 4. New York: *Oxford university press*, 2001.
[134] Davison, A., et al., Water safety plans: managing drinking-water quality from catchment to consumer. Water, Sanitation and Health, Protection and the Human Environment, World Health Organization, Geneva, 2005.

[135] Dufour, A. et al., Assessing Microbial Safety of Drinking Water, Improving Approaches and Methods, Published on behalf of the *World Health Organization and the Organization for Economic Co-operation and Development by IWA Publishing*. http://www.who.int/water_sanitation_health/dwq/9241546301full.pdf.

[136] Thoeye, C., et al., Methods used for health risk assessment. State of the art report health risks in aquifer recharge using reclaimed water. World Health Organization, Copenhagen, Denmark, pp. 17-53, 2003.

[137] Petterson, S., et al., QMRA methodology. Microrisk report April, 2006.

[138] National Research Council, Risk Assessment in the Federal Government: Managing the process. *National Academy Press*, Washington, DC, 1983.

[139] Machdar, E, et al., Application of quantitative microbial risk assessment to analyse the public health risk from poor drinking water quality in a low income area in Accra, Ghana. *Science of the Total Environment*, **449**, pp. 134-142, 2013.

[140] Haas, C. N., et al., Quantitative microbial risk assessment. *John Wiley & Sons*, 1999.

[141] Rosén, L., et al., Generic framework and methods for integrated risk management in water safety plans. Deliverable number D, **4**(3), 2007.

[142] World Health Organization, WHO human health risk assessment toolkit: chemical hazards, 2010. http://www.inchem.org/documents/harmproj/harmproj/harmproj8.pdf.

[143] Exposure assessment: http://www.ce.utexas.edu/prof/maidment/risk/rskchar3/tsld019.htm.

[144] Niedbalski, A.S, et al, Risk assessment in drinking water supplies of Sweden and Latvia. An overview within the Water Safety Plan framework. *Master of Science Thesis in the Master's Programme Infrastructure and Environmental Engineering*, Department of Civil and Environmental Engineering, Division of Water Environment Technology, CHALMERS University of Technology, Sweden, 2015.

[145] Davison A., et al., Water safety plans, WHO/SHE/WSH/02/09/ World Health Organization, Geneva, Switzerland, 2002.

[146] Bartram, J., Water safety plan manual: step-by-step risk management for drinking-water suppliers. World Health Organization, 2009.

Appendix A. Detailed Results of Risk Assessment Approaches

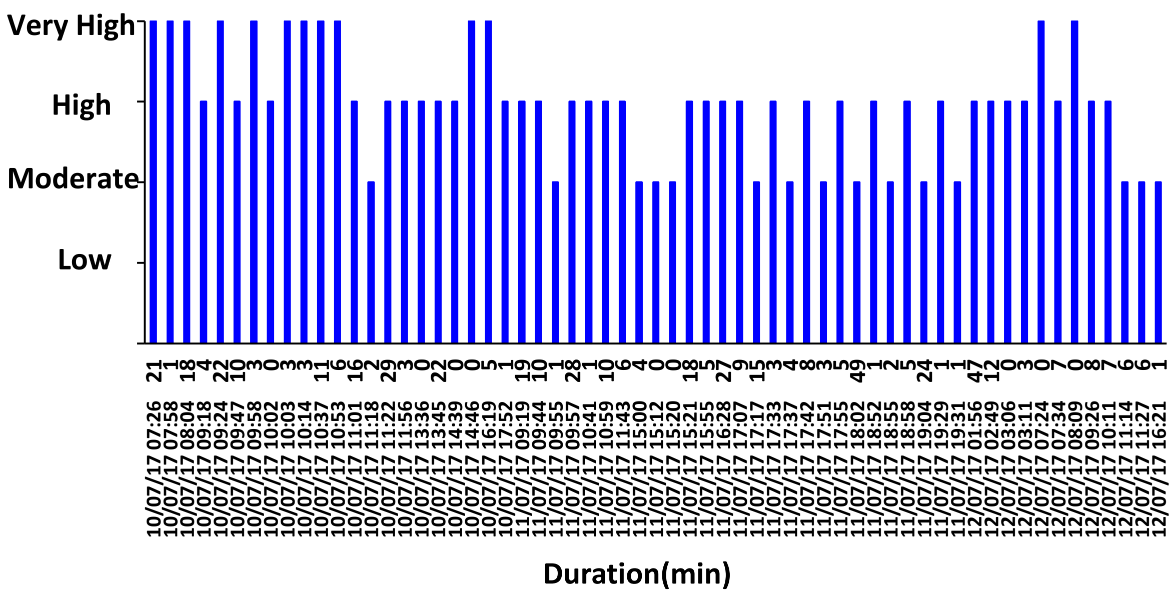


Figure A.1. Risk level classes with the corresponding duration after 10/07/17 07:13.

Table A.1. Risk and Priority level for S::CAN data according to Approach 1-Level 2.

Start of new level of an event	End of level	Risk Level	Weight of level	Duration (min)	Weight of duration	Priority Score	Priority level
6/7/2017 21:22	6/7/2017 21:23	High	3	1	1	3	L.P
6/7/2017 21:24	6/7/2017 21:24	Very High	4	0	1	4	I.P
6/7/2017 21:25	6/7/2017 22:38	High	3	73	3	9	H.P
6/7/2017 22:39	6/7/2017 22:40	Very High	4	1	1	4	I.P
6/7/2017 22:41	7/7/2017 1:41	High	3	180	4	12	H.P
7/7/2017 1:42	7/7/2017 1:45	Very High	4	3	1	4	I.P
7/7/2017 1:46	7/7/2017 1:46	High	3	0	1	3	L.P
7/7/2017 1:47	7/7/2017 1:47	Very High	4	0	1	4	I.P
7/7/2017 1:48	7/7/2017 1:52	High	3	4	1	3	L.P
7/7/2017 1:53	7/7/2017 6:40	Very High	4	287	4	16	U.P
7/7/2017 6:41	7/7/2017 7:10	High	3	29	1	3	L.P
7/7/2017 7:11	7/7/2017 7:18	Very High	4	7	1	4	I.P
7/7/2017 7:19	7/7/2017 8:07	High	3	48	2	6	I.P
7/7/2017 8:08	7/7/2017 8:09	Very High	4	1	1	4	I.P
7/7/2017 8:10	7/7/2017 8:42	High	3	32	2	6	I.P
7/7/2017 8:43	7/7/2017 8:44	Very High	4	1	1	4	I.P
7/7/2017 8:45	7/7/2017 8:50	High	3	5	1	3	L.P
7/7/2017 8:51	7/7/2017 8:51	Very High	4	0	1	4	I.P
7/7/2017 8:52	7/7/2017 9:23	High	3	31	2	6	I.P
7/7/2017 9:24	7/7/2017 9:24	Very High	4	0	1	4	I.P
7/7/2017 9:25	7/7/2017 11:40	High	3	135	4	12	H.P

7/7/2017 11:41	7/7/2017 11:44	Moderate	2	3	1	2	L.P
7/7/2017 11:45	7/7/2017 11:50	High	3	5	1	3	L.P
7/7/2017 11:51	7/7/2017 12:51	Moderate	2	60	3	6	I.P
7/7/2017 12:52	7/7/2017 12:57	High	3	5	1	3	L.P
7/7/2017 12:58	7/7/2017 13:02	Moderate	2	4	1	2	L.P
7/7/2017 13:03	7/7/2017 19:19	High	3	376	4	12	H.P
7/7/2017 19:20	7/7/2017 19:20	Very High	4	0	1	4	I.P
7/7/2017 19:21	7/7/2017 20:38	High	3	77	3	9	H.P
7/7/2017 20:39	7/7/2017 20:39	Very High	4	0	1	4	I.P
7/7/2017 20:41	7/7/2017 20:46	High	3	5	1	3	L.P
7/7/2017 22:48	7/7/2017 23:49	Very High	4	61	3	12	H.P
7/7/2017 23:59	7/7/2017 23:59	Very High	4	0	1	4	I.P
8/7/2017 13:15	8/7/2017 13:15	Very High	4	0	1	4	I.P
8/7/2017 13:53	8/7/2017 13:53	High	3	0	1	3	L.P
8/7/2017 13:54	8/7/2017 14:01	Very High	4	7	1	4	I.P
8/7/2017 14:02	8/7/2017 14:41	High	3	39	2	6	I.P
8/7/2017 14:42	8/7/2017 14:43	Very High	4	1	1	4	I.P
8/7/2017 14:44	8/7/2017 18:53	High	3	249	4	12	H.P
8/7/2017 18:59	8/7/2017 19:02	High	3	3	1	3	L.P
8/7/2017 19:10	8/7/2017 19:10	High	3	0	1	3	L.P
8/7/2017 19:16	8/7/2017 19:16	High	3	0	1	3	L.P
8/7/2017 19:23	8/7/2017 19:23	High	3	0	1	3	L.P
8/7/2017 19:30	8/7/2017 19:35	High	3	5	1	3	L.P
8/7/2017 19:40	8/7/2017 19:51	High	3	11	1	3	L.P
8/7/2017 21:26	8/7/2017 21:26	High	3	0	1	3	L.P
9/7/2017 1:08	9/7/2017 3:20	Very High	4	132	4	16	U.P
10/7/2017 1:04	10/7/2017 2:55	Very High	4	111	3	12	H.P
10/7/2017 7:13	10/7/2017 7:16	Very High	4	3	1	4	I.P
10/7/2017 7:26	10/7/2017 7:47	Very High	4	21	1	4	I.P
10/7/2017 7:58	10/7/2017 7:59	Very High	4	1	1	4	I.P
10/7/2017 8:04	10/7/2017 8:22	Very High	4	18	1	4	I.P
10/7/2017 9:18	10/7/2017 9:22	High	3	4	1	3	L.P
10/7/2017 9:24	10/7/2017 9:46	Very High	4	22	1	4	I.P
10/7/2017 9:47	10/7/2017 9:57	High	3	10	1	3	L.P
10/7/2017 9:58	10/7/2017 10:01	Very High	4	3	1	4	I.P
10/7/2017 10:02	10/7/2017 10:02	High	3	0	1	3	L.P
10/7/2017 10:03	10/7/2017 10:06	Very High	4	3	1	4	I.P
10/7/2017 10:14	10/7/2017 10:17	Very High	4	3	1	4	I.P
10/7/2017 10:37	10/7/2017 10:48	Very High	4	11	1	4	I.P
10/7/2017 10:53	10/7/2017 10:59	Very High	4	6	1	4	I.P
10/7/2017 11:01	10/7/2017 11:17	High	3	16	1	3	L.P
10/7/2017 11:18	10/7/2017 11:20	Moderate	2	2	1	2	L.P
10/7/2017 11:22	10/7/2017 11:51	High	3	29	1	3	L.P
10/7/2017 11:56	10/7/2017 11:59	High	3	3	1	3	L.P
10/7/2017 13:36	10/7/2017 13:36	High	3	0	1	3	L.P
10/7/2017 13:45	10/7/2017 14:07	High	3	22	1	3	L.P

10/7/2017 14:39	10/7/2017 14:39	High	3	0	1	3	L.P
10/7/2017 14:46	10/7/2017 14:46	Very High	4	0	1	4	I.P
10/7/2017 16:19	10/7/2017 16:24	Very High	4	5	1	4	I.P
10/7/2017 17:52	10/7/2017 17:53	High	3	1	1	3	L.P
11/7/2017 9:19	11/7/2017 9:38	High	3	19	1	3	L.P
11/7/2017 9:44	11/7/2017 9:54	High	3	10	1	3	L.P
11/7/2017 9:55	11/7/2017 9:56	Moderate	2	1	1	2	L.P
11/7/2017 9:57	11/7/2017 10:25	High	3	28	1	3	L.P
11/7/2017 10:41	11/7/2017 10:42	High	3	1	1	3	L.P
11/7/2017 10:59	11/7/2017 11:09	High	3	10	1	3	L.P
11/7/2017 11:43	11/7/2017 11:49	High	3	6	1	3	L.P
11/7/2017 15:00	11/7/2017 15:04	Moderate	2	4	1	2	L.P
11/7/2017 15:12	11/7/2017 15:12	Moderate	2	0	1	2	L.P
11/7/2017 15:20	11/7/2017 15:20	Moderate	2	0	1	2	L.P
11/7/2017 15:21	11/7/2017 15:39	High	3	18	1	3	L.P
11/7/2017 15:55	11/7/2017 16:00	High	3	5	1	3	L.P
11/7/2017 16:28	11/7/2017 16:55	High	3	27	1	3	L.P
11/7/2017 17:07	11/7/2017 17:16	High	3	9	1	3	L.P
11/7/2017 17:17	11/7/2017 17:32	Moderate	2	15	1	2	L.P
11/7/2017 17:33	11/7/2017 17:36	High	3	3	1	3	L.P
11/7/2017 17:37	11/7/2017 17:41	Moderate	2	4	1	2	L.P
11/7/2017 17:42	11/7/2017 17:50	High	3	8	1	3	L.P
11/7/2017 17:51	11/7/2017 17:54	Moderate	2	3	1	2	L.P
11/7/2017 17:55	11/7/2017 18:00	High	3	5	1	3	L.P
11/7/2017 18:02	11/7/2017 18:51	Moderate	2	49	2	4	I.P
11/7/2017 18:52	11/7/2017 18:53	High	3	1	1	3	L.P
11/7/2017 18:55	11/7/2017 18:57	Moderate	2	2	1	2	L.P
11/7/2017 18:58	11/7/2017 19:03	High	3	5	1	3	L.P
11/7/2017 19:04	11/7/2017 19:28	Moderate	2	24	1	2	L.P
11/7/2017 19:29	11/7/2017 19:30	High	3	1	1	3	L.P
11/7/2017 19:31	11/7/2017 19:32	Moderate	2	1	1	2	L.P
12/7/2017 1:56	12/7/2017 2:43	High	3	47	2	6	I.P
12/7/2017 2:49	12/7/2017 3:01	High	3	12	1	3	L.P
12/7/2017 3:06	12/7/2017 3:06	High	3	0	1	3	L.P
12/7/2017 3:11	12/7/2017 3:14	High	3	3	1	3	L.P
12/7/2017 7:24	12/7/2017 7:24	Very High	4	0	1	4	I.P
12/7/2017 7:34	12/7/2017 7:41	High	3	7	1	3	L.P
12/7/2017 8:09	12/7/2017 8:09	Very High	4	0	1	4	I.P
12/7/2017 9:26	12/7/2017 9:34	High	3	8	1	3	L.P
12/7/2017 10:11	12/7/2017 10:18	High	3	7	1	3	L.P
12/7/2017 11:14	12/7/2017 11:20	Moderate	2	6	1	2	L.P
12/7/2017 11:27	12/7/2017 11:33	Moderate	2	6	1	2	L.P
12/7/2017 16:21	12/7/2017 16:22	Moderate	2	1	1	2	L.P

Table A.2. Risk level for S::CAN data according to Approach 2-Level 1.

Start of Event	End of Event	Duration (min)	Weight of Duration	Average Turbidity (NTU)	Weight of Turbidity	Risk Score	Risk Level
06/07/2017 21:22	07/07/2017 20:46	1404	4	1,499	2	8	High
07/07/2017 22:48	07/07/2017 23:49	61	3	1,011	1	3	Low
07/07/2017 23:59	07/07/2017 23:59	0	1	1,001	1	1	Low
08/07/2017 13:15	08/07/2017 13:15	0	1	1,005	1	1	Low
08/07/2017 13:53	08/07/2017 18:53	300	4	1,017	1	4	Moderate
08/07/2017 18:59	08/07/2017 19:02	3	1	1,002	1	1	Low
08/07/2017 19:10	08/07/2017 19:10	0	1	1,004	1	1	Low
08/07/2017 19:16	08/07/2017 19:16	0	1	1,002	1	1	Low
08/07/2017 19:23	08/07/2017 19:23	0	1	1,002	1	1	Low
08/07/2017 19:30	08/07/2017 19:35	5	1	1,004	1	1	Low
08/07/2017 19:40	08/07/2017 19:51	11	1	1,003	1	1	Low
08/07/2017 21:26	08/07/2017 21:26	0	1	1,002	1	1	Low
09/07/2017 01:08	09/07/2017 03:20	132	4	1,014	1	4	Moderate
10/07/2017 01:04	10/07/2017 02:55	111	3	1,025	1	3	Low
10/07/2017 07:13	10/07/2017 07:16	3	1	1,056	1	1	Low
10/07/2017 07:26	10/07/2017 07:47	21	1	1,102	1	1	Low
10/07/2017 07:58	10/07/2017 07:59	1	1	1,008	1	1	Low
10/07/2017 08:04	10/07/2017 08:22	18	1	1,09	1	1	Low
10/07/2017 09:18	10/07/2017 10:06	48	2	1,07	1	2	Low
10/07/2017 10:14	10/07/2017 10:17	3	1	1,005	1	1	Low
10/07/2017 10:37	10/07/2017 10:48	11	1	1,045	1	1	Low
10/07/2017 10:53	10/07/2017 11:51	58	2	1,3	2	4	Moderate
10/07/2017 11:56	10/07/2017 11:59	3	1	1,004	1	1	Low
10/07/2017 13:36	10/07/2017 13:36	0	1	1,018	1	1	Low
10/07/2017 13:45	10/07/2017 14:07	22	1	1,021	1	1	Low
10/07/2017 14:39	10/07/2017 14:39	0	1	1,01	1	1	Low
10/07/2017 14:46	10/07/2017 14:46	0	1	1,012	1	1	Low
10/07/2017 16:19	10/07/2017 16:24	5	1	1,018	1	1	Low
10/07/2017 17:52	10/07/2017 17:53	1	1	1,041	1	1	Low
11/07/2017 09:19	11/07/2017 09:38	19	1	1,056	1	1	Low
11/07/2017 09:44	11/07/2017 10:25	41	2	1,052	1	2	Low
11/07/2017 10:41	11/07/2017 10:42	1	1	1,003	1	1	Low
11/07/2017 10:59	11/07/2017 11:09	10	1	1,013	1	1	Low
11/07/2017 11:43	11/07/2017 11:49	6	1	1,005	1	1	Low
11/07/2017 15:00	11/07/2017 15:04	4	1	1,004	1	1	Low
11/07/2017 15:12	11/07/2017 15:12	0	1	1,007	1	1	Low
11/07/2017 15:20	11/07/2017 15:39	19	1	1,008	1	1	Low
11/07/2017 15:55	11/07/2017 16:00	5	1	1,001	1	1	Low
11/07/2017 16:28	11/07/2017 16:55	27	1	1,082	1	1	Low
11/07/2017 17:07	11/07/2017 19:32	145	4	1,043	1	4	Moderate
12/07/2017 01:56	12/07/2017 02:43	47	2	1,03	1	2	Low
12/07/2017 02:49	12/07/2017 03:01	12	1	1,003	1	1	Low
12/07/2017 03:06	12/07/2017 03:06	0	1	1,004	1	1	Low

12/07/2017 03:11	12/07/2017 03:14	3	1	1,001	1	1	Low
12/07/2017 07:24	12/07/2017 07:24	0	1	1,007	1	1	Low
12/07/2017 07:34	12/07/2017 07:41	7	1	1,027	1	1	Low
12/07/2017 08:09	12/07/2017 08:09	0	1	1,008	1	1	Low
12/07/2017 09:26	12/07/2017 09:34	8	1	1,017	1	1	Low
12/07/2017 10:11	12/07/2017 10:18	7	1	1,145	1	1	Low
12/07/2017 11:14	12/07/2017 11:20	6	1	1,006	1	1	Low
12/07/2017 11:27	12/07/2017 11:33	6	1	1,017	1	1	Low
12/07/2017 16:21	12/07/2017 16:22	1	1	1,002	1	1	Low

Table A.3. Severity level for S::CAN data according to Approach 2-Level 2.

Start of Event	End of Event	Average Turbidity (NTU)	Weight of Turbidity	Average Chlorine (mg/l)	Weight of Chlorine	Severity Score	Severity Level
6/7/2017 21:22	7/7/2017 20:46	1.499	2	0.008	3	6	Minor
7/7/2017 22:48	7/7/2017 23:49	1.011	1	0.001	4	4	Minor
7/7/2017 23:59	7/7/2017 23:59	1.001	1	0	4	4	Minor
8/7/2017 13:15	8/7/2017 13:15	1.005	1	0.004	4	4	Minor
8/7/2017 13:53	8/7/2017 18:53	1.017	1	0.008	3	3	Insignificant
8/7/2017 18:59	8/7/2017 19:02	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:10	8/7/2017 19:10	1.004	1	0.009	3	3	Insignificant
8/7/2017 19:16	8/7/2017 19:16	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:23	8/7/2017 19:23	1.002	1	0.009	3	3	Insignificant
8/7/2017 19:30	8/7/2017 19:35	1.004	1	0.007	3	3	Insignificant
8/7/2017 19:40	8/7/2017 19:51	1.003	1	0.01	3	3	Insignificant
8/7/2017 21:26	8/7/2017 21:26	1.002	1	0.009	3	3	Insignificant
9/7/2017 1:08	9/7/2017 3:20	1.014	1	0.001	4	4	Minor
10/7/2017 1:04	10/7/2017 2:55	1.025	1	0	4	4	Minor
10/7/2017 7:13	10/7/2017 7:16	1.056	1	0	4	4	Minor
10/7/2017 7:26	10/7/2017 7:47	1.102	1	0	4	4	Minor
10/7/2017 7:58	10/7/2017 7:59	1.008	1	0	4	4	Minor
10/7/2017 8:04	10/7/2017 8:22	1.09	1	0.001	4	4	Minor
10/7/2017 9:18	10/7/2017 10:06	1.07	1	0.004	4	4	Minor
10/7/2017 10:14	10/7/2017 10:17	1.005	1	0.004	4	4	Minor
10/7/2017 10:37	10/7/2017 10:48	1.045	1	0.004	4	4	Minor
10/7/2017 10:53	10/7/2017 11:51	1.3	2	0.006	3	6	Minor
10/7/2017 11:56	10/7/2017 11:59	1.004	1	0.007	3	3	Insignificant
10/7/2017 13:36	10/7/2017 13:36	1.018	1	0.008	3	3	Insignificant
10/7/2017 13:45	10/7/2017 14:07	1.021	1	0.009	3	3	Insignificant
10/7/2017 14:39	10/7/2017 14:39	1.01	1	0.009	3	3	Insignificant
10/7/2017 14:46	10/7/2017 14:46	1.012	1	0.004	4	4	Minor
10/7/2017 16:19	10/7/2017 16:24	1.018	1	0.004	4	4	Minor
10/7/2017 17:52	10/7/2017 17:53	1.041	1	0.177	2	2	Insignificant
11/7/2017 9:19	11/7/2017 9:38	1.056	1	0.113	2	2	Insignificant
11/7/2017 9:44	11/7/2017 10:25	1.052	1	0.174	2	2	Insignificant
11/7/2017 10:41	11/7/2017 10:42	1.003	1	0.176	2	2	Insignificant
11/7/2017 10:59	11/7/2017 11:09	1.013	1	0.169	2	2	Insignificant

11/7/2017 11:43	11/7/2017 11:49	1.005	1	0.178	2	2	Insignificant
11/7/2017 15:00	11/7/2017 15:04	1.004	1	0.236	1	1	Insignificant
11/7/2017 15:12	11/7/2017 15:12	1.007	1	0.244	1	1	Insignificant
11/7/2017 15:20	11/7/2017 15:39	1.008	1	0.178	2	2	Insignificant
11/7/2017 15:55	11/7/2017 16:00	1.001	1	0.188	2	2	Insignificant
11/7/2017 16:28	11/7/2017 16:55	1.082	1	0.151	2	2	Insignificant
11/7/2017 17:07	11/7/2017 19:32	1.043	1	0.207	1	1	Insignificant
12/7/2017 1:56	12/7/2017 2:43	1.03	1	0.033	3	3	Insignificant
12/7/2017 2:49	12/7/2017 3:01	1.003	1	0.018	3	3	Insignificant
12/7/2017 3:06	12/7/2017 3:06	1.004	1	0.02	3	3	Insignificant
12/7/2017 3:11	12/7/2017 3:14	1.001	1	0.019	3	3	Insignificant
12/7/2017 7:24	12/7/2017 7:24	1.007	1	0	4	4	Minor
12/7/2017 7:34	12/7/2017 7:41	1.027	1	0.057	2	2	Insignificant
12/7/2017 8:09	12/7/2017 8:09	1.008	1	0	4	4	Minor
12/7/2017 9:26	12/7/2017 9:34	1.017	1	0.098	2	2	Insignificant
12/7/2017 10:11	12/7/2017 10:18	1.145	1	0.144	2	2	Insignificant
12/7/2017 11:14	12/7/2017 11:20	1.006	1	0.208	1	1	Insignificant
12/7/2017 11:27	12/7/2017 11:33	1.017	1	0.213	1	1	Insignificant
12/7/2017 16:21	12/7/2017 16:22	1.002	1	0.227	1	1	Insignificant

Table A.4. Risk level for S::CAN data according to Approach 2-Level 2.

Start of Event	End of Event	Duration (min)	Weight of Duration	Severity Level	Weight of Severity Level	Risk Score	Risk Level
6/7/2017 21:22	7/7/2017 20:46	1404	4	Minor	2	8	High
7/7/2017 22:48	7/7/2017 23:49	61	3	Minor	2	6	Moderate
7/7/2017 23:59	7/7/2017 23:59	0	1	Minor	2	2	Low
8/7/2017 13:15	8/7/2017 13:15	0	1	Minor	2	2	Low
8/7/2017 13:53	8/7/2017 18:53	300	4	Insignificant	1	4	Moderate
8/7/2017 18:59	8/7/2017 19:02	3	1	Insignificant	1	1	Low
8/7/2017 19:10	8/7/2017 19:10	0	1	Insignificant	1	1	Low
8/7/2017 19:16	8/7/2017 19:16	0	1	Insignificant	1	1	Low
8/7/2017 19:23	8/7/2017 19:23	0	1	Insignificant	1	1	Low
8/7/2017 19:30	8/7/2017 19:35	5	1	Insignificant	1	1	Low
8/7/2017 19:40	8/7/2017 19:51	11	1	Insignificant	1	1	Low
8/7/2017 21:26	8/7/2017 21:26	0	1	Insignificant	1	1	Low
9/7/2017 1:08	9/7/2017 3:20	132	4	Minor	2	8	High
10/7/2017 1:04	10/7/2017 2:55	111	3	Minor	2	6	Moderate
10/7/2017 7:13	10/7/2017 7:16	3	1	Minor	2	2	Low
10/7/2017 7:26	10/7/2017 7:47	21	1	Minor	2	2	Low
10/7/2017 7:58	10/7/2017 7:59	1	1	Minor	2	2	Low
10/7/2017 8:04	10/7/2017 8:22	18	1	Minor	2	2	Low
10/7/2017 9:18	10/7/2017 10:06	48	2	Minor	2	4	Moderate
10/7/2017 10:14	10/7/2017 10:17	3	1	Minor	2	2	Low
10/7/2017 10:37	10/7/2017 10:48	11	1	Minor	2	2	Low
10/7/2017 10:53	10/7/2017 11:51	58	2	Minor	2	4	Moderate

10/7/2017 11:56	10/7/2017 11:59	3	1	Insignificant	1	1	Low
10/7/2017 13:36	10/7/2017 13:36	0	1	Insignificant	1	1	Low
10/7/2017 13:45	10/7/2017 14:07	22	1	Insignificant	1	1	Low
10/7/2017 14:39	10/7/2017 14:39	0	1	Insignificant	1	1	Low
10/7/2017 14:46	10/7/2017 14:46	0	1	Minor	2	2	Low
10/7/2017 16:19	10/7/2017 16:24	5	1	Minor	2	2	Low
10/7/2017 17:52	10/7/2017 17:53	1	1	Insignificant	1	1	Low
11/7/2017 9:19	11/7/2017 9:38	19	1	Insignificant	1	1	Low
11/7/2017 9:44	11/7/2017 10:25	41	2	Insignificant	1	2	Low
11/7/2017 10:41	11/7/2017 10:42	1	1	Insignificant	1	1	Low
11/7/2017 10:59	11/7/2017 11:09	10	1	Insignificant	1	1	Low
11/7/2017 11:43	11/7/2017 11:49	6	1	Insignificant	1	1	Low
11/7/2017 15:00	11/7/2017 15:04	4	1	Insignificant	1	1	Low
11/7/2017 15:12	11/7/2017 15:12	0	1	Insignificant	1	1	Low
11/7/2017 15:20	11/7/2017 15:39	19	1	Insignificant	1	1	Low
11/7/2017 15:55	11/7/2017 16:00	5	1	Insignificant	1	1	Low
11/7/2017 16:28	11/7/2017 16:55	27	1	Insignificant	1	1	Low
11/7/2017 17:07	11/7/2017 19:32	145	4	Insignificant	1	4	Moderate
12/7/2017 1:56	12/7/2017 2:43	47	2	Insignificant	1	2	Low
12/7/2017 2:49	12/7/2017 3:01	12	1	Insignificant	1	1	Low
12/7/2017 3:06	12/7/2017 3:06	0	1	Insignificant	1	1	Low
12/7/2017 3:11	12/7/2017 3:14	3	1	Insignificant	1	1	Low
12/7/2017 7:24	12/7/2017 7:24	0	1	Minor	2	2	Low
12/7/2017 7:34	12/7/2017 7:41	7	1	Insignificant	1	1	Low
12/7/2017 8:09	12/7/2017 8:09	0	1	Minor	2	2	Low
12/7/2017 9:26	12/7/2017 9:34	8	1	Insignificant	1	1	Low
12/7/2017 10:11	12/7/2017 10:18	7	1	Insignificant	1	1	Low
12/7/2017 11:14	12/7/2017 11:20	6	1	Insignificant	1	1	Low
12/7/2017 11:27	12/7/2017 11:33	6	1	Insignificant	1	1	Low
12/7/2017 16:21	12/7/2017 16:22	1	1	Insignificant	1	1	Low

Table A.5. Risk level for S::CAN data, at Polytech'Lille, according to Approach 3.

Start of new severity level	End of severity level	Severity level	Weight of severity	Duration (min)	Weight of Duration	Risk score	Risk level
06/07/2017 21:22	06/07/2017 21:25	Insignificant	1	3	1	1	Low
06/07/2017 21:26	06/07/2017 21:31	Minor	2	5	1	2	Low
06/07/2017 21:32	06/07/2017 21:47	Moderate	3	15	1	3	Low
06/07/2017 21:48	06/07/2017 21:54	Minor	2	6	1	2	Low
06/07/2017 21:55	06/07/2017 21:58	Moderate	3	3	1	3	Low
06/07/2017 21:59	06/07/2017 22:29	Minor	2	30	2	4	Moderate
06/07/2017 22:30	06/07/2017 22:46	Insignificant	1	16	1	1	Low
06/07/2017 22:47	06/07/2017 23:36	Minor	2	49	2	4	Moderate
06/07/2017 23:37	06/07/2017 23:42	Moderate	3	5	1	3	Low

06/07/2017 23:43	07/07/2017 00:08	Minor	2	25	1	2	Low
07/07/2017 00:09	07/07/2017 07:42	Insignificant	1	453	4	4	Moderate
07/07/2017 07:43	07/07/2017 07:43	Minor	2	0	1	2	Low
07/07/2017 07:44	07/07/2017 11:40	Insignificant	1	236	4	4	Moderate
07/07/2017 11:41	07/07/2017 11:44	Moderate	3	3	1	3	Low
07/07/2017 11:45	07/07/2017 11:46	Minor	2	1	1	2	Low
07/07/2017 11:47	07/07/2017 11:48	Insignificant	1	1	1	1	Low
07/07/2017 11:49	07/07/2017 11:50	Minor	2	1	1	2	Low
07/07/2017 11:51	07/07/2017 12:52	Moderate	3	61	3	9	High
07/07/2017 12:53	07/07/2017 12:57	Minor	2	4	1	2	Low
07/07/2017 12:58	07/07/2017 13:02	Moderate	3	4	1	3	Low
07/07/2017 13:03	07/07/2017 13:10	Minor	2	7	1	2	Low
07/07/2017 13:11	07/07/2017 20:46	Insignificant	1	455	4	4	Moderate
07/07/2017 22:47	07/07/2017 23:49	Insignificant	1	62	3	3	Low
07/07/2017 23:54	07/07/2017 23:59	Insignificant	1	5	1	1	Low
08/07/2017 07:01	08/07/2017 07:01	Insignificant	1	0	1	1	Low
08/07/2017 13:15	08/07/2017 13:15	Insignificant	1	0	1	1	Low
08/07/2017 13:53	08/07/2017 19:16	Insignificant	1	323	4	4	Moderate
08/07/2017 19:21	08/07/2017 19:23	Insignificant	1	2	1	1	Low
08/07/2017 19:30	08/07/2017 19:35	Insignificant	1	5	1	1	Low
08/07/2017 19:40	08/07/2017 19:51	Insignificant	1	11	1	1	Low
08/07/2017 20:21	08/07/2017 20:21	Insignificant	1	0	1	1	Low
08/07/2017 21:18	08/07/2017 21:18	Insignificant	1	0	1	1	Low
08/07/2017 21:26	08/07/2017 21:26	Insignificant	1	0	1	1	Low
09/07/2017 01:08	09/07/2017 03:20	Insignificant	1	132	4	4	Moderate
10/07/2017 01:04	10/07/2017 02:55	Insignificant	1	111	3	3	Low
10/07/2017 07:13	10/07/2017 07:16	Insignificant	1	3	1	1	Low
10/07/2017 07:26	10/07/2017 07:47	Insignificant	1	21	1	1	Low
10/07/2017 07:58	10/07/2017 07:59	Insignificant	1	1	1	1	Low
10/07/2017 08:04	10/07/2017 08:22	Insignificant	1	18	1	1	Low
10/07/2017 09:18	10/07/2017 10:17	Insignificant	1	59	2	2	Low
10/07/2017 10:37	10/07/2017 10:48	Insignificant	1	11	1	1	Low
10/07/2017 10:53	10/07/2017 11:16	Insignificant	1	23	1	1	Low
10/07/2017 11:17	10/07/2017 11:17	Minor	2	0	1	2	Low
10/07/2017 11:18	10/07/2017 11:20	Moderate	3	2	1	3	Low
10/07/2017 11:22	10/07/2017 11:23	Minor	2	1	1	2	Low
10/07/2017 11:24	10/07/2017 11:59	Insignificant	1	35	2	2	Low
10/07/2017 13:36	10/07/2017 13:36	Insignificant	1	0	1	1	Low
10/07/2017 13:45	10/07/2017 14:07	Insignificant	1	22	1	1	Low
10/07/2017 14:39	10/07/2017 14:39	Insignificant	1	0	1	1	Low
10/07/2017 14:46	10/07/2017 14:46	Insignificant	1	0	1	1	Low
10/07/2017 16:19	10/07/2017 16:24	Insignificant	1	5	1	1	Low
10/07/2017 17:52	10/07/2017 17:53	Insignificant	1	1	1	1	Low
11/07/2017 09:19	11/07/2017 09:38	Insignificant	1	19	1	1	Low
11/07/2017 09:44	11/07/2017 10:25	Insignificant	1	41	2	2	Low
11/07/2017 10:41	11/07/2017 10:42	Insignificant	1	1	1	1	Low

11/07/2017 10:59	11/07/2017 11:09	Insignificant	1	10	1	1	Low
11/07/2017 11:43	11/07/2017 11:49	Insignificant	1	6	1	1	Low
11/07/2017 14:57	11/07/2017 15:07	Insignificant	1	10	1	1	Low
11/07/2017 15:12	11/07/2017 15:12	Insignificant	1	0	1	1	Low
11/07/2017 15:20	11/07/2017 15:39	Insignificant	1	19	1	1	Low
11/07/2017 15:55	11/07/2017 16:00	Insignificant	1	5	1	1	Low
11/07/2017 16:28	11/07/2017 16:57	Insignificant	1	29	1	1	Low
11/07/2017 17:07	11/07/2017 19:32	Insignificant	1	145	4	4	Moderate
12/07/2017 01:56	12/07/2017 03:01	Insignificant	1	65	3	3	Low
12/07/2017 03:06	12/07/2017 03:06	Insignificant	1	0	1	1	Low
12/07/2017 03:11	12/07/2017 03:14	Insignificant	1	3	1	1	Low
12/07/2017 07:24	12/07/2017 07:24	Insignificant	1	0	1	1	Low
12/07/2017 07:34	12/07/2017 07:41	Insignificant	1	7	1	1	Low
12/07/2017 08:09	12/07/2017 08:09	Insignificant	1	0	1	1	Low
12/07/2017 09:26	12/07/2017 09:34	Insignificant	1	8	1	1	Low
12/07/2017 10:11	12/07/2017 10:18	Insignificant	1	7	1	1	Low
12/07/2017 11:14	12/07/2017 11:20	Insignificant	1	6	1	1	Low
12/07/2017 11:26	12/07/2017 11:33	Insignificant	1	7	1	1	Low
12/07/2017 16:21	12/07/2017 16:22	Insignificant	1	1	1	1	Low