Dottorato di Ricerca in Ingegneria Civile e Architettura
Graduate School in Civil Engineering and Architecture

DICAr - Dipartimento di Ingegneria Civile e Architettura
Facoltà di Ingegneria, Università di Pavia
Department of Civil Engineering and Architecture
Faculty of Engineering, University of Pavia

XXX Cycle (XVI New Series)

In collaboration with NYU Tandon School of Engineering
Urban Infrastructure Institute

Monitoring, Early Detection & Warning Systems for contamination events in Water Distribution Networks

Ph.D. Thesis

Silvia Tinelli

Advisor
Prof. Carlo Ciaponi

Co-Advisors
Prof. Enrico Creaco
Prof. Ilan Juran

Reviewers
Prof. Alberto Campisano
Prof. Armando Di Nardo

Ph.D. Coordinator
Prof. Ferdinando Auricchio

March 2018

The Human mind has claimed for Water one
of its highest values - *the Value of Purity*
- Gaston Bachelard (1884 - 1962) -

Contents

Abstract

In the chain of water distribution, the network is the most complex element to be analyzed and managed to deliver safe water to the users due to the vast dispersion of the potential contamination spots.
For this reason, some countries, especially those most sensible to the terrorist attacks (USA, Israel, Europe) have already started research programs aimed at the development of an Online Water Quality Monitoring (OWQM) and of Early Warning Systems (EWSs). Both of them are based on sensors installed in selected nodes of the network and are capable of quickly detecting contamination events.
The implementation of EWSs paves the way to new interesting research topics, with particular reference to the technological aspects, to the employment of expert systems for the interpretation of the detected data, and to the definition of modeling tools for the design and management of the monitoring and alarm systems.
The Thesis focuses on some of these aspects, with the aim of contributing to a partial systematization of the knowledge required for the design and management of the aforementioned systems.
This Thesis can be divided into two parts.
The former part of the Thesis (**Chapters 1**, **2** and **3**) describes the general issues and the approach normally adopted in choosing the water parameters to be monitored. In particular, a wide overview of the currently available sensors for *in situ* and continuous automatic detection of physical, chemical and biological parameters of the water flowing through the pipeline is presented. However, due to the wide spectrum of possible contaminants, the almost real-time identification of risk situations in the delivered water is a very difficult scientific and technological challenge. In fact, while there are laboratory technologies

capable of measuring all the substances present in water, the analysis capabilities of the devices that can be used *in situ* for continuous and automatic monitoring are very limited. The current Thesis discusses this problem, which represents one of the most significant complexities in the implementation of the EWS, and the approaches generally adopted for its solution.

The latter part of the Thesis (**Chapters 4**, **5** and **6**) deals with some modeling aspects regarding the design and management of EWSs, introducing innovative proposals and developments.

In particular, the attention is given to the issue of determining the number and the optimal location of the sensors within the network. In fact, the effectiveness of the EWS depends on the number, as well as on the location of the sensors. For a pre-determined number of sensors, necessarily limited for budget reasons, the best placement is the one that maximizes its effectiveness, that is the ability of the system to reduce the impact of contamination accidents on public health. This is an optimization problem that must be addressed with reference to at least two conflicting objectives: the cost (to be minimized) and the system effectiveness (to be maximized). To resolve this optimization problem, it is useful to define all the contamination events that may potentially affect the network.

Each event is characterized by (i) the node, or nodes, where the contamination occurs (ii) the starting time of the same contamination (iii) its duration, and (iv) the value of its mss. Thus, the number of the potential events can become huge for extended and complex networks. For each contamination event, it is necessary to evaluate the propagation of the contaminant in the network through a hydraulic analysis, which includes quantitative and qualitative aspects. In addition, the optimization procedures are computationally burdensome when the number of the potential events to be considered is high. For these reasons, the ensemble of the contamination events to be taken into account in the calculation needs to be reduced. This can be done by selecting a small sample of events, but still representative of the global set. In the Thesis, a sampling procedure based on practical considerations regarding the topology and the network management is proposed. The application of these criteria to a case study showed that the final output (*Pareto front*) does not significantly change considering the reduced sample rather than the totality of the events.

Another important aspect associated with the optimization problem concerns the translation of the above-mentioned criterion, which reflects the minimum impact

of the contamination on public safety, into *Objective Functions* that can be quantitatively expressed.
For example, a possible objective function is the probability (to be maximized) that the contaminated flow passes through a monitored node in the network (detection likelihood). Alternatively, the objective functions can be expressed by other variables (to be minimized), such as the elapsed time between the contamination and its detection, the number of inhabitants that is reached by the contaminant, the number of people that receives a contaminant concentration higher than a certain threshold, the amount of the provided contaminated water, and the percentage of the non-detected contamination events. In the Thesis, this problem is addressed by examining how the choice of the objective functions affects the final results. To this purpose, two different variants were developed and compared to each others. Both of them adopt the number of installed sensors as the first objective function (to be minimized) but they differ in the choice of the second objective function. This function was assumed to be the detection likelihood (to be maximized) and the average contaminated population (to be minimized), in the former and latter variant respectively. The results of the optimizations, and the re-evaluations of the optimal solutions in terms of various effectiveness indicators for the water quality monitoring system, prove that the first variant (O. F. = detection likelihood) tends to produce better solutions in terms of detection likelihood and sensor redundancy. On the other hand, the second (O. F. = contaminated population) tends to produce better solutions in terms of contaminated population and event detection time. The choice between the two variants should thus be taken into account even in relation to the specific situation and the alarm-programmed interventions, depending on whether the preference should be given to the detection security or its promptness. The Thesis also shows that the two different variants give rise to sensitively different sensor locations.
For each contamination event considered in the optimization procedure, all the processing requires the assessment of the contaminant propagation in the network through a hydraulic analysis, which includes quantitative and qualitative aspects. To this purpose, in the optimization process the simplified hypothesis of *conservative contaminant* was adopted, neglecting the contaminant reactions that occur when it is combined with the other elements present in water. However, the issue of *non-conservative contaminant* was also addressed in the Thesis through numerical experiments, carried out by the EPANET Multi-Species Extension (EPANET-MSX) software. In details, they quantitatively

faced the chlorine decay as a result of *E. coli* contaminations. In addition to illustrating the numerical results, the Thesis describes the experimental apparatus and activities developed at the Civil Engineering and Geo-Environmental Laboratory of the Lille University (Villeneuve-d'Ascq, Lille - France) for the validation of the numerical models, developed by the Author.

Ultimately, the last chapter shows the technical feasibility of a smart prototype system for the early detection of biological contaminations within the network. This system will efficiently enable water utility managers to ensure a real-time adoption of water quality control procedures. To this end, an automated statistical model and Artificial Intelligence (AI) supported algorithms are presented and validated using chlorine data obtained from the numerical simulations above-mentioned. The developed algorithms exploit the concept of *expert pattern recognition*: an algorithm appropriately trained on the standard conditions of a system is able to recognize deviations from the normal conditions enough evident to constitute an anomaly. Among the available supervised learning models, advance pattern recognizers, such as the Support Vector Machines (SVMs), as well as the Artificial Neural Network (ANN), were tested and compared to each other. The results show not only an efficient anomaly detection and risk-based classification, but also the ability of the final output to visualize the contaminated nodes on the network map, according to a risk severity scale.

Sommario

Nell’ambito della filiera di produzione e di erogazione dell’acqua potabile, la rete di distribuzione idrica rappresenta l’elemento più complesso da analizzare e gestire per quanto riguarda la sicurezza qualitativa dell’acqua consegnata all’utenza, a causa della grande dispersione dei punti di potenziale contaminazione. Per questo motivo, soprattutto nei Paesi più sensibili al problema degli attacchi terroristici (Nord America, Europa, Israele), sono stati da tempo avviati programmi di ricerca finalizzati alla messa a punto di sistemi di monitoraggio continuo e di allarme precoce (EWS) basati su sensori, da installare in punti opportunamente scelti della rete, in grado di rilevare in tempi rapidi gli eventi di contaminazione.
L’implementazione di questi sistemi apre nuovi interessanti temi di ricerca con particolare riferimento agli aspetti tecnologici, all’implementazione di sistemi esperti per l’interpretazione dei dati rilevati, alla definizione di strumenti modellistici per la progettazione e la gestione dei sistemi di monitoraggio e di allarme.
La Tesi focalizza l’attenzione su alcuni di questi aspetti, con lo scopo di contribuire ad una pur parziale sistematizzazione delle conoscenze necessarie per la progettazione e la gestione dei sistemi sopra citati.
La Tesi può essere suddivisa in due parti.
La prima parte (**Capitoli 1**, **2** e **3**) descrive le problematiche generali e l’approccio generalmente adottato nella scelta dei parametri da monitorare. In particolare, viene presentata un’ampia disamina dei sensori attualmente disponibili per il rilevamento automatico *in situ* e in continuo dei parametri fisici, chimici e biologici dell’acqua che transita nella tubazione. In ragione dell’ampio spettro delle possibili sostanze contaminanti, l’identificazione in tempo pressoché reale di situazioni di rischio nell’acqua distribuita rappresenta, però, una sfida scientifica e tecnologica molto ardua. Mentre, infatti, esistono

tecnologie da laboratorio atte a misurare praticamente tutte le sostanze di interesse presenti nell'acqua, le capacità di analisi dei dispositivi utilizzabili *in situ* per il monitoraggio continuo e automatico sono molto limitate. La Tesi discute questo problema, che costituisce una delle maggiori difficoltà di implementazione dei EWS, e gli approcci generalmente adottati per la sua soluzione.
La seconda parte della Tesi (**Capitoli 4**, **5** e **6**) affronta poi, anche con proposte e sviluppi originali, alcuni aspetti modellistici riguardanti la progettazione e la gestione degli EWS.
In particolare, molta attenzione è dedicata al problema della definizione del numero e della localizzazione ottimale dei sensori nell'ambito della rete. L'efficacia del sistema EWS dipende infatti dal numero e dalla localizzazione dei sensori. Per un prefissato numero di sensori, necessariamente limitato per ragioni di costo, la migliore localizzazione è quella che ne massimizza l'efficacia, ovvero la capacità del sistema di ridurre l'impatto degli incidenti di contaminazione sulla salute pubblica. Si tratta di un problema di ottimizzazione che va affrontato con riferimento ad almeno due obiettivi fra loro in conflitto: il costo (che va minimizzato) e l'efficacia del sistema (che va massimizzato). Per la risoluzione del problema di ottimizzazione sopra indicato, è utile definire tutti gli eventi di contaminazione che potenzialmente possono interessare la rete. Ogni evento è caratterizzato dal nodo (o dai nodi) in cui avviene la contaminazione, dall'istante iniziale della stessa, dalla sua durata, dal valore della massa inquinante immessa in rete e quindi, per reti estese e complesse, il numero degli eventi potenziali può diventare enorme. Poiché, per ogni evento, è necessario valutare la propagazione del contaminante nella rete attraverso un'analisi del funzionamento idraulico che comprenda gli aspetti quantitativi e qualitativi e poiché anche le procedure di ottimizzazione sono molto onerose sotto il profilo computazionale quando il numero dei potenziali eventi da prendere in considerazione sia elevato, è necessario ridurre significativamente l'insieme delle situazioni di contaminazione di cui tenere conto nel calcolo. Ciò può essere fatto selezionando un campione di eventi ridotto, ma comunque rappresentativo dell'insieme globale. Nella Tesi è proposta una procedura di campionamento basata su considerazioni di tipo pratico relativamente alla topologia e alla gestione della rete. L'applicazione di questi criteri ad un caso studio ha mostrato che il risultato finale (fronte di Pareto) non cambia in modo significativo considerando il campione ridotto anziché la totalità degli eventi.
Un altro importante aspetto associato al problema di ottimizzazione riguarda la

traduzione in *Funzioni Obiettivo* esprimibili in termini quantitativi dell'obiettivo generale sopra indicato, che contempla il minimo impatto della contaminazione sulla salute pubblica.
Ad esempio, una possibile funzione obiettivo esprimibile in termini quantitativi corrisponde alla probabilità (da massimizzare) che il flusso contaminato transiti per un punto monitorato della rete (probabilità di rilevamento). In alternativa, la funzione obiettivo può essere espressa attraverso grandezze (da minimizzare) quali, ad esempio, il tempo intercorrente fra la contaminazione e il suo rilevamento, il numero degli abitanti che in questo tempo sono raggiunti dal contaminante, il numero degli abitanti che ricevono una concentrazione di contaminante superiore ad una determinata soglia, il quantitativo di acqua contaminata erogata, la percentuale degli eventi di contaminazione non rilevati. Nella Tesi questo problema è affrontato esaminando come la scelta della funzione obiettivo influenzi il risultato finale. A tal fine, sono state esaminate e fra loro comparate due diverse impostazioni, entrambe basate sull'impiego del numero dei sensori come prima funzione obiettivo (da minimizzare). Le due impostazioni si differenziano invece per la seconda funzione obiettivo che è stata assunta rispettivamente corrispondente alla probabilità di rilevamento (da massimizzare) e all'entità della popolazione raggiunta dal contaminante (da minimizzare). I risultati delle ottimizzazioni e le rivalutazioni delle soluzioni ottimali in termini di alcuni indicatori dell'efficacia del sistema di monitoraggio mostrano che la prima impostazione (F.O. = probabilità di rilevamento) produce soluzioni più efficaci per quanto riguarda la probabilità di rilevamento e il grado di ridondanza del sistema di monitoraggio. Per contro, la seconda impostazione (F.O. = n° utenti contaminati) produce soluzioni più efficaci con riferimento alla riduzione dell'entità della popolazione raggiunta dalla contaminazione e del tempo intercorrente fra l'inizio della contaminazione e il suo rilevamento. La scelta fra le due impostazioni va fatta quindi tenendo conto, anche in relazione alla situazione specifica e agli interventi programmati in caso di allarme, se sia preferibile privilegiare la sicurezza del rilevamento o la sua tempestività. La Tesi evidenzia anche che le due differenti impostazioni danno origine a localizzazioni dei sensori sensibilmente diverse fra loro.
Tutte le elaborazioni sopra indicate, richiedono, per ogni evento considerato nel processo di ottimizzazione, la valutazione della propagazione del contaminante nella rete attraverso un'analisi del funzionamento idraulico comprendente gli aspetti quantitativi e qualitativi. A tal fine, nel processo di ottimizzazione, è stata adottata l'ipotesi semplificata di *contaminante conservativo*, trascurando quindi

le alterazioni che il contaminante subisce quando si combina con altri elementi presenti nell'acqua. Nella Tesi, tuttavia, si è voluto affrontare anche il problema del *contaminante non conservativo*, mediante sperimentazioni numeriche, condotte attraverso il software EPANET Multi-Species Extension (EPANET-MSX), che hanno affrontato in termini quantitativi il decadimento del cloro per effetto di una contaminazione da *E. coli*. La Tesi, oltre a illustrare i risultati numerici, descrive anche gli apparati e le attività sperimentali messi a punto presso il Laboratorio di Ingegneria Civile e Geo-Ambientale dell'Università di Lille (Villeneuve-d'Ascq, Lille - Francia), per la validazione dei modelli di simulazione numerica, curata dalla scrivente.
Infine, nell'ultimo capitolo viene illustrata la fattibilità tecnica di un sistema intelligente atto a rilevare con rapidità contaminazioni biologiche nelle reti di distribuzione, consentendo agli enti gestori la definizione, in tempo reale, delle modalità di intervento per mantenere un'idonea qualità dell'acqua. A tal fine, nella Tesi, si sono presentati un modello statistico e algoritmi di apprendimento basati sull'Intelligenza Artificiale, validati mediante i dati che riproducono l'andamento della concentrazione del cloro nelle simulazioni numeriche sopra citate. Gli algoritmi sviluppati considerano il principio per cui una volta conosciuta la qualità "standard" dell'acqua in rete, definita dall'andamento tipico dei parametri chimico fisici misurati, le loro deviazioni dal range di "normalità" consentono di identificare celermente le anomalie. Tra le metodologie di apprendimento automatico (o "supervisionato"), sono state testate e confrontate le Macchine a Vettore di Supporto (dall'inglese Support Vector Machines-SVM) e la Rete Neurale Artificiale (dall'inglese Artificial Neural Network-ANN). I risultati mostrano non solo un rilevamento efficace delle anomalie e una classificazione in funzione del rischio generato, ma anche la capacità di offrire un output che garantisca la visualizzazione dei nodi contaminati sulla mappa della rete di distribuzione in esame, secondo una scala di gravità del rischio.

Introduction

In the chain of water distribution, the network is the most complex element to be analyzed and managed to deliver safe water to the users. This topic has recently been addressed with attention at legislative level. In fact, water introduced into the water distribution network can be subjected to various types of contamination in the network itself like, endogenous or exogenous, either accidental and deliberative (terrorist attack or sabotage), phenomena.
In general, the above-mentioned contamination risks equally affect all the parts of the network, resulting in difficulties for prevention and control. For this reason, some countries, especially those most sensible to the terrorist attacks (USA, Israel, Europe) have already started research programs aimed at the development of Online Water Quality Monitoring (OWQM) and Early Warning Systems (EWSs). Both are based on sensors installed in selected nodes of the network and are capable of quickly detecting contamination events.
The implementation of EWSs gives way to new interesting research topics, with particular reference to the technological aspects of the automatic online detection, to the employment of expert systems for the interpretation of the detected data, and to the definition of modeling tools for the system design and management.
This Thesis, which illustrates the research activity developed by the Author in this field, deals with some general, technological, and modeling aspects related to the application of online Monitoring and EWSs to the water distribution networks.
In particular:

After a brief legislative framework, **Chapter 1** explains the various types of contamination that may occur within distribution networks. Then, the general setup of the continuous monitoring systems and of the associated alarm systems are described, with particular reference to the selection of the observed parameters and to the expert systems for the interpretation of the detected data.

Chapter 2 provides a vast overview of the technologies currently available for the continuous monitoring of the water quality in the network, with particular reference to chemical, biological and radiological contaminations.

Chapter 3 investigates the main design and management aspects of the water quality monitoring systems in the network, with particular emphasis on the following issues:

- acquisition and transmission of the vast amount of data collected by sensors;
- optimal sensor locations within the network;
- identification of the contamination sources, once a contamination has been detected;
- definition of the urgent actions to be taken once a contamination has been detected;
- definition of the interventions to be implemented for the restoration of the regular distribution service.

Chapter 4 deals with the problem of the optimal location of monitoring stations within a water distribution network, with particular reference to the sampling of a small number of events, among all the potential contamination scenarios, to be taken into account in the optimization framework. For this problem, some innovative concepts and their applications to a case study are presented. In this chapter, the problem of selecting the objective functions to be introduced in the optimization is also addressed, using numerical comparisons. In all calculations, the assumption of conservative contaminant is considered.

After removing the conservation assumption, **Chapter 5** studies the actual behavior of the contaminants, once they have been dissolved in water. In particular, the results of numerical simulations carried out with engineering models (EPANET-MSX) are shown to analyze the *E. coli* fate and transport when chlorine is in the network, identifying the concentrations of the involved chemical/biological species. The experimental results available in the scientific literature, obtained in a pilot laboratory site created at the Civil Engineering and Geo-Environmental Laboratory of the Lille University (Villeneuve-d'Ascq, Lille - France), are also used to validate the numerical model.

Chapter 6 develops and demonstrates the technical feasibility of a prototype system for the early detection of biological contaminations within the network. This system will efficiently enable water operators to apply in real-time water quality control procedures. To this end, an automated statistical model and AI-supported algorithms are presented and validated using chlorine data obtained from the numerical simulations above-mentioned. In particular, the algorithms are developed in the field of the machine learning: after being trained on the standard conditions of water quality, they are able to recognize deviations from the baseline, identifying anomalies. The selected parameters for the training phase, thus for the definition of the standard conditions of water quality, are the chlorine and the TOC.

Chapter 1

Monitoring and early Detection Systems for contamination events in Water Distribution Networks: general approach

1.1 Introduction

The existing European legislation in the field of water intended for human consumption prescribes minimum requirements regarding the physical, chemical, biological and radiological characteristics that water must have throughout the all water distribution system (WDS).
To this end, following the European Directive 2015/1787, a new approach for the consumer safety is being imposed in Italy, implementing the Water Safety Plans (WSPs) in accordance with the model that has been introduced for over a decade from the World Health Organization (WHO, 2005).
WSPs aimed at assessing the risk of water contamination throughout the all production and supply chain (from capture, to distribution, and delivery), as well as defining the consequential management strategies.
The guidelines for the WSPs implementation are set out in a few documents of the World Health Organization (WHO, 2008; WHO, 2011), and in some international standards, such as EN 15975-2 (2013). In Italy, the guidelines for the WSPs implementation were formulated by the Italian National Institute of Health (Istituto Superiore di Sanità, 2014).
The main purpose of the WSPs is to ensure that the levels of the delivered water quality are adequate to protect human health. This objective is pursued through a

series of actions ranging from the analysis of the hydraulic system with its related contamination risks, to the definition of the monitoring measures and the action plans necessary for the priority risk and emergencies management.
Therefore, assessing the entire drinking water production and supply chain, the water distribution network is surely the most complex element to be analyzed and managed as regards the safety of the water quality delivered to the user.

1.2 Contamination events in Water Distribution Systems

Ignoring the eventual cases of contamination of the supply sources and the inadequate operation of the treatment plants, all the water entering in the water distribution network, even thought it has qualitative characteristics that meet the necessary requirements, can be subjected to contamination phenomena in the network itself, which are typically included in one of the following three main situations.

1.2.1 Endogenous contamination

An endogenous contamination is due to phenomena that occur inside the pipes, such as precipitation and flocculation of certain substances, decay of disinfectant agents with consequent bacterial growth, corrosion of the wall pipes, trialomethanes formation.
The development of bio-film that is formed on the inner pipe walls is also very important: in fact, it is a crucial problem in the control of the drinking water quality due to the presence and persistence of several microbial species (US-EPA, 2002). As highlighted by researchers of the Italian National Institute of Health (Bonadonna and Della Libera, 2005), the magnitude of the bio-film development is conditioned not only by the presence of nutrients, but also by the water residence time in the network and the water temperature. It is thus important that the water residence time in the network is the least possible. In terms of hydraulic network functionality this means that the velocities in the pipes have to be sufficiently high and the water path from the source nodes to the distribution ones has to be minimized.

1.2.2 Accidental contamination

An accidental contamination is usually related to the infiltration of dirty water, mud and other contaminants from the external environment. In general, this can occur with reference to two different situations:

i. Infiltration of contaminated water from the environment surrounding the pipeline due to hydraulic sealing defects (always present with lower or higher magnitude), especially nearby the joints.
ii. Backflow into the network of water coming from pressurized circuits erroneously connected to the main network.

The first case occurs when the pressure of dirty water in the surrounding external environment is higher than the one inside the pipe. Since generally the dirty water presents in the environment surrounding the pipeline is characterized by atmospheric pressure (or slightly higher), infiltrations occur when there is no pressure or even there is depression in the pipes. Therefore, with the exception of rough design errors that determine depression in some parts of the network, the internal pressure annulment is most of the time associated with the drainage of pipes related to maintenance works, which should be reduced to the required minimum. If possible, an alternative could be interventions without service interruptions.
The depression phenomenon in the pipes may also occur if water is drawn from the network by means of pumping systems without a hydraulic disconnection on the suction pipe. In order to reduce the risk of this kind of accidental contaminations, the elevation layout of the water distribution network should also be carefully designed with respect to other pipelines' structures. Particularly in the sewage systems, it is necessary to avoid the stagnation of contaminated liquids in contact with the pipelines of the WDS.

The second case of accidental contaminations is associated with erroneous connections between the drinking water network and other pressurized systems, which in turn are connected to a second network, conveyor of non-drinking water. In case there is a reversal of the pressure gradient between the drinking and non-drinking networks, the water from the pressurized plants returns to the water distribution network, unless an effective backflow preventer valve is set in. These situations of erroneous connections between different networks typically occur at household users employing private wells for non-drinking purposes, and at industrial users. In fact in these latter cases, the drinking water network is sometimes connected without a hydraulic disconnection to technical plants, working with non drinking process water.

1.2.3 *Intentional contamination associated with terrorism attacks or sabotage*

Intentional contamination actions can be carried out through the injection of chemical, biological and radioactive contaminants in one or more spots of the network.

Although structurally protected from intentional contamination because of their primarily underground location and their pressurized functioning, water distribution networks are one of the most vulnerable components in the supply of drinking water, according to many experts (US-EPA, 2005a). In fact, all devices (such as air-relief valves, bottom drains, hydrants, public water fountains) that connect buried pipes to the outside environment and above all, the user supply taps when connections are not equipped with suitable backflow preventer devices are potential input sites.

In general, the above explained contamination risks concern the all network, resulting in difficulties for the prevention and control. For this reason, some countries, especially those most vulnerable to the terrorist attacks (USA, Israel, Europe) have started for a long time research programs aimed at the development of an Online Water Quality Monitoring (OWQM) and Early Warning Systems (EWS). Both of them are based on sensors which are installed at selected nodes of the network and are capable of quickly detecting contamination events.

Various cases of intentional and also accidental water contamination that historically occurred advise us of the necessity to improve water monitoring. For instance, in 1972 a right-wing neo-Nazi group acquired 30–40 kg of typhoid bacteria cultures with the intention to use this against water supplies in Chicago (Kupperman and Trent, 1979). More recently, in 2000, workers at the Cellatex chemical plant in northern France dumped 5,000 litres of sulphuric acid into a tributary of the Meuse River, when they were denied workers' benefits (Gleick, 2006). In other episodes, the introduction of a contaminant would have affected nearly 4,000 households; officials suspect that the incident was related to professional work beyond vandalism (Groover, 2008) and again, few officers were arrested before executing a plan to poison a tank in Khao that supplies water to American troops in Jordan (Times, N. Y., 2003). Accidental water contaminations were also registered, as for example in Greenville, South Carolina, where the town's water supply was threatened with a castor oil poisoning. The attacker caused changes to the federal regulations regarding the number of hours that ground truckers were allowed to drive without rest (Gleick, 2006). Another exemplary incident is the contamination of drinking water with treated wastewater in Nokia (Finland) in 2007; this incident resulted in 8,453 cases of gastroenteritis, with costs exceeding 4.6 million Euros for clean-up (Williamson et al, 2014), reimbursed hospital expenses, claims for damages etc.

1.3 General approach of continuous monitoring and Early Warning Systems (EWSs)

The conventional monitoring of the water quality provided by the network is based on sampling at some taps, as well as on chemical and microbiological analysis performed in the laboratory (or on-site through suitable kits). In general, it allows a complete chemical and microbiological characterization of the supplied water, along with the research for almost any contaminant.
However, this type of monitoring is not capable of supporting an EWS due to the sampling which does not usually occur in a short term and due to the long time that characterizes some types of analysis (especially the microbiological ones).
Therefore, there is the need of a continuous monitoring system (or in any case characterized by a very dense temporal discretization) implemented through devices directly connected to the distribution network. They have to be capable of automatically performing the desired analysis *in situ* and in a very short time, as well as transmitting its results to a central control apparatus responsible for the outcomes interpretation and the implementation of the scheduled alarm actions.
An EWS is not limited to a collection of surveillance technologies; it is an integrated system for i) monitoring sensors (ii) water analysis (iii) interpretation and reporting of the results (iv) communication of the results in order to make decisions that are protective for public health and minimize unnecessary community concerns. The desired characteristics for an ideal EWS were

indicated by the Italian National Institute of Health (2004), the European Commission (2013) and the Environmental Protection Agency (EPA, 2005). Some of them include:

- high degree of automation, including automatic sample archiving;
- rapid response;
- detection of a sufficiently wide range of potential contaminants;
- acquisition, maintenance, and upgrades at an affordable cost;
- identification of the source contaminant and accurate prediction of the location, together with the concentration downstream of the detection point;
- minimal false-positives/false-negatives;
- function continuously;
- data acquisition at different locations of the network and their transmission to a processing center;
- equipment of an expert system capable of interpreting analytical results and providing support for the development of the strategies to contain the contamination effects.

Currently, there is no EWS with all the features listed above. However, there are some technologies that can be used to build an EWS as they show some basic characteristics: (i) functioning as an automated system that allows remote monitoring (ii) a rapid response, and (iii) the detection of contaminants maintaining acceptable sensitivity. Without these three characteristics, an EWS cannot be reputed an effective and reliable system. While emphasis is placed on these three characteristics, the other presented characteristics cannot be ignored designing an EWS. For example, the rate of false positive/false negative results and the sensitivity of the methods used to interpret the results should be considered.

The monitoring system is still the main component of the EWS thus, it must be carefully designed. Hasan et al. (2004) proposed a *tiered monitoring*, that consists of two stages: the first might provide a continuous screen for a range of contaminants that could pose a threat to public health. A positive result from the first stage would trigger confirmatory analysis using more specific and sensitive techniques, and a positive result from the confirmatory analysis would trigger a response action.

WaterSentinel project (US-EPA, 2005b)[1] considers that a complete EWS must include few important components, that are i) the online water quality monitoring; ii) the sampling and the analysis; iii) the enhanced security monitoring; iv) the consumer complaint surveillance; v) the public health surveillance.

1.4 Features of online contamination monitoring systems

Considering the wide spectrum of possible contaminants, the real-time identification of risks in WDSs is a very difficult scientific and technological challenge.

While there are laboratory technologies capable of measuring all of the substances in the water, the analysis abilities of devices that can be used *in situ* for continuous and automatic monitoring are very limited. A promising approach (Roberson and Morley, 2005; Janke et al., 2014) considers the continuous (or almost continuous) detection of the most common physical and chemical water parameters such as flow rate, turbidity, pH, temperature, conductivity, pressure, chlorine, fluoride, nitrate, particle count, Total Organic Carbon (TOC), Oxidation Reduction Potential (ORP).

Table 1.1 shows the top 10 parameters monitored online from raw to distributed water by different water companies around the world (BTO Report.028, 2008).

[1] The name comes from the collaboration of EPA and the Office of Water Security initiative-WSi and the program was developed in partnership with drinking water utilities and other key stakeholders involving the design, deployment, and evaluation of a EWS for drinking water systems.

Table 1-1. Top 10 parameters monitored online by drinking water companies in the USA, Belgium and the Netherlands, the United Kingdom and Australia (BTO Report.028, 2008).

Rate	Parameter USA (%) n=52		Parameter B-N (%) n=10		Parameter UK (%) n=7		Parameter Australia (%) n=6	
1	Flow rate	100	Flow rate	100	Flow rate	100	Flow rate	100
2	Turbidity	89	Turbidity	100	Turbidity	100	Turbidity	100
3	pH	79	pH	90	pH	100	pH	100
4	Water Temperature	77	Oxygen	90	Chlorine	100	Water Temperature	100
5	Conductivity	39	Water Temperature	80	Water Temperature	86	Free Chlorine	100
6	Particle count	37	Conductivity	60	Conductivity	72	Pressure	83
7	Fluoride	21	Ca/Mg/Hardness	50	Pressure	72	Conductivity	83
8	Oxygen	17	Biomonitors	50	Iron	72	Fluoride	83
9	Chlorine	14	Particle count	30	Oil in water	57	Particle count	83
10	TOC	14	Spectral Absorption	30	Nitrate	57	Total Chlorine	50

Through these investigations, basic information on the qualitative characterization (expected values) of the delivered water is acquired, together with the quantitative characterization in terms of pressures/flow rates of the delivery service, if reputed useful. The detection of water quality alterations due to contamination events occurs through the implementation of suitable software (expert systems) capable of interpreting the variations eventually measured in the values of the above mentioned chemical-physical parameters. Hall et al. (2007) investigated several water quality sensors in terms of their response to a contaminant that could be injected into the water. All the examined sensors were continuous online monitoring device and they were tested in a pilot-scale system, which was a re-circulating pipe loop distribution system simulator. The results showed that the most effective responses came from the free chlorine and the TOC sensors: the first sensors responded to all contaminants, although some contaminants did not react significantly with chlorine while the second sensors responded to all the organic (carbon-containing) compounds. Similar results were obtained within the project WaterSentinel (US-EPA, 2005b), which highlighted how the free chlorine and the TOC are potentially the most useful indicators of contamination, since they detected 28 out of the 33 tested baseline contaminants. In particular, the results illustrated that free chlorine is the most sensitive indicator of contamination, showing significant changes from baseline values at concentrations of one to two orders of magnitude below lethal concentrations. Also, TOC was indicated as a useful parameter for detecting the presence of many organic compounds, with a sensitivity ranging from a few tenths of a mg/L to more than 1 mg/L, depending on baseline levels and variability.

Based on these studies, US-EPA recommended the free chlorine, and the TOC as primary contamination indicators, while it suggested the Oxidation Reduction Potential (ORP), the pH, the conductivity and turbidity, as secondary indicators. In fact, ORP usually shows a behavior like the chlorine residual, of which it can corroborate an observed change. ORP is also employed in systems that use a chloramine residual disinfectant because certain oxidation reactions can take place without consuming chloramines. Conductivity and pH are both important to aqueous chemistry and they may be valuable in understanding observed changes in other parameters, such as free residual chlorine. Turbidity is an untrustworthy indicator of contamination but, as well as conductivity and pH, it may be considered for proving the understanding in the changes of other measured parameters.

Clark et. al (2002) also confirmed that chlorine residual and pH had been previously considered in research as surrogate candidates for on-line monitoring of distribution systems.

A summary of the most exhaustive study on this subject was offered by US-EPA (US-EPA, 2005a), while some guidelines for the design of on-line contamination detection systems were presented by Pikus (2004). Their consequent instruction is a more functional application of what was said above. It is based on the installation of water quality sensors located throughout the WDS combined with a public health surveillance system, as well as, with a customer complaint monitoring program for the detection of a wide range of contaminants (Janke et al, 2014).
Several typologies of new technologies will strongly impact the advancing online measurement of contaminants, but presently the field is not sufficiently mature to provide devices that would meet the needs of drinking water utilities. In fact, a new generation of on-line monitoring tools has emerged in recent years; however, an effective implementation of these tools has not been realized for a number of reasons: (i) they do not meet practical utility needs, (ii) their cost, reliability and maintenance are unsatisfactory, and (iii) the data handling and management along with the ability to produce meaningful operational information is still to be grasped (Van der Gaag and Volz, 2008).

1.5 Early Detection Software

The detection of a contamination event requires that the related variations in the values of the measured parameters can be distinguished from the normal daily and/or seasonally fluctuations (the so-called background noise), which are due to several reasons, such as the contribution of different supply sources variable with the demand. Another complexity lies in the fact that water quality may have different characteristics at various points of the distribution network in relation to the multiplicity of supply sources, the different materials and the age of the pipes. It is therefore necessary to use specific algorithms, essentially based on statistical methodologies, which are able to highlight abnormal variations in the observed parameters compared to normal fluctuations.
In practice, according to the approach described above and through a vast variety of water sensors, an EWS is based on a continuous acquisition of the values of the measured parameters and their transmission to a Supervisory Control and Data Acquisition (SCADA) integrated with an Early Detection software that reads and interprets the acquired data by distinguishing the abnormal variations

from the normal fluctuations. The abnormal variations then require prompt attention or intervention.
To this purpose, CANARY software (open source), developed by Sandia National Laboratories in collaboration with US-EPA, (Hart et al., 2007; US-EPA, 2010; US-EPA, 2012; Hagar et al., 2013) is often cited in the literature.
CANARY has been developed to provide both real-time, and off-line analysis tools, giving particular emphasis to the following features (i) the use of a standard format for input and output of water quality and operations data streams (ii) the ability to connect various detection algorithms, both in MATLAB and in compiled library formats for the testing and the evaluation by using a well defined interface (iii) an operation approach that simulates the utility operator mode iv) comparison of tools for different evaluation metrics, including Receiver Operating Characteristic (ROC) curves, time to detect, and false alarm rates.
Traditionally, water utilities use set points (thresholds) to identify changes in water quality parameters: set points provide alarms when the actual value of the water parameters goes above or below the set point value. For example, free chlorine levels nearby zero need to be communicated immediately to an operator. Hence, the discussion focuses on the deepening of the different detection algorithms used by CANARY to identify the water quality values that are significantly different from the background values whether or not they exceed the set point limits.
In fact, CANARY provides a platform within which different event detection algorithms can be developed and tested. These algorithms process the water quality data at each time step to identify anomalies in the water quality. CANARY works by reading in real-time (online) time series of data coming from any type of water sensor available on the market; it commonly uses from five to seven sensors, including free chlorine, pH, conductivity, TOC, ORP, temperature, and turbidity. For the analyzed time series, each quality signal (S) is constituted of the background water quality (B), any deviation from that background (D) due to an anomalous event, and the noise (N) intrinsic in the water quality monitoring system. All of the algorithms assume that past water quality observations can be used to accurately predict future water quality values under normal conditions.
The algorithms are planned to continuously update and learn the characteristics of the background signal in order to take them into consideration when a new water quality observation is presented.

Four main steps are involved in the event detection algorithms: (i) estimation of the future water quality values (ii) comparison of the estimated values with the measured values as they become available and calculation of the "residual" as the difference between estimated and observed values (iii) comparison of the residual at each time step against a threshold: those that exceed the threshold are "outliers", and (iv) use of the probability distribution (in particular the so-called Binomial Event Discriminator - BED algorithm) to determine the probability of an event from the number of outliers over a given number of time steps. Therefore, the final output is an indication of the probability of a water quality condition existing at each time step.

During the estimation phase, CANARY considers a pre-defined set of previous time steps to predict the values of the next time step; CANARY can also easily combined all the diverse signals characterized by different units of measurement. Regarding the estimation process, two approaches are available within CANARY, that is the linear filtering and the multivariate nearest neighbor.

The first one applies an optimal set of weights to each of the previously measured standardized observations for each water quality signal. The weights are calculated using an auto-covariance function independently computed for each signal, reflecting the importance of the previous value in the prediction of the next one.

The second approach considers the set of values at each time step as a point in a *n*-dimensional space. All of the data from previous time steps can be mapped as points in this space, and their mutual distance is evaluated. At each time step a new point is created and the closest point serves as the predicted value for this time step.

Starting from this process, the residual and the outliers are evaluated to get the final probability, as above explained.

A comparative analysis of the performance of CANARY and the other commercial Early Detection software (OptiDES, Ana :: tool, BlueBox, Event Monitor) has been published by US-EPA (2013b).

The Evaluation Center in Cincinnati (Ohio), together with the researchers in sensor industries such as the Hach Corporation in Loveland (Colorado), were also involved with water quality sensor testing, developing the *Guardian Blue EWS* to detect, alert, and classify a wide variety of threat contaminants in drinking WDSs (Kroll, 2006).

The optimal Event Detection System *(optiEDS*) monitors a set of water quality and operational data; once an abnormal combination of the monitored data set has been detected, the system alerts and reports the "suspicious" parameters. The basic algorithm of optiEDS uses trend analysis to monitor deviations from a steady parameters baseline. The innovation was here the possibility of embedding the water network operation logic into optiEDS, empowering the water utility's engineers and operators with specific knowledge of the system. The main capabilities of optiEDS are thus (i) the monitoring of a large set of water quality and operational parameters (ii) the real-time alarming for abnormal changes in the water quality (iii) the definition of a normal dynamic baseline for the monitored parameters, and (iv) the possible custom adjustments into a water network using the utility's knowledge.

Ana::tool evaluates measurement data that have been cleaned by the validation module. Once it has identified the normality of the analyzed data, it is able to trigger an alarm when a significant deviation from normality is detected, enabling the operators to timely react to faults in the monitored system. Combining, Static Alarms, Dynamic Alarms, Pattern Recognition and Spectral Alarms, ana::tool detects an alarm and let the users evaluate a feedback in order to learn which alarms are real and which ones are false (mostly associated with the normal changes in the water quality). Gradual composition changes (e.g., seasonal variations) are accounted for by automatic training on a moving time window. Among the main features of ana::tool, it should be highlighted the auto-correction of data based on threshold, outlier and noise analysis, as well as the capability of exploiting the enormous information contained in UV spectra, which provide the most sensitive and stable data source for event detection. It shows the ability of training itself on any type of data coming in, learning automatically which data are useful for event detection, and which ones are useless. Ana::tool also weighs automatically the results as appropriate when a variety of algorithms are applied in real-time analysis; finally, all the event information is aggregated into a "traffic light" output and a "deviation from normal” output.

Regarding *BlueBox***,** the EPA Challenge contributed to the development of the product. In fact, it acquired the ability to define and incorporate operational variables, such as discrete variables (e.g., indication of pumps and valves on/off status), or substantial changes in the measurements of operational parameters like flow, pressure and water direction. This allows the system to cross reference and correlate between suspected quality events and the operational environment, increasing the certainty and the accuracy of the alarm. BlueBox is also able to distinguish the cause of the alert, whether it is a result of a water quality event or an equipment malfunction. It can exchange data with any standard industrial automation system and, as the other detection software, presents a self-learning mechanism based on event classifications, facilitating the users with the classification "true-false". BlueBox can also do more than the other detection algorithms because it integrates several time parameters as part of the system inputs, enabling the detection of abnormalities based on seasonal parameters (time of the day/month of the year). The level of false alarms is thus reduced from seasonality effects.

As the other software, the *Event Monitor* (Hach Company) evaluates data from multiple sensors, interprets the significance of water quality deviations from the established baseline (e.g., deviations due to operational fluctuations), calculates a "fingerprint" of each system event registered in the software library, and provides a single trigger signal. Operators can adjust the trigger threshold, as well as other simple settings, and they can label event fingerprints for simplified identification if the event recurs. Also, Hach Event Monitor incorporates the ability to learn specific system dynamics, and the self-tuning capability, which modifies the definition of what constitutes an abnormality according to the variability encountered for a given time frame at a specific site. Both these features improve the water quality conditions, eliminating many false alarms due to the noise. Hach Company also developed an Agent Library to enhance the capabilities of the Event Monitor when used as part of the GuardianBlue Early Warning System. The Agent Library is capable of classifying threat contaminants so that they are easily differentiated from water quality events.

Since in this context CANARY is still undoubtedly the most used software, some results related to its application in the Cincinnati case study are given hereafter. Allgeier et al. (2008) reviewed the first year of operation for the Cincinnati Pilot’s online water quality Contamination Warning System (CWS). Allgeier reported that 3.7 alarms were generated per day across the network of 17 monitoring stations (15 were in the distribution system and 2 were located at the treatment plants) but they were triggered by regular operational changes for the most part. Consequently, the number of alarms was too high to be sustainable.

Later, in 2011, the same authors reported that 92% of the alerts were invalid, while the 8% valid were due to unusual plant conditions, changes in the process at the treatment plant, maintenance or repair activities in the distribution system, main breaks, or verified water quality anomaly with unknown causes (Allgeier et al., 2011). As a result, Allgeier et al. (2011) tested CANARY software, using the Cincinnati Pilot field and simulating several events (1,588). For the simulated events, their detection rate of the true positives was 40%, leaving the 60% as false negative. However, the authors showed that even if only 40% of the simulated contamination incidents were detected, those undetected scenarios caused low consequences.

The approach based on the analysis of the measured variations with reference to the most common physic-chemical parameters must be therefore enhanced by the search for specific contaminants that may cause a public health threat. In addition, investigating the opportunities to improve the event detection, Vugrin et al. (2009) used historical water quality data from the utility to identify recurring patterns and saved those patterns in a library that can be accessed during online operation. This pattern matching capability was implemented within CANARY in order to demonstrate a decrease in false alarms.

Finally, a significant false alarms decrease was noticed through the method proposed by Koch and McKenna (2011), according to which data can be combined from multiple stations considering the location and time of individual detections.

Kulldorff's scan test can thus find statistically significant clusters of detections, which reduce the false alarms resulting from background noise and indicate time, as well as source location of the contaminant.

Concluding, detection software has to be based on the evaluation of baseline water measured parameters. However, when changes are detected, additional analyses should be carried out in order to identify real contamination threats, which need conscientious and secure response activities.

Chapter 2

Water Quality Sensors

The needs of EWSs have encouraged the development of different types of devices based on various technologies (Storey et al, 2011; European Commission, 2013; Banna et al, 2014). Some of these devices are simple sensors already manufactured and marketed for a long time, such as those that can detect the most common chemical-physical water parameters (e.g., Chlorine, Total Organic Carbon, Turbidity, pH, Conductivity, etc.). Others, are monitoring stations which combine one or more sensors together.

There are also latest-generation sensors: some of which exploit very innovative physical-chemical principles (e.g., Refractive Index), others are sensors that directly detect specific chemical, biological or radioactive contaminants. For instance, the newest and most expensive one use the ability of Algae or Fluorescent Bacteria to differently react in the presence or absence of pollutants.

The most common devices are listed below divided into three categories:

- sensors, even multi-parametric, for detecting the physical/chemical water parameters (hereafter called water quality parameters);
- monitoring stations,
- sensors for detecting specific contaminants.

2.1 Sensor for the water quality parameters

Since the detection of water contaminations through the analysis of the most common water quality parameters (shown by EPA as possible indicators of water conditions) is still a worldwide used technique, in recent years several manufacturers have developed a vast variety of sensors.

Regarding the chlorine, James et al. (2005) developed a membrane-covered amperometric sensor, providing direct chlorine response without the need for chemical reagents. Examples of chlorine sensors are:

- *Series B20 Residual Chlorine Recorder* produced by Analytical Technology, Inc.;
- *AccuChlor 2 Residual Chlorine Measurement System* or *CL17 Free Residual Chlorine Analyzer* produced by Hach.

As for the Total organic Carbon (TOC) measure, Hach Co. developed the *Astro TOC UV analyzer*, which combines a chemical and Ultraviolet (UV) oxidation technique in a low-temperature reactor (James et al., 2005). Another available sensor for the TOC measurements is *Phoenix 8000 UV-Persulfate TOC Analyzer* developed by TeledyneTekmar.

James et al. (2005) also presented a method for the measurement of the turbidity, according to which it is measured with a 90◦ scatter nephelometer, using a Refractive Index (RI) light source for stability and a sealed flow chamber to reduce bubble formation. The incandescent light is directed from the sensor head assembly down into the turbidimeter body and is scattered by suspended particles in the sample.

Examples of turbidity sensors are:

- *4670 Series Turbidity System* produced by ABB Instrumentation;
- *WTM500 On-line Turbidimeter* developed by Sigrist.

The pH can be measured through a differential sensor, containing two glass pH electrodes, one for sensing and another in buffer to serve as a reference electrode or through an amperometric method.

Conductivity is continuously measured by a two-electrode cell/four-electrode conductivity sensor or through the conductance method.

The differential Oxidation Reduction Potential (ORP) sensor contains a platinum-sensing electrode and a separate glass electrode in buffer to serve as a reference electrode.

Examples of sensors that can concurrently measure pH, conductivity and temperature are:

- *Water Distribution Monitoring (WDM) PipeSonde In-Pipe Probe* developed by Hach;
- *Quanta-Display/Transmitter Multiparameter Water Quality Instrument* developed by Hydrolab.

Recently, a more advanced application of the sensors lies in the simultaneously measurement of several water parameters in order to distinguish changes in multiple parameters and promptly identify the contaminations. Therefore, the goal is to use multi-parameter water quality sensors, that is the so called multi-parametric probes. Typically, they are based on the following types of water monitoring methods (US-EPA, 2005b):

- colorimetric and membrane electrode for chlorine;
- thermistor for temperature;
- membrane electrode or optical sensors for Dissolved Oxygen (DO);
- potentiometric method for ORP;
- glass bulb electrode for pH;
- nephelometric method or optical sensor for turbidity;
- conductivity cell method for specific conductance;
- ion-selective electrodes for Cl-, NO_3 and NH_4^+.

Featuring fully automatic operation and remote connection, they can be directly installed on a pipeline, or they can be put in monitoring stations located close to the pipe and taking the water sample with a frequency of few minutes.

For example, an almost thorough list of multi-parametric probes is mentioned below, including their features (Highsmith, 2004; US-EPA, 2005b; European Commission, 2013):

- *Six-CENSE* developed by Dascore - it measures chlorine, monochloramine or dissolved oxygen, pH, temperature, conductivity, ORP/REDOX;
- *WDM Water Distribution Monitoring PipeSonde In-Pipe Probe* developed by Hach - it measures pH, ORP, conductivity, turbidity, dissolved oxygen, pressure, temperature;
- *(WDMP)Water Distribution Monitoring Panel* developed by Hach - it measures chlorine, conductivity, pH, turbidity, pressure, temperature;
- *Kapta 3000 AC4* developed by Veolia under the European SecurEau project (www.secureau.eu/) - it measures residual chlorine, pressure, temperature, electrical conductivity and turbidity;

- *Spectro::lyser* developed by the Austrian industry S::CAN - it measures a selection of parameters chosen by the user between: Total Suspended Solids (TSS), Nitrate-nitrogen (NO_3-N), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), TOC, Dissolved Organic Carbon (DOC), UV254, O_3, Hydrogen sulfide (H_2S), Assimilable Organic Carbon (AOC), color, turbidity, temperature and pressure;
- *Carbo::lyser* developed by S::CAN - it measures the organic carbon load, represented by parameters like Spectral Absorption Coefficient (SAC), TOC, COD, or BOD, and at the same time, turbidity or TSS;
- *Nitro::lyser II* developed by S::CAN - it measures TSS and NO_3-N or turbidity and NO_3-N;
- *Multi::lyser*, also developed by S::CAN, is a combination of carbo::lyser and nitro::lyser - it measures organic carbon and nitrate;
- *Ozo::lyser II* developed by S::CAN - it measures turbidity and ozone;
- *Sulfi::lyser II/III* developed by S::CAN - they measure TSS, HS, H_2S and TSS, HS, NO_3-N, H_2S, respectively;
- *UV::lyser* developed by S::CAN - it measures turbidity or TSS and up to 4 freely chosen wavelengths between 190 and 720 nm (measuring principle: UV-Vis spectrometry);
- *Ammo::lyser III pro* developed by S::CAN - it measures Ammonium-N (NH_4-N) and temperature. Likewise, Ammo::lyser IV pro+pH and Ammo::lyser IV pro+NO_3-N measure NH_4-N, temperature and pH (with potassium compensation) and NH_4-N, temperature and NO_3-N, respectively;
- *Ammo::lyser II eco* developed by S::CAN - it measures NH_4-N and temperature. Moreover, *ammo::lyser III eco+pH* additionally monitors pH; *ammo::lyser III eco+NO_3-N* also monitors NO_3-N; *ammo::lyser III eco+Cl* adds Chloride measurements; *ammo::lyser IV eco+pH+NO_3-N* additionally monitors pH and NO_3-N; *ammo::lyser IV eco+pH+Cl* includes pH and chloride measurements;
- *Chlori::lyser* developed by S::CAN - it measures free chlorine (Cl_2 + HOCl + OCl_-) or total chlorine (free chlorine + combined chlorine);
- *Chlodi::lyser* developed by S::CAN - it measures chlorine dioxide;
- *Hyper::lyser* developed by S::CAN as an amperometric sensor - it monitors hydrogen peroxide, while *peroxy::lyser*, also developed by S..CAN as an amperometric sensor, controls the peracetic acid;

- *Condu::lyser* developed by S::CAN - it measures conductivity and temperature;
- *Redo::lyser eco* developed by S::CAN - it measures ORP and temperature. *Redo::lyser pro* is then well performing within a high temperature range;
- *PH::lyser eco* developed by S::CAN - it measures pH and temperature. In addition, *PH::lyser pro* performs well within a high temperature and pH ranges;
- *Fluor::lyser* developed by S::CAN - it measures fluoride and temperature;
- *Soli::lyser* developed by S::CAN - it measures TSS;
- *Oxi::lyser* developed by S::CAN - it measures dissolved oxigen and temperature;
- *I::scan* developed by S::CAN - it measures turbidity, UV254 absorption, color, and TOC, using the high performance of a multi wavelength spectrophotometer;
- *EventLab* developed by Optiqua - it is equipped with a highly sensitive sensor for Refractive Index changes (RI), which is an effective indicator of water quality because when any substance is dissolved in water, it changes the refractive index of the water matrix in proportion to its own RI, as well as, its concentration.

2.2 Monitoring stations

In some cases several multi-parameter probes are aggregated into a single monitoring station in order to evaluate a vast collection of water parameters at the same time, and to provide a timely and effective response.
For examples (US-EPA, 2005b), in terms of drinking water regulation the S::CAN industry also designed the *micro::station* for the online monitoring of the water quality parameters. The *spectro::lyser*, few S::CAN probes and a controller are assembled with all required flow cells, mounting fittings and pipe working conditions into a compact and versatile system. The S::CAN *nano::station* presents a further step forward since it is a super-compact and versatile system, where as in the previous case, the users only have to connect it to the water supply through the "plug & measure" process to receive a prompt

variety of available information regarding the water parameters. It combines the *i::scan*, different S::CAN probes and a S::CAN controller.
For an illustration purpose, some S::CAN probes are reported in Figure 2-1 and the two monitoring stations are shown in Figure 2-2.

Figure 2-1. S::CAN Probes a) Multi::lyser b) Carbo::lyser c) Chlori::lyser d) pH::lyser e) i::scan (www.s-can.at)

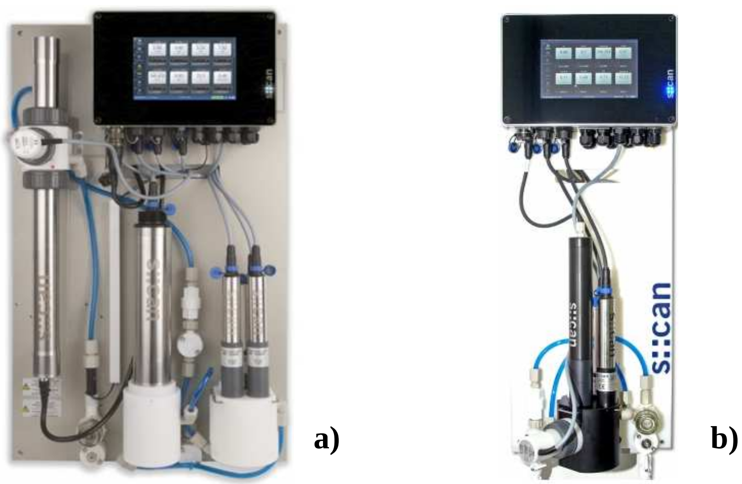


Figure 2-2. S::CAN Monitoring stations a) Micro::station b) Nano::station (www.s-can.at)

Also, the Hach Corporation developed its monitoring station, named *Water Distribution Monitoring Panel*. It combines several instruments into a single system for a more complete monitoring. The basic model includes the *Hach CL17 Chlorine Analyzer*, the *Hach 1720D Low Range Turbidimeter*, the *Hach/GLI pH Controller*, the *Hach/GLI Oxidation Reduction Potential Controller*, the *Hach/GLI Conductivity Controller*, and the *GEMS Pressure Sensor*. The expanded model also incorporates a *Hach Astro UV TOC analyzer* and an American *Sigma 900 MAX* auto sampler that can be activated to collect and archive samples when pre-specified set-point values are exceeded for any of the parameters being measured. The Hach *Distribution Monitoring Panel* continually measures these six or seven water quality parameters from a side stream of water in a municipal distribution system, and the results can be directly reported to the utility SCADA system.

The Hach *WDM PipeSonde In-Pipe Probe* (above explained) can be added to the system, being installed on any water pipe (at least eight inches diameter). It measures pressure, temperature, conductivity, turbidity, ORP, DO, chlorine concentration, TOC, and it is able to direct communicate with the SCADA system of a water utility. Therefore, the Hach *Event Monitor Trigger System* (Figure 2-3) allows an effective water monitoring, giving an alarm when water quality significantly deviates from the baseline.

Figure 2-3. Water Distribution Monitoring Panel with the Event Detection System by Hach Corporation (www.hach.com)

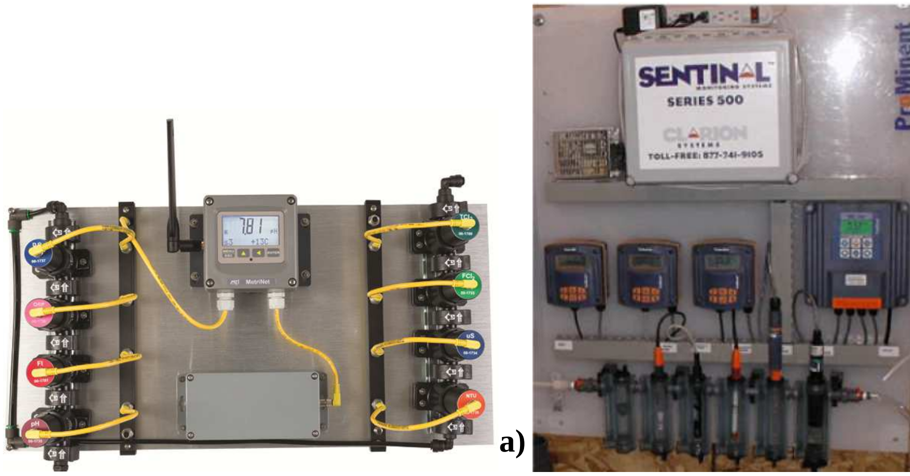
Dascore, Inc. developed a monitoring station named *Six-Cense*, which was designed for permanent insertion into a pressurized water main. It consists of electrochemical sensors mounted on a one squared inch ceramic chip layered with gold. It continuously measures six parameters, including chlorine or chloramine, DO, pH, ORP, conductivity, and temperature, using electrochemical methods, rather than reagents. The system can work remotely, with data reported to the utility SCADA system.

Emerson Process produced the *Model WQS Multi-Parameter Electrochemical/Optical Water Quality System (Model 1055 Solu Comp II)* which measures pH, conductivity, ORP, DO, free chlorine, and monochloramine by electrochemical methods. Two more parameters, turbidity and particle index, are evaluated through optical methods.

MetriNet, derived from *Network Metrics* and developed by Analytical Technology, Inc., let the user choose the desired parameters and integrate them in a monitoring package, suitable for continuous monitoring, alarming, and data collection. The system can measure free chlorine, combined chlorine, dissolved ozone, pH, ORP, conductivity, temperature, DO, and turbidity. The system provides several methods for delivering detected data, including cellular modem, Wi-Fi, wired Modbus, Ethernet/IP, or Profibus DP, as well as cloud-based data storage.

Sentinal, developed by Clarion Sensing Systems, integrated sensor data into a single display which can be remotely viewed. As in the case explained above, the user can choose among several parameters, including chlorine pH, temperature, flow, pressure, conductance, turbidity, ORP, DO, radiation, TOC, VOCs, and certain chemical weapons. Data can be transmitted via LAN or satellite link.

Finally, although designed for wastewater applications, *STIP-scan* produced by STIP Isco GmbH can be adapted to drinking WDSs and it can concurrently measure nitrate, Chemical Oxygen Demand (COD), TOC, Spectral Absorption Coefficient (SAC 254), total solids, turbidity and absorption in any specified range within the wavelength spectrum from 190 to 720 nm for detection of other compounds. The entire system is equipped with a controller, as well as, a bidirectional serial interface to transmit data. Examples of monitoring stations are represented in Figure 2-4.



a) b)

Figure 2-4 Examples of monitoring stations a) MetriNet b) Sentinal

2.3 New Detection Technologies

Even though sensors that measure the common water quality parameters are much more widespread, new devices have recently been developed that implement innovative detection techniques.
For examples, they are able to detect contaminants by utilizing the measure of the refractive index, the toxicity level etc.
For illustration purpose, the following common devices are reported, divided in three categories: the first is for the measurement of the toxicity level, the second concerns the detection of biological contaminants, and the third identifies radiological contaminations.

2.3.1 Toxicity Indicators

Most of the sensors that are able to detect specific chemical contaminants are intended for laboratory or *in situ* use, and cannot be directly installed on the distribution network pipelines.
However, few instruments that were born to detect chemical contaminations are also able to identify the toxicity level, being based on the use of microorganisms. They are named "biosensors" because they allow the detection of the presence of

water contaminants through the use of organisms: the changes in physiology or behavior of living organisms resulting from stresses induced by toxins are measured to indicate that there is an unusual condition in the water. Among the numerous biosensors, US-EPA (2005b) reported the *Tox Screen II* by Check Light and *MicroTox/DeltaTox* by Strategic Diagnostics for the online water monitoring.

In the following years, researchers from various industries noticed that many organisms are actually able to change their behavior in the presence of water pollutants. Mussels, for instance, change the frequency of valve opening and closing in response to toxins. *MosselMonitor*, developed by Delta Consult, can monitor chlorinated drinking water after a pre treatment to remove chlorine. Indeed, it is a biological EWS for continuous on-line monitoring of surface waters and drinking water, allowing for near real-time graphical presentation at a remote location or through Internet.

Then, few instruments are able to detect the photosynthetic activity of the algae: standardized algae are mixed with the sample water and the devices serve as a toxicity measurement, determining the percentage of active chlorophyll under illumination. In fact, damage to the algae (e.g., due to herbicides) causes a reduction in algae activity and activates an alarm. This principle is used by *Algae Toximeter* from BBE Moldaenken, by *LuminoTox* from Lab_Bell, and by *ALGControl* by MicroLAN.

Again, the *TOX control* developed by MicroLAN is a completely automated system that uses freshly cultivated lightemitting bacteria (*Vibrio fischeri*) as a biological sensor. The luminescence is measured before and after exposition to calculate the inhibition in percentage: as long as the sample toxicity is greater, the percentage light loss from the test suspension of luminescent bacteria increases. Other devices use the enzymes property of electron transport and the oxidative phosphorylation to monitor the redox state, or rather, the ratio of the concentration of the oxidized species, that is related to the toxic effects.

Finally, x-raies fluorescence technology can also be used for the detection of contaminant substances: ITN Energy Systems provided a smart, automatic early warning sensor to continuously trace levels of toxic metals in drinking water on a ppb scale.

For the purpose of illustration, examples of the described devices are reported below in Figure 2-5.

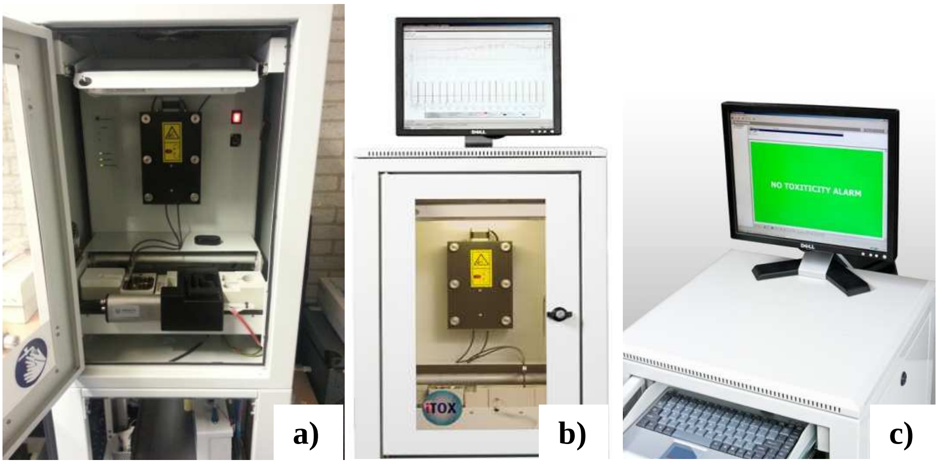


Figure 2-5 Examples of monitoring devices a) ALGControl b) TOXControl with its Integrated c) Toxicity Detection (www.microlan.nl)

In details, Figure 2-5a represents the *ALGControl*, which uses LEDs for fluorescence excitation. When chlorophyll molecules absorb light, a fraction of the energy absorbed is reemitted as fluorescence. As algae of the same classes contain a similar quantity and quality of pigments, it is possible to differentiate divisions of algae by their fluorescence excitation spectrum. Figure 2-5b shows the *TOXcontrol*, which uses a decrease in luminescence of the luminescent bacteria as an effect to measure the toxicity of water samples while the automated processing to detect an anomaly is illustrated in Figure 2-5c (www.microlan.nl).

2.3.2 Biological Contaminants

As in the previous case, even for biological contaminants most of the sensors are unable to detect specific pathogenic organism in drinking water due to the microbial culture time. Actually, culture methods are relatively slow, requiring at least 24 to 48 hours but the water monitoring should be rapid and prove results in two hours or less to be efficient. For this reason, only few technologies can be implemented for the online water monitoring and they cannot be specific for a single pathogen.

For instance, flow cytometry was used to distinguish some microorganisms on the basis of differential light scatter properties with the addition of fluorescent tags. In fact, a mono-disperse suspension of cells flows past a laser beam and the device measures properties of each cell, such as size, granularity, green fluorescence, red fluorescence, and far red fluorescence intensities. An example of this methodology is employed in *Microcyte Aqua* by BioDetect, which is a stationary device suited for online and continuous water surveillance.

In addition, *BioSentry* from JMAR Technologies is able to identify Cryptosporidium and other microorganism (e.g., Giardia) in water matrix particles by means of the light scattering technology, which is a simple scanning procedure that provides information about the presence of particles of a certain size (US-EPA, 2005b).

Finally, two sensors available on the market are able to detect a specific biological contaminant, the *E. coli*. The first one is called *COLIGUARD*, developed by the Austrian start-up Mb Online GmbH, and it indicates the *E. coli* by the optical analysis (luminescence) of the enzyme ß-Glucuronidase. A second version of the instrument also detects coliforms, analyzing the activity of ß-Galactosidase (European Commission, 2013). Similarly, the second sensor, which is the *BACTcontrol* from MicroLAN, measures the specific enzymatic activity of β-galactosidase (for coliforms), β-glucuronidase (for *E. coli*) and alkaline phosphatase as indicators of bacterial contaminations. The enzyme activity is detected by adding reagents that contain a fluorescent indicator: there is an increase in fluorescence when the enzyme is present in the sample. As the latter two sensors are the most innovative in their field being able to detect specific biological contaminants, they are represented in Figure 2-6.

Figure 2-6 Examples of specific biological detectors a) COLIGUARD b) BACTcontrol (www.mbonline.at, www.microlan.nl)

2.3.3 Radiological Contaminants

Radiation is another possible cause of contamination in WDSs. For this reason, the Federal Water Pollution Control Act (Clean Water Act), the Safe Drinking Water Act (SDWA), and the Maximum Contaminant Levels (MCLs) have recently addressed protection of water systems from radiation and other contaminants.

Until the last decade, radiation did not require continuous monitoring; however, since the terrorism is a major security concern in the U.S., as well as in many other nations, even the laws in this field have to become stricter and the real-time monitoring of radiation is turning out to be important for the immediate detection and response. The available systems detect the total amount of radiation (including alpha particles, beta/photon emitters and gamma radiation), alert operators but many of these do not identify the specific contaminant (US-EPA, 2005b). Concerning these relevant issues, general information are available on the EPA's website or on the Multi Agency Radiation Survey and Site Investigation Manual (US-EPA, 2000), developed by EPA, DOE, DOD, and the U.S. Nuclear Regulatory Commission.

Hence, among the various technologies available on the market, the following are those related to the online drinking water monitoring.

Technical Associates produces several radiation monitoring instruments which are sensitive to Alpha, Beta, Gamma, X-ray, Neutron and Positron Radiation.

Regarding the drinking water, it marketed (i) *MEDA-SP*, for the continuous monitoring of intentional contamination or accidental spills of gamma radiation into the water source (ii) *SSS-33DHC* and *SSS-33DHC-4193* to continuously monitor and detect tritium leakage, and (iii) *SSS-33M8 monitor194* for the monitoring of tritium in water (www.tech-associates.com).
The *3710 RLS Sampler192* is also avalable by Teledyne Isco, which detects radionuclides and continuously monitors water for all types of radiation.
For the purpose of illustration, *MEDA-SP* and *SSS-33DHC* are represented below in Figure 2-7.

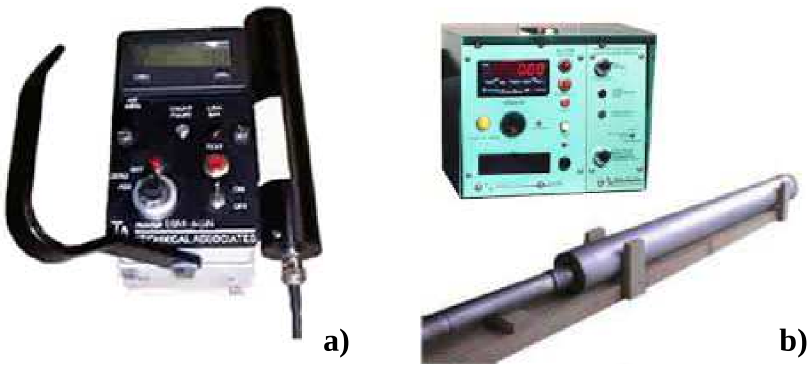


Figure 2-7 Examples of specific radiological sensors a) MEDA-SP b) SSS-33DHC (www.tech-associates.com)

2.4 Application Cases for Sensor Use

In order to demonstrate and prove what has been said in the previous paragraphs, some application cases of the use of the presented sensors are reported (www.s-can.at).
In these real applications, S::CAN also contemplates terminals and software to manage data, together with the installed monitoring stations. For examples, *moni::tool* is a new platform for the management of an almost unlimited number of stations, online probes, analyzers and parameters; *vali::tool* automatically detects, marks and corrects untrustworthy data, distinguishing outliers, noise and discontinuous data; *ana::tool* defines the normality for the baseline data, identifies unusual conditions and let an alarm start when a significant deviation is detected; the terminal *con::lyte* displays the readings of all S::CAN probes and

sensors connected on site through backlight LCD; *con::cube* and *con::nect* are respectively a compact terminal for data acquisition and a commercial notebook for operating the *spectro::lyser* (www.s-can.at).
It is important to note that the cost of the analyzed devices varies from a few thousands of Euros for the simpler multi-parameter probes up to some tens of thousands of Euros for probes that detect the presence of toxic substances and bacterial (e.g., Coliforms and *E. coli*). Their diffusion within the normal management of WDSs will only enhance the market, with consequent reduction of costs and interest in the production system, investing in the research and in the development of new technologies.

Over the last few years, the small communities of the First Nations in Canada had to face with the management of WDSs quality, as contamination events had reached a considerable number, causing a series of annoying accidents for the population. For this reason, the Canadian government launched a relevant project to provide safe drinking water to the communities of the First Nations, improving the infrastructure. More in detail, the government decided to install a remote monitoring of drinking water quality to ensure that industry operators can be promptly alerted in case of anomalies. This monitoring system consists of micro-stations, designed for the on-line monitoring of the most common water quality parameters. Each stations can combine up to four different s::can probes previously discussed (*spectro::lyser*, *ammo::lyser*, *chlori::lyser* and *pH::lyser*) and the terminal *con::cube*, being able to measure a total of ten parameters, including TSS, COD, BOD, Electric Current (EC), pH, temperature, NH_4, DO, NO_3-N, chlorine and free chlorine.
The stations were installed in combination with *moni::tool*.
The collected data are then transferred in real time to a computer network that connects all the Firs Nations.

In order to detect any problems caused by pipeline deterioration and to ensure good drinking water quality, the *i::scan* probe was installed in the city of Zurich (Swiss). It is a revolutionary and cost-effective spectrometer that uses the latest LED technology to measure the absorption spectrum, being able to quantify different parameters, that is turbidity, UV254, TOC, BOD and color. *I::scan* was installed in the pipes using a unique fixture that can withstand pressures up to 10 bar. The probe also has a valve to close the connection to the network: this

allows its removal from the pipe for inspections and cleaning procedures without interference with the regular water flow.
Here also, the probe was installed in combination with *moni::tool* in order to manage several monitoring stations.

In Bratislava, since the water quality from the sources is high, only the chlorination treatment is carried out, which excludes any microbiological growth during the distribution phase. To ensure that the possible contamination of one of the sources does not compromise the high water quality, the Bratislava Water Company (BVS) has sought a system capable of monitoring the various water sources coupled with robust event detection devices along the network to send an alarm in case of an event. To be able to evaluate a wide range of parameters, the UV *spectro::lyser* probe was installed and therefore, the measured parameters are: absorption spectrum (whole), TSS, turbidity, NO_3-N, COD, BOD, TOC, DOC, UV254, color, O_3, H_2S, Assimilable Organic Carbon (AOC), Benzene-Toluene-Xylene (BTX), temperature and pressure.
Here again, the probe was installed in combination with *Moni::tool*.

The city of Burgos (Spain) is one of the four demonstration sites of the "SmartWater4Europe" European project (SW4EU) project, in which the Author of this Thesis took part (as will be shown in the next chapters). The site is properly focused on the detection of the water quality anomalies, as well as on the integration of this information with the management of the WDS. For potable water monitoring, *Optiqua EventLab probes* have been distributed in the network, capable of detecting changes in any type of dissolved chemicals in real time. In the city of Burgos, the water provider is *Acciona Agua*, who was also responsible for the adoption of the S::CAN *nano::station* in *El Prat of Llobregat*. After seeing in the latter location the efficient and rapid response of this station to changes in salinity, turbidity and fouling, *nano::station* has been recently purchased also for the city of Burgos, together with the *con::cube* terminal.

BactControl was installed at the plant in Aigues de Barcelona, the water utility of Barcelona, where it showed excellent results.

Finally ,Vitens, the largest drinking water utility in the Netherlands, is currently installing a large-scale smart drinking water network in the province of

Friesland. More than 2,200 km of distribution network are currently fitted with 200 sensors that will measure the demand and quality of the drinking water in real time. In detail, eight S::CAN *nano::stations* were installed, measuring turbidity, color, UV254, TOC and DOC, conductivity, pH. Each station was made up of an *i::scan*, a *pH::lyser* and a *condu::lyser*. Data are transferred via 3G to the central Office in Leuwarden.
The Vitens initiative has also been part of the SW4E project.

Finally, it has to be pointed out that since these technologies are new and not yet widely tested, the choice between the different available sensors is challenging. In fact, the parameters that each type of sensor is able to measure, as well as its cost, are very clear but information about their reliability together with the costs of ordinary and extraordinary maintenance are still unknown. Choosing between the various types of sensors, the performance of any associated software packages must also be taken into account, as well as their ability to effectively interface and integrate with the information system serving the WDS to be monitored.

Chapter 3

Design and Management problems of Monitoring Systems

Once the sensors are defined, additional needs arise. Therefore, this chapter aims to discuss some aspects regarding the design and the management of an EWS. In details, concerning the EWS design, the following issues have to be addressed:

- definition of the sensor optimal locations along the WDS;
- acquisition and transmission of the enormous amount of data gathered together from the sensors.

Regarding the EWS management, other problems have to be taken into account:

- identification of the location of the contamination sources;
- assessment of the response acts consequent to a contamination event;
- restoration of the distribution network after a contamination event.

All of these issues are described below.

3.1 Sensors Placement

Considering the different types of sensors and the data they collect, the problem of determining the sensor optimal locations arises for the EWS in order to be efficient.

In fact, the contaminant, which can be accidentally or intentionally injected at any point of the distribution network, is predominantly propagated in relation to the hydraulic conditions (generated from the water demand). If a lot of sensors are installed in the network, the contamination short-time detection likelihood is high; on the contrary, if the sensors are few, or they are located at irrelevant points of the network, the contaminant could be detected after a long time from the injection time, or it may even not be detected if the flow that conveys the contaminant does not pass in the monitored spots. Therefore, the effectiveness of the EWS system depends on the number, as well as on the location of the sensors and their definition becomes a crucial aspect for the design of an EWS.

Thus, the current section will address this issue: for a pre-determined number of sensors[2], necessarily limited for cost reasons, the best localization is the one that maximizes its effectiveness, that is the ability of the system to reduce the impact of contamination accidents on public health.

The general criterion which reflects the minimum impact on public safety must be translated into objective functions that can be expressed quantitatively so that they can be introduced in the optimization models (Hart & Murray, 2010). For example, a possible objective function is the probability (to be maximized) that the contaminated flow passes through a monitored node in the network (detection likelihood). Alternatively, the objective functions can be expressed by other variables (to be minimized), such as the elapsed time between the contamination and its detection, the number of inhabitants that is reached by the contaminant, the number of people that receives a contaminant concentration higher than a certain threshold, the amount of the provided contaminated water, and the percentage of the non-detected contamination events.

Hart and Murray (2010) identified seven steps common to most of the optimization-based sensor placement strategies, including: (i) the definition of contamination risk to minimize consequences (e.g., public health consequences) (ii) the description of the sensor characteristics used in the warning system (iii) the selection of the objective performance (iv) the definition of the optimization objective (v) the formulation of the optimization model (vi) the application of an appropriate optimization strategy, and (vii) the implementation of the design. Hart and Murray (2010) also analyzed the literature state of the art, grouping the papers according to how the authors addressed each step.

In fact, the issue of the optimal location of sensors has been investigated for long time and it has been faced through both single-objective (Lee and Deininger, 1992; Kumar et al., 1997; Kessler et al., 1998; Woo et al., 2001; Al-Zahrani and

[2] Even though those located in a distribution network are monitoring stations, hereafter they will be simply called "sensors" for more brevity.

Moied, 2001; Ostfeld and Salomons, 2004, 2005; Berry et al., 2006, 2009; Propato, 2006; Shastri and Diwekar, 2006; Cheifetz et al., 2015) and multi-objective (McKenna et al., 2006; Ostfeld et al., 2008; Ostfeld and Salomons, 2006; Preis and Ostfeld, 2006; Dorini et al., 2006; Eliades and Polycarpou, 2006; Gueli, 2006; Huang et al., 2006; Wu and Walski, 2006) methodologies. Among the single-objective methodologies, Kessler et al. (1998) introduced a single-objective algorithm aimed at finding the best combination of sensors capable of providing a given level of service through a set covering algorithm. In the approach proposed by Kessler et al. (1998), the term "level of service" indicated the maximum volume of polluted water exposed at a concentration higher than a minimum hazard level and consumed before detecting the contamination. Ostfeld and Salomons (2004) used a similar approach, solving the optimization problem through a Genetic Algorithm while Ostfeld and Salomon (2005) extended their previous work by introducing uncertainties to the demands and the injected contamination events. Berry et al. (2006) introduced a mixed-integer programming (MIP) for sensor placement. Propato (2006) formulated a linear mixed-integer programming model to identify optimal sensor locations for early warning against accidental and intentional contaminations, considering few design objectives. Shastri and Diwekar (2006) introduced the study of uncertainties related to contamination location and demand at the time of the intrusion. Since changing water demand can cause changes in flow directions, contaminated nodes may also change; consequently, a change in demand by 25% was introduced. Cheifetz et al. (2015) proposed a greedy incremental sensor-placement approach to be used for sensor optimization in large real-world water system.

At the same time, Ostfeld et al. (2008) pointed out the importance of creating multi-objective algorithms to refine the problem of sensors optimal location; in this context several algorithms have been developed (Dorini et al., 2006; Eliades and Polycarpou, 2006; Gueli, 2006; Huang et al., 2006). In particular, Ostfeld and Salomons (2006) and Preis and Ostfeld (2006) used the multi-objective genetic algorithm NSGA-II (Deb et al., 2002). In this context, McKenna et al. (2006) investigated the perfect sensor assumption that shows an ability to indicate a positive contamination event as soon as any amount of contamination reaches the sensor. Indeed, they evaluated the impact of sensor detection limits and proved that the detection of events is dependent on the detection limit. The results of their research showed that a sensor detection limit of 0.01 times the average source concentration is adequate for maximum protection.

During the Battle of the Water Sensor Networks (BWSN) held in the United States, many other approaches were evaluated in order to face the problem of the sensor optimal location: a summary of the results can be found in Ostfeld et al. (2008). Later, Berry et al. (2009) addressed the problem of the imperfect sensors, taking into account the effect of false sensor readings: the inclusion of false negative/positive led to a non-linear formulation of the optimization problem and the results showed that the false sensor readings can have a significant impact on network safety. Nowadays, a wide variety of sensors is commercially available. Since the technology is advancing, the sensors are able to measure simultaneously an increasing amount of physical-chemical water parameters, considered to be crucial for detection of contamination events (e.g., US-EPA, 2012; Perelman et al., 2012; Arad et al., 2013; Oliker and Ostfeld, 2014a; 2014b). In the context of event detection, Perelman et al. (2012) utilized the artificial neural networks for studying the interplay between multivariate water quality parameters and detecting possible outliers: the results consist of alarms indicating a possible contamination event based on single and multiple water quality parameters. Arald et al. (2013) aimed to detect events by exploring the time series behavior of routine hydraulic and water quality measurements, developing a dynamic threshold scheme. Oliker and Ostfeld (2014a, 2014b) improved the contamination events detection ability including the support vector machines for the detection of outliers and a multivariate analysis for the examination of the relationships between water quality parameters and their mutual patterns.
However, despite the large research carried out in the field, a challenge is still unsolved: the potential contamination events in a real WDS with complex network topology are countless, since each of them is characterized by a different injection location, duration, mass rate and starting time. The large number of contamination events to be taken into account makes the problem of the optimal location of sensors intractable in practice. Hence, the necessity to set up a sampling method able to select the most representative events which can be considered in order to make the problem less burdensome to solve, arises. In this context Preis and Ostfeld (2008a) developed a heuristic procedure for sampling a set of contamination events: they reduced the initial contamination matrix size by using a statistical approach which selects representative events considering their geographical x/y coordinates, few specific injection mass rates, injection starting times and injection durations. Weickgenannt et al. (2010) introduced an importance-based sampling method to effectively classify the contamination

events based on their importance in terms of the total volume of water that is polluted in a given interval of time. Due to the complexity of WDSs the prediction of the performance under various conditions such as failure scenarios, detection of contamination intrusion sources and of sensor placement locations is difficult. Thus, Perelman and Ostfeld (2011) developed a graph theory connectivity based algorithm to simplify the system behavior. They introduced the notion of clustering in the context of topological/connectivity analysis and they suggested connectivity analysis for topological clustering of nodes, facilitating the nodes sampling for sensor optical locations. Chang et al. (2012) established a rule-based expert system where the two rules, accessibility and complexity, converge to a set of nodes for the final sensor locations based on four design objectives, including the expected time of detection, the expected population affected prior to detection, the expected consumption of contaminant water prior to detection, and the detection likelihood. Diao and Rauch (2013) presented a controllability analysis of the network as preprocessing method for sensor placement: it determines the nodes that have an outcome indication over a maximum number of downstream nodes. Rathi and Gupta (2016) formulated a simplified method that simultaneously maximizes two performance objectives, the demand coverage and the time-constrained detection likelihood, which were combined into a single objective by using weights; they also used Genetic Algorithm to obtain the final optimum sensor locations. Zhao et al. (2016) proposed a sensor placement algorithm based on greedy heuristics and convex relaxation and demonstrated significant performance by applying it to repeated sampling of random subsets of events.

The current research aims to face the optimal sensor-location problem as a bi-objective optimization problem where the number of sensors and the contaminated population are both minimized. Because the solution of the optimization problem requires definition of a set of possible contamination events, a sampling method was developed in order to select a reduced but representative set of events, making the problem solution computationally feasible.

Unlike the approaches described earlier, the proposed methodology no longer selects the representative events considering their geographical x/y coordinates, their importance in terms of location etc., but rather it takes into account practical information on network topology, together with the hydraulic characteristics of the network (as illustrated in the next chapters).

Ultimately, a software for locating the sensors along the water distribution network is available on the EPA website. In fact, the work of Uber et al. (2004) led to the development of the Threat Ensemble Vulnerability Assessment (TEVA) Research Program, which resulted in the most significant example of software in this filed: the TEVA-Sensor Placement Optimization Tool (SPOT) (U.S. EPA, 2013a; Morley et al., 2007; Murray et al., 2004). It is a software designed to optimally place a series of sensors, it allows the definition of a scenario of contamination, it simulates the spread of contaminant/contaminants throughout the water network and it analyzes the consequences, displaying the results in the form of charts and tables (Berry et al., 2012).
The software initially requires some input data to define the scenario of contaminations including, the chemical species to be injected, the injection locations, the effects on the population via the lethal dose or other methods that calculate the ingested contaminated volume, and the estimated population. Subsequently, the number of sensors to be located and the method used for their placement have to be specified: the solvers provide two options for the resolution of the optimization problem, that are the "GRASP (heuristic)" and the "Lagrangian"; GRASP is preferred as the Lagrangian requires much greater computational costs. Once all data are entered, the result shows the positioning of the number of sensors in the water distribution network.
The running time is highly influenced by the network size and the defined contamination scenarios but it is always a matter of minutes.
However, TEVA-SPOT software shows several limits listed below:

- In the input file any flow measurement unit is considered in gallons per minute;
- If very small doses are injected in the network, many nodes result with low concentrations that could therefore not be considered as dangerous (this aspect can be exceeded or reduced by specifying a value for the concentration thresholds in mg/L);
- For a proper functioning, TEVA-SPOT needs to be installed along with a compiler (Visual C or similar);
- If an EPANET Multi-Species Extension (EPANET-MSX, an extension of EPANET) file is used, the results cannot be displayed after the simulations;
- It is not easy to use since it involves the introduction of many parameters, often complicated to determine and to be familiar with for the users who are not specialists in the chemical/biological/medical sectors.

For these reasons, the software has not yet obtained a strong response from the scientific community and it is very hard to find (it is made available by EPA upon request).

3.2 Data Acquisition, Communication and Decision Making

As anticipated in the previous chapter, sensors application involves data management. In fact, once acquired, data must be validated, processed and transmitted to central units for their analysis. Since a large amount of data is involved, expert and automated systems are required in order to save time, increasing the analysis accuracy.

In this context, various data acquisition and transmission systems are known in literature.

Regarding data collection systems, the Supervisory Control and Data Acquisition (SCADA) systems are widely used for environmental monitoring. The SCADA systems are a computer-controlled type of Industrial Control System (ICS) that monitor and control physical industrial processes. SCADA systems historically distinguish themselves from the other ICS systems by being integrated into large-scale processes that can include multiple sites and large distances (Janke et al., 2014). These processes embrace industrial, infrastructure, and facility-based processes.

Since a SCADA system can often incorporate data from online or remote sensors in a cost effective manner (Mays, 2004), it has gained popularity for a long time among the largest water utilities for the control of the WDSs.

According to Panguluri et al. (2004), a water utility SCADA system usually consists of (i) a Human–Machine Interface (HMI) through which the human operator monitors and controls the process (ii) a supervisory computer system, gathering data on the process and sending commands control to the process (iii) Remote Terminal Units (RTUs) connecting to sensors in the process, and sending digital data to the supervisory system (iv) Programmable Logic Controllers (PLCs) (v) various process and analytical instrumentation.

Data acquisition begins at the RTU or PLC level, which includes meter readings and equipment status reports that are communicated to SCADA systems as required. Data are then compiled and formatted in such a way that a control room operator using the HMI can make supervisory decisions to adjust or override normal RTU or PLC controls. A HMI presents process data to a human operator: they are usually linked to the SCADA system's databases and software programs to provide trending, diagnostic data, and management information, such as scheduled maintenance procedures, logistics information, and detailed schematics for a particular sensor or machine. An important part of most SCADA implementations is alarm processing (i.e. determining when alarms should be activated if certain alarm conditions are satisfied). Once an alarm event has been detected, one or more actions are taken, such as the generation of e-mail or text messages to inform management or remote SCADA operators. SCADA systems were born as independent systems with no connectivity to other systems and they were later connected through a Local Area Network (LAN) to share information in real time. They were finally linked with Internet, becoming vulnerable to remote attack.

Once data are collected, the common approach initially used to manage them consisted in creating mathematical models to fit them with the available data but due to the nature of environmental phenomena (noise, non linearity, non stationary, missing data), the data often did not fulfill the hypothesis of these mathematical models.

Thus, a more recent approach consists in relying on the data to build predictive models[3], following firstly the verification and validation process.

Regarding these processes, in the last decade Carlson et al. (2004) proposed a comparison between data received from monitoring sites with data stored at the sensor locations to ensure accuracy and completeness. Following methods provide automated data filtering, such as the moving window averaging, which reduces random noise retaining a quick step response or the Gaussian, Blackman, and multiple-pass moving average that has demonstrated slightly better performance in the frequency domain at the expense of increased computation time. Other procedures are simple outlier detection like the ones used to find deviation from the regular condition of the network in terms of the most common physical/chemical water parameters.

The statistical learning theory (SLT) principle is a more modern theory, which combines knowledge of Artificial Intelligence (AI) learning theory, statistics, geo-statistics and time series analysis to provide tools for the analysis of these databases, called "environmental data mining" (European Project, 2002). Its

[3] Predictive modeling uses statistics to predict any type of unknown event or to guess the probability of an outcome given a set amount of input data (contamination event in this case).

result shows a new modeling of long and short term time events, identification and monitoring of extreme events applying Support Vector Machines for mono-class and multiclass problems, Support Vector Regression, Artificial Neural Networks, multi-scale Kernel Approach, and stochastic simulation in order to solve real environmental problems (case studies in water contamination analysis, air quality forecasting and risk assessment).

Following the data verification process, data should go through the transformation and analysis process (AWWARF, 2002). Data analysis is performed by specialized software and can take the form of univariate/multivariate analysis, Rule-Based Systems (RBS), or Case-Based Systems (CBSs)(US-EPA, 2005a). Univariate analysis considers one single variable, like for example a specific parameter or an instrument response that changes as a function of the water quality. On the contrary, multivariate analysis simultaneously uses inputs from all water parameters/instruments to detect data anomalies, minimizing false alarms. RBSs attempt to interpret information from a starting set of data and rules; they are usually characterized by IF-THEN rules, which provide real-time reasoning by looping through rules. CBSs operate by comparing a collection of current measurements to a database of historical measurements. Any deviations of the current state from past data are notified to the operator, who can run a predictive model to evaluate anomalous scenarios (Carlson et al., 2004).

Processed data should be transmitted to the central database through either hardwired or wireless systems; the latter one can use a variety of methods, including microwave, basic telephone modems, cellular telephone modems, or satellite. Wireless transmission may require a direct line of sight between the transmitter and the receiver, or the use of re-transmitters, also known as repeaters and amplifiers (AWWA Workshop, 2004; US-EPA, 2005a).

Since both acquisition and transmission systems usually require the use of Internet, the security plays a vital role: the existing monitoring system would need to be evaluated not only for its vulnerability to direct physical attacks but also to cyber attacks (e.g., tapping). Transmission of unencrypted data is another security risk thus, hardware and software should have encryption capabilities.

In this context, several researches have been led in order to study potential Internet disruptions and to develop plans for Internet recovery. For instance, the U.S. Government Accountability Office (GAO) was asked to identify examples of major Internet disruptions, together with the evaluation of laws and regulations for facilitating the recovery (Janke et al., 2014; U.S. GAO, 2006).

GAO found that a major disruption to the Internet could be caused by a cyber incident, such as a software malfunction or a malicious virus, a physical incident as a natural disaster or an attack that affects key facilities, and a combination of both cyber and physical incidents.
The security of SCADA systems were also investigated as they are seen as potentially vulnerable to cyber attacks. The two main threats are the unauthorized access to the control software, whether it is a human access or changes induced intentionally or unintentionally by virus infections along with other software threats residing in the control host machine, and the packet access to the network hosting SCADA devices with one's possibility to control or interrupt critical facility operations. For these reasons, more recently SCADA systems incorporate analog signals which require special drivers to accept data from monitors (e.g., particle counters) with digital signals.

Concluding, a data management plan should be implemented and deepened during the EWS design. Each data feature, including source, destination, collection, transmission and storage methods should be taken into consideration in detail to specifically illustrate how data flow through the system.
For the only purpose of illustration, an example of a generic data management plan is presented in Figure 3-1.

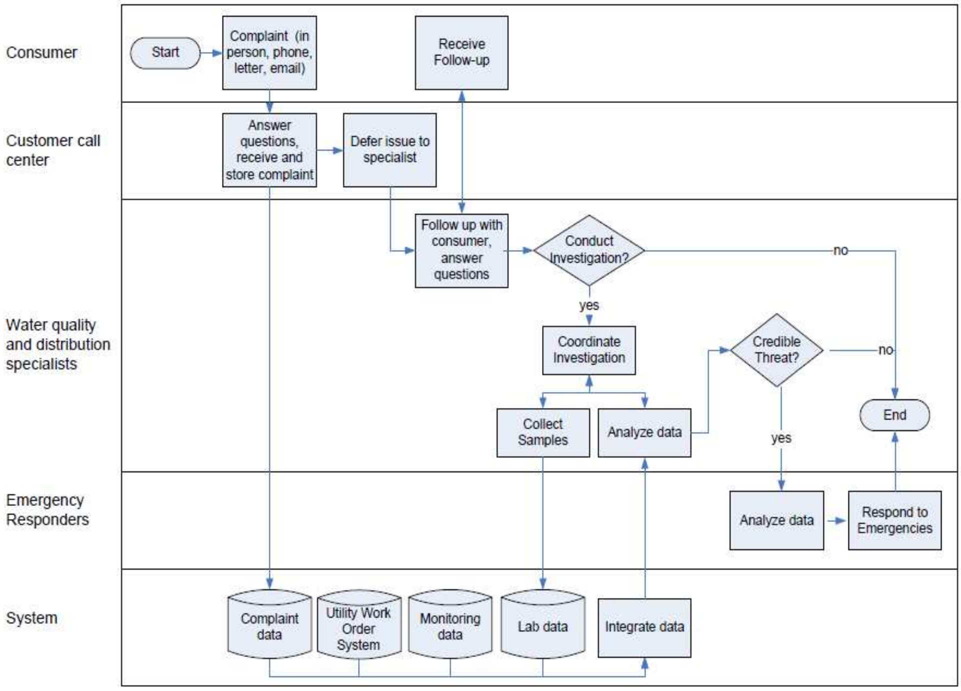


Figure 3-1. Utility Consumer Complaint Data Flow (US-EPA, 2005b)

Figure 3-1 shows the components of a data collection, transmission and integration system for consumer complaint surveillance: data types and formats should be listed, as well as data privacy, sensitivity, security, authorization, encryption, timeliness, cost, redundancy, and availability should be developed.

Finally, the last step of the communication pathway concerns the communication between data analysts and decision-makers during event detection. There are many possible communication mechanisms which may be employed, such as land-line, pager, cell phone, satellite communications, radio, television, internet, and emergency numbers. The type of communication mechanism is dependent on the information provider, source, recipient, content, format, timeliness, and other requirements. Anyway, the most common procedure would utilize voice calls, with supplementary data transmitted electronically. Communication during consequence management also shows many forms, depending on the message and target audience. Emergency broadcast warnings to the public can use well-established communication mechanisms, such as radio and television and/or internet websites (US-EPA, 2005b).

3.3 Contaminant Source Identification

As a result of what has been proven so far, water monitoring sensors are located only at some nodes of the network.
Once the EWS is realized, a model capable of revealing the possible contamination injection characteristics (e.g., injection sources, time, concentration, etc.) needs to be provided to face the occurrence of a contamination event.
In order to do that, the useful observations about water quality coming from sensors could be exploited. In fact, given a set of concentration observations at sensors in the network, an inverse problem can be constructed to identify the contaminant source characteristics (including location, strength, and release history).
Many research works have been already conducted to analyze the formulation of this inverse problem.
Starting with the review of the European projects in this field, around the end of the first decade of 2000, different strategies were implemented by the European Project SecurEau to identify the location of the contamination sources (www.secureau.eu). The considered approaches were: (i) a method based on the analysis of flow data (ii) a deterministic method based on successive positive readings of sensors (iii) methods based on Artificial Neural Networks (ANNs) for single and multiple contamination events, and (iv) stochastic methods, such as least-squares solving with Tikhonov regularization or minimum relative entropy solution (MRE).
The method based on the analysis of the nodes where contamination has been detected and on flow directions was tested in different cases. The results allow concluding that the method gives good results especially for the cases with a single contamination event. The method is fast and does not require any prior testing phases (www.secureau.eu).
The deterministic method based on successive positive readings of sensors concerned the analysis of the residence time of water in pipes and it only required a binary sensor status over time. The results for the localization of contamination sources are given sequentially, being updated each time a new sensor detected a change in contaminant concentration. In some situations this method enables the occurrence of false negatives and false positives thus, the ANNs algorithms were investigated: they identify the correct contamination source and predict the correct contamination time associated with each possible

contamination source, even in the case of large and complex WDSs. The method that extended the application of ANNs to multiple contamination scenarios achieved very satisfactory results for real WDSs. The method is generally able to correctly determine the simulated source and to define a very restricted set of possible contamination sources, even considering hydraulic scenarios with demand uncertainties. However, the estimation of the contamination time for scenarios characterized by demand uncertainties shows larger deviation compared to the simulated contamination sources. The time of required computation is generally very low, which has made this method very suitable for application in real contamination scenarios.

Within the SecurEau Project, the French research Institute Irstea, a member of the SecurEau Consortium, developed a two-step enumeration/exploration method (which is an inverse problem method) for the source identification base. Firstly, the input/output transport matrix was worked out with a backtracking method and then, minimum relative entropy method, without any assumption for the Probability Density Function (PDF) distribution, or the least squares method with Tikhonov regularization were used to refine the results and be a source as a confidence interval. The backtracking algorithm yielded good results giving very quickly the full list of potential node sources of contamination at the different times, and the in/out (transport) matrix returned the relation between the potential source and the detecting sensors. That matrix could then be used either on a minimum relative entropy method or a Tikhonov method: the real contaminant source is always determined as potential source, even though the minimum relative entropy method seems discriminating more the potential nodes than the Tikhonov method.

Regarding the last two approaches, numerous studies were already present in literature from the second half of the 90's. In fact, Islam et al. (1997) showed an inverse model for directly calculating the chlorine concentrations needed at the system sources in unsteady flow conditions for meeting a specified concentration value at a particular node in the network. The model used a one-dimensional chlorine transport equation which was discretized by using a four-point implicit finite difference scheme. The main weakness of the model is that it is suitable only for an even-determined case in which the number of unknowns (i.e., nodal and source concentrations) equals the number of equations (i.e., the one-dimensional transport equations along pipes and the mass balances equations at junctions). Al-Omari and Chandhry (2001) extended Islam’s model to the underdetermined case where the number of unknowns is greater than the number of available equations and there are more than one solution for which the prediction error was zero. Supplementary *a priori* information was added to the problem: a unique solution was generated by minimizing the Euclidian length of the solution vector subject to the equations that describe chlorine transport in the network. The input/output model based on the backtracking algorithm, which has been investigated by SecurEau project, was presented by Shang et al. (2002). As mentioned above, the model provided information about the relationships between water quality at input and output locations by tracking water parcels, and moving them simultaneously along their paths. The algorithm was geared mainly toward feedback control, providing information on all paths between pairs of nodes, and less toward contamination source identification using monitoring stations information.

Later, Bart G. van Bloemen Waanders et al. (2003) and Laird et al. (2005, 2006) introduced a large scale nonlinear programming approach that used real-time concentration information from an installed sensor grid to accurately determine the time and location of the contamination event. This approach introduced unknown, time dependent injection terms at every node in the network and formulated a quadratic program to solve for the time profiles of the injections. Van Bloemen Waanders et al. (2003) used a nonlinear least-squares minimization of the errors between the calculated and measured node concentrations at the sensor nodes with a regularization term to force a unique solution. The constraints in the optimization problem were the Partial Differential Equations (PDE) of the water quality model for the network.

Laird et al. (2005) discretized the problem, using an origin tracking algorithm to characterize the pipe time delays and remove the need to discretize along the length of the pipes. The resulting large scale nonlinear program was solved using a nonlinear interior point code and it provided good results in identifying a family of possible injection scenarios.

The following year, Laird et al. (2006) formulated the inverse problem of identifying the time and the unique injection scenarios, using concentration information from a sparse sensor grid and by means of a Mixed Integer Quadratic Program (MIQP). This formulation included constraints that limited the solution space and allowed the distinction between single and multiple injection locations.

Preis and Ostfeld (2006) solved the same problem through a hybrid Model Trees (MT), together with a Linear Programming (LP) scheme. The MT replaced EPANET through learning (i.e., training and cross validation), simulating the network response to different random contamination events, while the LP used the classification structures of the model trees linear rules to solve the inverse problem. In this context, the MT represented forward modeling and the LP on the linear tree structure allowed backward inverse modeling, being the contamination injections characteristics the unknown problem. In the same year, Zechman et al. (2006) recognized that the accuracy of the source characterization problem depends on the degree of non-uniqueness present in the study, since it may cause misidentification of the source characteristics. In fact, as more sensors are added to the network, the non-uniqueness is reduced and a unique solution may be identified.

Thus, a systematic search for a set of alternatives that are maximally different in solution characteristics can be used to address and quantify non-uniqueness. For this reason, Zechman et al. (2006) investigated the use of Evolutionary Algorithm (EA)-based alternatives generation procedures to quantify and address non-uniqueness presented in a contaminant source identification problem for a water distribution network.

At the same time, Di Cristo and Leopardi (2006) used time-varying concentration measurements to identify the source location of an accidental contamination. In particular, as nodal demand uncertainty in input data and errors in concentration measurements determine a high level of uncertainty in the analyzed inverse problem, an analysis based on a Monte Carlo procedure was performed. The results showed a good identification frequency of the right pollution source node also at high uncertainty levels; however, the results depended on the number and the location of water quality measurements. The maximum coverage criterion appeared as a good method for selecting measurement location.

The EPANET water distribution system simulator (Rossman, 2000) has been also exploited to solve a nonlinear contaminant source as it provides a convenient platform for implementing the approach of Guan and Aral (1999) and Aral et al. (2001) in a WDS. In fact, Guan et al. (2006) coupled EPANET with an optimization code, solving the contaminant-source identification and release-history problem. In details, EPANET was firstly used to simulate concentrations at *a priori* selected monitoring locations with release histories of potential contaminant sources. Then, the optimization model was used as a predictor-corrector algorithm to identify the sources and their release histories based on similarity of responses between simulation results and measured data at

the selected monitoring locations. This information exchange was developed as a closed-loop system that yielded a rapidly converging algorithm.
More recently, Liu et al. (2011) proposed an adaptive dynamic optimization technique (ADOPT) for providing a real-time response and they investigated a new multiple population–based search that used an Evolutionary Algorithm (EA); the multiple populations were designed to maintain a set of alternative solutions that represented various non-unique solutions. Facing the non-uniqueness in the solution, the procedure was coupled with a systematic method to identify a set of alternative solutions that were as different as possible in the solution space. Thus, at any stage of the solution procedure, possible solutions that best describe the observations were determined and were used as starting solutions for subsequent searches as more information became available.

In addition, several approaches have been used to develop software tools for the simulation of contamination transport (www.secureau.eu). The following approaches have been considered: (i) an off-line software tool based on equations governing bacterial re-growth that is affected by sorption, desorption, chlorine and substrate concentration (ii) a software tool considering sorption developed using MATLAB, Visual Basic for Applications (VBA) and EPANET with models for the evaluation of contaminant concentrations, and (iii) an on-line software tool that uses flow direction data for tracking contamination spread.
The off-line software tool is supported with a model developed through EPANET-MSX (refer to Shang et al., 2008). The model contains differential equations defining functions of attached bacteria, bulk bacteria, substrate and chlorine concentrations as a function of time and also considers the phenomena of pathogen adsorption/desorption. The model parameters are user-adjustable as various types of contaminants have different adsorption and desorption coefficients. The graphical user interface of the EPANET software has been added to the model, allowing the operator to modify model parameters, to set initial conditions (e. g. contamination sources), to view results in graphical or tabular form, as well as to visualize the distribution of contamination over the network. The advantage of the off-line software tool is that it contains a comprehensive model that besides convection takes into account adsorption/desorption and re-growth of bacteria, as well as chlorine and substrate concentration. The model can be used to run simulations and study

contamination development, long-term effects including sorption, effects of chlorine disinfectant addition and substrate concentration.
The software tool developed on Matlab and VBA enables testing the effect of sorption phenomena on contamination spread in drinking water distribution systems and studying long-term behavior of a partially adsorbed contaminant in a drinking water distribution system. It has been shown that the proposed method is suitable for the study of the effects of the sorption phenomena in the modeling of the transport of contaminants in real drinking WDSs.
The on-line software tool enables running simulations of contamination transport in a water distribution network based on flow direction data. The concept of the on-line software tool is based on the idea that in case of contamination accident the affected area of the network is mostly determined by flow directions rather than flow magnitudes. Flow direction data can be obtained by means of flow direction sensors or by hydraulic simulation. A combined approach (flow direction sensors installed in some pipes, simulation for other pipes) is also possible. The advantage of the method is that if flow direction sensors are used, the software tool uses real-time data from the network and therefore, it provides more robust simulation results.

3.4 Response after Contamination

An important management problem concerns the intervention phase that follows the occurrence of the contamination event.
For this reason, the existing technical regulations (e.g., the ISO 11830 "Guide on crisis management process") states that drinking water supply should manage the response and the restoration of WDSs after contamination events to ensure secure and hygienically proper drinking water to the customers.
This section will address the response problem, while the next one will deal with the recovery, following a contamination event.
The response can be implemented in different ways such as a simple alerting of the population or the injection of substances into the system that can neutralize the effects of the contaminant. Alternatively, the devices that control the flow and the functioning of the system (e.g., isolation valves and hydrants) can be managed in order to limit the diffusion of the contaminant with its consumption by users. This represents an alternative to the more conservative approach that

imposes the shutdown of the entire water supply, considered not feasible if the WDS is not divided into District Metered Areas (DMAs) or for many customers with special needs (hospitals, clinics, factories, etc.).
However, the teams that water utility managers usually have available to intervene following a contamination event are not so many. Thus, there is a need to determine the procedure and the order that must be executed to intervene, deciding, for example, which valves should be closed or which hydrants should be opened.
This issue leads to the formulation of different optimization problems, giving birth to several studies in literature.
The first, Baranowski and LeBoeuf (2006) proposed three different optimization techniques (an unconstrained and a constrained first-order reliability method, as well as, a parameter estimation method) for determining the optimal nodal demand to reduce the contaminant concentration within the network after the detection. In the same year, Poulin et al. (2006) aimed to minimize the risk that contaminated water is consumed, to identify the valves to be closed safely containing the contaminated water as well as proceeding with the isolation actions, and to define a set of operations to efficiently flush contaminated water from the network for the quick returning to the normal operation conditions. Therefore, they proposed a heuristic algorithm based on simple rules, capable of marking and isolating the contaminated zones through the simultaneous closure of a certain number of valves in the system with the assumption of an unlimited number of response teams. Since the response to a contamination event intrinsically involves conflicting objectives (e.g., isolation of some network areas versus operation costs or citizens' need), Pries and Ostfeld (2008a) introduced a multi-objective procedure to develop an optimal response, minimizing the contamination mass consumed after the first sensor detection and the total number of operations (i.e. valves closure and hydrants opening) required for the isolation and the flushing of the contamination from the network. The study solved the optimization problem through the Non-Dominated Sorted Genetic Algorithm-II (NSGA-II); it also assumed that the number of response teams is unlimited, all the operations take place simultaneously, and the characteristics of the contamination event (i.e. location, time, duration etc.) are known *a priori*. Similarly, Guidorzi et al. (2009) proposed a procedure based on two consecutive optimization process: the first one defines the position of a given number of sensors, minimizing the percentage of undetected contamination events and the volume of contaminated

water consumed up to the beginning of the response operations; the second one identifies the hydrant-opening, together with the valve-closing operations to be carried out for a generic configuration of sensors, resulting from the first optimization procedure. Alfonso et al. (2010) similarly proposed a multi-objective procedure for a preliminary identification of the operations to be activated in order to minimize their number and the contaminated volume consumed after the detection. While, once the operations are identified, Guidorzi et al. (2009) also suggested an *a posteriori* analysis to determine the sequence according to which the operations should be activated based on the number of response teams actually available, Alfonso et al. (2010) did not take into account the problem of the best operation time. Both of the last two cited works assumed that the characteristics of the contamination event are not known a *priori*. Finally, Alvisi et al. (2012) recalled the study of Guidorzi et al. (2009), locating the sensors in the network, activating them when one of these triggers an alarm and developing a procedure which enables the automatic identification of the optimal scheduling of a set of devices (hydrants to be opened and valves to be closed) in order to minimize the contaminated volumes consumed by users after a contamination detection (the source is assumed unknown). In this study, the constraints were represented by the number of available response teams and the maximum speed at which these teams could travel along the roadway; the optimization process was based on a genetic algorithm (GA) which interacted with a Mixed Integer Linear Programming (MILP) solver and which is coupled with an hydraulic/quality simulator to calculate the contaminated water consumed.

The effectiveness of any response strategy largely depends on the length of time needed to implement the required actions since response means intervening and ending any consequences on public health. Numerous researchers have also investigated the influence of response time on the magnitude of public health consequences (Janke et al., 2006; Skadsen et al., 2008; Murray et al., 2008), showing that an increase from 12 to 48 hours in the response delay can reduce the effectiveness of a warning system by 50%, or more (up to 70%).

Bristow and Brumbelow (2006) finally analyzed the temporal and procedural space between the detection of an anomaly in the water quality and the response decisions, including the process by which decision-makers confirm contamination and activate the initial phases of an emergency response plan. The results showed that the cumulative time required to detect the contamination event, perform emergency response, and address the compliance process can

take a considerable amount of time, generally on the order of magnitude of days, where the contaminant verification and the response transmission are the most significant sources of delay (Bristow and Brumbelow, 2006).
Consequently, an effective and preventive monitoring system is essential in a WDS as it provides crucial information for timely intervention by limiting damage to citizens.

3.5 Recovery after Contamination

Once a contamination event has been proved, it appears necessary to promptly intervene in order to limit the negative effects of the contamination itself on public health. While the response involves the isolation of the contaminated areas and the development of strategies to reduce the consumed contaminated water, the recovery phase imposes pipe wall cleaning and decontamination.
This section will discuss the latter issue, that is, the time period in which all the actions necessary to restore the network back to the regular operating conditions are implemented.
In particular, the contaminant does not have to be zero but only be below a certain limit (detection limit, acceptable level, etc.). Hence, the cleaning procedures must lead the WDS to an acceptable low concentration of the contaminant in the water and in the deposits. Only if this criterion is met, the clearance can be given to distribute drinking water to the customers and to return to the routine operation.
The recovery problem has been investigated in several European projects, as for instance in the SecurEau Project, already mentioned. In details, it proposed a strategy for pipe wall cleaning and decontamination to be carried out *in situ* (i.e. inside the pipe), flushing neutralized contaminants out of the system.
If cleaning is very intensive, pipes could be damaged and contamination leaked out to groundwater thus, SecurEau avoided to apply aggressive methods. However, flushing should have been more effective than traditional methods, which were mostly focusing on removing loose deposits because corrosion layer inside the pipe absorbed some of contaminants and the only way to remove it was to flush it out with incrustation layer.
Several methods for cleaning drinking water distributions systems were taken into account; after that, simple solution "how to deal with adsorbed Chemical,

Biological, Radiological and Nuclear (CBRN) agents" were proposed (www.secureau.eu).
The following steps should be always followed to ensure efficient cleaning:

- the selection of the most suitable decontamination method as a function of the contaminant location;
- when the surface is associated with loose deposits and/or bio-films, traditional techniques (e.g., water/air flushing) let this surface layer to be removed;
- the selection of deeper methodologies for the deposits when an effective diffusion of the CBRN agents is presented.

Both spores, non-spore forming bacteria, and viruses were used by SecurEau as models for testing decontamination procedures, which are summarized below. Shock-chlorination was studied by adding high concentration of chlorine and keeping it to reach the optimal Concentration multiplied by Time (CT) value: although reasonable efficacy was observed in water, bio-films were not effectively removed. In fact, results showed that disinfection with shock chlorination was an effective method for neutralizing bacteria but the removal of bio-films was not possible without treatments to activate them. Ultrasound cavitations may be used to detach spores from surfaces towards a further procession (i.e. DNA analysis or quantification) but it should be followed by other disinfection methods, which together has been too expensive to use in case of contamination of WDSs. One of the most promising results on surface disinfection was the regime alternating between the free chlorine (200 mg/L) and the sodium hydroxide (1.5%): this technique was based on the spore disinfection in a bulk and afterwards, the releasing of spores adhered to the surface.
Advanced oxidation process was successfully tested in SecurEau to take advantage of iron and cupper in WDSs and in bio-films, while mercury was used as model substance of inorganic agent, testing several methods (water flushing with chlorinated or non chlorinated water, ice pigging). Chemicals as release agents of radiological agents from pipe material and real pipe deposits were tested: sodium bicarbonate was the most effective chemical for cleaning compared to other decontamination chemicals.
Moreover in SecurEau, several solutions and strategy for effective *in situ* cleaning based on using simple reagents were selected:

- Chemicals such as pesticides as well as pathogens and autochthonous bacteria could be removed using Hydrogen Peroxide solutions;

- Removal of resistant microorganisms, such as Bacillus spores, from pipe surfaces could be achieved by alternating treatment with sodium hydroxide and chlorination;
- For removal of radionuclides, desorption by sodium bicarbonate solution and flushing of the system for safe storage offered a remediate option;
- Ice slugs and gravel in combination with water flushing were effective methods for removing both loose deposits and corrosion layer of pipes.

Chapter 4

Early Detection Systems with the hypothesis of conservative contaminants - Optimal Sensors Placement

As was explained in the previous introductory chapters, water distribution networks can be equipped with water quality monitoring systems to successfully detect potential contamination events. These systems include sensors installed at strategic locations, selected in such a way as to guarantee early warning and reduced impact (Walski et al., 2003).
The issue of the optimal placement of sensors, which is usually dealt with a multi-objective approach, is crucial for network management and protection, and hides various pitfalls. One of them lies in the definition of the significant contamination events. In fact, contamination events can occur at any node of the network and at any time of the day, with whatever values of duration and mass. All this generates a very high number of potential contamination scenarios. However, taking account of all of them may be exceedingly demanding from the computational viewpoint.
Another pitfall lies in the suitable choice of objective functions to be considered. In fact, numerous objective functions were formulated in the scientific literature (e.g., Ostfeld et al., 2008; Presis and Ostfeld, 2008b), including number of installed sensors, as a surrogate for the cost, and event detection time, contaminated population and sensor redundancy, as surrogates for system reliability. Though all these variables could be simultaneously considered in the same optimization framework, optimization techniques lose resolution

effectiveness as the number of objective functions grows (Creaco et al., 2016). Therefore, while considering a total number of two conflicting objectives, associated with the cost and reliability of the monitoring system respectively, the issue of which single pair of objective functions can give the best results in the context of the optimal placement of sensors in WDNs arises.
In the chapter, first details will be given about the quantity and quality simulations that need to be carried out to represent the network behavior during contamination events (section 4.1). Then, a procedure for the sampling of the significant contamination events will be presented in the context of bi-objective optimization, by adopting an explicative pair of objective functions (section 4.2). Unlike other approaches described in the scientific literature (Preis and Ostfeld, 2008b; Weickgenannt et al., 2010; Perelman and Ostfeld, 2011; Chang et al., 2012; Diao and Rauch, 2013; Rathi and Gupta, 2016; Zhao et al., 2016), this procedure has the peculiarity of being based on practical information on network topology and operation (drawn from Tinelli et al., 2017a). The chapter ends with section 4.3, in which an analysis is reported concerning the impact of the objective function selection on the optimal placement of sensors (drawn from Tinelli et al., 2017b; Tinelli et al., under review).

4.1 Quantity and Quality Simulations for Contamination Events

After a set of potential contamination events has been defined, quantity and quality simulations need to be carried out to model network behavior during each of them. These simulations can be carried out through such software as EPANET.
The quantity simulation can be carried out *una tantum* with reference to the typical day of network operation prior to the set of quality simulations. This reflects the common assumption that hydraulics is not affected by contaminant propagation.
Then, thanks to quality simulations, it is possible to identify how the contaminant propagates in the network for each contamination event.
In EPANET, water quality is solved through a system of one-dimensional (1D) advection-reaction pipe equations, and perfect mixing is considered at junction nodes. This approach is considered acceptable in the search for optimal sensor locations in which the contaminant is often considered conservative.

Nevertheless, under specific hydraulic conditions, a better representation of the water-quality processes, above all with regards to mixing at the junctions, could be obtained by taking account of more accurate modeling, such as computational fluid dynamics (CFD) (Braun et al., 2015 offers an example). CFD and other accurate modeling approaches, which are too burdensome to be considered in the sensor design phase, remain valid and useful tools to be adopted for an *a posteriori* analysis of the results.

For each contamination event, EPANET is able to calculate the fate of the contaminants injected into the network, identifying which nodes are reached and when they are reached following the initial injection time.

Starting from these simulations, two matrices can be calculated, providing information about the network behavior during the various contamination events. These two matrices are (i) the contamination matrix (introduced for the first time by Kessler at al., 1998), and (ii) the time matrix. They both have as many rows and columns as the number of nodes and contamination events, respectively. In detail, for each contamination event, the contamination matrix helps distinguishing reached and unaffected nodes (matrix values equal to 1 and 0, respectively). The time matrix gives the time interval for the generic node to be reached, following the initial instant of contamination. For unaffected nodes, this time interval is $+\infty$.

Explicative examples of the two matrices are illustrated in Figure 4-1.

By simple manipulations on these matrices, it is possible to assess the performance of a generic system of sensors installed in the network. In detail, it is possible to assess how many events are detected by the system in a total group of events, and then to calculate the event detection likelihood. Other performance functions that can be potentially evaluated thanks to these matrices include the average time of detection and the sensor redundancy, that is how many sensors are on average able to detect the generic event. By leaning on info on the number of inhabitants connected to the network nodes, the contamination and time matrices also enable quantifying the average contaminated population for the group of contamination events considered.

As the following sub-sections will show, the previous performance indicators can be usefully adopted in the context of multi-objective optimization, for the search of the optimal sensor locations in the network.

Node	**Contamination Events**				
	1	**2**	**...**	**...**	$\mathbf{n_{events}}$
1	0	1	...	...	0
2	1	0	...	...	1
...	...	...	...	...	...
n_{nodes}	0	1	...	...	0

a)

Node	**Contamination Events**				
	1	**2**	**...**	**...**	$\mathbf{n_{events}}$
1	$+\infty$	t_{12}	...	...	$+\infty$
2	t_{21}	$+\infty$	...	...	$t_{2n.events}$
...	...	t_{ij}	...	...	..
n_{nodes}	$+\infty$	$t_{n.node2}$	...	...	$+\infty$

b)

Figure 4-1. Explicative examples of a) contamination matrix b) time matrix

4.2 Sampling Problem

Due to the initial operations reported in section 4.1, the optimization procedures are computationally very burdensome when the number of the considered potential events is high. For these reasons, the contamination scenarios to be taken into account in the calculations must be reduced. Thus, a method for defining a small significant set of contamination events, which is representative for the totality of the events, was set up. Each possible contamination event is characterized by certain values of injection location, starting time, mass rate, and duration. Therefore, the sampling was done for each of the event characteristics, as it is explained in the following sub-paragraphs.

4.2.1 Sampling of contamination events

Injection location

Although all the network nodes could theoretically be potential injection locations, an algorithm was developed using the graph theory to select the representative nodes that should be considered for the injection of contaminants. A preliminary step of the algorithm is the opening of network loops and source interconnection paths, which can be carried out through the minimum spanning tree algorithm (Kruskal, 1956) and/or other procedures accounting for pipe diameters and water discharges. Subsequently, the representative nodes of the system are selected as a function of their gradually increasing distance from the source nodes. In fact, the distance from the source node is a variable with more hydraulic meaningfulness than the nodal geographical x/y coordinates of Preis and Ostfeld (2008b). Specifically, sampling is done with a prefixed frequency, i.e. one out of two, three, four (and so forth) nodes, along the path outgoing from the network source(s). In each path, the closest nodes to the sources were always accounted for. Subsequently, the sampling is modified through the two following steps, which force selection of the dead-ends at the expense of the close nodes:

i. Dead-ends are included in the list of selected nodes;
ii. Nodes adjacent to dead-ends, whether previously sampled, are excluded from the list if they are serial nodes.

Inclusion of dead-ends is important because they are the final nodes of the network where the generic contamination events can be detected. In fact, let the generic water path in the network be considered. The dead end is the only node able to detect all the contamination events that have injections along this path.
The network in Figure 4-2 is provided as an example for the application of the selection with one out of two and one out of three sampling frequency.
The network has 13 nodes with one source node placed at Node 1. The application of the minimum spanning tree algorithm leads to the removal of pipe 13 for loop opening. Subsequently, nodes are selected according to their gradually increasing distance from the source node.

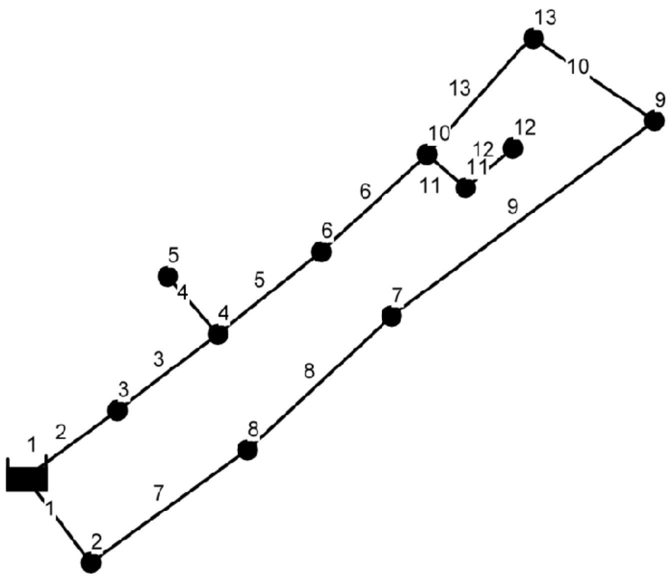


Figure 4-2 Explicative WDS for visualization of the sampling frequency

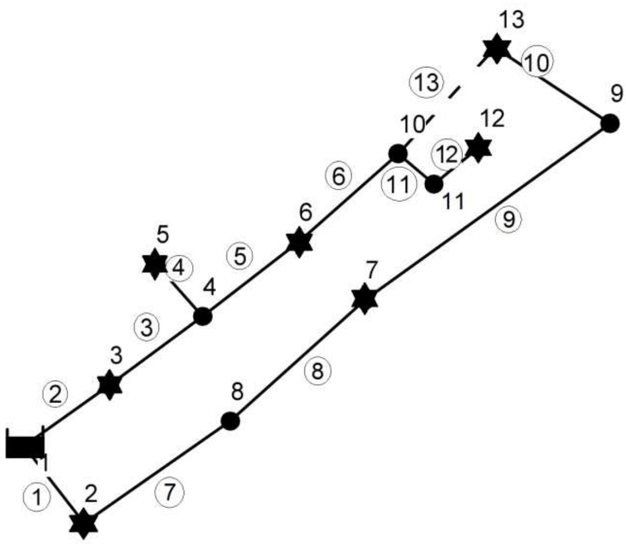


Figure 4-3. Selected nodes considering a sampling frequency for the injection nodes equal to 2 in an explicative WDS. The source node is indicated with a box and the dashed line indicates the pipe that is removed for loop opening. Node numbers close to the nodes. Pipe numbers inside circles and close to the pipes.

As shown in Figure 4-3, the closest nodes to the tank (node 2 and 3) are selected. Therefore, considering a frequency of one out of two nodes, the selected nodes are 2, 3, 5, 6, 7, 12 and 13. In particular, being a dead-end, node 12, which should not have been included as a result of the frequency sampling, is finally included in the list at the expense of node 11, which is a serial node adjacent to node 12. In the same way, it would have to proceed by selecting a frequency of 3, 4, and so on.

A remark must be made about the injection location sampling, which may fail to select important crossroad nodes or supernodes, which, as defined by Deuerlein et al. (2014), belong to several paths. In this context, even if an important node is missed in the sampling, information about this node is not lost; in fact, contamination events will always reach it through paths including other nodes sampled by the algorithm. Furthermore, exclusion from contamination location sampling does not prevent the generic node from being a good sensor location.

Starting time

Taking as benchmark the typical day of WDS operation, every instant could theoretically be the starting time of contamination. This means that considering the whole day sampled with a 0.5-h step, there could be 48 potential starting times. The sampling of the starting times is carried out based on the WDS operation phases, detected as a function of pipe-water discharges, which can vary based on nodal demand, source head patterns, and switching on/off of pumps.

In detail, these phases can be identified by detecting the times when the water discharges in network pipes vary significantly. Assuming that the water discharge in a pipe follows the daily trend shown in Figure 4-4, the instants associated with significant changes in the flow are 0, 7, and 15 h.

Three phases are then detected for the pipe, that is, Phase 1 from 0 to 7 h, Phase 2 from 7 to 15 h, and Phase 3 from 15 to 24 h.

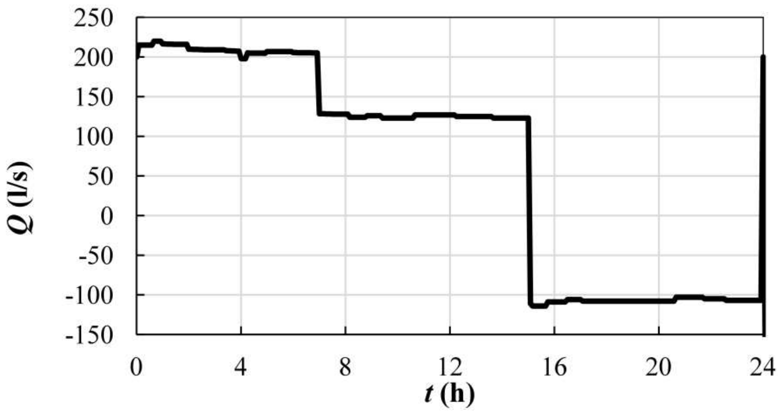


Figure 4-4. Water discharge in a WDS pipe during a typical day

The instants of significant flow variation for the whole network are obtained by putting in a single timeline the instants of significant flow variation in all the network pipes. The generic phase is then detected as the time slot between two successive instants in the timeline. Then, a representative instant can be selected for each phase, i.e., either the initial instant or an inner instant in which the pipe-water discharges are closest to the average values in the phase. These representative instants are selected as significant starting times for the sampled contamination events.
In the assessment of any objective functions, the operating-phase durations can be used as weights to be associated with contamination events.

<u>Mass rate</u>

The contaminant advection-reaction equations are linear if the contaminant is conservative or first-order reactions are used (often the case with the optimal sensor location). The consequences of this aspect are easily shown through the explicative example of the network in Figure 4-5, with one source node (Node 1) and six demanding nodes (Nodes 2–7).

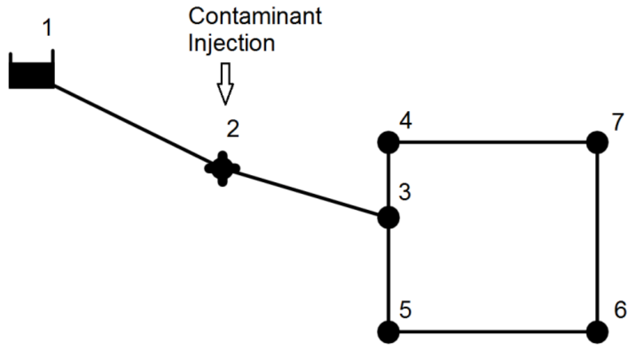


Figure 4-5. Explicative water distribution network for sampling contamination mass rates

In this network, which features a constant demand of 26 L/s, two contamination events are considered with the same injection location (node 2) and duration (4 h) and differing in the mass rate - only for explicative purposes, equal to 50 gr/min (32.05 mg/L) and 200 gr/min (128.21 mg/L) in the two events, respectively. The separate effects of the two events are shown in Figure 4-6.

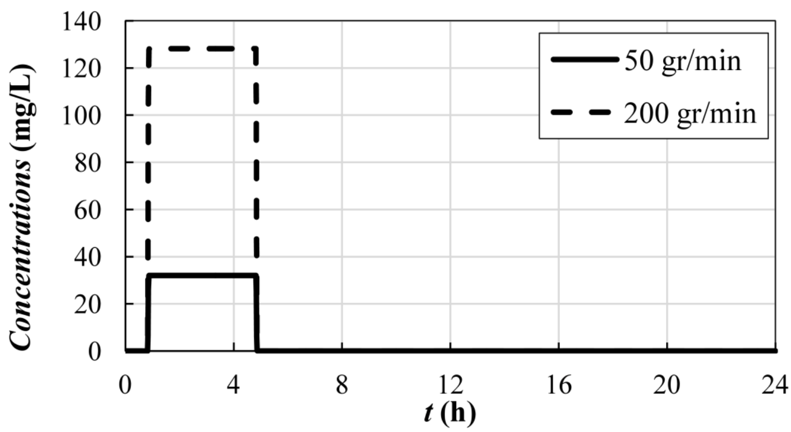


Figure 4-6. Trend of the contaminant concentration at node 7 in response to the injected masses of 50 gr/min and 200 gr/min in the explicative water distribution network shown in Figure 4-5

In detail, this graph reports, for Node 7, the trends of the contaminant concentration in response to the two events.

The results clearly show that the two trends are proportional and one can be obtained from the other through multiplication by a factor equal to the mass rate ratio, which is 4. Considering the two trends in Figure 4-6, where the contaminant concentration rises almost instantaneously to the highest value (very high variation speed), a sensor placed in Node 7 would be able to detect the event in both cases, as long as its sensitivity is small enough (usual assumption in the framework of optimal sensor placement) to detect the contamination. Nevertheless, in real networks, cases in which the contaminant concentration is so low as to be under the sensor sensitivity are less dangerous in terms of network safety.

In light of the linearity of the contaminant advection-reaction equations, the mass of injected contaminants does not influence certain variables, such as number of contaminated nodes or contaminated population, if the pollutant concentrations are high enough to be detected by the sensors. Therefore, if the focus is on the number of contaminated nodes and/or the contaminated population, rather than on the contamination concentration, the average of the possible masses can be sampled as a representative value.

Furthermore, when the contamination concentration is also relevant, the WDS quality simulation can always be carried out only for one contamination mass rate. The results associated with other values can then be derived by taking advantage of the linearity of the equations, as discussed earlier.

<u>Duration</u>

Under conditions of constant (or slightly variable) pipe water discharges (as it occurs in every WDS operation phase), the nodes reached by the generic contamination do not change as a function of the event duration; only the contaminants residence time into a single node can change. This is easily demonstrated likewise the previous sub-section, by separately injecting 200 gr/min for 120, 300 and 600 min in the explicative water distribution network shown in Figure 4-5. The trend of the contaminant concentrations in response to the three injections at Node 7 are illustrated in Figure 4-7.

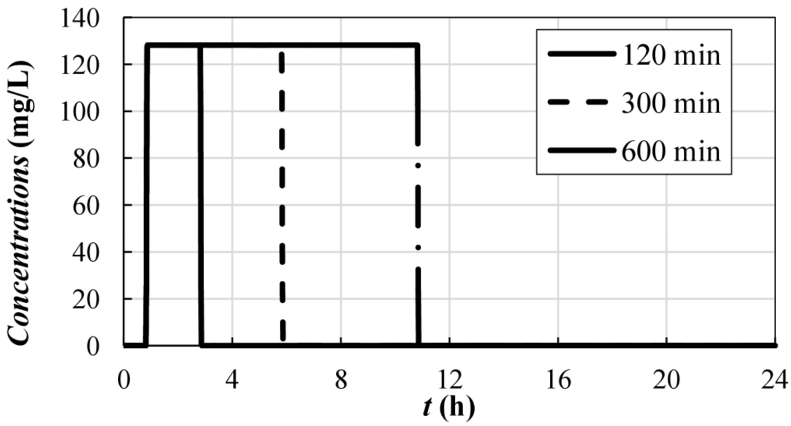


Figure 4-7. Trend of contaminant concentrations at node 7 for the injected durations of 120, 300 and 600 min in the explicative water distribution network shown in Figure 4-5

The results clearly show that the concentration trends are only shifted along time, confirming that the event duration only affects the residence time of the contaminants in the nodes. The average duration can then be sampled from a list of possible contaminant event durations.

Furthermore, the long duration events can be regarded as a succession of short duration events. Therefore, a single short duration, shorter than the network operating phase durations, can be sampled. It must be underlined that an event lying astride two consecutive operation phases can always be decomposed into the combination of two events, each of which is fully lying inside a single operation phase. Either composing element is banally decomposable into a series of events equal to the short duration event used for the sampling.

4.2.2 Optimal Sensor Location

It is evident that application of the aforementioned sampling method enables a significant reduction in the number of contamination events to be considered, and, therefore, in the size of the contamination and time matrices.

The effectiveness of the sampling method is hereinafter tested in the problem of optimal sensor placement. The idea is to define a total set of contamination events, which are then sampled through the procedure described above. Then, the effectiveness of the sampling procedure must be proven through the comparison of the results of optimal sensor placement in the total and sampled groups of events.

Regarding the objective functions to adopt in the optimization, a lot of them have been explained in literature to face the problem of the optimal sensor-location. In order to demonstrate the proposed sampling method, two generic functions are selected: the sensor redundancy, f_1, and contaminated population, f_2, resulting in a bi-objective formulation of the stated problem.

The objective function f_1 is related to the number n_r of sensors that can detect the generic contamination event within a time interval Δt_{red} (to be specified) following the first event detection. For a generic contamination event r, n_r is equal to 0 if no sensor can detect the event; it is equal to 1 if only one sensor can detect the event; it is equal to 2 if two sensors can detect the contamination event in close times, that is, the first detection sensor and an extra sensor that detects the event within a time interval Δt_{red} following the first event detection; generalizing the concept, n_r is equal to x when, besides the first detection sensor, there are other x-1 sensors able to detect it within Δt_{red} following the first detection. After assessing n_r for each contamination event, f_1 is calculated as the weighted average value of n_r:

$$f_1 = \frac{\sum_{r=1}^{S} w_r n_r}{\sum_{r=1}^{S} w_r}, \tag{4.1}$$

where S is the total number of contamination events and w_r is a weight coefficient associated with the generic contamination event. This coefficient is set to 1 if no event sampling has been carried out. Otherwise, it is set equal to the operating phase duration. A large value of f_1 is associated with a large redundancy in the system. This means that, on average, there are numerous sensors able to detect the generic contamination event in the system in a short time interval between one another. Therefore, should a sensor fail, another sensor would be able to give the warning in its place.

The objective function f_2 is related to the contaminated population *pop* before the first detection of the generic contamination event. In the generic contamination event r, the nodes contaminated before the first event detection can be evaluated, and pop_r can be assessed by summing up the inhabitants served by the contaminated nodes. In details, the evaluation of the contaminated population takes into account the residential population belonging to each single node, neglecting indeed the urban mobility that varies according to the time. It is assumed that a warning is given to interrupt network service in a reaction time interval Δt_{react} after the event detection. Hereinafter, Δt_{react} is set to 0 for simplifying purposes, but can be set to other values without loss of validity of the whole methodology. After assessing *pop* for each contamination event, f_2 is calculated as the average value of pop_r:

$$f_2 = \frac{\sum_{r=1}^{S} w_r pop_r}{\sum_{r=1}^{S} w_r} \qquad (4.2)$$

For the generic location of sensors in the network, the objective functions can be assessed through simple manipulations on the contamination and time matrices explained in section 4.1. Functions f_1 and f_2 are minimized simultaneously as mutually contrasting objectives in the bi-objective optimization process. In detail, the minimization of the former yields benefits of system cost, whereas minimization of the latter impacts positively on the system security. Therefore, the optimization results consist of a Pareto front of compromised solutions.

Regarding the optimal sensor placement, efficient algorithms can be used to find a global optimum when specific objective functions are used. For example, Kessler et al. (1998) and Ostfeld and Salomons (2004) solved a set-covering problem, whereas Propato and Piller (2006) solved a MILP problem. Additionally, it is possible to solve with a greedy algorithm in very efficient manner, even for large networks (e.g., Cheifetz et al., 2015), with an additional optimal sensor added at each iteration. Nevertheless, although being able to guarantee only the near-optimality of the solutions, genetic algorithms have the advantage of being easily implementable with whatever objective functions, even in the multi-objective framework. Therefore, for the bi-objective optimization of this paper, non-dominated sorting genetic algorithm II (NSGAII) (Deb et al., 2002) was chosen.

In NSGAII population individuals, the number of genes is equal to the number of network nodes where sensors can be installed. Each gene can take one of two possible values, 0 and 1, which stand for absence and presence of the sensor in the node associated with the gene, respectively. At each NSGAII generation starting from the initial population, the parent population is selected based on its fitness. The algorithm then generates the offspring population through crossover and mutation from the parent population. After being obtained as a combination of the parent and offspring populations, the new population is sorted according to fitness criteria, with the best individuals chosen in order to keep the total

number of population individuals constant during generations. The process is repeated until the maximum number of generations.
To ensure robustness of the end solutions found, which are expected to be close to the global optima, a certain number (n_{par}) of NSGAII runs can be carried out in parallel. The ultimate solutions are then put together and some solutions are sampled on the basis of their fitness. The sampled solutions can be used inside the population of new parallel NSGAII runs. This process can be repeated for a certain number of times (n_{iter}).

4.2.3 Case Study

The presented method was developed and applied to a WDS, that is the pipe network model used as benchmark in the Battles of Water Networks of the last Water Distribution Systems Analysis (WDSA) conferences (Marchi et al., 2014). The pipe and node characteristics for this district were reported by Creaco and Pezzinga (2015). The number of inhabitants connected to each network node is reported in Table 4-1.
The choice of the NSGAII settings was made based on the results of preliminary simulations unreported here. In particular, the NSGAII run was carried out considering a population of 50 individuals and a maximum number of 50 generations.
Furthermore, both n_{par} and n_{iter} were set to 5; Δt_{red}, useful for the evaluation of f_1 [Eq. (4.1)],was set to 0.5 hr.

Table 4-1. Inhabitants connected to the nodes in the first case study.

Node	Inhabitants	Node	Inhabitants
1	0	24	282
2	213	25	165
3	288	26	271
4	341	27	215
5	353	28	300
6	100	29	7
7	59	30	38
8	233	31	46
9	148	32	0
10	149	33	193
11	196	34	237
12	330	35	196
13	167	36	298
14	97	37	32
15	20	38	35
16	88	39	160
17	352	40	314
18	22	41	270
19	141	42	220
20	131	43	135
21	182	44	93
22	141	45	0
23	39	46	0

As shown in Figure 4-8, the network of the first case study had 45 demanding nodes, 52 pipes, and 1 tank. In the lowest node in the layout, the water input from a pumping station was considered as a negative demand, as previously done by Creaco and Pezzinga (2015).

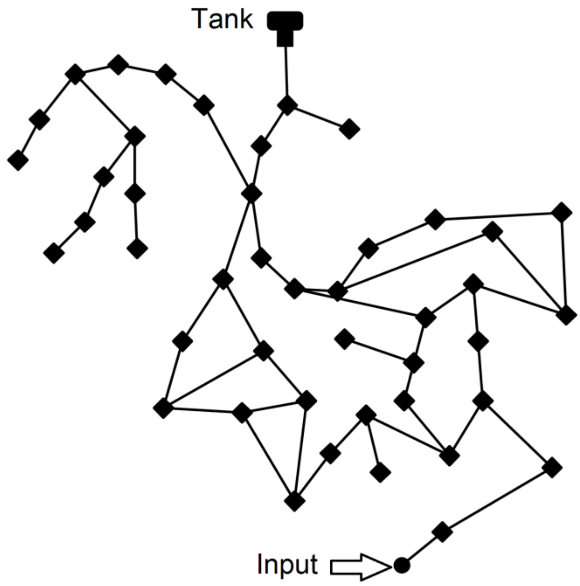


Figure 4-8. Case study layout

In this case study, the following assumptions were made to define the whole sets of contamination events:

i. All the nodes except for the tank and the input node, that is 44 nodes, were considered possible injection locations;
ii. Possible injections were assumed to occur every 30 minutes, leading to 48 possible values of the starting time in the day;
iii. Mass injection rate offered four possible values of 50, 200, 350, and 500 gr/min;
iv. Injection duration offered five possible values of 60, 220, 380, 500, and 600 min.

Assumptions 2, 3 and 4 were taken from the work of Preis and Ostfeld (2008b).
Therefore, the *S* total number of contamination events was 44*48*4*5= 42,240. Once *S* was set, the contamination and time matrices could be evaluated, as explained in the "Sampling Methodology" section.
In network modeling for the construction of the total contamination and time matrices, the multiplying coefficients used for nodal demands were expressed through 1-day-long patterns with 24 hourly steps. Because injections were assumed to take place during the first day of network operation, the simulations had to be conducted for 3 days, because the highest residence time in the

network is approximately 24 h. This was done to make sure that even contaminants injected close to the sources at the last instant of the first day had enough time to leave the network. Furthermore, the fact that the simulation duration is superior to the residence time inside the network (Piller et al., 2015) is sufficient to avoid the influence of initial conditions on the numerical concentration solution.
Sampling for the selection of the most representative contamination events was carried out on all the variables, i.e., location, starting time, mass rate, and duration. The optimizations were carried out to search for solutions up to a number of sensors equal to the number of nodes with positive demand, i.e., 44.

4.2.4 Results and Discussion

This section presents the results for the bi-objective optimal placement of sensors, aimed at minimizing simultaneously sensor redundancy and the contaminated population in the discussed case study.
The scenario considering the total number of contamination events is indicated as S0. The method proposed for sampling the events was applied considering various scenarios (S1, S1a, S1b, S1c, S1d, and S2) to reduce the size of the contamination and time matrices. The sampling was done considering (1) frequencies of one out of two or one out of three for the injection location, (2) representative starting times of 0, 5, and 18 h, (3) intermediate mass injection rate of 200 gr/min, and (4) smallest event duration of 60 min.
Scenarios S1, S1a, S1b, S1c, and S1d were obtained considering the location sampling frequency of one out of two nodes, whereas Scenario S2 was obtained considering the location sampling frequency of one out of three nodes. Subsequently, 23 and 19 possible injection locations were sampled in S1, S1a, S1b, S1c, and S1d on the one hand (Figure 4-9a) and S2 on the other hand (Figure 4-9b), respectively. Furthermore, in S1 and S2, all variables other than the injection location, i.e., starting time, mass rate, and duration, were sampled at the same time. In S1a, S1b, S1c, and S1d, the sampling concerned one variable at a time instead. The features of all sampling scenarios are reported in Table 4-2. This table shows that, compared to Scenario S0, a large reduction in the number of events is obtained through the sampling method, above all in Scenarios S1 and S2 operating on all the sampled variables at the same time.

Table 4-2. Features of the sampling scenarios in the first case study.

scenario	injection sampling with one out of two frequency	injection sampling with one out of three frequency
S0		
S1	X	
S1a	X	
S1b		
S1c		
S1d		
S2		X

scenario	starting time sampling	mass rate sampling	duration sampling	number of events
S0				42,240
S1	X	X	X	69
S1a				22,080
S1b	X			2,640
S1c		X		10,560
S1d			X	8,448
S2	X	X	X	57

In fact, Scenarios S1 and S2 are made up of 69 and 57 events, respectively, which are smaller than the number of events in S0 (42,240) by three orders of magnitude.

Genetic algorithm (GA) applications enabled deriving the Pareto fronts of optimal solutions in the trade-off between sensor redundancy and contaminated population in all the scenarios. Figure 4-9 reports the Pareto front obtained in Scenarios S1a, S1b, S1c, and S1d in comparison with that of S0. As expected, each front shows decreasing values of the contaminated population as the sensor redundancy, and therefore the number of installed sensors, increases. Furthermore, the best benefits in terms of contaminated population are obtained up to a redundancy of 2.5 sensors. Analysis of Figure 4-9 reveals that the fronts obtained in Scenarios S1a, S1b, S1c, and S1d are close to the S0 front. As expected considering the linearity of the advection-reaction equations, the fronts are almost coincident in Figure 4-9c associated with the mass sampling. However, in the other cases, the differences between the fronts are small.

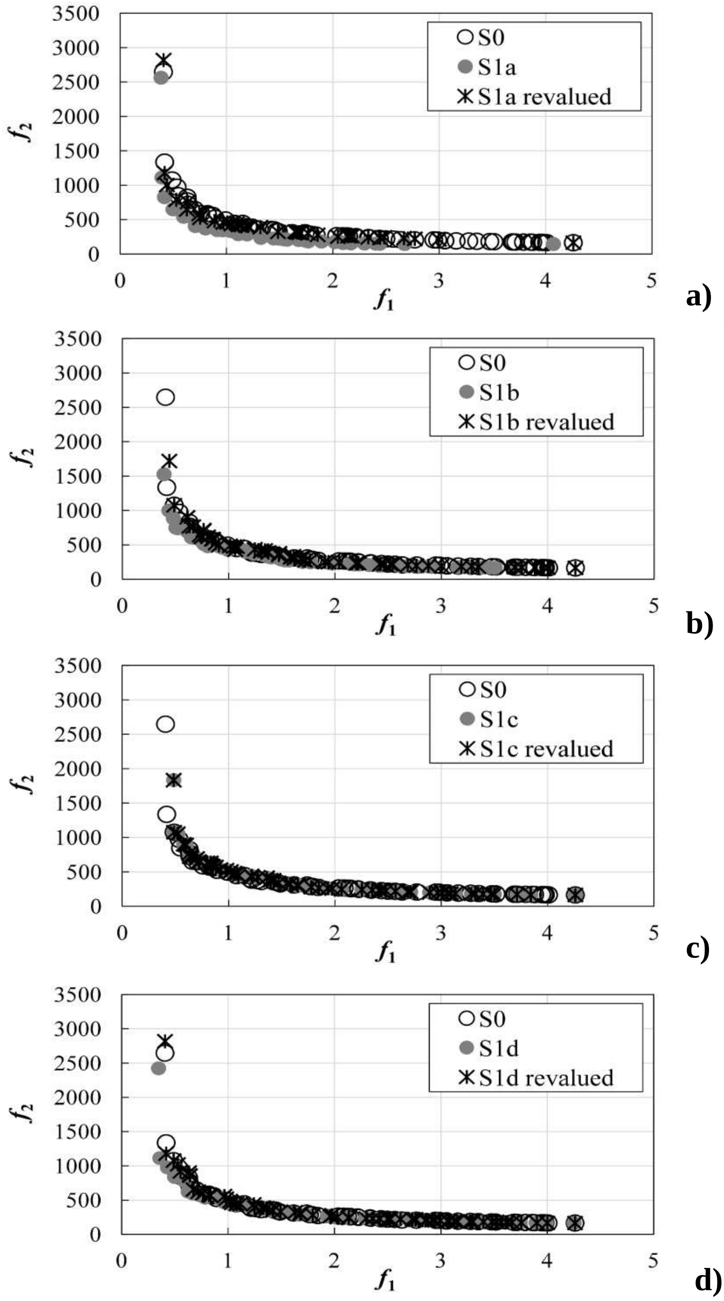


Figure 4-9. Pareto fronts of optimal solutions in the tradeoff between sensor redundancy (f_1) and contaminated population (f_2) in Scenarios S0 and (a) S1a; (b) S1b; (c) S1c; (d) S1d; curves of optimal S1a, S1b, S1c, and S1d solutions revalued in Scenario S0

To better compare the results obtained in S1a, S1b, S1c, S1d, and S0, the optimal sensor locations obtained in S1a, S1b, S1c, and S1d were tested in S0, considering the totality of contamination events. This led to the curve of revalued solutions in Figure 4-9. The closeness of these curves to the Pareto front of S0 attests to the fact that the performance of the optimal sensor-location solutions obtained in S1a, S1b, S1c, and S1d do not decay when tested against the totality of events of S0. Similar remarks can be made as in the comparison of Scenarios S1 and S2, featuring sampling on all the variables, with Scenario S0 (Figure 4-10).

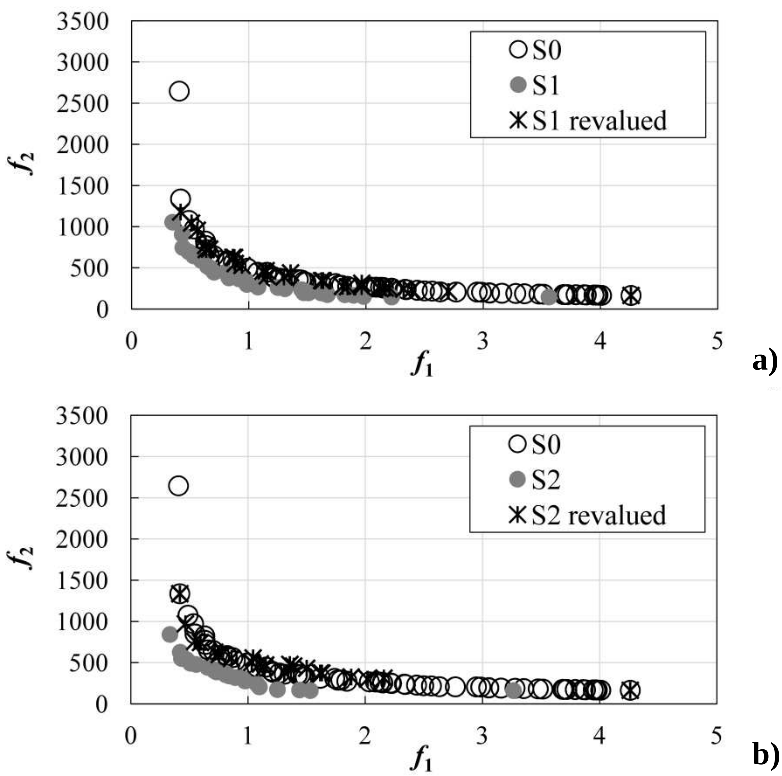


Figure 4-10. Pareto fronts of optimal solutions in the tradeoff between the sensor redundancy (f_1) and contaminated population (f_2) in Scenarios S0 and (a) S1; (b) S2; curves of S1 and S2 optimal solutions revalued in Scenario S0

As an example of the obtained solutions, Figure 4-11 shows identical locations of four sensors in Scenarios S0 and S1, with values of f_1 and f_2 equal to 0.63 and

772.21, respectively, assessed based on the total number of contamination events (S0).

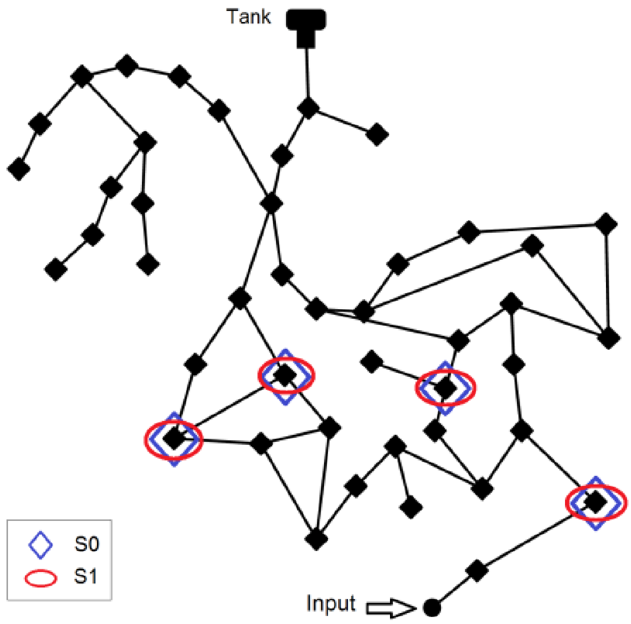


Figure 4-11. Optimal location of four sensors in Scenarios S0 and S1

Figure 4-11 shows that one of the four sensors is located close to the water input, whereas the other three are halfway between the input and the tank. The four locations were selected by the optimizer to promptly detect the generic contamination event, wherever it takes place in the network, and in an attempt to compromise the contaminated population with the sensor redundancy. The results shown in Figure 4-11 corroborate the previous findings concerning the representativeness of the sampled events and the effectiveness of the sampling method.

4.3 Selection of the objective functions

In section 4.2, the problem of the optimal sensor location was formulated in mathematical terms, defining decision variables, objective functions, constraints, and any possible modeling assumptions.
Regarding the objective functions, two of them were selected only for the purpose of demonstrating the proposed sampling methodology. However, there are many objective functions that can be used and their choice significantly influences the results of the optimization process. For this reason, the impact of the objective function selection was investigated in the optimal placement of water quality sensors.
This research is hereafter explained: a bi-objective optimization is used to search for the sensor optimal locations in the network and it has been applied to a real WDS.

4.3.1 Description of Objective Functions

Several competing design objectives have been used for sensor placement. In fact, on one hand, some objective functions minimize the cost of the system, such as the number of installed sensors.
On the other hand, many objective functions can be taken into account to minimize the impact of contamination events on public health, such as:

- the detection likelihood (to be maximized);
- the redundancy, which is related to the number of sensors that can detect the generic contamination event within a specified time interval, following the first event detection (to be maximized)
- the population exposed to a contaminant or the number of individuals receiving a dose above a fixed threshold (to be minimized);
- the detection time, defined as the elapsed time from the beginning of the contamination event to the instant of the detection in the first sensor (to be minimized);
- the extent of the contamination in the pipe network (to be minimized);
- the percentage of the contamination incidents not detected (to be minimized);

- the volume of the delivered water after the contamination or the volume of the contaminated water consumed prior to detection;
- etc.

Like in section 4.2, the optimization problem has been formulated as a bi-objective problem but two different variants are here used. Both of them adopt for the first objective function f_1, the number of installed sensors, as a surrogate for the total cost of the monitoring system. The difference between the two variants lies in the choice of the second objective function f_2, which accounts for the performance of the monitoring system.
In detail, the former variant considers the detection likelihood, which is the probability of events being detected by at least one of the installed sensors. This function to be maximized inside the optimization is calculated as follows:

$$f_2 = \frac{1}{S} \sum_{r=1}^{S} \mathrm{d_r}, \tag{4.3}$$

where S is the total number of potential contamination events considered in the analysis. Variable d_r is equal to 1 if at least one sensor detects the r-th contamination event; otherwise, it is equal to 0.
The second variant, instead, uses the average population contaminated before the first detection of the generic event. This function to be minimized inside the optimization is expressed as explained above [Eq. (4.2)].
For each solution considered inside the optimization process, variables d_r and pop_r can be assessed through simple manipulations on the matrices evaluated in section 4.1.
Since Genetic Algorithms (GA) have the advantage of being easily implementable with whatever objective functions, NSGA-II is still used to solve the optimization problem.
As the objectives clearly compete against each other, the output of the optimization still consists of a set of trade-off solutions, that is the Pareto front. Various criteria can be used by the decision maker to select the ultimate solution, such as a constraint in f_1 or f_2. Otherwise, the knee-point in the Pareto front can be identified, in which an increase in the cost of the monitoring systems f_1 is no longer paid back by a significant benefit in terms of f_2.

4.3.2 Objective Functions Application

The case-study considered in this work is a network in Northern Italy (Guidorzi et al., 2009; Creaco and Franchini, 2012), made up of 536 demanding nodes, 825 pipes and 2 reservoirs (layout in Figure 4-12).

Figure 4-12. Layout of Ferrara Network

The sampling procedure was applied to select a representative set of contamination events. As a result, all 536 demanding nodes of the WDS were considered possible injection locations. Single values of mass rate and injection duration, equal to 200 g/min and 60 min, respectively, were considered following the assumption that contamination events should be massive. Only one representative starting time was accounted for, that is 8:00 a.m., because preliminary analyses showed the network to have a single operating condition (i.e., no flow inversion at any pipes). The overall number *S* of contamination events was then equal to 536.

As in the case study reported in section 4.2.3, the system water demand was assumed to vary with hourly steps. Therefore, 1-day-long patterns were used for the demand multiplying coefficients and the simulations were run for 3 days.

The NSGAII settings were chosen based on the results of preliminary simulations unreported here, which enabled obtaining a trade-off between accuracy of the results and computational overhead. In detail, n_{in} and the maximum number of generations were both set to 500. Furthermore, n_{par} and n_{iter} were both set to 5.

The graphs in Figure 4-13 show the Pareto fronts of optimal trade-off solutions in the two variants of optimization. In graph a), associated with the first variant, a monotonous trend of $f_2(f_1)$ is shown, in which a significant benefit in terms of detection likelihood (f_2) is obtained as the number of installed sensors (f_1) increases up to about 10, which is close to the knee of the front. A further increase in f_1 does not yield significant benefits. Compared to graph a), the main difference of graph b) lies in the monotonous decreasing trend of the contaminated population f_2. The position of the knee of the front in the results of the second variant is also close to f_1=10.

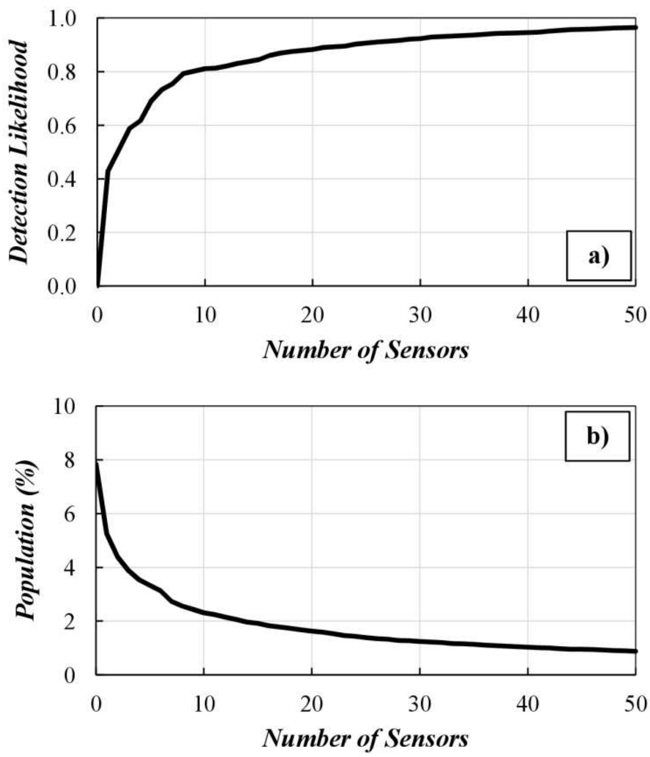
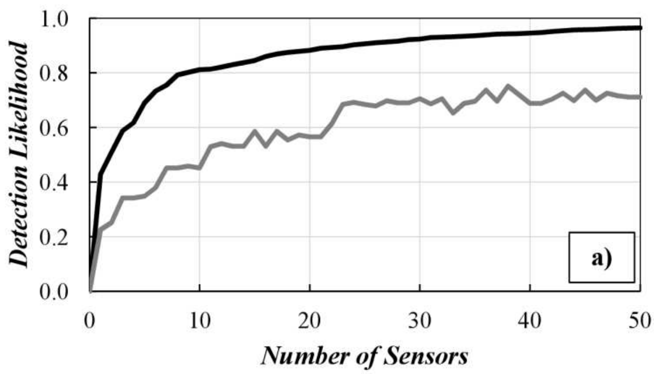


Figure 4-13. Pareto fronts obtained in the first a) and second b) variant of optimization

To thoroughly compare the solutions obtained in the two variants of optimization, these solutions were re-evaluated in terms of four effectiveness indicators for the water quality monitoring system. Besides the detection likelihood and contaminated populations, evaluated over the whole group of *S* events through equations 1 and 2 respectively, the detection time and the sensor redundancy were adopted as a benchmark. These two additional indicators, instead, were assessed over the sub-group of detected events, that is the events that are detected by at least one sensor. In detail, the detection time is the average time elapsing between the contamination start and the time instant when the first sensor is reached. The redundancy is defined as explained above [Eq. (4.1)], which contributes to the safety of the monitoring systems.

The graphs in Figure 4-14 show the curves of re-evaluated solutions plotted against the number of installed sensors. Looking at the solutions of the first variant of optimization, the curve in graph a) coincides with the Pareto front in Figure 4-14a and then features a monotonous increasing trend. The trend of the curves in the other graphs is not strictly monotonous since the contaminated population (graph b), the detection time (graph c) and the sensor redundancy (graph d) were not objective functions in the first variant of optimization. In fact, optimal solutions are usually sub-optimal when re-evaluated in terms of different indicators from the objective functions used in the optimization.



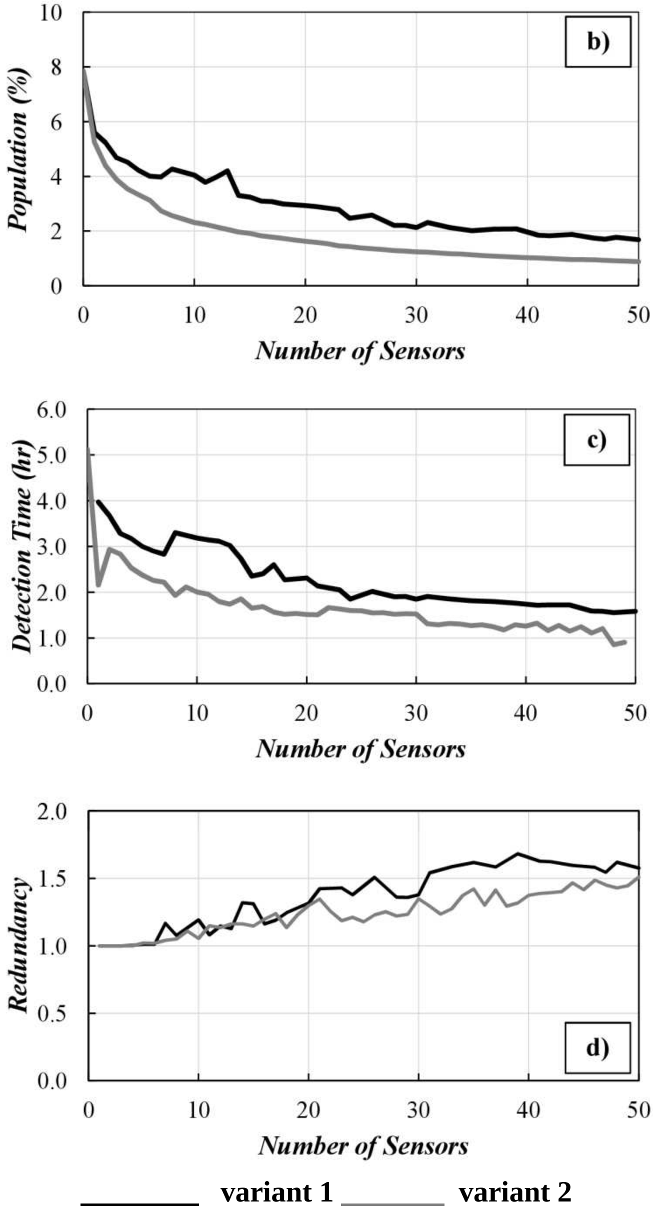


Figure 4-14. Solutions obtained in the two variants of optimization, re-evaluated in terms of a) detection likelihood, b) contaminated population, c) detection time and d) sensor redundancy

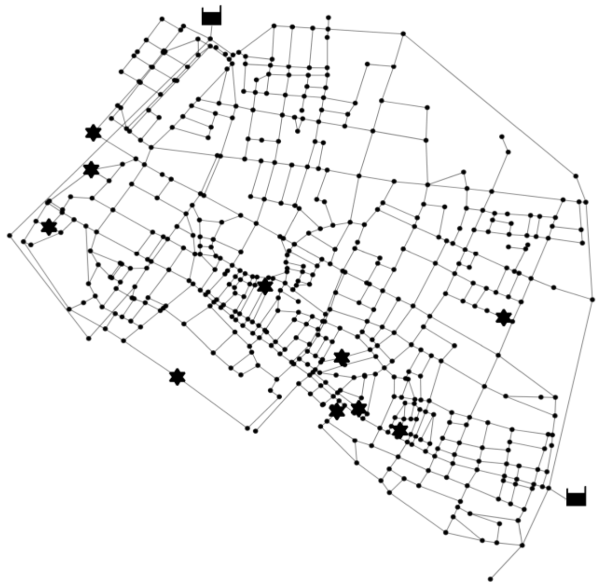
Analogously, looking at the solutions of the second variant of optimization, the curve in graph b) coincides with the Pareto front in Figure 4-13b and then features a monotonous decreasing trend. The trend of the curves in the other graphs is not strictly monotonous since the detection likelihood (graph a), the detection time (graph c) and the sensor redundancy (graph d) were not objective functions in the second variant of optimization. However, as Table 4-3 shows, the four effectiveness indicators are always strongly intercorrelated in both variants of optimization.

Table 4-3. Correlation coefficient between the objective functions used in the two variants of optimization and the four effectiveness indicators.

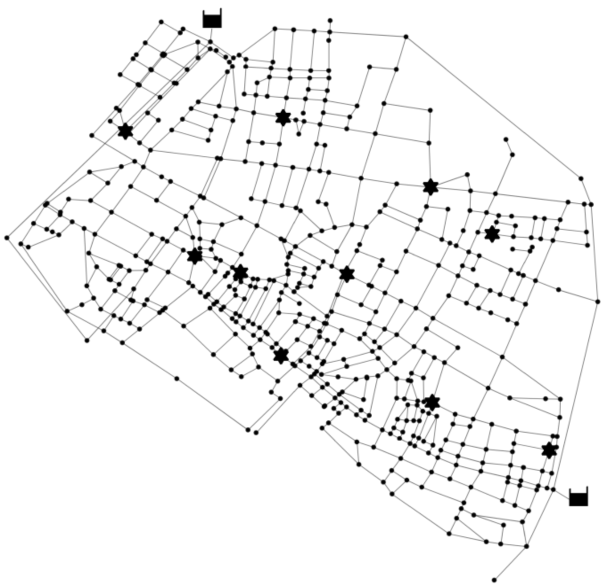
	First Variant	**Second Variant**
	Detection Likelihood	**Contaminated Population**
Detection Likelihood	1.00	-0.96
Contaminated Population	-0.94	1.00
Detection Time	-0.90	-0.84
Sensor Redundancy	0.83	0.84

Overall, the analysis of the results in Figure 4-14 shows that neither variant of optimization is superior. In fact, the first variant yields solutions that better perform in terms of detection likelihood and sensor redundancy, both positive indicators of the effectiveness of the monitoring system (black line above grey line in graphs a) and d). The second variant, instead, produces better performing solutions in terms of contaminated population and detection time, both inverse indicators of the effectiveness of the monitoring system (grey line below black line in graphs b and c). However, by leaning on graphs such as those in Figure 4-14, water utility managers can choose the ultimate solution for *in situ* installation based on their budget (which impacts the number of installed sensors), on the effectiveness indicator they prefer and on the degree of effectiveness they aim to reach in terms of the various indicators. As an example, the solution obtained through variant 1 with 10 sensors has a detection likelihood of 0.81, a contaminated population of 4%, a detection time of 3.2 hr and a redundancy of 1.2. The solution obtained through variant 2 with 10 sensors, instead, features almost halved detection likelihood (0.45) and contaminated population (2.3%), a lower detection time (2.1 hr) and a similar sensor redundancy (1.1). It is important to underline that the objective function related to the contaminated population *pop* provides the number of inhabitants that are averagely reached by the contamination, whose health effect depends on many factors such as, the type of contaminant, its concentration, and the alarm timeliness. In most of the possible contamination events, a significant damage to the affected population can be reasonably excluded thus, its non-zero value can be accepted, as long as it is sufficiently small.

Another criterion that can be adopted for the choice concerns the location of the sensors in the various optimal solutions. As an example, Figure 4-15 enables analysis and comparison of the results of the two variants of optimization, in terms of optimal placement of 10 sensors. Figure 4-15a shows that the placement obtained in the first variant is made up of sensors located in the intermediate area of the network, that is at the maximum hydraulic distance from either reservoir. This happens because most water paths outgoing from the reservoirs converge to this area. Therefore, the placement of sensors in this area is essential for maximizing the event detection likelihood. In the second variant, sensors are more scattered over the whole layout at gradually increasing distance from the reservoirs (Figure 4-15b), to guarantee early warning and therefore reduced impact in terms of contaminated population.



a)



b)

Figure 4-15. Optimal locations of 10 sensors for the a) first and b) second variant of optimization

As a confirmation of the results described above for 10 sensors, Figure 4-16 shows the optimal locations of 5 and 15 sensors for the two variants.

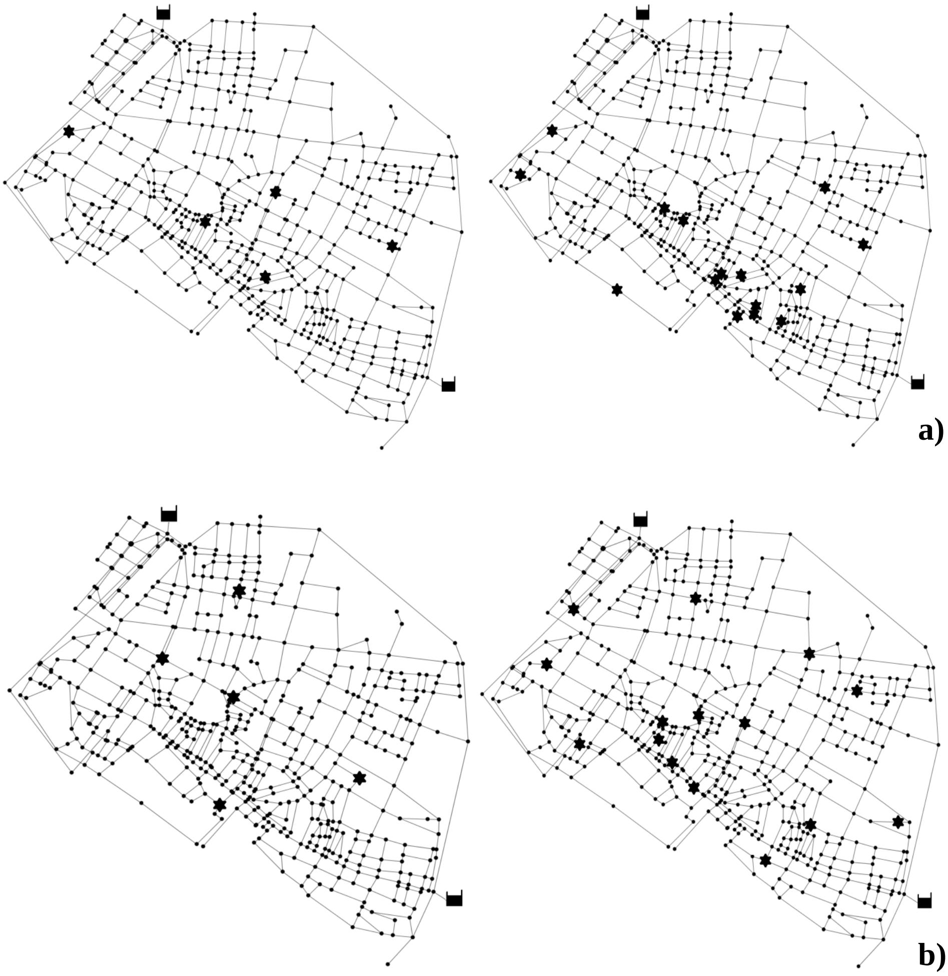


Figure 4-16. Optimal locations of 5 and 15 a) first and b) second variant of optimization

Chapter 5

Modeling and Measuring non-conservative contaminants

In the previous chapters contaminants were assumed to be conservative, i.e. their changes when dissolved in water were neglected. Though being a good assumption of the first attempt in the context of optimal sensor placement, the attention received by water quality topics worldwide, also inside Water Safety Plans (WSPs), spurred the writer to abandon this assumption for better analyzing the actual behavior of the contaminants. In fact, once dissolved in water, the chemical or biological substances can react with each other, with pipe walls, as well as with water, or also transform themselves into other compounds, still propagating throughout the water distribution network. Some substances may also precipitate in pipes; thus, the assumption of conservative contaminants would overestimate too much the propagation effects. For examples, *Arsenic Pentoxide* (As_2O_5) is often used as a solution in the production of herbicides, metal adhesives or insecticides. It quickly dissolves in water forming *Arsenic Acid* (H_3AsO_4), which is characterized by high toxicity. The pesticide *Chlorpyrifos* (CP) is a moderately toxic insecticide (toxicity category 2), able to oxidize in the presence of free chlorine (Duirk and Collette, 2006). Consumption of CP causes malfunction of the nervous system and may result in death in the case of large consumption. Again, the powerful insecticide *Parathion* (PA) is capable of oxidizing to *Paraxon* (PAO) in the presence of free chlorine (Duirk et al., 2009). Consumption of extremely small volumes of *Paraxon* (PA) (3-5 mg/kg body weight) results in death.

Therefore, it is very important to consider the reactions that occur following the injection of contaminants into the network.
The simulations presented so far have been performed with EPANET, which carries out hydraulic and quality simulations in water distribution networks even for extended periods of time. Within the Quality Module, EPANET enables the modeling of (i) a non-reactive tracer material through the network over time (ii) the movement and fate of a reactive material as it grows (e.g., a disinfection by-product) or decays (e.g., residual chlorine) with time (iii) the age of water throughout a network (iv) the reactions both in the bulk flow and at the pipe wall (v) time-varying concentration or mass inputs at any location in the network. However, EPANET is not able to model the behavior of a generic contaminant injected into the network because it is limited to track the transport and fate of just a single chemical species, such as fluoride used in a tracer study or free chlorine used in a disinfectant decay study (Rossman, 2000).
For this reason, the research continued using another software, that is EPANET Multi-Species Extension (Shang et al., 2008).
It is an EPANET plug-in that allows modeling of any system of multiple, interacting chemical and biological species (Shang and Uber, 2008). In particular, EPANET-MSX computes the flow transported volume and apply dynamic reactions within each pipe segment and storage tank over the defined time step. It takes into account either bulk species and surface species, enabling modeling of the interaction between any contaminants with bulk species and pipe wall surface. The water quality contained in the system can be modeled using principles of conservation of mass coupled with reaction kinetics.
Consequently, EPANET-MSX can simulate any injected substances in the network but sets of differential-algebraic equations are used, along with all the required kinetic constants, equilibrium equations, and bulk/wall coefficients.
Being aware that not all the potentially injected substances can be considered, this chapter presents the research carried out by Tinelli, Juran & Cantos (2017), Tinelli and Juran (2017) because it faces the problem of the presence of the Escherichia Coli bacterium (*E. coli*) in WDSs. Since chlorine is added to drinking water in order to kill certain bacteria and other microbes, the research had to analyze the fate and transport of *E. coli* when chlorine is in the network, identifying the adequate kinetic models for the chlorine decay.
Once an appropriate EPANET-MSX model was defined, it had to be tested for its proper functioning and accuracy before its application to a real distribution network. To face this problem, a pilot laboratory site was created at the

University of Lille, able to test the behavior of the injected chemical/biological substances. In fact, in order to confirm the validity of the multi-species model, the pilot lab allowed the comparison with the numerical simulations reproducing the real conditions of WDSs, especially in terms of materials, velocity and pressure, and performing the same tests that had been modeled through EPANET-MSX (Abdallah, 2015).

5.1 Numerical Modeling of Chemical-Biological Contaminations

The common goal of the US National Drinking Water Regulation and the European Drinking Water Regulation is to protect public health by monitoring the level of the specific chemical/biological substances (e.g., chlorine, arsenic, iron, *E. coli*) in public water systems. Thus, several researches related to the survival and transport of these substances have followed the increasing concern for intentional intrusion of contaminants into drinking WDSs.

For example, even though many efforts have been made in the last decades, the modeling of chlorine is still complex, as it relies on the accuracy of hydraulic models to describe flows as well as flow velocities (Blokker et al., 2008; Pasha and Lansey, 2010) and on the adequacy of chlorine decay kinetic models (Fisher et al., 2011).

In addition, contamination of water by accidental entry of biological matter is very likely. Therefore, a multi-species model is used in the present research to analyze the fate and transport of the *E. coli* bacterium, incorporating the chlorine inactivation and its consequent decay through interaction with the organic matter itself.

The multispecies model was proposed by Pemmasani (2012) according to previous laboratory studies where the *E. coli* was grown in a nutrient broth, called Tryptic Soy Broth, and chlorinated water solution. The Tryptic Soy Broth has been assumed as a composition of three species, deriving a four-species model with eight parameters (TSB ratio/TSB reaction rate coefficients), where they can be determined using optimization and parameter estimation techniques (Bacteriological Analytical Manual, 1998).

In the EPANET-MSX model, all the reactions between the species are expressed in the form of differential equations; the *E. coli* inactivation because of the

chlorine is represented by the chlorine first order decay from a predefined initial chlorine concentration.
The differential equations were formulated as follows:

$$\frac{dC}{dt} = -k_1T_1C - k_2T_2C - k_3T_3C - kC \quad (5.1)$$

$$\frac{dT_1}{dt} = -k_1r_1T_1C \quad (5.2)$$

$$\frac{dT_2}{dt} = -k_2r_2T_2C \quad (5.3)$$

$$\frac{dT_3}{dt} = -k_3r_3T_3C \quad (5.4)$$

$$\frac{dE}{dt} = -k_eCE, \quad (5.5)$$

where:

- k_1= fast TSB reaction rate coefficient;
- k_2= medium TSB reaction rate coefficient;
- k_3= slow TSB reaction rate coefficient;
- k= Chlorine reaction rate based on the initial chlorine concentration;
- r_1= fast TSB pseudo-stoichiometric ratio;
- r_2= medium TSB pseudo-stoichiometric ratio;
- r_3= slow TSB pseudo-stoichiometric ratio;
- k_e= *E. coli* inactivation coefficient;
- C= Chlorine;
- E= *E. coli*;
- T_1= Fast reacting TSB;
- T_2= Medium reacting TSB;
- T_3= Slow reacting TSB.

Fractional coefficients of T1, T2, T3 are f1, f2, f3 respectively i.e., $f_1+f_2+f_3 =1$.

The k_e inactivation *E. coli* coefficient was predetermined by exploiting other studies (Rice et al., 1999), and it is equal to 0.10349 L/mg*s. This coefficient estimates the rate of *E. coli* decay due to interaction with the residual chlorine present in the WDS.
The influence of the initial chlorine concentration on the kinetics decay of the chlorine itself is considered within the model using the chlorine first order reaction (*k*). It was firstly derived from Uber et al. (2003) for an initial chlorine

concentration of 5.25 mg/L but the used value were drawn from Pemmasani (2012).
Table 5.1 lists the parameter values used in the model.

Table 5-1. Values of the Model Parameters

Symbol	Value	Description
k_1	0.04816	Fast TSB reaction rate coefficient, $L*mL^{-1}s^{-1}$
k_2	0.00953	Medium TSB reaction rate coeff., $L*mL^{-1}s^{-1}$
k_3	$9.07E^{-6}$	Slow TSB reaction rate coefficient, $L*mL^{-1}s^{-1}$
r_1	0.10218	Fast TSB pseudo-stoichiometric ratio, $L*mg^{-1}$
r_2	0.00630	Medium TSB pseudo-stoichiometric ratio, $L*mg^{-1}$
r_3	0.00201	Slow TSB pseudo-stoichiometric ratio, $L*mg^{-1}$
f_1	0.6414	Fast TSB partition coefficient
f_2	0.0718	Medium TSB partition coefficient
f_3	0.2868	Slow TSB partition coefficient
k	$1.620E^{-6}$	Cl bulk reaction rate coefficient, /sec

The final outcome of the EPANET-MSX simulations is made up of (i) the trend representation of the considered species, such as chlorine and *E. coli* (ii) the calculation of their concentrations at all nodes/pipes of the network over the analysis time period (iii) the evaluation of the parameters specified in the input files (e.g., pH, temperature, TOC, etc.).
All the simulations let the user impose the coefficients of the reactions between the different chemical and biological species (refer to Shang et al., 2008).

5.2 Experimental Activities for Chemical-Biological Contaminations

This section illustrates the pilot laboratory site created at the University of Lille.
As already mentioned, the laboratory was built to test chemical reactions and transformations that occur once the *E. coli*, as well as the chlorine, have been injected into the network. In addition, being connected to some sensors, the pilot

lab can also test the ability and the sensitivity of sensors available on the market in the detection of chemical or biological contamination in drinking water.

5.2.1 Pilot laboratory site at the University of Lille

As part of the European project Smart Water for Europe (SW4EU), the pilot laboratory site was built at the Laboratory of Civil Engineering and Geo-Environment, located in the Campus of Lille University (a scientific city in Northern France).

The pilot lab is 61 meters long and it includes: pipes for the circulation of water, tanks for the filling, a draining and a pump system, connections for the equipment to be tested through water quality sensors, valves for the control of flow direction, a device for the injection of contaminants, instrumentations for the control of water pressure and velocity.

Figure 5-1 shows the pilot laboratory scheme.

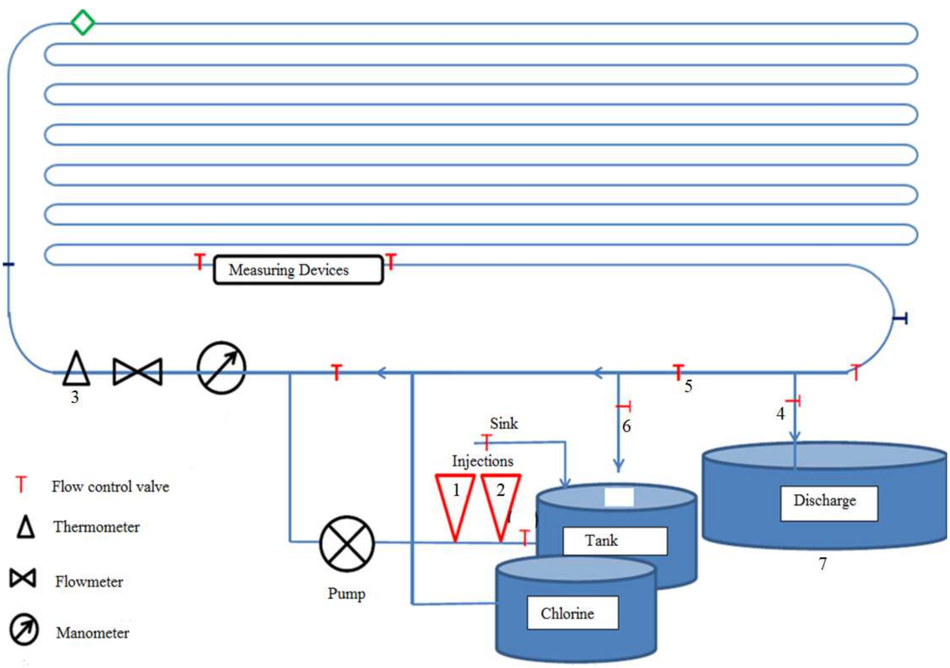


Figure 5-1. Scheme of the pilot laboratory site (Abdallah, 2015)

Opaque dual-layer pipes have been used in the pilot system to prevent the entry of light and the formation of bio-films. These pipes are 16 mm in diameter and they are made of aluminum on the outside and plastic on the inside.
A 40-liter tank feeds the system with a pump station. An additional chlorine tank is connected to the system to directly introduce chlorine.
Two specific funnels (1) and (2) have been used to inject chemicals and biological agents. The water quality sensors used in the experiments have been connected in line at 41 meters from the injection points.
Several manometers have been added to the circuit to permanently measure the pressure at any point of the circuit.
The flow of water is controlled by several valves and continuously measured with an automatic flow meter (3). The laboratory layout permits the circulation of water in an open circuit by opening the valve (4) and closing the valve (5), or in a closed circuit by opening the valve (6) and closing the valve (4).
During the open-loop experiments, the water is transmitted to an external tank for the discharge and treatment of the polluted water (9).
Finally, backflow preventer valves, and safety shut-off valves have been installed to prevent the return of contaminated water to the distribution system.

5.2.2 Simulation Processes

During the simulation processes, the feeding of the system is always supplied through the tank without direct connection to the main network in order to prevent the return of contaminated water. The flow rate is set by adjusting the individual valves in the downstream section of the pump.
Figure 5-2 shows the sequence of the performed analysis:

- The inlet tank is filled with tap water;
- The connection tap of the injection is open: contaminants are injected by feeding the two funnels, and then by letting the water go throughout the pump system;
- The pump allows the flow of the contaminated water in the circuit;
- The monitoring instruments are connected to the network, measuring the water quality parameters.

The other parameters (e.g., pressure, speed, ...) are also continuously monitored.

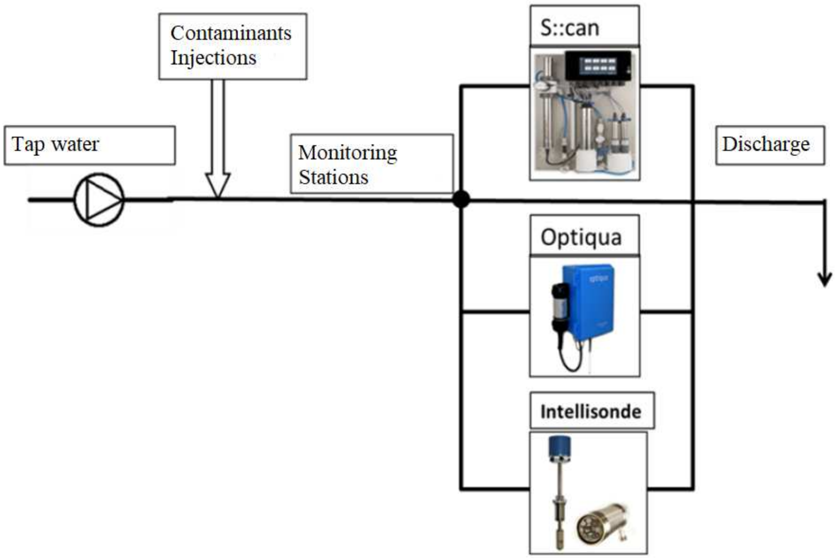


Figure 5-2. Sequence of the performed analysis (Abdallah, 2015)

As part of the European project SW4EU, *S::can*, *Optiqua* and *Intellisonde* were selected for the monitoring of water quality, determining the difference between the normal variation of background and a contamination event.

In particular, the *S::can station* combines several instruments, that is the *Spectro::Lyser*, the *S::can* probes (*i::scan, chlori::lyser, pH::lyser, condu::lyser*), and the terminal *con::cube*; these components are assembled in one single compact panel.

Optiqua EventLab measures the presence of substances in the water through the change in the refractive using the principle of interferometry, and *Intellisonde* uses electrochemical/optical technologies to control conductivity, temperature, pH, free and total chlorine, dissolved oxygen, oxidation-reduction potential, redox, turbidity and color.

Regarding the analysis, the tests can be carried out according to two protocols:

- Open circuit: the water is directly discharged for the treatment and the circuit is continuously supplied from the reservoir;
- Closed circuit: the valve (4) of the discharge tank is closed, the network is supplied to compensate only the water used in the monitoring instruments of the water quality.

For all the experiments, by setting the valves of the system, the flow rate is kept constant and equal to 3 L/min, the pressure is 2 bar in the downstream section of the pump. The measurements are continuously recorded.

5.3 Comparison between Numerical Modeling and Experiments

This section illustrates the comparison between the results obtained by the numerical analysis and those that come from the experimental activities.
In the experimental activities Abdallah (2015) stated that *EventLab* showed high reliability in detecting low concentrations of chemical contaminants and no ability in the detection of biological contaminants, while the *S::can* probes (especially the *spectro::lyser*) were able to detect either chemical and biological contaminants, for bacterial concentrations above 10^5UFC/mL). Therefore, the results obtained from the *S::Can* monitoring station were considered.

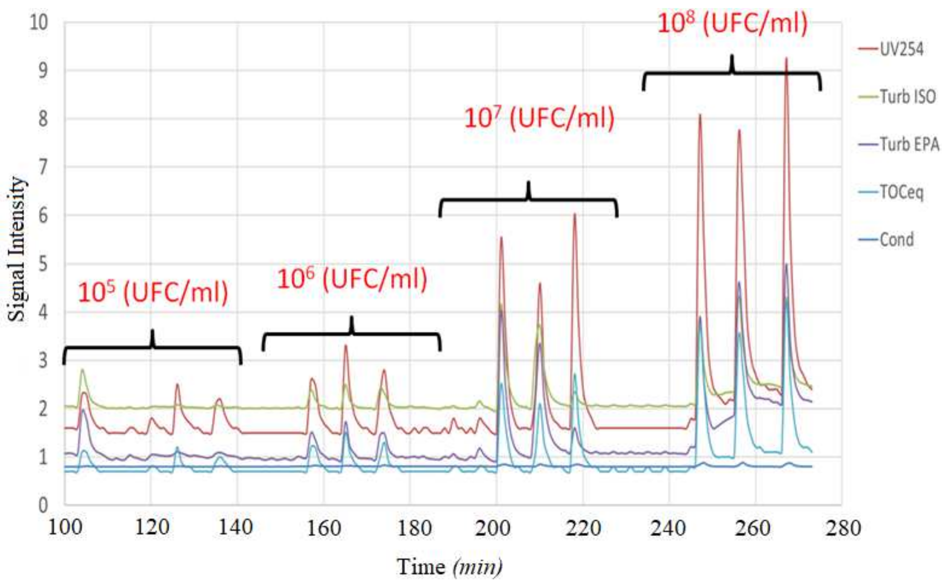


Figure 5-3. Results of *S::can* probes with different *E. coli* injections (Abdallah, 2015)

Figure 5-3 shows the trend of the parameters measured by the *S::can* probes according to different *E. coli* concentrations (from 10^5 up to 10^8 UFC/mL): the parameters start to change when the *E. coli* concentration reaches at least 10^5 UFC/mL.

In addition, Figure 5-4 shows that contaminant concentrations below this level are not detectable also by chlorine sensors. In fact, for an initial chlorine concentration of 0.25 mg/L, the chlorine level decreases by 0.05 mg/L after an *E. coli* injection of 10^5 CFU/mL. The chlorine decrease reaches 0.15 mg/L (60% of the initial concentration) with an *E. coli* injection of 10^8 CFU/mL.

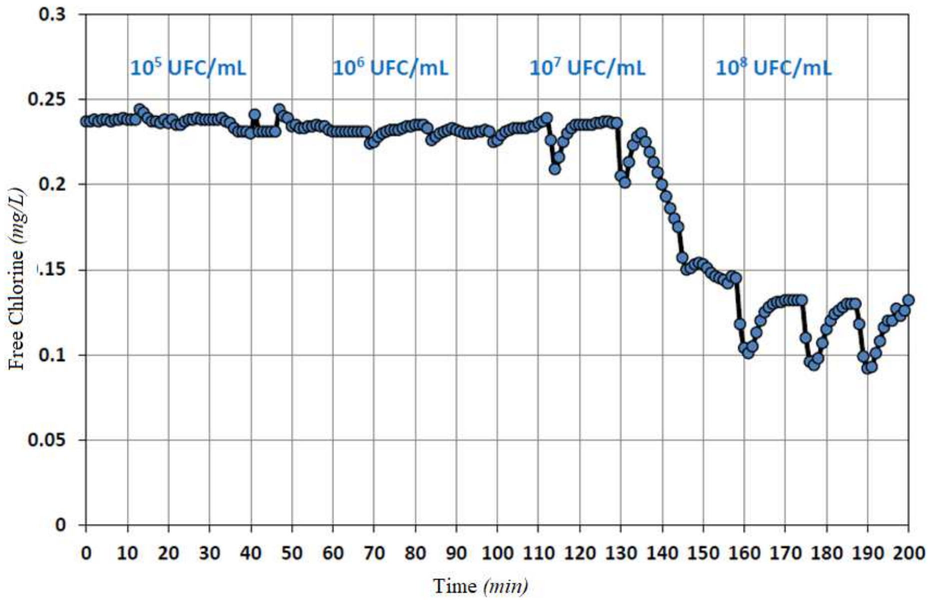


Figure 5-4. Free chlorine variation as a function of time in response to the injection of different *E. coli* concentrations (Abdallah, 2015)

Finally, Figure 5-5 illustrates the chlorine decrease with an *E. coli* injection of 10^5 CFU/μL. The *E. coli* was injected in the pilot lab and different initial chlorine concentrations were tested: 0,3 mg/L, 0,5 mg/L, and 1 mg/L.

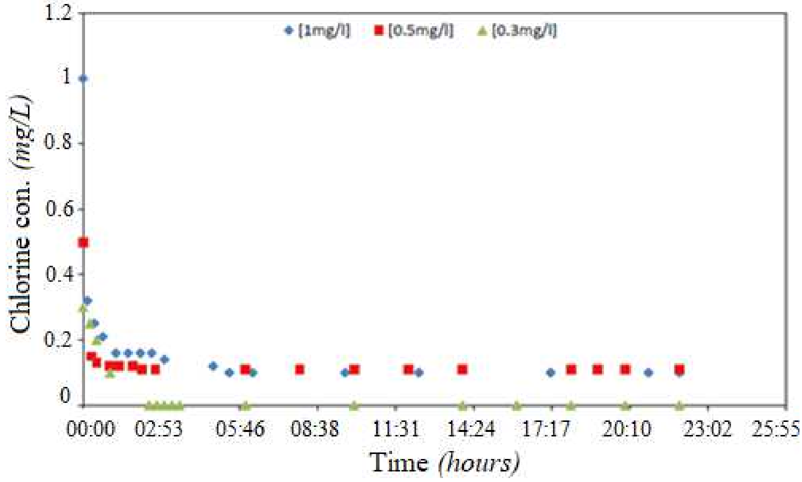


Figure 5-5. Chlorine concentrations with *E. coli* injection of 10^5 CFU/μL according to the experimental activities (Abdallah, 2015)

As regards the numerical analysis, EPANET-MSX requires three input files, that is (i) EPANET .inp file for the network definition (ii) .msx file which contains all the equations and reactions of the problem to be scanned (iii) .rpt file that is the file which the simulation results are saved in. Therefore, being known the pump curve and all the geometric-hydraulic characteristics of the network, the WDS of the laboratory had to be modeled in EPANET, as shown in Figure 5-6.

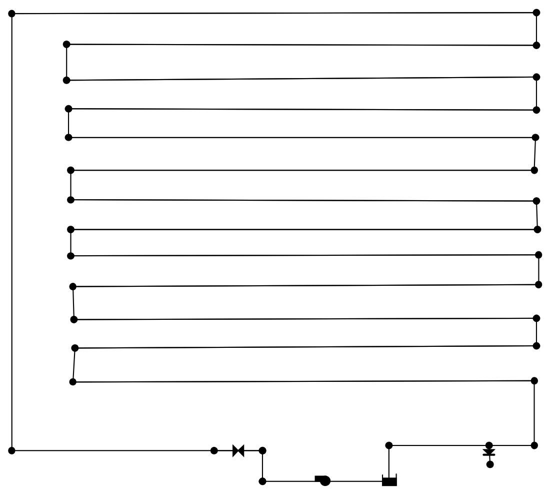


Figure 5-6. Layout of the EPANET model for the laboratory WDS

In details, the system water demand was assumed to vary with hourly steps, the simulations were run for 24 hours with the same features of the experimental activities thus, they considered:

- initial chlorine concentration of 0.3 mg/L, 0.5 mg/L and 1 mg/L;
- *E. coli* injection equal to 10^5 UFC/μL.

The comparison between the experimental activities and the numerical model is illustrated in Figure 5-7: the comparison is reported in terms of chlorine decrease, for an initial chlorine concentration of 0.3 mg/L (and an *E. coli* injection of 10^5 CFU/μL).

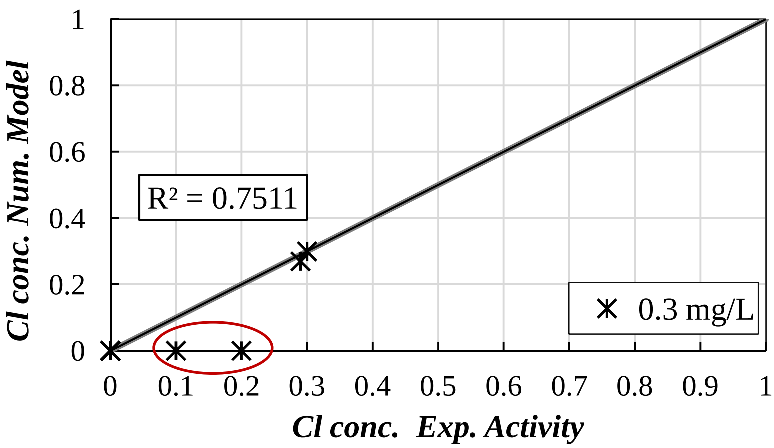


Figure 5-7. Comparison of chlorine concentrations between the experimental activity and the numerical model with *E. coli* injection of 10^5 CFU/μL

Figure 5-7 shows that chlorine concentrations reach zero in both of the two analysis but the chlorine decay is faster with the numerical model than the experimental activity. In fact, according to the numerical model, chlorine arrives at zero more rapidly. For this reason, the two red circled points in Figure 5-7 differ from the bisector: they specifically represent the delay of the experiments in achieving zero, as compared to the numerical model. These two points are also responsible for the reduction of the factor R^2, which is still satisfactory for water quality analyses.

An improvement of the Equations (5.1-5.5) and an enhancement in the model calibration could lead to a better matching of the two chlorine decreases.

Approximately the same results were obtained comparing the two studies for initial chlorine concentrations of 0.5 and 1 mg/L with an *E. coli* injection of 10^5 CFU/μL.

Therefore, both of the two studies demonstrate that (i) the chlorine decreases when the bacterium is injected into the network but only with an initial chlorine concentration of 0.3 mg/L the zero is reached, if the *E. coli* injection is equal to 10^5 CFU/μL (ii) the time required to reach the residual chlorine value increases with the decrease in the initial concentration of chlorine, and subsequently (iii) the detection of microbial contamination of the order of 10^5 Colony-Forming Unit (CFU)/μL is faster for chlorine concentrations between 0.5 and 1 mg/L.

Even using the numerical model, it is proved that chlorine does not drop to zero if the concentration of *E. coli* decreases. In fact, considering an initial chlorine concentration of 0.3 mg/L and injecting 10^7 CFU/mL of *E. coli*, the chlorine drops but it does not reach zero, as represented in Figure 5-8.

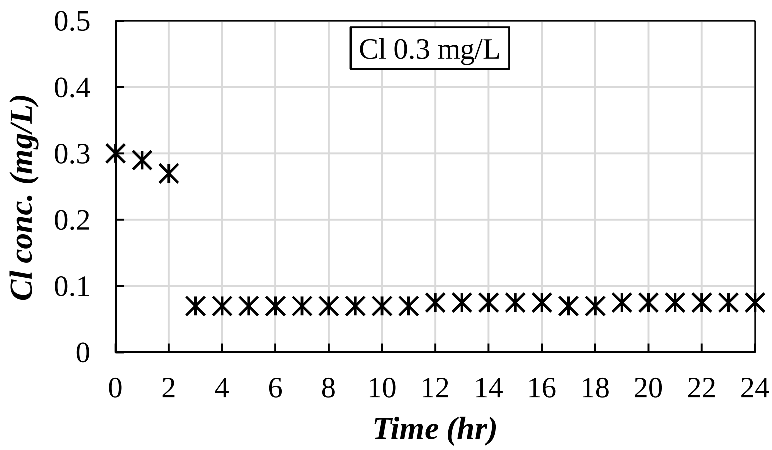


Figure 5-8. Chlorine concentrations with *E. coli* injection of 10^7 CFU/mL according to the numerical model

Thus, Figure 5-8 corroborates the experimental activity in terms of the threshold above which the bacteriological species is significantly detected.

Hence, the type of the injected contaminant and, above all, its dynamic interaction with the fluid and with the pipe walls, are of fundamental importance.

This aspect identifies EPANET-MSX as an essential support for qualitative studies in WDS because, unlike the traditional software used in the literature

(such as EPANET), it is able to describe the behaviour of the most common chemical/physical water reactions during *E. coli* presence and identify variation patterns in the biological and chemical parameters.
Finally, since currently there is no direct measure for biological contaminations in the water, the chlorine can be monitored to detect *E. coli* injections, and more in general, looking at all the measured water quality parameters, a pattern recognition able to distinguish the normal conditions of water quality from anomalies is recognized.

5.4 Extension of the study to a real water distribution network

Once the EPANET-MSX numeric model has been validated, it was possible to apply it to a real case study: the water distribution network of the Lille Cité Scientifique Campus.

5.4.1 Case Study

The Lille Campus is located in the town of Villeneuve d'Ascq (Northern France) from 1967, although its academic roots date back to 1562.
It covers an area of 110 hectares and it comprises 145 buildings with very different uses (teaching and research buildings, administrative buildings and university residences). It welcomes 25,000 users, including 4,000 students who live in the campus, presented below. The water supply system of the Lille Scientific Campus is relatively old: it was lain during the campus construction. The network is nearly 15 km long. The pipes are mainly made of cast iron with diameters ranging from 20 to 300 mm. The water network includes 49 fire hydrants, 250 valves, 93 Automatic Meters Readers (AMR) measure hourly water consumption, 5 pressure sensors and 2 Virtual District Metering Areas (VDMA). Therefore, the campus can be compared to a small city.
An EPANET hydraulic model was created in order to simulate the behavior of the network towards bio-contamination injections. EPANET Lille network has 393 nodes, 412 pipes and 5 reservoirs as shown in Figure 5-9.

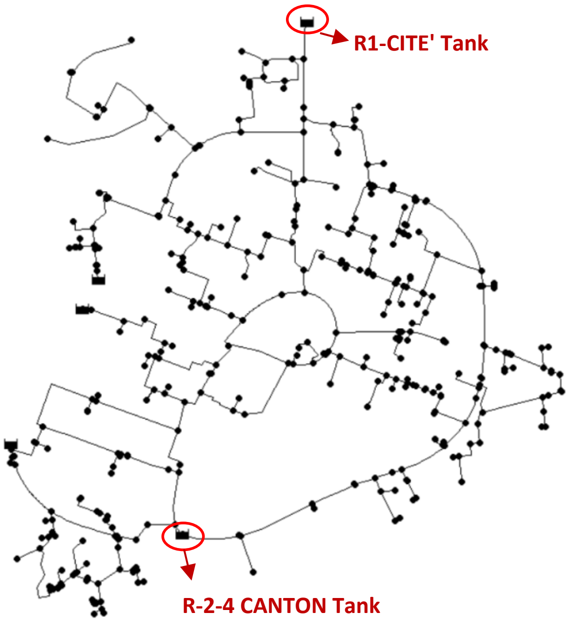


Figure 5-9. Network layout of Lille Cité Scientifique Campus

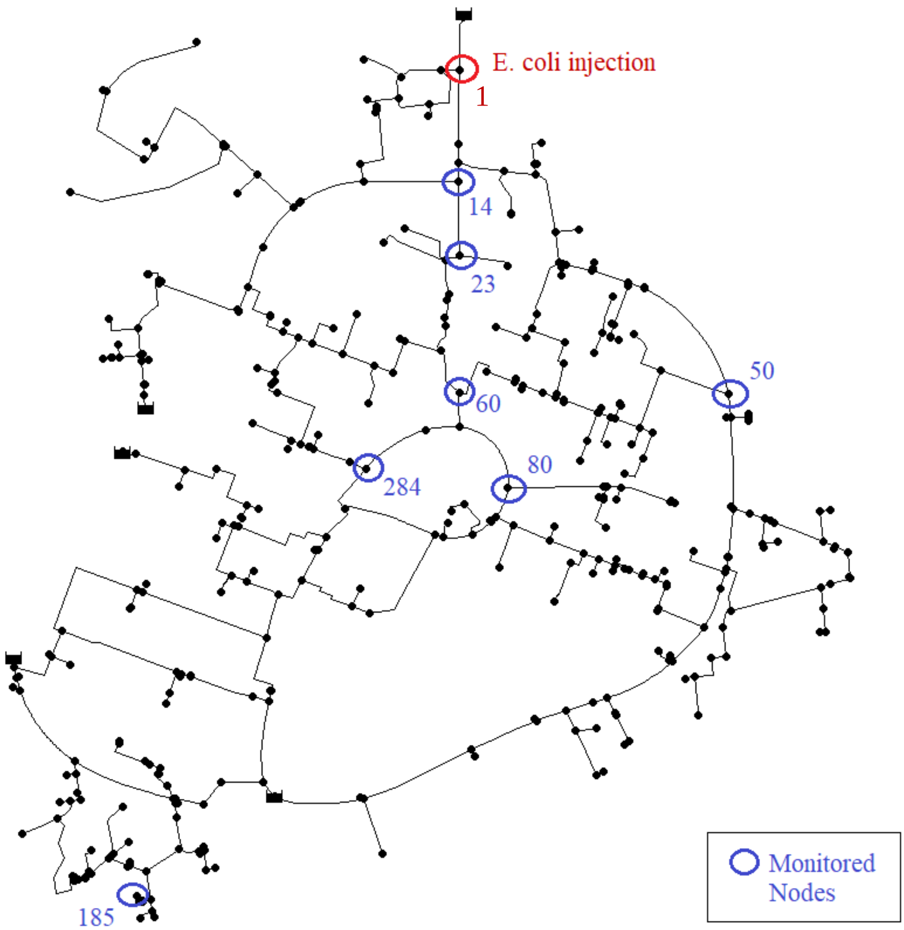
After selecting the species to be injected into the network (*E. coli* and chlorine), the knowledge of the hydraulic system and the solution of the advection/reaction equations, which can be carried out through such software as EPANET, determine the contaminants transport and fate.

Initially, the system water demand was assumed to vary with hourly steps. Therefore, 1 day long patterns were used for the demand multiplying coefficients and the simulations were run for 24 hours with the EPANET software. All calculations were performed setting the traditional SI units. However, this setting did not enable the understanding of the exact concentration of the pollutants in the network as a result of the complex reactions between the injected chemical/biological species and the reactions related to the liquid mass in contact with the pipe walls. For this purpose, the system water demand was assumed to vary with a 10 minutes step.

The EPANET-MSX multi-species model was used for running the simulations: as said, it provides the evaluation of the chlorine and the *E. coli*.

5.4.2 Results and Discussion

In a first stage of the analysis, two scenarios have been analyzed in order to test the transportation of the contaminants in the network. 10^8 CFU (as units diluted in 1 mL) of *E. coli* was injected in the nodes adjacent to the two tanks represented in Figure 5-9, that is R-1-CITE' Tank and R-2-4CANTON Tank. An initial chlorine concentration equal to 0.5 mg/L was assumed in all the nodes. For the sake of brevity, only the analysis of the injections close to the first tank, that is the injection in node 1, is reported (refer to Tinelli and Juran, 2017 for the full analysis) .
Figure 5-10 visualizes the injection location together with the monitored nodes and shows the results of *E. coli* and chlorine trends.



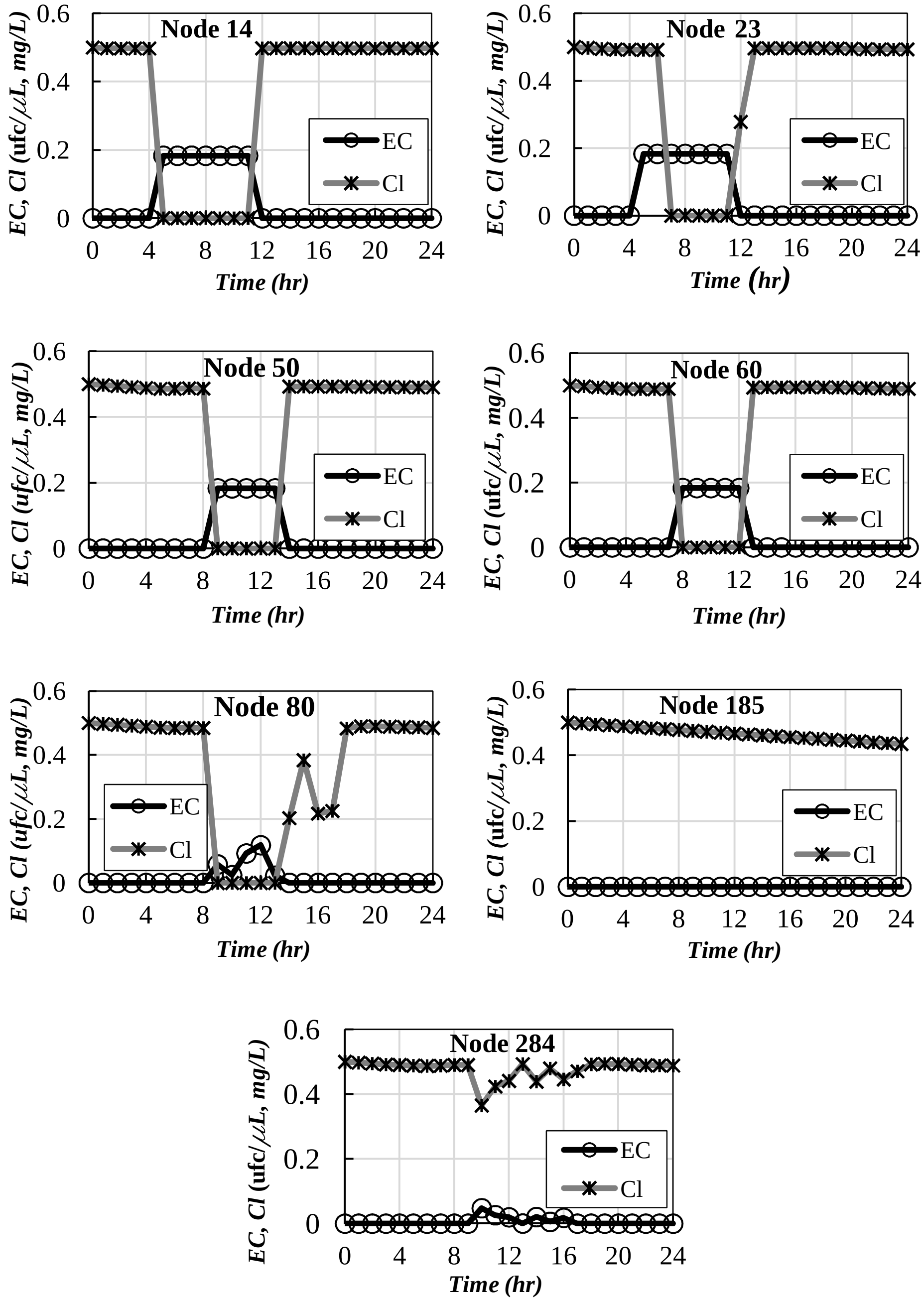


Figure 5-10. *E. coli* injection in node 1 with *E. coli* and chlorine concentrations in the monitored nodes

In details, Figure 5-10 illustrates the chlorine decomposition, following the injections of *E. coli* at a single point.
As long as the distance from the injection location increases, the contaminant effect on the chlorine is less evident. Starting from node 14, it demonstrates an immediate contaminant effect on the chlorine. Then, nodes 23, 50, and 60 illustrate a delay in the chlorine decrease, while node 80 shows a drop in the chlorine even if the *E. coli* does not reach its peak. Only a small quantity of contaminant with an apparent decrease in the chlorine is reported in node 284. Ultimately, node 185 is too far from the injection location to be influenced; here, the chlorine decrease is only due to the chlorine demand[4].
Therefore, the analysis showed a clear possibility to detect *E. coli* bacteria by analyzing the level of the chlorine in the network. In fact, the *E. coli* presents an immediate effect on the network: if *E. coli* quantity increases, the level of total chlorine directly decreases.

In a second stage of the analysis, due to the absence of a pumping system capable of injecting chlorine in the distribution network, an initial chlorine concentration of 1 mg/L was assumed. Some of the monitored nodes were analyzed before and after *E. coli* injections in order to (i) illustrate the chlorine trend without any contaminant injections (ii) corroborate the sudden reaction between chlorine and *E. coli*.

[4] Chlorine demand is the difference between total chlorine added in the water and residual chlorine. It is the amount which reacts with organic materials and other compounds present in water prior to disinfection.

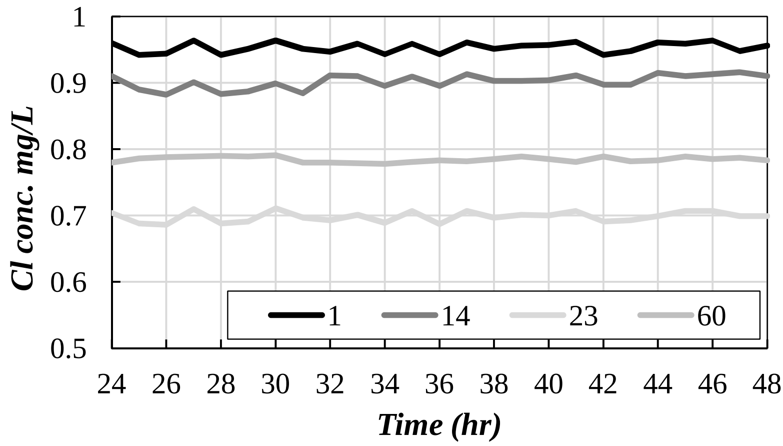


Figure 5-11. Chlorine concentrations in some of the monitored nodes after the stabilization

As expected (Hallama et al., 2002), after an initial drop due to the chlorine reactions with organic materials, tank walls and substances present in the water (e.g., metals), the chlorine reaches its stabilization. Thus, the remaining chlorine is the total chlorine divided into: i) the amount of chlorine that has reacted with nitrates and is unavailable for disinfection which is called combined chlorine and, ii) the free chlorine, which is the chlorine available to inactivate disease-causing organisms, and used as a measure to determine the drinkability of water.

Hence, Figure 5-11 shows the chlorine trend after the stabilization. As for nodes 1, 14 and 23, as long as the distance from the tank increases, the chlorine concentrations decrease due to its reactions and its dilution. The chlorine trends are also very similar, showing the slight daily variation.

The furthest node, that is node 60, reveals a concentration higher than node 23, probably due to the change in dilution phenomena (also related to the flow rates).

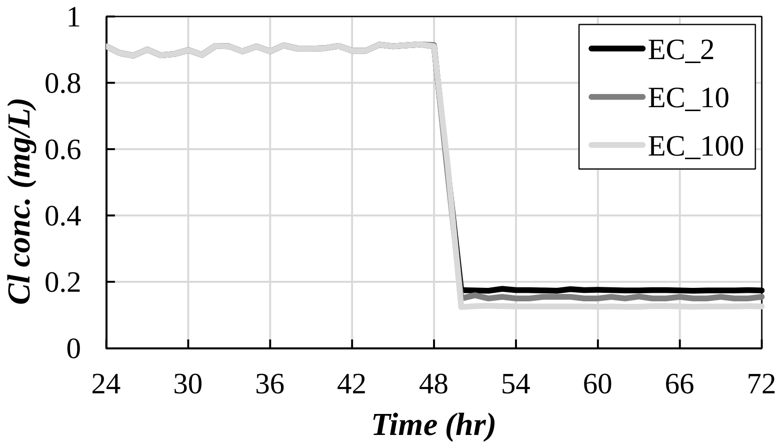


Figure 5-12. Chlorine concentrations on node 14 after different *E. coli* injections

EPANET-MSX simulations of *E. coli* injections were repeated. The bacterium was injected in node 1 (see Figure 5-10) at the 48th hour with different concentrations: 2, 5 and 100 CFU/μL.

Figure 5-12 shows the chlorine trend in the presence of the three *E. coli* concentrations for node 14. It demonstrates that the chlorine drops down in the presence of the bacterium, without reaching the zero. The results are consistent with the laboratory experiments and their simulations, proving that (i) the chlorine, which is one of the most common disinfectants, is consumed when it reacts with pathogens (ii) the chlorine drops down to zero when the *E. coli* injections reach the order of magnitude of 10^8 CFU/mL.

Chapter 6

Prototype System for Early non-specific Biological Contaminations Detection

After studying the sensitivity of the quality sensors to detect biological accidental or deliberate contaminations in WDSs, as the quality sensors produce a huge amount of data, a rapid, as well as smart data processing procedure has to be established for the development of a risk assessment model, together with a decision support system. In details, the model should be able to early detect biological contaminations in drinking water network in order to efficiently enable water operators to ensure real-time water quality control management.

Consequently, this procedure should consist of pre-processing/processing, training, validation, and forecasting phases, filtering out data anomalies and false alarms.

In addition, Chapter 5 discussed the feasibility of detecting non-specific biological anomalies (such as *E. coli*) through the use of the chlorine trend to develop a prototype system for early non specific bio-contamination detection. For this purpose, the numerical simulations of the chlorine decay trend during injection of *E. coli* were compared with the laboratory model test results performed at the University of Lille.

In the absence of field data of bio-contamination, the EPANET-MSX model was used for scenario-simulations to produce numerical data, hereafter named Chlorscan: they simulate the effect of bio-anomaly scenarios on chlorine concentration in water distribution networks.

Therefore, this chapter first presents an automated statistical based approach to detect bio-anomalies in a generic WDS (drawn in Tinelli, Juran & Cantos, 2017), starting from a database, which consists of Chlorscan time series (section 6.1). The same procedure becomes more accurate in the rest of the chapter (section 6.2), where the statistical data analysis approach is supported by the use of the Artificial Intelligence (AI). In fact, by exploiting some principles such as the pattern recognition, the detection process is able to recognize a contamination scenario in quasi-real-time. To do this, not only chlorine but also different water quality parameters are analyzed so that the bio-contamination signature, its likelihood and severity can be detected when a contamination event occurs.

6.1 Statistical based Early Detection

6.1.1 Chlorscan Data Analysis

Typical Chlorscan data are considered as relevant indicators of potential non-specific bio-contamination, thus they can be used as input for a risk assessment system in WDSs. They require a first processing step, which consists in identifying the chlorine trend after its stabilization, as shown in Figure 5.11. Taking as input the numerically simulated Chlorscan data, which are aggregated into a continuously updated data file, false alarms due to the initial data temporal variability are filtered out.
A multi-spots approach is used to compare the anomalies detected by the Chlorscan at different locations over the water distribution network.

Starting from a database, which consists of Chlorscan time series, statistical tests are implemented to establish the 1st, 2nd, and 3rd Standard Deviation (STDs). Afterwards, the analysis requires the following steps:

- Chlorscan data are normalized to the average (F) to filter out any outlier that might be generated;
- 5 threshold levels (Insignificant, Low, Moderate, High, Very High) of normalized Chlorscan data are defined to establish likelihood and risk severity levels corresponding to the amplitude of normalized concentration deviation from the average (ΔF);
- The Likelihood is defined as a function of the ΔF amplitude and the elapsed time period (ΔT in Hours) of the detected anomaly;

- Using the selected thresholds of the state parameters (ΔF; ΔT) the likelihood matrix is established, as shown in Figure 6-1a;
- The risk severity matrix is obtained, using a similar process, considering respectively the normalized Chlorscan concentration deviation data (ΔF) and the exposure period (ΔT_{ex} in hours). Therefore, using the selected thresholds of the state parameters (ΔF; ΔT_{ex}) the severity matrix is established, as shown in Figure 6-1b.

a)

Likelihood Matrix					
	Hours				
ΔF%	1	2	3	4	>4
0-4%					
4-10%					
10-20%					
20-30%					
>30%					

Likelihood Scale (0-100%)	
0 – 10 %	Insignificant
10- 30 %	Low
30-60 %	Moderate
60-90%	High
>90	Very High

b)

Severity Matrix					
	Exposure Hours				
ΔF%	0-1	1-3	3-12	12-24	>24
0-4%					
4-10%					
10-20%					
20-30%					
>30%					

Severity Scale (1-5)	
1	Insignificant
2	Low
3	Moderate
4	High
5	Very High

Figure 6-1. a) Likelihood matrix and Likelihood scale; b) Severity matrix and Severity scale

- Using the likelihood scale and the severity scale the risk matrix is defined as shown in Figure 6-2.

Risk Assessment Matrix					
Likelihood Scale	**Severity Scale**				
	1	2	3	4	5
0-10%					
10-30%					
30-60%					
60-90%					
>90%					

Risk Scale (0-1)	
0-0.1	Insignificant
0.1-0.2	Low
0.3-0.6	Moderate
0.6-0.9	High
0.9-1	Very High

Figure 6-2. Risk scale and Risk assessment matrix

- Finally, using the risk scale, the time series of the Risk Indicator is defined based on the appropriate state color of the time step.

6.1.2 Application of Chlorscan Analysis

The methodology was deployed in the Lille University Campus, already explained in Chapter 5 (section 5.4).

In order to demo-illustrate the application of the proposal analysis, numerical Chlorscan data simulating the effect of bio-anomaly scenario on chlorine concentration in two nodes (nodes 23 and 80) of the network were used.

In particular, they were obtained from 8 hours EPANET-MSX simulations, using a reporting time step of 10 minutes. An initial chlorine concentration equal to 0.4 mg/L along with an *E. coli* injection of 10^5 CFU/mL were assumed.

Thus, the data were classified as follows:

- Chlorscan1: node 23;
- Chlorscan2: node 80.

The numerical Chlorscan data are reported from the 24th hour of the analysis to avoid the initial variability of the chlorine, as shown for the previous simulations (Figure 5.11).

Table 6.1 illustrates the numerical Chlorscan data that were employed in the analysis: each Chlorscan value is reported with reference to the sequential number of the time series to be easily represented in the procedure application.

Table 6-1. Chlorscan data used in the statistical based analysis.

Seq. N. of Times Series	Chlorscan 1 - Node 23		Chlorscan 2 - Node 80	
	Time	Chlor. (mg/l)	Time	Chlor. (mg/l)
1	24:00:00'	0.164	24:00:00'	0.182
2	24:10:00'	0.164	24:10:00'	0.182
3	24:20:00'	0.163	24:20:00'	0.182
4	24:30:00'	0.163	24:30:00'	0.183
5	24:40:00'	0.166	24:40:00'	0.183
6	24:50:00'	0.166	24:50:00'	0.183
7	25:00:00'	0.162	25:00:00'	0.183
8	25:10:00'	0.161	25:10:00'	0.184
9	25:20:00'	0.154	25:20:00'	0.184
10	25:30:00'	0.157	25:30:00'	0.184
11	25:40:00'	0.152	25:40:00'	0.184
12	25:50:00'	0.151	25:50:00'	0.185
13	26:00:00'	0.158	26:00:00'	0.185
14	26:10:00'	0.154	26:10:00'	0.185
15	26:20:00'	0.157	26:20:00'	0.185
16	26:30:00'	0.153	26:30:00'	0.185
17	26:40:00'	0.151	26:40:00'	0.186
18	26:50:00'	0.145	26:50:00'	0.186
19	27:00:00'	0.146	27:00:00'	0.185
20	27:10:00'	0.125	27:10:00'	0.185
21	27:20:00'	0.093	27:20:00'	0.187
22	27:30:00'	0.080	27:30:00'	0.187
23	27:40:00'	0.075	27:40:00'	0.186
24	27:50:00'	0.072	27:50:00'	0.141
25	28:00:00'	0.074	28:00:00'	0.123
26	28:10:00'	0.078	28:10:00'	0.118
27	28:20:00'	0.068	28:20:00'	0.122
28	28:30:00'	0.072	28:30:00'	0.118
29	28:40:00'	0.075	28:40:00'	0.119
30	28:50:00'	0.078	28:50:00'	0.119
31	29:00:00'	0.069	29:00:00'	0.141
32	29:10:00'	0.156	29:10:00'	0.123
33	29:20:00'	0.162	29:20:00'	0.118
34	29:30:00'	0.163	29:30:00'	0.122
35	29:40:00'	0.159	29:40:00'	0.118
36	29:50:00'	0.159	29:50:00'	0.119
37	30:00:00'	0.160	30:00:00'	0.119
38	30:10:00'	0.153	30:10:00'	0.121
39	30:20:00'	0.161	30:20:00'	0.121
40	30:30:00'	0.161	30:30:00'	0.12
41	30:40:00'	0.161	30:40:00'	0.12
42	30:50:00'	0.162	30:50:00'	0.119
43	31:00:00'	0.163	31:00:00'	0.119
44	31:10:00'	0.162	31:10:00'	0.12
45	31:20:00'	0.162	31:20:00'	0.12
46	31:30:00'	0.160	31:30:00'	0.119
47	31:40:00'	0.161	31:40:00'	175
48	31:50:00'	0.161	31:50:00'	179
49	32:00:00'	0.161	32:00:00'	0.184
50	32:10:00'	0.160	32:10:00'	0.183

For the purpose of showing the proposed statistical procedure, the first time series, that is Chlorscan1is used.

Therefore, Figures 6-3, and 6-4 illustrate the Chlorscan data analysis, as follows:

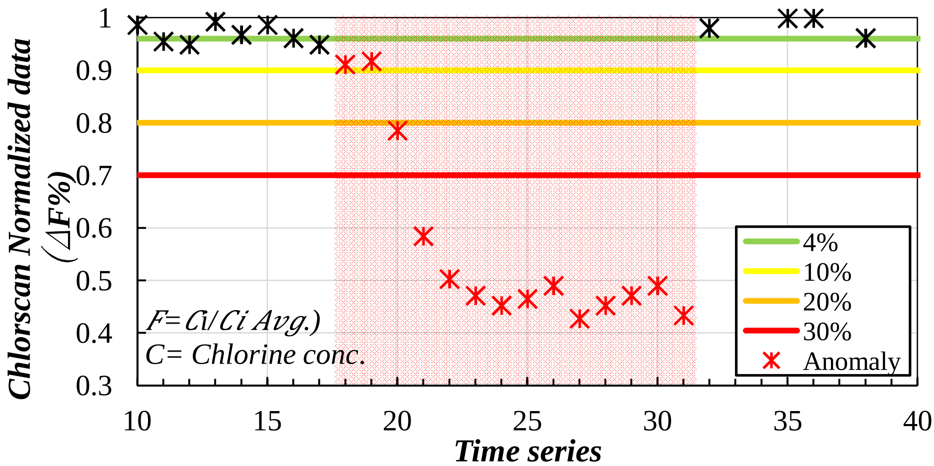


Figure 6-3. Normalization of Chlorscan data to the average (F) and Contamination Likelihood Assessment

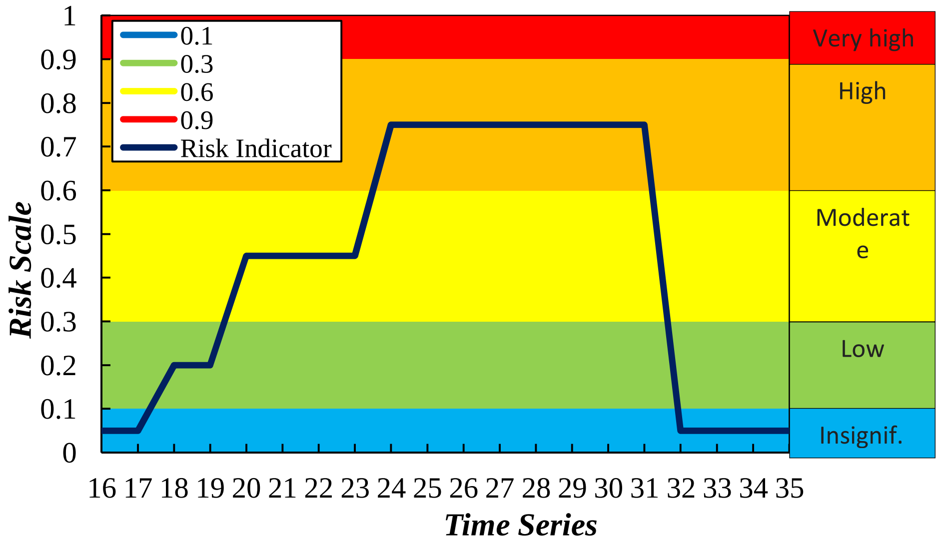


Figure 6-4. Risk Indicator with its scale for Lille Case Study

In details, Figure 6-3 shows the numerical trend of the normalized data in percentage along with time.
In particular, consistently with the results of the laboratory test and the numerical simulations presented in Chapter 5, an initially increasing anomaly is detected. It tends to be constant once the peak is reached, with a residual of about 40% of the initial default value without any bio-anomaly. The statistical tools of the 1^{st}, 2^{nd}, and 3^{rd} standard deviations were used as default values for identifying the thresholds for the likelihood and the severity levels, reaching a high orange color-coded likelihood on the likelihood scale with ΔF greater than 30%.
Operators can input threshold levels based on their experience.
Figure 6-4 illustrates the time series of the risk indicator with its scale, taking into account the likelihood and the severity scale. It is evident that the risk indicator indicates a high orange color-coded risk level and an alarm should be emitted for the water utility operators in order to support the decisions makers in their resolution for the public community.

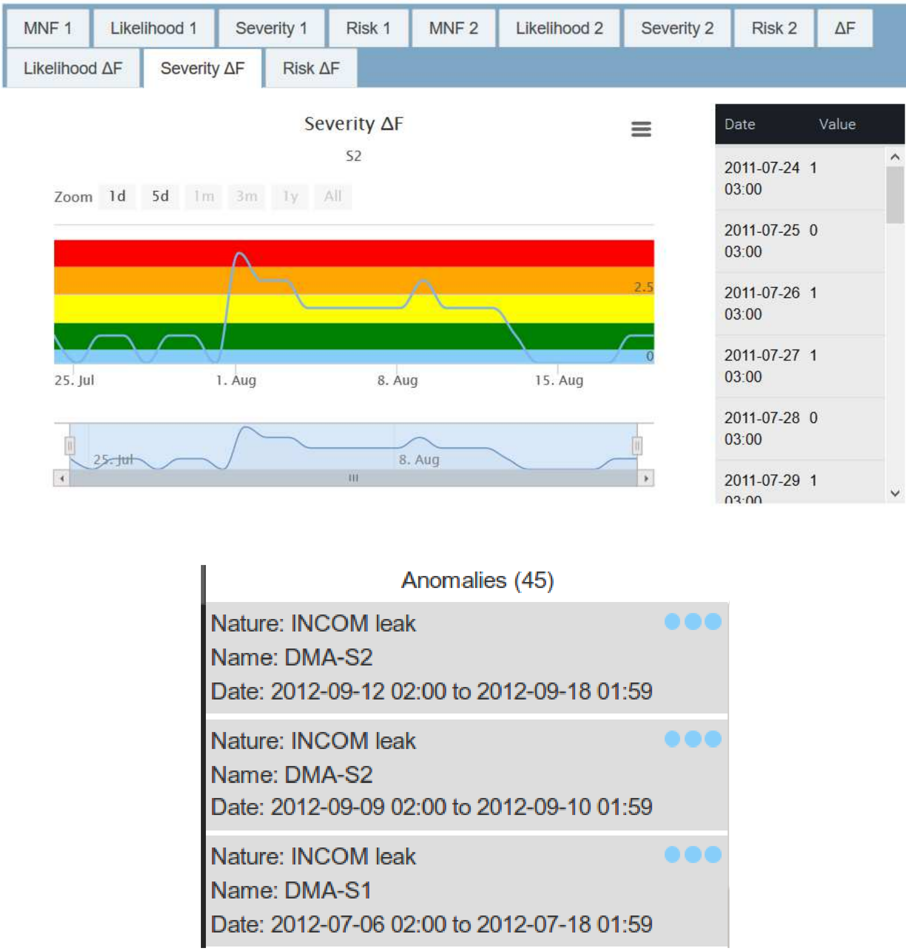
Within the framework of the SW4EU project, the W-SMART Association[5] acting as an integrator for the research conducted in several universities including University of Lille, NYU, and the University of Pavia has engaged the collaboration with the French University ESIEA ("Ecole D'Ingénieurs Du Monde Numérique" located in Paris) for the development of the Bio-CON Prototype System and its support software.
The software is able to run scenarios that are loaded as Excel data and it quickly provides the risk matrix to detect anomalies in the data. The analyzed WDS is also displayed in the software through the user accessibility and the Google Earth visualization in such a way that the detected anomalies can be geo-located on the map, according to the color-based procedure. In fact nodes can remain grey if no anomaly occurs or they can appear from blue to red as a function of the anomaly severity, as shown in Figure 6-5.

[5] *W-Smart Association* is an International Association of Water & Wastewater Utilities for Sustainable Water Security located in France.

Figure 6-5. Color-based detection system

The alarm system is finally based upon the identified risk matrix: the levels of likelihood-severity-risk are plot on graphs, and an alarm panel displays each anomalies with its related day, type and location.

Figure 6-6a illustrates how the user can display the likelihood-severity-risk graph for each uploaded dataset, while Figure 6-6b indicates an example of the final alarm panel with the detected anomalies.



a)

b)

Figure 6-6. An example of a) likelihood-severity-risk graphs for uploaded dataset b) alarm panel with the detected anomalies

6.2 Statistical based Early Detection

6.2.1 *Multi-parameters AI-based Analysis*

Trying to improve the accuracy and the efficiency of the analysis, in a second stage the research dismissed the statistical approach and it developed, adapted and demonstrated the feasibility of an Artificial Intelligence (AI) based smart system to ensure quasi real-time quality control for early chemical and/or bio-contamination detection. Instead of monitoring only the chlorine, the methodology became multi-parameters in order to have a clearer fingerprint of the contamination scenarios, reducing the error detection.

In this context, advance pattern recognizers, such as Support Vector Machines (SVMs), and innovative sensing technology solutions, as Artificial Neural Network (ANN), were used (Girolami, 2002; Smola, 2004).

In machine learning the SVMs are supervised learning models (Cortes and Vapnik, 1995) with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one out of two categories, a SVM training algorithm builds a model that assigns new examples to one category or the other, becoming a non-probabilistic binary linear classifier (Cui and Wang, 2010).

Hence, the designed SVM identified two different classes, that is (+1) or (0) respectively for "anomaly" or "no-anomaly" classification. However, due to the fact that a single SVM only resolves two classifier problems, the research set up an SVM model composed by several classifiers to distinguish different states of anomaly levels (Mamo et al., 2014). Thus, diverse signatures were defined in such a way that the proposed SVM could classify the incoming unknown data as belonging to one specific signature, based on the anomaly severity.

The definition of each of the anomaly signature started according to the already explained matrix: the 1st, the 2nd and the 3rd STDs[6] of the input data were used to distinguish the values of the Amplitude (A), while the time identified the different values of the Duration (D). In particular, the Amplitude has three levels, that is A1 included in the range between the 1st and the 2nd STD, A2 between the 2nd and the 3rd STD, and A3 above the 3rd STD. Durations are simply divided in 4 groups, as a function of the duration in hours.

	D (hours)			
A	1	2	3	4
A1 (1std-2std)				
A2 (2std-3std)				
A3 (>3 std)				

Figure 6-7. Matrix used for the definition of the anomaly signature

[6] Starting from a database, which consists of multi-parameters time series (such as Chlorine, TOC, pH etc.), the 1st, the 2nd and the 3rd STDs were calculated to find the data deviation from the normal conditions (without any contamination).

Figure 6-7 illustrates the definition of the anomaly signatures, according to the Amplitude and Duration levels.
Using the classification of the signatures, the risk scale was defined as shown in Figure 6-8.

	>90%	A2D4-A3D3-A3D4
	60-90%	A1D4-A2D3-A3D2
	30-60%	A1D3-A2D2-A3D1
	10-30%	A1D2-A2D1
	0-10%	A1D1

Figure 6-8. Risk scale for the classification of the SVM output

Figure 6-8 shows how the risk of each signature is classified according to a color-based risk scale.
In details, to be able to classify new upcoming data from the sensors, the designed SVM had initially to be trained. To this purpose, the SVM was coded in Matlab, using functions already implemented, such as "svmstruct" and "svmtrain". In particular, "svmstruct" contains information about the trained SVM classifier, which is the actual data separator.
The used command line was *"SVMStruct= svmtrain (Training, Group)"*, where the *Training* was the matrix of training data and the *Group* is the grouping variable (numeric, logical vector or matrix) representing a class label.
According to the SVMs theory, each training data item is plotted as a point in *n*-dimensional space, with *n* equal to the number of available features. Then, the "svmtrain" uses an optimization method to define an hyperplane, which linearly separates the *n*-dimensional data into two classes, being a discriminative classifier. The optimal hyperplane maximizes the distances between nearest data point: the distance is called *"Margin"*, and the selection of the hyperplane with the higher margin demonstrates the best robustness.

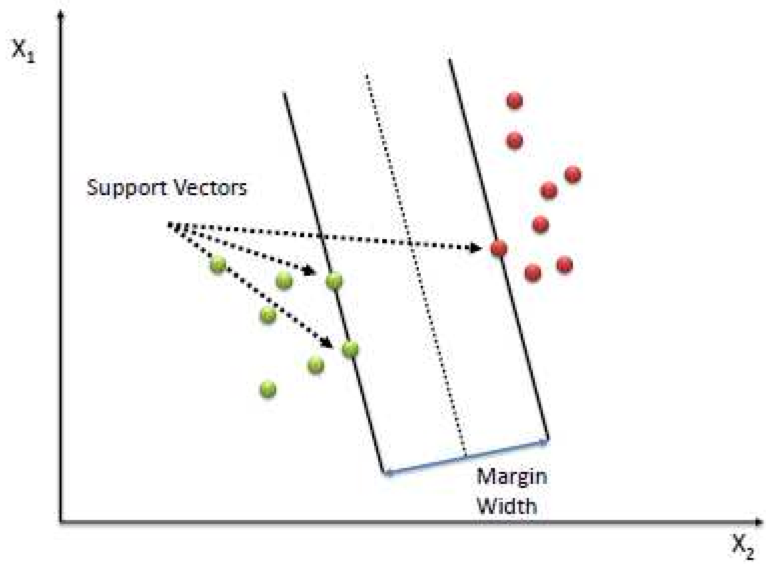


Figure 6-9. Maximum-margin hyperplane for an SVM trained with samples from two classes.

Figure 6.9 shows the optimal hyperplane which is used for the classification, after being trained from two data categories: samples on the margin are called the Support Vectors because they actually define the hyperplane.

Sometimes data cannot be linearly separable thus, SVMs introduce the notion of a "Kernel induced feature space", which casts data into a higher dimensional space, where data are separable.

In this case as explained below, the Kernel function was considered linear (Boswell, 2002).

Therefore, the required inputs were:

- X - Matrix of predictor data, where each row was one monitored node of the network (thus, rows were equal to the number of nodes), and each column was one parameter (e.g., Chlroine, TOC etc.);
- Y - Array of class labels with each row corresponding to the value of the corresponding row in X. Y was indeed a column vector, whose values were "+1" or "0", according to the belonged category of "anomaly" or "no-anomaly";
- Kernel Function - The default value was 'linear' for two-classes learning, which separates data by a hyperplane;
- Class Names - It distinguished between the negative and positive class, or specified which classes to be included in the data. The "0 class" was here

the "negClass" (no-anomaly) class, while the "+1" was the "posClass" (anomaly) class.

The resulting trained model ("svmmodel" in Matlab) contained the optimized parameters from the SVM algorithm, enabling the classification of the new data as "anomaly" or "no anomaly". Once the model was trained, it became capable of predicting the specific signature of the new upcoming data from the sensors.
Since the proposed methodology investigated the physical/chemical water parameters in every single node, it was able to identify the anomalies at each single node. In particular, the identification of the node anomaly signatures was made based on the assigned input and the upcoming data from the sensors. Therefore, the i_{th} SVM was design to recognize one of the presented anomaly signature, providing as output the stated "+1, or positive" if the contamination was actually present in the specific node, "+0, or negative" otherwise.
Figure 6-10 illustrates the steps of the Multi-class SVM anomaly detector: if an anomaly is detected, the procedure continues with the identification of the severity level, according to the defined signatures.

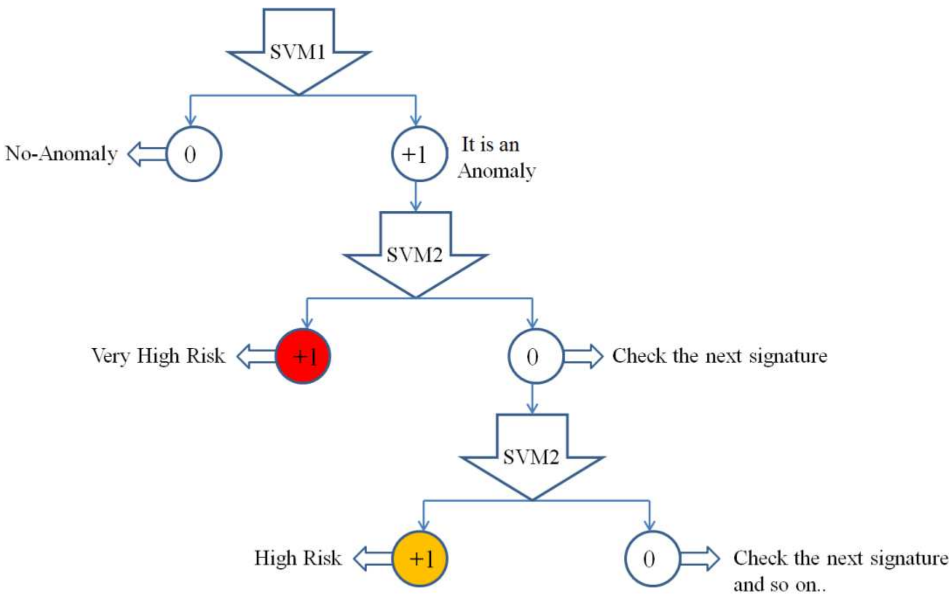


Figure 6-10. Scheme of the Multi-class SVM anomaly detector

In order to have a comparison for the obtained results, the ANN was applied (Kohavi and Provost, 1998). Like the SVM, the ANNs are used to solve a wide variety of tasks, as pattern recognition. The ANN recognition problem is being posed as a classification task, where the classes are defined by the system designer. The approach of the ANN learns from a set of examples (training set) and adapts themselves to the data.

To define the pattern recognition problem, functions already implemented in Matlab were used; particularly, the *Neural Network Pattern Recognition Tool* was applied. It is an information processing that is inspired by the biological nervous system: it is composed of a large number of highly interconnected processing elements (neurons) working for pattern recognition, through a learning process (Basu et al., 2010). Neurons are organized in layers so that signals can travel from the first (input), to the last (output), passing the information and adjusting the network to reflect that information.

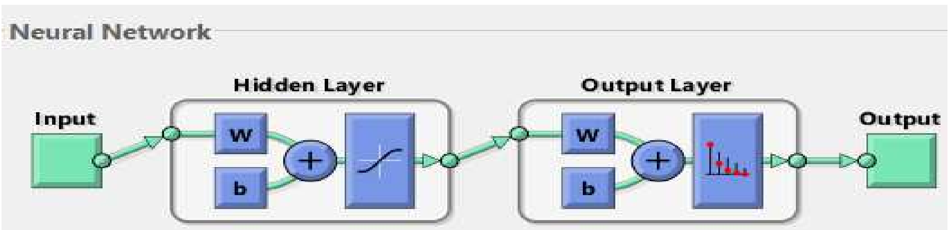


Figure 6-11. Structure of Neural Network Pattern Recognition in Matlab

Figure 6-11 illustrates the structure of the Neural Network Pattern Recognition implemented in Matlab: two-layers feed-forward network with output neurons can classify vectors arbitrarily well, given enough neurons in its hidden layer.

Therefore, regarding the inputs and targets for the ANN pattern recognition problem, a set of vectors was required as columns in a matrix. Then, another set of target vectors was needed, indicating the classes which the input vectors were assigned to. In details:

- X - Matrix of predictor data, where each row was one parameter (e.g., Chlroine, TOC etc.) and each column was one monitored node of the network (thus, rows were equal to the number of nodes);
- Y - Array of class labels: when there were only two classes, each scalar target value was set to either "0" or "+1", indicating which class the corresponding input belonged to. As in the SVM, the values were "+1" or "0", according to the category of "anomaly" or "no-anomaly".

Once the input was defined, the pattern recognition tool was able to (i) train the network (ii) evaluate its performance using Cross-Entropy and percent misclassification error, and (iii) analyze the results using visualization tools, such as Confusion Matrices and Receiver Operating Characteristic curves (ROC curve).

6.2.2 Application of AI-based Analysis

The data analysis was deployed in the Lille University Campus, already explained in Chapter 5 (section 5.4).
Using the illustrated Chlorscan data and the data obtained during the testing of the numerical analysis described in the previous chapter (section 5.4), a further database was built: TOC was calculated, starting from chlorine and according to the modeling of chlorine residuals in WDSs. In fact, Ahn et al. (2012) proposed a mathematical model of chlorine bulk decay based on multiple regression analysis. The dependent variables were the initial chlorine concentration, temperature and TOC; the independent and dependent variables were first formed into natural logarithms, and then the coefficients of the dependent variables were identified by multiple regression analysis. Consequently, TOC was evaluated using the Equations 6.1 and 6.2:

$$C = C_0 \exp\left(-K_b t\right), \tag{6.1}$$
$$K_b = 0.0488 \cdot T^{0.3668} \cdot C_0^{-1.3268} \cdot TOC^{0.5979}, \tag{6.2}$$

where, C is the chlorine concentration at time *t* (mg L^{-1}), C_0 is the initial chlorine concentration (mg L^{-1}), K_b is the chlorine bulk decay coefficient (d^{-1}) calculated according to the first order model, T is the temperature (°C), and TOC is the TOC concentration.
Chlorine and TOC were thus used as input data to run multi-variables analysis for specific bio-contaminations, that is *E. coli* injections at a mono-spot and/or multi-spots.

Regarding the SVM, a matrix of predictor data for mono-spot analysis (chlorine and TOC) was input, as reported for example in Figure 6-12a.
The array of class labels was defined according to the statistical process: for each row of the matrix, the proposed model included the evaluation of chlorine and TOC in order to classify the row into one of the two different classes, that is "+1" or "0" respectively for "anomaly" or "no-anomaly" classification. As

described, the recognition of each of the anomaly signature started according to Amplitude (A) and the Duration (D). Thus, the array of class labels was a vector made up of "+1" or "0", as shown in figure 6-12b.
Figures 12a and 12b report an example of the SVM coding input: starting from the chlorine concentrations obtained in section 5.4 for the node 14 (shown in Figure 5.12), the TOC was derived along with the array of the two classes.

Time	Chlor. (mg/l	TOC (mg/l C)
24:00:00'	0.91	0.094442518
25:00:00'	0.89	0.13454182
26:00:00'	0.882	0.152427912
27:00:00'	0.901	0.11167289
28:00:00'	0.883	0.150134206
29:00:00'	0.887	0.141125045
30:00:00'	0.899	0.115682715
31:00:00'	0.884	0.147857078
32:00:00'	0.911	0.092610237
33:00:00'	0.91	0.094442518
34:00:00'	0.895	0.123899808
35:00:00'	0.909	0.096291244
36:00:00'	0.895	0.123899808
37:00:00'	0.913	0.088995025
38:00:00'	0.903	0.107728842
39:00:00'	0.903	0.107728842
40:00:00'	0.904	0.105781479
41:00:00'	0.911	0.092610237
42:00:00'	0.897	0.119758343
43:00:00'	0.897	0.119758343
44:00:00'	0.915	0.085445643
45:00:00'	0.91	0.094442518
46:00:00'	0.913	0.088995025
47:00:00'	0.916	0.083695653
48:00:00'	0.91	0.094442518
49:00:00'	0.559	1.979576277
50:00:00'	0.1237	16.81320793
5100:00'	0.1267	16.49202293
52:00:00'	0.1277	16.38719236
53:00:00'	0.1267	16.49202293
54:00:00'	0.1257	16.59795599
55:00:00'	0.1257	16.59795599
56:00:00'	0.1257	16.59795599
57:00:00'	0.1257	16.59795599
58:00:00'	0.126	16.56605916
59:00:00'	0.126	16.56605916
60:00:00'	0.1253	16.64064226
61:00:00'	0.1257	16.59795599
62:00:00'	0.1254	16.6299538
63:00:00'	0.1254	16.6299538
64:00:00'	0.1267	16.49202293
65:00:00'	0.1267	16.49202293
66:00:00'	0.1264	16.52368621
67:00:00'	0.1254	16.6299538
68:00:00'	0.1257	16.59795599
69:00:00'	0.1261	16.55544923
70:00:00'	0.1257	16.59795599
7100:00'	0.1267	16.49202293
72:00:00'	0.1257	16.59795599

a)

Class
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
0
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1

b)

Figure 6-12. Coding Input for a) Chl/TOC SVM data b) Array of class labels

Matlab functions were used to train and cross-validates the SVM model for the two-classes (binary) classification.

After importing the datasets (either the Chlorscan time series and the data obtained in section 5.4) into the code, data were randomly divided, assigning 70% of the database for the training phase, and 30% for the testing phase.
Hence, the results displayed the properties of SVM model, including:

i. the class order which was "+1" for the negative class, and "0" for the positive class;
ii. the SVM classifier that was the radial basis kernel in this case;
iii. the testing phase, that estimated the out-of-sample misclassification rate;
iv. the "Class Loss" (so-called in the Matlab functions), that is the classification rate, was approximately $4^{0}/_{00}$.

Finally, the main output of the research was the visualization of the contaminated nodes in the network, according to the color-based analysis, as shown in Figure 6-13.

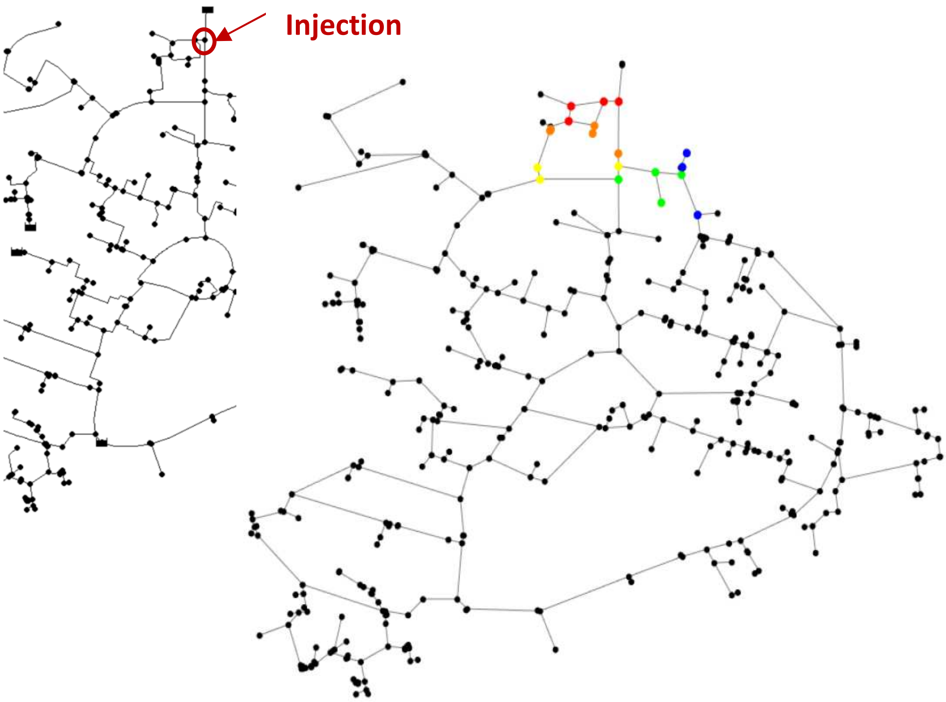
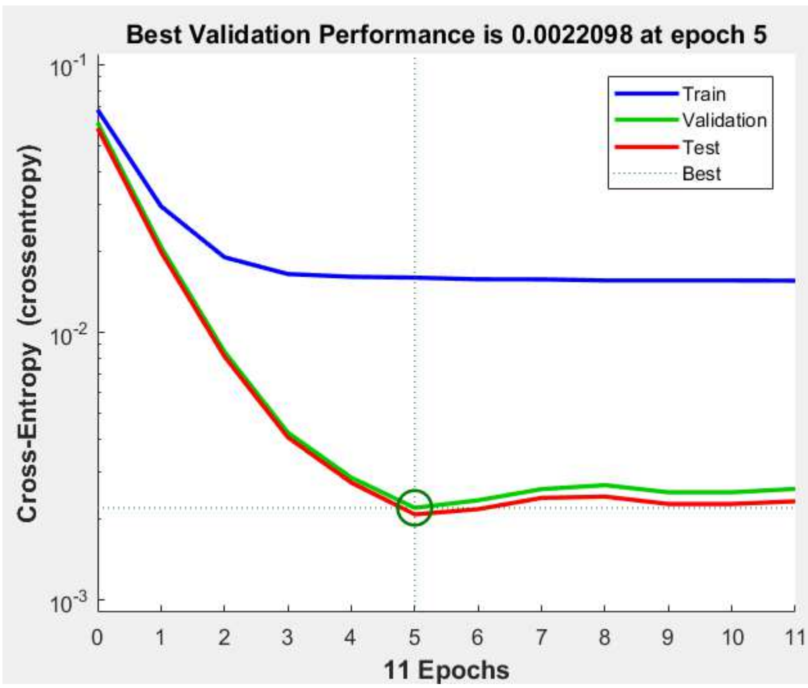


Figure 6-13. Color-based approach visualization for mono-spot analysis

The same was done applying the ANN in order to compare the obtained results. To define a pattern recognition problem, Chlorine and TOC were input as rows in a matrix. Then, another set of row vectors were arranged to indicate the classes to which the input vectors were assigned (ANN required the same SVM input but the matrices were transposed).

During the coding phase, the *Neural Network Pattern Recognition Tool* classified inputs into a set of target categories: it randomly divided up the samples into training (70%), validation (15%), and testing (15%). The latter two options were assumed following the default options. Also, in the phase of building the network architecture, the number of neurons was assumed following the defaults options (10) because the network performed well after training.

Hence, the *Neural Pattern Recognition* application created, trained a network, and evaluated its performance using Cross-Entropy error and Confusion Matrix. In particular, the Cross-Entropy error defines the error in classification problem, while the Confusion Matrix is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives. These allow detailed analysis in terms of accuracy.



a)

b)

Figure 6-14. ANN Results a) Algorithm performance b) Confusion Matrix

Figure 6-14 shows the best validation performance of the artificial network (6-14a) and the Confusion Matrix applied to the current analysis (6-14b). Error was approximately $2^{0}/_{00}$, confirming the efficiency and the accuracy of the described methodology. In fact, minimizing Cross-Entropy (as shown in Figure 6-14a) the classification results are accurate because it means no error and a value of zero error means no misclassifications in the incoming data. This is reflected in the Confusion Matrix in which the false negative, as well as the false positive are zero.

Finally, the difference between the SVM and the ANN was 0.0018, approximately $2^{0}/_{00}$.

The same process was done for multi-injections analysis. Regarding the SVM, matrix of predictor data and the array of class labels for multi-spots analysis were the same but they showed multiple changes in the values of Chlorine and TOC, according to the injections.
Figure 6-15 shows the multi-injections analysis carried out at the Lille Campus and characterized by contamination injection at nodes 1 and 204. In details, it displays the contaminated nodes according to the color-based procedure, and it indicates the difference with the respect to the mono-spot analysis.

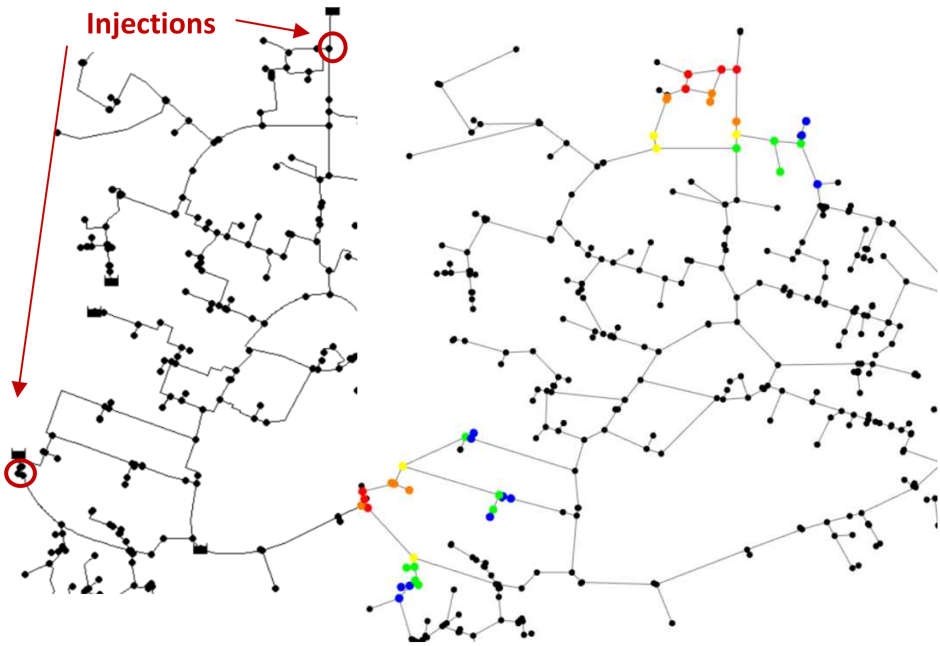


Figure 6-15. Color-based approach visualization for multi-injections analysis

Considering the multi-parameters analysis, an efficient and reliable bio-anomaly detection method based on AI simulations was developed and the final output supports the anomaly visualization, both temporal and spatial. In fact, the Chlorine and the TOC were input as time series in which every value corresponded to the predefined time step (hourly pattern time step in this case). Therefore, the tested advanced pattern recognizers demonstrate an improvement of data control in order to sustain water utilities with a secure decision support system.

Conclusions

Regarding the Water Safety Plans (WSPs), which have been introduced about a decade ago by the World Health Organization, the analysis of contamination risk in water distribution networks has taken particular importance.
In particular, since the contamination events which may potentially affect the distribution network are distributed throughout all the network and, so, they are hardly predictable and controllable, the increasingly focused attention on security issues has created a great interest in monitoring and Early Warning Systems (EWS) applied to water distribution networks.
For this reason, there are already research programs (such as the European project SW4EU, to which this Thesis has provided a scientific contribution) aimed at the realization of alarm systems based on appropriate sensors to be installed all over the network, capable of analyzing and interpreting the results in real time. Some of these sensors are already designed and marketed; in this Thesis, an updated state of the art is presented about these equipments. In particular, it is highlighted that nowadays instruments for real-time monitoring of water quality include either very simple and inexpensive systems suitable for the detection of the most common physical-chemical parameters (e.g.: pressure, pH, temperature, conductivity, chlorine, etc.), and also more complex systems, such as those based on UV spectrometry and on toxicity or biological contaminants assessment systems.
However, the state of the art highlights many still open problems.
Among these, it is particularly important to underscore the difficulty of sensor technology currently available to cope with the wide spectrum of substances that can potentially contaminate the delivered water.
The possible breakthroughs are the technological development that leads to the availability of less expensive and more polyvalent sensors on one hand, and the development of new interpretative models that allow detection of a large

spectrum of contaminations on the other hand, starting from a limited number of measured parameters.

In addition, once the sensors to be installed in the network are chosen in relation to the risks to be addressed, the problem of the definition of the sensors number and optimal locations arises. This is a typical multi-objective optimization problem, where the objective to be minimized is the number of sensors (directly associated to the cost of the system), and the objective to be maximized is the ability of the system to reduce the impact of contaminations on the public health. Since the solution of the optimization problem requires the definition of all the contamination events that may potentially affect the network, and due to the fact that each possible contamination event is characterized by certain values of injection location, starting time, mass rate and duration, the complexity of the algorithms for the design of monitoring systems causes many difficulties in large-scale network applications. Therefore, this Thesis proposes and tests a procedure based on practical considerations on network topology and operation for sampling the most representative contamination events. The procedure is applied to one case study. The results do not vary significantly when the sampled contamination events are used inside the optimization, instead of the totality of the possible contamination events.

In addition, the Thesis examines how the choice of the objective functions to be used in the optimization process affects the final results. To this end, two different variants of optimization were considered. Both variants feature the total number of sensors as first objective function to minimize, like a surrogate for the cost of the monitoring system. The two variants differ, instead, in the second objective function, which is the likelihood of contamination event detection (to be maximized) and the contaminated population (to minimized) for the former and latter variant, respectively.

The results of the optimizations, and the re-evaluations of the optimal solutions in terms of various effectiveness indicators for the water quality monitoring system, prove that the first variant tends to produce better solutions in terms of detection likelihood and sensor redundancy. The second, instead, tends to produce better solutions in terms of contaminated population and event detection time. However, all the effectiveness indicators are well intercorrelated in the solutions of the optimizations. The ultimate choice of water utility managers is based on their preferences. In fact, minimizing the contaminated population yields benefits in terms of detection time and thus mainly contributes to the system’s early warning capacity. On the other hand, maximizing the detection likelihood strongly impacts on the system redundancy and therefore contributes to the system safety. A further difference between the two variants of analyzed optimization lies in the placement of sensors in the network layout. In fact, whereas the first variant tends to locate the sensors in the area where most water paths converge, the second produces a more scattered distribution over the layout.

The subsequent part of the Thesis removed the assumption used in the first part of conservative contaminant.

Though being a good assumption of the first attempt in the context of optimal sensor placement, the attention received by water quality topics worldwide spurred the writer to abandon this assumption for better analyzing the actual behavior of the contaminants. In this context, the Thesis identifies the software EPANET-MSX as an essential support for qualitative studies in WDSs because, unlike other software used in literature, it describes dynamic interactions between contaminants, water and pipe/tank walls.

In fact, the multi-species numerical model used in the EPANET-MSX software to carry out the numerical simulations is able to show an initial drop in the chlorine due to its reactions with organic materials, tank walls and other substances present in the water. Then, the model reports the residual chlorine stabilization, and it correctly corroborates the fact that the presence of natural or injected organic matter (like *E. coli*) in the WDSs plays a vital role in the fate and transportation of chlorine. In particular, the numerical simulations demonstrate in a laboratory case study that *E. coli* injections can be detected if bacterial concentrations reach a concentration equal to 10^5UFC/mL and if they result in significant reduction of the free chlorine to a residual level, which depends either on *E. coli* and chlorine concentrations. For instance, the chlorine drops down to zero when the initial chlorine concentration is 0.3 mg/L and the *E. coli* injection is at least equal to 10^5 CFU/μL. Chlorine decays but does not reach zero if the concentration of the injected *E. coli* decreases (or the initial chlorine concentration increases). In addition, the Thesis describes the experimental apparatus and activities developed at the Civil Engineering and Geo-Environmental Laboratory of the Lille University (Villeneuve-d'Ascq, Lille - France) for the validation of the numerical models. The two studies yield the same results, confirming the threshold above which the bacteriological species is significantly detected and showing that the time required to reach the residual chlorine value increases with the decrease in the initial concentration of chlorine.

Consequently, the detection of microbial contamination of the order of 10^5 (CFU)/μL is faster for chlorine concentrations between 0.3 and 0.5 mg/L, rather than 1 mg/L.

Finally, knowing the feasibility of the early detection of non-specific biological anomalies (such as *E. coli*) through the use of the chlorine trend, chlorine measurements were exploited to develop an automated smart prototype system for the early anomaly detection, which efficiently enables water operators to apply in real-time management water quality control procedures, as well as a preemptive decision making process. In particular, an automated statistical model and AI-supported algorithms were developed and validated using chlorine data obtained from the numerical simulations above-mentioned. According to the statistical procedure, simulated chlorine data are statistically evaluated to define a risk indicator for the anomaly detection, which is lastly visualized by a color-based procedure.
Then, the AI-based algorithms exploit the concept of expert pattern recognition: an algorithm appropriately trained on the standard conditions of the water quality is able to recognize deviations from the normal conditions, identifying an anomaly. Chlorine and TOC are the parameters used to train the algorithms, and the Support Vector Machines (SVMs), as well as the Artificial Neural Network (ANN), are the supervised learning models that are tested and compared. In both proposed models, the results prove an efficient anomaly detection together with a risk-based classification of the detected anomalies. In fact, following the definition of each anomaly class based on the duration and the severity of the anomaly itself, the SVMs show a classification error of $4^0/_{00}$, which decreases to approximately $2^0/_{00}$ using the ANN. The results of the ANN are corroborated by the Confusion Matrix, which is a table with two rows and two columns that reports the number of false positives, false negatives, true positives, and true negatives.
Ultimately, the main output of the research is the visualization of the contaminated nodes on the network map, according to a color-based risk severity scale.

An extension of the optimization problem can involve multiple and simultaneous injections of contaminant at multiple points. Furthermore, analysis of spreading

contamination during short transients with significant pipe flow changes, which would require use of unsteady flow modeling, is another field to explore.
Regarding the non conservative contaminants, future developments may concern the study of other species (like pesticides, herbicides, etc.) in order to have a general picture of the water quality parameters, if a contamination event occurs.
Finally, a software can be developed associating ArcGIS with the advance pattern recognizers in order to contribute the decision support system of the water utilities managers.

It should be noted that the applications of early warning solutions to water distribution networks involve many skills (chemical, biological, hydraulic, electronic, computer and mathematical). The highly interdisciplinary connotation of the problem consequently implies significant difficulties in triggering an effective collaboration between the various involved cultural areas.
However, it is reasonable to foresee that if early monitoring and warning systems are increasingly used in the regular management of WDSs in relation to the growing security demand, the interaction between different disciplines will become more and more effective, determining a significant and positive impact on the advancement of knowledge and technologies.

Acknowledgements

My PhD work has been very rewarding, but onerous at the same time thus, I would like to acknowledge some people for their support over the three-years research period.

The first special thanks are for two Italian Professors: my Advisor, Professor Carlo Ciaponi, for his technical and human teachings during the whole period, and my Co-Advisor, Professor Enrico Creaco, for his precious suggestions that encouraged my scientific and personal growth.

I would also like to thank Professor Ilan Juran from the NYU-Tandon School of Engineering, Co-Advisor abroad of my PhD work. Notwithstanding his many worldwide commitments, he always showed great interest in my research, allowing me to cooperate with a highly valued University at international level. Many thanks are also due to Professor Isam Sharour from the University of Lille, for access to the Lille demonstration site and data sharing, as well as W-SMART Association for access to the operator experience of water utilities.

Outside the Academic context, the first heartfelt thanks are for my Parents: Marzia and Giorgio. They supported my stay in Pavia at Collegio Volta, making me feel the Family close. Together with my Parents, special thanks are for my Aunt Rossana and my Brother Davide, for their constant presence.

During these years in Pavia I met several people, each of them enriched me a little. Therefore, I would like to thank them all with a particular remark to Luigi and Sara, the colleagues I shared the office with.

Among my many friends, special thanks go to Anastasia, Eleonora, and Giulia, who shared my first 30s with me.

My experiences in NY will always stay in my memories because I really found another "Big Family" over there. Among all the people I met, I want to thank my

American relatives (especially Barbara and Aunt Gina), along with my friends, Teddy and Wilmer. They have been my Family for months with their hospitality and psychological encouragements.

I desire to conclude with particular thanks to the people that, together with my Parents, grew Davide and I up, giving us unconditioned love: my Grandparents. Three of them, Nonna Wilma, Nonno Renzo and Nonno Enrico will not be at my dissertation but Nonna Bruna is still always ready to tell me "welcome home".

Index of Figures

Index of Tables

References

Abdallah, A. (2015). "Reseaux d'eau potable: surveillance de la qualite de l'eau par des capteurs en ligne." *PhD Thesis* (University of Lille, France).

Ahn, J. C., Lee, S. W., Choi, K. Y., and Koo, J. Y. (2012). "Application of EPANET for the determination of chlorine dose and prediction of THMs in a water distribution system." *Sustain. Environ. Res.*, 22(1), 31-38.

Alfonso, L., Jonoski, A., and Solomatine, D. (2010). "Multiobjective optimization of operational responses for contaminant flushing in water distribution networks." *J. Wat. Res. Plann. Mngmnt*. 136 (1), 48-58.

Allgeier, S. C., Lee, Y., Umberg, K., and Tyree, M. (2008). "Operational experience with water quality event detection during the Cincinnati water security initiative pilot." In: Proceedings of the *AWWA Water quality technology Conference*, Cincinnati, 16–20 Nov. American Water Works Association, Denver (USA).

Allgeier, S. C., Umberg, K., Johnson, R., and Sahni, N. (2011). "Detection of distribution system contamination incidents using online water quality monitoring." In: Proceedings of the AWWA *Water quality technology Conference*, Phoenix, 13–17 Nov. American Water Works Association, Denver (USA).

Al-Omari, A. S., and Chandhry, M. H. (2001). "Unsteady-state inverse chlorine modeling in pipe networks." *J. Hydraul. Eng.*, 127(8), 669–677.

Alvisi, S., Franchini, M., Gavanelli, M., and Nonato, M. (2012). "Near-optimal scheduling of device activation in water distribution systems to reduce the impact of a contamination event." *J. Hydroinf*. 4 (2) 345-365.

Al-Zahrani, M., and Moied, K. (2001). "Locating optimum water quality monitoring stations in water distribution system." *World Water and Environmental Resources Congress*, ASCE, Reston, VA (USA).

Arad, J., Housh, M., Perelman, L., and Ostfeld, A. (2013). "A dynamic thresholds scheme for contaminant event detection in water distribution systems." *Water Res*., 47(5), 1899–1908.

Aral, M. M., Guan, J., and Maslia, M. L. (2001). "Identification of contaminant source locations and release history in aquifers." *J. Hydrologic Eng*., 6(3), 225-234.

AWWA Workshop, (2004). "Contamination Monitoring Technologies." American Water Works Association, Richmond, VA (USA).

AWWARF, (2002). "Online Monitoring for Drinking Water Utilities (Project #2545)". Report prepared by: E. Hargesheimer, O. Conio, and J. Popovicova, *American Water Works Association Research Foundation*, Richmond, VA (USA).

Bacteriological Analytical Manual (BAM), (1998). *8th Edition*, Revision A.

Banna, M. H., Imran, S., Francisque, A., Najjaran, H., Sadiq, R., Rodriguez, M., and Hoorfar, M. (2014). "Online Drinking Water Quality Monitoring: Review on Available and Emerging Technologies." Critical Reviews in *Environmental Science and Technology*, 44: 1370-1421, ISSN 1547-6537 online.

Baranowski, T. M., and LeBoeuf, E. J. (2006). "Consequence management optimization for contaminant detection and isolation". *J. Wat. Res. Plann. Mngmnt*. 132 (4), 274-282.

Basu, J. K., Bhattacharyya, D., and Kim, T. (2010). "Use of Artificial Neural Network in Pattern Recognition International." *J. Software Engineering and Its Applications*, Vol. 4, No.2.

Berry, J. W., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J. P. (2006). "Sensor placement in municipal water networks with temporal integer programming models." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2006)132: 4(218), 218–224.

Berry, J. W., Carr, R. D., Hart, W. E., Leung, V. J., Phillips, C. A., and Watson, J. P. (2009). "Designing contamination warning systems for municipal water networks using imperfect sensors." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2009)135: 4(253), 253–263.

Berry, J., Boman, E., Riesen, L. A., Hart, W. E., Phillips, C. A., Watson, J. P. (2012). "User's Manual TEVA-SPOT Toolkit. Version 2.5.2." *Scalable Algorithms Dept. & Discrete Math and Complex Systems Dept., Sandia National Laboratories Albuquerque.*

Blokker, E. J. M., Vreeburg, J. H., Buchberger, S. G., and Van Dijk, J. C. (2008). "Importance of demand modelling in network water quality models: a review." *Drink. Water Eng. Sci. Discuss*. 1, 1-20.

Bonadonna, L., and Della Libera, S. (2005). "Problemi di alterazione microbiologica: biofilm e biofuling (Fenomeno, cause, effetti)". *Congress of "Influenza dei sistemi di distribuzione sulla qualità dell'acqua potabile"* in collaboration with *Fondazione AMGA* and *UNIBS*, November 18th, Genova (ITA).

Bowsell, D. (2002). "Introduction to Support Vector Machines". Available at *http://dustwell.com/PastWork/IntroToSVM.pdf*

Braun, M., Bernard, T., Ung, H., Piller, O., and Gilbert, D. (2015). "CFD Modeling of Contaminant Mixing at Junctions for an online Security Management Toolkit in Water Distribution Networks." *J. Water Supply: Research and Technology* - AQUA, IWA, 64(5), 504–515.

Bristow, E., and Brumbelow, K. (2006). "Delay between sensing and response in water contamination events." *J. Infrastruct. Syst*. 12(2):87-95.

BTO Report.028, (2008). *Joint research programme of the Dutch and Belgium* (Dutchspeaking regions) *water works*.

Carlson, K., Byer, D., and Frazey, J. (2004). "Section 5.2 - Candidate Instruments and Observables. Section 5.3. - Models." *Draft of Methodology and Characteristics of Water System Infrastructure Security*.

Chang, N. B., Pongsanone, N. P., and Ernest, A. (2012). "Optimal sensor deployment in a large-scale complex drinking water distribution network: Comparisons between a rule-based decision support system and optimization models." *Comput. Chem. Eng*., 43(Aug), 191–199.

Cheifetz, N., Sandraza, A. C., Feliers, C., Gilbert, D., Piller, O., and Lang, A. (2015). "An incremental sensor placement optimization in a large real-world water system." *Procedia Eng.*, 119, 947–952.

Clark, R. M., Haught, R., and Panguluri, S. (2002). "Remote Monitoring and Network Models: Their Potential for Protecting the Nation's Water Supplies." Presented at *Workshop on Advanced Technologies in Real-Time Monitoring and Modeling for Drinking Water Safety and Security*, Newark, New Jersey, (USA).

Cortes, C., and Vapnik, V. (1995). "Support-vector networks." *Machine Learning*. 20 (3), 273–297.

Creaco, E., and Franchini, M. (2012). "Fast network multi-objective design algorithm combined with an a-posteriori procedure for reliability evaluation under various operational scenarios." *Urban Water J*., 9(6), 385–399.

Creaco, E., Franchini, M., and Todini, E. (2016). "The combined use of resilience and loop diameter uniformity as a good indirect measure of network reliability." *Urban Water J.*, 13(2), 167-181.

Creaco, E., and Pezzinga, G. (2015). "Embedding linear programming in multi objective genetic algorithms for reducing the size of the search space with application to leakage minimization in water distribution networks." *Environ. Modell. Software*, 69(Jul), 308–318.

Cui, J., and Wang, Y. (2010). "A novel approach of analog fault classification using a support vector machines classifier." *Metrology and Measurement Systems*, Volume 17, 4.

Deb, K., Pratap, A., Agrawal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NGSA-II." *IEEE Trans. Evol. Comput.*, 6(2), 182–197.

Deuerlein, J., Piller, O., and Montalvo, I. (2014). "Improved Real-time Monitoring and Control of Water Supply Networks by Use of Graph Decomposition." *Procedia Engineering*, 89(0), 1276-1281

Diao, K., and Rauch, W. (2013). "Controllability analysis as a pre-selection method for sensor placement in water distribution systems." *Water Res.*, 47(16), 6097–6108.

Di Cristo, C., and Leopardi, A. (2006). "Uncertainty Effects on Pollution Source Location in Water Networks." *Eighth Annual Water Distribution Systems Analysis Symposium (WDSA)*, August 27-30, Cincinnati, Ohio, (USA).

Dorini, G., Jonkergouw, P., Kapelan, Z., Di Pierro, F., Khu, S. T., and Savic, D. (2006). "An efficient algorithm for sensor placement in water distribution systems." Proc., *8th Annual Water Distribution Systems Analysis Symp.*, ASCE, Reston, VA (USA).

Duirk, S. E., and Collette, T. W. (2006). "Degradation of chlorpyrifos in aqueous chlorine solutions: pathways, kinetics and modeling." *Environ. Sci. Technol*. 40, 546-551.

Duirk, S. E., Desetto, L. M., and Davis, G. M. (2009). "Transformation of organophosphorus pesticides in the presence of aqueous chlorine: kinetics, pathways and structure e activity relationships." *Environ. Sci. Technol*. 43, 2335-2340.

Eliades, D., and Polycarpou, M. (2006). "Iterative deepening of Pareto solutions in water sensor networks." Proc., *8th Annual Water Distribution System Analysis Symp*., ASCE, Reston, VA (USA).

European Commission (Raich J.) (2013). "Review of sensors to monitor water quality ." Thematic group "Chemical and biological risks to the water sector" – *European Reference Network for Critical Infrastructure Protection (ENRNCIP) Project*, European Commission.

European Project, (2002). "Environmental data mining: learning algorithms and statistical tools for monitoring and forecasting." http://cordis.europa.eu/project/rcn/65948_en.html

Fisher, I., Kastl, G., and Sathasivan, A. (2011). "Evaluation of suitable chlorine bulk-decay models for water distribution systems." *Water Research* 45, 4896-4908.

Girolami, M. (2002). "Mercer kernel-based clustering in feature space." *IEEE Transactions on Neural Networks*, 13(3), 780-784.

Gleick, P. H. (2006). "Water and terrorism." *The World's Water*, 1-28.

Groover, (2008). "Better Sharing of Water Utility Security Incidents Still Needed."*http://www.miwater.org/awwa/Whats_New_Archives/092003%20security%20bulletin.htm*

Guan, J., and Aral, M. M. (1999). "Optimal remediation with well locations and pumping rates selected as continuous decision variables." *J. Hydrol.*, 221, 20-42.

Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006). "Identification of Contaminant Sources in Water Distribution Systems Using Simulation–Optimization Method: Case Study." *J. Water Resour. Plann. Manage.*, 132(4), 252-262.

Gueli, R. (2006). "Predator–prey model for discrete sensor placement." Proc., *8th Annual Water Distribution System Analysis Symp.*, ASCE, Reston, VA (USA).

Guidorzi, M., Franchini, M., and Alvisi, S. (2009). "A multi-objective approach for detecting and responding to accidental and intentional contamination events in water distribution systems." *Urban Wat*. 6 (2), 115-135.

Hall, J., Zaffiro, A. D., Marx, R. B., Kefauver, P. C., Krishnan, E. R., Haught, R. C., and Herrmann J. G. (2007). "On-line water quality parameters as indicators of distribution system contamination." *J. Am. Water Works Ass.* 66-77.

Hallama, N. B., Westa, J. R, Forstera, C. F, Powellb, J. C, and Spencerc, I. (2002). "The decay of chlorine associated with the pipe wall in water distribution systems." *Water Research*, Volume 36, 3479–3488.

Hagar, J., Murray, R., Haxton, T., Hall, J., and McKenna, S. (2013). "The Potential for Using the CANARY Event Detection Software to Enhance Security and Improve Water Quality." Proc., *Environmental & Water Resources Institute*, May 19-23, 2013, Cincinnati, OH (USA).

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

Hart, W. E., and Murray, R. (2010). "Review of Sensor Placement Strategies for Contamination Warning Systems in Drinking Water Distribution Systems". *J. Water Resources Planning and Management*, 136(6):611-619.

Hasan, J., States, S., and Deininger, R. (2004). "Safeguarding the security of public water supplies using early warning systems: A brief review." *J. of Contemporary Water Research & Education*, 129(1), 27-33.

Hatchett, S., Uber, J., Boccelli, D., Haxton, T., Janke, R., Kramer, A., Matracia, A., and Panguluri, S. (2011). "Real- time distribution system modeling:

development, application, and insights." In: Savic, D.A., Kapelan, Z., and Butler, D. Proceedings of the *11th International Conference on Computing and Control for the Water Industry*, Exeter, UK. Centre for Water Systems, University of Exeter, Exeter (UK).

Highsmith, A. K., Mcshane, W., Margolis, S., and Smith, D. L. (2004). "Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System." Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System, Chapter: Appendix: *Methodology and Characteristics of Water System Infrastructure Security 5.1 Contaminants and Concentrations*, Publisher: American Society of Civil Engineers, pp.A3-A30.

Huang, J., McBean, E. A., and James, W. (2006). "Multiobjective optimization for monitoring sensor placement in water distribution systems." Proc., *8th Annual Water Distribution System Analysis Symp.*, ASCE, Reston, VA (USA).

Islam, M. R., Chandhry, M. H., and Clark, R. M. (1997). "Inverse modeling of chlorine concentration in pipe networks under dynamic condition."*J. Environ. Eng.*, 123 (10), 1033–1040.

Istituto Superiore di Sanità, (2014). "Linee guida per la valutazione del rischio nella filiera delle acque destinate al consumo umano secondo il modello dei Water Safety Plan" (by L. Lucentini, L. Achene, V. Fuscoletti, F. Nigro Di Gregorio & P. Pettine) – 89 p., *Rapporti ISTISAN 14/21 – ISSN: 1123-3117.*

James, R., Dindal, A., Willenberg, Z., and Riggs, K. (2005). "Environmental technology verification report: Analytical technology." *Inc. Q45wq Continuous Multiparameter Water Quality Monitor*, Battelle, Columbus, OH (USA).

Janke, R., Murray, R., Uber, J., Taxon, T. (2006). "Comparison of physical sampling and real-time monitoring strategies for designing a contamination warning system in a drinking water distribution system." *J. Water Resour. Plann. Manage.* 132(4):310-313

Janke, R., Morley, K., Uber, J., and Haxton, T. (2011). "Real-time modeling for water distribution system operation: integrating security developed technologies with normal operations." In: Proc., *American Water Works Association, water security conference and distribution systems symposium*, Nashville, TN, 11–14 Sept. American Water Works Association, Denver (USA).

Janke, R., Tryby, M. E., and Clark, R. M. (2014). "Protecting Water Supply Critical Infrastructure: An Overview." In *Securing Water and Wastewater Systems (R.M. Clark and S. Hakim eds.),* Springer International Publishing Switzerland.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1998)124:4(192), 192–198.

Koch, M. W., and McKenna, S. A. (2011). "Distributed sensor fusion in water quality event detection." *J. Water Res Pl*. ASCE 137(1):10-19.

Kohavi, R., and Provost, F. (1998). "Glossary of terms". *Machine Learning*. 30, 271–274.

Kroll, D. J. (2006). "Securing our water supply: protecting a vulnerable resource.", *PennWell Corp,* Tulsa.

Kruskal, J. B. (1956). "On the shortest spanning subtree of a graph and the traveling salesman problem". Proc. *of the American Mathematical Society,* 7: 48–50.

Kumar, A., Kansal, M. L., and Arora, G. (1997). "Identification of monitoring stations in water distribution system." *J. Environ. Eng.*, 10.1061/(ASCE)0733-9372(1997)123:8(746), 746–752.

Kupperman, R. H. and Trent, D. M. (1979). "Terrorism: Threat, Reality, Response." *Hoover Institution Press,* Stanford, CA (USA).

Laird, C. D., Biegler, L.T., Van Bloemen Waanders, B. G., and Bartlett, R. A. (2005). "Contamination source determination for water networks." *J. Water Resour. Plann. Manage*, Vol. 131, Issue 2 (March 2005)

Laird, C. D., Biegler, L. T., and Van Bloemen Waanders, B. G. (2006). "Mixed Integer Approach for Obtaining Unique Solutions in Source Inversion of Water Networks". *J. Water Resour. Plann. Manage*, Vol. 132(4).

Lee, B. H., and Deininger, R. A. (1992). "Optimal locations of monitoring stations in water distribution systems." *J. Environ. Eng*., 10.1061/(ASCE)0733-9372(1992)118,1(4), 4–16.

Liu, L., Ranjithan, S. R., and Mahinthakumar, G. (2011). "Contamination Source Identification in Water Distribution Systems Using an Adaptive Dynamic Optimization Procedure." *J. Water Resour. Plann. Manage*., 137(2): 183-192.

Mamo, T. G., Juran, I., and Shahrour, I. (2014). "Municipal Water Pipe Network Leak Detection and Monitoring system Using Advanced Pattern Recognizer Support Vector Machine (SVM)" *J. Pattern Recognition Research,* Volume 9, No. 1.

Marchi, A., Salomons, E., Ostfeld, A., Kapelan, Z., Simpson, A. R., Zecchin, A. C., ... Asadzadeh, M. (2014). "The battle of the water networks II (BWN-II)." *J. Water Resour. Plann. Manage.,* 10.1061/(ASCE)WR.1943-5452.0000378, 04014009-1-14.

Mays, L. W. (2004). "Cyber threats and IT/SCADA and System Vulnerability". Chapter (*McGraw-Hill Professional*), *Access Engineering.*

McKenna, S. A., Hart, D. B., and Yarrington, L. (2006). "Impact of sensor detection limits on protecting water distribution systems from contamination events." *J. Water Resour. Plann. Manage*., 10.1061/(ASCE)0733-9496(2006)132:4(305), 305–309.

Morley, K., Janke, R., Murray, R., and Fox, K. (2007). "Drinking water contamination-warning systems: water utilities driving water security research." *J. Am. Water Works Ass*., 99(6):40-46.

Murray, R., Janke, R., and Uber, J. (2004). "The threat ensemble vulnerability assessment program for drinking water distribution system security." In: Proc., *ASCE-EWRI World Water and Environmental Resources Congress*, Salt Lake City, UT, June 27-July 1, 2004.

Murray, R., Janke, R., Hart, W.E., Berry, J.W., Taxon, T., and Uber, J. (2008). "Sensor network design of contamination warning systems: a decision framework." *J. Am. Water Works Ass*. 100(11):97-109.

Oliker, N., and Ostfeld, A. (2014a). "A coupled classification-evolutionary optimization model for contamination event detection in water distribution systems." *Water Res*., 51(15), 234–245.

Oliker, N., and Ostfeld, A. (2014b). "Minimum volume ellipsoid classification model for contamination event detection in water distribution systems." *J. Environ. Model. Software*, 57(Jul.), 1–12.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2004)130:5(377),377–385.

Ostfeld, A., and Salomons, E. (2005). "Securing water distribution systems using online contamination monitoring." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2005)131:5(402), 402–405.

Ostfeld, A., and Salomons, E. (2006). "Sensor network design proposal for the battle of the water sensor networks (BWSN)." Proc., *8th Annual Water Distribution System Analysis Symp*., ASCE, Reston, VA (USA).

Ostfeld, A., Uber, J., Salomons, E., Berry, J., Hart, W., Phillips, C., Watson, J., Dorini, G., Jonkergouw, P., Kapelan, Z., Di Pierro, F., Khu, S., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S., Barkdoll, B., Gueli, R., Huang, J., McBean, E., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., Van-Briesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G., Wu Z., and Walski T. (2008). "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms." *J. Water Resour. Plann. Manage*., 10.1061/(ASCE)0733-9496(2008)134:6(556),556–568.

Panguluri, S., William, P.R. Jr., and Clark, R. M. (2004). "Cyber threats and IT/SCADAsystem vulnerability." In: Mays LW (ed) *Water supply systems security*. Mc Graw-Hill, New York. pp 5.1-5.18.

Pasha, M. F. K., and Lansey, K., (2010). "Effect of parameter uncertainty on water quality predictions in distribution systems-case study." *J. Hydroinformatics*, 12, 1-21.

Pemmasani, L. (2012). "Analysis of Fate and Transport of E. coli in Chlorinated Drinking Water Distribution System in the Presence of Natural Organic Matter." *M.S. Thesis*, Graduate School of the University of Cincinnati, Cincinnati, OH (USA).

Perelman, L., and Ostfeld, A. (2011). "Topological clustering for water distribution systems analysis." *J. Environ. Software*, 26(7), 969–972.

Perelman, L., Arad, J., Housh, M., and Ostfeld, A. (2012). "Event detection in water distribution systems from multivariate water quality time series." *Environ. Sci. Technol*., 46(15), 8212–8219.

Pikus, I. M. (2004). "Interim Voluntary Guidelines for Designing an Online Contaminant Monitoring System."ASCE, *Am. Water Works Ass. Richmond, VA (USA).*

Piller, O., Deuerlein, J., Gilbert, D., and Weber, J.M. (2015). "Installing fixed sensors for double calibration and early-warning detection purposes." *Procedia Eng*., 119, 564–572.

Poulin, A., Mailhot, A., Grondin, P., Delorme, L., and Villeneuve, J. P. (2006). "Optimization of operational response to contamination in water networks." In Proc. *8th AnnualWater Distribution Systems Analysis Symposium*. University of Cincinnati, Cincinnati, OH (USA).

Preis, A., and Ostfeld, A. (2006). "Contamination Source Identification in Water Systems: A Hybrid Model Trees–Linear Programming Scheme." *J. Water Resour. Plann. Manage.*, 2006, 132(4), 263-273.

Preis, A., and Ostfeld, A. (2008a). "Multiobjective contaminant response modeling for water distribution systems security." *J. Hydroinf.*, 10 (4), 267-274.

Preis, A., and Ostfeld, A. (2008b). "Multiobjective contaminant sensor network design for water distribution systems". *J. Water Resour. Plann. Manag.*, 134 (4), 366.

Propato, M., and Piller, O. (2006). "Battle of the water sensor networks." Proc. *8th Annual Water Distribution System Analysis Symp.*, Univ. of Cincinnati, ASCE, Reston, VA (USA).

Rathi, S., and Gupta, R. (2016). "A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks." KSCE *J. Civil Eng.*, 20(2), 597–608.

Rice, E. W., Clark, R. M., and Johnson, C. H. (1999). "Chlorine Inactivation of Escherichia coli O157:H7". *Emerging Infectious Diseases,* 5(3), 461-463.

Roberson, J. A., and Morley, K. M. (2005). "Contamination warning systems for water: an approach for providing actionable information to decision-makers". *American Water Works Association*, Richmond, VA (USA).

Rossman, L. A. (2000). "EPANET 2 User Manual". *U.S. Environmental Protection Agency*, Washington, D.C., EPA/600/R-00/057

Shang, F., Uber, J. G., and Polycarpou, M. M. (2002). "Particle backtracking algorithm for water distribution system analysis." *J. Environ. Eng.,* 128(5), 441–450.

Shang, F., and Uber, J. G. (2008). "Modeling Reaction and Transport of Multiple Species in Water Distribution Systems." *Environmental Science and Technology* 42(3), 808-814.

Shang, F., Uber, J. G., and Rossman, L. (2008). "EPANET Multi-Species Extension (EPANET-MSX) Software and User's Manual." *U.S. Environmental Protection Agency*, Washington, DC, EPA/600/C-10/002.

Shastri, Y., and Diwekar, U. (2006). "Sensor placement in water networks: A stochastic programming approach." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2006)132:3(192), 192–203.

Skadsen, J., Janke, R., Grayman, W., Samuels, W., TenBroek, M., Steglitz, B., and Bahl, S. (2008). "Distribution system on-line monitoring for detecting contamination and water quality changes." *J. Am. Water Works Ass.* 100(7):81-94.

Smola, A. J., and Schölkopf, B. (2004). "A tutorial on support vector regression." *Statistics and computing*, 14(3), 199-222.

Storey, M. V., Van Der Gaag, B., and Burns, B. P. (2011). "Advances in on-line drinking water quality monitoring and early warning systems." *Water Research*, 45, 741-747.

Times, N. Y. (2003). "Iraqi Agents Held in Plot to Poison Water Supply."

Tinelli, S., Creaco, E., and Ciaponi, C. (2017a). "Sampling significant contamina-tion events for optimal placement in water distribution systems" *J. Water Resour. Plann. Manage.*, 143(9): 04017058.

Tinelli, S., Creaco, E., and Ciaponi, C. (2017b). "Contaminazioni nelle reti idriche: sistemi di monitoraggio e di allarme." *Tecniche per la difesa del suolo e dall'inquinamento* – Proc. *XXXVIII Edition*, June 21st-26th, Cosenza (IT), EdiBios.

Tinelli, S., Creaco, E., and Ciaponi, C. (-) "Impact of Objective Function Selection on Optimal Placement of Sensors in Water Distribution Networks" *J. Water Resour. Plann. Manage.*, under review.

Tinelli, S., and Juran, I., (2017). "Numerical Modeling of Early Bio-Contamination in a Water Distribution System and Comparison with

Laboratory Experiments", Proc. *ASCE International Conference on Sustainable Infrastructure - ICSI*, NYC, NY (USA).

Tinelli, S., Juran, I., and Cantos, W. P. (2017). "Early detection System of non specific bio-contaminations in Water Distribution Systems" Proc. *IWA Efficient 2017*, July 18th-20th, Bath (UK) & in press in *J. Water Science and Technology: Water Supply*.

Uber, J. G., Boccelli, D. L., Summers, R. S., and Tryby, M. E. (2003). "Maintaining Distribution System Residuals Through Booster Chlorination." *American Water Works Association Research Foundation*, 111-113, Richmond, VA (USA).

Uber, J., Murray, R., and Janke, R. (2004). "Use of systems analysis to assess and minimize water security risks." *J. Contemp. Water Res. Edu.*, 129(1):34–40.

US-EPA, (2000). "Multi Agency Radiation Survey and Site Investigation Manual." (EPA 402-R-97-016), *U.S. Environmental Protection Agency*, Washington, DC (USA).

US-EPA, (2002). "Health Risks from Microbial Growth and Biofilms in Drinking Water Distribution Systems". *U.S. Environmental Protection Agency, Office of Ground Water and Drinking Water*, Washington, DC (USA).

US-EPA, (2005a) . "Technologies and Techniques for Early Warning Systems to Monitor and Evaluate Drinking Water Quality: A State-of-the-Art Review." *U.S. Environmental Protection Agency*, Washington, DC, EPA/600/R-05/156 (USA).

US-EPA, (2005b). "Water Sentinel System Architecture (Draft, version 1.0)". *U.S. Environmental Protection Agency*, Washington, DC, EPA 817/D-05/003 (USA).

US-EPA, (2010). "Water Quality Event Detection Systems for Drinking Water Contamination Warning Systems." *U.S. Environmental Protection Agency*, Washington, DC, EPA 600/R-010/036 (USA).

US-EPA, (2012). "CANARY Users Manual version 4.3.2." *U.S. Environmental Protection Agency*, Washington, DC, EPA 600/R-08/040B (USA).

US-EPA, (2013a). "Homeland security and EPA's strategic plan." *www.epa.gov/homelandsecurityportal/strategicplan.htm*. Accessed April 23.

US-EPA, (2013b). "Water Quality Event Detection System Challenge: Methodology and Findings." *U.S. Environmental Protection Agency, Office of Water*, Washington, DC, EPA 817-R-13-002 (USA).

US-Government Accountability Office (U.S. GAO), (2006). "Internet infrastructure: DHS faces challenges in developing a joint public/private recovery plan". GAO-06-672, *U.S. Government Accountability Office, Washington*, DC (USA).

Van Bloemen Waanders, B. G., Bartlett, R. A., Biegler, L.T., L. T. Biegler, and Laird, C. D. (2003). "Nonlinear programming strategies for source detection of municipal water networks." Presented at *EWRI Conference*, Philadelphia, PA (USA).

Van Der Gaag, B., and Volz, J. (2008). "Real-time On-line Monitoring of Contaminants in Water: Developing a Research Strategy from Utility Experiences and Needs". *KIWA Water Research*, Nieuwegein (NL).

Vugrin, E., McKenna, S. A., and Hart, D. (2009). "Trajectory clustering approach for reducing water quality event false alarms." In: Proc., *World Environmental and Water Resources Congress*, Kansas City, Missouri, May 17th-21st. American Society of Civil Engineers, Reston, VA (USA).

Walski, M., Chase, D., Savic, D., Grayman, W., Beckwith, S., and Koelle, E. (2003). "Advanced water distribution modelling and management." Waterbury, CT: *Haestad Press*.

Weickgenannt, M., Kapelan, Z., Blokker, M., and Savic, D. A. (2010). "Risk based sensor placement for contaminant detection in water distribution system." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000073, 629–636.

WHO, (2005). "Water Safety Plans: Managing drinking-water quality from catchment to consumer". *World Health Organization*, Geneva, Swiss.

WHO, (2008). "Water Safety Plan Manual: Step-by-step risk management for drinking-water suppliers". *World Health Organization*, Geneva, Swiss.

WHO, (2011). "Guidelines for Drinking Water Quality (Fourth Edition)". *World Health Organization*, Geneva, Swiss.

Williamson, F., Van Den Broeke, J., Koster, T., Koerkamp, M. K., Verhoef, J. W., Hoogterp, J., Trietsch, E., and De Graaf, B. R. (2014). "Online water quality monitoring in the distribution network". *Water Convention at the Singapore International Water Week 2014*.

Woo, H. M., Yoon, J. H., and Choi, D. Y. (2001). "Optimal monitoring sites based on water quality and quantity in water distribution systems.", *World Water and Environmental Resources Congress*, ASCE, Reston, VA (USA).

Wu, Z. Y., and Walski, T. (2006). "Multiobjective optimization of sensor placement in water distribution systems." Proc., *8th Annual Water Distribution System Analysis Symp.*, ASCE, Reston, VA (USA).

www.hach.com

www.mbonline.at

www.microlan.nl

www.s-can.at

www.secureau.eu

www.tech-associates.com

Zechman, E. M., Brill, E. D., Mahinthakumar G., and Ranjithan, S. (2006). "Addressing Non-uniqueness in a Water Distribution Contaminant Source Identification Problem". Proc. *8th Annual Water Distribution Systems Analysis Symposium (WDSA)*, August 27-30. Cincinnati, OH (USA).

Zhao, Y., Schwartz, R., Salomons, E., Ostfeld, A., and Poor, H. V. (2016). "New formulation and optimization methods for water sensor placement." *Environ. Modell. Software*, 76(Feb), 128–136.