Doctoral Program in Chemical and Biological Engineering

Contamination in Drinking Water Distribution Systems: Some Approaches to Forward and Inverse Modelling

Dissertation submitted by

Diogo Moreira da Costa

Department of Chemical Engineering

Thesis supervised by

Professor Fernando Gomes Martins

Professor Luís Manuel Ferreira Melo

To my Parents,

my Brother,

Grandmother Mariana,

and Joana.

Acknowledgments

The research leading to this thesis has received funding from the European Community's Seventh Framework Programme (FP7/2007-2011) under grant agreement n° 217976 {Project "SecurEau"}.

I want to start by thanking my supervisors, Professor Fernando Martins and Professor Luís Melo, for the opportunity to do this PhD thesis, for all the support and guidance they gave me throughout this work. Particularly, I would like to express a very special thanks to Professor Fernando Martins. It was a pleasure to work under your supervision and I am very grateful for everything I had the opportunity to learn from you.

I must also acknowledge all partners from the project "SecurEau", with whom I had the opportunity to collaborate. My participation in this project has set the perfect environment for the development of my thesis. A special thanks to Doctor Olivier Piller, Doctor Denis Gilbert and Hervé Ung from Irstea, who gave me the opportunity to work with them for a very interesting and fruitful period. I would also like to thank Professor Luís Miguel Madeira for his contribution to my work on the study of sorption phenomena.

A very special thanks to my friends, whose support and friendship gave me always a great help. I am very lucky for having such a great group of fantastic friends that it would be impossible to thank each one individually. However, I have to give a special acknowledgement to some of them.

Many thanks to David Gonçalves, one of the main witnesses of the effort I made for this thesis. It was really nice to work in such a friendly environment and I thank you for all the help you gave me, especially in my first experiences with Matlab.

I must also to thank João Cravo for our talks on artificial neural networks and for all the technical support he gave me. Now I know how to compile a Matlab application!

A very special thanks to Daniel Cardoso for his friendship, which lasts for as long as I can remember, and for the contribution he gave me from his experience in the operation of drinking water distribution systems.

Many thanks to Nuno Salvaterra for being such a great friend. From a technical/scientific point of view, your friendship was pretty much worthless for me but you surely managed to compensate with your always amazing companionship.

I am also very grateful to Luís Matias for being the best friend that anyone could ask. Ever since I met you, I have been lucky to enjoy your friendship. Especially in the most difficult moments, you were always there trying to cheer me up and I will never forget that.

The most special thanks goes out to my all family. I want to thank my parents for the education, support and affection they gave me my all life. I will always regard both of you as two perfect examples of what parents should be. Big thanks also to my brother. I could write a thousand words to express my gratitude and it wouldn't be enough, so let me just tell you that I am really happy and proud of having you as brother. To my Grandmother Mariana, I am deeply grateful for all the support you have always gave me. Since I was a little kid, you have always made me feel like I was the most beloved grandson in the world.

Finally, I would like give an exclusive thanks to Joana. I have the great luck of having you in my life in the last 4 years. You gave an invaluable contribution to the success of this work. Thank you for your unconditional support, patient and tenderness.

Abstract

Contamination of drinking water distribution systems (DWDSs) is a threat that can have major effects in public health besides economic and social consequences. DWDSs are highly vulnerable to deliberate attacks due to the difficulty inherent to the surveillance and protection of large and complex networks. In case of a contamination event, it is very important to locate the contamination source as fast as it is possible to assess the propagation of the contamination for taking the necessary security measures.

The main objective of this thesis was to study the simulation of the spread of contaminants in a DWDS and to develop methods for the localization of contamination sources after contamination events, based on information provided by a surveillance system. The work performed during this period enabled the development of two approaches for simulating the transport mechanisms of contaminants in DWDSs, the development of deterministic methods for the localization of contamination sources and the application of artificial neural networks (ANNs) in the development of methods for the localization of contamination sources.

A software tool was developed to implement an analytical approach for the simulation of the advective transport of contaminants, considering psudo-first order reaction terms, for steady hydraulic conditions. In addition, another software tool was developed to simulate the contaminant reactive transport along DWDSs considering sorption phenomena. This method demonstrated to be a relevant contribution for the study of the effects of the sorption phenomena in the modelling of the transport of contaminants in real DWDS.

The results achieved with the deterministic methods proved that these methods are effective in the search of the correct locations and times of the contamination, despite being based only on the analysis of the residence time of water in pipes. One of these methods also enables the study of the effect of false positives or false negatives at the sensors. Since only binary sensor status over time is required, these methods were considered suitable for application in real case scenarios. The methods based on the application of artificial neural networks also achieved very satisfactory results for real DWDSs. The methods were generally able to determine correctly the simulated source and to define a very restricted set of possible contamination sources, even when considering hydraulic scenarios with demand uncertainties. The time of computation required was very low, which makes these methods very suitable for application in real contamination scenarios.

Keywords: Drinking water distribution Systems, contamination source identification, water quality modelling, artificial neural networks.

Resumo

Contaminação de sistemas de distribuição de água potável (SDAPs) é uma ameaça que pode ter grandes repercussões na saúde pública, além de consequências económicas e sociais. SDAPs são altamente vulneráveis a ataques deliberados devido às dificuldades inerentes à vigilância e proteção de redes grandes e complexas. Na eventualidade da ocorrência de uma contaminação, é muito importante localizar a fonte de contaminação o mais rapidamente possível, para que seja possível avaliar a propagação da contaminação de modo a tomar as medidas de segurança necessárias.

O objetivo principal deste trabalho foi estudar a simulação da propagação de contaminantes num SDAP e desenvolver métodos para a localização de fontes de contaminação com base em informações fornecidas por um sistema de vigilância. O trabalho realizado durante este período permitiu o desenvolvimento de duas abordagens para simular os mecanismos de transporte de contaminantes em SDAPs, o desenvolvimento de métodos determinísticos para a localização de fontes de contaminação, a aplicação de redes neuronais artificiais para o desenvolvimento de métodos para a localização de fontes de contaminação e o desenvolvimento de um método para a otimização da seleção de pontos de amostragem.

Foi desenvolvida uma aplicação informática para implementar uma solução analítica para a simulação do transporte advectivo de contaminantes, considerando reação de primeira ordem, para condições hidráulicas constantes. Além disso, outra aplicação informática foi também desenvolvida para simular o transporte reativo de contaminantes ao longo SDAPs, tendo em consideração fenómenos de sorção. Este método demonstrou ser um contributo importante para o estudo dos efeitos dos fenómenos de sorção na modelação do transporte de contaminantes em SDAPs reais.

Os resultados obtidos com os métodos determinísticos provaram que estes métodos são eficazes para a procura dos locais e instantes de contaminação, apesar de apenas se basear na análise do tempo de residência da água nos tubos. Um destes métodos também permitiu o estudo do efeito de falsos positivos ou falsos negativos nos sensores. Uma vez que é apenas necessário o estado binário dos sensores ao longo do tempo, estes métodos foram considerados adequados para a aplicação em cenários reais. Os métodos baseados na aplicação de redes neurais artificiais também alcançaram resultados muito satisfatórios para SDAPs. Os métodos foram geralmente capazes de determinar corretamente a(s) fonte(s) simulada(s) e de definir um conjunto muito restrito de possíveis fontes de contaminação, mesmo considerando cenários hidráulicos com incertezas nos consumos de água. O tempo de computação necessário foi muito baixo, o que faz com que estes métodos sejam muito adequados para aplicação em situações reais.

Palavras-chave: Sistemas de distribuição de água potável, identificação de fontes de contaminação, modelação da qualidade da água, redes neuronais artificiais.

Contents

List of figures

List of Tables

Nomenclature

t	Time
x	Distance along the pipe
l_i	Length of pipe i
u_i	Flow velocity in pipe i
Q_i	Volumetric flow rate in pipe i
$C_i(x,t)$	Concentration at position x of pipe i in instant equal to t
Q_{Ads}	Rate of mass of contaminant adsorbed per unit mass of deposit
m_{dep}	Mass of sorbent
V	Water volume in dx
k_1	Pseudo-first order sorption rate constant
C_{eq}	Contaminant concentration in the water at the equilibrium
C_b	Contaminant concentration in the bulk
$r(C_i(x,t))$	Reaction rate
k	Kinetic coefficient
n	Reaction order
q	Contaminant concentration in the deposit
k_d	Desorption rate coefficient of first order
q_e*	Final equilibrium concentration of contaminant in the solid for the desorption process
q_0	Contaminant concentration in the deposit at the beginning of the desorption phase
$C_p(t)$	Concentration at node p in instant equal to t
$V_{\tan k}(t)$	Volume of water in a storage tank in instant t
pi	Set of the incoming links to the node p
q_{eq}	Contaminant concentration at the deposit at the equilibrium
Q_0	Maximum contaminant concentration at the deposit
B	Langmuir constant related to adsorption capacity
Cr	Courant number

Greek Letters

Δ	Interval

Index

p	Node
i	Link

Part I: Introduction

1 Framework

This thesis was carried out between 2008 and 2013 at the Laboratory for Process, Environmental and Energy Engineering (LEPAE) in the Chemical Engineering Department of the Faculty of Engineering, University of Porto. This work was performed under the European project "SecurEau", entitled "*Security and decontamination of drinking water distribution systems following a deliberate contamination*" (Project nº 217976).

1.1 Motivation and Relevance

Water is a fundamental resource for human and economic welfare and modern society depends on complex, interconnected water infrastructure to provide reliable safe water supplies and to remove and treat wastewater (Gleik, 2006). Thus, contamination of drinking water distribution systems (DWDSs) is a threat that can have major effects in public health, besides economic and social consequences.

Water resources and systems are attractive targets because there is no substitute for water. Thus, the risk of a terrorist attack at water systems is real, as it has already happened in the past (Gleik, 2006). The consequences of such an action can be grim and the risk of casualties, social disruptions and disarray is high.

With emerging security issues, drinking water utilities are facing new challenges. Within their mission statement, not only should they seek to provide sufficient quantities of high quality water, but they should also be concerned with security issues. Water networks should be operated in such a way to protect against, detect and respond to man-made and natural threats and disasters (Poulin, 2006).

DWDSs are highly vulnerable to deliberate attacks due to the difficulty inherent to the surveillance and protection of large and complex networks. Since it is not realistic to measure the water quality in each node of a DWDS, one reasonable approach to guarantee the safety of the water consumers is to have a warning system constituted by a set of sensors positioned throughout the network to detect changes in water quality.

When a contaminant is introduced in a DWDS, it might spread very quickly through large areas of the network. Thus, in case of a contamination event, it is very important to locate the contamination source as fast as it possible. Besides that, the localization of the contamination source has great significance to predict the distribution of the contaminant and to take the necessary security measures.

The detection of water quality deterioration in DWDSs and the problem of the localization of the contamination sources call for the development of new, sensitive and rapid methodologies.

Therefore, new simulation tools and procedures need to be developed and made available to water utilities, to help detect and manage contamination events in practical applications.

1.2 Objectives

The main objective of this thesis was to study the occurrence of contaminations in DWDSs.

This was accomplished by two main tasks:

a) Development of methods for the simulation of the spread of contaminants in a DWDS;

b) Development of methods for the localization of contamination sources after contamination events were detected based on information provided by a surveillance system.

1.3 Thesis Outline

The work presented in this thesis is divided in six main parts.

Part I is constituted by Chapters 1, 2, 3, 4 and 5. Chapter 1 presents the subject relevance, the objectives of this work and the thesis structure. Chapter 2 presents an overview of the previous research performed in the field of water quality modelling in DWDSs and some software tool packages available for the simulation of DWDSs. Chapter 3 presents a literature review on the subject of localization of contamination sources in DWDSs. Chapter 4 presents an introduction to the application of artificial neural networks to the problems under study. Finally, in Chapter 5, several examples of DWDSs, which are used as case studies in this thesis, are presented.

Part II addresses the simulation of the water quality behaviour in DWDSs. Chapter 6 presents an analytical approach to the modelling of the advective transport of contaminant in DWDSs with pseudo-first order reaction, considering steady hydraulic conditions. Chapter 7 presents a numerical solution for the simulation of the advective transport of contaminants in DWDSs considering the occurrence of reaction and sorption phenomena at the water-pipe wall interface, for specified dynamic hydraulic conditions.

Part III proposes deterministic methods for the localization of contamination sources based on the analysis of the residence time of water in pipes. Chapter 8 presents a method based on the analysis of the residence time of water in pipes for fixed time intervals. Chapter 9 proposes a different method based on the information given by successive positive readings of the sensors. This chapter also addresses the occurrence of false positives or false negatives.

Part IV addresses the problem of the localization of contamination sources in DWDSs through the application of artificial neural networks. Chapter 10 presents a method suitable for single contamination events. Chapter 11 presents strategies to apply this method to large DWDSs and Chapter 12 presents an extension of the method for multiple contamination scenarios.

Part V goes over the main conclusions of this work and identifies the research areas in which future work can be performed to enhance the results presented in this thesis.

2 Drinking Water Distribution Systems

2.1 Modelling and Simulation of DWDS

DWDSs are typically buried infrastructures, in which just a small fraction of its components can be frequently examined. On the other hand, the actual capacity for monitoring these systems, through measurements of pressure, flow rate or concentration of water quality parameters, for example, is greatly insufficient in time and space, because of the complexity, the number of points of consumption and to the variability of the consumption (Coelho et al., 2006).

Models are especially important for DWDSs due to their complex topology, frequent growth and change, and sheer size. DWDSs simulation, which replicates the dynamics of an existing or proposed system, are commonly performed when it is not practical for the real system to be directly subjected to experimentation, or for evaluating a system before it is actually built. In addition, for situations in which water quality is an issue, directly testing a system may be costly and a potentially hazardous risk to public health. Simulations can be used to anticipate problems in proposed or existing systems without disrupting the actual system. Operators need to be prepared to deal with a very wide range of emergencies. Planning ahead for these emergencies by using a model may prevent service from being compromised, or may at least minimize the extent to which customers are affected. Modelling is an excellent tool for emergency response planning and contingency (Strafaci, 2003).

The study of water quality aspects within drinking-water distribution systems is of great significance as it plays an important role for assuring the quality of the water that is delivered to the consumers. Computer-based mathematical models are useful tools for evaluating the water quality changes in drinking-water distribution systems.

In the last two decades, the investigation in this area was centred mainly in the simulation of chlorine decay (Rossman et al., 1994; Clark et al., 1995; Ozdemir and Ger, 1999; Al-Omari and Chaudhry, 2001; Ozdemir and Ucak, 2002) in drinking-water distribution systems. Currently available chlorine decay and propagations models treat the pipe segments as if they were plug flow reactors. The radial and axial dispersion are generally neglected in most developed models (Ozdemir and Ucak, 2002).

Several approaches have been taken to numerically model the transport of contaminants in DWDSs, either in steady or dynamic hydraulic conditions. Steady-state models use the law of mass conservation to determine the ultimate concentration distribution of contaminants assuming that the distribution system reaches hydraulic equilibrium (Boulos et al., 1995). These models can only

provide intermittent assessment capabilities, which make the simulation of the water quality behaviour less accurate (Rossman and Boulos, 1996). Dynamic models rely on a system simulation approach to determine the movement and the fate of contaminants under time-varying demand, supply and hydraulic conditions (Boulos et al., 1995). These models provide a more realistic approximation to the actual operation of DWDSs since it considers time-varying conditions in the simulation of DWDSs (Rossman and Boulos, 1996).

Boulos et al. (1992) proposed an explicit solution for modelling water quality parameters, such as chemical concentrations and water age, considering steady hydraulic behaviour. In 1993, Boulos and Altman also proposed algorithms for directly determining a variety of blended water quality parameters under steady-state. The developed algorithms are formulated analytically from mass balance relationships as contingent linear boundary value problems in conjunction with a topological sorting methodology and aimed to enhance the DWDSs water quality management (Boulos and Altman, 1993).

Rossman et al. (1993) developed an explicit water quality modelling algorithm for tracking dissolved substances in water distributions networks, considering dynamic systems. The algorithm was based on a mass-balance relation within pipes that considered both advective transport and reaction kinetics. The proposed method allowed simulating spatial and temporal distribution of substances in water distribution networks. Later, Rossman et al. (1994) presented a mass-transfer-based model for predicting chlorine decay in drinking-water distribution networks. These authors considered first-order reactions of chlorine occurring in the bulk flow and at the pipe wall. The model was able to explain observed phenomena in previous chlorine decay studies, such as higher decay rates in smaller pipes or in higher velocity flows. The chlorine decay model was incorporated into the EPANET, which is a software tool able to perform dynamic water quality simulations on complex pipe networks. Boulos et al. (1994) also presented a method for simulating the transport of conservative substances in DWDSs under dynamic hydraulic conditions. These authors proposed an event-driven method that determines the optimal segmentation scheme with the smallest number of segments necessary to perform the simulation, achieving a water quality modelling less sensitive to the structure of the DWDS and to the length of the simulation. Boulos et al. (1995) presented an extension of the method of Boulos et al. (1994) to enable the simulation of the transport of reactive species. However, they stated that the water quality models can only be effective for the simulation of reactive species if the reaction mechanisms are properly defined.

Mau et al. (1996) developed an analytical approach for modelling the water quality within storage tanks or reservoirs based on material mass balance that considers transport, mixing and kinetic

reaction processes. The model-generated results showed good agreements with the observed field measurements.

Dynamic models can be classified spatially as either Eulerian or Lagrangian and temporally as time-driven or event-driven. Eulerian models divide the pipe network in fixed control volumes and register changes as water flows through them while Lagrangian models track changes in a series of discrete parcels of water as they travel through the network. Time-driven models update the state of the network at fixed time intervals while event –driven models update the state of the network only when a change occurs, such as when a new parcel of water reaches the end of the pipe (Rossman and Boulos, 1996).

Rossman and Boulos (1996) made a comparison between four numerical methods (two Eulerian and two Lagrangian approaches, considering both time-driven and event-driven approaches) for modelling the water quality behaviour in DWDS. Results showed that the Lagrangian methods were more efficient for simulating chemical transport and for modelling the water age, while the Eulerian methods were more memory-efficient. In 2004, Munavalli and Kumar made a comparison between Lagrangian time-driven methods and event-driven methods for varying concentration tolerance and water quality time step and proposed a new hybrid method for improving the accuracy of these methods. They concluded that the time-driven methods were affected by both concentration tolerance and water quality time step, while event-driven methods depended on concentration tolerance. The proposed hybrid method proved to be less sensitive to these parameters and required a reasonable computational effort.

The numerical discretization used to model the advection in water networks provides a good solution for Gaussian initial distribution but produces artificial diffusion when steep gradients are simulated. The finite-difference methods have become more popular for one-dimensional problems due to their simplicity and modest computational effort. Explicit finite-difference techniques are generally simply programmed, with the time step size usually restricted by the Courant stability condition. The majority of implicit finite-difference methods are unconditionally stable; however they are significantly more complex and require extra parameters (Islam and Chaudhry, 1997).

Islam and Chaudhry (1997) presented a study of the application of seven finite-difference methods and one polynomial interpolation scheme to the solution of the transport equation for both Gaussian and non-Gaussian initial distributions. The results obtained with each method were compared with the exact solution. Almost every method gave good results for a Gaussian initial distribution. However, for steep gradient concentrations some of the methods produced high oscillations. The third and

fourth-order methods produced the best result for this scenario but required additional computation time and programming became more complex.

The majority of the models presented so far used the extended-period simulation technique to solve the time-varying flow conditions in the network. Islam and Chaudhry (1998) presented a dynamic model to compute the spatial and temporal distribution of substances in a network under slowly varying flow conditions. In this method, first the flow conditions are computed and then the substances concentration is simulated, with a separation between advection and dispersion. However, the results were not satisfactory when the flow became more unsteady and reverse flows occurred.

Ozdemir and Ger (1999) evaluated the effects of the difference between day time and night time operations. They considered that dispersion may also be taken into account, so they developed an unsteady 2-D convective dispersive model and compared the model output with experimental results. However, this work demonstrated that this procedure does not lead to any enhancements in these studies. In 2002, Ozdemir and Ucak developed a model for evaluating chlorine decay in drinking-water distribution networks using a simplified two dimensional expression to model the chlorine transport. The decay equation within a single pipe included the bulk-flow reaction, radial diffusion and pipe wall reaction of chlorine.

Zierolf et al. (1998) developed an input-output model that expresses the chlorine concentration at a given pipe junction and time as a weighted average of exponentially decayed values of the concentrations at all adjacent upstream junctions. The exponential decay models chlorine reactions and the weighted average reflect the effect of mixing at the pipe junction. Measured data from real systems are used to calibrate the model off-line. The model finds all paths from treatment stations to each measured point, so the reaction rate associated with chlorine decay at the pipe wall can be adjusted to improve predicted chlorine concentrations, increasing the model accuracy. Shang et al. (2002) presented a particle backtracking algorithm (PBA), which is a simpler and more efficient version from the input-output model developed by Zierolf et al. in 1998. Besides that, this model is extended to allow the analysis of storage tanks and multiple quality inputs. The main advantage of this algorithm is the capability of analysing specific paths and its characteristics since they are made explicit by this tool.

Alhumaizi et al. (2003) analysed the set of governing equations that describes the reaction-diffusion-convection system for a homogeneous tubular reactor, considering both steady-state and dynamic systems, using different standard reduction techniques, such as finite-difference, orthogonal collocation and finite-element methods. Results demonstrated that, for cases dominated by convection and reaction, high-resolution discretization methods, such as essentially non oscillatory (ENO) and

total variation diminishing (TVD), eliminate oscillations and are efficient for tracking steep moving fronts.

Different types and order of finite difference methods were identified to model the accumulation and the convection derivatives. First-order finite-difference methods result in monotonic and stable solutions but they are also strongly dissipative, giving less accurate solutions for strongly convective systems. On the other hand, higher-order difference methods are less dissipative but prone to numerical instabilities. Significant improvements have been made to the accuracy of these models. One common approach to reduce non-physical oscillations near discontinuities is to add numerical diffusion that should be tuned to be large enough near discontinuities but small enough elsewhere to maintain high-order accuracy. TVD schemes and flux and slope-limiters approaches, such as ENO, weighted essentially non-oscillatory (WENO) schemes and Superbee flux corrector, are examples of these methods (Alhumaizi, 2004). This author made a comparison between several finite difference schemes developed to solve the convection-reaction problem. The results demonstrated that first-order schemes require large grid numbers to improve the solution accuracy, which might not be practical for large DWDSs, while higher-order methods caused great oscillations near discontinuities. High resolution TVD, WENO schemes and the Superbee flux corrector methods were tested and proved to be accurate for solving cases with steep concentration profiles. In 2007, Alhumaizi analysed the strengths and limitations of five flux-limiters to simulate numerically the behaviour of a homogenous tubular reactor with steep moving fronts. All methods were successful in solving the cases with steep concentrations profiles without giving negative concentrations. The Superbee limiter was found to be the fastest scheme for simulating the sharp front of the model for all cases.

Hallam et al. (2002) stated that the chlorine decay rates in DWDSs for bulk and wall demands should be modelled separately because they depend on different factors. They developed a methodology for the laboratory determination of the wall decay rate. The results indicated that wall decay rates were limited by chlorine transport for pipe materials with high reactivity and by the pipe material characteristics for pipe materials with low reactivity. In general, water velocity increased wall decay rates though the statistical confidence is low for low reactivity pipes. A moderate biofilm coating did not influence the wall decay rate for low reactivity pipes.

Munavalli and Kumar (2003) presented an inverse model for determining the water quality parameters for both bulk and pipe wall reactions and the source strength of chlorine necessary to maintain the specified chlorine residual at a target node. The method was applied to a DWDS operating with steady-state hydraulic behaviour. In 2005, the same authors extended this method for DWDSs with dynamic hydraulic behaviour (Munavalli and Kumar, 2005). The application of the method to a real

DWDS demonstrated that the quantity, quality and location of measurements nodes played an important role in the estimation of parameters.

Yang et al. (2008) tried to define contaminant-chlorine reactions taking place during the transport of water in pipes to characterize the hydraulic dispersion of non-reactive chemicals, to improve the detection of contaminants using water quality sensors and to establish a model for predicting the fate and transport of a "slug" of a reactive contaminant. They performed a series of experiments in a pilot-scale network and made a comparison with the results of simulations. The results showed good agreements and enabled to conclude that the residual chlorine loss curve and its geometry are useful tools to identify the presence of a contaminant "slug" and infer its reactive properties in adaptative contamination detections.

In water systems, sensitivity analyses are used for studying the estimation of physical parameters by empirical techniques. These analyses derive from and are highly dependent on the unsteady advection-reaction equations for quality modelling (Gancel et al., 2006). Gancel et al. (2006) proposed a splitting method for solving simultaneously the sensitivities and the advection-reaction equation that describes the quality modelling. The advection term is solved with an Eulerian scheme using a TVD criterion and the ordinary differential equations with reaction are solved with an implicit third-order Runge-Kutta scheme. The method was validated for a test network and proved to be suitable to take into account several types of concentration fronts (either smooth or sharp). In 2010, Fabrie et al. identified some weaknesses for improvement, mainly on three categories: the lack of measurements, the difficulty to estimate accurately the velocities and the complexity of the reaction. The new solution is more efficient in considering the unsteady behaviour of the system and takes into account inertial terms. Furthermore, the derivatives for calibration are directly computed for the reaction term, in a more accurate and efficient way. The model presented enables to perform the global sensitivity analysis of the system. The importance of the sensitivity analysis is also shown as part of the calibration process on a real network.

In the last 20 years, several authors studied the problem of simulation of the transport of contaminants in DWDSs. Several different numerical approaches were tested aiming to achieve accurate models for the advection-reaction system that describes the contaminant transport in DWDSs. Several works studied the interactions between substances and the pipe walls, usually as an additional reaction term. However, little work has been done regarding the incorporation of sorption phenomena in the governing equations. This work tries to give a contribution on this specific field of research.

2.2 Simulation Tools Available

There are several open-source and commercial software packages able to simulate DWDSs. These tools are mainly oriented for tasks of modelling and simulation of the hydraulic behaviour in DWDSs. Models for studying water quality behaviours, especially for the commercial packages, are add-in tools and closed boxes, which makes it difficult to understand their features but in general their simulation capabilities are limited.

This section presents several software packages used in modelling and simulation of DWDS, giving the available information for the characteristics of the software tools in what concerns to the quality aspects. A particular relevance was given to EPANET, since this specific software tool was widely used to provide the hydraulic data for the DWDS used as case studies in this thesis.

2.2.1 EPANET

This software performs extended period simulation of hydraulic and water quality behaviour in drinking-water distribution systems. These systems are built by pipes, nodes, pumps, valves, reservoirs and tanks (Rossman, 2000).

EPANET can model systems of any size, computes friction head loss using different equations - Hazen-Williams, Darcy-Weisbach or Chezy-Manning - (Rossman, 2000), allows minor head losses for bends or fittings and computes pumping energy and cost. EPANET models constant or variable speed pumps, various types of valves including shutoff valves, check valves, pressure regulating valves and flow control valves, storage tanks of any shape and allows multiple demand categories at nodes, each with its own pattern of time variation.

The basic data to be introduced in EPANET software are:

- For reservoirs: hydraulic head (equal to the water surface elevation if the reservoir is not under pressure;
- For junctions: elevation above some reference (usually mean sea level) and water demand (rate of withdrawal from the network);
- For tanks: bottom elevation (where water level is zero), diameter (or shape if non-cylindrical) and initial, minimum and maximum levels;
- For pipes: start and end nodes, diameter, length, roughness coefficient and status (open close, or contains a check valve);

- For pumps: start and end nodes, and pump curve (the combination of heads and flows that the pump can produce).

In addition to hydraulic modelling, EPANET provides water quality modelling capabilities. The quality models allow evaluating the movement of a non reactive tracer material through the network over time and the movement and the fate of a reactive material along the network. This software models the reaction mechanisms, both in the bulk flow and in the pipe wall, using several order kinetics to model reactions in the bulk flow and zero or first order kinetics for reaction at the pipe wall. The global reaction rate coefficients can be specific for each pipe and the wall reaction rate coefficients can be correlated with pipe roughness. It is also possible to determine the effects of concentration or mass input at any location in the network. The models available for storage tanks can simulate different behaviours, such as, complete mixing, plug flow and as two compartment reactors.

Todini's approach to a hydraulic node-loop system, also known as *Gradient Method*, is used by EPANET to solve the flow continuity and head loss equations which characterize the hydraulic state of the pipe network at a given point in time. The hydraulic head lost by water flowing in a pipe due to friction with the pipe walls can be computed using one of the following formulas: Hazen-Williams, Darcy-Weisbach and Chezy-Manning (Rossman, 2000).

The governing equations for EPANET's water quality solver are based on the principles of conservation of mass conjugated with reaction kinetics. The equations involve:

- Advective transport in pipes;
- Mixing in storage facilities;
- Bulk flow reactions;
- Pipe wall reactions;
- System of equations;
- Lagrangian transport algorithm;

EPANET also provides an integrated set of conditions for editing network input data, running hydraulic and water quality simulations, and viewing the results in a variety of formats, such as colour-coded network maps, data tables, time series graphs and contour plots. This software is available as freeware.

2.2.2 EPANET MSX

There are several types of water quality problems that cannot be accurately modelled by using the single-species capabilities of current EPANET program. EPANET-MSX is an extension to EPANET that enables it to model complex reaction schemes between multiple chemical and biological species in both the bulk flow and at the pipe wall.

This extension, which is also available as freeware, allows modelling chemical reactions such as auto-decomposition of chloramines to ammonia, the formation of disinfection by-products, biological re-growth, combined reaction rate constants in multi-source systems and mass transfer limited oxidation-pipe wall adsorption reactions (Shang et al, 2008).

The EPANET-MSX system is supplied as two different formats: a stand-alone console application (epanetmsx.exe) that can run standard water quality analysis without any additional programming effort and a function library (epanetmsx.ddl) that is used with the original EPANET function library (epanet2.ddl) to produce customised programming applications. In both formats, the user must prepare two input files to run a multi-species analysis. One of these files is a standard EPANET input file that describes the hydraulic characteristics of the network being analysed (EPANET-MSX will ignore any water quality information that might be in these file). The second file is a special EPANET-MSX file that describes the species being simulated and the chemical reaction/equilibrium model that govern their dynamics (Shang et al, 2008).

In the EPANET-MSX Manual nothing is mentioned about the introduction of a diffusion term into the mass balance equations. The only possibility would be the introduction of an expression in the "terms" section, which had to be an approximation to the diffusion term.

2.2.3 Porteau

PORTEAU is a tool used to model the behaviour of looped main networks distributing or transporting water under pressure. It provides a decision-making tool for designing and managing drinking water supply networks. The general principle is simple. It consists in creating a diagram of the network in question composed of pipes and nodes. The data about the different components of the system is entered to ensure that the complete network infrastructure and conditions of use are represented in order to stick as close to reality as possible.

For what concerns water quality, three types of results are supplied: the concentration of a product (chlorine for example) the age of the water and the origin of the water. The results can be displayed

either in the form of a table at each of the time steps (5 min to 1 h) for the whole network, or in the form of a table over the whole day by pipe or by node, or in the form of variation curves over the day by pipe and by node on the network diagram (URL: http://porteau.irstea.fr/).

The algorithm in PORTEAU for solving the reactive transport is a “Hybrid” method described in Fabrie et al (2010).This software is available as freeware.

2.2.4 Piccolo

Piccolo is a general software application that simulates flow problems in networks. PICCOLO works out the velocity, pressure and flow rates according to the network data. Simulations can be run for steady-state or dynamic conditions. A calibrated model can be used for master planning, case studies and water quality studies, such as influenced areas, chlorination, origin of water and blended sources. Advanced water quality studies may be carry out for risk assessment of bacterial re-growth, chloramines, nitrites and nitrification level assessment, settling / fouling rates assessment (URL: http://www.safege.fr/fr/nos-metiers/logiciels/logiciels/).

2.2.5 SynerGEE Water

SynerGEE Water is a simulation software package used to model and analyze water distribution systems. SynerGEE is highly flexible. It is possible to choose the level of detail from simple hydraulic analysis of a single pressure zone to the twin substance propagation of water quality in a multi-zoned system.

Additionally, SynerGEE can model complex control arrangements for pumps, valves and regulators in any operational scenario. Optional modules are available for more advanced modelling requirements such as area isolation, reliability analysis, subsystem management and calibration.

Compared to the Epanet algorithm for solving the hydraulics, the main difference is that when the change in flow, from previous iteration to the next, is less than one tenth of one percent of the total flow in the system, the network is considered solved. While SynerGEE solves and checks all variables to a tolerance, the Epanet solution is simply an average of flows to within 1/10th of one percent of all flows. There is no mass balance to double check nor is there any “forcing” to bring the solution within compliance.

For water quality, SynerGEE provides the unique capability to model two substances simultaneously within a distribution system. Both non-reactive (e.g. nitrates) and reactive (e.g. chlorine or THMs) substances can be simulated with a range of user-defined settings such as: Separate bulk and wall reactivities, limiting concentrations, source water reactivity blending, and temperature dependent reactivities (URL: http://www.gl-group.com/en/water/SynerGEEWater.php).

2.2.6 WaterGEMS

WaterGEMS (Water Distribution Modeling and Management) is a hydraulic and water quality modelling solution for water distribution systems with advanced interoperability, geospatial model-building, optimization, and asset management tools.

WaterGEMS is a comprehensive and easy to use water distribution modelling solution featuring interoperability across stand-alone, ArcGIS, AutoCAD, and MicroStation environments. From constituent concentration analyses, to energy consumption and capital cost management, WaterGEMS provides an environment for engineers to analyze, design, and optimize water distribution systems (URL: http://www.bentley.com/en-US/Products/WaterGEMS).

2.2.7 H2ONET

Spanning all platforms, from AutoCAD to ArcGIS to the Web, H2ONet products are stand-alone software for complete modelling, analysis, design, rehabilitation and optimisation of water distribution and supply systems. H2ONET helps in identifying the best combination of network improvements that meet target system 20 hydraulic design/performance criteria at minimum cost. It is possible to conceive and evaluate effective and economical design, rehabilitation, and enhancement alternatives for upgrading and modifying existing water distribution piping systems for improved performance. The H2ONET Designer allows determining cost-effective rehabilitation, replacement, strengthening, and expansion options to reliably supply projected demands at adequate levels of service, considering any modelling condition time frame (e.g. maximum day), multiple design scenarios (e.g., find the single optimum solution that provides the standard of service under peak day for normal operation + average day under a failure scenario), multiple fire flow scenarios, and complete extended period simulation (EPS) designs (e.g. 24-hour operational design). (URL: http://www.mwhsoft.com/page/p_product/net/net_overview.htm).

3 Contamination Sources Identification

The response to a contamination of DWDSs is a great challenge, in which protection and surveillance represent only one of the aspects to take into attention. The detection of water quality deterioration in DWDSs calls for the development of new, sensitive and rapid methods.

In case of deliberate or accidental contaminations of DWDSs, it is important to rapidly identify probable localizations of point sources of contaminations and subsequently the contaminated areas, allowing corrective actions to be performed, once the delimitation of affected areas is specified

The solution to this problem is obtained through inverse modelling techniques, which allow estimating the localization of point sources of contaminations. Some of these techniques consider the analysis of the concentration profiles in several check points along the network and several works have been published based on this strategy. In this chapter, these techniques are divided in two main groups: a) optimization based approaches; and b) other approaches.

3.1 Optimization Based Approaches

Standard simulation problems assume known contaminant injections and solve the propagation of the contaminant throughout the network. Simulation of the output state of a model based on known inputs is referred to as the forward problem. On the other hand, optimization based approaches try to find the unknown inputs that generate a partially known output state. The injection characteristics are unknown and are determined based on concentration measurements from the network. This kind of problems is inherently ill-conditioned and poses unique difficulties that are not present in the forward problem (Laird et al., 2005).

Van Bloemen Waanders et al. (2003) started to investigate the use of optimization techniques to locate potential contamination sources given a concentration and velocity profile. They considered the source inversion problem as a nonlinear programming problem, determining the unknown sources at the network nodes through the minimization of the differences between the calculated and target concentrations, using EPANET to simulate the behaviour of DWDSs. Furthermore, intrusive optimization and sensitivity analysis techniques were identified as suitable for evaluating the effect of various parameters in the computational effort. However, the performance of the method and the quality of the solution was affected by the source location. Laird et al. (2004a) improved this approach using a direct simultaneous approach that converge the network model and the optimization problems simultaneously. A similar work was also published in 2005 (Laird et al., 2005). These authors

presented a distinction between direct sequential and direct simultaneous methods. Direct sequential methods (van Bloemen Waanders et al., 2003) discretize the independent variables only evaluating the model and the objective function for each iteration, while direct simultaneous methods fully discretize all the unknown variables in the problem and solve the resulting system as a large scale optimization problem with algebraic constraints. The solution to the forward problem is converged only once, achieving significant computational gains over the direct sequential approach. The simultaneous approach required an explicit mathematical representation of the discretized water quality model. An algorithm similar to the particle backtracking algorithm, presented by Zierolf et al. (1998) and extended by Shang et al. (2002), was presented for reducing the water quality down to its essential elements, calculating impact coefficients that describe the concentration of selected outputs as functions of network sources and tank concentrations. This algorithm reformulates the pipe constraints of the water quality models reducing the size of the discretized problem and provides a straightforward mathematical representation of the pipe boundary concentrations. They also identified some important areas of future work, such as the determination of optimal sensor location, the testing of the reliability of the formulation against sensor failure or noise in flow rates or sensor measurements and the correct description of the reaction rate in pipes. Some results were presented for a real municipal DWDS with almost 500 nodes to demonstrate the effectiveness of the proposed simulation. However Laird et al. (2004b) verified that this approach did not scale up indefinitely to very large DWDSs. Thus, they presented a dynamic optimization approach for solving the problem of contamination sources identification for very large DWDSs by performing the optimization on a smaller subdomain of the DWDSs. The approach considers the hydraulic behaviour and the sensor measurements of the entire DWDS but formulates the dynamic optimization problem for a subset of nodes. The results proved that this approach was effective and is able to scale up for very large DWDSs. Laird et al. (2006) presented a mixed-integer quadratic program to refine the solution developed by Laird et al. in 2005. Due to the sparseness of the sensor grid, the problem of contamination source identification has no unique solutions. This method estimates the number of likely injection locations and tries to extract the likely scenarios from the family of no unique solutions. This two-phase approach was tested on a realistic municipal water network model and achieved good agreements between the set of possible injection scenarios calculated and the actual simulated injections.

Mann et al. (2012) presented a mixed-integer linear programming formulation for the problem of contamination source identification using discrete (yes/no) measurements available from sparse manual grab samples at limited points in time and space. An origin-tracking approach was used to develop the water quality model, which was efficiently and exactly reduced prior to the formulation of the contamination source identification problem. The results showed that this method solve efficiently

a real-time setting on large networks with over 10000 nodes, considering over 150 time discretizations.

Preis and Ostfeld (2006) presented a coupled model trees-linear programming algorithm for contaminant source identification in water distribution systems. Several contamination scenarios were developed in EPANET (Rossman, 2000) to define a model trees linear rule classification structure. Linear programming was then used to solve the inverse problem of contamination source identification. This method provided an estimation of the time, location, and concentration of the contamination sources. Later, Preis and Ostfeld (2007) addressed the challenge of achieving unique solutions for water distribution systems of large size, coupling genetic algorithms with EPANET. The objective function considered was the minimization of the least-squares of the differences between simulated and measured contaminant concentrations, with the intrusion location, starting time, duration and mass rate as decision variables. The developed methodology was demonstrated through base runs and sensitivity analyses using three water distribution systems examples of increasing complexity. The main limitations were the highly intensive computational effort required by this method and the assumptions that the flows in the pipes are known and the monitoring stations are perfect. In 2010, they investigated the effect of uncertainties in the sensor measurements in the problem of contamination sources characterization and presented a modified genetic algorithm scheme (Preis and Ostfeld, 2010). The developed model was implemented using three sensor types: perfect sensors, sensors transmitting fuzzy measured information and sensors indicating only a contamination presence. Two example applications of increasing complexity were presented to demonstrate the performance of the proposed methodology, showing the trade-offs between the sensor types and the model abilities to receive a unique solution to the source identification problem.

The effects of the uncertainties in the management of water resources systems have been widely studied (Chung et al, 2009; Li et al., 2008; Li et al., 2010; Li et al., 2011; Qin et al., 2007) and Dong et al. (2012) have presented a review of the existent techniques for the development of scenarios in water research management. Torres et al. (2009) have shown that an understanding of hydraulic uniqueness of water distribution systems is important for building robust risk models and/or performing vulnerability assessments and reasonable uncertainties in model inputs produce high degrees of uncertainty in estimated exposure levels. Thus, the parameter uncertainty is a source of error that may create large disparity in water distribution, particularly in the water demand, which can fluctuate widely and is unpredictable (Huang and McBean, 2009). These uncertainties may be derived from random feature of resource conditions and natural processes, errors in estimated modelling parameters or imprecision or fuzziness human-induced (Li et al., 2011). In fact, the modelling approaches are not able to provide an exact representation of the contaminant spread in DWDSs. These approaches

provide an approximation that is always affected by modelling uncertainty and, therefore, cannot be handled by the available model calibration approaches (Huang and McBean, 2009).

Preis and Ostfeld (2011) considered that for real applications only a small portion of the hydraulic data is known and the only information available might be a binary sensor status. Thus, they presented a methodology for the inclusion of hydraulics uncertainty in contamination source identification. The proposed method is based on a previous contamination source detection model developed by the authors coupled with a statistical framework for quantifying the uncertainty of a contamination source detection outcome. The results showed that in all case scenarios that were evaluated the contamination area and the approximated time of injection was revealed.

Guan et al. (2006) presented a simulation-optimization method to solve a nonlinear contaminant source and release history identification problem for a complex water distribution system. This method used the EPANET software application to simulate the occurrence of contaminations. Results for arbitrarily selected monitoring locations were used in a continuous optimal predictor-corrector algorithm to identify the contamination sources and their release histories. The optimization model used as corrector was designed to identify the similarity between the simulation response and measured data at the monitored sites. Results showed that the approach was effective, efficient and robust in identifying contamination sources and their release histories.

Hill et al. (2006) presented a least squares formulation for the problem of localization of contamination sources, considering only Boolean type sensors that were considered to be closer approximation to real sensors. A comparison was made between the solutions achieved considering continuous readings and Boolean measurements to demonstrate de robustness of the method. The results demonstrated that it was possible to achieve good results for the localization of contamination sources since there wasn’t any relation between the network time delays and the magnitude of the concentrations.

Di Cristo and Leopardi (2008) formulated a methodology for identifying the source location as an optimization problem, linearized using the water fraction matrix concept. The method starts from the concentration data to select a group of candidate nodes, from which the source location is identified minimizing the differences between simulated and measured concentrations. An uncertainty analysis was presented to demonstrate the methodology robustness against uncertainties in concentration measurements and water demands.

Vankayala et al. (2009) also studied the problem of contamination source identification, under water demand uncertainty. A simulation-optimization approach was presented that used EPANET as simulator and a stochastic and a noisy genetic algorithm (GA) as optimizers, minimizing the

difference between the simulated and observed concentrations at the sensor nodes. The noisy GA proved to be more robust and required less computational effort to achieve a solution for the contamination source identification.

Mou et al. (2010) applied a simulation-optimization method to backtrack contamination sources in a laboratory water distribution system, using sodium hypochlorite solution as a substitute for contaminants. Results for different input parameters were compared, such as water network topology structure and simulation time horizon, and influencing factors of this algorithms and their effects were discussed. Tryby et al. (2010) formulated the monitoring design problem as a nonlinear combinatorial optimization problem and solved using a genetic algorithm. The source inversion performance of an optimized monitoring network relative to networks designed using different methods was evaluated. The results demonstrated that the number of sensors installed in the network is more important than the method used to locate them when the source identification problem is underdetermined.

Propato et al. (2010) presented another approach for solving this problem based on two main steps. First, linear algebra was employed to select the potential contamination sources through contamination source pruning. Second, the minimum relative entropy method was used for evaluating the probability of each potential contamination source. The solution was a space-time contaminant concentration probability density function accounting for the various possible contamination sources that may be responsible for the data registered at the sensors.

Liu et al. (2011) presented an adaptive dynamic optimization technique for identifying the contamination source in real time following a contamination event, through a multiple population-based search that uses an evolutionary algorithm. The results showed that the algorithm adaptively converges to the best solutions, given the observed data.

Cugat (2012) considered a DWDS where the contaminant intrusion could occur at a limited number of nodes. The method is based on the data collected at a defined set of sensors. The corresponding infinite-dimensional optimization problem was defined in a Hilbert space setting, with the addition of a quadratic regularization term is added in the objective function to guarantee the obtainment of a unique solution. Under certain assumptions, the computation of the solution on a discrete time grid is performed by solving finite-dimensional linear least squares problems This method was considered useful to minimize potential impacts of contamination emergencies on consumers by helping to select locations to flush the contaminant out of the distribution network.

3.2 Other Approaches

Davidson et al. (2005) proposed two methods that use supervisory control and data acquisition to create connectivity matrixes that contain the worst-case projection of the potential spread of contamination obtained by combining the effects of all possible scenarios. The first creates the connectivity matrices based on operating modes and the second on fundamental paths. Results showed that the methods had similar results, however the method of fundamental paths was more efficient computationally.

Neupauer et al. (2005) presented a backward modelling approach that uses probability density functions to identify the source node and release time of a contamination. The probability density functions are obtained from the data collected an installed set of sensors. The backward model proved to be effective for steady flow conditions and for a single, instantaneous source of contamination.

Dawsey et al. (2006) proposed a methodology based on Bayesian belief network (BBN) for the characterization of contamination events, reducing false positive sensor detections in DWDSs. The methodology uses distribution system simulations and conservative transport to estimate conditional prior probabilities for contaminant introductions. . In addition, the simulations identify the upstream nodes that are more likely to result in positive detections. They presented a case study to show how data from sensors and other sources can be interpreted with a BBN to distinguish between routine false positive sensor detection and a true system contamination event.

Khanal et al. (2006) introduced a dimensionless exposure index as a simple global measure of network response. The fraction of the population at risk of contaminant exposure was estimated at the end of a 72 h simulation period for several contamination scenarios. Simulation results were used to categorize network injection nodes on the basis of their potential to expose downstream consumers. Furthermore, a generalized sensitivity analysis was performed to determine the sensitivity of network response to four dynamic network variables. The exposure levels were more sensitive to variations in base demand and injection mass. Tank storage capacity was important in certain cases, while injection duration tended to be least important.

Quesson et al. (2009) investigated the application of acoustic sensors for the detection of contamination events, through the classification of aberrant sounds in the DWDSs. Models were developed for the sound propagation in PVC pipes and computer simulations were carried out to specify detection ranges and field experiments were carried out to characterise background noise and suspicious sounds. An acoustic monitoring demonstrator was developed and tested for the detection and classification of aberrant sounds and it has shown promising system results for protection of buildings.

ANNs have been also used in the identification of contamination sources in water systems. Kim et al. (2008) applied ANNs models to identify the contamination source in case of an accidental or deliberated release of Escherichia coli 15597 in a water system. Results showed that dispersion patterns of E. coli were positively correlated to pH, turbidity and conductivity. The ANNs models identified successfully the contamination source with 75% accuracy based on the pre-programmed relationships between E. coli transport patterns and contamination sources. However, as far as it is known, this is the only published work that uses ANNs in helping to predict contamination source locations in DWDSs.

Huang and McBean (2009) presented a method to identify the location and time of an intrusion event, based on limited sensor data through the application of a data mining approach in conjunction with a maximum likelihood procedure. They also demonstrated that uncertainties in water demand, sensor measurement, and modelling, are highly relevant and necessary to be considered in the contamination identification problem. The proposed method took 3 minutes to identify multiple injections for a 285 node water distribution network with five sensors. According to these authors, this approach is suitable to identify the contamination source for a simple water network; for a large and complex system, more sophisticated algorithms may be required.

Yang et al. (2009) explored a real-time event adaptive detection, identification and warning methodology based on the information collected by conventional water quality sensors. They performed several pilot-scale pipe flow experiments with different chemical and biological contaminants at different concentration levels. Contaminant signals were enhanced and background noise was reduced in time-series plots, through adaptive transformation of the sensor outputs, leading to detection and identification of all simulated contamination events. Then, the relative changes calculated from adaptively transformed residual chlorine measurements were quantitatively related to contaminant-chlorine reactivity in drinking water. The results showed that the tested contaminants were distinguishable based on kinetic and chemical differences.

Zechman and Ranjithan (2009) described a specific implementation of a method using evolutionary strategies, a population-based heuristic global search algorithm. The method was constructed using a tree-based encoding design to enable the representation of decision vectors and a set of associated genetic operators that enabled an efficient search. Results showed that the algorithm had good performance and showed a robust behaviour for several examples of water distribution systems.

De Sanctis et al. (2010) developed a method that only required a binary sensor status over time. This method consisted in a particle backtracking algorithm to identify the water flow paths and travel times leading to each sensor measurement. It was assumed that the locations and times connected to positive

sensor measurements and not connected to negative measurements were the possible sources, assuming no false positive/negative readings and an accurate hydraulic model. The output for any node and time interval is of three kinds: safe (not a possible source); unsafe (possible source); or unknown (insufficient data to determine, thus could be safe or unsafe).

Koch and McKenna (2011) proposed an approach for combining data from multiple sensors to reduce false background alarms. They used the Kulldorff's scan test to find statistically significant clusters of detections, considering the location and time of individual detections as points resulting from a random space-time point process. The results showed that the scan test can detect significant clusters of events, reducing the occurrence of false alarms caused by background noise by three orders of magnitude using the scan test. The clusters can also help to indicate the time and source location of the contaminant.

Di Nardo et al. (2012) simulated a simple backflow attack with cyanide being introduced into a real-water system, defining the most dangerous introduction points for a contaminant incident. The vulnerability of the DWDS was analysed by computing the lethal dose of cyanide ingested by users and the total pipe length in the DWDS and the effects of network partitioning and district isolation was assessed. Results demonstrate the sectorization of the DWDS can significantly decrease contaminant spread and protect part of the users from the contaminated water. However, some simulations also showed that this procedure could have negative effects in terms of the exposition to contaminants, thus it was concluded that further investigation was required to design water districts for DWDS security and safety.

Eliades and Polycarpou (2012) proposed a computational approach based on decision trees for selecting a sequence of nodes in the DWDS to perform expanded sampling, for evaluating the water contamination impact and isolating the source-area with as few quality samples taken as possible. A simplified and a benchmark water distribution system were used to demonstrate the functioning of the proposed procedure.

Liu et al. (2012a) presented a method for a contamination characterization algorithm by coupling a statistical model with a heuristic search method. The statistical model is used to identify potential locations for the contaminant intrusion and the heuristic search method aims at further refining contaminant source characteristics. Two illustrative examples of DWDSs demonstrated the ability of the method in adaptively discovering contaminant source characteristics as well as evaluating the degree of non-uniqueness of solutions.

Liu et al. (2012b) introduced a hybrid method for the real-time characterisation of a contaminant source, given sensor measurements in DWDSs. This method integrates a simulation-optimisation

approach with a logistic regression and a local improvement method to accelerate the convergence. The results of numerical experiments demonstrate the efficiency of the proposed hybrid method for contaminant source characterisation.

Tao et al. (2012) proposes a rapid identification methodology for determining the location, starting time, and injection rates at different time intervals in a DWDS. The proposed methodology identifies the key characteristics of the contamination by matching the dynamic patterns of the simulated and measured concentrations. Results showed that if the data collected by the sensor is minimal, a greater number of redundant contamination source nodes will be present. More data is necessary to effectively use this method for locating multiple sources of contamination in DWDSs.

Shen and McBean (2012) stated that the identification of contamination sources had two major issues: the occurrence of false negatives, in which the method fails to identify the true contamination source, and false positives, when the method wrongly identifies a location that was not the true contamination source. These authors presented a data mining procedure based on a database constituted by the first-detection times at the sensors. Results showed that increasing the number of scenarios in the database always reduces the false-negative rate of each sensor, and usually reduces the number of false positives.

3.3 Chapter Conclusions

The works mentioned above applied a wide range of approaches to the problem of the localization of contamination sources in DWDSs, which indicates that this important research field is not yet consolidated. A limitation associated to some of these approaches is that the time of computation might be very high, due to the complexity of the mathematical formulations. Although some of the works mentioned require a reasonably low time of computation, they depend on the analysis of the concentration profiles and make the initial assumption that the sensors should be able to evaluate the contaminant concentration, which might be difficult to achieve in real scenarios.

Furthermore, the majority of the published work assumes a well-calibrated hydraulic model with a known water demand at each node at certain time. The challenges concerning parameter uncertainty are in general not addressed, although some of these works have shown applications to deal with measurement uncertainty.

Additionally, the majority of the these works do not present detailed information concerning the level of restriction achieved for the location of the possible contamination sources and the deviation between the estimates obtained by the methods and the real times of contamination.

This thesis tries to make a contribution presenting different strategies to overcome some the weaknesses here identified.

4 Artificial Neural Networks

Artificial Neural Networks (ANNs) are mathematical models which have a distributed and parallel architecture, consisting of processing units, analogous to neurons, with multiple connections, analogous to dendrites and axons. Figure 4.1 presents the analogy between ANNs and biological neural networks. ANNs try to mimic the brain capacities in two aspects: acquiring the knowledge from its environment by a learning process and storing the knowledge acquired in the learning process by the interneuron connection strength, also known as synaptic weights (Haykin, 1999).

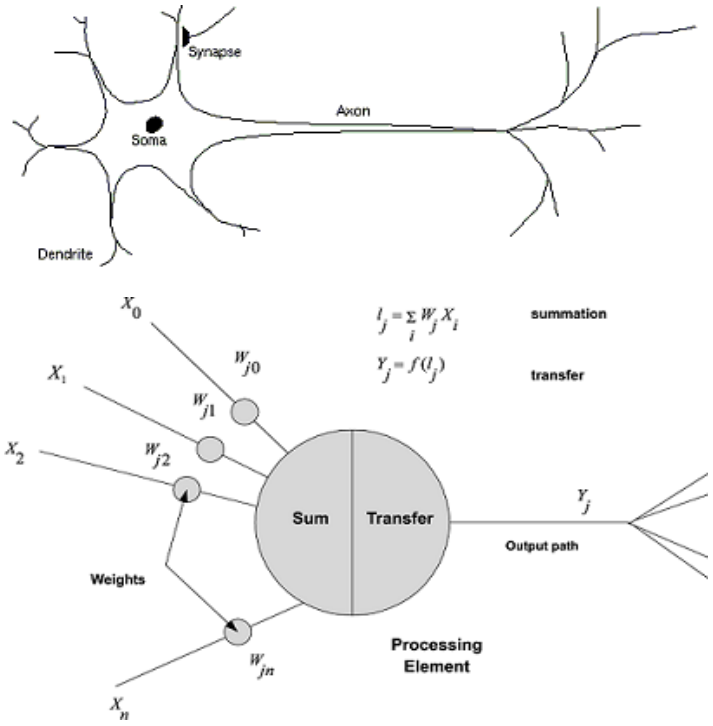

Figure 4.1 – ANNs versus biological neural networks.

ANNs are developed based on the assumption that:

- Information processing occurs at the neurons;´
- Signals are passed between neurons over connection links;
- Each link as an associated weight;
- Each neuron applies an activation function to its input to determine its output signal (Fausett, 1994).

An ANN is mainly characterized by its architecture, learning process and activation functions (Fausett, 1994). These characteristics are present, with more detail, in Sections 4.1, 4.2 and 4.3.

4.1 Architectures

Often, it is convenient to visualize neurons as arranged in layers, where neurons behave in the same manner. Within each layer, neurons usually have the same activation function and the same pattern of connection to other neurons (Fausett, 1994).

The arrangement of neurons into layers constitutes the network architecture. In general, it is possible to identify three main classes of architectures:

- Single-Layer Feedforward Networks;
- Multilayer Feedforward Networks;
- Recurrent Networks.

4.1.1 Single-Layer Feedforward Networks

A single-layer network has one layer of connection weights. Thus, there is an input layer that is connected to an output layer of neurons, but not vice versa. The network is strictly a feedforward type (Haykin, 1999). Figure 4.2 shows a schematic representation of this type of architecture.

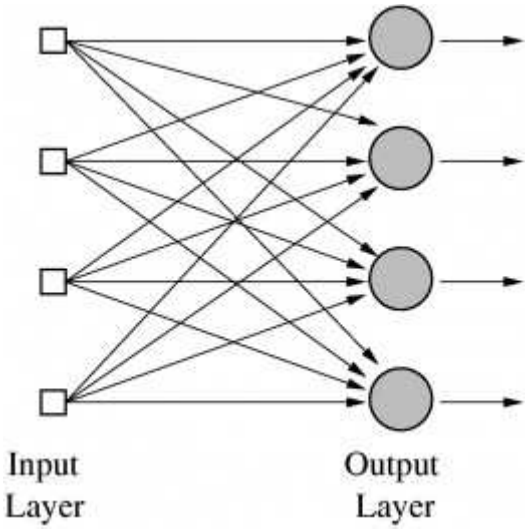

Figure 4.2 – Representation of a single-layer feedforward network.

4.1.2 Multilayer Feedforward Networks

A multilayer feedforward network is constituted by one or more layers of nodes (usually denominated as hidden layers) between the input and the output layers. Multilayer networks can solve more complicated problems than can single-layer networks, but the training may be more difficult (Fausett, 1994). By adding one or more hidden layers, the network is enabled to extract higher-order statistics, which can be a valuable ability when the size of the input layer is large. These ANNs are commonly referred to as multilayer perceptrons (MLPs); (Haykin, 1999).

Typically the neurons in each layer of the network have as their input the outputs of the previous layer. The set of outputs of the neurons in the output layer constitutes the overall response of the network to the inputs provided to the input layer. Figure 4.3 shows the schematic representation of a multilayer feedforward network with one hidden layer.

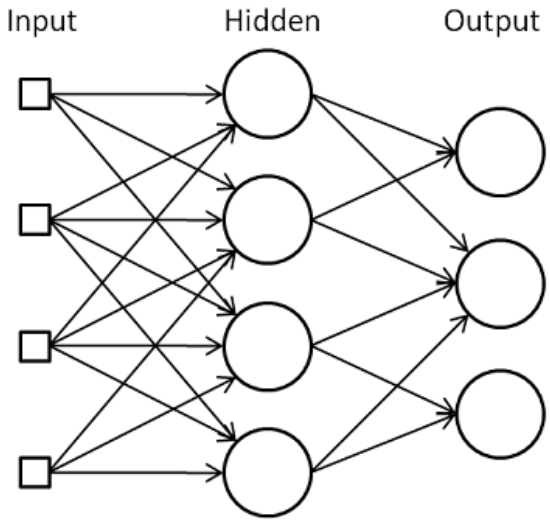

Figure 4.3 – Representation of a multilayer feedforward network.

4.1.3 Recurrent Networks

The main difference between the recurrent networks and single or multilayer feedforward networks is the existence of at least one feedback loop, regardless the presence of hidden nodes. The presence of feedback loops has great impacts on the performance and learning capability of the network (Haykin, 1999).

Figures 4.4 and 4.5 show examples of recurrent networks with single and multiple layers, respectively.

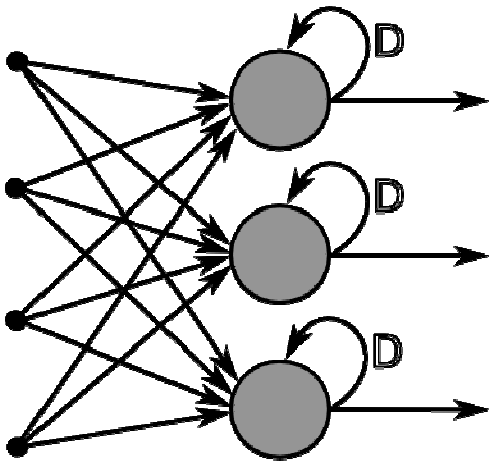

Figure 4.4 – Representation of a recurrent network with a single layer.

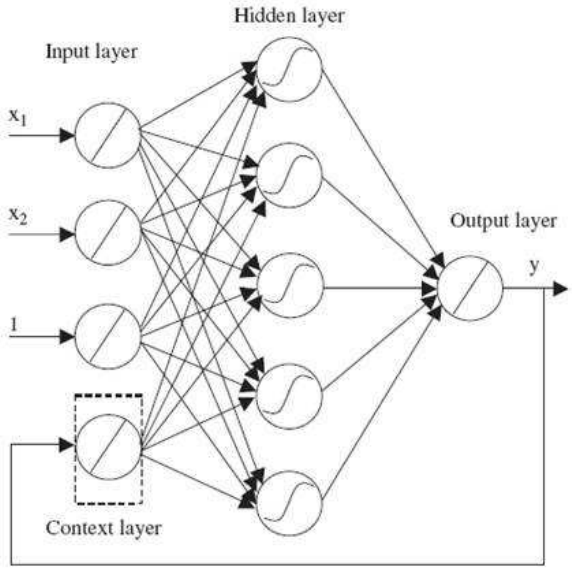

Figure 4.5 – Representation of a recurrent network with multiple layers.

4.2 Activation Functions

The basic operation of the ANNs is based on the processing of information performed by neurons, presented in Figure 4.6. A neuron, receives a set of connection links with associated synaptic weights. Then, the activation function is applied to the sum of all input signals, weighted by each corresponding weight. Some neurons, as it is also shown in Figure 4.6, include a bias, which has the effect of increasing or decreasing the input to the activation function (Haykin, 1999).

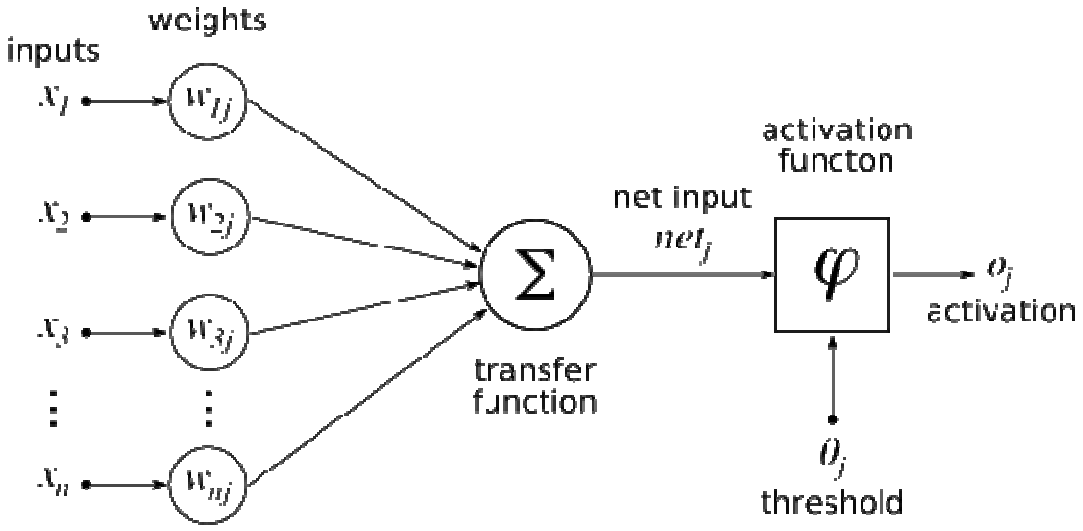

Figure 4.6 – Schematic representation of a neuron.

There are several choices for the activation functions, which can be divided in three main types: threshold functions, piecewise-linear functions and sigmoid functions (Haykin, 1999).

4.2.1 Threshold Functions

With this type of activation function, which is defined by Equation 4.1 and is described in Figure 4.7, the output of a neuron is 1 for non-negative values and 0, otherwise.

$$\varphi(v) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{4.1}$$

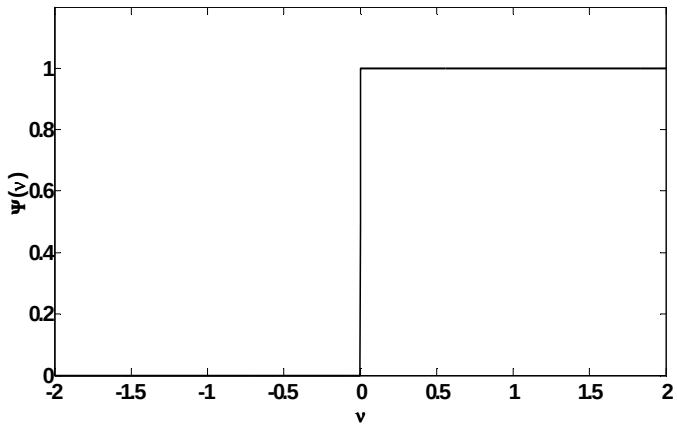

Figure 4.7 – Threshold function.

4.2.2 Piecewise-linear Functions

An example of a piecewise-linear function is given by Equation 4.2 and described in the Figure 4.8. This type of activation function is a linear combiner in the linear region but reduces to a threshold function if the amplification factor υ is infinitely large.

$$\varphi(v) = \begin{cases} 0, & v < -\frac{1}{2} \\ \frac{1}{2} + v, & -\frac{1}{2} > v > +\frac{1}{2} \\ 1, & v \geq +\frac{1}{2} \end{cases} \tag{4.2}$$

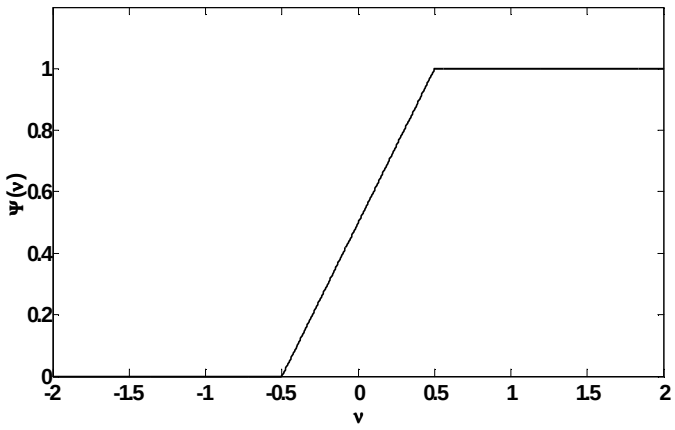

Figure 4.8 – Piecewise-linear function.

4.2.3 Sigmoid Functions

Sigmoid functions are the most common form of activation function used in the development of ANNs. It is defined as a strictly increasing function that presents a balance between linear and nonlinear behaviour. Log-sigmoid is an example of this type of function. The parameter *a* of Equation 4.3 is the slope parameter of the sigmoid function. In the limit, the sigmoid function becomes simply the threshold function.

$$\varphi(v) = \frac{1}{1+e^{-a\,v}} \tag{4.3}$$

The examples of activation functions presented so far vary between 0 and 1. The hyperbolic tangent function (also known as tan-sigmoid) has a different behaviour, since it varies between -1 and +1, which might be desirable for some applications. This function is given by equation 4.4 and is described by Figure 4.10.

$$\varphi(v) = \tanh(v) \tag{4.4}$$

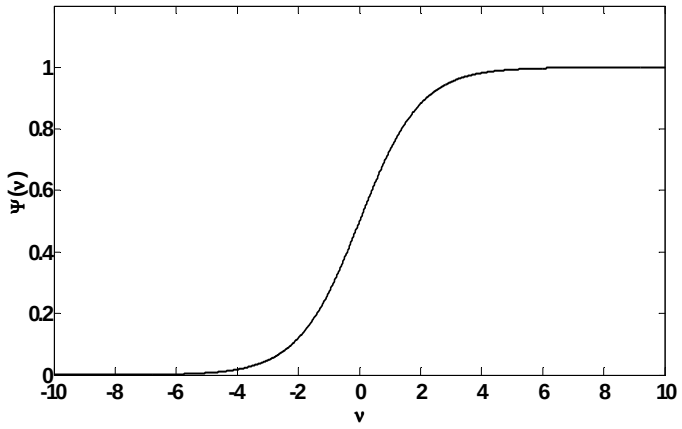

Figure 4.9 – Log-sigmoid function.

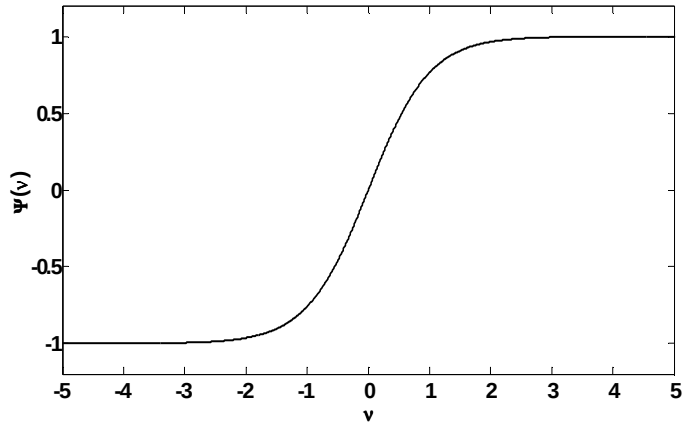

Figure 4.10 – Tan-sigmoid function.

4.3 Learning Process

Another very important characteristic of the ANNs is the training, which is the process of setting the values of the synaptic weights. There are two main types of training: supervised and unsupervised, whose basic features are presented in the following sections.

4.3.1 Supervised Training

In the supervised training, training is accomplished by providing a set of input vectors with its associated targets for the output vector. The network weights are then adjusted according to a learning algorithm (Fausett, 1994).

The most popular algorithm is known as the error back-propagation algorithm, and it has been widely applied to MLPs for solving some difficult and diverse problems. The objective of this training method, as it is the case with most ANNs, is to develop an ANN able to respond correctly to the input patterns used in the training (memorization) and to generalise its knowledge for inputs that are similar, but not identical, to that used in the training (generalization) (Fausett, 1994). Most training algorithms involve an iterative procedure for minimization of an error function, adjusting the weights following a sequence of steps. These steps are constituted by two distinct stages. In the first stage, the derivatives of the error function with respect to the weights are evaluated. The important contribution of the back-propagation technique is in providing a computationally efficient method for evaluating such derivatives. In the second stage, the derivatives are then used to compute the adjustments to be made to the weights. (Bishop, 1995).

The learning process is carried out by an optimization process where, based on the training data or examples, weights and biases are updated looking for minimizing the error between the predicted variable and the real one. After the training phase, a validation phase must be conducted to evaluate the generalization capacity of the final network configuration. This evaluation is done through a cross-validation test (usually the data set is divided into three subsets, for training, validation and test). Even though, it is important to remark that the optimization problem is not convex, and for this reason the optimization process is not a simple task. On the other hand, the generalization or prediction power of the network will be better when the training data correspond to a wide range of problem situations.

After the training and validation phases, ANNs are able to classify input patterns to an acceptable level of accuracy even if they were never used during the training process (Ham and Kostanic, 2001).

4.3.2 Unsupervised Training

In unsupervised training, a set of inputs are provided but no target vectors are specified. The adjustment of the synaptic weights is performed obeying to a set of learning rules. From numerous originally random local interactions within the ANN, in response to the input patterns, there emerges global order (Ham and Kostanic, 2001).

4.4 Development of ANNs

Several authors summarized the main steps that should be performed in the application of ANNs, from conceptualization to design and implementation (Basheer and Hajmeer, 2000; Maier et. al, 2000; Maier et al., 2010).

In 2000, Basheer and Hajmeer outlined some issues that should be addressed before beginning with training of an ANN, with focus on ANNs trained with a back-propagation algorithm:

- Database size and partioning;
- Data preprocessing, balancing and enrichment;
- Data normalization;
- Input/output representation;
- Network weight initialization;
- Back-propagation learning rate;
- Back-propagation momentum coefficient;
- Transfer function;
- Convergence criteria;
- Number of training cycles;
- Training modes;
- Hidden layer size;
- Parameter optimization.

Maier and Dandy (2000) and Maier et al. (2010) noted that the methods for developing ANN models are not yet well established. These authors have outlined some steps and guidelines for applying these modelling techniques in an efficient and reliable way, introducing taxonomies of approaches for each one:

- Input selection;
- Data division;
- Model architecture selection;
- Model structure selection;
- Model Calibration;
- Model evaluation;

The development of ANNs performed for some of the methods proposed in this thesis followed a procedure similar to those presented by Maier and Dandy (2000) and Maier et al. (2010).

4.5 Main Advantages of ANNs

The greatest advantage of ANNs over traditional modelling techniques is their capability to model complex, non-linear processes without having to express a relationship between input and output variables (Chau, 2006). This reflects a different approach for computing when compared to other methods, which involve the development of computer programs. In a computer program, the programmer has to specify every step executed by the computer and it is a process that takes time and resources. On the other hand, ANNs are able to provide the correct results based on training examples, not being necessary to program every theoretically possible case that may occur in the problem (Dayhoff, 1990).

Besides the qualities mentioned before (adaptability and nonlinearity), ANN have another important one, its robustness and resistance tending to be fault-tolerant. This means, that even if a neuron or a connection is damaged, it will only decrease a little the overall performance of the network, because the information is distributed in total neural network. These reasons have contributed to become ANN as a powerful tool used in different interdisciplinary areas (Ham and Kostanic, 2001).

Computational structures based on biological systems may grant superior performances for certain problems, which include labelling problems, scheduling problems, search problems and other constraint satisfaction problems. These problems are characterized by some or all of the following properties: a high-dimensional problem space; complex, unknown or mathematically unmanageable relationships between problem variables; and a solution space that might be empty, contain a unique solution or, most typically, contain more than one solution.

4.6 Main Applications of ANNs

During the last decades, there has been a substantial increase in the application of ANNs in the solution of a wide range of problems. These mathematical models are very suitable for tasks involving incomplete-data sets, fuzzy or incomplete information, and for highly complex and ill-defined problems, where humans usually decide on an intuitional basis. The most important tasks in which ANN can be applied may be divided in classification, forecasting, control systems and optimisation and decision making (Kalogirou, 2000).

Several works were found in very different fields, such as process control (Willis et al., 1992; Martins and Coelho, 2000; Hussain and Ramachandran, 2006; Peng et al., 2007; Ławrynczuk, 2009), image processing (Egmont-Petersen et al., 2002; Li et al., 2002), medicine (Khan et al., 2001; Lisboa, 2002;

Ahmed, 2005; Di Luca et al., 2005; Lisboa and Taktak, 2006), speech production and recognition (Tebelskis, 1995; Altun et al., 2003; Al-Alaoui et al., 2008; Dede and Sazli, 2010; Abe and Nakamatsu, 2012), engineering (Farkas and Géczy-Víg, 2003; Cerri et al., 2006; Singh et al., 2007; Souliotis, 2009;), modelling and forecasting (Dibike and Coulibaly, 2006; Gómez- Sanchis et al., 2006; Sousa et al., 2006; Argiriou et al., 2007; Sousa et al., 2007), and business applications (Li, 1994; Poh et al., 1998; Quaddus and Ktinn ,1999; Kamruzzaman, 2004).

ANNs have been widely applied for the prediction and forecasting of water resources and other environmental processes, due to their great flexibility of implementation to accurately represent the behaviour of relatively poorly understood processes such as the complex and non-linear dynamics of water quality within water distribution systems Bowden et al., 2006; May et al., 2008a; May et al., 2008b; Seródes, et al., 2001) and the problem of water-demand forecasting for real-time operation of water supply systems (Odan and Reis, 2012).

In 2000, Maier and Dandy presented a review of 34 papers dealing with the use of neural network models for the prediction and forecasting of water resources variables. They concluded that ANNs have the potential to be a useful tool for the prediction and forecasting of water resources variables. However, they also identified a need to develop guidelines which identify the circumstances under which particular approaches should be adopted and how to optimise the parameters that control them. In 2010, Maier et al. presented another review of 210 papers, published between 1999 and 2007, which focus on the prediction of water resource variables in river systems. They concluded that methods used for determining model inputs, appropriate data subsets and the best model structure are generally obtained in an ad-hoc fashion and, despite a significant amount of research activity on the use of ANNs, there is still a need for the development of robust ANN model development approaches. They also concluded that multilayer perceptrons are the most popular model architecture, while gradient based methods are used almost exclusively.

ANNs have been also used in the identification of contamination sources in groundwater (Sahoo et al., 2005; Sahoo et al., 2006; Singh and Datta, 2006) and in DWDSs (Kim et al., 2008). However, as far as it is known, the paper published by Kim et al., already mentioned in the previous chapter, is the only published work that uses ANNs in helping to predict contamination source locations in water systems.

5 DWDSs used as case studies

Several examples of DWDS, with different size and level of complexity, were used as examples to demonstrate the performance of the methods presented in this thesis. This chapter presents a brief description of each DWDS used as case study.

5.1 Network A

Figure 5.1 presents a theoretical example of a DWDS, which was used in a preliminary phase of this work The hydraulic characteristics of the network are listed in Tables 5.1, 5.2 and 5.3. The demand of each node is determined by multiplying its base demand by the respective demand pattern presented in Table 5.3. The hydraulic behaviour of the Network A was analysed, for a period of 6 hours, using the EPANET Programmer's Toolkit in Visual Basic and the hydraulic data was exported to MATLAB. Table 5.4 presents the definition of the different time periods performed by the hydraulic analysis of the EPANET.

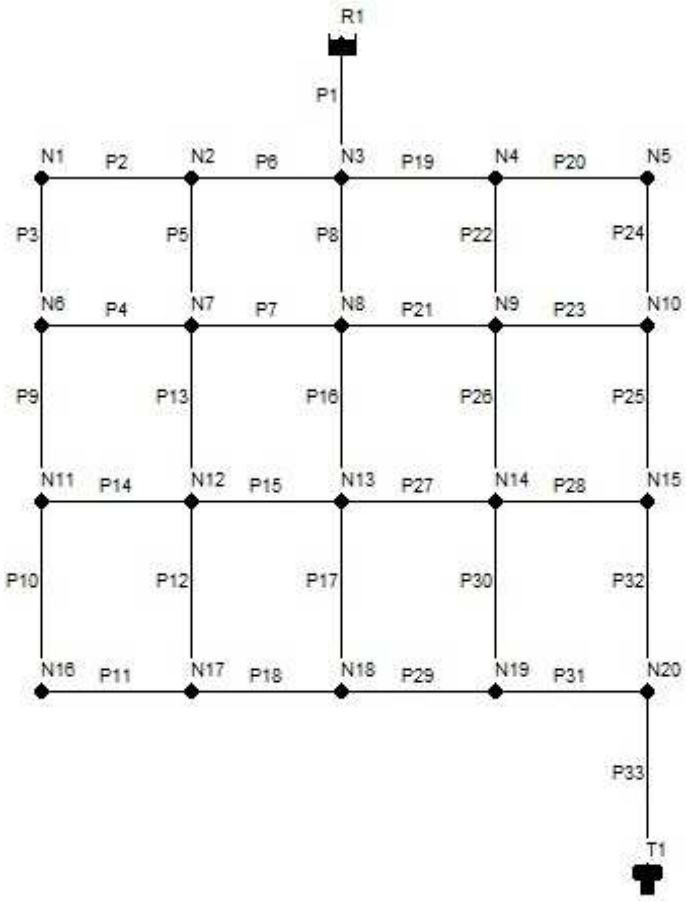

Figure 5.1 – Representation of Network A.

Table 5.1- Network A nodes characteristics.

Node	Elevation (m)	Base demand (m^3/h)	Pattern
R1	600	-	-
N1	480	10	1
N2	380	10	1
N3	280	10	1
N4	180	10	1
N5	80	10	1
N6	460	15	2
N7	360	15	2
N8	260	15	2
N9	160	15	2
N10	60	15	2
N11	440	5	3
N12	340	5	3
N13	240	5	3
N14	140	5	3
N15	40	5	3
N16	420	10	4
N17	320	10	4
N18	220	10	4
N19	120	10	4
N20	20	10	4
T1*	0	-	-

Note: Tank T1 has a diameter of 50 m, the initial level of water is 10 m and the minimum and maximum levels are 10 m and 50 m, respectively.

Table 5.2 - Network A pipes characteristics.

Link	Length (m)	Diameter (mm)	Link	Length (m)	Diameter (mm)
P1	100	400	**P18**	150	200
P2	200	300	**P19**	100	250
P3	200	300	**P20**	100	250
P4	200	300	**P21**	100	250
P5	200	12	**P22**	200	250
P6	150	300	**P23**	100	250
P7	150	300	**P24**	200	250
P8	200	300	**P25**	150	200
P9	150	250	**P26**	150	200
P10	100	200	**P27**	100	200
P11	200	200	**P28**	100	200
P12	100	200	**P29**	100	200
P13	150	250	**P30**	100	200
P14	200	250	**P31**	100	200
P15	150	200	**P32**	100	200
P16	150	12	**P33**	100	400
P17	100	200	-	-	-

Table 5.3 - Demand patterns (Network A).

Pattern\Time period	[0, 3] h	[3, 6] h	[6, 9] h	[9, 12] h
Pattern 1	0.5	1	0.5	1
Pattern 2	0.4	0.8	1.2	0.6
Pattern 3	2.5	1.25	2	1
Pattern 4	0.5	1.5	2.5	1

Table 5.4 - Duration of time periods (Network A).

Time period	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Start time (s)	0	3600	7200	10800	14400	14462	18000
End Time (s)	3600	7200	10800	14400	14462	18000	21600

5.2 Network B

Figure 5.2 presents a DWDS constituted by 1 reservoir, 40 junction nodes, 50 pipes and 4 valves. The hydraulic behaviour of the DWDS was analysed using the EPANET Programmer's Toolkit in Visual Basic and the hydraulic data was exported to MATLAB. This hydraulic analysis revealed the existence of 144 different hydraulic patterns per day. In each pattern, there are variations in velocity and in direction of the water in several pipes of the network. The analysis period considered was 1 day, i.e., 24 h.

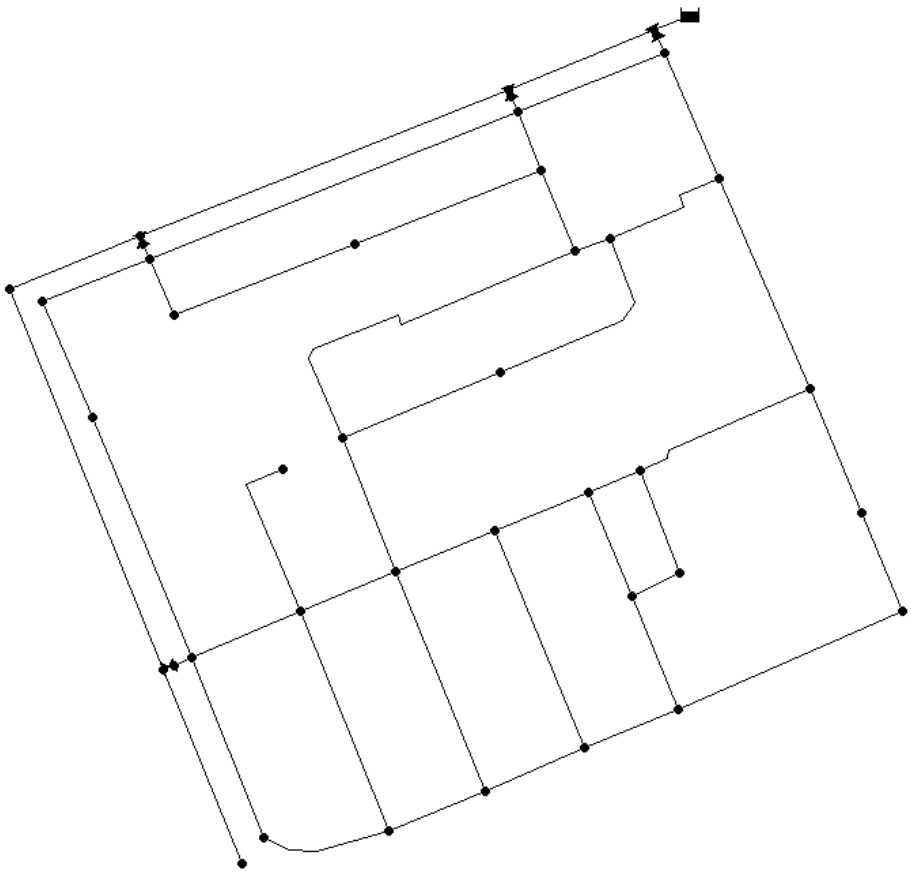

Figure 5.2 - Representation of Network B.

5.3 Network C

Figure 5.3 presents the Network C, constituted by 578 nodes (including 1 reservoir and 3 tanks) and 927 links. The hydraulic behaviour of the DWDS was analysed using the EPANET. This hydraulic analysis revealed the existence of 25 different hydraulic patterns per day. In each pattern, there are variations in velocity and in direction of the water in several pipes of the network.

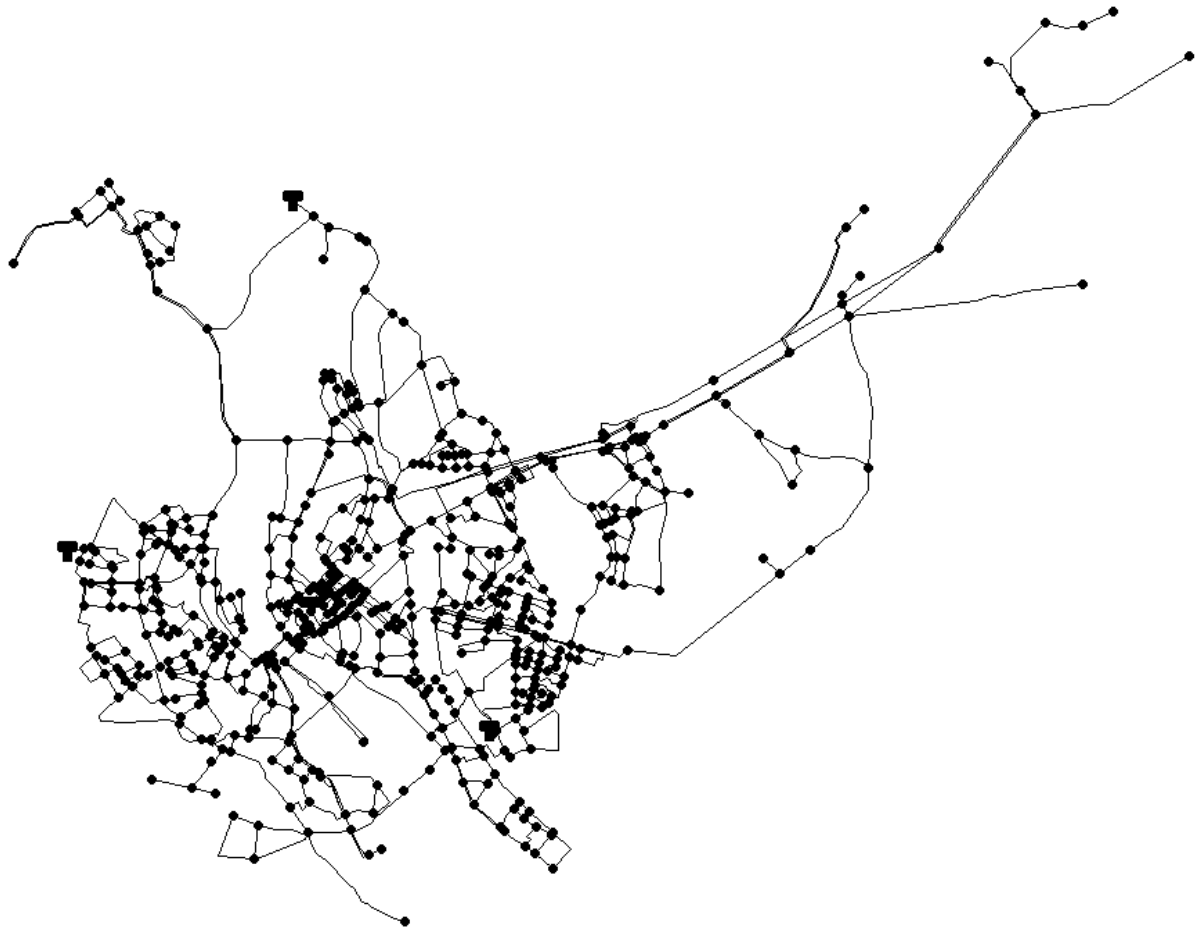

Figure 5.3 - Representation of Network C.

5.4 Network D

Figure 5.4 presents the Network D, a DWDS of an urban area constituted by 10 reservoirs, 8846 junction nodes, 1 tank, 9991 pipes, 7 pumps and 31 valves that was used as an example to explain the proposed method and to show its performance. The hydraulic behaviour of the DWDS was also analysed using the EPANET. This hydraulic analysis revealed the existence of 96 different hydraulic patterns per day. In each pattern, there are variations in velocity and in direction of the water in several pipes of the network.

Figure 5.4 – Representation of Network D.

Part II: Developed Methods for Simulation of Contaminations in Drinking Water Distribution Systems

6 Analytical approach for the advective transport phenomenon with reaction

6.1 Governing Equations

The model used in the evaluation of contaminant concentrations was based on the phenomena of advective transport and reaction in the pipes. This model considers that a dissolved substance will travel down the length of a pipe with the same average velocity as the carrier fluid at the same time reacting at some given rate. So, the advective transport within a pipe is represented with the following equation:

$$\frac{\partial C_i(x,t)}{\partial t} = -u_i \frac{\partial C_i(x,t)}{\partial x} + r\left(C_i(x,t)\right) \tag{6.1}$$

where $C_i(x,t)$ is the concentration (mass/volume) in pipe i as a function of distance x and time t, u_i is the flow velocity (length/time) in pipe i and r is the rate of reaction (mass/volume/time) that is function of concentration (Rossman et al., 1993).

An analytical solution was obtained for the partial differential equation given by Equation 6.1. It is assumed that a pseudo-first order reaction rate is able to simulate the decay of contaminant:

$$r\left(C_i(x,t)\right) = -k\, C_i(x,t) \tag{6.2}$$

Equation 6.3 is obtained by the substitution of the Equation 6.2 in Equation 6.1:

$$\frac{\partial C_i(x,t)}{\partial t} = -u_i \frac{\partial C_i(x,t)}{\partial x} - k\, C_i(x,t) \tag{6.3}$$

Applying the Laplace Transform to the variable t, Equation 6.3 is transformed in Equation 6.4:

$$s\,\overline{C_i}(x,s) - C_i(x,0) = -u_i \frac{d\,\overline{C_i}(x,s)}{d\,x} - k\,\overline{C_i}(x,s) \tag{6.4}$$

The next step is to define the boundary conditions. It is assumed that the initial contaminant concentration is 0 for all the length of the pipe.

$$C_i(x,0)=0 \tag{6.5}$$

Substituting this initial condition in Equation 6.4, Equation 6.6 is obtained:

$$s\,\overline{C_i}(x,s)=-u_i\frac{d\,\overline{C_i}(x,s)}{d\,x}-k\,\overline{C_i}(x,s) \tag{6.6}$$

Equation 6.6 is a differential equation which can be solved by direct integration:

$$-\frac{1}{u_i}\int d\,x+f(s)=\int\frac{d\,\overline{C_i}(x,s)}{\overline{C_i}(x,s)(s+k)} \tag{6.7}$$

The integration is performed without integration limits. A function $f(s)$ is added as a constant of integration, not dependent of the variable x, but as a function of the other independent variable s.

Integrating the Equation 6.7 and rewriting the same equation, it is possible to obtain an explicit form for $\overline{C_i}(x,s)$:

$$\overline{C_i}(x,s)=\frac{1}{s+k}\exp\left[(s+k)\left(-\frac{x}{u_i}+f(s)\right)\right] \tag{6.8}$$

Considering $g(s)=\exp[(s+k)\,f(s)]$, the Equation 6.9 is obtained:

$$\overline{C_i}(x,s)=\frac{1}{s+k}\exp\left[(s+k)\left(-\frac{x}{u_i}\right)\right]g(s) \tag{6.9}$$

It is still impossible to invert Equation 6.9 for time domain because $g(s)$ is an unknown function. This function can be determined by applying a second boundary condition. This boundary condition must be the contaminant concentration at the point $x=0$ along the time. As example, Equation 6.10 describes a pulse with amplitude A that occurs between the times t_1 and t_2.

$$C_i(0,t)=A[H(t-t_1)-H(t-t_2)] \tag{6.10}$$

The transformation of this equation to Laplace domain is:

$$\overline{C_i}(0,s)=A\left[\frac{\exp(-t_1\,s)}{s}-\frac{\exp(-t_2\,s)}{s}\right] \tag{6.11}$$

Applying this boundary condition to Equation 6.9 allows the evaluation of $g(s)$, i.e.:

$$\frac{1}{s+k}\exp\left[(s+k)\left(-\frac{0}{u_i}\right)\right]g(s)=A\left[\frac{\exp(-t_1\,s)}{s}-\frac{\exp(-t_2\,s)}{s}\right] \tag{6.12}$$

$$g(s)=A(s+k)\left[\frac{\exp(-t_1\,s)}{s}-\frac{\exp(-t_2\,s)}{s}\right] \tag{6.13}$$

Substituting Equation 6.13 in Equation 6.9 and rearranging:

$$\overline{C_i}(x,s)=\frac{A}{s}\exp\left(-\frac{k\,x}{u_i}\right)\left[\exp\left[-\left(\frac{x}{u_i}+t_1\right)s\right]-\exp\left[-\left(\frac{x}{u_i}+t_2\right)s\right]\right] \tag{6.14}$$

It is possible to invert Equation 6.14 for time domain. Equation 6.15 is the analytical solution for the concentration profile in a plug flow chemical reactor modelled by Equations 6.1 and 6.2.

$$C_i(x,t)=A\exp\left(-\frac{k\,x}{u_i}\right)\left[H\left[t-\left(\frac{x}{u_i}+t_1\right)\right]-H\left[t-\left(\frac{x}{u_i}+t_2\right)\right]\right] \tag{6.15}$$

An equation for the description of concentration profile in a plug flow chemical reactor, such as Equation 6.15, can be obtained for any conditions at the point x=0. It is only necessary to redefine the Equation 6.10, that describes the concentration at point x=0 and repeat the procedure to evaluate the function $g(s)$ and subsequently the function $\overline{C_i}(x,s)$.

For instance, it is possible to define the concentration profile of another pipe, which is connected to the end of the present pipe with contamination profile as described by the Equation 6.15. Denominating the first pipe as pipe 1 and the second as pipe 2, the initial concentration is set as 0 for all length of pipe 2, as it has been done for pipe 1. Having the same boundary condition at t=0, Equations 6.5 to 6.9 are also used for pipe 2.

It is assumed that the concentration at the start of the pipe 2 is equal to the concentration at the end of pipe 1 - $x=l_1$ - (l_1 is the length of pipe 1). To determine the second boundary condition, it is necessary to use the Equation 6.14, setting the variable x as l_1.

$$\overline{C_2}(0,s)=\overline{C_1}(l_1,s)=\frac{A}{s}\exp\left(-\frac{k\,l_1}{u_1}\right)\left[\exp\left[-\left(\frac{l_1}{u_1}+t_1\right)s\right]-\exp\left[-\left(\frac{l_1}{u_1}+t_2\right)s\right]\right] \tag{6.16}$$

Applying this boundary condition to Equation 6.9, a new function $g(s)$ is created:

$$g(s) = (s+k)\frac{A}{s}\exp\left(-\frac{k\,l_1}{u_1}\right)\left[\exp\left[-\left(\frac{l_1}{u_1}+t_1\right)s\right]-\exp\left[-\left(\frac{l_1}{u_1}+t_2\right)s\right]\right] \tag{6.17}$$

Modifying the Equation 6.9 with the information given by Equation 6.17 and rearranging:

$$\overline{C_2}(x,s) = \frac{A}{s}\exp\left[-\left(\frac{k\,l_1}{u_1}+\frac{k\,x}{u_2}\right)\right]\left[\exp\left[-\left(\frac{l_1}{u_1}+\frac{x}{u_2}+t_1\right)s\right]-\exp\left[-\left(\frac{l_1}{u_1}+\frac{x}{u_2}+t_2\right)s\right]\right] \tag{6.18}$$

The inverse of the Equation 6.18 enables to evaluate the concentration profile for the pipe 2.

$$C_2(x,t) = A\exp\left[-\left(\frac{k\,l_1}{u_1}+\frac{k\,x}{u_2}\right)\right]\left[H\left[t-\left(\frac{l_1}{u_1}+\frac{x}{u_2}+t_1\right)\right]-H\left[t-\left(\frac{l_1}{u_1}+\frac{x}{u_2}+t_2\right)\right]\right] \tag{6.19}$$

6.2 The Software Tool

A software tool was developed in MATLAB and Visual Basic for Applications, incorporating the hydraulic analysis performed by EPANET software (Rossman, 2000) with models for the evaluation of contaminant concentrations solved using the analytical approach described previously, under steady hydraulic conditions.

The application is divided in the following tasks:

1. Network design: The network is designed with the EPANET software, introducing the characteristics and the information associated to the equipment. For example, it is necessary to define pump curves in EPANET.

2. Setting the necessary data to characterize the reaction within the pipes.

3. Definition of the perturbations, providing the information needed to fully characterize the perturbations.

4. Run hydraulics analysis using the EPANET Programmer's Toolkit, which is a dynamic link library that allows the developers to incorporate EPANET's functions in their own applications.

5. Perform the evaluation of contaminant concentrations, based on the equations described before. The hydraulic data needed to run the evaluation of contaminant concentrations is provided by EPANET results computed previously.

Figure 6.1 shows a diagram that explains the interaction between MATLAB, Visual Basic for Applications (VBA) and EPANET software, during the execution of steps listed above.

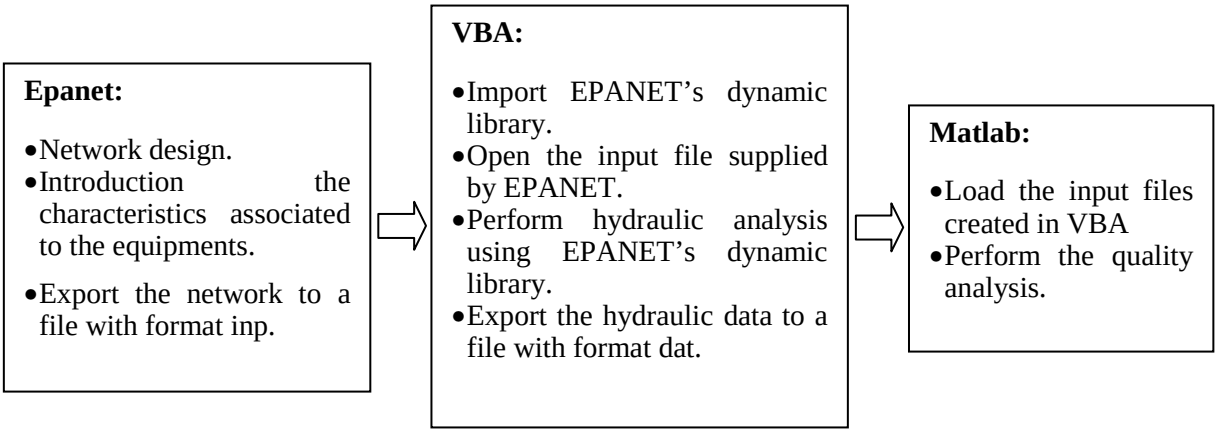

Figure 6.1 - Interaction between MATLAB, Visual Basic for Applications and EPANET.

The program starts introducing the following information:

- Network scheme designed by EPANET software in a file with the format *inp*;
- Reaction rate coefficient and order of reaction;
- Nodes where perturbations occur;
- Start time and duration of each perturbation;
- Amount of contaminant introduced at network per time unit.

As output, the software applications return an expression for describing the concentration profile for each physical component of the network, at the water.

The software tool is executed sequentially and incorporates the following steps:

Step 1. Import the EPANET Programmer's Toolkit, EPANET's dynamic link library.

Step 2. Open the input file that provides the description of the network in study and run a complete extended period analysis.

Step 3. Get the information related with nodes (node type, hydraulic head and water demand) and links (length, velocity, diameter, start and end nodes).

Step 4. Sort the links for decreasing order of hydraulic head at the start node (if there were nodes with the same hydraulic head at the start node, sort those nodes for decreasing order of hydraulic head at the end node), with exception of links, which have a reservoir as start node that have to be placed first in the order of resolution, and pumps.

Step 5. Set the concentration at each node and in all length of each pipe as 0, for $t = 0$, the first boundary condition.

Step 6. Define the perturbations during the simulation time, which simulates the deliberate contaminations.

Step 7. Run the evaluation of contaminant concentrations for each link following the order of resolution defined in Step 5:

Firstly, it is necessary to perform the mass balance, in Laplace domain, at the start node for each pipe that enables the calculation of $\overline{C_i}(0,s)$, the second boundary condition.

$$\overline{C_i}(0,s) = \overline{C_p}(s) = \frac{\sum_{i \in pi} Q_i \, \overline{C_i}(l_i, s)}{\sum_{i \in pi} Q_i} \tag{6.21}$$

where $\overline{C_p}(s)$ denotes the contaminant concentration at junction node p, pi is the set of incoming links to node p, Q_i is the volumetric flow rate in link i and l_i is the length.

Step 8. After the determination of contaminant concentration at the start node, it is necessary to determine the expression that describes the concentration profile for all values of x at the bulk, following Equation 6.22.

$$\overline{C_i}(x,s) = \exp\left(-\frac{x}{u_i}(s+k)\right)\overline{C_i}(0,s) \tag{6.22}$$

6.3 Case Study

One case study is presented to demonstrate the performance of this software tool. Network C, presented in Section 5.3, was analysed. It was simulated a contamination occurring at Node A, starting at t=0 h with duration of 1 h (pulse of 50 g/s). The contaminant concentration at the water was evaluated for the entire DWDS, considering no reactive transport or a pseudo-first order reaction with $k = 1 \times 10^{-4} s^{-1}$.

Figure 6.3 presents the area of the DWDS where it Node A, which is flagged with a red diamond, is located (see Figure 6.2). The Link I, which starts at Node B and is presented in green, was used to demonstrate the features of the proposed software tool. There are 3 different paths, presented in Figure 6.3 in different colours, to travel from Node A to Node B.

Figure 6.2 - Real DWDS.

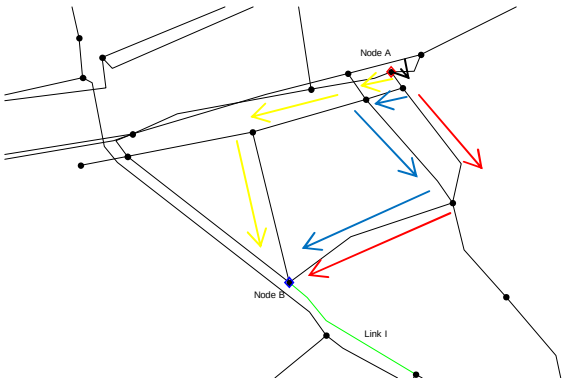

Figure 6.3 - Vicinity of the Node A.

Figure 6.4 shows the comparison between the contaminant concentration at the water in Node A and Node B, for the example without reaction. Figure 6.5 presents the same comparison with a different scale to enable the distinction of the second an third peaks observed at Node B.

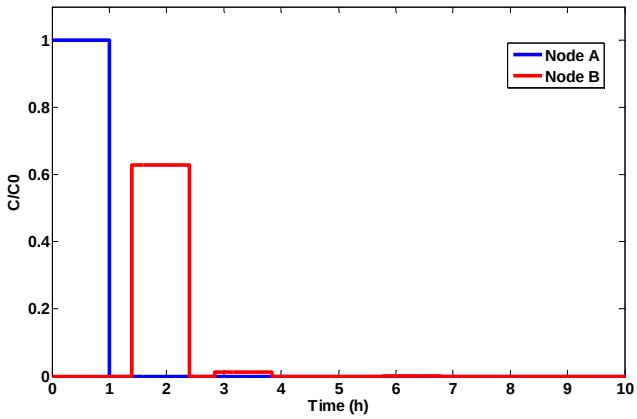

Figure 6.4 - Contaminant concentration at the water in Node A and Node B, without reaction.

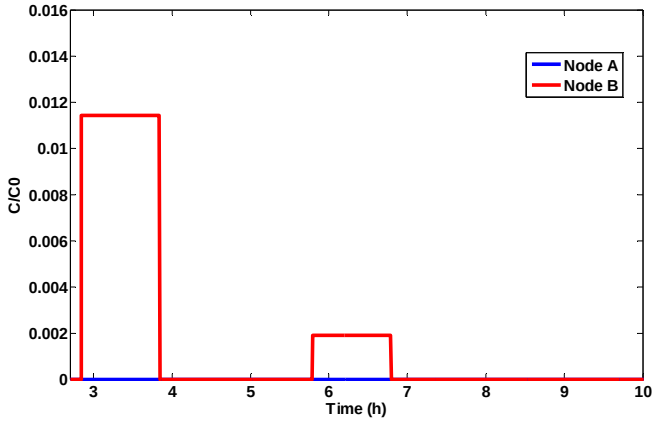

Figure 6.5 - Contaminant concentration at the water in Node A and Node B, without reaction, in a different scale.

There are three peaks, one for each flow path, occurring at the corresponding travel times. The first peak corresponds to the path presented in red in Figure 6.4 and occurs at 01:24 h; the second peak, which corresponds to the flow path presented in blue, occurs around 2:50 h; the third peak corresponds to the flow path presented in yellow and occurs at 05:48 h. The contaminant that reaches Node B through the paths corresponding to the second and third peaks is much diluted, thus these peaks are only perceptible in Figure 6.5. After the transporting of the contaminated fronts, it is observed that the contaminant concentration at the water is 0.

Figure 6.6 shows the comparison between the results obtained considering reactive and non-reactive transport for the contaminant concentration at Node B. Figure 6.6 also presents the same comparison with a different scale to enable the distinction of the second an third peaks observed at Node B.

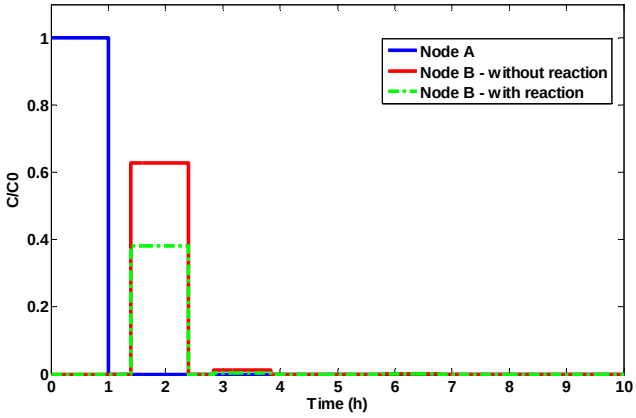

Figure 6.6 - Contaminant concentration at the water in Node A and Node B, with reaction.

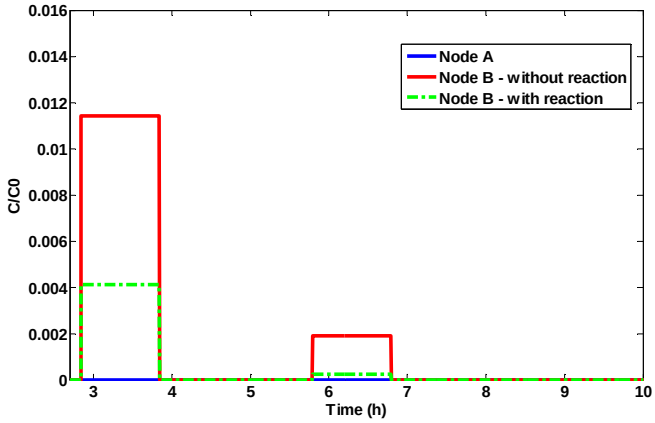

Figure 6.7 - Contaminant concentration at the water in Node A and Node B, with reaction, in a different scale.

In these figures, it is possible to observe that the contaminated fronts arrive at Node B at the same time considering reactive transport. The difference is related with the amplitude of peaks, which is obviously lower considering the decay described by the pseudo-first order reaction term.

6.4 Chapter conclusions

A software tool was developed to implement an analytical approach for the simulation of the advective transport of contaminants, considering pseudo-first order reaction terms.

The approach enabled the study of the effects of advective transport and reaction in the behaviour of contaminants in DWDSs.

The case study performed for a real DWDS demonstrated that the proposed method is suitable for providing the analytical solution for the modelling of the transport of contaminants in DWDS, considering steady hydraulic conditions.

7 Numerical approach for the simulation of contaminant reactive transport along drinking water distribution systems considering sorption phenomena

7.1 Introduction

The model used in the evaluation of contaminant concentrations was based on the phenomena of advective transport and reaction in the pipes (Rossman et al., 1993), considering the occurrence of sorption phenomena at the pipe walls. This model considers that a dissolved substance will travel down the length of a pipe with the same average velocity as the carrier fluid, at the same time reacting with chemical and microbial components in the water, at some given rate, while part of the contaminant can be transferred to the pipe walls and adsorb at either the surface material or at any layer (deposit) attached to this surface. Once at the liquid-solid interface, the contaminant will penetrate, to a higher or lower degree, into the adsorbent, depending on the characteristics of the sorbate and of the adsorbent.

So, the advective transport and reaction in the pipes is given by Equations 7.1 and 7.2, considering the occurrence of adsorption at the pipe walls:

$$\frac{\partial C_i\left(x,t\right)}{\partial t} = -u_i \frac{\partial C_i\left(x,t\right)}{\partial x} + r\left(C_i\left(x,t\right)\right) - Q_{Ads}\frac{m_{dep}}{V} \tag{7.1}$$

where Q_{Ads} is the rate of the mass of contaminant adsorbed per unit mass of deposit (mass of contaminant/mass of deposit/time), m_{dep} is the mass of solid phase (mass of sorbent) and V is the water volume (volume).

The equation of the pseudo-first order adsorption model is expressed by:

$$Q_{Ads} = \frac{k_1 V}{m_{dep}}\left(C_b - C_{eq}\right) \tag{7.2}$$

where k_1 is the pseudo-first order adsorption rate constant (time^{-1}), C_{eq} is the contaminant concentration in the water when equilibrium saturation is reached (mass of contaminant/volume) and

C_b is the contaminant concentration at the bulk. In this approach, it was assumed that C_b is the average between $C_i\,(x,t)$ and $C_i\,(x+dx,t)$, in which dx is integration interval for x.

Considering a generic equation for the decay reaction rate as given by Equation 7.3, where k is the rate kinetic coefficient and n is the reaction order.

$$r\left(C_i\,(x,t)\right) = -k\; C_i^{\,n}\,(x,t) \tag{7.3}$$

Equation 7.1 is transformed in Equation 7.4:

$$\frac{\partial C_i\,(x,t)}{\partial t} = -u_i\;\frac{\partial C_i\,(x,t)}{\partial x} - k\; C_i^{\,n}\,(x,t) - k_1\left(\frac{C_i\,(x,t)+C_i\,(x+dx,t)}{2} - C_{eq}\right) \tag{7.4}$$

Desorption can be modelled as a first order phenomena (as regards the amount of contaminant remaining in the deposit):

$$-\frac{dq}{dt} = k_d\,(q - q_e\,*) \tag{7.5}$$

Where k_d is the desorption rate coefficient of first order (time^{-1}), $q_e\,*$ is the final equilibrium concentration of contaminant in the solid (mass of contaminant/mass of deposit) for the desorption process and q is the contaminant concentration in the deposit (mass of contaminant/mass of deposit), at any time t.

The correspondent integrated equation is:

$$q = q_0 - (q_0 - q_e\,*)\left[1 - \exp(-k_d\,t)\right] \tag{7.6}$$

where q_0 is the contaminant concentration in the deposit at the beginning of the desorption phase.

The equations presented to sorption phenomena at the deposits can be used similarly for modelling the sorption phenomena at the biofilm.

The adsorption isotherms relate the amount of contaminant in the sorbent as a function of its concentration in the water at constant temperature. These isotherms are used to define the equilibrium conditions throughout the sorption phenomenon.

One of the most frequently used isotherms is the Langmuir isotherm, which can be expressed by Equation 7.7:

$$\frac{q_{eq}}{Q_0} = \frac{B\, C_{eq}}{1 + B\, C_{eq}} \tag{7.7}$$

where q_{eq} is the contaminant concentration at the deposit at the equilibrium (mass of contaminant/mass of deposit), Q_0 is the maximum contaminant concentration at the deposit and B is the Langmuir constant related to adsorption capacity.

7.2 Numerical Approach

The numerical solution is achieved applying the first order upwind scheme method for solving Equation 7.4. The first-order upwind method uses a one-sided finite difference in the upstream direction to approximate the convection term and can be expressed as follows (Alhumaizi, 2004):

$$C_i\left(x+dx,t+dt\right) = C_i\left(x+dx,t\right) - Cr\left(C_i\left(x+dx,t\right) - C_i\left(x,t\right)\right) \tag{7.8}$$

The Courant number, defined by Equation 7.9, should be lower than 1 to guarantee the convergence of the method.

$$Cr = u\,\frac{\Delta t}{\Delta x} \tag{7.9}$$

Replacing the convective term of Equation 7.4 by Equation 7.8, Equation 7.10 is obtained:

$$\begin{aligned} &C_i\left(x+dx,t+dt\right) = C_i\left(x+dx,t\right) - Cr\left(C_i\left(x+dx,t\right) - C_i\left(x,t\right)\right) \\ &- k\, C_i{}^{n}\left(x,t\right) - k_1\left(\frac{C_i\left(x,t\right) + C_i\left(x+dx,t\right)}{2} - C_{eq}\right) \end{aligned} \tag{7.10}$$

7.3 The Software Tool

A software tool was developed in MATLAB and Visual Basic for Applications, incorporating the hydraulic analysis performed by EPANET software (Rossman, 2000) with models for the evaluation of contaminant concentrations solved using the numerical approach described previously.

The application is divided in tasks similar to the ones already presented in Section 6.2. The only difference is in Task 2: besides the characterization of the reaction within pipes, it is also necessary to characterize the sorption phenomena and set the integration steps. The interaction between MATLAB, Visual Basic for Applications and EPANET software may also be represented by the diagram shown in Figure 6.1.

The program starts introducing the following information:

- Network scheme designed by EPANET software in a file with the format *inp*;
- Simulation time;
- Integration steps for space and time variables;
- Reaction rate coefficient and order of reaction;
- Adsorption isotherms;
- Sorption kinetic parameters;
- Nodes where perturbations occur;
- Start time and duration of each perturbation;
- Amount of contaminant introduced at network per time unit.

As output, the software applications return the concentration profiles for each physical component of the network, at the water and at the deposits.

The software tool is executed sequentially and incorporates the following steps:

Step 1. Import the EPANET Programmer's Toolkit, EPANET's dynamic link library.

Step 2. Open the input file that provides the description of the network in study and run a complete extended period analysis.

Step 3. Get the information related with nodes (node type, hydraulic head and water demand) and links (length, velocity, diameter, start and end nodes).

Step 4. Sort the links for decreasing order of hydraulic head at the start node (if there were nodes with the same hydraulic head at the start node, sort those nodes for decreasing order of hydraulic head at the end node), with exception links, which have a reservoir as start node that have to be placed first in the order of resolution, and pumps.

Step 5. Set the concentration at each node and in all length of each pipe as 0, for $t = 0$.

Step 6. Define the perturbations during the simulation time, which simulates the deliberate contaminations.

Step 7. Run the evaluation of contaminant concentrations for each link following the order of resolution defined in Step 4, for each time step during the simulation time:

Firstly, it is necessary to perform the mass balance at the start node for each pipe which allows calculating $C_i(0,t)$ through the Equation 7.10. This calculation is performed once in each time step..

$$C_i(0,t) = C_p(t) = \frac{\sum_{i \in pi} Q_i \, C_i(l_i,t)}{\sum_{i \in pi} Q_i} \tag{7.11}$$

where $C_p(t)$ denotes the contaminant concentration at junction node p, pi is the set of incoming links to node p, Q_i is the volumetric flow rate in link i and l_i is the length.

Step 8. After the determination of contaminant concentration at the start node, it is necessary to compute the concentration profile for all values of x at the bulk and at the deposit, through the following procedure:

Step 8.1. Determine Ceq using the adsorption isotherm.

Step 8.2. Evaluate $C_i(x+dx,t+t+dt)$ considering only advective transport with reaction, neglecting the sorption. Compare $C_i(x+dx,t+dt)$ with Ceq.

Step 8.2.1. If it is higher, adsorption occurs and $C_i(x+dx,t+dt)$ is evaluated using the Equation 7.10. The contaminant concentration at the deposit $q(x,t+dt)$ is updated, taking into account the amount of contamination transferred from the water to the deposit.

Step 8.2.2. If it is equal, the system is at equilibrium, there is no sorption and the previous evaluation of $C_i(x+dx,t+dt)$, performed at Step 8.2, is used. The contaminant concentration at the deposit $q(x,t+dt)$ remains constant.

Step 8.2.3. If it is lower, desorption occurs and $q(x,t+dt)$ is evaluated using the Equation 7.6. The contaminant concentration at the bulk $C_i(x+dx,t+dt)$ computed at the Step 8.2 is updated, taking into account the amount of contamination transferred from the deposit to the water.

Step 9. Calculate the contaminant concentration at tanks and nodes that aren't the start nodes for any link. The mass balance for a tank is given by the Equation 7.12:

$$\frac{d\left(V_{\tan k}(t)C_p(t)\right)}{dt} = \sum_{i\in pi} Q_i\ C_i\ (l_i,t) - \sum_{j\in pj} Q_j\ C_p\ (t) + r\left(C_p(t)\right)V_{\tan k}(t) - Q_{Ads}\ m_{dep} \qquad (7.12)$$

where $V_{tank}(t)$ is the volume of water in the tank, pj is the set of links that leave the tank p and Q_j is the volumetric flow rate in link *j*.

Equation 7.13 is a generic equation for the reaction rate, similar to Equation 7.3 but dependent of concentration at the node.

$$r\left(C_p(t)\right) = -k\ C_p{}^n(t) \qquad (7.13)$$

With the substitution of Equation 7.13 in Equation 7.10 and developing the derivative, Equation 7.14 is obtained.

$$\begin{aligned} &V_{\tan k}(t)\frac{dC_p(t)}{dt} + C_p(t)\frac{dV_{\tan k}(t)}{dt} = \\ &\sum_{i\in pi} Q_i\ C_i\ (l_i,t) - \sum_{j\in pj} Q_j\ C_p\ (t) - k\ C_p{}^n(t)V_{\tan k}(t) - Q_{Ads}\ m_{dep} \end{aligned} \qquad (7.14)$$

It is possible to write an expression that relates the volume in the tank with the time, using the explicit form:

$$V_{\tan k}(t) = V_0 + \left(\sum_{i\in pi} Q_i - \sum_{j\in pj} Q_j\right)t \qquad (7.15)$$

where V_0 is the initial volume at the tank. Using the Equation 7.15, the derivative of volume is obtained by:

$$\frac{dV_{\tan k}(t)}{dt} = \sum_{i\in pi} Q_i - \sum_{j\in pj} Q_j \qquad (7.16)$$

The derivative of the concentration at the node p in order to time can be rewrite using the finite difference method. Using this approximation, Equation 7.17 is obtained:

$$\frac{d\,C_p(t)}{d\,t} = \frac{C_p(t+\Delta t) - C_p(t)}{\Delta t} \tag{7.17}$$

The substitution of the Equations 7.16 and 7.17 in Equation 7.14 leads to:

$$\begin{aligned} &V_{\tan k}(t)\frac{C_p(t+\Delta t) - C_p(t)}{\Delta t} + C_p(t)\left(\sum_{i\in pi} Q_i - \sum_{j\in pj} Q_j\right) = \\ &\sum_{i\in pi} Q_i\, C_i(l_i,t) - \sum_{j\in pj} Q_j\, C_p(t) - k\, C_p{}^n(t)\, V_{\tan k}(t) - Q_{Ads}\, m_{dep} \end{aligned} \tag{7.18}$$

It is also possible to obtain an equation in an explicit form in relation to $C_p(t+\Delta t)$, rearranging the Equation 7.18:

$$C_p(t+\Delta t) = C_p(t) + \frac{\Delta t}{V_{\tan k}(t)}\left[\sum_{i\in pi} Q_i\left(C_i(l_i,t) - C_p(t)\right) - Q_{Ads}\, m_{dep}\right] - k\, C_p{}^n(t)\,\Delta t \tag{7.19}$$

These 9 steps allow performing the evaluation of contaminant concentrations, presenting results for the contaminant concentration at each node or pipe length step in the network.

7.4 Case Studies

Two case studies are presented to demonstrate the performance of this software tool. Both case studies were evaluated considering a contamination by paraquat at 20 ºC. The reaction term was not considered because the purpose of this case study was to evaluate the effect of the sorption in the behaviour of the systems. The information regarding the sorption kinetic parameters and adsorption isotherms were taken from the sorption experiments for paraquat in APP1 deposits presented in Deliverable 4.2 of "SecurEau" project (confidential report).

The Langmuir isotherm (as it was presented at Equation 7.7) at 20 ºC is given by Equation 7.20, where q_{eq} and Q_0 are presented in mg of contaminant/g of deposit, B in litres of water/g of deposit and C_{eq} in mg of contaminant/ litres of water.

$$q_{eq} = 5.7\,\frac{0.6\,C_{eq}}{1 + 0.6\,C_{eq}} \tag{7.20}$$

7.4.1 Case Study 1

The first case study was performed for a pipe with 5000 m of length and 0.2 m of diameter. The velocity was considered constant at 0.3 m/s. One pulse with the contaminant concentration of 30 g/m^3 was simulated at the entrance of the pipe, starting at 1h with duration of 5 h.

Two different sets of sorption kinetic parameters were considered. In the first scenario, the sorption kinetic parameters presented in Deliverable 4.2 for paraquat in APP1 deposits were used. As there were not any available results for the desorption of paraquat, the desorption rate coefficient was estimated as 10 times lower than the adsorption rate coefficient and the final equilibrium concentration of contaminant in the solid for the desorption process was estimated as 100 lower than the maximum amount of contaminant that can be adsorbed, for Scenario 1. In the Scenario 2, the adsorption and desorption rate coefficients were defined as 10 times higher than the parameters used in Scenario 1, to enable a clearer demonstration of the effect of the adsorption in the contaminant concentration profiles. Table 7.1 presents the sorption kinetic parameters used in both scenarios.

Table 7.1– Sorption kinetic parameters for Scenarios 1 and 2.

Scenario 1		Scenario 2	
k_1 (min^{-1})	$7.5x10^{-4}$	k_1 (min^{-1})	$7.5x10^{-3}$
q_{ads} (mg_{cont}/g_{dep})	5	q_{ads} (mg_{cont}/g_{dep})	5
k_d (min^{-1})	$7.5x10^{-5}$	k_d (min^{-1})	$7.5x10^{-4}$
q_{des} (mg_{cont}/g_{dep})	$5x10^{-2}$	q_{des} (mg_{cont}/g_{dep})	$5x10^{-4}$

Figures 7.1 and 7.2 present the contaminant concentration in the water and in the deposits, respectively, at different positions of the pipe as function of time, for Scenario 1.

Figures 7.3 and 7.4 present the contaminant concentration in the water and in the deposits, respectively, at different instants as function of the position in the pipe, for Scenario 1.

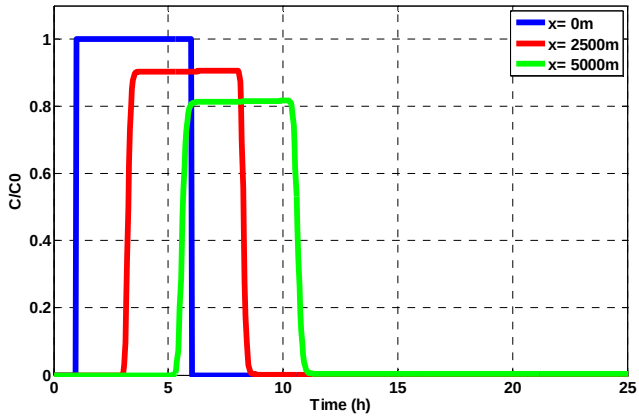

Figure 7.1 - Contaminant concentration in the water as function of time (Scenario 1)

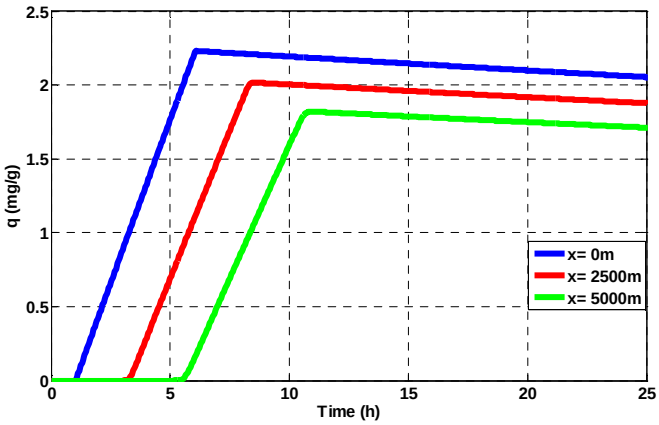

Figure 7.2 - Contaminant concentration in the deposits as function of time (Scenario 1).

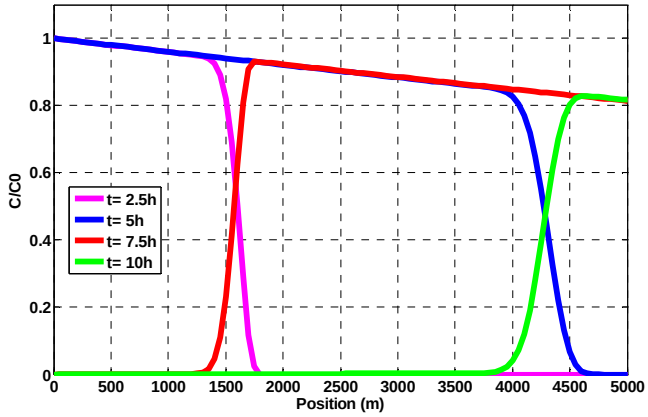

Figure 7.3 - Contaminant concentration in the water as function of position (Scenario 1).

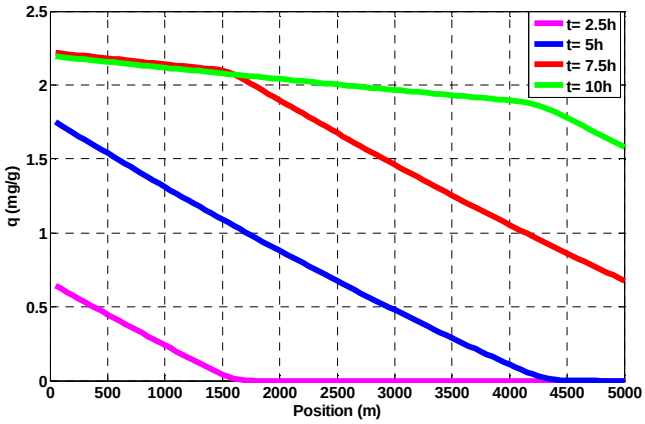

Figure 7.4 - Contaminant concentration in the deposits as function of position (Scenario 1)

In Figure 7.1, it is possible to observe that the maximum contaminant concentration in the water is decreasing along the pipe. This happens because as the contaminated water travels along the pipe, the contaminant is getting adsorbed to the deposits, so the contaminant concentration in the water becomes lower. The effect of desorption is negligible, when the contaminated water leaves the pipe the deposits release contaminant to the clean water, but at a low rate that cannot be observed at the scale of the figure.

In Figure 7.2, in the adsorption phase, the contaminant concentration in the deposits increases. As soon as the introduction of contaminant is finished, clean water starts to circulate and desorption occurs, with a consequent decrease of the contaminant concentration in the deposits.

Figure 7.3 presents the decay of the contaminant concentration in the water throughout the pipe, because as the water travels throughout the pipe with unsaturated deposits, there is a continuous adsorption of contaminant in the deposits.

Figure 7.4 shows that the contaminant concentration in the deposits is higher in the beginning of the pipe. This is explained by the fact that the driving force of the adsorption is higher in the beginning of the pipe, as the contaminant concentration in the water is higher.

Figures 7.5 and 7.6 present the contaminant concentration in the water and in the deposits, respectively, at different positions of the pipe as function of time, for Scenario 2.

Figures 7.7 and 7.8 present the contaminant concentration in the water and in the deposits, respectively, at different instants as function of the position in the pipe, for Scenario 2.

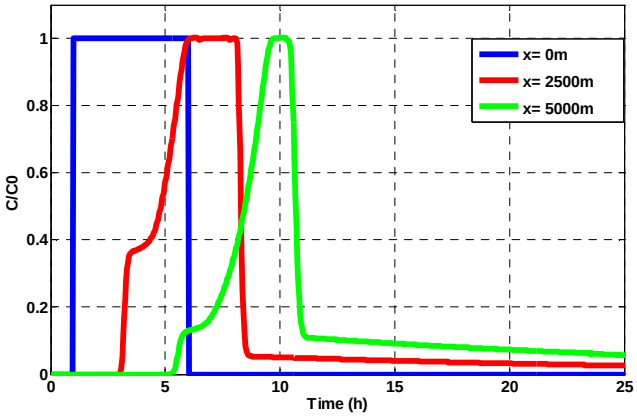

Figure 7.5 - Contaminant concentration in the water as function of time (Scenario 2)

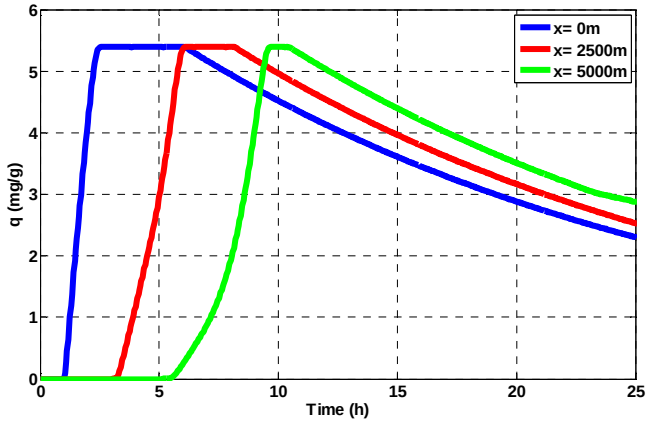

Figure 7.6 - Contaminant concentration in the deposits as function of time (Scenario 2)

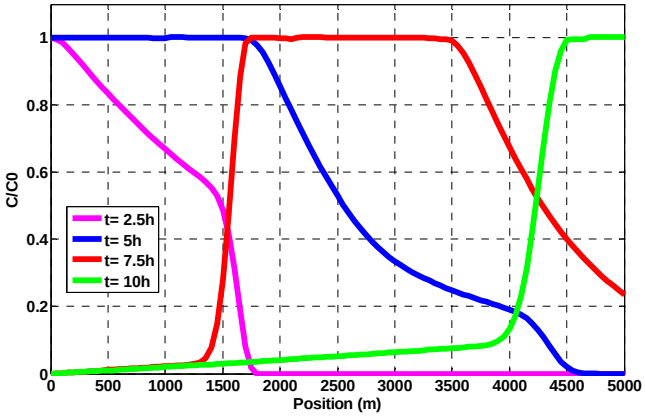

Figure 7.7 - Contaminant concentration in the water as function of position (Scenario 2)

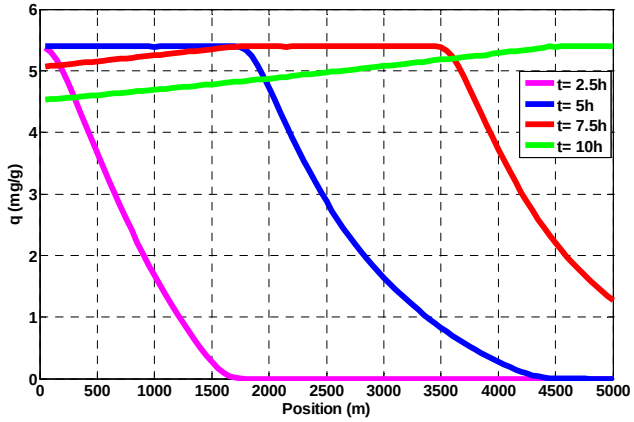

Figure 7.8 - Contaminant concentration in the deposits as function of position (Scenario 2)

In Figure 7.5, it is possible to observe the effect of the adsorption of contaminant to the deposits. However, in contrast with what happens in Scenario 1, the deposits reach the saturation point. Thus, after reaching the saturation point, as more contaminated water is passing through the pipe there is no adsorption, and the contaminant concentration at different positions throughout the pipe reaches the same maximum value as at the entrance of the pipe. The effect of the desorption is observed, when the contaminated water leaves the pipe the deposits release contaminant to the clean water at a higher rate, thus a significant contaminant concentration is still observed after the passing of the contaminated front.

The effect of the saturation of the deposits is clearly demonstrated in Figure 7.6. It is observed an increase of the contaminant concentration in the deposits until it reaches the saturation point, when there is no adsorption anymore. Finally after the passing of the contaminated front, desorption occurs and contaminant is released from the deposits to the clean water.

Figures 7.7 and 7.8 demonstrate the instants when the different parts of the pipe reach the saturation point.

7.4.2 Case Study 2

The second case study considered the same contamination scenario defined in the Section 6.3.

Figure 7.9 shows the comparison between the contaminant concentration in the water in Node A and Node B, using the sorption kinetic parameters defined for Scenario 2. Figure 7.10 presents the same comparison with a different scale to enable the distinction of the three peaks observed at Node B.

Figures 7.11 and 7.12 present the comparison between these results and the results obtained with the analytical approach without reaction, to highlight the effects of sorption phenomena.

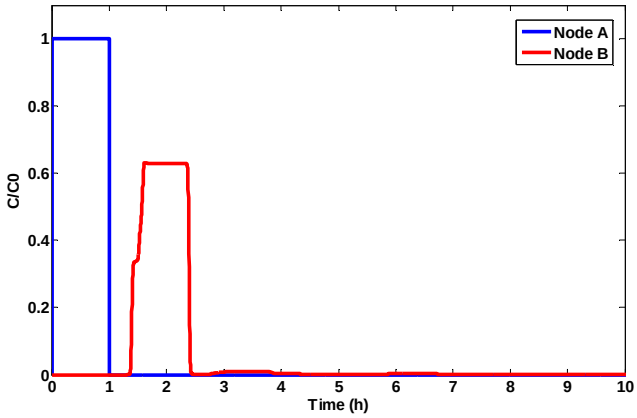

Figure 7.9 - Contaminant concentration in the water in Node A and Node B

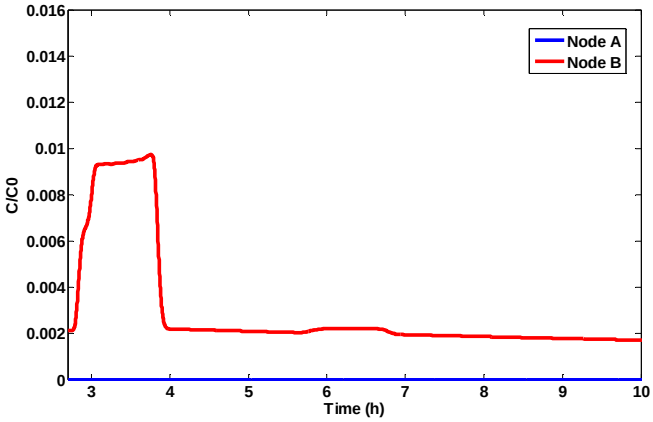

Figure 7.10 - Contaminant concentration in the water in Node A and Node B, in a different scale.

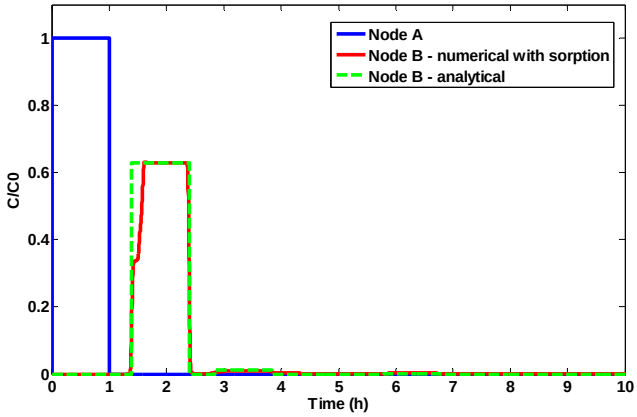

Figure 7.11 – Comparison between results obtained with the numerical approach considering sorption phenomena and results obtained with the analytical approach.

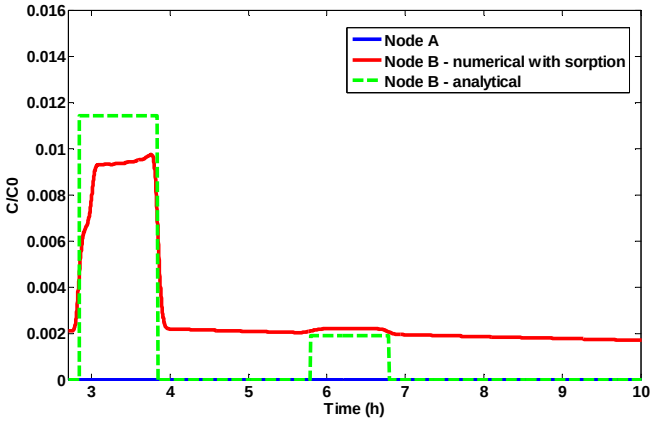

Figure 7.12 – Comparison between results obtained with the numerical approach considering sorption phenomena and results obtained with the analytical approach, in a different scale.

There are 3 peaks, one for each flow path, occurring at the corresponding travel times, as it was previously shown for the analytical approach. The second and third peaks are only observed in Figures 7.10 and 7.12, due to the dilution that the contaminant suffers through the corresponding paths. After the transporting of the contaminated fronts, it is observed that the contaminant concentration in the water is not 0, due to the occurrence of desorption.

It is possible to observe the effect of the adsorption in the shape of the peaks, as it was already described for the Case Study 1.

Figure 7.13 presents the comparison between the contaminant concentration in the deposits in the beginning of the links that start at Node A and Node B.

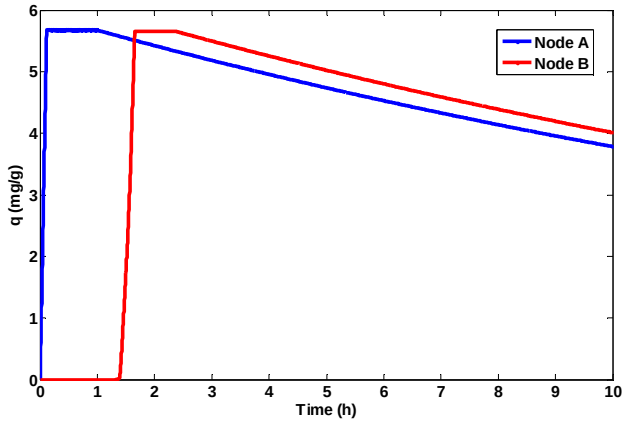

Figure 7.13 - Contaminant concentration in the deposits in the beginning of the links that start at Node A and Node B

It is also possible to observe that the contaminant concentration in the deposits in beginning of the link that starts at Node A increases during the transporting of the contaminated front until it reaches the saturation point. When clean water resumes flowing into the pipe, desorption may start occurring more significantly and the contaminant concentration in the deposits decreases.

Regarding the deposits in beginning of the link that starts at Node B, it is verified an increase of the contaminant concentration when the first contaminated front arrives. The deposits reach the saturation point, so the contaminant concentration remains steady until clean water resumes flowing into the pipe, when desorption occurs and the contaminant concentration decreases. The second and third contaminated fronts are much diluted, as it was already discussed. Thus, the arrival of this contaminated front does not interrupt the desorption, as the contaminant concentration in the water was lower than the C_{eq}.

7.5 Chapter Conclusions

A software tool was developed to simulate the contaminant reactive transport along DWDSs considering sorption phenomena.

The approach followed enabled the study of the effects of sorption phenomena in the behaviour of contaminants in DWDSs.

For contaminants with low sorption rate coefficients, the effect of the sorption phenomena might be negligible. However, for higher rate coefficients sorption phenomena can change significantly the transport of contaminants throughout the network and the desorption can be responsible for a longstanding release of contaminants.

The case study performed for a real DWDS demonstrated that the proposed method is suitable for the study of the effects of the sorption phenomena in the modelling of the transport of contaminants in real DWDS.

Part III: Deterministic Methods for the Localization of Contamination Sources in Drinking Water Distribution Systems

8 A Method Based on the Residence Time of Water in Pipes for the Localization of Contamination Sources in a Drinking Water Distribution System

8.1 Chapter Overview

The majority of the existent methods for the localization of contamination sources require the knowledge of the contaminants' concentrations throughout the network. This information is not easy to obtain, because the contaminants may be unknown and the sensors are usually not capable to measure the concentration of the contaminants that were deliberately or accidentally introduced.

The main aim of this chapter is to propose an alternative solution scheme for the localization of contamination sources, presenting a method based on the analysis of the residence time of water in pipes. With this approach, it is not necessary to have accurate values for the contaminant concentrations in sensors. The method is based on binary sensor status over the time, a type of information more reachable in real situations.

8.2 Methods Description

This section describes three algorithms, which correspond to the proposed approach to solve the problem of the localization of possible contamination sources in DWDS. The first algorithm (*Algorithm A*) allows the description of the different paths that a contaminated water parcel can take from a contamination source. The second algorithm (*Algorithm B*) allows the evaluation of all possible contamination sources that could explain a positive reading in each sensor. *Algorithm B* runs in reverse time. These algorithms are based on similar concepts to the particle backtracking algorithm presented by Zierolf et al. (1998) and Shang et al. (2002).

In both algorithms, it is assumed that the velocity and flow direction are known for all pipes, in each hydraulic period contained in a defined time horizon, *Thorizon*, correspondent to the interval analysed after the contamination. The algorithms study the distribution of contaminants in a DWDS between an initial instant and a *Thorizon*, assuming non reactive transport. These two algorithms are needed to run the third proposed algorithm (*Algorithm C*) to perform the localization of contamination sources based on the information gathered by sensors.

8.2.1 Algorithm A

Given a possible contamination source *S* associated to an instant of contamination *T* (*T* belongs to hydraulic period *i*), initialize the matrix *Distribution* with *S* as *Node* and *T* as *Tcont*.

Step 1 - Verify the size *N(i,Node)* of the set of links *LinksPost* that have each new value of *Node* as upstream node, at hydraulic period *i*. If $N(i, Node) = 0$ for each *Node*, finish the algorithm.

Step 2 - For each path, verify if the water parcel which leaves *Node* at *Tcont* reaches the downstream node *D* before the end of the hydraulic period t_i.

If $Tcont + \tau_i^p \leq t_i$ (τ_i^p is the residence time in pipe *p* at hydraulic period *i*) add *D* and the associated time $Tcont + \tau_i^p$ to the matrix *Distribution* and actualize the values of *Node* and *Tcont*.

a) Otherwise:

Step 2.1 - Calculate the distance travelled by the water during the hydraulic period *i*: $x_i^p = u_i^p \times (t_i - Tcont)$; u_i^p is the velocity of water in pipe *p* at hydraulic period *i*.

Step 2.2 - Consider the next hydraulic period (*i+1*) and add the distance travelled during the period until reach the period *i** in which $\sum_{j=i:i^*} x_j^p \leq 0$ or $\sum_{j=i:i^*} x_j^p \geq l^p$ (l^p is the length of the pipe).

Step 2.3 - If $\sum_{j=i:i^*} x_j^p \leq 0$: *Node* = *D*; $Tcont = t_{i^*} + \dfrac{\sum_{j=i:i^*} x_j^p}{u_{i^*}^p}$.

If $\sum_{j=i:i^*} x_j^p \geq l^p$: *Node* = *D*; $Tcont = t_{i^*} - \dfrac{l^p - \sum_{j=i:i^*} x_j^p}{u_{i^*}^p}$.

Step 2.4 – Add the new values of *Node* and *Tcont* to the matrix *Distribution*.

Step 3 – If $Tcont < Thorizon$, return to *Step 1*. Otherwise, finish the algorithm.

8.2.2 Algorithm B

Given a positive reading at sensor *X* with an associated instant of registration *Y* (*Y* belongs to hydraulic period *i*), initialize the matrix *Track* with *X* as *Source* and *Y* as *Tsource*.

Step 1 - Verify the size *N(i,Source)* of the set of links *LinksPrev* that have each value of *Source* as downstream node, at hydraulic period *i*. If $N(i, Source) = 0$ for each *Source*, finish the algorithm.

Step 2 - For each path, verify if the water parcel which arrives at *Source* at *Tsource* reaches the upstream node *U* before the beginning of the hydraulic period i, t_{i-1}.

a) If $Tsource - \tau_i^p \geq t_{i-1}$, add *U* and the associated time $Tsource - \tau_i^p$ to the matrix *Track* and actualize the values of *Source* and *Tsource*.

b) Otherwise:

Step 2.1 - Calculate the distance travelled by the water during the hydraulic period *i*: $x_i^p = u_i^p \times (Tsource - t_{i-1})$;

Step 2.2 - Consider the previous hydraulic period and add the distance travelled during the period until reach the period *i** in which $\sum_{j=i:i^*} x_j^p \leq 0$ or $\sum_{j=i:i^*} x_j^p \geq l^p$.

Step 2.3 - If $\sum_{j=i:i^*} x_j^p \leq 0$: *Source* = *U*. $Tsource = t_{i^*} + \dfrac{\sum_{j=i:i^*} x_j^p}{u_{i^*}^p}$.

If $\sum_{j=i:i^*} x_j^p \geq l^p$: *Source* = *U* $Tsource = t_{i^*} - \dfrac{l^p - \sum_{j=i:i^*} x_j^p}{u_{i^*}^p}$.

Step 2.4 – Add the new values of *Source* and *Tsource* to the matrix *Track*.

Step 3 – If $Tsource > 0$, return to *Step 1*. Otherwise, finish the algorithm.

8.2.3 Algorithm C

Step 1 - Define the number and localization of sensors able to cover the entire network. This step is only necessary for simulated situations, when there isn't installed any set of sensors in the DWDS.

Step 2 - For each sensor, verify if the contaminant concentration was modified and register the time for the first modification *T*.

Step 3 - For each sensor that detected at least one contamination, backtrack the contaminated water parcel that arrived firstly at the sensor, using the Algorithm B determining every possible origin of this water parcel, in short every possible contamination location *Source*, with its respective time of contamination *Tsource*.

Step 4 - Search for possible contaminations that could explain all the information gathered by the sensors:

> *Step 4.1:* Create a vector *A* constituted by the sensors that detected at least one contamination. Create a matrix *B* composed by the entire set of sensors coupled with the times of their first detection or, in cases which there were no detections, the time horizon.
>
> *Step 4.2:* Define *C* as the set of possible contaminations associated with the first sensor presented in vector *A*, and compare it with each set of possible contaminations associated to the remaining sensors belonging to vector *A*. Remove the first sensor in A and, for each remaining sensors, if there is any possible contamination coincident with an element of C, eliminate the others elements of *C* which are not coincident and remove that sensor from vector *A*. If, for a given sensor, there is no contamination coincident, it is concluded that the sensor registered another contamination. Thus, that sensor is kept in vector *A* for further evaluation and *C* is maintained unmodified.
>
> *Step 4.3:* For each possible contamination represented in *C*, track the water parcel that first leaves the contamination location, using the Algorithm A. If it is verified that the contaminated water parcel would reach any sensor before the coupled time of detection presented in matrix *B*, eliminate this possible contamination.
>
> *Step 4.4:* Verify the number of elements of the vector *A*. If it is empty, save the information related to the contaminations stored in *C* and the algorithm is finished. Otherwise, save the information related to the contaminations stored in *C*, set $Ncont = Ncont + 1$ and return *Step 4.2*.

8.3 Results and Discussion

To demonstrate the performance of the proposed method, two different networks are analysed. Network A, presented in Section 5.1, was used to explain in detail the procedure. Network C, presented in Section 5.3, was used to show the performance of the algorithms for real applications.

For both networks, the occurrence of contamination at given contamination sources was simulated. A set of sensors was defined for each network, and the results from the simulation for sensor locations were used to test the performance of the algorithms.

8.3.1 Network A

This example considered the existence of 3 sensors located at nodes N8, N15 and N18, as it can be seen in Figure 8.1. The algorithm was tested for two simultaneous contaminations at nodes R1 and N14, occurring at t= 5 h (18000s). The transport of the contaminants throughout the network was simulated, using Algorithm A, to determine the time of the first detection for each sensor. The sensors registered the changes in contaminant concentration at 18850 s, 21513 s and 21000 s, respectively.

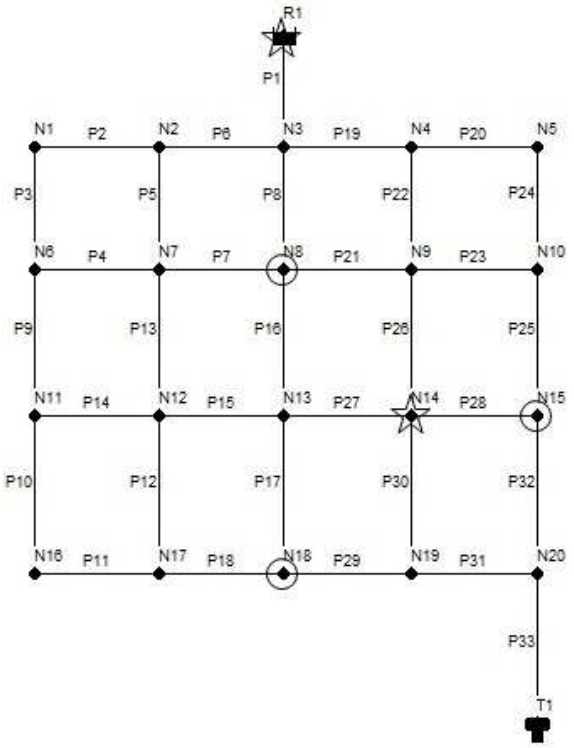

Figure 8.1 – Location of sensors and contamination sources in Network A.

Applying the Algorithm B it was possible to evaluate all possible contamination sources that could explain a positive reading in each sensor. For instance, for the sensor located at N18: $Source = N18$; $Tsource = 21000$; $Track = [N18 \quad 21000]$.

Tsource was contained in time period 7; $N(7, N18) = 2$; $LinksPrev = [P17 \quad P18]$;

$Tsource - \tau_7^{P17} = 21000 - 970 = 20030; t_6 = 18000;\ U = N13;$

$Tsource - \tau_7^{P18} = 21000 - 2836 = 18164; t_6 = 18000; U = N17;$

$$Track' = \begin{bmatrix} N18 & N13 & N17 \\ 21000 & 20030 & 18164 \end{bmatrix}.$$

Repeating this procedure for the 2 new couples $[Source \quad Tsource]$, the updated transposed matrix *Track* was: $$Track' = \begin{bmatrix} N18 & N13 & N17 & N12 & N8 & N14 & N16 & N12 \\ 21000 & 20030 & 18164 & 18654 & 16315 & 18000 & 14434 & 17535 \end{bmatrix}.$$

Running the Algorithm B while $Tsource > 0$ and $N(i, Source) > 0$, the set of possible contaminations that explain the sensor positive reading was achieved. Tables 8.1, 8.2 and 8.3 present the sets of possible contaminants which correspond to sensors located at N8, N15 and N18, respectively.

Table 8.1- Possible contaminations related to sensor located at N8.

N8	N3	R1
18850	18209	18000

Table 8.2 - Possible contaminations related to sensor located at N15.

N15	N10	N9	N5	N8	N4	N4
21513	20743	19568	18881	18850	18200	18272
N3	N3	N3	R1	R1	R1	
18209	17928	18000	18000	17719	17790	

Table 8.3 - Possible contaminations related to sensor located at N18.

N18	N13	N17	N12	N8	N14	N16	N12	N7	N11	N3
21000	20030	18164	18654	16315	18000	14434	17535	17796	14440	15674
N9	N14	N11	N7	N11	N6	N2	N8	N6	R1	N8
17275	14455	14407	16677	14428	14285	14377	16903	14412	15465	16556
N4	N9	N13	N6	N2	N8	N6	N1	N3	N3	N1
15907	14442	14439	14379	14336	15785	14400	14245	14347	16263	14371
N3	N3	N8	N4	N12	N8	N1	N3	N3	N1	N2
15915	15634	14427	14402	14423	14363	14338	14306	15144	14359	14204
R1	R1	N2	R1	R1	N3	N3	N7	N11	N3	N2
14338	16053	14331	15706	15425	14402	14392	14400	14356	14338	14298
R1	R1	N2	N3	N3	R1	R1	N6	N2	N8	N6
14297	14934	14319	14174	14301	14393	14383	14222	14192	14355	14328
R1	N3	N3	R1	R1	N1	N3	N3	N1	R1	R1
14329	14268	14289	14166	14292	14182	14162	14330	14287	14259	14280
N2	R1	R1	N2	N3	N3	R1				
14141	14153	14321	14247	14111	14217	14102				

To start the algorithm for the localization of contamination sources, vector *A* and the matrix *B* were defined *and* matrix *C was initialized*:

$$A = [N8 \quad N15 \quad N18];\ B = \begin{bmatrix} N8 & N15 & N18 \\ 18850 & 21513 & 21000 \end{bmatrix}; C = \begin{bmatrix} N8 & N3 & R1 \\ 18850 & 18209 & 18000 \end{bmatrix}.$$

The comparison between the possible contaminations represented in C with the set of possible contaminations related to sensor located at N15 verified that every element in *C* coincided with the set of possible contaminations. *A* and *C* were updated:

$$A = [N18]; C = \begin{bmatrix} N8 & N3 & R1 \\ 18850 & 18209 & 18000 \end{bmatrix}.$$

From the comparison of the possible contaminations represented in C with the set of possible contaminations related to sensor located at N18, it was possible to conclude that there were no coincident contaminations. This result determined that the sensor located at N18 registered another contamination. Applying the Algorithm A to the first contamination presented in *C*:

$Node = N8$; $Tcont = 18850$; $Distribution = [N8 \quad 18850]$.

Tcont was contained in time period 7; $N(7, N8) = 3$; $LinksPost = [P7 \quad P16 \quad P21]$;

$Tcont + \tau_7^{P7} = 18850 + 892 = 19742; t_7 = 21600;\ D = N7$;

$Tcont + \tau_7^{P16} = 18850 + 3715 = 22565; t_7 = 21600; Tcont + \tau_7^{P16} > t_7$, the contaminant did not reach the downstream node during time horizon.

$Tcont + \tau_7^{P21} = 18850 + 718 = 19568; t_7 = 21600;\ D = N9$;

$$Distribution' = \begin{bmatrix} N8 & N7 & N9 \\ 18850 & 19742 & 19568 \end{bmatrix}.$$

Repeating this procedure for the 2 new couples $[Node \quad Tcon]$, the actualized transposed matrix *Distribution* was:

$$Distribution' = \begin{bmatrix} N8 & N7 & N9 & N12 & N10 & N14 \\ 18850 & 19742 & 19568 & 20600 & 20743 & 20293 \end{bmatrix}$$

Running the Algorithm B while $Tcont < Thorizon$ and $N(i, Node) > 0$, it was possible to evaluate the distribution of the contaminant from the given possible source throughout the entire DWDS. Tables 8.4, 8.5 and 8.6 present the matrix *Distribution* for each possible contamination of *C*.

Table 8.4 - Distribution of contaminant from source [N8 18850].

N8	N7	N9	N12	N10	N14	N17	N15	N19
18850	19742	19568	20600	20743	20293	21230	21513	21129

Table 8.5 – Distribution of contaminant from source [N3 18209].

N3	N2	N8	N4	N1	N7	N9	N5	N9	N6
18209	18824	18850	18481	19803	19742	19568	19091	19849	21014
N12	N10	N14	N10	N10	N14	N17	N15	N19	N19
20600	20743	20293	20953	21024	20574	21230	21513	21129	21410

Table 8.6 – Distribution of contaminant from source [R1 18000].

R1	N3	N2	N8	N4	N1	N7	N9	N5	N9	N6
18000	18209	18824	18850	18481	19803	19742	19568	19091	19849	21014
N12	N10	N14	N10	N10	N14	N17	N15	N19	N19	
20600	20743	20293	20953	21024	20574	21230	21513	21129	21410	

By analysing each matrix *Distribution,* it was concluded that in case of the occurrence of a contamination at any point of *C* the contaminant would not reach any sensor before the coupled time of detection presented in matrix *B*. So there were no possible contaminations to eliminate in *C*. Then, it was possible to conclude that there was one contamination with 3 possible contamination sources described by each couple of matrix *C*.

$$C = \begin{bmatrix} N8 & N3 & R1 \\ 18850 & 18209 & 18000 \end{bmatrix}.$$

Since vector *A* was not yet empty, the occurrence of another contamination was considered. Matrix *C* was initialized as the set of possible contaminations related to sensor located at N18 and this sensor was eliminated from vector *A*. There were no more sensors in vector A and the algorithm skips Step 4.2. Running Algorithm B for each possible contamination source presented in *C* enables the evaluation of the distribution of the contaminant for the different scenarios. After analysing all matrixes Distribution, it was concluded that the contaminant would reach at least one sensor before the coupled time of detection presented in matrix *B,* for several possible contamination sources. It was necessary to update *C* by eliminating those elements:

$$C = \begin{bmatrix} N14 & N14 & N16 & N17 & N18 \\ 14455 & 18000 & 14434 & 18164 & 21000 \end{bmatrix}.$$

Vector *A* was empty so the algorithm was stopped.

8.3.2 Network C

This example considered two different sets of sensors; Set 1 (50 sensors) and Set 2 (10 sensors). Set 2 is a subset of Set 1. The localization of the sensors was defined without obeying to any specific criterion, having only the objective to cover the majority of the DWDS. The results for each set were analysed for different time horizons: 3 h, 6 h, 12 h and 24 h.

The method was tested for 5 simultaneous contaminations, at nodes A, B, C, D and E, occurring at t=0 h. The transport of the contaminants throughout the network was simulated, using Algorithm A, for determining the times of the first detection for each sensor.

Figure 8.2 presents the localization of the simulated contamination sources, represented by coloured diamonds, and the localization of the sensors. The sensors belonging to Set 2 are represented by green triangles, which also belong to Set 1 along with the sensors represented by green squares.

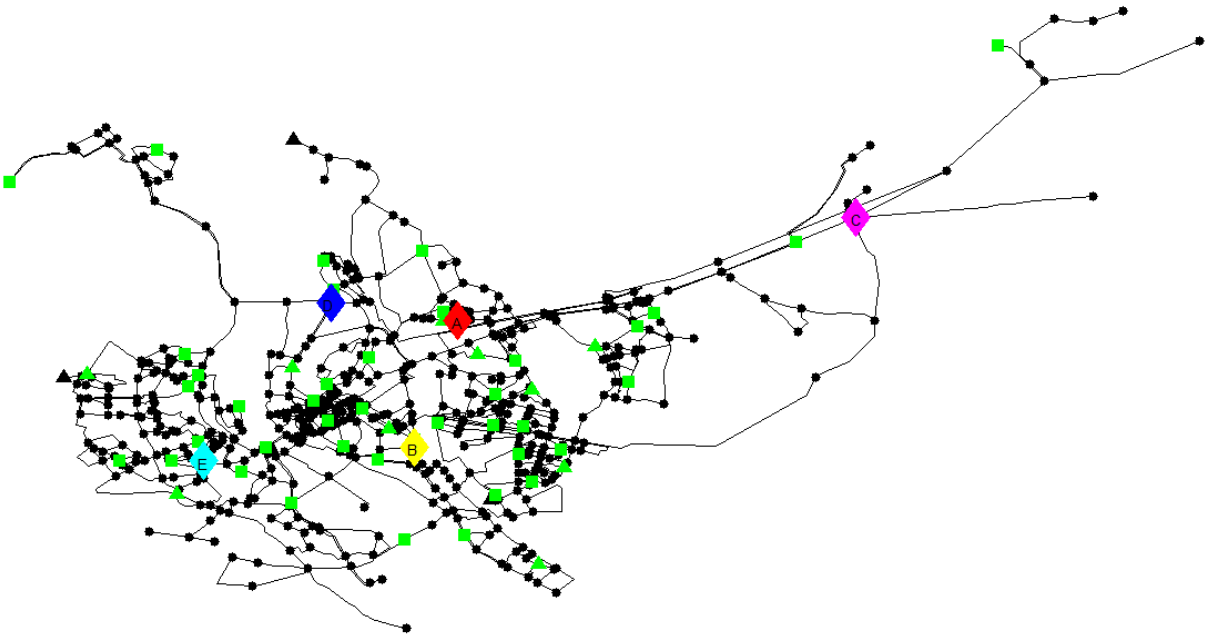

Figure 8.2 - Location of sensors and contamination sources in Network C.

Figures 8.3-8.6 show the results obtained by the proposed method for the different time horizons, using the information provided by Set 1. Figures 8.7-8.10 present the results using the information provided by Set 2. In these two groups of figures, the possible contamination sources are presented as diamonds with the same colour associated to each contamination in Figure 8.2. The sensors that detected one contamination are represented as circles with the same colour correspondent to the associated contamination. The sensors that didn't detect any contamination are presented as green circles. In some cases, for instance in both contaminations presented in Figure 8.3, the sensor which detected the contamination was, at the same time, a possible contamination source, and was represented in diamond shape. In addition, the occurrence of different contamination sources with the same location but with a different instant of contamination is also possible. In Figures 8.3-8.10, these multiple possibilities of contamination sources at one single node are represented only by one diamond. For instance, Figure 8.10 shows 78 possibilities for the location of the contamination E but, in fact, there are 134 possible contamination sources for that contamination. For a time horizon of three hours, only three sensors belonging to Set 1 detected contaminations (Figure 8.3). Applying the proposed method to the information gathered by the sensors, it was concluded that there were two contaminations, each one with two possible contamination sources. Extending the analysis for a time horizon of six hours, there are three more sensors that detected contaminations (Figure 8.4). With this additional information, the occurrence of another two contaminations was verified, one with two and the other with 13 possible contamination sources. For this time horizon, it was still not possible to restrict the possibilities of contamination sources for the contaminations detected previously. For a time horizon of 12 hours, there are 13 sensors that detected contaminations and it was possible to identify the position of the five contaminations simulated (Figure 8.5). Besides that, the possibilities of contamination sources for several contaminations are more restrict. For instance, contamination E (cyan) has 13 possible contamination sources for a time horizon of six hours, and only one for a time horizon of 12 hours. Finally, for a time horizon of 24 hours, 30 sensors detected contaminations. All the simulated contaminations were successfully identified and the number of possibilities of contamination source was reduced to two for the contamination B, and to one for the remaining cases (Figure 8.6).

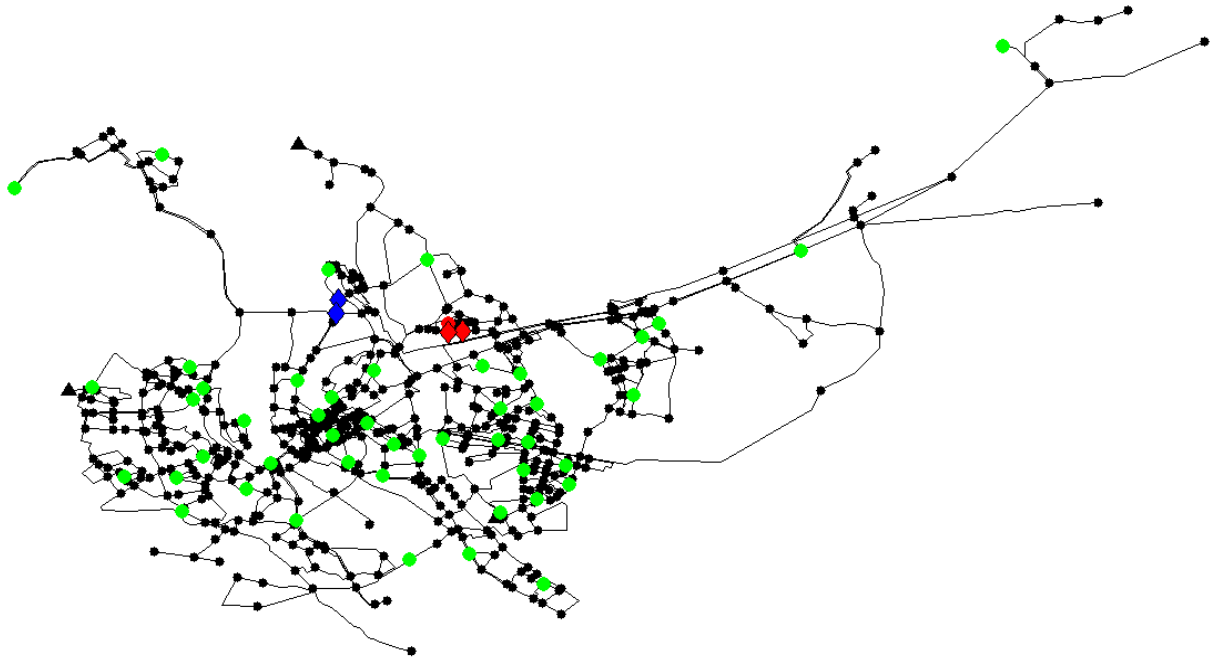

Figure 8.3 - Results for a time horizon of 3 hours – Set 1.

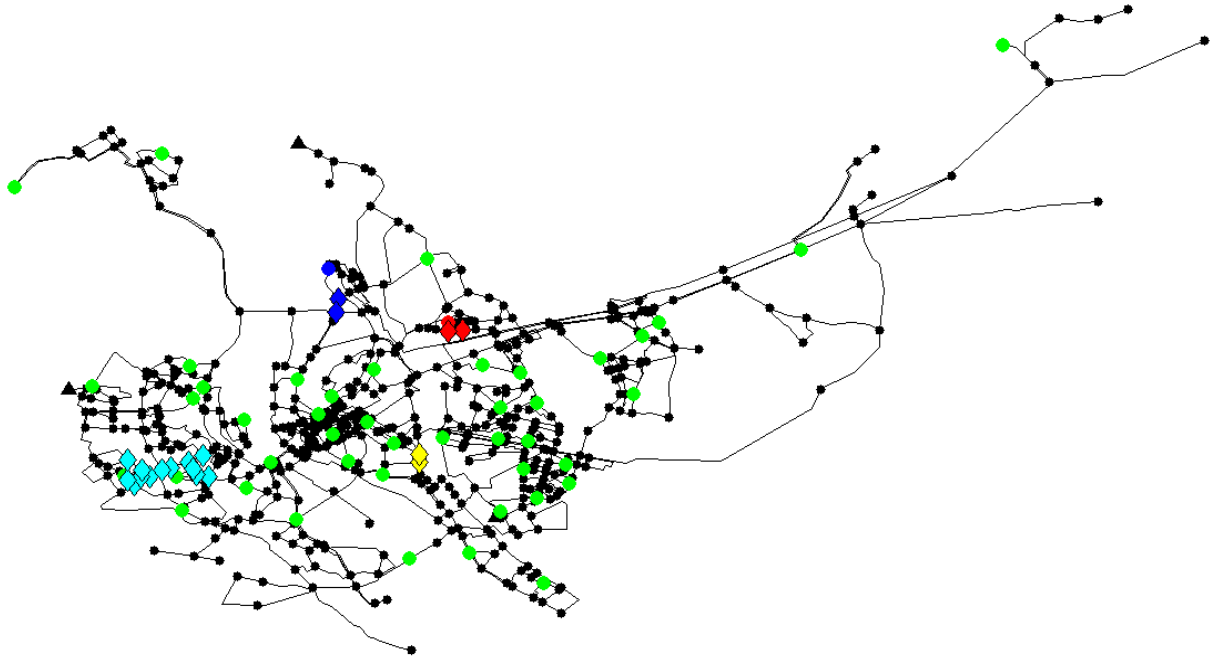

Figure 8.4 - Results for a time horizon of 6 hours – Set 1.

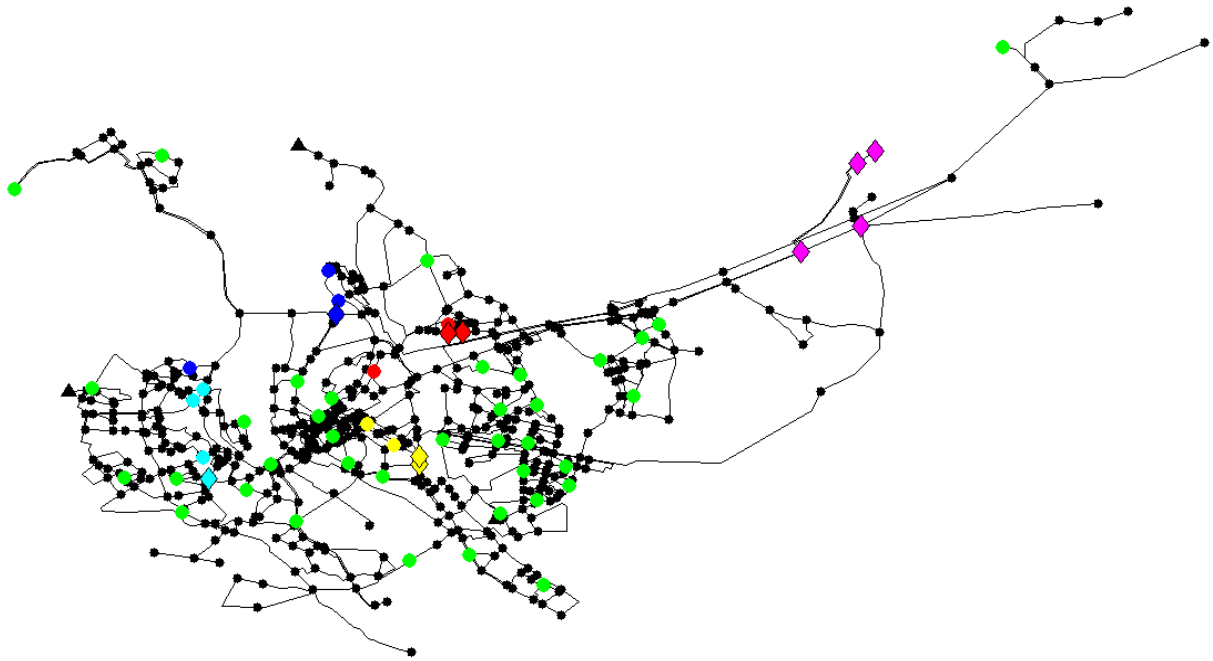

Figure 8.5 - Results for a time horizon of 12 hours – Set 1.

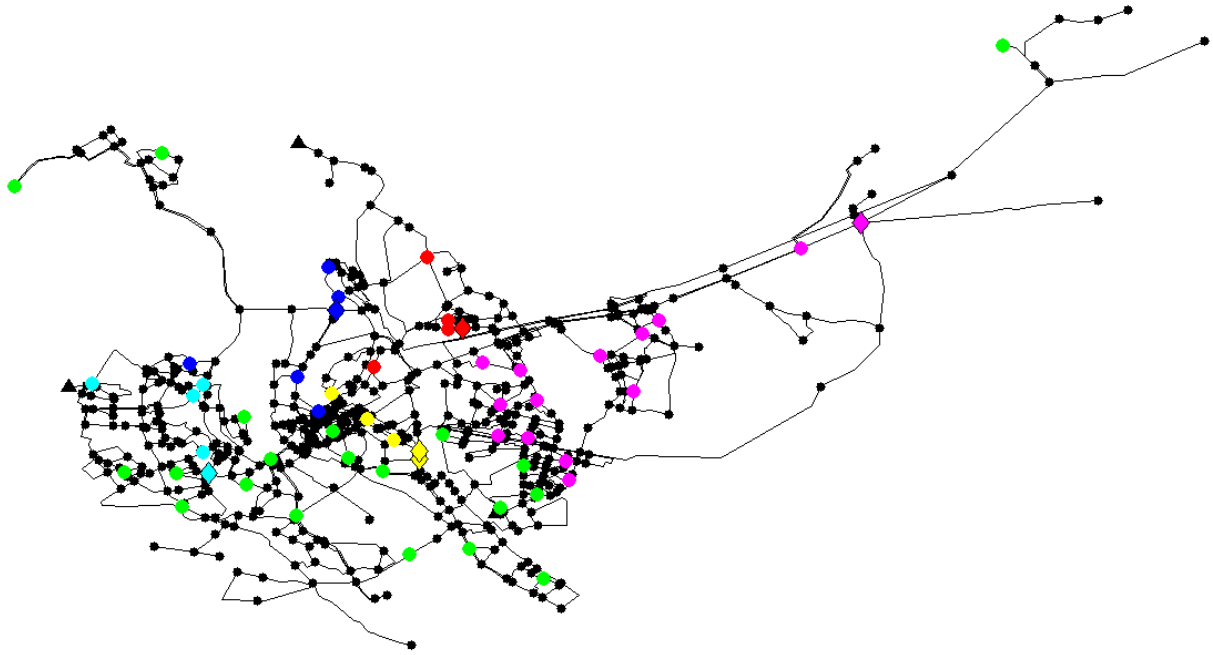

Figure 8.6 - Results for a time horizon of 24 hours – Set 1.

Performing the same analysis of the results obtained with Set 2, it is observed that, for time horizons of three and six hours, only one sensor detected a contamination. The occurrence of a contamination was verified, with the existence of two possible contamination sources (Figures 8.7-8.8). For a time horizon of 12 hours another sensor detected a contamination allowing the localization of another contamination, with 12 possible contamination sources (Figure 8.9). Finally, eight sensors registered

contaminations for a time horizon of 24 hours. All contaminations simulated previously were detected. Contamination A (red) has two, contamination B (yellow) has 11, contamination C (magenta) has one, contamination D (blue) has 19 and contamination E (cyan) has 134 contamination possible contamination sources.

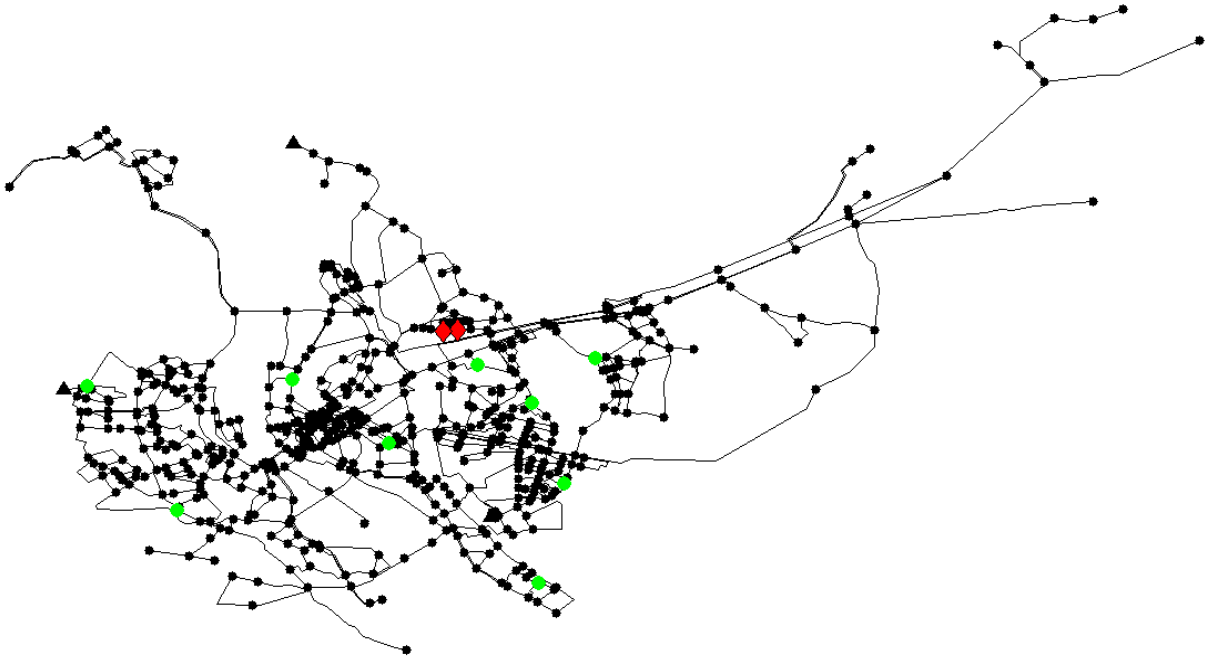

Figure 8.7 - Results for a time horizon of 3 hours – Set 2.

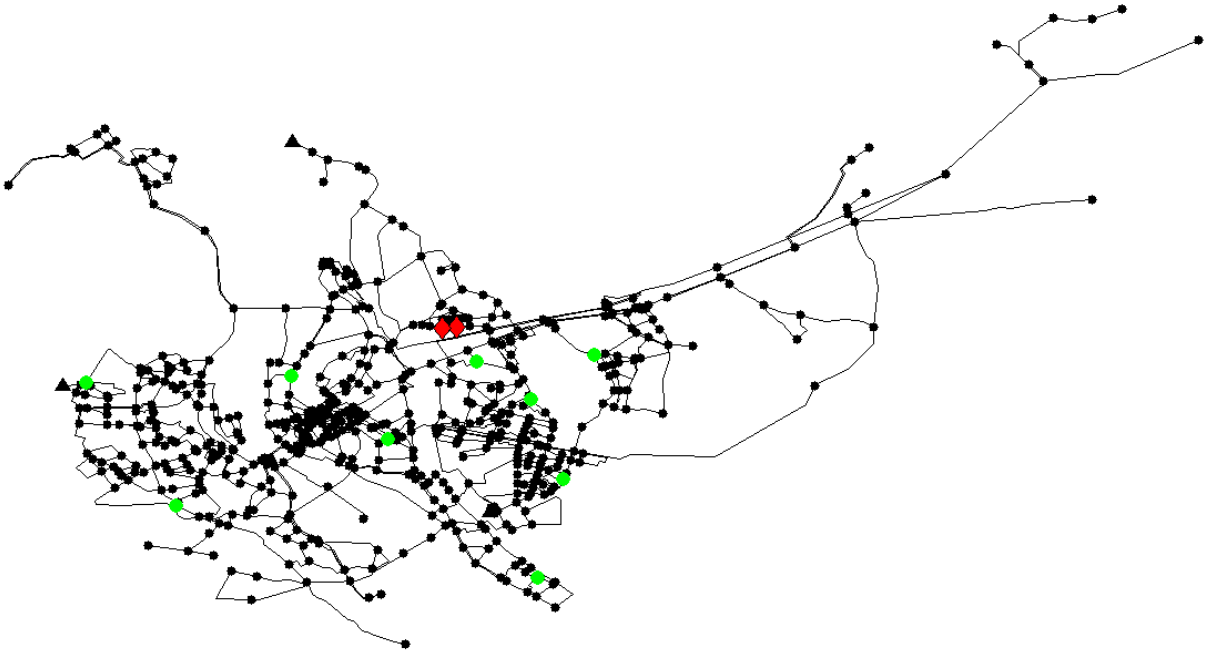

Figure 8.8 - Results for a time horizon of 6 hours – Set 2.

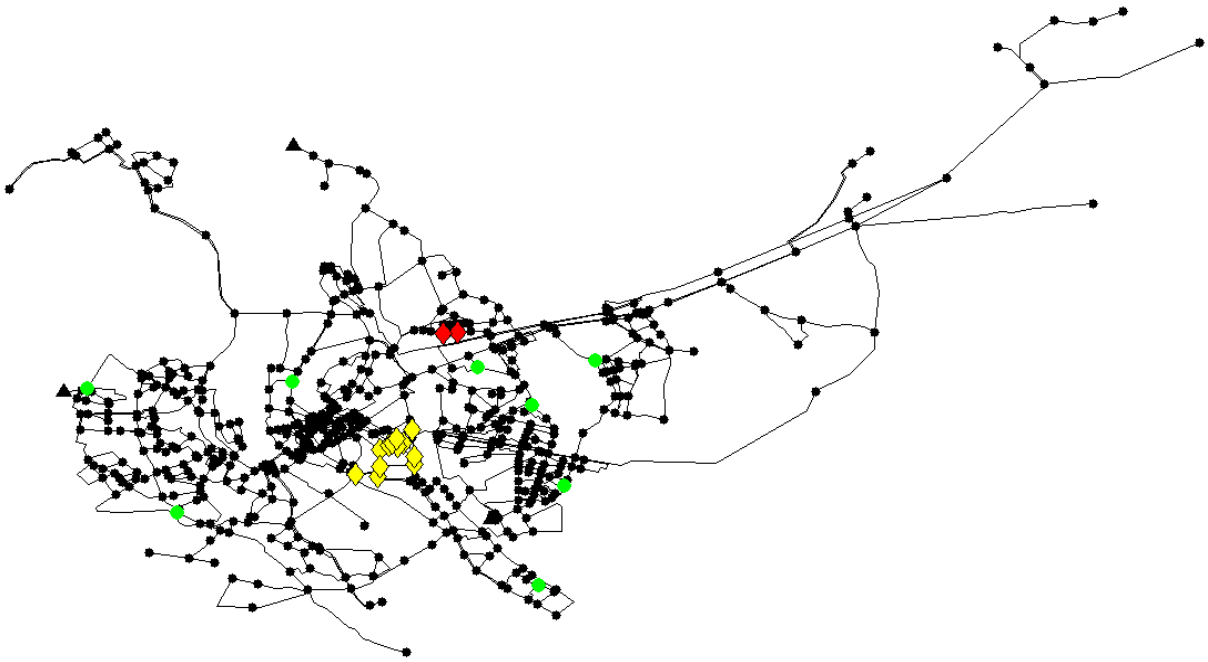

Figure 8.9 - Results for a time horizon of 12 hours – Set 2.

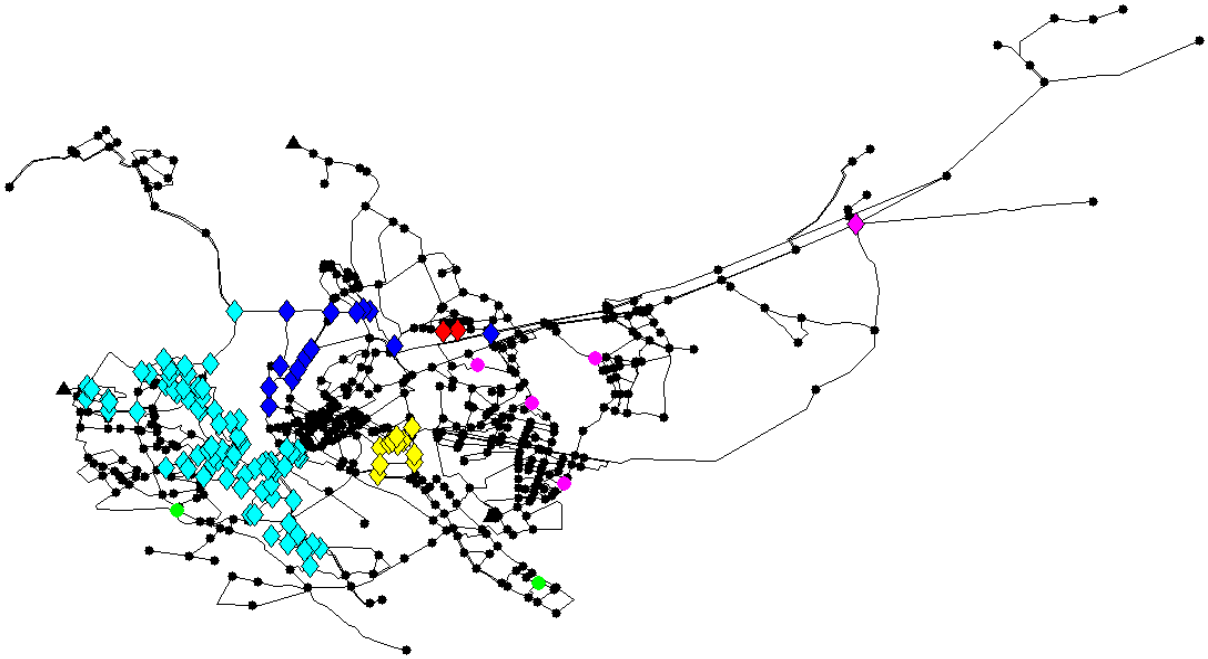

Figure 8.10 - Results for a time horizon of 24 hours – Set 2.

The results achieved by the proposed method, for each time horizon, whatever the set of sensors considered, contained the correct location and instant of each contamination simulated previously.

Through the analysis of the Figures 8.3-8.10, it is possible to observe that: a) for smaller time horizons, there are several contaminations that were not detected until the end of analysis interval; b)

for larger time horizons it was possible to detect a larger number of contaminations and to narrow the set of possible contamination sources.

This happens because, as time passes by, there are more sensors detecting contaminations, and it is possible to assume that the sensors that don't detect any contamination have negative readings for larger intervals. These two facts give more information to restrict the search area.

Comparing the results obtained for the Set 1 with the results obtained for Set 2, it is possible to verify: a) Set 2 needs larger time horizons to detect all contaminations; b) Set 1 is able to give more restrict results for the localization of the contamination sources.

This was expected because, since Set 1 has more sensors than Set 2, there is more information to perform the localization of contamination sources with the Set 1 than with Set 2.

8.4 Chapter Conclusions

A method based on the analysis of the residence time of the water in pipes was proposed to perform the localization of contamination sources in DWDS after an accidental or deliberate contamination.

Several approaches that are described in the literature need to have accurate readings of the contaminant concentration in sensors, which may be a difficult goal to achieve. This method tried to overcome this limitation, being based on only on the residence time of water in pipes and requiring only a binary sensor status over time.

Some tests were run to check the performance of the proposed method for a real DWDS. A scenario with multiple contaminations was simulated to generate the information that would be registered by the sensors throughout the DWDS. The results achieved, for each time horizon, whatever the set of sensors considered, contained the correct location and instant of each contamination previously simulated.

It was possible to observe that: a) for smaller time horizons, there are several contaminations that may not be detected until the end of the analysis interval; b) for larger time horizons, generally, it is possible to detect a larger number of contaminations and to narrow the set of possible contamination sources. Comparing the results obtained for the Set 1 with the results obtained for Set 2, it was possible to verify: a) smaller sets of sensors generally need larger time horizons to detect all contaminations; b) larger sets of sensors are generally able to give more restrict results for the localization of the contamination sources.

9 Localization of Contamination Sources in Drinking Water Distribution Systems: A Method based on Successive Positive Readings of Sensors

9.1 Chapter Overview

Some of the previous works are dependent of the information available for a certain fixed time interval, which may not be suitable for real scenarios, where an answer as quick as possible is required. From the spread of a contaminant in a DWDS, it is possible to obtain different amounts of information depending on the time interval analysed. In this chapter, a different method is proposed to perform the localization of contamination sources based on the information given by successive positive readings of the sensors. With this dynamic approach, it is possible to perform the localization of the contamination sources based on only the first sensor that detected a contamination, and then update the information when more information is available. The proposed method performs the localization of the multiple contamination sources regardless the duration of the contaminant injections, since it assumes the existence of an online surveillance system with continuous information about the contaminant concentration. It is also possible to evaluate the time of contamination, assuming that the hydraulic models is calibrated adequately. These features allow the distinction between any set of simultaneous multiple contaminations detected.

This chapter also addresses the problem of the occurrence of false positives and false negatives at sensors. It is considered as false negative a sensor that fails to detect a contamination that occurs on a DWDS and should be detected by that sensor. On the other hand, a false positive is considered when a sensor wrongly detects a contamination, without the occurrence of any event that could explain that detection. The sensors responses were assumed to be continuous. These situations can have a great influence on the localization of contamination sources, resulting in erroneous outcomes for this problem. The proposed method enables the verification of the occurrence false negatives or false positives and the identification of the sensors that may suffer these effects.

The proposed method is easily applied in real scenarios since only a small amount of information is needed, and its solution can be updated along the time.

9.2 Methods Description

In this section, it is described an algorithm to perform the localization of contamination sources based on the information given by successive positive readings of the sensors, *Algorithm D*.

Algorithms *A* and *B*, presented in Chapter 8, are needed to run this algorithm. *Algorithm A* is used to perform the description of the different paths that a contaminated water parcel can take from a contamination source. *Algorithm B* is used to evaluate all possible contamination sources that could explain a positive reading in each sensor.

This algorithm is able to perform the localization of the contamination source based on only the first sensor that detected a contamination. When additional positive readings are received from other sensors, the information is updated and the possible contaminations locations are redefined. Even without new positive readings, it is possible to update the results obtained for the possible contamination sources associated to the contaminations previously detected.

9.2.1 Algorithm D

Step 1 - Define the number and localization of sensors able to cover the entire network. This step is only necessary for simulated situations, when no set of sensors is installed in the DWDS.

Step 2 – For the set of sensors defined, verify which sensor was the first detecting a change in the contaminant concentration and register the instant in which that change happened *T*. Initialize the number of contaminations, $Ncont = 1$.

Step 3 – Define the *search interval* that will set the size of the initial *analysis interval*. The initial *analysis interval* is defined between an initial threshold (*T-search interval*) and the instant of the first detection (*T*). The search interval has to be defined to ensure that the *analysis interval* includes the correct instants of the contaminations.

Step 4 - Backtrack the contaminated water parcel that arrived firstly at the sensor, using the *Algorithm B* to determine every possible origin of this water parcel, in short, every possible contamination location *Source*, with its respective time of contamination *Tsource*.

Step 5 - Create the matrix *B* associated to the first contamination constituted by the entire set of sensors coupled with the instant of the detection *T*.

Step 6 - For each possible contamination, constituted by the pair [*Source*, *Tsource*], track the water parcel that first leaves the contamination location, using the *Algorithm A*. If it is verified that the contaminated water parcel would reach any sensor before the coupled time of detection presented in matrix *B*, eliminate this possible contamination.

Step 7 - The remaining possible contaminations are the result for the localization of contamination sources for the first detection, *Contamination(1)*.

Step 8 – Associate the sensor and the instant of the detection to the contamination in study.

Step 9 - For each new detection of a change in the contaminant concentration, at an instant *T*:

Step 9.1 – Update the *analysis interval* with the new instant of detection *T*.

Step 9.2 – Perform the *Step 4*. Define *C* as the set of possible contaminations associated with the sensor that registered the last detection.

Step 9.3 – Compare the elements of *C* with the set of possible contamination sources correspondent to each *Contamination*. If there is a matrix *Contamination* with possible contamination sources coincident with elements of *C*, eliminate the others elements of *Contamination* and update the matrix *B* with the instant of the last detection for the sensors that had not registered this contamination (the sensors which had registered the contamination in study keep the instant of their last detection). If there is any contamination source coincident, it is concluded that the sensor registered another contamination. Update the number of contaminations, $Ncont = Ncont + 1$. *Contamination(Ncont)=C*. Associate the sensor and the instant of detection to the new contamination.

Step 9.4 - Create the matrix *B* associated to the new contamination constituted by the entire set of sensors coupled with the new instant of the detection *T*.

Step 9.5 – Perform the *Step 6* for all elements of each matrix *Contamination*.

Step 9.6 – Output: Updated sets of possible contamination sources for each contamination detected.

Step 10 – It is also possible to update the results obtained for the localization of the possible contamination sources of the contaminations detected without new detections:

Step 10.1 – Update the *analysis interval* and each matrix *B* with the new limit of analysis *T* for the sensors that had not registered the respective contamination.

Step 10.2 – Perform the *Step 6* for all elements of each matrix *Contamination*.

Step 10.3 – Output: Updated sets of possible contamination sources for each contamination detected.

9.2.2 False Positives and False Negatives

If false negatives occur and the contaminations are still detected by other sensors, this method has as outcome a set of associated sensors with detections but without any possible contamination sources. After verifying the occurrence of the false negatives, it is possible to determine the corresponding sensors: these are the only sensors that can eliminate alone all the possible contamination sources considered for the contamination that should have been detected.

In the event of false positives at sensors, the outcome of this method would be the determination of new contaminations, which could explain the false positives and respect the negative results at the other sensors. If the contaminations localized after the false positives were supposed to be detected by other sensors, all the possible contamination sources would be eliminated in the *Step 6* of the *Algorithm C* and the outcome would be sensors with a detection but without any contamination associated. If the contaminations localized after the false positives were not supposed to be detected by other sensors, the information available is not sufficient to verify the occurrence of false positives through this method.

The occurrence of false positives and false negatives can have similar outcomes. However, if there is more than one sensor with detections associated to the same contamination but without possible contamination sources, the occurrence of a false negative is verified. One sensor in that situation can be explained either by a false positive on that sensor or by a false negative on a sensor that would be in the flow path of the contamination identified.

9.3 Results and Discussion

Network C was analysed to show the performance of the proposed method. The different sets of sensors defined in Chapter 8 were considered: *Set 1* (50 sensors) and *Set 2* (10 sensors). The method was also tested for the same 5 simultaneous contaminations, at nodes A, B, C, D and E, occurring at t=0 h (see Figure 8.2).

Table 9.1 and Figures 9.1-9.3 show the results obtained by the proposed method using the information provided by the *Set 2*. The results are presented similarly to the results presented in Figure 8.3-8.10.

Table 9.1 - Results for the localization of contamination sources - Set 2.

Detection	Sensor	Contamination	Time (s)	A	B	C	D	E
1	100	A	3303	7	0	0	0	0
2	150	B	26630	7	21	0	0	0
5	100	C	40595	7	8	10	0	0
8	400	D	49834	7	7	6	36	0
11	200	E	56589	6	7	6	10	89
14	450	C	62541	6	7	2	10	69
16	350	C	63850	6	7	2	10	69

Analysing the data gathered by the *Set 2*, the first detection observed occurred at $T = 3303\,s$. The *search interval* was considered to be 6 hours, thus the initial *analysis interval* was set between $t_{initial} = 68103s$ from the previous day and $t_{final} = 3303\,s$. After first detection (Figure 9.1), the localization of 1 contamination (A) was determined with the existence of 7 possible contamination sources. The second detection occurred at $T = 26630\,s$ (Figure 9.2), then the *analysis interval* was updated with the new t_{final}. The *analysis interval* was updated each time a new sensor registered a contamination. With this additional information, it was not possible to restrict the results obtained for the localization of the contamination A, but the occurrence of another contamination (B) was verified with 21 possible contamination sources. The sequence of sensors that registered changes in the contaminant concentration enabled the localization of contaminations that were not detected before but it was not sufficient to improve the results for the localization of the contaminations already detected. The fifth detection enabled the detection of the contamination C, with 10 associated possible contamination sources. Meanwhile, previous detections enable the decrease in the number of possible contamination sources for contamination B to 8. Contamination D was located after the eighth detection, with 36 possible contamination sources. At this point, this methodology enabled decreasing the number of possible contamination sources for contaminations B and C to 7 and 6, respectively. After 11 detections, it was already possible to detect all five simulated contaminations (Figure 9.3). Contamination E has 89 possible contamination sources and it was possible to restrict the results for

the localization of the contaminations A and D to 6 and 10 possible contamination sources, respectively.

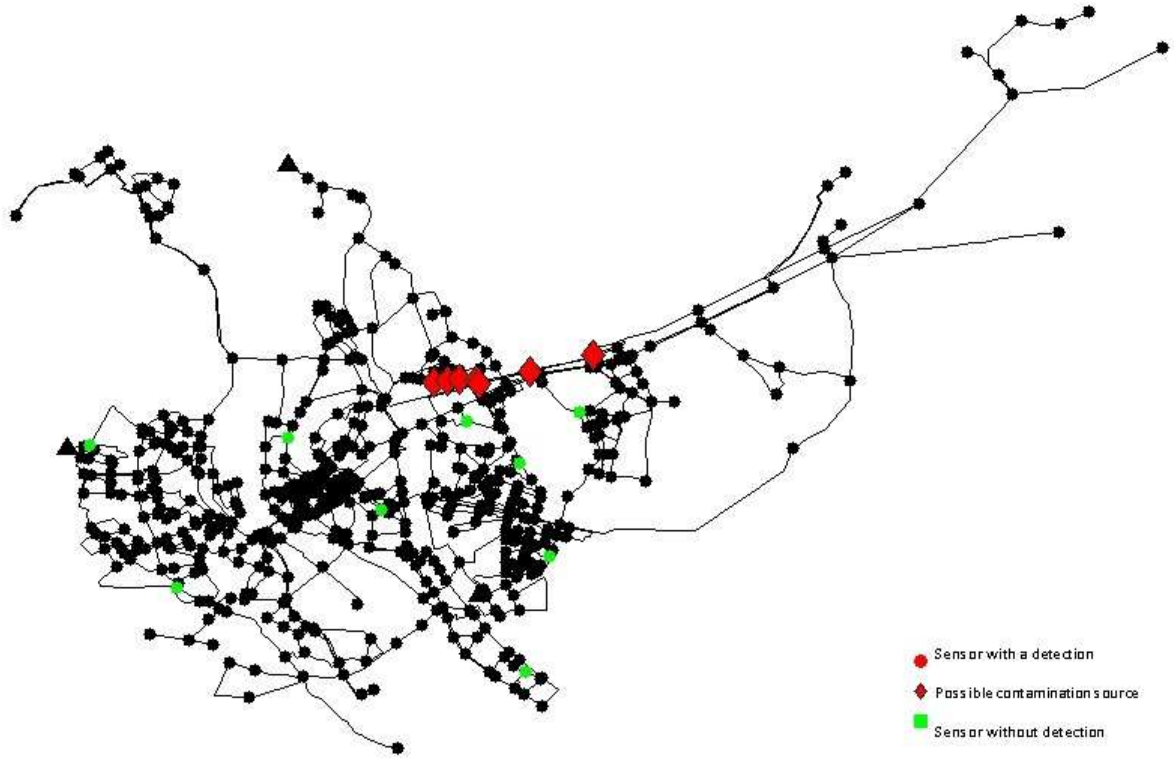

Figure 9.1. Results obtained after 1 detection – *Set 2*.

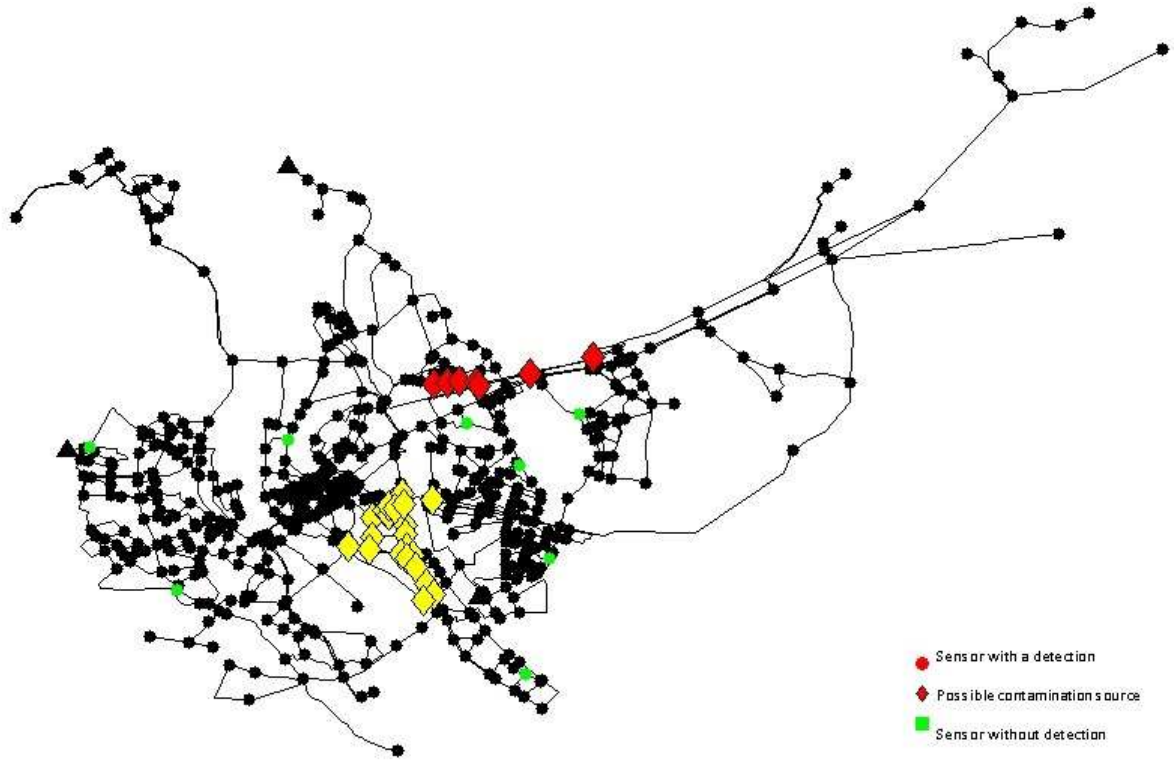

Figure 9.2. Results obtained after 2 detections – Set 2.

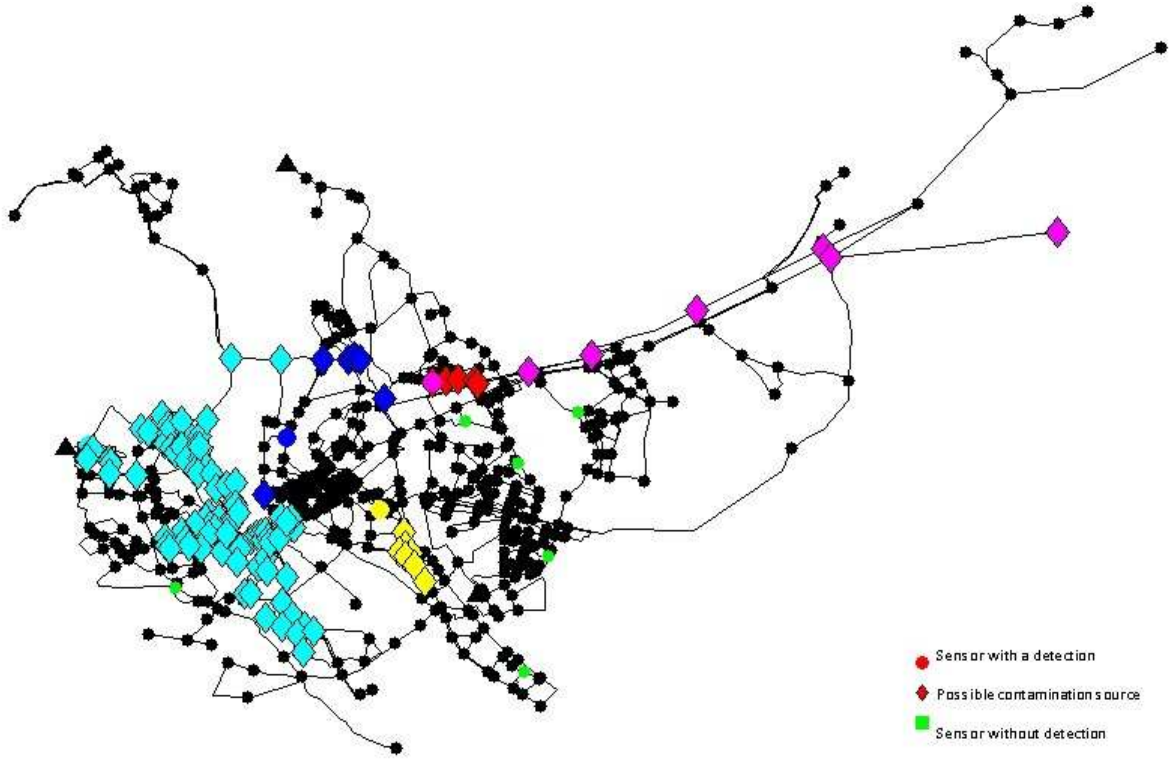

Figure 9.3. Results obtained after 11 detections – Set 2.

Table 9.2 show the results obtained by the proposed methodology using the information provided by the *Set 1*. The first detection observed occurred at $T = 3303\,s$. The *search interval* was again considered to be 6 hour and the *analysis interval* was set between $t_{initial} = 68103\,s$ from the previous day and $t_{final} = 3303\,s$. After the first detection, the localization of 1 contamination (A) was determined with the existence of 7 possible contamination sources, as it has already happened with the *Set 2*. Once again, the *analysis interval* was updated each time a new sensor registered a contamination. The second detection, occurring at $T = 5150\,s$, enabled the localization of a second contamination (D) with 5 possible contamination sources. The fifth and sixth detections enabled the localization of the contaminations B and E with, respectively, 7 and 16 possible contamination sources associated. The number of possible contamination sources for the contamination B decreased to 4. The last contamination (C) was detected by the 22nd detection, at $T = 39043s$, with 5 possible contamination sources associated. At $T = 56589s$, the number of possible contaminations were reduced to 3, 7, 2, 4 and 1 for contaminations A, B, C, D and E, respectively.

Table 9.2 - Results for the localization of contamination sources - Set 1.

Detection	Sensor	Contamination	Time (s)	A	B	C	D	E
1	100	A	3303	7	0	0	0	0
2	410	D	5150	7	0	0	5	0
5	500	B	17790	7	7	0	4	0
6	10	E	18327	7	7	0	4	16
22	470	C	39043	4	7	5	4	1
58	200	E	56589	3	7	2	4	1

The results achieved by the proposed methodology, for each situation analysed, whatever the set of sensors considered, contained the correct locations and the correspondent instants of the contaminations simulated previously.

Through the analysis of Tables 9.1 and 9.2, it is possible to observe that as new sensors detect changes in contaminant concentration, other contaminations are detected and the possibilities for the location of contamination sources may be more restricted. This is due to the increase of the amount of information given by new positive readings in sensors and to the extension of the *analysis interval*. Comparing the results obtained for the Set 2 with the results obtained for Set 1, it is possible to verify that:

a) *Set 1* took less time to detect all simulated contaminations than *Set 2*. This was expected because since as *Set 1* had more sensors with positive readings, there was more available information to perform a faster detection of the simulated contaminations.

b) *Set 1* enabled the achievement of more restricted results for the localization of possible contamination sources for all simulated contaminations. This happens because with *Set 1* there are more sensors detecting contaminations and more sensors that do not detect any contamination as well. These two facts give more information to restrict the search area.

The effect of the occurrence of false negatives in this methodology performance was also studied. As case study, it was tested the occurrence of a false negative at the sensor that should be responsible for the 16th detection, considering the surveillance system defined by the *Set 2*. As it is presented in Table 9.1, this detection is associated to the contamination C. If the results were updated after this false negative, either following *Step 9* or *Step 10* from *Algorithm C*, contaminations A, B, D and E would

still be identified. This happens because these contaminations are not associated with the false negative. However, analysing the previous possible contamination sources for contamination C, considering the new *analysis interval*, the proposed methodology would eliminate every possibilities, by performing the *Step 6* from *Algorithm C*. Thus, there would be several detections without any associated possible contamination source, which constitutes proof of the existence of a false negative. The only sensor that would be able to eliminate every possible contamination sources associated to the contamination C is the sensor #350.

The effect of the occurrence of false positives in this methodology performance was tested by the introduction of false positives occurring at sensor #100 and sensor #300, at $T = 10000\,s$, to the previous simulations (Table 9.3).

Table 9.3 - Results for false positives at sensor #100 and sensor #300.

Detection	Sensor	Contamination	Time (s)	A	B	C	D	E	F	G
1	100	A	3303	7	0	0	0	0	0	0
FP 1	100	-	10000	7	0	0	0	0	7	0
FP 2	300	-	10000	7	0	0	0	0	7	6
17	350	C	63996	6	7	2	10	69	5	6
18	400	A	65129	6	7	2	10	69	0	6

Analysing the information given by the false positives, two different contaminations (F and G) were localized with 7 and 6 possible contamination sources, respectively. After 17 detections, both false positives were still considered to be real contaminations and the proposed methodology tries to reduce the number of possible contamination sources for both cases. However, after 18 detections, with the associated extension in the *analysis interval*, the number of possible contamination sources associated to the contamination F was reduced to 0. This happens because a contamination that was detected at sensor #100, at $T = 10000\,s$, should be detected at sensor #400 at $T = 64656s$. This situation is an anomaly that can either be explained by the occurrence of a false positive at sensor #100 or a false negative at sensor #400.

After 18 detections, it was still not possible to detect any anomaly associated with the contamination G, detected after the false positive 2.

An important final remark is related with the definition of the *search interval*. For instance, if the *search interval* was set as 0.5 h, the initial *analysis interval* would be defined between 1503 s and 3303 s. The correct instants of contamination would be excluded of this *analysis interval*, thus the proposed methodology would not be able to perform a correct localization of the contamination sources.

9.4 Chapter Conclusions

A methodology based on the analysis of the residence time of the water in pipes was proposed to perform the localization of contamination sources in DWDS**,** through the analysis of the information given by successive positive readings on the sensors.

Several other approaches that are described in the literature need to have accurate readings of the contaminant concentration in sensors, which may be a difficult goal to achieve. This methodology tried to overcome this limitation. It is based only on residence time of water in pipes and it only requires a binary sensor status over time. The results for the localization of contamination sources are given sequentially, being updated each time a new sensor detects a change in contaminant concentration. This is an important feature since it enables the reduction of the computation time and the amount of information needed to perform the localization of contamination sources. Furthermore, in some situations, this methodology enables the verification of the occurrence of false negatives and false positives.

Some tests were run to check the performance of the proposed methodology for a real DWDS. A scenario with multiple contaminations was simulated to generate the information that would be registered by the sensors throughout the DWDS. The results achieved, for each set of sensors considered, contained the correct locations and the instants of the contaminations previously simulated. It was possible to observe that as new sensors detect changes in contaminant concentration, other contaminations may be detected and the possibilities for the location of contamination sources may become focused on a more restricted area.

Comparing the results obtained for *Set 1* with the results obtained for *Set 2*, it was possible to verify, as expected that: a) larger sets of sensors generally need smaller *analysis intervals* to detect all contaminations; b) larger sets of sensors are generally able to provide results with more restricted areas for the localization of the contamination sources.

The case study with the occurrence of a false negative showed that the proposed methodology enables the confirmation of the existence of this anomaly if more than one sensor detected the contamination that should have been detected by the sensor which has a false negative. It was observed that the occurrence of false negatives did not affect the results related with the real detections and it was possible to detect the sensor which suffered this anomaly.

In the first case study with the occurrence of a false positive, the existence of an anomaly that could be explained either by a false positive or a false negative was verified. The information available was not sufficient to distinguish between these options through the proposed methodology. In the second case study, it was not possible to verify the occurrence of a false positive, since there was not any other detection associated to the possible contamination sources determined through the analysis of the false positive. It was observed that the occurrence of false positives did not affect the results related with the real detections.

Part IV: Application of Artificial Neural Networks for the Localization of Contamination Sources in Drinking Water Distribution Systems

10 Localization of Contamination Sources in Drinking Water Distribution Systems through the Application of Artificial Neural Networks

10.1 Chapter Overview

This chapter describes the application of ANNs in problems of pattern recognition and non-linear regression to the problem of localization of contamination sources in DWDSs. ANNs are applied to predict the possible localization points of contamination sources. The main advantages of using this approach when compared with the applications presented are lower time of computation and the ability to predict the contamination sources based only on the time of the first detection at a sensor, without the requirement of having an accurate evaluation of the contaminant concentration at the sensors.

Thus, the work here presented aims to apply ANNs: a) for determining the probability of a candidate source of contamination being actually the real source of contamination based on the detection pattern; b) for estimating the corresponding time of contamination for each possible contamination source based on the times of detection of the sensors.

10.2 Methods Description

The procedure proposed in this work is divided in two main stages:

1. Determine the probability of each node being the contamination source based on the detection pattern;
2. Verify each possible contamination source determined in stage 1 and estimate the corresponding time based on the times of detection of the sensors. Update, if necessary, the probability of each node being the contamination source. For this stage, two different approaches are presented, being designated by Approaches A and B, respectively, which can be implemented separately or used to validate each other.

Figure 10.1 presents a schematic diagram with the interactions between these two stages. Both stages require the creation of databases with the information associated to a large number of contamination scenarios. These scenarios shall include contaminations at each node in the network, for several

instants of contamination. *Algorithm A,* described in Section 8.2.1, was used to simulate those contamination scenarios.

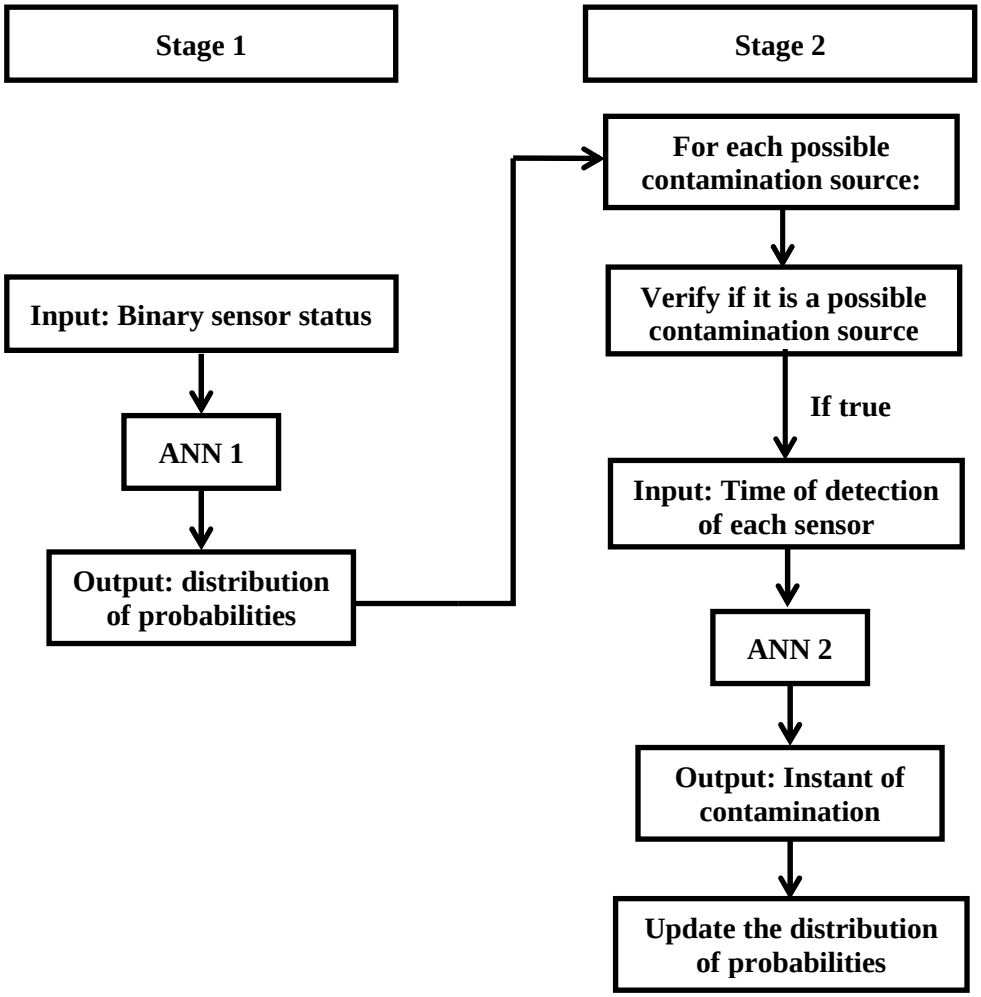

Figure 10.1 - Diagram of the steps involved in solving the problem of the localization of the contamination sources.

10.2.1 Stage 1

The aim of this stage is to have an initial pruning of the set of nodes that are possible contamination sources. Thus, the selected output for the ANN was the probability of each node being the contamination source, setting the number of nodes of the DWDS as the number of outputs of the ANN. For the selection of the input variables, two different approaches were tested considering: a) the time of detection at each sensor (the time of detection was set as $2 \times Thorizon$ when the contamination was not detected by a sensor); b) the binary status for each sensor (1 if the sensor detected the contamination, 0 otherwise). The results obtained considering the time of detection at each sensor were not satisfactory in contrast with the results obtained considering the binary status of

the sensors, so the second alternative was selected. The information provided as an input is a vector with the size of the number of sensors, in which each element presents the status of the corresponding sensor. Meanwhile, the target is the identification of the node where the contamination occurred, defined as a vector with the number of elements equal to the number of nodes of the DWDS. For each contamination scenario, all elements of the target vector are equal to 0 excluding the element corresponding to the node where the contamination was performed.

Thus, for this stage, it is necessary to create a database with the detection pattern (binary status for each sensor) associated to each contamination scenario simulated. The data was attributed entirely to training, since every possible combination of detections at the sensors is included in the data set. It is relevant to note that this stage is a problem of pattern recognition and thus the objective is to achieve ANN structures able to relate a detection pattern to a set of potential possible contamination source. In addition and since the inputs are discrete values and there are several cases when the same inputs are associated with different target values, the ANN structures have no risk of becoming too specific.

The ANNs models were developed using MATLAB software. The selected model architecture was the multilayer perceptron (MLP), since this architecture has previously been widely used for pattern recognition applications (Daqi and Genxing, 2003). The ANNs were constituted by three layers: an input layer, a hidden layer and an output layer. The input layer is constituted by a number of neurons equal to the number of sensors and the output layer constituted by a number of neurons equal to the number of nodes. The number of neurons in the hidden layer and the transfer functions for hidden and output neurons were selected by trial and error method looking for the minimum of sum of squared errors. A deterministic approach using a first-order gradient back-propagation method was implemented to calibrate each ANN, minimizing the sum of squared errors.

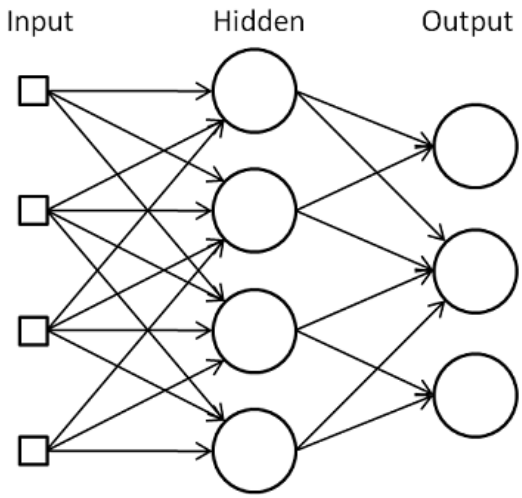

Figure 10.2 -Typical topology of a feedforward artificial neural network.

10.2.2 Stage 2

The aim of this stage is to estimate the time of contamination associated to each location based on the times of detection of the sensors, so each ANN developed has time of contamination associated to a node as output. For the selection of the input variables, two different approaches were tested considering: a) an ANN per sensor, for each node, with the time of detection at that sensor as the input; b) an ANN for each node, with a vector constituted by the time of detection at each sensor as the input (the time of detection was set as $2 \times Thorizon$ when the contamination was not detected by a sensor). The results obtained following both approaches were satisfactory, so both alternatives are described in Sections 10.3.1 and 10.3.2. Meanwhile, the target for each ANN is the time of contamination associated to the corresponding node.

For this stage, the necessary databases are constituted by the time of the first detection at the sensors, associated to each contamination scenario simulated. The data was divided in 60% for training, 20% for testing and 20% for validation. The division was made following a physics based approach, ensuring that each set of data contained points representative of every data subsets.

The ANNs models were developed using MATLAB software. The selected model architecture was the multilayer perceptron (MLP), since this architecture has been the most used architecture in water resources applications (Maier et al., 2010). The ANNs were also constituted by three layers: an input layer, a hidden layer and an output layer, following the topology presented in Figure 10.2. The output layer is constituted by 1 neuron, for both approaches.

The number of neurons in the hidden layer and the transfer functions for hidden and output neurons were selected by trial and error method looking for the minimum of sum of squared errors. A deterministic approach using a first-order gradient back-propagation method was implemented to calibrate each ANN, minimizing the sum of squared errors.

Approach A

For each node of the DWDS, an ANN was trained per sensor that shall detect contaminations occurring at the node under study. The input of each ANN is the time of the first detection at the corresponding sensor, so the input layer is constituted by 1 neuron. The output is the time of contamination associated to the node under study.

Through this approach, it is possible to perform several estimates for the time of contamination associated to a source candidate based on the time of the first detection of each sensor that detected the

contamination. After estimating the different times of contamination associated to the node considered as a possible contamination source, it is possible to verify if that node remains a candidate. If the estimates obtained using the different ANNs are similar, the node is confirmed as a possible contamination source and the average between the estimates obtained by each ANN is chosen as the time of contamination associated to the node. If there is a discrepancy between the estimates obtained by the different ANNs, it is concluded that the node under study is not a possible contamination source. This discrepancy is due to the fact that the set of values considered for the time of first detection at the sensors does not correspond to a contamination on the node in study.

After verifying which nodes remain as possible contamination sources, the corresponding values of probability calculated in Stage 1 are updated, ensuring that the set of possible contamination sources has a probability density function equal to 1.

Approach B

In this approach, an ANN was trained for each node of the DWDS. Each ANN has the time of first detection at each sensor as input. When a contamination is not detected by a sensor, the time of detection at this sensor is set as $2 \times Thorizon$. As in the Approach A, the output is the time of contamination associated to the node in study.

This approach also enables to verify if the node under study remains as a possible contamination source. It is necessary to verify the position of the time of the first detection of one sensor within the training set used to train the ANN. Then, it is evaluated if the ratios between the times of the first detection given by the sensors are close to the ratios between the times of the first detection in the correspondent position in the training set. If this condition is verified, the node is a possible contamination source and the ANN associated to this node is used to estimate the corresponding time of contamination. Otherwise, the set of values considered for the time of first detection at the sensors are not corresponding to a contamination on the node under study, so it is no longer considered as a possible contamination source.

After verifying which nodes remain as possible contamination sources, the corresponding values of probability calculated in Stage 1 are updated, ensuring that the set of possible contamination sources has a probability density function equal to 1.

10.2.3 Example of Application

The Network B, presented in the Section 5.2 was used as case study. A set of 3 sensors, presented in Figure 10.3 as squares, were defined to test the performance of the proposed method. Several ANNs were trained to perform the tasks defined in each stage of the method, following the different approaches. The databases used in the training of these ANNs were constituted by contamination scenarios occurring at each node at each 10 min.

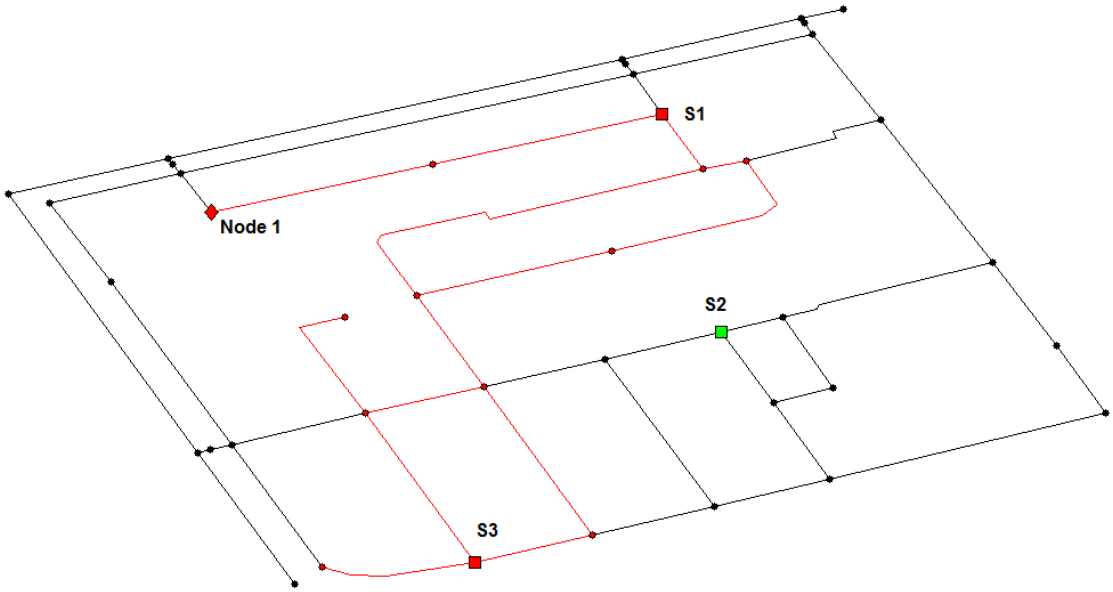

Figure 10.3 - Representation of the DWDS in study

10.2.4 Creation of the Databases - Example of a Contamination Scenario

A contamination occurring at Node 1 (Figure 10.3 in diamond shape), at t = 8.33 h was simulated through the application of the algorithm presented in Section 8.2.1. Analysing the contamination plume generated in this contamination scenario (presented in red), it is verified that the contaminant would be detected by sensors S1 and S3 at t = 10.00 h and t = 8.84 h, respectively. This contamination would remain undetected by the sensor S2.

Figure 10.4 presents the creation process of the databases and explains the contribution of each contamination scenario to the inputs and targets of each ANN.

In the Stage 1, all elements of the target vector are equal to 0 excluding the element corresponding to Node 1, which is set as 1.

Meanwhile, in the Stage 2 following the Approach A, this contamination scenario contributed to the training of 2 different ANNs having the time of the first detection of each corresponding sensor as inputs and the time of contamination associated to Node 1 as target. Since sensor S2 does not detect any contamination occurring at Node 1, any ANN has been trained for this sensor. Following Approach B, an ANN was trained with the set of the times of the first detection of each sensor as input and the time of contamination as target. As the sensor S2 does not detect the contamination, the time of first detection associated to that sensor was as set as $2 \times Thorizon$, 48 h.

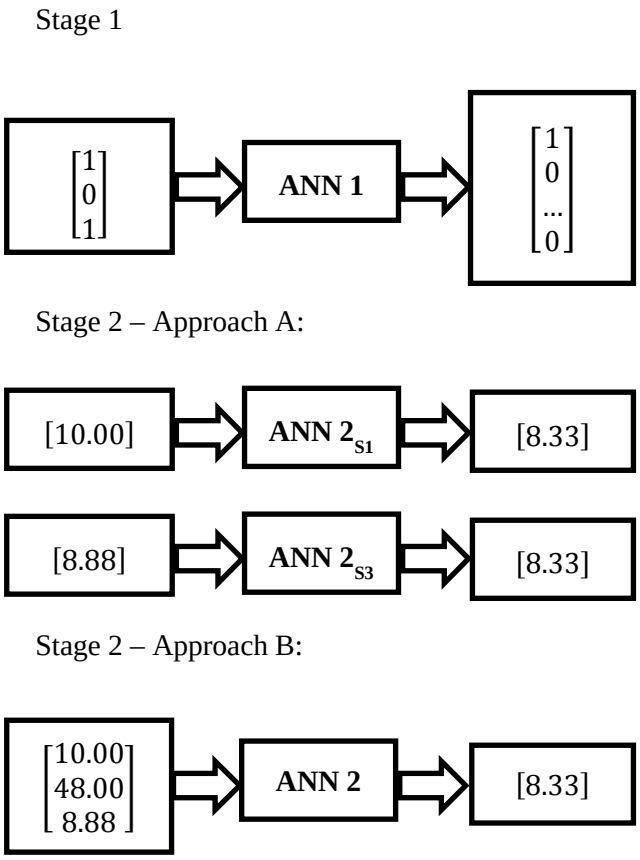

Figure 10.4 - Example of the contribution of a contamination scenario to the inputs and targets of each ANN.

10.3 Results and Discussion

The ANNs necessary to perform the proposed method were developed for the DWDS presented in the Section 10.2.3. The number of neurons in the hidden layer and the transfer functions for hidden and

output neurons were selected as it was described in Sections 10.2.1 and 10.2.2. For the Stage 1, the hidden layer was constituted by 20 neurons and a logarithmic sigmoid and a linear transfer function were selected for the hidden and output neurons, respectively. For both different approaches developed for Stage 2, a logarithmic sigmoid and a linear transfer function were also selected for the hidden and output neurons, respectively. The hidden layer is constituted by 3 and 6 neurons for approaches A and B, respectively.

A contamination occurring at Node 1 (Figure 10.3), at t = 3.47 h was simulated. This contamination scenario had not been used in training the ANNs. Using the algorithm presented in Section 8.2.1, the contamination plume was determined and it is verified that sensors S1 and S3 detect a contamination at t = 7.67 h and t = 4.86 h, respectively. Sensor S2 does not detect any contamination.

Beginning with the Stage 1, the input to the ANN1 is the vector [1 0 1]′. The output of the ANN1 is a vector with 41 elements. Among them, there are 9 nodes with a probability of being the contamination source equal to 0.11. The remaining nodes have a probability of being the contamination source equal to 0, so they are not possible contamination sources.

In the Stage 2, each possible contamination source determined in the Stage 1 has to be verified. In Sections 10.3.1 and 10.3.2, Nodes 1 and 2 (both nodes were indicated in Stage 1 as possible contamination sources), will be presented as examples of possible contamination sources confirmed and eliminated by Stage 2, following the different approaches developed for this stage.

10.3.1 Case Study - Approach A

Following the Approach A, the time of contamination associated to each possible contamination sources is estimated using the ANN trained with the information related with each single sensor.

Using the ANN 2_{S1} trained for the Node 1 with the input t_{S1} = 7.67 h, the estimate for the time of contamination associated to Node 1 is t = 3.41 h. Using the ANN 2_{S3} trained for Node 1 with the input t_{S3} = 4.86 h, the estimate obtained is t = 3.47 h. The deviation of these two values in relation to the average time of contamination is 0.02 h, so it is verified that Node 1 is still a possible contamination source. The average between the estimates obtained by each ANN is chosen as the time of contamination associated to Node 1. Applying the same procedure to the Node 2, the estimate for the time of contamination associated to Node 2, using the ANN 2_{S1}, is t = 5.29 h and the estimate obtained with the ANN 2_{S3} is t = 3.29 h. The deviation of these two values in relation to the average time of

contamination is 1.00, so it is verified that the Node 2 is not a possible contamination source. The results obtained following this Approach for Nodes 1 and 2 are presented in Table 10.1.

Table 10.1- Results achieved following Approach A for Nodes 1 and 2.

	Node 1			**Node 2**		
	Estimate (h)	**Average (h)**	**Deviation (h)**	**Estimate (h)**	**Average (h)**	**Deviation(h)**
S1	3.51	3.49	0.02	5.29	4.29	1.00
S3	3.47		0.02	3.29		1.00

By analysing the 9 possible contamination sources determined in the step 1, it is possible to verify that only 6 of them are possible contamination sources. So the probability of each possible contamination source confirmed in the step 2 is updated to 0.167. The estimates of the time of contamination associated to each possible contamination source are presented in Table 10.2.

Table 10.2 - Estimates of time of contamination associated to each possible contamination source obtained following both approaches

Node	**Probability**	**Estimate A (h)**	**Estimate B (h)**	**Expected result (h)**
1	0.167	3.49	3.48	3.48
5	0.167	4.68	4.72	4.72
6	0.167	4.69	4.70	4.72
7	0.167	4.83	4.79	4.80
8	0.167	4.86	4.86	4.86
9	0.167	4.24	4.21	4.24

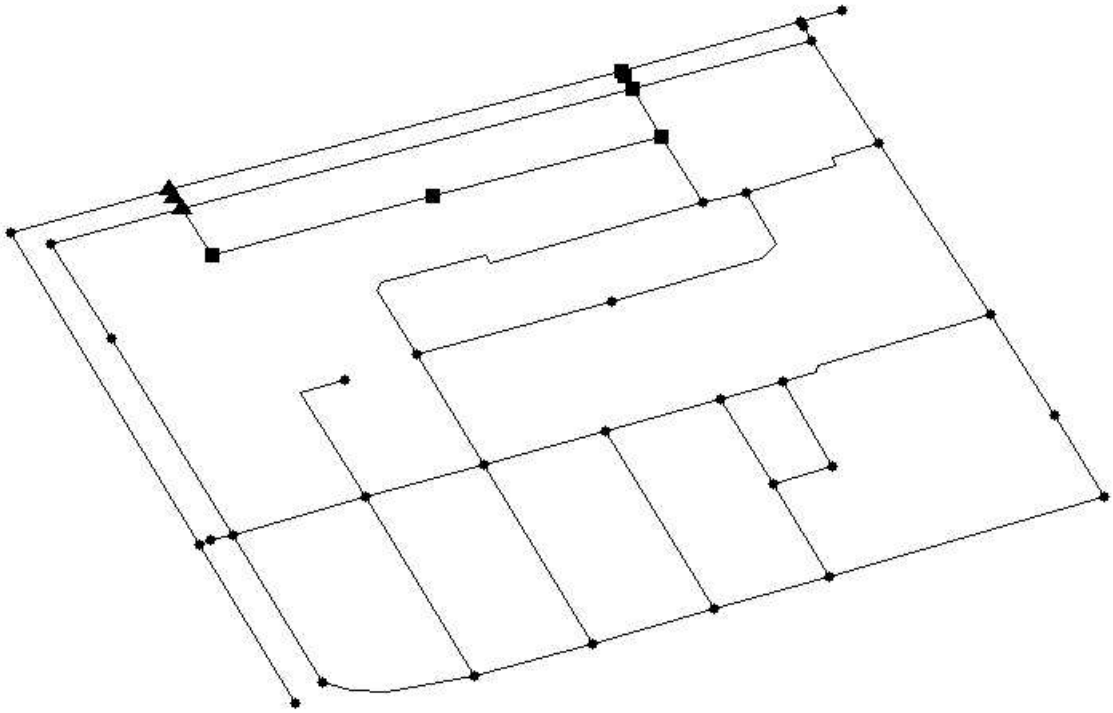

Figure 10.5 - Results obtained by the proposed method for the localization of contamination sources in DWDS.

Figure 10.5 presents the possible contamination sources verified at step 2 as squares and the possible contamination sources eliminated as triangles.

10.3.2 Case Study - Approach B

Following the Approach B, the time of contamination associated to each possible contamination sources is estimated using a single ANN trained with the information given by the entire set of sensors.

To verify if the Node 1 is a possible contamination source, sensor S1 is chosen as the reference sensor to find the position of the time of the first detection of that sensor within the training set. Table 10.3 shows the interval of contamination scenarios at Node 1 in which the time of the first detection at the sensor S1 is contained. This table presents the times of the first detection at each sensor, as well as the ratios between those times and the time of the first detection at the sensor S1, which was defined as reference sensor. It is observed that the ratios between the times of the first detection of the

contamination scenario under study are similar to the ratios of the training set, so it is concluded that the Node 1 is a possible contamination source. Using the ANN trained for the Node 1 with the vector $[7.67 \;\; 48 \;\; 4.86]'$ as input, the estimate for the time of contamination obtained is t = 3.48 h.

Table 10.3 - Training set position for Node 1 – Approach B

Time (h)	t_{S1} (h)	t_{S2} (h)	t_{S3} (h)	t_{S2} / t_{S1}	t_{S3} / t_{S1}
3.33	7.57	48.00	4.76	6.26	0.63
?	7.67	48.00	4.86	6.26	0.63
3.50	7.69	48.00	4.88	6.26	0.63

Still using the sensor S1 as reference sensor, Table 10.4 shows the interval of contamination scenarios at Node 2 in which the time of the first detection at the sensor S1 is contained. The ratio t_{S3} / t_{S1} associated to the contamination scenario in study is quite different from the same ratio associated to the same zone in the training set, so it is concluded that the Node 2 is not a possible contamination source.

Table 10.4 - Training set position for Node 2 – Approach B

Time (h)	t_{S1} (h)	t_{S2} (h)	t_{S3} (h)	t_{S2} / t_{S1}	t_{S3} / t_{S1}
5.17	7.57	48.00	6.54	6.26	0.86
?	7.67	48.00	4.86	6.26	0.63
5.33	7.73	48.00	6.66	6.26	0.86

By analysing the 9 possible contamination sources determined in the step 1, it is possible to verify that only 6 of them are possible contamination sources, as it already occurred when using the Approach A. The probability of each possible contamination source confirmed in the step 2 is updated to 0.167 and the estimates of the time of contamination associated to each possible contamination source are presented in Table 10.2 as well. By analysing the results presented in the Table 10.2, good agreements are observed between the estimates determined by each approach and the expected results for the time of contamination associated to each source candidate. Figure 10.5 presents the results obtained following this approach.

10.4 Chapter Conclusions

A method was developed to address the problem of the localization of contamination sources after deliberate contaminations in DWDSs through the application of ANNs. The work here presented emphasizes the advantages of the application of ANNs in the problem of the localization of contamination sources.

Two different approaches were developed to determine the probability of each node of the DWDS being the contamination source and to estimate the corresponding time of contamination for nodes with probabilities different from zero. Good agreements were observed between the estimates determined by each approach and the expected results for the time of contamination associated to each source candidate. The case study performed with a simplified DWDS shows that the proposed method is able to identify the correct contamination source and to predict the correct time of contamination associated to each possible contamination source.

One important feature of this method is that it is able to find the contamination sources without having any information about the contaminant concentration throughout the DWDS. This type of information is really difficult to get, especially in a deliberate contamination event, since the contaminant introduced in the DWDS might be unknown, making it difficult to assess its concentration. Following this method, the search for a contamination source can be triggered by any anomaly registered at a sensor. On the other hand, in spite of the computational effort necessary to create the databases and to train the ANNs, this procedure only has to be carried out once. The time of computation required for performing the proposed method following each one of the different approaches is minimal (just a few seconds), which is a great advantage in this problem that demands the computation of the results as quick as possible. Thus, this method can be applied in real situations to predict the location of possible contamination sources and its associated time of contamination by following both approaches and using the two set of results to validate each other.

11 Application of ANNs to the Problem of Localization of Contamination Sources: Strategies to Address Challenges Created by Large Drinking Water Distribution Systems

11.1 Chapter Overview

Chapter 10 discussed a method that tried to address some of the limitations identified in previous works concerning the subject of the localization of contamination sources in DWDSs, presented in Chapter 3. A limitation associated to some of these approaches is concerned with the time of computation, which might be very high, due to the complexity of the mathematical formulations. Another limitation is related with the initial assumption that the sensors should be able to evaluate the contaminant concentration, which might be difficult to achieve for real scenarios.

The method proposed by in Chapter 10 is able to find the contamination sources without having any information about the contaminant concentration throughout the DWDS. Following this method, the search for a contamination source can be triggered by any anomaly registered at a sensor. The time of computation required for performing this method is minimal (just a few seconds), which was presented as a great advantage in this problem that requires responses as quick as possible. The results obtained with this method were demonstrated with a simplified DWDS with very satisfactory results, without considering the existence of water demand uncertainties. However, the application of this method to real DWDS brings different challenges, mainly on two levels. One is concerned with the size of some DWDSs because the ANNs for DWDSs with a high number of junction nodes are too complex. The other is concerned with the hydraulic behaviour of the DWDS, which can take very different shapes throughout the day. For instance, the hydraulic conditions during the night period are significantly different from the hydraulic conditions during the day.

This chapter proposes strategies that address these two challenges. To deal with the high number of junction nodes that constitutes the real DWDS, the entire set of junction nodes is divided in clusters to decrease the complexity of the ANNs applied in this stage. Concerning the existence of very different hydraulic behaviours of the DWDS, the different scenarios that are used for the training of the ANNs are also divided in subsets, clustering scenarios with similar hydraulic behaviours for improving the accuracy of the ANNs. Furthermore, the effect of water demand uncertainties in the performance of the proposed method is also evaluated.

11.2 Methods Description

The proposed method aims to predict the localization of contaminations in DWDSs, based on the information given by a surveillance system constituted by a set of sensors, through the application of ANNs in the event of a single contamination scenario. The proposed method is a modified version of the method presented in Chapter 10, to enable the application of ANNs to large DWDS. Due to the high number of junction nodes in a real DWDS it is not reasonable to have an output for the probability that each one of them has of being the contamination source. As it was already referred above, the proposed strategies are based on cluster analysis (CA), which is a classification method that organizes a set of objects in clusters, grouping objects which show a high degree of similarity at the same cluster, while objects belonging to different clusters are as dissimilar as possible (Kaufman and Rousseeuw, 1990). The ideal number of clusters may be determined graphically through a dendrogram, a tree diagram commonly used in CA (Manly, 1994; McKenna, 2003).

The procedure proposed in this work is also divided in two main stages:

1. Determine the probability that each cluster has of including the contamination source based on the detection pattern. Evaluate the probability associated to each node considering the probability of the respective cluster and the number of junction nodes which belong to that cluster.
2. Verify each possible contamination source determined in Stage 1 and estimate the corresponding time based on the times of detection of the sensors. Update, if necessary, the probability of each node being the contamination source. For this stage, two different approaches are presented, being designated by Approaches A and B, respectively. These approaches can be used separately or together, to validate each other.

Both stages require the creation of databases with the information associated to a large number of contamination scenarios. These scenarios shall include contaminations at each node in the network, for several instants of contamination. The algorithm that was used to simulate those contamination scenarios is presented in Section 8.2.1.

11.2.1 Stage 1

The aim of this stage is to have an initial pruning of the set of nodes that are possible contamination sources. The entire set of junction nodes is divided in clusters. A matrix with the Euclidean distances between each pair of EPANET coordinates of junction nodes has to be created. Then, a hierarchical

clusters tree is created using the group average algorithm. A maximum number of clusters has to be defined and the clusters are constructed, based on the hierarchical clusters tree, using distance as a criterion. Through this criterion, the method seeks the smallest height at which a "horizontal cut" through the tree will leave the maximum number of clusters or fewer clusters.

The selected output for the ANN is the probability that each cluster has of including the contamination source, setting the number of clusters of the DWDS as the number of outputs of the ANN. The input variable selected is the binary status for each sensor (1 if the sensor detected the contamination, 0 otherwise). The information provided as an input is a vector with the size of the number of sensors, in which each element presents the status of the corresponding sensor. Meanwhile, the target is the identification of the cluster that include the node where the contamination occurred, defined as a vector with the number of elements equal to the number of clusters in which the DWDS was divided. For each contamination scenario, all elements of the target vector are equal to 0 excluding the element corresponding to the cluster that includes the node where the contamination was performed.

After evaluating the probability that each cluster has of including the contamination source, the probabilities are divided by the number of junction nodes that belongs to the respective cluster to determine the probability that each node has of being the contamination source.

11.2.2 Stage 2

The objective of this stage is to estimate the time of contamination associated to each node belonging to the clusters that were indicated in the Stage 1 as the ones which have higher probability of including the simulated source, based on the times of detection of the sensors. Thus, a minimum threshold for this probability has to be set.

The development of the ANNs in this stage follows the same procedure as it was described in Section 10.2.2. However, to improve the accuracy of the estimation of the time of contamination, the set of contamination scenarios that constitutes the database created to train the ANNs was divided in clusters, following a two stages procedure described below, for the approaches A and B, respectively.

Figure 11.1 presents the schematic representation of the stages that constitute the procedure developed for the clustering of the set of scenarios.

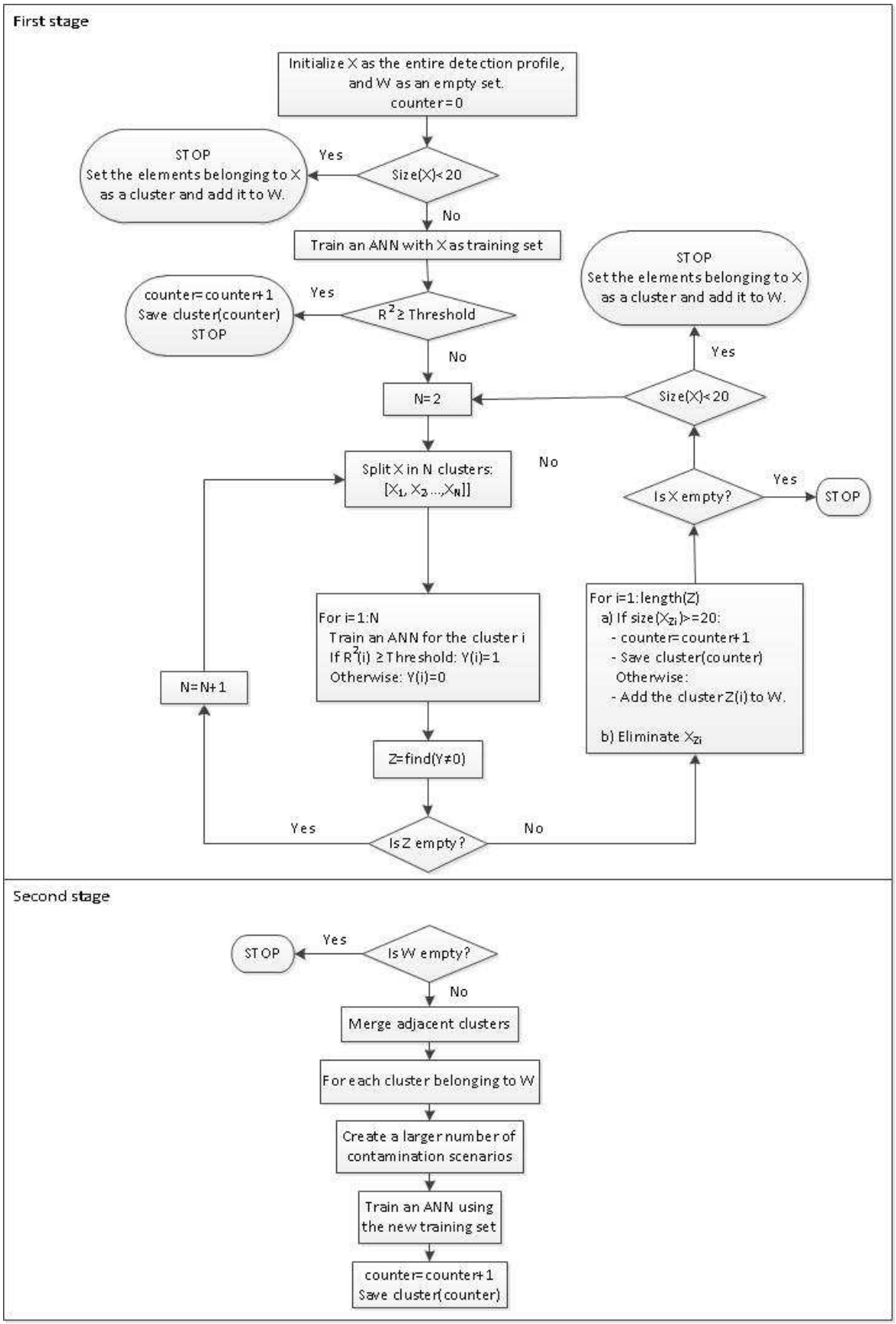

Figure 11.1 - Schematic representation of the stages that constitute the procedure developed for the clustering of the set of scenarios.

Approach A

For each node of the DWDS, the set of contamination scenarios was divided in a suitable number of clusters and an ANN was trained per sensor that shall detect contaminations occurring at the node under study, for each cluster of scenarios. The input of each ANN is the time of the first detection at the corresponding sensor, so the input layer is constituted by 1 neuron. The output is the time of contamination associated to the node under study.

The first stage of the proposed procedure is constituted by the following steps:

Step 1 – For each sensor that detects a contamination that occurred at a given node, initialize X as the entire detection profile (the time of contamination and its associated time of detection). Initialize a counter of the number of clusters as 0.

Step 2 – Verify the number of elements that constitutes the matrix X. If it has less than 20 elements, stop the algorithm and the elements belonging to X are set as cluster, which is added to W. Otherwise, continue to Step3.

Step 3 –Train an ANN using X as the training set. The time of detection at the sensor is the input and the time of contamination at the source is the output.

Step 4 – Verify if the correlation coefficient is higher than a minimum threshold that it is considered as sufficient for each ANN. If it is true, update the counter of the number of clusters (counter=counter+1) and associate the ANN to the corresponding cluster. Otherwise update the number of clusters, *N*, to 2.

Step 5 – Split X in N clusters: $[X_1, X_2,\ldots,X_N]$. A matrix with the Euclidean distances between each pair of time of contamination/time of detection is created. Then, a hierarchical clusters tree is created using the minimum variance algorithm. The clusters are constructed, based on the hierarchical clusters tree, using distance as a criterion, considering a maximum number of clusters N.

Step 6 – For each cluster, train an ANN and verify if the correlation coefficient (R^2) is higher than the minimum threshold. If it is true, set Y(i) =1. Otherwise, Y(i) =0.

Step 7 – Set Z as the set of elements of Y different from 0. If Z is empty, update the number of clusters as *N=N+1* and return to Step 5. Otherwise, for each cluster belonging to Z:

Step 7.1 – If size($X_{Z(i)}$)>20, update the counter of the number of clusters (counter=counter+1) and associate the ANN to the corresponding cluster. Otherwise, the cluster Z(i) is added to W.

Step 7.2 – Eliminate $X_{Z(i)}$.

Step 8 – Verify if X is empty. It that is true, stop the algorithm. Otherwise, proceed to Step 9.

Step 9 - Verify the size of *X*. If it has less than 20 contamination scenarios, stop the algorithm and the elements belonging to X are set as cluster, which is added to W. Otherwise, set *N* as 2 and return to Step 5.

The clusters that constitutes W, for each sensor that detects the node in analysis, will be the subject of the second stage of the procedure.

Step 1 – Merge adjacent clusters.

For each cluster belonging to W:

Step 2 – Simulate a larger number of contamination scenarios, considering scenarios with the time of contamination belonging to the interval defined for the cluster. Set the results of these simulations as the database used for the training of the ANN.

Step 3 – Train an ANN using the new set of contamination scenarios.

Step 4 – Update the counter of the number of clusters (counter=counter+1) and associate the ANN to the corresponding cluster.

After applying this procedure to the entire set of junction nodes, there is a developed ANN for each cluster of contamination scenarios, for each sensor that detects a contamination that occurred at a given junction node. With this set of ANNs, it is possible to perform several estimates for the time of contamination associated to a source candidate based on the time of the first detection of each sensor that detected the contamination.

For each detection, it is necessary to verify in which cluster of contamination scenarios the detection is included. After these verifications, an estimation of the different times of contamination associated to the node considered as a possible contamination source is performed, enabling to verify if that node remains a candidate. If the estimates obtained using the different ANNs are similar, the node is confirmed as a possible contamination source and the average between the estimates obtained by each ANN is chosen as the time of contamination associated to the node. If there is a discrepancy between the estimates obtained by the different ANNs, it is concluded that the node under study is not a possible contamination source. This discrepancy is due to the fact that the set of values considered for the time of first detection at the sensors does not correspond to a contamination on the node in study.

After verifying which nodes remain as possible contamination sources, the corresponding values of probability calculated in Stage 1 are updated, ensuring that the sum of the probability of the remaining possible contamination sources associated to each cluster remains equal to the value calculated at the Stage 1.

Approach B

For each node of the DWDS, the set of contamination scenarios was divided in a suitable number of clusters, as it happens in Approach A, and an ANN was trained for the junction node under study, for each cluster of scenarios. The input of each ANN is the time of first detection at each sensor as input. When a contamination is not detected by a sensor, the time of detection at this sensor is set as *2x Thorizon* (*Thorizon* is the analysis interval). As in the Approach A, the output is the time of contamination associated to the node in study.

The procedure implemented for the division of the contamination scenarios in clusters is also presented by the Figure 11.1. The procedure is similar to the one that is presented for the Approach A. However, there are differences in the Steps 1 and 5 of the First Stage. In the Step1, X is initialized as the detection profile of the entire set of sensors (the time of contamination and the associated time of detections at each sensor). In Step 5, the division of the contamination scenarios in clusters is preceded by principal components analysis to create a new variable able to explain the largest fraction of the original data variability (Sousa et al., 2007; Pires et al., 2008). Then, a matrix with the Euclidean distances between each pair of time of contamination/"new variable" is created, instead of considering pairs of time of contamination/time of detection.

This approach also enables to verify if the node under study remains as a possible contamination source. It is necessary to verify in which cluster of contamination scenarios the detection pattern is included and then the position of the time of the first detection of one sensor within the training set used to train the respective ANN. Then, it is evaluated if the ratios between the times of the first detection given by the sensors are close to the ratios between the times of the first detection in the correspondent position in the training set. If this condition is verified, the node is a possible contamination source and the ANN associated to this node is used to estimate the corresponding time of contamination. Otherwise, the set of values considered for the time of first detection at the sensors are not corresponding to a contamination on the node under study, so it is no longer considered as a possible contamination source.

After verifying which nodes remain as possible contamination sources, the corresponding values of probability calculated in Stage 1 are updated, following the same procedure described for the Approach A.

11.3 Example of application

The Network D, presented in Section 5.4, was used as case study. A set of 60 sensors, presented in Figure 11.2 as black circles, were defined to test the performance of the proposed method.

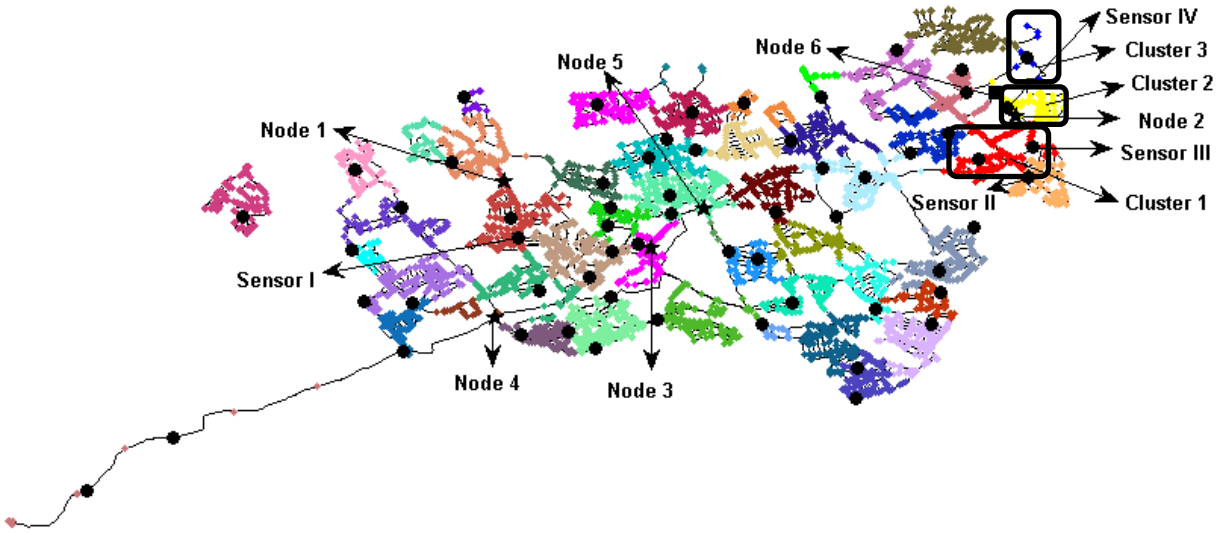

Figure 11.2 – Division of the Network D in clusters.

The junction nodes, reservoirs and tanks that constitute the DWDS were divided in clusters. A trial and error procedure took place to verify which number of clusters would be reasonable to ensure that the computational effort associated to the development of the ANN required to perform the Stage 1 of the procedure is manageable and, at the same time, that the number of clusters is not so low as to render Stage 1 worthless. It was verified that a division in 50 clusters would meet both criteria, so the entire set of junction nodes is divided in clusters, following the procedure presented in Section 11.2.1. The clusters of nodes are presented in Figure 11.2 in different colours.

Several ANNs were trained following the procedures described in Sections 11.2.1and 11.2.2 to perform the tasks defined in each stage of the method, considering the different approaches. The databases used in the training of these ANNs were constituted by contamination scenarios occurring at each node at each 5 min.

A demonstration of the procedures developed for the division of the contamination scenarios in clusters and training of the respective ANNs is presented in Sections 11.3.1 and 11.3.2, for approaches A and B, respectively.

11.3.1 Demonstration – Approach A

The development of ANNs for the detections at the sensor I associated to contaminations occurring at the Node 1, presented in the Figure 11.2 in star shape, is considered for demonstrating the procedure developed for the division of the contamination scenarios in clusters, for the Approach A. The threshold value for the correlation coefficient was set as 0.995.

Analysing the set of contamination scenarios, *X*, it was observed that, during the analysis interval, the sensor I detects several contaminations that occur at the Node 1 between t=2:55 h and t=41:55 h, with some discontinuities. The counter of the stored clusters was initialized as 0. It was verified that *X* was not empty, and its size was higher than 20.

An ANN was trained considering the entire detection profile, as it was described in the step 3 of the first stage, reaching a R^2 of 0.991. Figure 11.3 presents the comparison between the target data (presented as circles) and the ANN results (presented as a solid line).

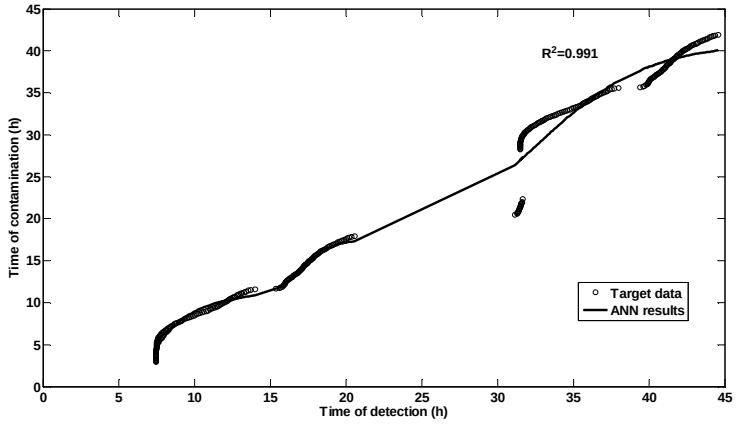

Figure 11.3 - Results obtained for the ANN associated to the pair Node 1/Sensor I – 1 cluster.

As the R^2 was lower than the defined threshold, the number of clusters was updated to 2 and the *X* was divided in two clusters (Figure 11.4), as it was described in the step 5. One ANN was trained for each cluster and the comparison between the target data and the ANN results is presented with a presentation similar to Figure 11.3. The R^2 of the ANN trained for the cluster A was higher than the

threshold while the R^2 of the ANN trained for the cluster B was lower. There was one cluster with an acceptable R^2, so the algorithm proceeded to Step 7.1. It was verified that the cluster 1 had more than 20 contamination scenarios, so the counter of stored clusters was set as 1, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

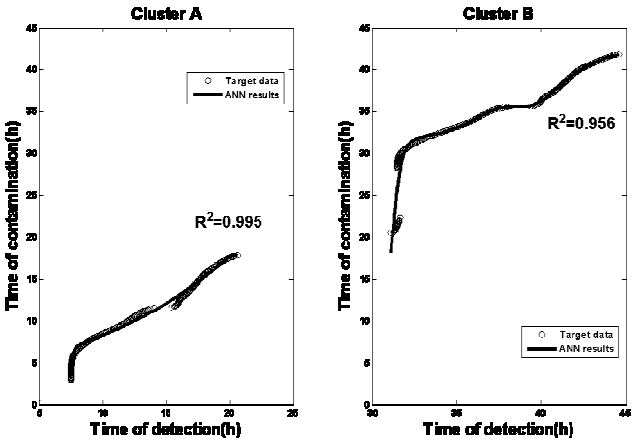

Figure 11.4 - Results obtained for the ANN associated to the pair Node 1/Sensor I – 2 clusters.

It was verified that X was still not empty and it had more than 20 elements, so *N* remained as 2 and the algorithm returned for the Step 5. The elements of *X* were divided in 2 clusters (Figure 11.5). One ANN was trained for each cluster and it was verified that the R^2 of the ANN trained for the cluster B was higher than the threshold and the cluster had more than 20 contamination scenarios. The counter of stored clusters was updated to 2, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

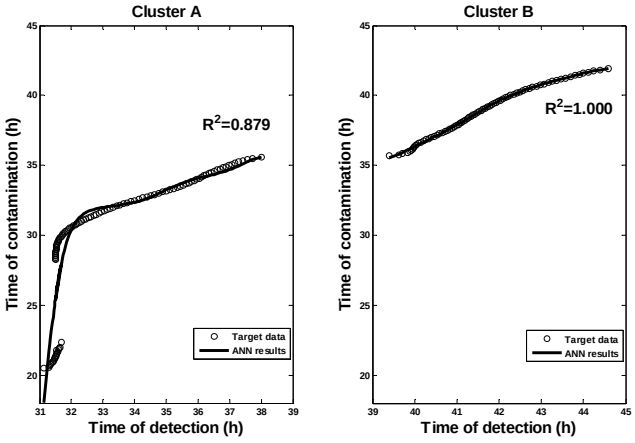

Figure 11.5 - Results obtained for the ANN associated to the pair Node 1/Sensor I – 3 clusters.

As *X* was still not empty and had more than 20 elements, *N* remained again as 2 and the algorithm returned for the Step 5. The elements of *X* were divided in 2 clusters (Figure 11.6). One ANN was trained for each cluster and it was verified that the R^2 of the ANN trained for the cluster B was higher than the threshold and the cluster had more than 20 contamination scenarios. The counter of stored clusters was updated to 3, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

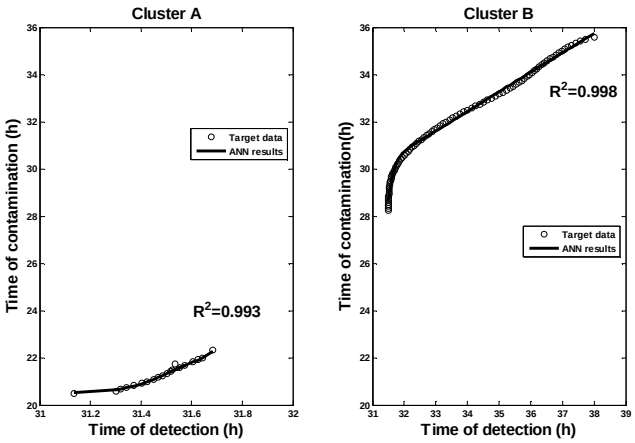

Figure 11.6 - Results obtained for the ANN associated to the pair Node 1/Sensor I – 4 clusters.

X was still not empty, but as it had only 20 elements, dividing it in two clusters would not create any cluster that meet the criteria to be stored. Thus, the cluster was added to *W* and the algorithm proceeded to the second stage.

New contamination scenarios were created for the cluster that belongs to *W*, with a time step of 30 seconds between t=20:30 h and t=22:20 h. An ANN was trained with the new training set (Figure 11.7). The R^2 of the ANN was 0.995, the counter of stored clusters was updated to 4 and the ANN associated to the cluster was stored. The procedure for the development of ANNs for the Node 1 was considered as complete. Figure 11.8 presents the results of the division of the contamination scenarios associated to Node 1 in clusters, for Approach A.

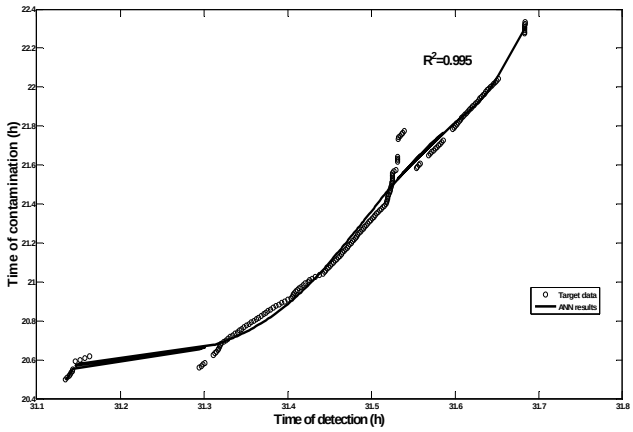

Figure 11.7 - Results obtained after the application of the second stage to cluster 4.

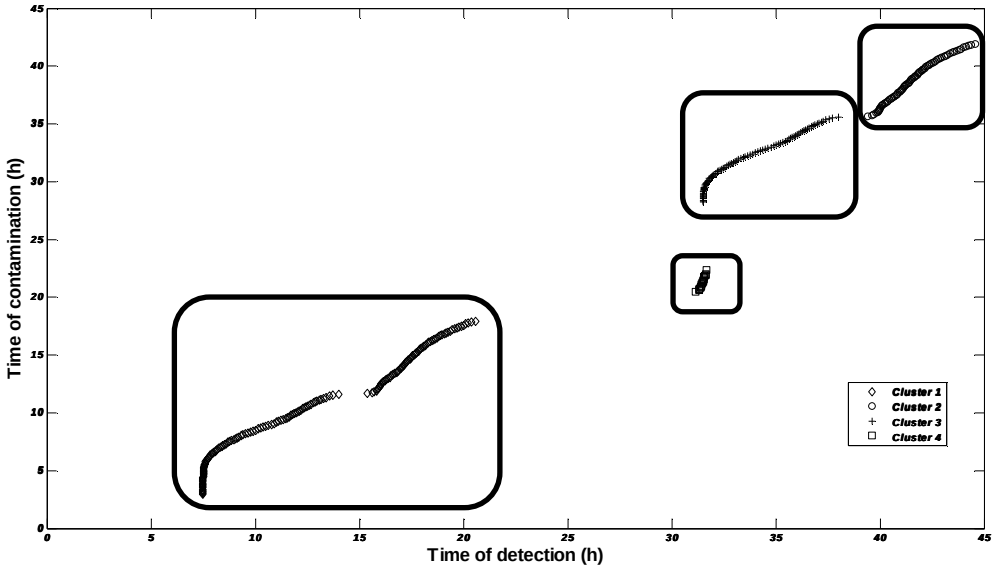

Figure 11.8 - Division of the training set of the pair Node 1/Sensor I in clusters.

11.3.2 Demonstration – Approach B

The development of ANNs, associated to contaminations occurring at the Node 2, presented in the Figure 11.2 in star shape, is considered for demonstrating the procedure developed for the division of the contamination scenarios in clusters, for the Approach B. Contaminations at this node might be detected by sensors II and III. The threshold value for the correlation coefficient was set as 0.995.

The contamination scenarios simulated that were detected by the system of sensors, X, were set as the training set. The counter of the stored clusters was initialized as 0. It was verified that *X* was not empty, and its size was higher than 20.

An ANN was trained considering the entire detection profile, reaching a R^2 of 0.993. Figure 11.9 presents the comparison between the target data (presented as circles) and the ANN results (presented as plus marks). Principal components analysis was used to create a new variable able to explain over 80% of the original data variability.

The number of clusters was updated to 2 and the X was divided in two clusters (Figure 11.10), as it was described in the step 5, considering the Euclidean distances between each pair of time of contamination/"new variable". One ANN was trained for each cluster and it was verified that the R^2 of the ANN trained for the cluster A was higher than the threshold while the R^2 of the ANN trained for the cluster B was lower. There was one cluster with an acceptable R^2, so the algorithm proceeded to Step 7.1. It was verified that the cluster A had more than 20 contamination scenarios, so the counter of stored clusters was set as 1, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

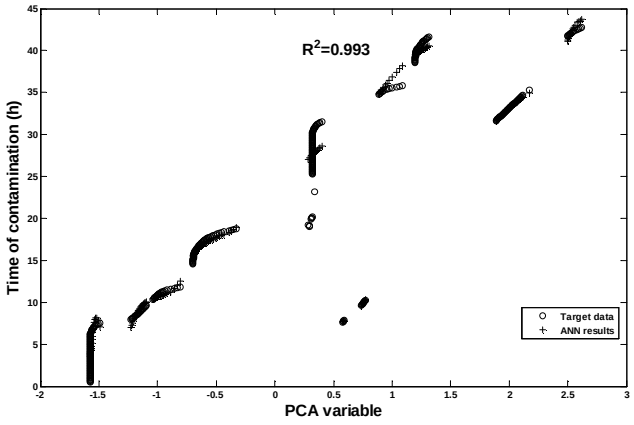

Figure 11.9 - Results obtained for the ANN associated to the Node 2 – 1 cluster.

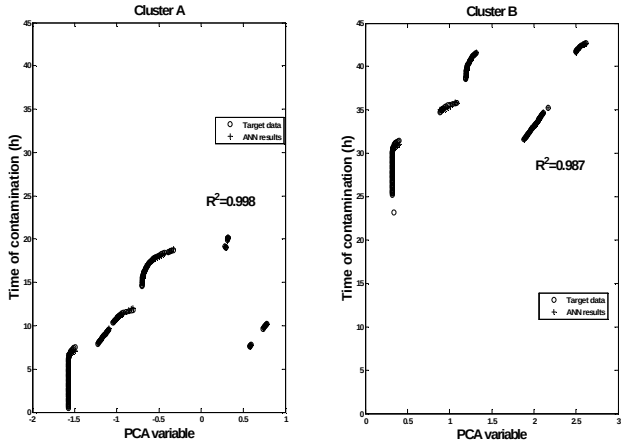

Figure 11.10 - Results obtained for the ANN associated to the Node 2 – 2 clusters.

It was verified that X was still not empty and it had more than 20 elements, so *N* remained as 2 and the algorithm returned for the Step 5. The remaining elements of *X* were divided in 2 clusters (Figure 11.11) and one ANN was trained for each cluster. The R^2 of the ANN trained for the cluster B was higher than the threshold and the cluster had more than 20 contamination scenarios, so the counter of stored clusters was updated to 2, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

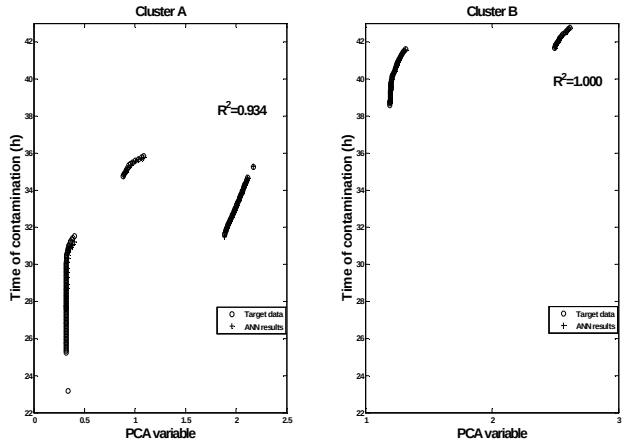

Figure 11.11 – Results obtained for the ANN associated to the Node 2 – 3 clusters.

As *X* was still not empty and had more than 20 elements, *N* remained again as 2 and the algorithm returned for the Step 5. The remaining elements of *X* were divided in 2 clusters (Figure 11.12) and one ANN was trained for each cluster. The R^2 of the ANN trained for the cluster B was higher than the threshold and the cluster had more than 20 contamination scenarios, so the counter of stored clusters was updated to 3, the ANN associated to the cluster was stored and the contamination scenarios that belong to the cluster were eliminated from *X*.

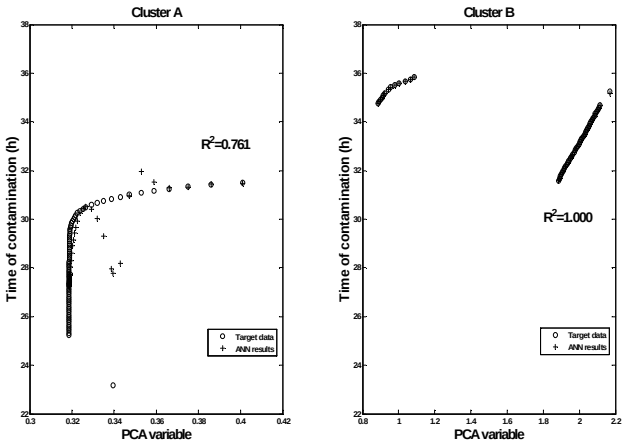

Figure 11.12 - Results obtained for the ANN associated to the Node 2 – 4 clusters.

As *X* was still not empty and had more than 20 elements, *N* remained again as 2 and the algorithm returned for the Step 5. The elements of *X* were divided in 2 clusters (Figure 11.13) and one ANN was trained for each cluster. Both ANNs had a R^2 lower than the threshold, so *N* was updated to 3 and the algorithm returned for the Step 5. The elements of *X* were divided in 3 clusters (Figure 11.14) and one ANN was trained for each cluster. Clusters A and B had a R^2 higher than the threshold, so the algorithm proceeds to Step 7.1. Clusters A and B had less than 20 elements (19 each cluster), so these clusters were added to *W*. The contamination scenarios that belong to these clusters were eliminated from *X*.

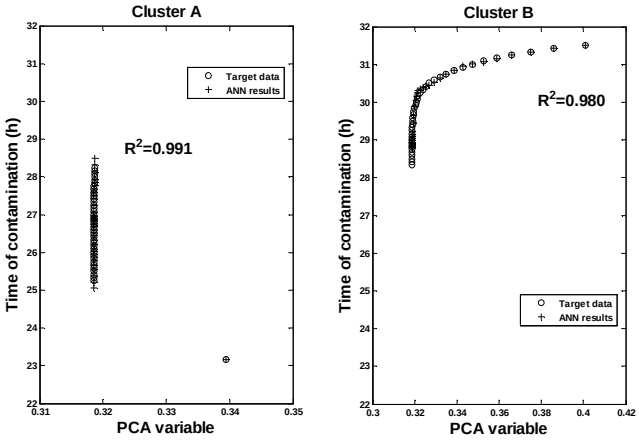

Figure 11.13 - Results obtained for the ANN associated to the Node 2 – 5 clusters.

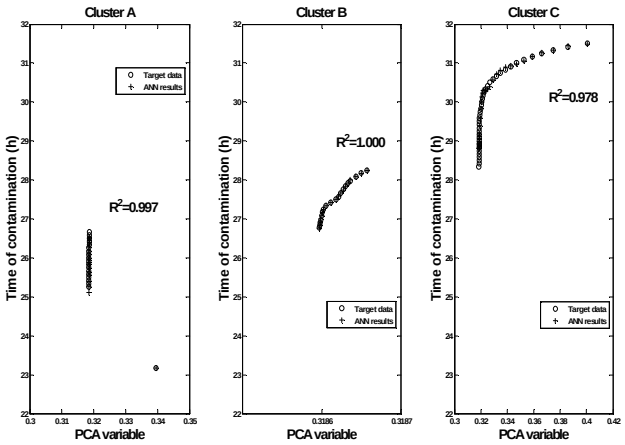

Figure 11.14 – Results obtained for the ANN associated to the Node 2 – 6 clusters.

As *X* was still not empty and had more than 20 elements, *N* was set again as 2 and the algorithm returned to the Step 5. The remaining elements of *X* were divided in 2 clusters (Figure 11.15) and one ANN was trained for each cluster. Both ANNs had a R^2 higher than the threshold, so the algorithm proceeds to Step 7.1. The cluster B had less than 20 elements (13), so this cluster was added to *W*. The cluster A had more than 20 elements, so the counter of stored clusters was updated to 4 and the ANN associated to the cluster was stored. The contamination scenarios that belong to these clusters were eliminated from *X*. As *X* became empty, the first stage was completed and the algorithm proceeds to the second stage.

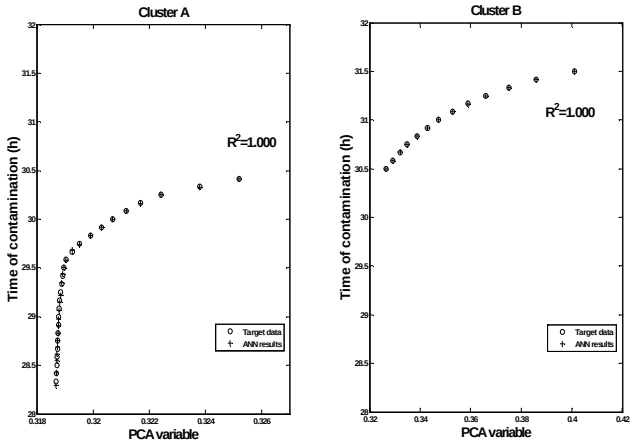

Figure 11.15 - Results obtained for the ANN associated to the Node 2 – 7 clusters.

New contamination scenarios were created for both clusters that belongs to *W* for the node in analysis, with a time step of 30 seconds between t=23:10 h and t=26:40 h, t=26:45 h and t=28:15 h, and t=30:30 h and t=31:30 h. An ANN was trained for each cluster with the new training sets (Figure 11.16). The R^2 of the ANNs were 0.999, 1.000 and 1.000 for clusters A, B and C, respectively. The counter of

stored clusters was updated for each new cluster and the ANN associated to each cluster was stored. The procedure for the development of ANNs for the Node 2 was considered as complete. Figures 11.17 and 11.18 present the results of the division of the contamination scenarios associated to Node 1 in clusters, for Approach B.

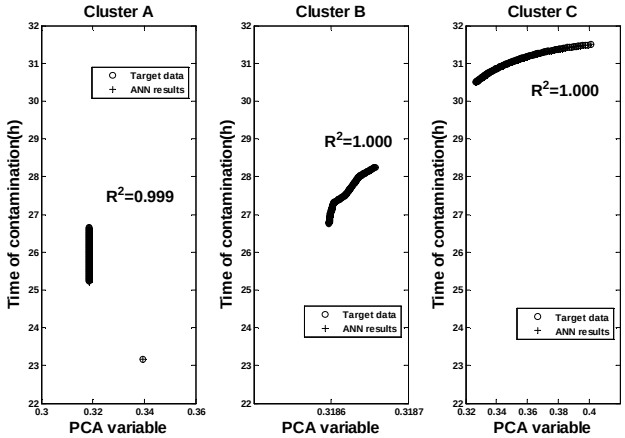

Figure 11.16 - Results obtained after the application of the second stage to clusters 5, 6 and 7.

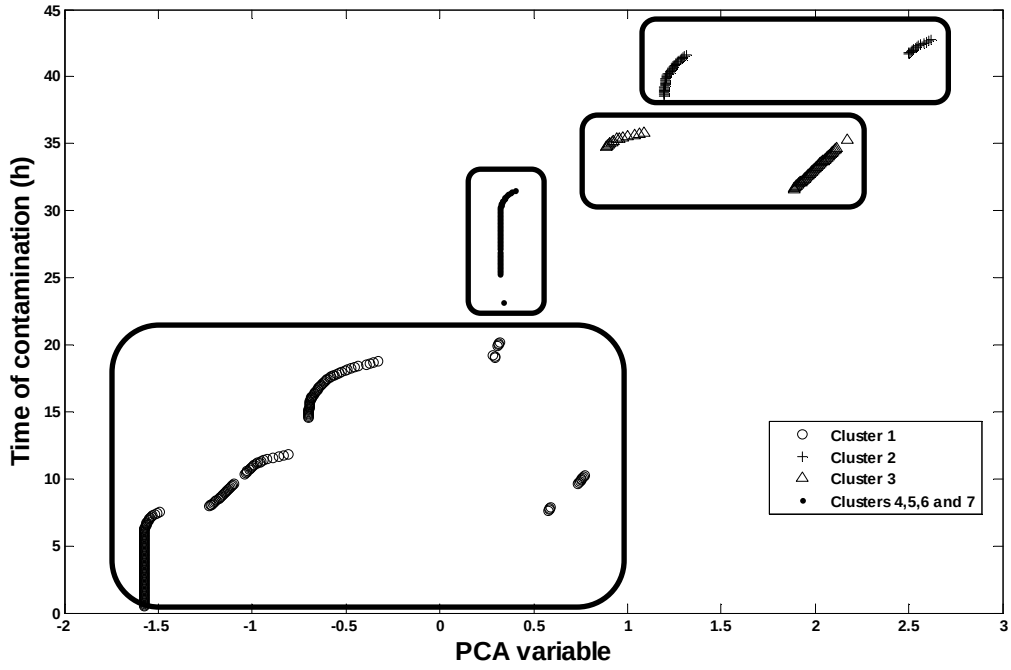

Figure 11.17 - Division of the training set of the Node 2 in clusters.

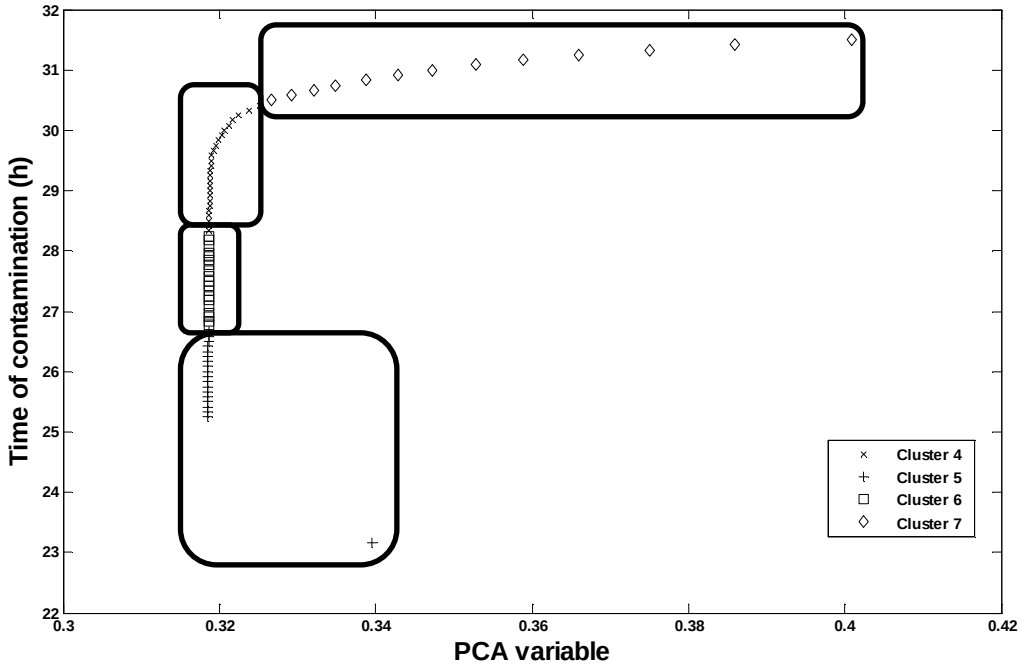

Figure 11.18 - Clusters 4, 5, 6 and 7 of the training set of the Node 2.

11.4 Results and Discussion

The ANNs necessary to perform the proposed method were developed for the ideal hydraulic model of the DWDS presented in the Section 5.4. For the Stage 1, the hidden layer was constituted by 150 neurons and a logarithmic sigmoid and a linear transfer function were selected for the hidden and output neurons, respectively. For both different approaches developed for Stage 2, due to the high amount of junction nodes, the trial and error method was not applied to all junction nodes. A logarithmic sigmoid and a linear transfer function were selected as the transfer functions for hidden and output neurons for every node, for both approaches. The hidden layer was constituted by 3 neurons for approach A. For approach B, the heuristic estimate given by the Equation 3 in the work presented by Gao et al. (2010), which relates the number of neurons in the hidden layer with the number of input and output neurons were applied (Zhang et al., 1997).

A contamination occurring at Node 2 (Figure 11.2), at t=15:00 h was simulated, considering the ideal hydraulic scenario. Using the algorithm presented in Section 8.2.1, the contamination plume was determined and it was verified that Sensor II and Sensor III detect a contamination at t=22:03 h and t=18:21 h, respectively. The remaining sensors didn't detect any contamination.

Beginning with the Stage 1, the input to the ANN_1 was a vector constituted by 60 elements, which were all set as 0 except the elements corresponding to Sensor II and Sensor II, which were set as 1. The output of the ANN_1 is a vector with 50 elements, each one corresponding to one cluster. The probability that each cluster has of including the contamination source was divided by the number of junction nodes that belongs to the respective cluster. Table 11.1 presents the results for the clusters which have a probability of including the contamination source higher than 1%, which was the minimum threshold that had to be defined. The remaining clusters have a probability very close to 0, so were considered negligible. These 3 clusters (Figure 11.2) constitute 96.6% of the distribution of probabilities performed by the ANN. The Stage 1 enabled reducing the number of possible contamination sources from 8857 (the entire DWDS) to 447. It was observed that the simulated contamination source (Node 2) does not belong to the cluster with the higher probability.

Table 11.1 – Results after Stage 1

	Probability (cluster)	**Number of nodes**	**Probability (each node)**
Cluster 1	72.6%	275	0.26%
Cluster 2	21.9%	156	0.14%
Cluster 3	2.4%	16	0.15%

In the Stage 2, each possible contamination sources determined in the Stage 1 has to be verified. In Sections 11.4.1 and 11.4.2, Nodes 2 and 6 (both nodes belong to Cluster 2) will be presented as examples of possible contamination sources confirmed and eliminated by Stage 2, following the different approaches developed for this stage.

11.4.1 Approach A

Following the Approach A, the time of contamination associated to each possible contamination sources is estimated using the ANNs trained with the information related with each sensor that detects contamination at that node. To evaluate if the estimates obtained by the ANNs associated to different sensors are similar it was defined that the maximum value of deviation between each estimate and the average of all estimates was 1 hour. If any of the estimates presented a higher value of deviation in relation with the average, the corresponding possible contamination source would be considered eliminated. Table 11.2 presents the results obtained for Nodes 2 and 6, following Approach A.

Table 11.2 – Results achieved following Approach A for Nodes 2 and 6.

	Node 2			**Node 6**		
	Estimate (h)	**Average (h)**	**Deviation (h)**	**Estimate (h)**	**Average (h)**	**Deviation (h)**
Sensor II	14:52	15:1 9	0:27	12:48	13:32	0:44
Sensor III	15:46		0:27	14:16		0:44
Sensor IV				-	-	-

Beginning with Node 2, it is verified that contaminations occurring at this node might be detected at Sensor II and Sensor III. Analysing the detection at the Sensor II, firstly it was necessary to verify in which cluster the detection at the sensor was included. The division of the set of contamination scenarios was not performed in the development of the ANNs, following Approach A, since the ANN trained in the Step 3 of the procedure had a R^2 higher than 0.995, the minimum threshold defined for this parameter. Thus, there was just one cluster for this pair node/sensor. It was verified that the detection at t=22:03 h was covered by the training set, so the time of contamination given by the respective ANN was evaluated. The estimate obtained was t=14:52 h. Repeating the same procedure for the analysis of the detection at the Sensor III, it was verified that there was just one cluster for this pair node/sensor, by the same reasons as for the pair Node 2/Sensor II. It was verified that the detection at t=18:21 h was covered by the training set, and the estimate obtained for the time of contamination was t=15:47 h. The average time of contamination was t=15:19 h. The deviation between the estimates and the average was 27 min, lower than 1 hour, so Node 2 remained as a possible contamination source.

Node 6 might be detected by Sensors II, III and IV, so an evaluation of each sensor must be performed. Analysing the detection at the Sensor II, it was verified that there was just one cluster for the pair Node 6/Sensor II. The detection at t=22:03 h was covered by the training set and the estimate obtained for the time of contamination was t=12:48 h. There was also just one cluster for the pair Node 6/Sensor III, and the corresponding training set covers the detection occurring at t=18:21 h. The estimate obtained for the time of contamination was t=14:16 h. Sensor IV did not detect any contamination, so it was necessary to verify if the contamination source responsible for the detections at Sensors II and III was supposed to be detected at Sensor IV. T the training set of the pair Node 6/Sensor IV was analysed to verify if it contained a contamination occurring in the vicinity of t=13:32 h (the average between the estimates associated to Sensors II and III). As it was confirmed, the Node 6

was eliminated as a possible contamination source, since a contamination at this Node that was responsible for the detections at Sensors II and II, would trigger a detection at Sensor IV as well.

By analysing the 447 possible contamination sources determined in Stage 1, it is possible to verify that only 45 of them are possible contamination sources. The final results for the localization of contamination sources, following this approach are presented in Figure 11.19. The set possible contamination sources obtained following the Approach A are presented by the groups of red and pink dots, the sensors that detect the contamination as yellow circles and the sensors that don't detect any contamination as green circles.

The probability of each possible contamination source was updated. For instance, after Stage 1, Node 2 had a probability of 0.14%; after the Stage 2, this value was updated to 1.46%.

The method, following this approach, required a time of computation of 5.0 s in a 3.10 GHz processor.

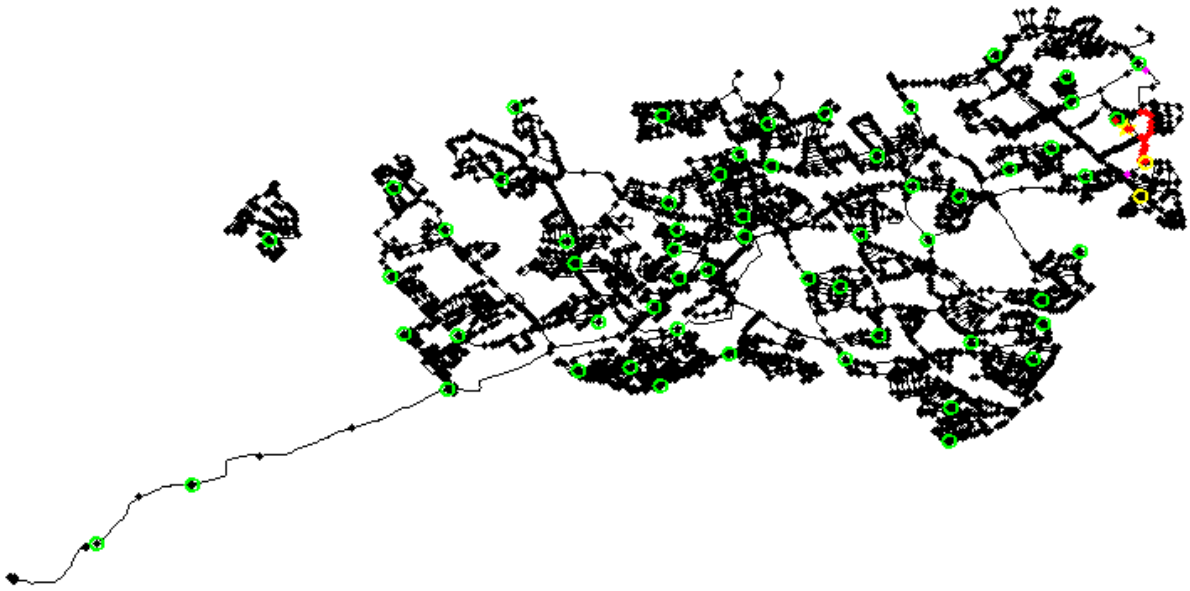

Figure 11.19 - Figures obtained for the case study following approaches A and B.

11.4.2 Approach B

Following the Approach B, the time of contamination associated to each possible contamination sources is estimated using a single ANN trained with the information given by the entire set of sensors. To evaluate if the node was a possible contamination source, the position of the time of the first detection of one sensor within the training set used to train the ANN was verified and the ratios

between the times of the first detection given by the sensors were compared with the ratios between the times of the first detection in the correspondent position in the training set. A maximum value of deviation of 5% was defined to evaluate if the node in analysis was a possible contamination source. If there was a higher deviation between the rations, the corresponding possible contamination source would be considered eliminated.

To verify if the Node 2 is a possible contamination source, Sensor II was defined as the reference sensor to find the position of the time of the first detection of that sensor within the training set. Table 11.2 shows the interval of contamination scenarios at Node 2 in which the time of the first detection at the Sensor II was contained. This table presents the times of the first detection at each sensor, as well as the ratios between those times and the time of the first detection at Sensor II, which was defined as reference sensor. Firstly it was necessary to verify in which cluster the detection profile was included, since the set of contamination scenarios was divided in seven clusters as it was shown previously (Figure 11.18). It was verified that the detection profile was contained in cluster 1 of the training set, thus the corresponding ANN was used to determine if Node 2 remained as a possible contamination source and to estimate the associated time of contamination. It was observed that the ratios between the times of the first detection of the contamination scenario under study did not show almost any deviation in relation to the ratios of the training set (Table 11.3), so it was concluded that the Node 2 remained as a possible contamination source. Using the ANN trained for the Node 2 with the vector $[22{:}03 \;\; 18{:}21]'$ as input, the estimate for the time of contamination obtained was t=15:04 h.

Table 11.3 – Training set position for Node 2 – Approach B.

Time (h)	**t_{II} (h)**	**t_{III} (h)**	**t_{III} / t_{II}**
15:00	22:02	18:20	0.83
?	22:02	18:20	0.83
15:05	22:03	18:21	0.83

Still using Sensor II as the reference sensor, Table 11.4 shows the interval of contamination scenarios at Node 6 in which the time of the first detection at the Sensor II was contained. It was verified that the training set was divided in just one cluster for Node 6, thus the corresponding ANN was used. The ratio t_{IV} / t_{II} associated to the contamination scenario in study is quite different (deviation much larger than 5%) from the same ratio associated to the same zone in the training set, since Sensor IV did not detect any contamination, so it was concluded that the Node 6 was not a possible contamination source.

Table 11.4 - Training set position for Node 6 – Approach B

Time (h)	t_{II} (h)	t_{III} (h)	t_{IV} (h)	t_{III} / t_{II}	t_{IV} / t_{II}
15:00	22:02	18:20	13:48	0.83	0.63
?	22:02	18:20	96:00	0.83	4.36
15:05	22:03	18:21	13:54	0.83	0.63

By analysing the 447 possible contamination sources determined in Stage 1, it is possible to verify that only 43 of them are possible contamination sources. The final results for the localization of contamination sources, following this approach are presented as red dots in Figure 12. All the possible contamination sources obtained following the Approach B were contained in the set of possible contaminations sources obtained following the Approach A.

The probability of each possible contamination source was updated. After Stage 1, Node 2 had a probability of 0.14% and, after the Stage 2, this value was updated to 1.46%.

The method, following this approach, required a time of computation of 1.7 s in a 3.10 GHz processor.

Comparing the results obtained by the different approaches, it is verified that both have defined very similar sets of possible contamination sources. There are only 2 possibilities that are included in the results obtained by Approach A that are not included in the results obtained by the Approach B. In fact, those 2 nodes should not be considered as possible contamination sources but obeyed to the criterion defined for the Approach A. The criterion defined for the Approach B was able to eliminate those possibilities.

The deviation between the simulated time of computation and the estimated results was 19 minutes for the Approach A and 4 minutes for the Approach B. Approach B required a lower time of computation.

11.4.3 Case Study – Effects of Water Demand Uncertainty

The robustness of both approaches against water demand uncertainty was evaluated. Single contamination scenarios were tested with contaminations occurring at Nodes 1, 2, 3, 4 and 5, at 6 h. These contamination scenarios were evaluated considering two different types of hydraulic behaviour: a) an ideal situation, with no model error and no parameter error; b) 100 different hydraulic scenarios, with no model error but with water demand uncertainty. The ANNs necessary to apply the proposed

approaches were developed using the ideal hydraulic parameters. The hydraulic behaviours with demand uncertainty were generated at the Bordeaux Centre of Irstea (Institut national de recherche en sciences et technologies pour l'environment et l'agriculture), through the Monte Carlo method that was used to generate several plausible demand time patterns considering a level of uncertainty of 10%.

The tests were performed in a 3.10 GHz processor.

The performance of the method was evaluated, considering 4 indicators:

1. Target the simulated source;
2. Number of possible contamination sources obtained;
3. Difference between the obtained estimates and the simulated time of contamination;
4. Time of computation required.

Tables 11.5 and 11.6 present the results obtained following the Approach A, considering a maximum value of deviation between each estimate and the average of all estimates of 3 hours, and Tables 11.7 and 11.8 present the results obtained following the Approach B, considering a maximum deviation of 50% in the ratios between the times of the first detection given by the sensors in the correspondent position in the training set. These maximum values, defined for each approach, are higher than the values used in Sections 11.4.1 and 11.4.2 to enable the evaluation of contamination scenarios under demand uncertainty. Tables 11.5 and 11.7 present the results obtained for the tests considering the ideal hydraulic scenario. Tables 11.6 and 11.8 present the average results obtained considering 100 hydraulic scenarios under demand uncertainty. Each column of the tables evaluates the respective approach in terms of each one of the parameters mentioned above.

Table 11.5 – Results obtained following the Approach A considering the ideal hydraulic scenario.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**	**Parameter 4 (s)**
Node 1	100	89	00:10	4.4
Node 2	100	97	00:45	7.1
Node 3	100	155	00:01	3.1
Node 4	100	27	00:04	18.7
Node 5	100	15	00:02	2.4
Average	100	77	00:12	7.1

Table 11.6 – Average results obtained following the Approach A considering 100 hydraulic scenarios under demand uncertainties.

	Parameter 1 (%)	Parameter 2	Parameter 3 (h:min)	Parameter 4 (s)
Node 1	100	96	00:29	3.9
Node 2	100	97	02:23	3.3
Node 3	86	152	00:07	2.7
Node 4	100	27	00:04	18.6
Node 5	100	15	00:02	2.2
Average	97.2	75	00:38	6.2

Table 11.7 – Results obtained following the Approach B considering the ideal hydraulic scenario.

	Parameter 1 (%)	Parameter 2	Parameter 3 (h:min)	Parameter 4 (s)
Node 1	100	58	0:06	1.6
Node 2	100	47	0:04	0.9
Node 3	100	105	0:02	2.0
Node 4	100	4	0:07	0.9
Node 5	100	7	0:01	0.7
Average	100	44	00:04	1.2

Comparing the results obtained by both approaches, it is verified that even though both approaches were able to solve the problem of the localization of contamination sources very quickly, the Approach B always required a lower time of computation.

For ideal hydraulic scenarios, both approaches detected all the tested contamination scenarios. The approach B achieved generally better results in terms of the restriction of the set of possible contamination sources and the accuracy of the estimated time of contamination. The Approach A only achieved a better estimate for the time of contamination for the contamination scenario occurring at Node 3.

Table 11.8 – Average results obtained following the Approach B considering 100 hydraulic scenarios under demand uncertainties.

	Parameter 1 (%)	Parameter 2	Parameter 3 (h:min)	Parameter 4 (s)
Node 1	100	58	00:03	1.3
Node 2	100	47	03:16	1.2
Node 3	86	105	00:08	1.9
Node 4	100	4	00:03	0.8
Node 5	100	7	00:01	0.5
Average	97.2	42	00:42	1.1

For hydraulic scenarios with water demand uncertainty, Approaches A and B were able to find the correct contamination source in 97.2% of the contamination scenarios. Both approaches always found the correct location of the contamination source for the scenarios occurring at Nodes 1, 2, 4 and 5 and found the correct location of the contamination source in 86% of the scenarios occurring at Node 3. However, it was verified that in the remaining 14% scenarios occurring at Node 3 the contaminations were not detected by any sensor.

Results obtained following the Approach B constituted a more restricted set of possible contamination sources. It is not possible to observe a general trend concerning the accuracy of the estimated time of contamination, although the average deviation of the estimates obtained following the Approach A is lower. Both approaches achieved less accurate estimates of the time of contamination for some of the case scenarios occurring at Node 2, considering hydraulic scenarios with water demand uncertainty. In these situations, some of the detections were contained only in the vicinity of the corresponding clusters, thus an extrapolation had to be performed.

11.5 Chapter Conclusions

Strategies were developed for addressing the challenges created by large DWDSs. The main challenges were concerned with the existence of a high number of junctions and with the highly irregular behaviour of the DWDSs. The strategies were developed to enable the application of both approaches presented in Chapter 10. To deal with the high number of junctions, a cluster analysis

based on Euclidean distances was applied. As future work, this strategy may be improved by ensuring that the entire set of nodes that belongs to each cluster is hydraulically interconnected.

In Sections 11.4.1 and 11.4.2 a case study is presented and good agreements are observed between the estimates determined by each approach and the simulated time of contamination. This case study shows that the proposed method is able to identify the correct contamination source and to predict the correct time of contamination associated to each possible contamination source, even in the case of a large and highly complex DWDS.

The method was able to find the contamination sources without having any information concerning the absolute values of the contaminant concentration throughout the DWDS. Thus, the effect of measurements uncertainty is mitigated. As in the method presented in Chapter 10, the search for a contamination source can be triggered by any anomaly registered at a sensor.

The case studies presented demonstrated that both approaches required a very low time of computation to obtain the results for real DWDS, generally less than 5 seconds in a 3.10 GHz processor, which is a great improvement to the solution of this problem that demands the computation of the results as quick as possible.

The tests performed with contamination scenarios considering water demand uncertainty demonstrated that the method has shown good performance even in situations that are not described by the hydraulic model used in the development of the ANNs. Those tests showed different advantages of each approach but were not conclusive regarding the question of which was the most suitable for application in real situations.

Nevertheless, it is concluded that this method is suitable for application in real case scenarios. Since both approaches require a very low time of computation and demonstrated good performance in the tests performed under water demand uncertainty, but sometimes presented different results, it was concluded that it would be wise to follow both approaches and use the two set of results to validate each other.

11.6 Acknowledgements

The author is thankful to Dr. Olivier Piller from the Bordeaux Centre of Irstea for the generation of the hydraulic scenarios under water demand uncertainty.

12 Application of Artificial Neural Networks to the Problem of Localization of Contamination Sources: Extension to multiple source scenarios

12.1 Chapter Overview

Chapters 10 and 11 present methods, constituted by two different approaches that are able to find the contamination sources without having any information about the contaminant concentration throughout the DWDS, through the application of ANNs.

However, the methods presented in these two chapters are only suitable for single contamination scenarios. In the event of a multiple contamination event, it would not be possible to have prior knowledge about the contamination that is associated to each detection. In fact, when the search for the localization of the contamination sources is started the number of contaminations will probably be unknown. Thus, the initial pruning performed at the first stage and the application of the Approach B in the second stage are not suitable for applying in a multiple contamination scenario, since there is not enough information regarding the detection profile of each contamination.

Thus, in this chapter, a new algorithm, based on the Approach A of the second stage form the previous chapters, is proposed to extend the application of ANNs for solving the problem of localization of contamination sources in multiple contaminations scenarios.

The effect of water demand uncertainties in the performance of the proposed method is also evaluated in this chapter.

12.2 Methods Description

The proposed method aims to predict the localization of contaminations in DWDSs, based on the information given by a surveillance system constituted by a set of sensors, through the application of ANNs in the event of a single or multiple contamination scenarios. The method is based on the Approach A of the second stage from the already mentioned method presented in Chapter 10.

This procedure requires the creation of databases with the information associated to a large number of contamination scenarios. These scenarios shall include contaminations at each node in the network, for

several instants of contamination, constituting a database that is used to develop the necessary ANNs as it was described in Chapter 11.

Section 12.2.1 describes the steps that constitute the proposed method.

12.2.1 Algorithm for the Localization of Contamination Sources for a Single/Multiple Contamination Event

Step 1 – After the first detection, initialize the number of contaminations, N=1.

Step 2 – A new contamination was detected, so for each node, in which a contamination can be detected by the surveillance system, verify if the sensor in analysis is supposed to detect a contamination occurring at that node. If it is not, the node is not a possible contamination source. Otherwise, proceed to Step 3.

Step 3 – Verify at each cluster of the training set used in the development of the ANNs for the pair node/sensor in analysis the detection is contained. If it is not contained in any cluster, the node is not a possible contamination source. Otherwise, proceed to Step 4.

Step 4 – Calculate the estimate of the time of contamination using the ANN developed for the pair node/sensor in study with the time of the detection being evaluated. Associate the detection and the estimates related to each remaining possible contamination source to the contamination N.

Step 5 – For each possible contamination source, verify if that contamination was supposed to be detected by other sensors that have not detected any contamination. If this is true, that contamination source is not considered possible anymore.

Step 6 – For each new sensor registering a contamination:

> *Step 6.1* – Determine the most suitable detection to be associated to the previously detected contamination in analysis. For that detection, calculate the estimates of the time of contamination using the ANN developed for each pair node/sensor, following the procedure described in Steps 3 and 4.
>
> *Step 6.2* – Calculate the average between each estimate and the time of contamination associated to the respective contamination source. If the difference between the estimate and the average is higher than a maximum value (that needs to be defined), the node associated to the estimate is not a possible contamination source, for the contamination that is being

evaluated. Otherwise, the node remains as a possible contamination source and the associated time of contamination is updated to the average value calculated in this step.

Step 6.3 – If there isn't any remaining possible contamination source after the comparison performed at Step 6.2, the detection in analysis is not associated to the contamination used in the comparison. Otherwise, the set of possible contamination sources associated is updated considering the results of the comparison performed at Step 6.2.

Step 6.4 – For each remaining possible contamination source, verify if that contamination was supposed to be detected by other sensors that have not detected any contamination. If this is true, that contamination source is not considered possible anymore.

Step 6.5 – After finishing the evaluation of every previously detected contaminations, through the procedure describe from Step 6.1 to Step 6.4, if there isn't any contamination associated to the detection in analysis, it is considered the existence of a new contamination, so the number of contaminations is updated, N=N+1, and the algorithm returns to the Step 2. Otherwise, the sensor in analysis is associated to the respective contamination and the analysis is finished.

Step 7 – If the first detection was not selected at Step 6.1, it is considered the existence of a new contamination, so the number of contaminations is updated, N=N+1, and the algorithm returns to the Step 2 for analysing that detection.

Step 8 – If there is any new sensor registering a contamination, for each remaining possible contamination source from the different contaminations previously detected, verify if that contamination was supposed to be detected by other sensors that have not detected any contamination until the final time of the analysis. If this is true, that contamination source is not considered possible anymore. Stop the algorithm.

12.3 Results and Discussion

The Network D, presented in Section 5.4, was used as case study. Figure 12.1 presents the locations of Nodes 1 (red), 2 (blue), 3 (yellow) 4 (pink) and 5 (light blue), in which contaminations were simulated, occurring at 06:00 with a duration of 1 h. 10 different combinations of these 5 single contamination scenarios, described in the Table 12.1, were also simulated. The robustness of the proposed method against water demand uncertainty was evaluated. The contamination scenarios were also evaluated considering two different types of hydraulic behaviour: a) an ideal situation, with no model error and no parameter error; b) 100 different hydraulic scenarios, with no model error but with

water demand uncertainty. The hydraulic behaviours with demand uncertainty were the same files that were used in Chapter 11.

The same ANNs that were trained to perform the case studies following the Approach A in Chapter 11 were used to estimate the time of contamination based on the times of detection of each sensor that detects contaminations at a given junction node.

Table 12.1 – Description of scenarios.

Scenario	Nodes
1	1
2	2
3	3
4	4
5	5
6	1+2
7	2+3
8	3+4
9	4+5
10	1+2+3
11	2+3+4
12	3+4+5
13	1+2+3+4
14	2+3+4+5
15	1+2+3+4+5

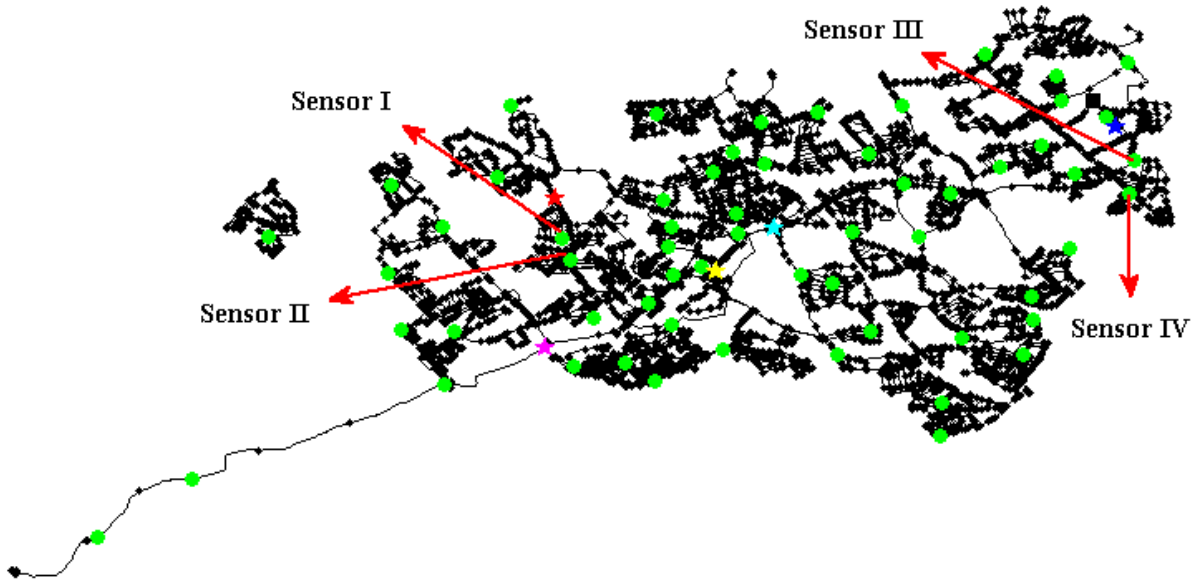

Figure 12.1 - Real DWDS used as case study.

Section 12.3.1 presents a detailed description of an application of the proposed method to a multiple contamination scenario. Section 12.3.2 presents the results achieved for the contamination scenarios presented in Table 12.1. The tests were performed in a 3.10 GHz processor.

12.3.1 Algorithm demonstration

The evaluation of Scenario 6, with simultaneous contaminations occurring at Nodes 1 and 2, considering ideal hydraulic behaviour, is presented to demonstrate the running of the algorithm. Table 12.2 shows the detections associated to this contamination scenario.

Table 12.2 – Detections of Scenario 6, considering ideal hydraulic conditions.

Sensor	Time of detection (h:min)
I	07:15
II	07:42
III	07:50
IV	09:42
IV	09:47

The first detection occurs at Sensor I, at 07:15. The number of contaminations was initialized, N=1. Each node that can be detected by Sensor I was evaluated to determine which time of contamination was associated to the detection in analysis, following a procedure similar to the one presented for the Approach A of the Second Stage in Chapter 11. Firstly, it was necessary to verify in which cluster the detection at the sensor was included for each node. After that, the ANN developed for the chosen cluster was used to determine the time of contamination associated to node in analysis.

After applying this procedure, a set of 218 possible contamination sources, presented in Figure 12.2 as red dots, was defined. It was necessary to determine if each possible contamination source was supposed to be detected by other sensors that had not detected any contamination. After this verification, it was possible to restrict the set of possible contamination sources to 71 nodes, presented in Figure 12.3 as red dots. The analysis of the first detection was finished with the identification of the Contamination 1 with 71 possible sources.

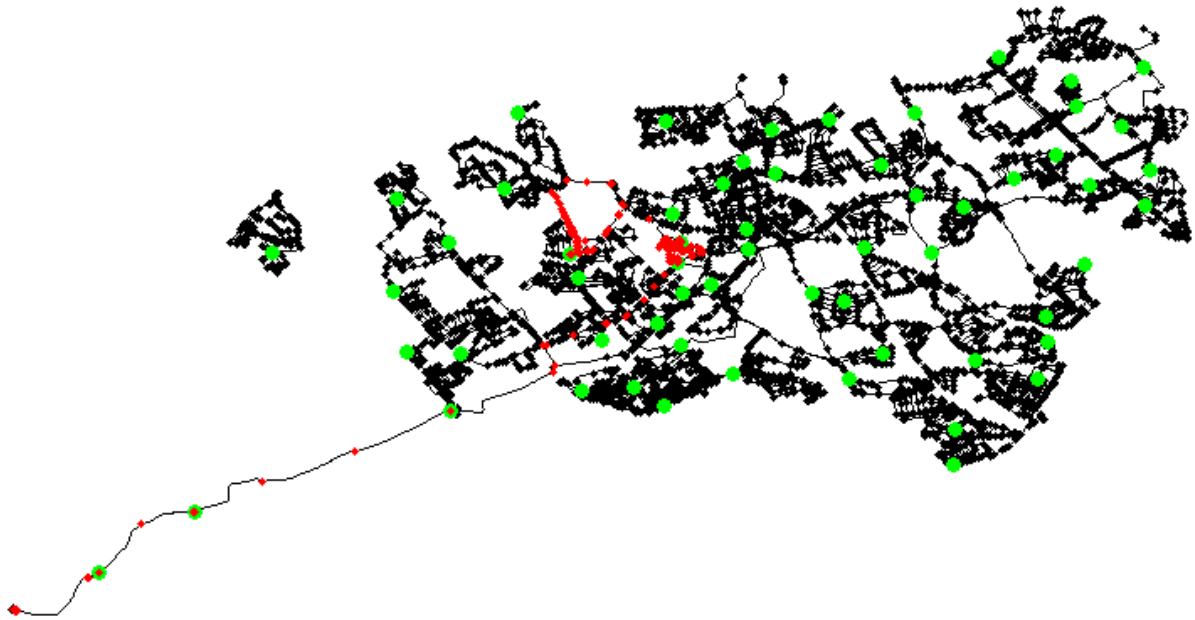

Figure 12.2 - Results achieved from the analysis of the first detection after Step 4.

Figure 12.3 - Results achieved from the complete analysis of the first detection.

The second detection occurs at Sensor II, at 07:42. Each possible contamination source associated to the Contamination 1 was evaluated using the ANNs developed for each pair node/Sensor II. 7 nodes that didn't have any ANN developed for detections at Sensor II (because Sensor II never detects contaminations occurring at these nodes) were eliminated from this evaluation. The average between the estimates obtained from the evaluation of the detections at both sensors was calculated and the deviation between each estimate and the average was evaluated. All 64 remaining possible contamination sources evaluated had a deviation in relation to the average lower than the defined threshold (3 hours). Thus, the times of contamination associated to each possible contamination source of the Contamination 1 was updated to the average between the estimates obtained from both detections. After verifying if each possible contamination source was supposed to be detected by other sensors that had not detected any contamination, it was possible to restrict the set of possible contamination sources to 63 nodes, presented in Figure 12.4 as red dots. The analysis of the second detection was finished with another detection of the Contamination 1 and the restriction of the set of possible sources to 63 nodes.

Figure 12.4 - Results achieved from the complete analysis of the second detection.

The third detection occurs at Sensor III, at 07:50. Evaluating each possible contamination source associated to the Contamination 1, it was possible to conclude that any possibility would be detected by Sensor III. Thus, the detection in analysis is not associated to the Contamination 1 and the procedures contained in the different sub steps of the Step 6 might be skipped. The number of contaminations was updated, N=2, and the algorithm returned to the Step 2.

Each node that can be detected by Sensor III was evaluated to determine which time of contamination was associated to the detection in analysis, as it had been already done for the detection at Sensor I.

After applying this procedure, a set of 109 possible contamination sources, presented in Figure 12.5 as blue dots, was defined. It was also necessary to determine if each possible contamination source was supposed to be detected by other sensors that had not detected any contamination. After this verification, it was possible to restrict the set of possible contamination sources to 91 nodes, presented in Figure 12.6 also as blue dots. The analysis of the third detection was finished with the identification of the Contamination 2 with 91 possible sources.

Figure 12.5 - Results achieved from the analysis of the third detection after Step 4.

Figure 12.6 - Results achieved from the complete analysis of the third detection.

Sensor IV registered detections at 09:42 and 09:47. Analysing the possible contamination sources associated to the Contamination 1, it was concluded that any possibility would be detected by that sensor. Thus, there was no point in determining which detection would be more suitable to explain the Contamination 1 since that sensor clearly was not associated to the Contamination 1. Analysing the databases used in the training of the ANNs associated to the Sensor IV for each possible contamination source associated to Contamination 2, it was concluded that the detection that occurred at 09:42 was more suitable to explain the majority of the possible contamination sources.

Each possible contamination source associated to the Contamination 2 was evaluated using the ANNs developed for each pair node/Sensor IV. 5 nodes that didn't have any ANN developed for detections at Sensor IV (because Sensor IV never detects contaminations occurring at these nodes) and were eliminated from this evaluation. The average between the estimates obtained from the evaluation of the detections at both sensors was calculated and the deviation between each estimate and the average was evaluated. All 86 remaining possible contamination sources evaluated also had a deviation in relation to the average lower than 3 hours and the times of contamination associated to each possible contamination source were updated to the average between the estimates obtained from both detections. After verifying if each possible contamination source was supposed to be detected by other sensors that had not detected any contamination, it was possible to restrict the set of possible contamination sources to 43 nodes, presented in Figure 12.7 as blue dots. The analysis of the second detection was finished with another detection of the Contamination 2 and with the restriction of the associated set of possible sources to 43 nodes.

Figure 12.7- Results achieved from the complete analysis of the fourth detection.

Since there weren't any new detection, the sets of possible contamination sources were updated considering a final time of analysis of 48:00 to finish the algorithm. The method was able to restrict the set of possible contamination sources of Contaminations 1 to 56 nodes, presented in Figure 12.8 as red dots. This step did not enable any further restriction of the set of possible contamination sources associated to Contamination 2, presented in Figure 12.8 as blue dots, which remained with 43 nodes.

Figure 12.8 - Results achieved at the end of the algorithm.

12.3.2 Case studies

The performance of the method was evaluated, as in Chapter 11, considering 4 indicators:

1. Target the real source;
2. Number of possible contamination sources obtained;
3. Difference between the obtained estimates and the real time of contamination;
4. Time of computation required.

Tables 12.3 and 12.4 present the results obtained for parameters 1, 2 and 3, considering a maximum value of deviation between each estimate and the average of all estimates of 3 hours. Table 12.3 presents the results obtained for the tests considering the ideal hydraulic scenario and Table 12.4 presents the average results obtained considering 100 hydraulic scenarios under demand uncertainty. Table 12.5 presents the time of computation and the average time of computation required by the method to analyse each scenario (parameter 4) for ideal hydraulic conditions and for the hydraulic scenarios under demand uncertainty, respectively.

Table 12.3 - Results obtained considering the ideal hydraulic scenario.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	57	00:10
Node 2	100	44	00:45
Node 3	100	116	00:01
Node 4	100	4	00:04
Node 5	100	11	00:02
Average	100	50	00:13

Table 12.4 - Average results obtained considering 100 hydraulic scenarios under demand uncertainties.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	56	00:28
Node 2	100	42	01:42
Node 3	89	115	00:09
Node 4	98	4	00:04
Node 5	99	10	00:02
Average	96	48	00:32

The proposed method achieved the correct solution in 100% of the case scenarios simulated under ideal hydraulic conditions, with an average of 50 possible contamination sources and 13 minutes of deviation between the estimates and the simulated time of computation.

For case scenarios simulated considering hydraulic behaviour with water demand uncertainty, 96% of the simulated contaminations. The success rate was very close to 100% except for contaminations occurring at Node 3, in which the method still targeted the correct source in 89% of the case scenarios. In this case scenario, the intersection between different contamination events, caused by the higher level of uncertainty, does not enable the correct localization of all real contamination sources. The set of possible contamination sources was very similar to the set defined for ideal hydraulic conditions,

with an average of 50 possible contamination sources. The deviation between the estimates and the simulated times of contamination was higher (average deviation of 32 minutes) than the deviation registered with ideal hydraulic scenarios, especially for contaminations occurring at Nodes 1 and 2.

Table 12.5 – Time of computation required by the method.

Scenario	Ideal (s)	Uncertainty (s)
1	22	19
2	13	13
3	6	6
4	18	24
5	16	14
6	37	34
7	24	22
8	21	35
9	48	47
10	50	45
11	39	55
12	58	56
13	79	79
14	74	74
15	141	132

Analysing the Table 12.5, it is possible to observe that there was no significant difference between the times of computation required for evaluating contamination scenarios with ideal hydraulic conditions and contamination scenarios with water demand uncertainty. The time of contamination seems to be lower for single contamination scenarios (Scenarios 1-5) and registers its highest value for Scenario 15 (contaminations at 5 nodes).

Comparing the results presented in Tables 12.6 and 12.7, it is possible to observe that, for scenarios simulated considering ideal hydraulic conditions, the proposed method has a very similar performance

for single and multiple contamination scenarios. The simulated contamination source belonged always to the set of possible contamination sources and the deviation between the estimates and the simulated times of contamination is very similar. The only observed difference is in the number of possible contamination sources. For some nodes, the results achieved for the set of possible contamination sources were less restrict for multiple contamination (average of 52 possible contamination sources against 44 for single contamination scenarios).

Table 12.6 - Results obtained for single contamination scenarios, considering the ideal hydraulic scenario.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	56	00:10
Node 2	100	44	00:45
Node 3	100	105	00:01
Node 4	100	4	00:04
Node 5	100	9	00:02
Average	100	44	00:12

Table 12.7 – Average results obtained for multiple contamination scenarios, considering the ideal hydraulic scenario.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	58	00:10
Node 2	100	44	00:45
Node 3	100	117	00:01
Node 4	100	4	00:04
Node 5	100	11	00:02
Average	100	52	00:12

Tables 12.8 and 12.9 show the results obtained considering 100 hydraulic scenarios under demand uncertainties for single and multiple contamination scenarios, respectively. The method was able to

detect 99% of the single contamination scenarios, with an average of 43 possible contamination sources and an average deviation of 29 minutes between the estimates and the simulated times of contamination. For multiple contamination scenarios, the method was able to successfully locate 96% of the contaminations. The success rate is almost as high as for single contamination scenarios for every contamination scenario, except for the ones occurring at Node 3, which still were correctly located in 88% of the respective scenarios. No significant differences were observed in the number of possible contamination sources and in the deviation between the estimates and the simulated times of contamination.

Table 12.8 - Results obtained for single contamination scenarios, considering 100 hydraulic scenarios under demand uncertainties.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	55	00:29
Node 2	100	42	01:42
Node 3	100	105	00:10
Node 4	98	4	00:04
Node 5	99	9	00:02
Average	99	43	00:29

Table 12.9 – Average results obtained for multiple contamination scenarios, considering 100 hydraulic scenarios under demand uncertainties.

	Parameter 1 (%)	**Parameter 2**	**Parameter 3 (h:min)**
Node 1	100	56	00:29
Node 2	100	42	01:42
Node 3	88	116	00:09
Node 4	98	4	00:04
Node 5	98	11	00:02
Average	96	49	00:32

A comparison was made between the results achieved for single contamination scenarios (Tables 12.6 and 12.8) and the results presented for the same contamination scenarios in Tables 11.5 and 11.6 of the Chapter 11.

For scenarios considering ideal hydraulic behaviour, the proposed method achieved the same success rate (100%) and the same deviation between the estimates and the simulated times of contamination. Regarding the parameter 2, the proposed method enabled a higher restriction of the set of possible contamination sources in all scenarios (average of 44 possible contamination sources against 77 from the previous work). The time of computation required from the proposed method was higher in the majority of the scenarios, but still remained lower than 30 s.

For scenarios considering water demand uncertainty, achieved a higher average success rate (99% against 97%), especially for contaminations occurring at Node 3. The success rate was slightly lower for Nodes 4 and 5. As it already occurred for single contamination scenarios, the proposed method was able to get a more restrict set of possible contamination sources in all scenarios (average of 43 possible contamination sources against 75 from the previous work). Regarding the parameter 3, the proposed method achieved a lower average deviation between the estimates and the simulated times of contamination, however it is not possible to observe a consistent trend in this parameter. The time of computation required from the proposed method was higher in all scenarios, but still remained lower than 30 s.

Some of the differences presented by some of the results obtained by the different methods are related with a difference in the definition of the contamination scenarios. In Chapter 11, the contaminations were considered as instantaneous injections, while this method was applied to scenarios that considered injections with 1 h of duration. Thus, in some scenarios there are more sensor detections to evaluate. For instance, for Node 3, 14% of the scenarios considering water demand uncertainty are not detected for instantaneous injections, but are detected when the injection is simulated during 1 h.

12.4 Chapter conclusions

A new algorithm, based on the Approach A of the second stage form the methods presented in Chapters 10 and 11, was proposed to extend the application of ANNs for solving the problem of localization of contamination sources in multiple contaminations scenarios.

In Section 12.3.2, case studies are presented, with the simulation of single or multiple contamination scenarios, considering either ideal hydraulic behaviour or water demand uncertainty.

The method localized successfully the simulated sources of contamination for all single or multiple contamination scenarios in which ideal hydraulic behaviour was considered. For single and multiple contamination scenarios considering water demand uncertainty, the method successfully localized the great majority of the contamination sources. The restriction of the sets of possible contamination sources was very similar for all types of hydraulic conditions. The method achieved more accurate estimates of the times of contamination for scenarios that considered an ideal hydraulic condition; however, it still achieved fairly approximate estimates for scenarios considering water demand uncertainty. The effect of water demand uncertainty does not have any observable effect on the required time of computation. It is observed that the evaluation of single contamination scenarios is quicker than the evaluation of multiple contamination scenarios. This was expected since the time of computation depends on the number of detections to be evaluated. It is possible to conclude that the method achieved better results for scenarios with ideal hydraulic conditions, as it was expected, but still registered a very good performance for scenarios considering water demand uncertainty.

The method registers very similar performances for single and multiple contamination scenarios, when ideal hydraulic conditions are considered. The only observed difference is in the level of restriction of the set of possible contaminations sources, in which results might be better for single contamination scenarios, since more detections associated to each contamination might be analysed.

For scenarios considering water demand uncertainty, the method registered better performance for single contamination scenarios. However, the success rate is almost as high as for single contamination scenarios for every contamination scenario, except for the ones occurring at Node 3. In several hydraulic scenarios with water demand uncertainty, the method was able to localize correctly the contamination occurring at Node 3 when this contamination was simulated alone but failed to achieve the same performance when other contaminations were simulated simultaneously. This is due to the effect of the water demand uncertainty, which sometimes leads to greater changes in the hydraulic behaviour of the DWDS, leading to times of detections at the sensors quite different from the ones used in the training of the ANNs.

The method achieved also some improvements in the results obtained for single contamination scenarios. Comparing the results from this method with the results presented in Chapter 11 for the same scenarios, it is concluded that the proposed method enabled a higher restriction of the set of possible contamination sources for scenarios with ideal hydraulic behaviour. For scenarios considering water demand uncertainty, it achieved a higher average success rate and a more restrict set of possible contamination sources in all scenarios. The time of computation required by the proposed method was higher in the majority of the scenarios, but still remained lower than 30 s.

Part of the differences presented by some of the results obtained by the various methods are related to a difference in the definition of the contamination scenarios. In Chapter 11, the contaminations were considered as instantaneous injections, while this method was applied to scenarios that considered injections with 1 h of duration. Thus, in some scenarios there are more sensor detections to evaluate.

The proposed method is able to extend the application of ANN to the localization of contamination sources in multiple contamination scenarios, even for real and highly complex DWDS. The method was also able to find the contamination sources without having any information concerning the absolute values of the contaminant concentration throughout the DWDS, mitigating the effect of measurements uncertainties.

The proposed method generally identifies the correct contamination source in the majority of the evaluated scenarios and good agreements are observed between the estimates determined by the method and the simulated times of contamination, without significant decrease of performance in the application to multiple contamination scenarios.

The time of computation is slightly higher than the time required by the method presented in Chapter 11, but it still remained in a very reasonable level (on average, less than 1 minute). Thus, it is concluded that this method is a very suitable solution for application in real case scenarios.

12.5 Acknowledgements

The author is thankful to Dr. Olivier Piller from the Bordeaux Centre of Irstea for the generation of the hydraulic scenarios under water demand uncertainty.

Part V: Conclusions and Future Work

13 Main Conclusions and Suggested Future Work

13.1 Main Conclusions

This thesis was aimed to study the occurrence of contaminations in DWDSs. The work performed during this period had the following major outcomes:

1- The development of two approaches for simulating the transport mechanisms of contaminants in DWDSs;
2- The development of deterministic methods for the localization of contamination sources;
3- The application of ANNs in the development of methods for the localization of contamination sources;

Regarding the simulation of the transport mechanisms of contaminants in DWDSs, a software tool was developed to implement an analytical approach for the simulation of the advective transport of contaminants, considering pseudo-first order reaction terms. This approach proved to be suitable for providing the analytical solution for the modelling of the transport of contaminants in DWDS, considering steady hydraulic conditions. In addition, a software tool was developed to simulate the contaminant reactive transport along DWDSs considering sorption phenomena. This method also demonstrated to be a relevant contribution for the study of the effects of the sorption phenomena in the modelling of the transport of contaminants in real DWDS.

The challenge of the localization of contamination sources in DWDSs was addressed following two types of approach: one deterministic and the other based on the application of ANNs.

Two deterministic methods based on the analysis of the residence time of water in pipes were developed, considering fixed time intervals or based on successive detections at sensors. One of the limitations identified in several works presented in the review of the works performed in the past on this subject was the dependency on the existence of accurate concentration profiles at the sensors. This type of information is really difficult to get, especially in a deliberate contamination event, since the contaminant introduced in the DWDS might be unknown making it difficult to evaluate its concentration. The deterministic methods developed in this work were able to achieve reliable solutions for the characterization of the contamination sources requiring only a binary sensor status over time, avoiding the need of having accurate readings of the contaminant concentrations at the sensors. Furthermore, it was possible to confirm, as it was expected, that the constitution of the set of sensors available in the surveillance system has great influence in the performance of this kind of

methods. The method based on successive detections at sensors is not affected by the occurrence of false or positive readings at sensors and enables the verification of the occurrence these abnormalities.

A method was developed to address the problem of the localization of contamination sources after deliberate single contamination event in DWDSs through the application of ANNs. Additionally, strategies were developed for addressing the challenges created by large DWDSs. The main challenges were concerned with the existence of a high number of junctions and with the highly irregular behaviour of the DWDSs.

Two different approaches were developed to determine the probability of each node of the DWDS being the contamination source and to estimate the corresponding time of contamination for nodes with probabilities different from zero. One important feature of this method is that it is also able to find the contamination sources without having any information about the contaminant concentration throughout the DWDS. Following this method, the search for a contamination source can be triggered by any anomaly registered at a sensor. The case studies performed showed that both approaches required a very low time of computation to obtain the results for real DWDS, generally less than 5 seconds in a 3.10 GHz processor, which is a great improvement to the solution of this problem that demands the computation of the results as quick as possible. The tests performed with contamination scenarios considering water demand uncertainty demonstrated that the method has shown good performance even in situations that are not described by the hydraulic model used in the development of the ANNs.

The method that extended the application of ANNs for multiple contamination scenarios achieved very satisfactory results for real DWDSs. The method was generally able to determine correctly the simulated source and to define a very restricted set of possible contamination sources, even when considering hydraulic scenarios with demand uncertainties. However, the estimations of the time of contamination for scenarios under demand uncertainties showed larger deviation in relation with the simulated contamination sources. The time of computation required was generally very low, which makes this method very suitable for application in real contamination scenarios.

13.2 Suggested Future work

Concerning the development of methods for the simulation of the spread of contaminants in DWDSs, the application of other numerical schemes, described in previous works mentioned in Chapter 2 as more suitable for modelling steep concentration profiles considering advective transport phenomenon, could improve the accuracy and the reliability of the results obtained

The field of localization of contamination sources in DWDS is not yet consolidated and there is still plenty of room for improvement.

The performance of the deterministic methods developed in this thesis still need to be evaluated against the existence of water demand uncertainties. Besides that, future work can be performed to reduce the computational effort required by these methods.

All the methods developed in this thesis for the localization of contamination sources provide estimations for the location and instant of contamination. Further work could be performed to enhance their outcomes, with the introduction of an estimation of the duration and amplitude of the contamination. Furthermore, a differentiation of the risk levels associated to each node of the DWDS could be included to prioritize the verification of critical nodes. Moreover, proposed methods were developed base on the assumption that the sensors have no detection limit. Future improvements could avoid the need for this assumption.

The installation of a pilot scale network, with a reasonable level of complexity, could be very useful for testing the performance of methods developed for the simulation of the water quality and localization of contamination sources in DWDSs

Finally, the majority of the methods developed for the localization of contamination sources rely on the information collected by a surveillance system constituted by water quality sensors. However the implementation of a comprehensive surveillance system constituted by water quality sensor is still conditioned by the high installation costs associated. Thus, the number of sensors available for installation tends to be kept as low as possible to achieve reasonable installation and operation costs. As consequence, even for successfully detected contamination events, the information provided by the sensors may not be sufficient to achieve a satisfactorily restricted set of possible contamination sources. Future research could be performed to achieve a further restriction of the set of possible contamination sources by extending the evaluation of the contamination events to the analysis of water/deposits samples gathered at some selected points among the initial set of possible contamination sources.

References

Abe, J. M., Nakamatsu, K., 2012. Paraconsistent Artificial Neural Networks and Pattern Recognition: Speech Production Recognition and Cephalometric Analysis. Advances in Reasoning-Based Image Processing Intelligent Systems, 29, 365-382.

Ahmed, F. E., 2005. Artificial neural networks for diagnosis and survival prediction in colon cancer. Molecular Cancer, 4-29 (doi:10.1186/1476-4598-4-29).

Al-Alaoui, M. A., Al-Kanj, L., Azar, J., Yaacoub, E., 2008. Speech Recognition using Artificial Neural Networks and Hidden Markov Models. IEEE Multidisciplinary Engineering Education Magazine, 3(3), 77-86.

Al-Omari, A. S., Chaudhry, M. H., 2001. Unsteady-state inverse chlorine modeling in pipe networks. Journal of Hydraulic Engineering, 127, 669-677.

Altun, H., Curtis, K.M., Yalcınoz, T., 2003. Neural learning for articulatory speech synthesis under different statistical characteristics of acoustic input patterns. Computers and Electrical Engineering, 29, 687–702.

Alhumaizi, K., Henda, R., Soliman, M., 2003. Numerical analysis of a reaction-diffusion-convection system. Computers and Chemical Engineering, 27(4), 579-594.

Alhumaizi, K., 2004. Comparison of finite difference methods for the numerical simulation of reacting flow. Computers and Chemical Engineering, 28, 1759-1769.

Alhumaizi, K., 2007. Flux-limiting solution techniques for simulation of reaction–diffusion–convection system. Communications in Nonlinear Science and Numerical Simulation, 12(6), 953–965.

Argiriou, A. A., 2007. Use of neural networks for tropospheric ozone time series approximation and forecasting – a review. Atmospheric Chemistry and Physics Discussions, 7, 5739–5767.

Azwar, Hussain, M. A., Ramachandran, K. B., The study of neural network-based controller for controlling dissolved oxygen concentration in a sequencing batch reactor. Bioprocess and Biosystems Engineering, 28, 251–265.

Boulos, P. F., Altman, T., Sadhal, 1992. Computer modeling of water quality in large multiple source networks. Applied Mathematical Modelling, 16(8), 439–445.

Boulos, P. F., Altman, T., 1993. Explicit Calculation of Water Quality Parameters in Pipe Distribution Systems. Civil Engineering Systems, 10(3), 187-206.

Boulos, P. F., Altman, T., Jarrige, P. A., Collevati, F., 1994. An event driven method for modelling contaminant propagation in water networks. Applied Mathematical Modelling, 18(2), 84-92.

Boulos, P. F., Altman, T., Jarrige, P. A., Collevati, F., 1995. Discrete Simulation Approach for Network-Water-Quality Models. Journal of Water Resources Planning and Management, 121(1), 49–60.

Bowden, G. J., Nixon, J. B., Dandy, G. C., Maier, H. R., Holmes, M., 2006. Forecasting chlorine residuals in a water distribution system using a general regression neural network. Mathematical and Computer Modelling, 44, 469–484.

Cerri, G., Borghetti, S., Salvini, C., 2006. Neural management for heat and power cogeneration plants. Engineering Applications of Artificial Intelligence, 19, 721–730.

Chau, K., 2006. A review on the integration of artificial intelligence into coastal modeling. Journal of Environment Management, 80,47-57.

Chung, G., Lansey, K., Bayraksan, G., 2009. Reliable water supply system design under uncertainty, Environmental Modelling & Software, 24, 449-462.

Clark, R. M., Rossman, L. A., Wymer, L. J., 1995. Modeling distribution system water quality: regulatory implications. Journal of Water Resources Planning and Management, 121, 423-428.

Coelho, S. T., Loureiro, D., Alegre, H., 2006. Modelação e Análise de Sistemas de Abastecimento de Água. Instituto Regulador de água e Resíduos, Laboratório Nacional de Engenharia Civil, 2006.

Cugat, M., 2012. Contamination Source Determination in Water Distribution Networks. SIAM Journal on Applied Mathematics, 72 (6), 1772-1791.

Daqi, G., Genxing, Y., 2003. Influences of variable scales and activation functions on the performances of multilayer feedforward neural networks. Pattern Recognition, 36, 869-878.

Davidson, J., Bouchart, F., Cavill, S., and Jowitt, P., 2005. Real-Time Connectivity Modeling of Water Distribution Networks to Predict Contamination Spread. Journal of Computing in Civil Engineering, 19(4), 377-386.

Dawsey, W., Minsker, B., VanBlaricum, V., 2006. Bayesian Belief Networks to Integrate Monitoring Evidence of Water Distribution System Contamination. Journal of Water Resources Planning and Management, 132(4), 234-241.

Dayhoff, J. E., 1990. Neural Network Architectures – An Introduction. Van Nostrand Reinhold, U.S.A..

De Sanctis, A. E., Shang, F., Uber, J. G., 2010. Real-time identification of possible contamination sources using network backtracking methods. Journal of Water Resources Planning and Management, 136 (4), 444-453.

Deliverable 4.2 - Experimental data of sorption kinetics, "SecurEau" project, October 2012 (confidential report).

Di Cristo, C., Leopardi, A., 2008. Pollution Source Identification of Accidental Contamination in Water Distribution Networks. Journal of Water Resources Planning and Management, 134(2), 197-202.

Dede, G., Sazlı, M. H., 2010. Speech recognition with artificial neural networks. Digital Signal Processing, 20, 763–768.

Di Luca, M., Grossi, E., Borroni, B., Zimmermann, M., Marcello, E., Colciaghi1, F., Gardoni1, F., Intraligi, M., Padovani, A., Buscema, M., 2005. Artificial neural networks allow the use of simultaneous measurements of Alzheimer Disease markers for early detection of the disease. Journal of Translational Medicine, 3-30 (doi:10.1186/1479-5876-3-30).

Di Nardo, A., Di Natale, M., Guida, M., Musmarra, D., 2012. Water Network Protection from Intentional Contamination by Sectorization. Water Resources Management, doi: 10.1007/s11269-012-0133-y.

Dibike, Y. B., Coulibaly, P., 2006. Temporal neural networks for downscaling climate variability and extremes. Neural Networks, 19(2), 135-44.

Dong, C., Schoups, G., van de Giesen, N., 2012. Scenario development for water resource planning and management, Technological Forecasting & Social Change, (article in press).

Egmont-Petersen, M., de Ridder, D., Handels, H., 2002. Image processing with neural networks—a review. Pattern Recognition, 35(10), 2279–2301.

Eliades, D. G., Polycarpou, M. M., Water Contamination Impact Evaluation and Source-Area Isolation Using Decision Trees. Journal of Water Resources Planning and Management, 138(5), 562-570.

Fabrie, P., Gancel, G., Montazavi, I., Piller, O., 2010. Quality Modeling of Water Distribution Systems Using Sensitivity Equations. Journal of Hydraulic Engineering, 136 (1), 34-44.

Farkas, I., Géczy-Víg, P., 2003. Neural network modelling of flat-plate solar collectors. Computers and Electronics in Agriculture, 40, 87 -102.

Gancel, G., Mortazavi, I., Piller, O., 2006. Coupled numerical simulation and sensitivity assessment for quality modelling for water distribution systems. Applied Mathematics Letters, 19(12), 1313–1319.

Gao, P., Chen, C., Qin, S., 2010. An Optimization Method of Hidden Nodes for Neural Network, 2010 Second International Workshop on Education Technology and Computer Science.

Gleick, P. H., 2006. Water and terrorismo. Water Policy, 8(6), 481–503.

Gómez-Sanchis, J., Martín-Guerrero, J. D., Soria-Olivas, E., Vila-Francés, J., Carrasco, J. L., del Valle-Tascón, S., 2006. Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration. Atmospheric Environment, 40(32), 6173–6180.

Guan, J., Aral, M. M., Maslia, M. L., Grayman, W. M., 2006. Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. Journal of Water Resources Planning and Management, 132 (4), 252-262.

Hallam, N. B., West, J. R. Forster, C. F., Powell, J. C., Spencer, I., 2002. The decay of chlorine associated with the pipe wall in water distribution systems. Water Research, 36(14), 3479–3488.

Ham, F. M., Kostanic, I., 2001. Principles of Neurocomputing for Science & Engineering. McGraw-Hill Companies, New York, U.S.A..

Haykin, S., Neural Networks – A Comprehensive Foundation, 1999. Prentice-Hall, Canada.

Hill, J., van Bloemen Waanders, B. G., Laird, C. D., 2006. Source Inversion with Uncertain Sensor Measurements. 8th Annual Water Distribution systems Analysis Symposium, Cincinnati, Ohio.

Huang, J. J., McBean, E. A., 2009. Data Mining to Identify Contaminant Event Locations in Water Distribution Systems, Journal of Water Resources Planning and Management, 466-474.

Islam, M. R., Chaudhry, M. H., 1997. Numerical solution of transport equation for applications in environmental hydraulics and hydrology. Journal of Hydrology, 191(1–4), 106–121.

Islam, M. R., Chaudhry, M. H., 1998. Modeling of Constituent Transport in Unsteady Flows In Pipe Network. Journal of Hydraulic Engineering, 124 (11), 1301-1320.

Kamruzzaman, J., 2004. ANN-Based Forecasting of Foreign Currency Exchange Rates. Neural Information Processing - Letters and Reviews, 3(2), 49-58.

Kalogirou, S. A., 2000. Applications of artificial neural-networks for energy systems. Applied Energy, 67, 17-35.

Kaufman, L., Rousseeuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, Inc.

Khan, J., Wei, J. S., Ringnér, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., Meltzer, P. S., 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine, 7(6), 673-679.

Khanal, N., Buchberger, S. G., McKenna, S. A, 2006. Distribution System Contamination Events: Exposure, Influence, and Sensitivity. Journal of Water Resources Planning and Management, 132(4), 283-292.

Kim, M., Choi, C. Y., Gerba, C. P., 2008. Source tracking of microbial intrusion in water systems using artificial neural networks. Water Research, 42, 1308-1314.

Koch, M. W., McKenna, S. A., 2011. Distributed Sensor Fusion in Water Quality Event Detection. Journal of Water Resources Planning and Management, 137 (1), 10-19.

Laird, C.D., Biegler, L. T., van Bloemen Waanders, B. G., Bartlett, R.A.,2004a. Time Dependent Contamination Source Determination for Municipal Water Networks using Large Scale Optimization. (URL: http://www.cs.sandia.gov/~bartv/papers.html) – accessed in 2010/10/28.

Laird, C.D., Biegler, L. T., van Bloemen Waanders, B. G., Bartlett, R.A., 2004b. Time Dependent Contamination Source Determination: A Network Subdomain Approach for Very Large Water Networks. World Water & Environmental Resources Congress, Salt Lake City.

Laird, C.D., Biegler, L. T., van Bloemen Waanders, B. G., Bartlett, R.A., 2005. Contamination source determination for water networks. Journal of Water Resources Planning and Management,131 (2), 125-134.

Laird, C.D., Biegler, L. T., van Bloemen Waanders, B. G., 2006. Mixed integer approach for obtaining unique solutions in source inversion of water networks. Journal of Water Resources Planning and Management, 132 (4), 242-251.

Ławrynczuk, M., 2009. Efficient nonlinear predictive control of a biochemical reactor using neural models. Bioprocess and Biosystems Engineering, 32, 301–312.

Li, E. Y., 1994. Artificial Neural Networks and Their Business Applications. Information & Management, 27(5), 303-313.

Li, S., Kwok, J. T., Wang, Y., 2002. Multifocus image fusion using artificial neural networks. Pattern Recognition Letters, 23(8), 985–997.

Li, Y. P., Huang, G. H., Nie, S. L., Liu, L., 2008. Inexact multistage stochastic integer programming for water resources management under uncertainty. Journal of Environmental Management, 88, 93-107.

Li, Y. P., Huang, G. H., Nie, S. L., 2010. Planning water resources management systems using a fuzzy-boundary interval-stochastic programming method. Advance in Water Resources, 33, 1105-1117.

Li, Y. P., Huang, G. H., Nie, S. L., Chen, X., 2011. A robust modelling approach for regional water management under multiple uncertainties. Agricultural Water Management, 98, 1577-1588.

Lisboa, P. J. G., 2002. A review of evidence of health benefit from artificial neural networks in medical intervention. Neural Networks, 15(1), 11–39.

Lisboa, P. J., Taktak, A. F. G., 2006. The use of artificial neural networks in decision support in cancer: A systematic review. Neural Networks, 19(4), 408–415.

Liu, L., Ranjithan, S. R., Mahinthakumar, G., 2011. Contamination Source Identification in Water Distribution Systems Using an Adaptive Dynamic Optimization Procedure. Journal of Water Resources Planning and Management 137(2), 183-192.

Liu, L., Zechman, E. M., Mahinthakumar, G., Ranjithan, S. R., 2012a. Coupling of logistic regression analysis and local search methods for characterization of water distribution system contaminant source. Engineering Applications of Artificial Intelligence, 25, 309–316.

Liu, L., Zechman, E. M., Mahinthakumar, G., Ranjithan, S. R., 2012b. Identifying contaminant sources for water distribution systems using a hybrid method. Civil Engineering and Environmental Systems, 29(2), 123-136.

Maier, H. R., Dandy, G. C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software 15, 101-124.

Maier, H. R., Jain, A., Dandy, G. C., Sudheer, K. P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software 25, 891-909.

Manly, B.F.J., 1994. Multivariate Statistical Methods - A Primer, second ed. Chapman and Hall, London, 129–133.

Mann, A. V., McKenna, S. A., Hart, W. E., Laird, C. D., 2012. Real-time inversion in large-scale water networks using discrete measurements. Computers and Chemical Engineering, 37, 143– 151.

Martins, F.G., Coelho, M. A. N., 2000. Application of feedforward artificial neural networks to improve process control of PID-based control algorithms. Computers and Chemical Engineering, 24, 853-858.

Mau, R. E., Boulos, P. F., Bowcock, R. W., 1996. Modelling distribution storage water quality: An anlytical approach. Applied Mathematical Modelling, 20(4), 329–338.

May, R. J., Dandy, G. C., Maier, H. R., Nixon, J. B., 2008. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Environmental Modelling & Software 23, 1289-1299.

May, R. J., Maier, H. R., Dandy, G. C., Fernando, T. M. K. G., 2008. Non-linear variable selection for artificial neural networks using partial mutual information. Environmental Modelling & Software 23, 1312-1326.

McKenna, J.E., 2003. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modelling & Software 18 (3), 205–220.

Mou, L., Menglin, W., Jie, L., Shen, D., 2010. Investigation on backward tracking of contamination sources in water supply systems – case study. 2010 2nd Conference on Environmental Science and Information Application Technology.

Munavalli, G. R., Kumar, M. S. M., 2003. Water Quality Parameter Estimation in Steady-State Distribution System. Journal of Water Resources Planning and Management, 129(2), 124-134.

Munavalli, G. R., Kumar, M. S. M., 2004. Modified Lagrangian method for modeling water quality in distribution Systems. Water Research, 38(13), 2973-2988.

Munavalli, G. R., Kumar, M. S. M., 2005. Water quality parameter estimation in a distribution system under dynamic state. Water Research, 39(18), 4287-4298.

Neupauer, R. M., Records, M. K., Ashwood, W. H., 2010. Backward Probabilistic Modeling to Identify Contaminant Sources in Water Distribution Systems. Journal of Water Resources Planning and Management, 136(5), 587-591.

Odan, F. K., Reis, L. F. R., 2012. Hybrid Water Demand Forecasting Model Associating Artificial Neural Network with Fourier Series. Journal of Water Resources Planning and Management, 138(3), 245-256.

Ozdemir, O. N., Ger, A. M., 1999. Unsteady 2-D chlorine transport in water supply pipes. Water Research, 33, 3637-3645.

Ozdemir, O., N., Ucak, A., 2002. Simulation of chlorine decay in drinking-water distribution systems. Journal of Environmental Engineering, 128, 31-39.

Peng, H., Yang, Z. J., Gui, W., Wu, M., Shioya, H., Nakano, K., 2007. Nonlinear system modeling and robust predictive control based on RBF-ARX model. Engineering Applications of Artificial Intelligence, 20,1–9.

Pires, J. C. M., Martins, F. G., Sousa, S. I. V., Alvim-Ferraz, M. C. M., Pereira, M. C., 2008. Prediction of the Daily Mean PM_{10} Concentrations Using Linear Models. American Journal of Environmental Sciences, 4 (5), 445-453.

Poh, H. L., Yao, J., Jasic, T., 1998. Neural Networks for the Analysis and Forecasting of Advertising and Promotion Impact. International Journal of Intelligent Systems in Accounting, Finance & Management, 7, 253–268.

Poulin, A., Mailhot, A., Grondin, P., Delorme, L., Villeneuve, J. P., 2006. Optimization of Operational Response to Contamination in Water Networks. 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, USA.

Preis, A., Ostfeld, A., 2006. Contamination source identification in water systems: a hybrid model trees-linear programming scheme. Journal of Water Resources Planning and Management, 132 (4), 263-273.

Preis, A., Ostfeld, A., 2007. A contamination source identification model for water distribution system security. Engineering Optimization, 39(8), 941–951.

Preis, A., Ostfeld, A., 2010. Genetic algorithm for contaminant source characterization using imperfect sensors. Civil Engineering and Environmental Systems, 25(1), 29-39.

Preis, A., Ostfeld, A., 2011. Hydraulic uncertainty inclusion in water distribution systems contamination source identification. Urban Water Journal, 8(5), 267-277.

Propato, M., Sarrazy, F., Tryby, M., 2010. Linear algebra and minimum relative entropy to investigate contamination events in drinking water distribution systems. Journal of Water Resources Planning and Management, 136 (4), 483-492.

Qin, X.S., Huang, G. H., Zeng, G.M., Chakma, A., Huang, Y. F., 2007. An interval-parameter fuzzy nonlinear optimization model for stream water quality management under uncertainty. European Journal of Operational Research, 180, 1331-1357.

Quaddus, M. A., Ktinn, M. S., 1999. Businccs Applications of Artificial Neural Networks: An Updated Review and Analysis. Neural Information Processing, 1999. Proceedings. ICONIP '99, 819-824.

Quesson, B. A., Sheldon-Robert, M. K., Vloerbergh, I. N., Vreeburg, J. H. G., 2009. Acoustic Monitoring of Terrorist Intrusion in a Drinking Water Network. Proceedings of the 10th Annual Water Distribution Systems Analysis Conference, WSDA.

Rossman, L. A., Boulos, P. F., Altman, T., 1993. Discrete volume-element method for network water-quality models. Journal of Water Resources Planning and Management, 119, 505-517.

Rossman, L.A., Clark, R., M., Grayman, W. M., 1994. Modeling chlorine residuals in drinking-water distribution systems. Journal of Environmental Engineering, 120, 803-820.

Rossman, L. A., Boulos, P. F., 1996. Numerical Methods for Modeling Water Quality in Distribution Systems: A Comparison. Journal of Water Resources Planning and Management, 122 (2), 137-146.

Rossman, L. A., 2000. EPANET Users Manual. National Risk Management Research Laboratory, Office of Research and Development, United States Environmental Protection Agency, Cincinnati, Ohio.

Sahoo, G. B., Ray, C., Wade, H. F., 2005. Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. Ecological Modelling. 183(1), 29-46.

Sahoo, G. B., Ray, C., Mehnert, E., Keefer, D. A., 2006. Application of artificial neural networks to assess pesticide contamination in shallow groundwater. Science of the Total Environment, 367(1), 234-251.

Seródes, J. B., Rodriguez, M. J., Ponton, A., 2001. Chlorcast□: a methodology for developing decision-making tools for chlorine disinfection control. Environmental Modelling & Software 16, 53-62.

Shang, F., Uber, J. G., Plycarpou, M. M., 2002. Particle backtracking algorithm for water distribution system analysis. Journal of Environmental Engineering, 128 (5),441-450.

Shang, F., Uber, J.G., Rossman, L.A., 2008. EPANET Multi-Species Extension User's Manual. National Risk Management Research Laboratory, Office of Research and Development, United States Environmental Protection Agency, Cincinnati, U.S.A..

Shen, H., McBean, E., 2012. False Negative/Positive Issues in Contaminant Source Identification for Water-Distribution Systems. Journal of Water Resources Planning and Management, 138(3), 230–236.

Singh, R. M., Datta, B., 2006. Artificial neural network modeling for identification of unknown pollution sources in groundwater with partially missing concentration observation data. Water Resources Management 21(3), 557-572.

Singh, V., Gupta, I., Gupta, H.O., 2007. ANN-based estimator for distillation using Levenberg–Marquardt approach. Engineering Applications of Artificial Intelligence, 20(2), 249-259.

Souliotis, M., Kalogirou, S., Tripanagnostopoulos, Y., 2009. Modelling of an ICS solar water heater using artificial neural networks and TRNSYS. Renewable Energy, 34, 1333–1339.

Sousa, S.I.V., Martins, F.G., Pereira, M.C., Alvim-Ferraz, M.C.M., 2006. Prediction of ozone concentrations in Oporto city with statistical approaches. Chemosphere, 64, 1141–1149.

Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., Pereira, M. C., 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environmental Modelling & Software, 22, 97-103.

Strafaci, A. et al., 2003. Advanced Water Distribution Modeling and Management. Haestad Methods, U.S.A..

Tao, T., Lu, Y. J., Fu, X., Xin, K. L., 2012. Identification of sources of pollution and contamination in water distribution networks based on pattern recognition. Journal of Zhejiang University SCIENCE A, 13(7), 559-570.

Tebelskis, J., 1995. Speech Recognition using Neural Networks. PhD Thesis.

Torres, J. M., Brumbelow, K., Guikema, S. D., 2009. Risk classification and uncertainty propagation for virtual water distribution systems, Reliability Engineering and System Safety, 94, 1259-1273.

van Bloemen Waanders, B. G., Bartlett, R.A., Laird, C.D., Biegler, L. T., 2003.Nonlinear Startegies for Source Detection of Municipal Water Networks. EWRI Conference, Philadelphia 2003.

Tryby, M. E., Propato, M., Ranjithan, S. R., 2010. Monitoring Design for Source Identification in Water Distribution Systems. Journal of Water Resources Planning and Management, 136 (6), 637-346.

Vankayala, P., Sankarasubramanian, A., Ranjithan, S. R., Mahinthakumar, G., 2009. Contaminant Source Identification in Water Distribution Networks Under Conditions of Demand Uncertainty. Environmental Forensics, 10(3), 253-263.

Willis, M. J., Montague, G.A., Di Massimo, C., Tham, M.T., Morris, A.J., 1992. Artificial neural networks in process estimation and control. Automatica, 28(6), 1181–1187.

Yang, Y. J., Goodrich, J. A., Clark, R. M., Li, S. Y., 2008. Modeling and testing of reactive contaminant transport in drinking water pipes: Chlorine response and implications. Water Research, 42(6-7), 1397-1412.

Yang, Y. J., Haught, R. C., Goodrich, J. A., 2009. Real-time contaminant detection and classification in a drinking water pipe using conventional water quality sensors: Techniques and experimental results. Journal of Environmental Management, 90(8), 2494–2506.

Zhang, L., Jian, J. H., Liu, P., Liang, Yi.Z., Yu, R. Q., 1997. Multivariate nonlinear modelling of fluorescence data by neural network with hidden node pruning algorithm, Analytica Chimica Acta, 344, 29-39.

Zechman, E. M., Ranjithan, S. R., 2009. Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. Journal of Water Resources Planning and Management, 135 (5), 334-343.

Zierolf, M. L., Plycarpou, M. M., Uber, J. G., 1998. Development and autocalibration of an input-output model of chlorine transport in drinking water distribution systems. IEEE Transactions on Control Systems Technology, 6 (4), 543-552.

http://www.advantica.biz/default.aspx?page=321 (accessed in February 2009).

http://www.bentley.com/en-US/Products/WaterGEMS (accessed in February 2009).

http://www.mwhsoft.com/page/p_product/net/net_overview.htm (accessed in February 2009).

http://porteau.irstea.fr/Index.html (accessed in February 2009).

http://www.safege.fr/fr/nos-metiers/logiciels/logiciels/ (accessed in February 2009).