MODEL-BASED LEAK LOCALIZATION IN SMALL WATER SUPPLY NETWORKS

CASE STUDY IN DMA LEIMUIDEN

J. MOORS

TU Delft
Delft University of Technology
oaseo
drinkwater

Model-based Leak Localization in Small Water Supply Networks

Case study in DMA Leimuiden

By

J. Moors

in partial fulfilment of the requirements for the degree of

Master of Science

in Civil Engineering

at the Delft University of Technology,

to be defended publicly on Thursday April 14, 2016 at 09:30 AM.

Student ID: 4024214

Graduation committee:

Prof. Dr. Ir. J. P. van der Hoek

Dr. L. Scholten

J. den Besten (Oasen)

Prof. Dr. Ir. N. van de Giesen

TU Delft

ACKNOWLEDGEMENTS

This research is the final work of my study watermanagement. I really enjoyed working on this topic. This thesis was not possible without the help of many others. First of all, I like to thank my graduation committee for their support and useful comments. Jan Peter van der Hoek for helping me in the search of a graduation topic. Special thanks go to my supervisors Lisa Scholten and Jurjen den Besten. I like to thank Lisa for her guidance in the scientific process and keeping me on track. It was also very special to give Lisa some 'Dutch language class' in the middle of the night while opening fire hydrants. I also like to thank Jurjen for his enthusiasm and believe in the leak localization from the beginning. These increased my own enthusiasm and believe in the project. He also learned me that everything is possible and especially that you should not be afraid to just try things and see what happens.

I also like to thank all my colleagues from Oasen. And especially many thanks to the mechanics who helped me in the field; Patrick van Beckum, Peter van Houwelingen, Jeroen Kegge and Kevin Holthuijsen. Special thanks to Mario Wildschut, my roommate at Oasen, for all his cups of tea, fruit talks, Christmas decoration and the leak localization trophy. During my research it was very nice to discuss with him the importance of model-based leak localization for practical applications.

And of course I also like to thank my friends and family. I really enjoyed my student time, which was not possible without them. Many of them helped me with writing the report and supported me to finish everything. Finally, I also thank Vincent for all his support the last years that actually already started with the help of choosing my master's program.

Delft, 5 April 2016

Janneke Moors

ABSTRACT

Small leaks in water supply networks often remain undiscovered, resulting in large amounts of lost water. Moreover, small leaks can grow larger over time and may result in pipe bursts, having negative consequences for the surroundings. If leaks are detected in an early stage, supply interruptions and damage due to flooding can be prevented. At this moment, District Metered Areas (DMAs) are used to detect leaks. A DMA is a small part of the network of which the inflow and outflow is monitored. A leak is mostly detected when the minimum night flow (MNF) increases. With labor-intensive step-testing and acoustic equipment, the leak is eventually localized. Despite the smaller size of a DMA compared to the whole network, localizing the leak is still a time-consuming job. An automatic leak localization method is required to decrease the search area and hence localize small leaks earlier.

In this research, the automatic leak localization method of Quevedo et al. (2011) is validated in DMA Leimuiden (the Netherlands). The leak localization method uses flow and pressure measurements at the inlet(s) of a DMA, pressure measurements inside the DMA and a hydraulic model to localize the leak. The method is based on measuring the additional head loss that is caused by the occurrence of a leak. It is recommended to perform the leak localization method in the night to avoid effect of unknown consumptions that influence the pressure. A prerequisite of the localization method is a detailed consumption distribution of the inflow for the hydraulic model. The goal of this research is to study the need for a detailed consumption distribution model in a DMA with a small MNF compared to the leak size (MNF: 4.5 m^3/h, leak size: 5.2 m^3/h, 7.5 m^3/h and 15 m^3/h). Hence, the research question is: *How is the performance of the model-based leak localization method, proposed by Quevedo et al. (2011)[1], influenced by the consumption distribution in a District Metered Area (DMA) with a relative small MNF compared to the leak size?* To study this influence, two consumption distribution models were compared. The 'uniform model' divides the consumption equally between all customer connections. The 'factorized model' uses a more accurate approach that divides the consumption proportionally to the billing information. The leak localization method was applied to eight artificial leaks that lasted 15 minutes and measurements of one day of a real leak (5.2 m^3/h). The artificial leaks were created by opening fire hydrants on three different locations at three different times (night, morning peak and daytime) with a leak size of 7.5 m^3/h and 15 m^3/h. The artificial leaks are localized with thirteen pressure measurement devices (six smart meters at customers and seven pressure loggers on fire hydrants). The localization of the real leak is only with the smart meters.

Leak localization results of the artificial leaks showed that the performance was the same with both consumption distribution models during the night. The consumption distribution does not influence the performance of the leak localization because this small night consumption flow had almost no effect on the measured pressures. The leak was localized well in two out of three artificial leak locations during the night. The third leak location, in the looped part of the network, had a poor performance. Here the leak resulted in a small head loss change at the sensors. During the morning peak, the leak caused a larger head loss and the leak localization with the factorized model was able to localize the leak. Results of the real leak showed a high fluctuation of leak localization performance every 15 minutes for both models. This might be caused by unexpected consumptions in the DMA. The performance became more robust (less fluctuation over the day) when the leak localization method of multiple time analysis steps (15 minutes) is combined to 1 hour. The accumulation of 24 leak localization outputs of one day resulted in localization of the leak 11 meters from the actual leak when executed with the factorized model and 35 meters with the uniform model.

[1] *Method based on correlation of measurements of pressure sensors in a water distribution network and data from a hydraulic model, proposed by Quevedo et al., (2011).*

To conclude, the leak localization method performs the same with both consumption models in case of low flow conditions and when leak localization results of the real leak for a whole day are combined. The performance of the leak localization method depends on the location of the leak. For some leak locations more flow in the system is required to create detectable head loss at the sensor locations. Uncertainties in the model cause larger pressure variations with higher flow conditions and a more detailed model must be used when there is more flow (morning peak). Too short measurement periods make the leak localization result sensitive to unexpected consumption inside the DMA. An accumulation of hourly results of a whole day makes the method more robust and gave satisfying performance irrespective of the used consumption models and with only 6 pressure sensors inside the network.

It is recommended to apply the leak localization method in more DMAs and evaluate the performance. In this way, more insight is gained in the performance depending on leak location and network configuration. When these leak localizations result in poor performances, more research is needed on the influence of improving the consumption distribution or the use of different time intervals to execute the method.

SAMENVATTING

Kleine lekken in drinkwaternetwerken blijven vaak onopgemerkt en zorgen daardoor voor grote lekverliezen. Daarnaast kunnen kleine lekken groter worden en overlast veroorzaken voor de omgeving. Als lekken in een vroeg stadium opgemerkt worden kan onderbreking van de levering aan klanten en schade voorkomen worden. Op dit moment worden onder andere *District Metered Areas* (DMAs) gebruikt om lekken te detecteren. Een DMA is een gedeelte van het netwerk waarbij al het instromende en uitstromende water gemonitord wordt. Lekken worden vaak gedetecteerd als een verhoging van het nachtverbruik gemeten wordt. De uiteindelijke leklokalisatie wordt uitgevoerd met behulp van arbeidsintensieve methodes zoals *step testing* en akoestisch leklokalisatie apparatuur. Ondanks dat een DMA kleiner is dan het gehele netwerk vergt de leklokalisatie veel tijd. Een automatische leklokalisatie methode is nodig om het zoekgebied kleiner te maken en kleine lekken sneller te lokaliseren.

In dit onderzoek is de automatische leklokalisatie methode van Quevedo et al. (2011) gevalideerd in DMA Leimuiden (Nederland). De leklokalisatie methode maakt gebruik van verbruik- en drukmetingen bij de ingaande leiding(en) van een DMA, drukmetingen binnen het DMA en een hydraulisch model van het DMA. De leklokalisatie methode is gebaseerd op het meten van drukval die veroorzaakt wordt door een lek. Het wordt aangeraden om de leklokalisatie toe te passen tijdens laag verbruik omdat dit zorgt voor zo min mogelijk ruis in de gemeten drukken. Een voorwaarde om de leklokalisatie methode toe te passen is een gedetailleerd consumptie verdeling van het verbruik in het model. Het doel van dit onderzoek is om de noodzaak van een gedetailleerd consumptie verdeling model te bestuderen wanneer het nachtverbruik van een DMA laag is ten opzichte van de lekgrootte (nachtverbruik: 4.5 m^3/h, lekgroottes: 5.2 m^3/h, 7.5 m^3/h en 15 m^3/h). Hieruit volgt de onderzoeksvraag: *Hoe is de prestatie van de leklokalisatie methode, voorgesteld door* Quevedo et al. (2011)[2]*, beïnvloed door de consumptie verdeling in een DMA met een relatief klein nachtverbruik in vergelijking tot de lekgrootte?* Om de invloed van consumptie verdeling te bestuderen zijn twee consumptie distributie modellen vergeleken. Het 'uniforme model' verdeeld het verbruik uniform tussen alle klantaansluitingen. Het 'gefactorizeerde' model gebruikt een realistischere insteek en verdeeld het verbruik proportioneel naar het jaarverbruik. De leklokalisatie methode is toegepast op acht kunstmatige lekken met een duur van 15 minuten en metingen van een echt lek voor 24 uur. De kunstmatige lekken zijn gemaakt door brandkranen open te zetten op drie verschillende locaties tijdens drie momenten van de dag (nacht, ochtendpiek en overdag) en met twee verschillende lekgroottes (7.5 m^3/h en 15 m^3/h). De kunstmatige lekke worden gelokaliseerd met behulp van dertien druk meters (zes slimme watermeters bij klanten thuis en zeven drukloggers op brandkranen). Het echte lek van 5.2 m^3/h wordt alleen met de zes slimme watermeters gelokaliseerd.

De leklokalisatie resultaten van de kunstmatige lekken gaven tijdens nachtmetingen bij beiden consumptie distributie modellen een gelijk beeld. De consumptie distributie had 's nachts nauwelijks invloed op de prestatie van de leklokalisatie omdat het kleine nachtverbruik haast geen effect heeft op de gemeten drukken. 's Nachts zijn de lekken in twee van de drie gevallen goed gelokaliseerd. Het derde lek kon slecht worden gelokaliseerd. Dit lek bevond zich in het vermaasde deel van het netwerk en het lek veroorzaakte een kleine drukval bij de druk sensoren. De drukval was groter tijdens de ochtendpiek en het gefactorizeerde model kon het lek toen wel lokaliseren. Resultaten voor elke tijdstap van 15 minuten van het echte lek laten gedurende een dag een grote fluctuatie in leklokalisatie prestatie zien. De prestatie werd robuuster (minder fluctuatie over de dag) wanneer meerdere tijdstappen werden gecombineerd tot 1 uur. De 24 leklokalisatie resultaten van een dag werden vervolgens gecombineerd tot een resultaat. Met het gecombineerde resultaat en het gefactorizeerde model werd het lek 11 meter van het daadwerkelijke lek gelokaliseerd en 35 meter met het uniforme model.

Geconcludeerd kan worden dat de leklokalisatie methode hetzelfde presteert voor beiden consumptie

[2] *Methode gebaseerd op de correlatie van drukmetingen in een drinkwaternetwerk en data van een hydraulisch model, voorgesteld door* Quevedo et al. (2011).

modellen in het geval van laag verbruik en wanneer de resultaten van een hele dag worden gecombineerd. De prestatie van de leklokalisatie methode hangt af van de leklocatie binnen het netwerk. Voor sommige leklocaties is meer stroming vereist om voldoende drukval te creëren op de locaties van de sensoren. Onzekerheden in het model veroorzaken een grotere variatie in druk en een gedetailleerder model is nodig in de ochtendpiek om deze stroming te verdelen. Korte meetperioden maken de leklokalisatie resultaten gevoelig voor onverwachte consumptie binnen het DMA. Het wordt daarom ook geadviseerd om metingen van een hele dag te gebruiken.

Het wordt aangeraden om de leklokalisatie methode in meer DMAs toe te passen en de prestatie te evalueren. Op deze manier wordt meer inzicht verworven in de afhankelijkheid van de leklocatie en netwerkstructuur op de prestatie. Wanneer de leklokalisatie prestatie van nieuwe lekken onvoldoende is, meer onderzoek is nodig op de invloed van het verbeteren van de consumptie distributie en het gebruik van verschillende tijdsintervallen om de leklokalisatie uit te voeren.

CONTENTS

ABBREVIATIONS

DMA = District Metered Area

DWN = Drink Water Network

Fact. = Factorized

FP = False Positive

FSM = Fault Signature Matrix

FSMCHC = column from Fault Signature Matrix (FSM) that results in the Highest Correlation

GA = Genetic Algorithm

GIS = Geographical Information System

LC = Leak Correlation

Loc. = Location

MNF = Minimum Night Flow

N/A = Not Applicable

NRW = Non-Revenue Water

RMSE = Root Mean Square Error

SCADA = Supervisory Control and Data Acquisition

Std. = Standard deviation

T_a = scenario analysis time step

T_s = sensor sample time

T_w = diagnosis sliding window

Unif. = Uniform

Δ =Change in…

1. INTRODUCTION

Part of the water distributed by a drinking water utility is unprofitable and is called non-revenue water (NRW). NRW can be divided in unbilled authorized consumption, apparent losses and real losses (Lambert & Hirner, 2000). Unbilled authorized consumption consists of e.g. firefighting, flushing of mains and sewers. Apparent losses contain unauthorized consumption like theft and are a result due to inaccuracies of water meters. Real losses are due to physical leaks on for instance transmission mains, distribution mains and service connections (Lambert & Hirner, 2000). The average percentage of non-revenue water (NRW) in the Netherlands is 5%. The expectation for the Netherlands is that the major part is due to real losses. Compared to other developed countries the amount of NRW is low (Vewin, 2013; Rosario-Ortiz et al., 2016). Among other reasons, this is due to low operating pressures, rapid reparation after a reported leak and good maintenance. Additionally, the Netherlands has a relative new infrastructure (post-war), compared to for example England (Ofwat, 2007). Several sources expect that deterioration of the network may cause a higher leakage frequency (Kleiner & Rajani, 2001; Tao et al., 2013).

Not all leaks are the same: a distinction can be made between bursts (high volume, short duration events), leaks (low volume, long duration events) and 'background' leaks (small continuous leakages) (Bakker, 2014). Consumers normally report bursts that cause floods. The leaks that do not cause water to come up at the surface, can continue unreported for a long time, resulting in large volumes of lost water. Additionally, moving ground particles around a small leak tend to enlarge a leak (Puust et al., 2010 and Appendix A). These leaks can in the end have the same consequences as large bursts (Mense, 2015). Earlier and automatic leak detection can save water and prevents small leaks turning into bursts.

In order to detect leaks automatically, measurements in the field are needed. The smaller the monitored area, the easier it is to detect a leak automatically. The network is divided into smaller areas, called 'District Metered Areas' (DMAs), since there is a relation between the size of leaks that can be detected automatically and the size of the monitored area (Bakker, 2014). It is already proven in the literature that it is possible to detect leaks in DMAs by making a simple water balance that compares expected demand and actual water use (Bakker, 2014; Romano et al., 2012). Additionally, an increase of the minimum night flow (MNF) can be used for the detection of new leaks (Farley et al., 2008; Oasen, 2015a). An increase in the MNF is often the result of a leak.

To reduce the search area within a DMA, step-testing is often applied. Step-testing is based on redirecting flows of sub areas within a DMA by systematically opening and closing valves. This technique can reduce the search area but is not suitable for all DMAs. Furthermore, several steps are needed that are carried out during multiple nights (Puust et al., 2010). Reducing the search area does not solve the problem of finding the exact location of a leak and therefore, apart from leak detection, an additional leak localization technique is needed to localize the leak (Bakker, 2014). Leak localization is frequently done with acoustic equipment such as listening rods, leak correlators, leak noise loggers and non-acoustic methods such as gas injection, ground penetrating radar technology and infrared photography (Li et al., 2015). These methods are very accurate but it takes a long time to find a leak (Puust et al., 2010; Li et al., 2015). Additionally, acoustic equipment is less effective for new pipe materials like PVC. Noise is transmitted less far in these materials, resulting in an increased searching time (Gao et al., 2005; Li et al., 2015). Even with small DMAs, leak localization results in a time-consuming and labor-intensive process.

To reduce the search area and make the leak localization process less time consuming, software-based methods are needed next to the hardware based methods. These can be divided in non-numerical modeling methods and numerical modeling methods. A leak is causing more flow in a pipe. This results in more friction and consequently a larger head loss and different pressures within the network, compared to a situation without a leak (Pérez et al., 2011b). Non-numerical models use Artificial Intelligence techniques which need historical pressure data to train them (Li et al., 2015). The drawback of these techniques is that large amounts of training data are not

always available (yet) (Li et al., 2015; Tao et al., 2013).The numerical modeling methods use a model and compare simulated results with field data (Li et al., 2015). One of these methods is based on pressure transients in pipes. After the appearance of a leak, the pressure wave can accurately be localized by the use of sensors with high sampling intervals. This method is very sensitive to the exact configuration of the whole network, which is often unknown in real systems. Practical application is mainly limited to single large pipelines (Puust et al., 2010; Li et al., 2015). A model-based leak localization method that proved success in practice is the model-based leak localization of Quevedo et al. (2011) (Mirats-Tur et al., 2014). This method uses inflow and pressure at the DMA inlet(s), pressure sensors inside the DMA and a hydraulic model (Quevedo et al., 2011).

Automatic leak detection and localization techniques can reduce the area of interest to make it easier for traditional district audits to find the leak (Puust et al., 2010). Besides large volumes of water, this can save time and money.

1.1 PROBLEM STATEMENT

In 1992, Pudar and Ligget already raised the question whether leaks can be located with pressure measurements and introduced the pressure sensitivity matrix. The change in pressure after a leak at every location in the network is stored in this sensitivity matrix. Pudar & Liggett (1992) stated that at least in theory, leak localization is possible by analyzing the sensitivity matrix. In practice the success depends on the network configuration the accuracy of network characteristics (e.g. pipe roughness) and the accuracy of estimated demand patterns (Pudar & Liggett, 1992).

The leak localization method of Quevedo et al. (2011) uses the sensitivity matrix (Pudar & Liggett, 1992) and is based on the standard theory of model-based fault diagnosis (Gertler, 1998). Model-based leak localization seeks for pressure anomalies between field measurements and simulated values from a hydraulic model in a 'leak-free situation'. Fault isolation is used to identify the most probable location of the leak (Pérez et al., 2011b). For the leak localization method, measured pressures are compared with expected simulated pressures to create a list of residuals and localize the leak. In practice, residuals are never zero due to demand uncertainties in the model. In Pérez et al. (2009) the residuals are binarised by the use of a threshold. By binarising the residuals information is lost and Quevedo et al. (2011) improved this method by directly using the residuals.

The model-based leak localization method of Quevedo et al. (2011) is validated for three leaks in the DMA Nova Icària in Barcelona (Mirats-Tur et al., 2014). The leaks were located within a circle of 150 meter radius (Mirats-Tur et al., 2014). The MNF of DMA Nova Icària is approximately 75.6 m^3/h (Casillas et al., 2015). The MNF is important since it is recommended to localize the leak in night hours when there is less consumption and less noise in the pressure data (Pérez et al., 2009). During nighttime most consumers are inactive, with the exception of some toilet flushing, washing machines, etc. (Farley et al., 2008).

One of the limiting factors of the model-based leak localization method is the need for accurate estimates of the spatial consumption distribution by customers (Mirats-Tur et al., 2014; Meseguer et al., 2014). An accurate consumption distribution is important because this influences the pressure variations in the model and therefore the performance of the leak localization method. Leimuiden is a DMA in the Netherlands, where additional water meters are installed within the DMA to apply automatic leak localization (Oasen, 2015b). The MNF of DMA Leimuiden is only 6% of the MNF of DMA Nova Icària (4.5 m^3/h, 75.6 m^3/h MNF of DMA Leimuiden, Nova Icària respectively). The smallest localized leak in DMA Nova Icària (14.4 m3/h) is three times the amount of the MNF of Leimuiden. The relatively large leak flow is flowing to one location and the rest (normal MNF) is distributed to all customers. It is expected that this low consumption flow does not influence the pressures much and therefore it is questioned if an accurate consumption distribution is needed when the leak is large compared to the MNF.

The research question is:

How is the performance of the model-based leak localization method, proposed by Quevedo et al. (2011)[3], influenced by the consumption distribution in a District Metered Area (DMA) with a relative small minimum night flow (MNF) compared to the leak size?

The sub questions are:

- How can the performance of the leak localization method be defined?
- Is it possible to apply the model-based leak localization method in DMA Leimuiden?
- How does the time of the day influence the performance of the leak localization method?
- How does the location of a leak within a network influence the performance of the leak localization method?

1.2 RESEARCH GOAL AND SCOPE

The goal of this research is to validate the leak-localization method in a DMA with a low MNF and study the need for a detailed consumption model as an input to the model-based leak localization. It is important to know whether a detailed consumption distribution is still necessary when the magnitude of the leak is high compared to the total flow. The assumption is that it is easier to observe a deviation between the simulated pressures and the pressure measurements if the magnitude of a leak is high compared to the total inflow of the DMA.

To study the influence of the consumption distribution on the performance of the leak localization method two simple consumption distribution models are developed. The simplest approach is the uniform model. The uniform model distributes the consumption inflow equally between all customers. The second model is the factorized model. This model divides the consumption in the same way as proposed in Quevedo et al. (2011): the customers get a factor relative to the overall consumption based on the billing information. Hereby, the assumption is that a customer who uses more water in one year, also uses more water at every time-step. For the factorized model the billing information of 2014 is used. The factorized model is still a simple approach where all customers get the same consumption pattern. In the leak localization in DMA Nova Icària different consumption patterns are used based on the type of water usage (i.e. domestic, industrial) (Meseguer et al., 2014). Another possible improvement is the use of telemetry data (Pérez et al., 2011a).

Next to the need of an accurate consumption distribution, this research aims to create insight in the performance of the leak localization method depending on location of the leak in the network and time of the day the leak localization is executed. Results can be used to improve practical application of the leak localization method.

1.3 STUDY AREA: DMA LEIMUIDEN

The leak localization method is applied in DMA Leimuiden (Figure 1.1). This DMA consist of the village Leimuiden and is located in the Mid-West of the Netherlands. The responsible drinking water utility of this area is Oasen. DMA Leimuiden is located in the most Northern part of their service area.

[3] *Method based on correlation of measurements of pressure sensors in a water distribution network and data from a hydraulic model, proposed by Quevedo et al., 2011*

Leimuiden is a small village of 1.46 km^2 with 4,285 inhabitants (1835 households) (CBS, 2015). The network length is 26 kilometers and it consists mainly of PVC (54%). The average age is 35 years. During the research, the water of DMA Leimuiden is supplied by one inlet (green arrow in Figure 1.1). The delivered monthly volume is 20,000 m^3 to 1932 customers.

Artificial leaks are created by opening fire hydrants to apply the leak localization method. A few weeks after the artificial leaks, a real leak was present in DMA Leimuiden. Measurements during the real leak are used as well to apply the leak localization method.

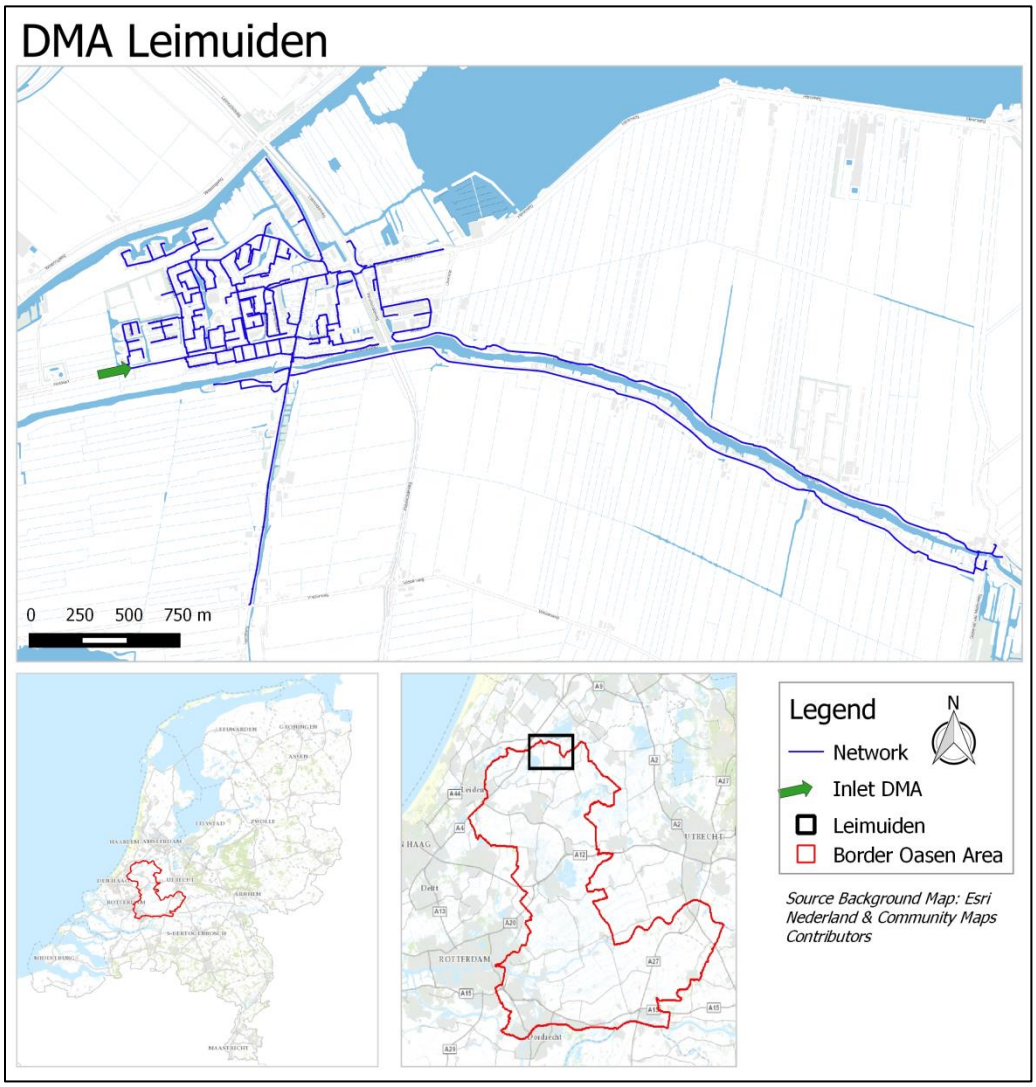

Figure 1.1 Case study location: DMA Leimuiden in the service area of Oasen, the Netherlands

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 SOFTWARE

The model-based leak localization requires a hydraulic model. In this research the hydraulic model EPANET 2.0 (EPANET) is used. EPANET is public domain software, developed by the U.S. Environmental Protection Agency. It can perform extended period simulation of hydraulic behavior within pressurized pipe networks (Rossman, 2000). EPANET is widely applied in practice by water supply companies (EPA, 2005) and in scientific research (e.g. Hatchett et al., 2010; Arunkumar & Nethaji Mariappan, 2011; Farina et al., 2014).

EPANET is less suitable to set-up the model from the GIS database of Oasen (DiaGIS). Therefore, during the set-up of the model, the commercial hydraulic model InfoWorks WS (InfoWorks) is used (see paragraph 2.2.3). After the set-up of the model, the leak localization method is applied in EPANET. The advantage of EPANET is that it can be integrated with other applications by using the EPANET Programmer's Toolkit. This integration is needed in order to automatically perform many network analyses that are desired for the model calibration and leak localization method.

The integration with the EPANET Programmer's Toolkit is done with the programming language Python. This is a freely available, general-purpose, high-level programming language (Python, 2016). For this research the Scientific PYthon Development EnviRonment (Spyder) is used and Python version 2.7.

2.1.2 MEASUREMENT DEVICES

The measurements needed for the leak localization method are flow and pressure measurements at the inlet of the DMA and pressure measurements inside the DMA (inner pressure sensors).

2.1.2.1 DMA INLET METER

At the inlet of the DMA, flow and pressure measurements are needed to fix the boundary conditions when the model is simulated (Sanz & Pérez, 2014). The meter at the inlet of DMA Leimuiden is installed in 2014 (Figure 2.1). The meter consists of a flow meter and a pressure meter. The flow is logged every 15 minutes and data of one day is sent digitally in the subsequent night. The flow meter in the DMA meter has a measurement uncertainty of 0.1%. The resolution of the pressure measurement device is 0.1 bar (=1.0197 mWc). For the leak localization method this is too insensitive. Therefore, a 'Tee piece' is attached to one inlet to install a temporary pressure logger (specifications in next paragraph: 2.1.2.2).

Figure 2.1 Example DMA inlet meter

2.1.2.2 PRESSURE LOGGER

The pressure loggers that are used are the PrimeLog+. These can be attached to a fire hydrant (Figure 2.2). The measurements are stored on an internal memory that has to be read out manually. The uncertainty is 0.1% and pressure is measured every second. The resolution of this device is not specified but based on measurements it is 0.002 mWc.

Figure 2.2 Pressure logger attached to a fire hydrant

2.1.2.3 SMART METER

The smart meter consists of three measurement devices that can be attached to an installed in front of an existing water meter of a customer (Figure 2.3). This smart meter can measure temperature and pressure of the water and water flow into the house. This information is send to Oasen, using the internet connection of the customer. The advantage is that the information is directly available and the smart meter costs 200 euros compared to a fixed meter of 7000-8000 euros. The disadvantage is the dependency on the willingness of the customer to participate.

For this research, only the pressure measurements of the smart meters are used when the customer does not use water. As long as the customer does not use water, the pressure at the house is the same as the pressure in the network. The pressure device used is a pressure transmitter named JUMO MIDAS C08 and has an uncertainty of 0.35%. The pressure is measured every five seconds. The resolution of this device is not specified but is reported with a resolution of 0.01 bar (=0.10197 mWc).

Figure 2.3 Smart meter attached to and installed in front of the normal water meter at a household

2.2 METHODS

To execute the leak localization method, pressure measurements in the network and simulated pressures from a calibrated hydraulic model are compared by means of correlation (green, yellow and blue square in Figure 2.4). In order to mathematically represent the real network of DMA Leimuiden by a model, it is built with links (representing pipes) and nodes (consumption points). The model is used to model a leak at every node in the network, resulting in different pressures compared to the situation without a leak. The nodes with a high correlation between the pressure measurements and pressure simulations, indicate the area of the leak.

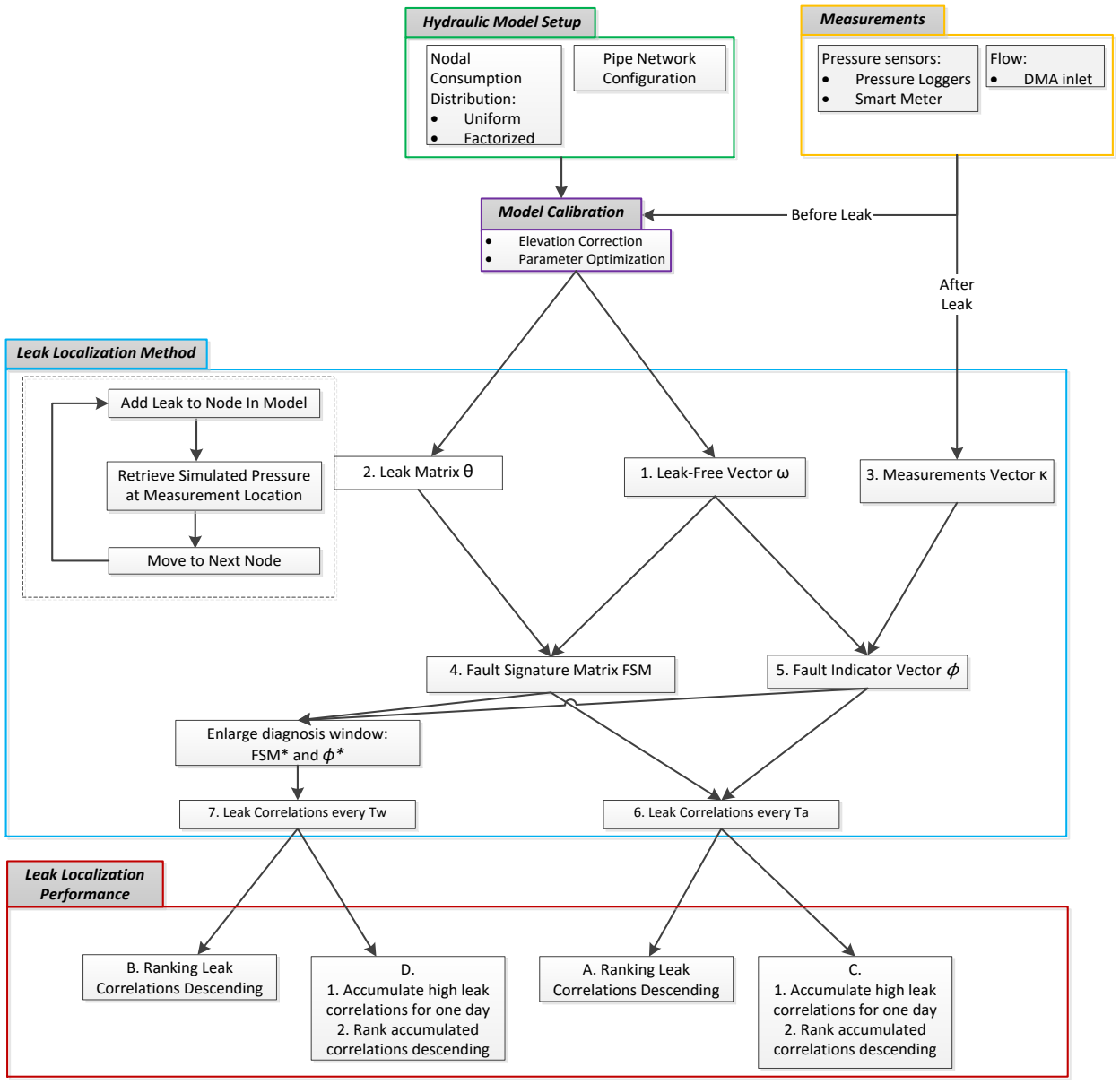

Figure 2.4 Overview leak localization method

2.2.1 LEAK LOCALIZATION

The steps of the leak localization method are described below. The numbers and letters of the steps, correspond to the blocks inside the blue and red square in Figure 2.4. The numbers are related to the leak localization method (blue square). The letters correspond to the result of the leak localization method (red square). More details of the leak localization method can be found in Quevedo et al. (2011) and Meseguer et al. (2014).

The leak localization method is time dependent and bounded by the flow and pressure measured at the inlet of the DMA. To execute the leak localization method three different time steps are needed (Meseguer et al., 2014). These are the sensor sample time (T_s), scenario analysis time step (T_a) and the diagnosis sliding window (T_w). T_s is the sample time of the DMA instrumentation. T_a is the time step used to compute a leak localization result (step 1-6 in Figure 2.4). T_a needs to be smaller or equal to T_s. When T_a is smaller than T_s, the sensor measurements are averaged. This reduces the effect of 'nuisance noise' affecting the measurements (Meseguer et al., 2014). When T_w is larger than T_a the effect of one time analysis step that is affected by anomalies, is reduced and the leak localization method gets more robust (Meseguer et al., 2014).

In this research the smallest possible T_a is 15 minutes due to the flow measurements. The cumulative flow is stored every 15 minutes. The pressure measurements are sampled every second (pressure loggers) and every 5 seconds (smart meters). Both are averaged to 15 minutes. The artificial leaks last for 15 minutes and T_a is T_w. The measurements of the real leak are available for three days. Again the T_a is 15 minutes and T_w is set to 1 hour ($4*T_a$).

Leak localization method T_a:

1. First, the leak-free vector ω is computed during a leak-free scenario at time instant *t*. The leak-free vector contains of pressures ($\hat{p}_0$), at the sensor-nodes, as if there is no leak in the DMA. The size of the leak-free vector depends on the number of pressure sensors inside the network (*ns*).

$$\boldsymbol{\omega}(t) = \begin{bmatrix} \hat{\boldsymbol{p}}_{1_1}(t) \\ \vdots \\ \hat{\boldsymbol{p}}_{1_{ns}}(t) \end{bmatrix} \quad (1)$$

2. Subsequently, the model is used to one by one simulate a leak at the nodes (*np)* in the model. The leak is simulated as additional consumption. After simulation of a leak in one node, the modelled pressure ($\hat{p}$) at each sensor-node (ns) is stored in the leak matrix θ. The number of rows depends on the number of pressure sensors inside the network (*ns*). The number of columns depends on the number of nodes (*nn*).

$$\boldsymbol{\theta}(t) = \begin{bmatrix} \hat{\boldsymbol{p}}_{1_1}(t) & \cdots & \hat{\boldsymbol{p}}_{nn_1}(t) \\ \vdots & \ddots & \vdots \\ \hat{\boldsymbol{p}}_{1_{ns}}(t) & \cdots & \hat{\boldsymbol{p}}_{nn_{ns}}(t) \end{bmatrix} \quad (2)$$

3. The third value for the sensor-nodes consists of the pressure measurements. The average pressure for one time analysis step is stored in the measurements vector κ.

$$\boldsymbol{\kappa}(t) = \begin{bmatrix} \boldsymbol{p}_1(t) \\ \vdots \\ \boldsymbol{p}_{ns}(t) \end{bmatrix} \quad (3)$$

4. The leak-free vector ω and the leak-matrix θ are used to compute the theoretical fault signature matrix FSM ϵ $\Re^{ns\ x\ np}$. The FSM contains the modelled differences between the leak-free pressures ($\hat{p}_0$) and the pressures ($\hat{p}$) after a modelled leak.

$$\mathbf{FSM}(t) = \begin{bmatrix} \hat{p}_{1_1}(t) - \hat{p}_{10}(t) & \cdots & \hat{p}_{nn_1}(t) - \hat{p}_{10}(t) \\ \vdots & \ddots & \vdots \\ \hat{p}_{1_{ns}}(t) - \hat{p}_{ns0}(t) & \cdots & \hat{p}_{nn_{ns}}(t) - \hat{p}_{ns0}(t) \end{bmatrix} \quad (4)$$

5. The fault indicator vector ϕ contains residuals and is obtained by subtracting the measurements, p ϵ $\Re^{ns}$, from their model-estimates in a leak free scenario, $\hat{p}$ ϵ $\Re^{ns}$, on time instant t.

$$\boldsymbol{\phi}(t) = \begin{bmatrix} p_1(t) - \hat{p}_{10}(t) \\ \vdots \\ p_{ns}(t) - \hat{p}_{ns0}(t) \end{bmatrix} \quad (5)$$

6. Subsequently, the leak correlation is calculated by taking the Pearson's correlation coefficient[4], $\rho_{\varphi,FSM_j}(t)$. All the columns of the FSM are correlated with the fault indicator vector, resulting in ρ_{φ,FSM_j}. The maximum value in this vector represents the most probable node to have a leak.

$$\boldsymbol{LC} = \max_{j} \left(\rho_{\varphi,FSM_j}(t) \right), \ j = 1, \dots, np \quad (6)$$

A. By sorting the leak correlations descending, a ranking of the nodes with the most probable leak locations is created.

Leak localization method when $T_w = \tau T_a$ where $\tau \in \mathbb{N}$ and $\tau > 1$:

7. The $\phi^*(t)$ and $FSM^*(t)$ are created by concatenating the $\phi^*(t)$ and $FSM^*(t)$ of the time analysis steps. $\phi^*(t)$ and $FSM^*(t)$ are used to calculate the leak correlations with the person correlation.

$$\boldsymbol{\phi}^*(t) = \begin{bmatrix} \boldsymbol{\phi}(t) \\ \vdots \\ \boldsymbol{\phi}(t-\tau) \end{bmatrix} \quad (7)$$

$$\mathbf{FSM}^*(\mathbf{t}) = \begin{bmatrix} \mathbf{FSM(t)} \\ \vdots \\ \mathbf{FSM(t-\tau)} \end{bmatrix} \quad (8)$$

$$\boldsymbol{LC} = \max_{j} \left(\rho_{\varphi^*,FSM_j^*}(t) \right), \ j = 1, \dots, np \quad (9)$$

B. By sorting the leak correlations of every T_w descending, a ranking of the nodes with the most probable leak locations is created.

---

[4] Pearson's correlation coefficient $\rho_{x_i x_j}$ between two variables x_i, x_j may be defined as: $\rho_{x_i x_j} = cov(\boldsymbol{x_i}, \boldsymbol{x_j}) / \sqrt{cov(\boldsymbol{x_i}, \boldsymbol{x_i}) cov(\boldsymbol{x_j}, \boldsymbol{x_j})}$, where $cov(\boldsymbol{a}, \boldsymbol{b}) = E\left[(\boldsymbol{a} - \bar{a})(\boldsymbol{b} - \bar{b})\right]$ is the covariance function between two variables **a** and **b**, being $\bar{a} = E(\boldsymbol{a})$ and $\bar{b} = E(\boldsymbol{b})$, respectively.

The leak localization method has a result every T_a or T_w. To have one result for a whole day they can be accumulated. Only correlations above 0.5 are considered (Pérez et al. (2014a).

C. An accumulated leak correlation is created by accumulating the high leak correlations (Ic>0.5) of very T_a for a whole day for each node. By sorting the accumulated leak correlations of every T_a descending, a ranking of the nodes with the most probable leak locations is created.
D. An accumulated leak correlation is created by accumulating the high leak correlations (Ic>0.5) of very T_w for a whole day for each node. By sorting the accumulated leak correlations of every T_w descending, a ranking of the nodes with the most probable leak locations is created.

2.2.2 LEAK LOCALIZATION PERFORMANCE

The result of every leak localization is a list of all nodes and their corresponding (accumulated) leak correlation. One approach to judge the performance of the leak localization method is to look at the distance between the node with the highest leak correlation and the actual leak (Pérez et al., 2011c). In the perfect situation (all flows in the model are correct), the node with the highest leak correlation is the location of the actual leak. Figure 2.5 shows an example where the distance between the leak and the node with the highest leak correlation is 414 meter (distance when the pipes are followed). This is relatively close to the leak location. Although without the leak correlations of the rest of the nodes, it is hard to define the search area. The model-based leak localization method is meant to find a leak area, rather than the precise location of the leak (Quevedo et al., 2011). The distance to the highest leak correlation can be small resulting in a good performance while the other high correlating nodes can be in the opposite direction of the leak (red nodes in Figure 2.5).

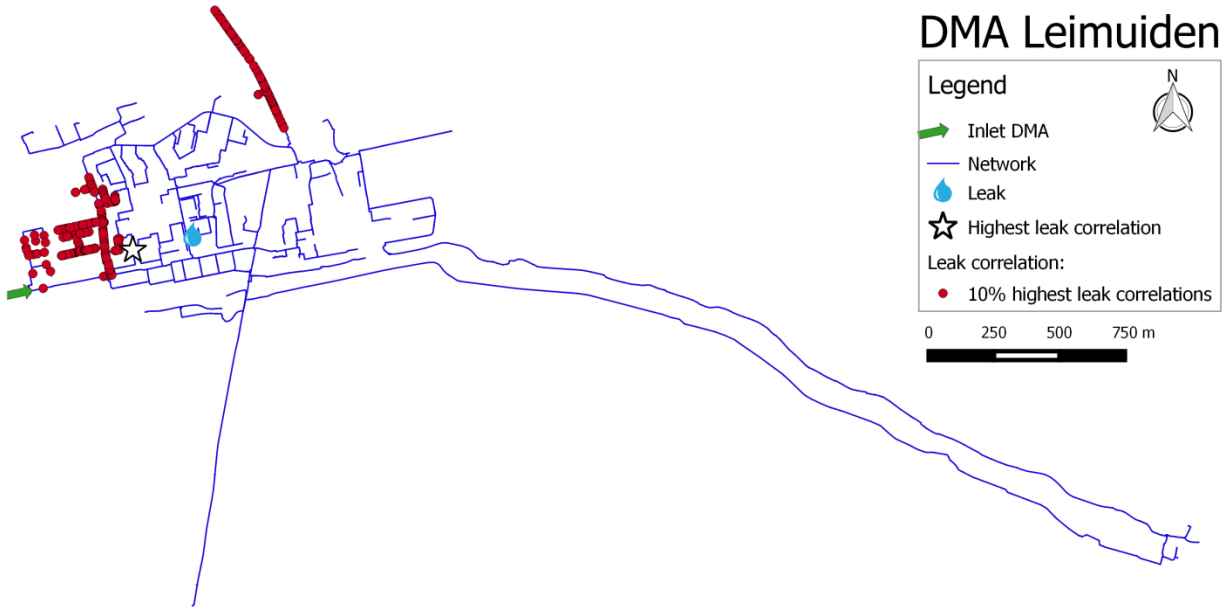

Figure 2.5 Leak localization result example. The star represents the node with the highest leak correlation. The red nodes represent the highest 10% leak correlations.

In Pérez et al. (2014b) the mean distance to the gravity center of those nodes with correlations over 99% of the maximum correlation is used. In this way the result of the method is more robust than only looking at the

maximum correlation node (Pérez et al., 2014b). The disadvantage is that the nodes that have high correlations can be spread out in the network, resulting in a gravity center accidently close to the leak (Figure 2.6).

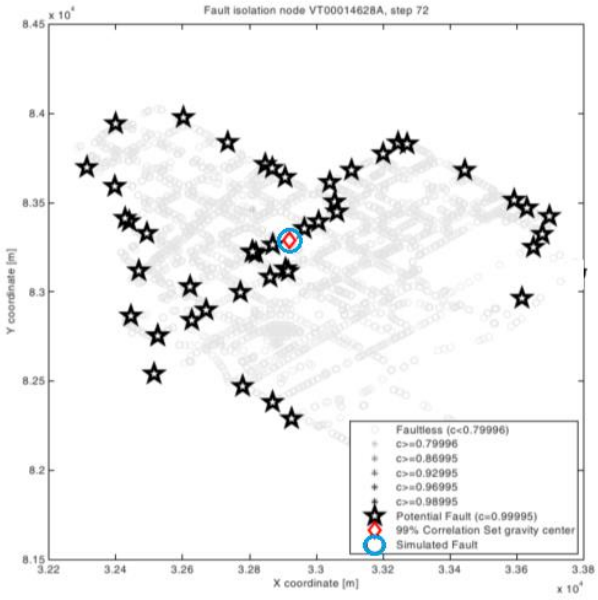

Figure 2.6 Fictive example of gravity center accidently close to leak. The stars represent the nodes with a high leak correlation. The red diamond is the 99% correlation set gravity center and the blue circle the 'real leak'. Figure is taken from Pérez et al. (2014b) and the location of the 'real leak' (simulated fault) is adapted. This is done to show the possibility of having potential faults spread out over the DMA with the 99% correlation set gravity center accidently close to the 'real leak'.

In this research the performance of the leak localization with the two consumption distribution models is compared by the percentage of false positive (FP) nodes. These nodes have a higher leak correlation value than the node in the model where the actual leak was. In practice, this group of FP nodes should be as small as possible since these are locations that are searched first without finding the leak. Instead of comparing the exact percentages of FP nodes, the results are divided in three classes. If the percentage of FP nodes is underneath 5%, the performance is classified as good, between 5 and 15% as reasonable and above 15% as bad. The class-division makes the comparison of the results fairer because the amounts of nodes are not evenly distributed over the area (Appendix E). The same amount of nodes can result in a different surface area in the network depending on the location in the network.

2.2.3 HYDRAULIC MODEL SETUP DMA LEIMUIDEN

In order to build the consumption distribution models, structural and geometrical date are needed. Structural data contains information about the position of pipes and location of household connections. These are complemented with geometrical data such as diameter and roughness of the pipes. In order to perform calculations operational data and model parameters are needed. Operational data are the pressures at the inlet of the DMA and the consumption by the customers. These are the boundary conditions. Model parameters are needed to perform the hydraulic calculations and are for example the type of head loss formula, specific gravity and accuracy criteria for the convergence of equations that govern the network hydraulics. The consumption by the customers is the only difference for the two consumption distribution models.

2.2.3.1 PIPE NETWORK CONFIGURATION

The configuration of the pipe network, including structural and geometrical data, is exported from a geographical information system (DiaGIS) to an ESRI shapefile and used to set-up the model in InfoWorks. In the exported shapefile, all pipes are represented, including the pipes that are out of order or removed. Only the pipes of DMA Leimuiden with status 'In use', are selected and used for the model set-up. All pipes with length 0 are removed and it is checked if everything is connected to each other.

Both structural data and geometrical network data can be directly imported in InfoWorks. For the model set-up; pipe diameter, pipe material and pipe age are imported together with the structural data of the pipe. Regarding the diameter and roughness, additional processing is needed to determine the correct inner diameters and Darcy-Weisbach roughness coefficient based on pipe age and pipe material. Details are described in Appendix B and Appendix C.

A node is automatically created at the junction of two different pipes or at valves, fire hydrants or customer connections. As long as there is no consumption by (a) customer(s) at this point (node), nothing happens and the water simply flows through.

To account for consumption by customers, 'customer points' are added in InfoWorks. The locations of the customer points are obtained from the 'basic administration of addresses and buildings' ('Basisregistraties Adressen en Gebouwen', Kadaster, 2015). The 'customer points' (black squares) and related nodes (green dots) are shown in Figure 2.7. The customer connections have a base consumption factor that is used to distribute the water in the DMA (uniform model: Equation 12, factorized model: Equation 13 in paragraph 2.2.3.2). The 'customer points' with consumption factor are allocated to the nearest node since exact house pipe connections are not in the model. Multiple houses can be connected to one node, resulting in an aggregated consumption for that node. These models are exported to Epanet.

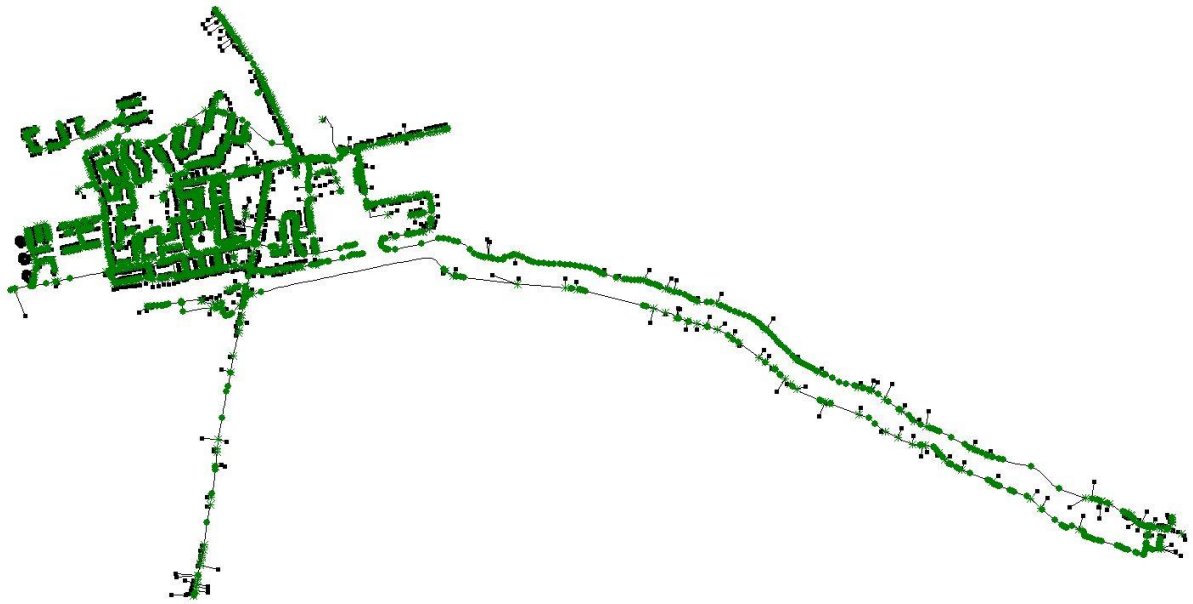

Figure 2.7 Customer connections in DMA Leimuiden. Connections between customer points (black squares) and nodes (green dots) in InfoWorks.

2.2.3.2 NODAL CONSUMPTION DISTRIBUTION

In Epanet, the DMA inlet is modelled as a reservoir. The measured pressures at the DMA inlet are used as water levels of the reservoir. The supply from the reservoir depends on the customer consumption in the network. The measured consumption (Figure 2.8: left) is used to create the consumption pattern in Epanet (Figure 2.8: right):

$$\boldsymbol{Consumption\ Pattern}\ (t) = \frac{Q(t)}{3.6} / Ncust \qquad (10)$$

The measured inflow of the DMA is distributed among the 1891 customer connections and is different due to the base factor for the uniform model and the factorized model. For every time instant the consumption pattern is multiplied with the base consumption factor.

$$\boldsymbol{Consumption\ Node}\ (t) = \boldsymbol{Consumption\ Pattern}(t) * \boldsymbol{Base\ Consumption\ Factor} \qquad (91)$$

The nodes without consumption have a base consumption factor of 0 (e.g. nodes at the junction of two different pipes or fire hydrants). The base consumption factor is different for the two consumption distribution models:

$$\boldsymbol{Base\ Factor\ Uniform\ Consumption\ Model} = 1 \qquad (102)$$

or

$$\boldsymbol{Base\ Factor\ Factorized\ Consumption\ Model} = \frac{\boldsymbol{Billed\ Consumption\ Node\ 2014}}{\boldsymbol{Average\ Billed\ Consumption\ All\ Nodes\ 2014}} \qquad (113)$$

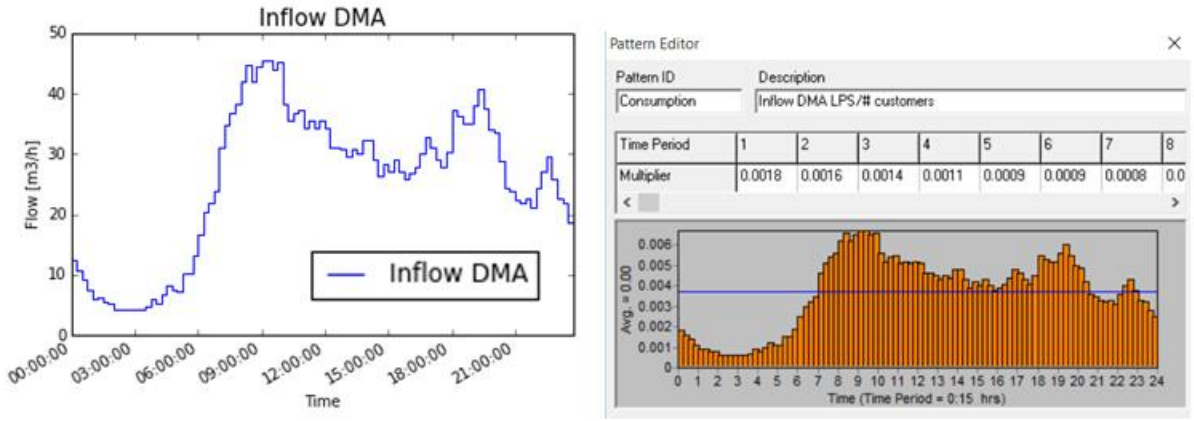

Figure 2.8 Consumption pattern in Epanet. Left: measured inflow of DMA Leimuiden during one day. Right: measurements converted to l/s/customer and set as consumption pattern in Epanet.

2.2.3.3 MODEL PARAMETERS

The default values of EPANET are used for all model parameters (e.g. specific gravity, relative viscosity and max. number of trials) except for the formula to calculate the head loss and are described in the EPANET 2 User's Manual (Rossman, 2000). Head loss in the network is caused by friction and local losses. In this research the Darcy-Weisbach formula is used to calculate the head loss caused by friction (Equation 14, page 14). Local losses can be calculated with a minor loss coefficient. Roughness and minor loss coefficient are both unsure for the network and need to be calibrated. It is not possible to distinguish between the parameters because they both affect the head loss. In this research only roughness is taken into account. It is expected that the roughness values become higher

than the theoretical values since the local losses are also included in this Darcy-Weisbach roughness coefficient. The formula used for head loss due to friction is:

$$h_l\,(friction) = 0.0252\frac{f(\varepsilon,d,q)}{d^5}Lq^2 \qquad (14)$$

With:

h_l = head loss
f = friction factor (dependent on ε, d and q) (Rossman, 2000)
ε = Darcy-Weisbach roughness coefficient
d = pipe diameter
q = flow rate
L = pipe length

Another default setting is to continue if the system gets unbalanced. 'A System Unbalanced condition can occur when EPANET cannot converge to a hydraulic solution in some time period within its allowed maximum number of trials. This situation can occur when valves, pumps, or pipelines keep switching their status from one trial to the next as the search for a hydraulic solution proceeds (Rossman, 2000)'. In the consumption models there are no pumps or valves present but the system can get unbalanced with certain values of the Darcy-Weisbach roughness coefficient. If the default setting of 'continue' is used, another 10 trials with no links status changes allowed are executed in an attempt to achieve convergence (Rossman, 2000). The default setting is used for the sensor deployment optimization and for the leak localizations. The calibration of the models to execute the leak localization is performed with a model that stops if the system gets unbalanced.

2.2.4 MODEL CALIBRATION

One of the prerequisites to perform the leak localization is a well-calibrated model (Meseguer et al., 2014). During the calibration process, observed pressures and simulated pressures are compared. Subsequently, model-parameters are adjusted to minimize the differences during calibration (Clemens, 2001). In this research only the model parameter Darcy-Weisbach roughness coefficient is adjusted.

Before the consumption distribution models are calibrated, the raw data from the field measurements is checked visually on completeness, min-max and drifts in the measurements. This is needed in order to remove erroneous measurements caused by instruments or other unforeseeable factors.

2.2.4.1 ELEVATION CORRECTION

The pressure measurements are compared with the simulated pressures at the locations of sensors in the model. In the model all node elevations are set to 0 while at the measurement locations there are elevation differences up to 4 meter (van der Zon, 2011). One way to compare the measurements and the modelled results is to compare the heads and use the ground levels and the elevations of the sensors below ground level for the elevation of the sensor. This information can be retrieved by measurements at the sensor locations but contains errors up to 0.5 meter. To work around these large errors in elevation, the elevations were calculated using the uncalibrated model (before parameter-optimization) during minimum consumption. During the night there is almost no flow and the influence of the roughness of the pipe is small. Most pipes in DMA Leimuiden have laminar flow during the night (Figure K.6). Epanet uses the Hagen-Poisseuille formula to calculate the friction factor. This factor does not depend on the Darcy-Weisbach roughness coefficient (Appendix D and Rossman, 2000). The measurements of the pressure at the inlet are used as a boundary pressure for the model. The difference between the pressure measurements

and the modelled pressures during low consumption are used to set the elevation of the nodes that have a pressure sensor in the real network.

2.2.4.2 PARAMETER OPTIMIZATION

Pipe roughness coefficients and nodal consumptions are often used for parameter optimization to calibrate the model. The consumption by every household in real-time is unknown and is the most uncertain parameter in this model due to constant variability (Sanz & Pérez, 2014). The consumption factor is kept constant in both consumption distribution models because the purpose of this study is to analyze the difference in performance of the leak localization method with the uniform and factorized consumption distribution. The Darcy-Weisbach roughness coefficient (pipe roughness) is used for the parameter optimization.

An extended period analysis of one day is used to calibrate for a range of different times. Least-squares are used to objectively judge the residuals between the measured pressures and the modelled pressures (Clemens, 2001). The objective is to minimize the residuals by adjusting the roughness. The objective function for one day of 96 measurements (every 15 minutes for 24 hours) is formulated as:

$$\boldsymbol{min} \sum_{t=1}^{96} \sum_{n=1}^{ns} (\boldsymbol{p_t} - \boldsymbol{\hat{p}_t})^2 \qquad (15)$$

With ns: total number of sensors, p_t: measured pressure and $\hat{p}_t$:simulated pressure on time instant t.

To minimize the least-squares the roughness is adapted in two manners. The first manner is multiplying the first estimated roughness that is based on material and age of the pipe (Appendix B) by a multiplication factor. Pipes with different materials and age still have a different roughness but can only change with the same multiplication factor. In this way the final roughness is dependent on the first estimate. The roughness is expected to be higher than the first estimate since local losses are not taken into account. The second manner uses one roughness for all pipes, irrespective of the material or age and is called the system roughness. With the second manner the roughness is not dependent on a first estimate but is one value for all pipes in the network.

Since there are two manners, two objective functions are created. In both cases the minimum of this objective function is taken. The multiplication factor is varied between 0 and 100 and the uniform roughness between 0 and 16. The sums of least squares of both manners are compared and the multiplication factor or system roughness that results in the smallest sum of least squares is used in the model.

2.2.5 SENSOR DEPLOYMENT INNER PRESSURE SENSORS

The performance of the leak localization method and the location of the sensors are interdependent (Bonada et al., 2014a; Bonada et al., 2014b; Meseguer et al., 2015). The initial set-up of measurements was planned with 15 smart meters. The locations of the smart meters were planned based on the current modelled preferred stream directions and locations with high stream velocities. At these points the change of pressure is the most sensitive, because the pressure changes in a quadratic way with the flow (Equation 14, page 14). Furthermore, the sensors were planned at the outer areas and on long pipe segments. After the locations were chosen, all respective customers were called to ask whether they wanted a smart meter. This was a time-consuming procedure with a low success-rate. Three out of fifteen customers accepted a smart meter.

As a new approach 500 e-mails were sent to randomly picked customers in the DMA Leimuiden. The e-mail contained information about the smart-meter, practical requirements (e.g. accessibility to electricity and internet connection) and the question if they are interested. Approximately 10% of the approached customers gave a positive reply. From these interested customers, twelve customers were selected based on the previous

planned locations of the smart meters. Due to holidays, practical unsuitability and firmware problems, only six of the planned smart meters were installed and working before the artificial leak tests.

In addition to the smart meters, seven pressure loggers on fire hydrants are installed. In total 107 fire hydrant locations are present in DMA Leimuiden (Appendix E). The location of the additional sensors is defined with the optimal model-based sensor location deployment procedure and uses a Genetic Algorithm (Bonada et al., 2014b). This procedure is developed to improve the performance of the leak localization method of Quevedo et al. (2011) and is focused on two requirements:

1. A high distinguishability among potential leaks. A leak causes a pressure change at the locations of the sensors. This is called the leak signature of a leak. Leaks close to each other cause the same leak signature and form a group. 'A group is a set of columns that have the same signature, which correspond to a set of leak locations that cannot be isolated because we cannot distinguish them (Bonada et al., 2014b).'

2. A strong robustness in front of model-reality divergences and other uncertainties (Bonada et al., 2014b). The pressure differences simulated with the model are very precise resulting in a unique response for every node and always 3218 groups. In practice this is not the case due to sensor resolution. Therefore the modelled results are truncated (Bonada et al., 2014b).

The model-based sensor location deployment starts with an extended Fault Signature Matrix (FSM_{ext}). This matrix has as many rows as possible pressure sensor locations and as many columns as possible leak locations. In the case of DMA Leimuiden this results in a matrix of 107 rows and 3218 columns. There are 107 fire hydrant locations where a pressure logger can be installed and the model consists of 3218 nodes where a leak can be modelled. The objective is to make the largest group with the same leak response as small as possible with an optimal combination of sensors. This results in the FSM_{sens}. The group size determines the practical search area of a leak. If the measurement devices are distributed in a way that the largest group is as small as possible, a high distinguishability among potential leaks is created (Bonada et al., 2014b). An example of a FSM_{sens} is given in Figure 2.9. In this small example there are nine possible leak locations and two locations with a sensor (node 2 and 7). The largest group contains three nodes (node 1, 2 and 3) that cause the same pressure change at the theoretical measurement locations.

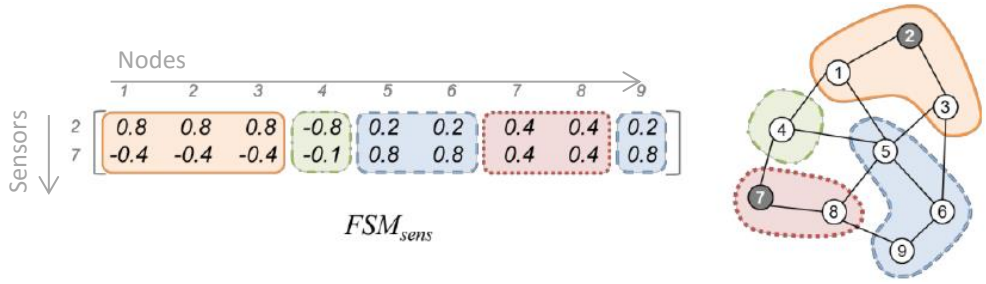

Figure 2.9 Example of a FSM_{sens}. Equal leak responses at the sensors (nodes 2 and 7) form one group. The largest group has leak response 0.8 and -0.4 for three nodes. Figure from Bonada et al. (2014b).

Requirement 1 (page 16), making the largest group with the same pressure response after a leak as small as possible, makes the deployment of the sensors an optimization problem. In most networks, however, the nodes are not equally distributed over the network causing a smaller group of nodes, not necessarily corresponds to a spatially smaller area to search for the leak. Different optimisation objectives can be used, linked to topological

(number of leaks), topographical (geographical coordinates) or sensibility (FSM values) indicators (Bonada et al., 2014a).

For this research the simplest objective of minimizing the largest group with the same leak response is used. The density of the nodes is the highest in the looped center of the DMA (Appendix E). This results in a higher 'weight' of this part compared to other locations since there are more nodes. This small advantage is compensated by the fact that the leak localization method has more difficulties in locating a leak in a highly looped area (Pérez et al., 2011b).

The groups with the same leak response are different during the day due to different flows. To determine the locations for the pressure loggers on the fire hydrants the situation where the largest groups are expected is used. This is with the uniform model and based on the minimum night flow (MNF) and the smallest leak size (7.5 m^3/h) that is used during the artificial leaks. During the night and with smaller leak sizes the pressure drop due to a leak is lower, which results in more locations with the same leak response.

Requirement 2 (page 16) is fulfilled by truncating the pressure responses to 1 decimal. This is based on the resolution of the pressure sensor devices (0.1 m for the smart meters, see Paragraph 2.1.2.3).

The sensor deployment optimization with the objective of making the largest group as small as possible uses a Genetic Algorithm (GA) to determine the sensor locations (Bonada et al., 2014b). With 107 possible locations and 7 pressure loggers, more than 1 billion different combinations of sensor locations are possible. GAs can be used to 'obtain optimal respectively near optimal solutions for the problem (Steffelbauer et al., 2014)'.

For this research, the Python package DEAP is used (Fortin et al., 2012). The GA parameters are; population size: 4000, crossover fraction: 0.5, individual mutation probability: 0.2, attribute mutation probability: 0.1, number of generations: 20. In Bonada et al. (2014b) values for the parameters are given but not discussed; population size: 1000, elite count: 2, crossover fraction 0.8, termination condition: 10 generations with a cost changer lower than 1e-6. The larger the population size the longer the computation size. This was not a problem and more individuals are evaluated. The crossover fraction is lower but in this research mutation is added. In this research the termination condition is fixed and it is only visually checked if the fitness function of the obtained best result was flat in the last generations.

2.2.6 LEAK LOCATIONS

The leak localization method is validated for artificial leaks at three different locations and one real leak. The leak localization method is applied both with the uniform model and factorized model. An overview of all leak scenarios is shown in Figure 2.10 and explained below.

Artificial leaks are created by opening a fire hydrant on three different locations, at different times and with two different leak sizes. The locations are chosen after the sensor deployment optimization. All three locations are in a different leak response group. The three locations of the artificial leaks are shown in Figure 2.11 (location I, II, III). The first location is in the highly looped part of the network in a residential area. The second location is less looped and situated at a small industrial area. The third location is in an outer branch. Pictures of the artificial leak locations can be found in Appendix G.

As described in the objective of this research and as recommended in literature, the artificial leaks are performed during the night since the effect of the unknown exact distribution of the consumption is the smallest during night (Pérez et al., 2009). Even though there is a leak, the total head loss is small. During peak hours, when there is more flow, the head loss is much larger. A leak during peak flow causes a larger head loss than during the night since the head loss increases quadratically with the flow (Appendix D). If the actual distribution of the

consumption is modelled well enough, the leak is localized easier during the peak hours than during the night. To test the influence of the relative leak size the artificial leaks are apart from the night, executed during the morning peak and during the daytime (after the morning peak). During the morning peak only location I and III are tested because there was not enough time to fit three locations with two leak sizes in one morning peak.

The head loss that is necessary for the leak localization method depends, among others, on the size of the leak. If the leak is too small there is not enough head loss to detect at the sensors. The model-based sensor deployment with a leak of 7.5 m^3/h already resulted in a large group that has the same leak response. This results in a large searching area and therefore this is taken as the lower limit. The leak of 15 m3/h is chosen as an upper limit. It is assumed that leaks above this size easily show up at the surface and therefor no leak localization is necessary.

The MNF of DMA Leimuiden increased in the night of 26 November 2015 with 5.2 m^3/h (Figure 2.12). Measurements were performed for three continuous days (2015-12-07 15:00 till 2015-12-10 12:00). The leak occurred after creation of the artificial leaks and the pressure loggers were already removed. The leak localization is performed with the smart meters only.

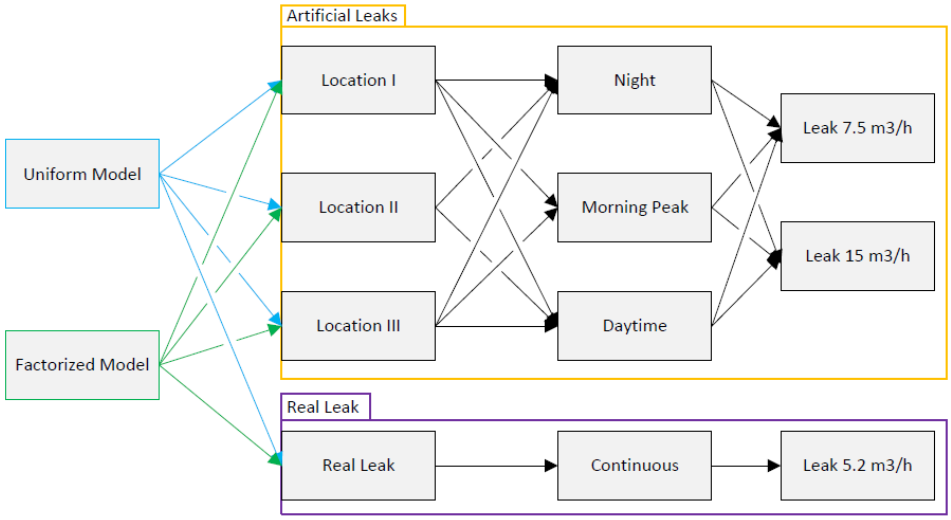

Figure 2.10 Leak scenarios artificial leaks and real leak

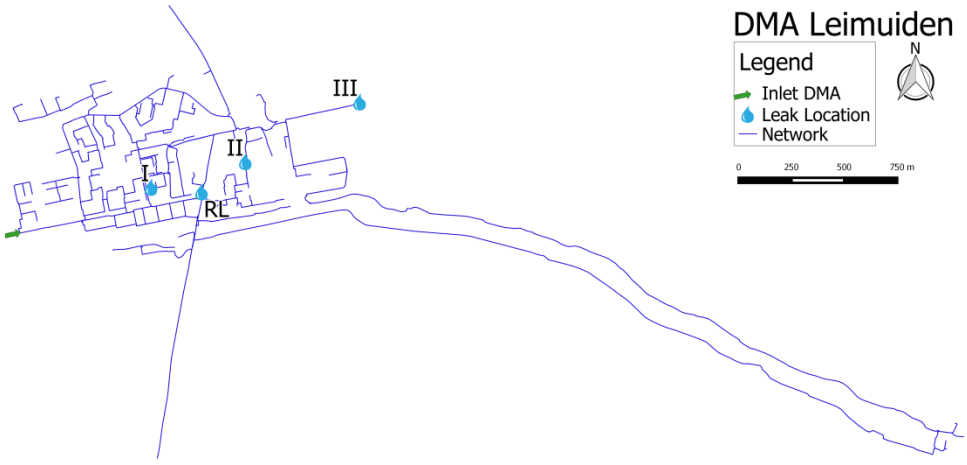

Figure 2.11 Artificial leak locations (location I, II, III) and the real leak location (RL)

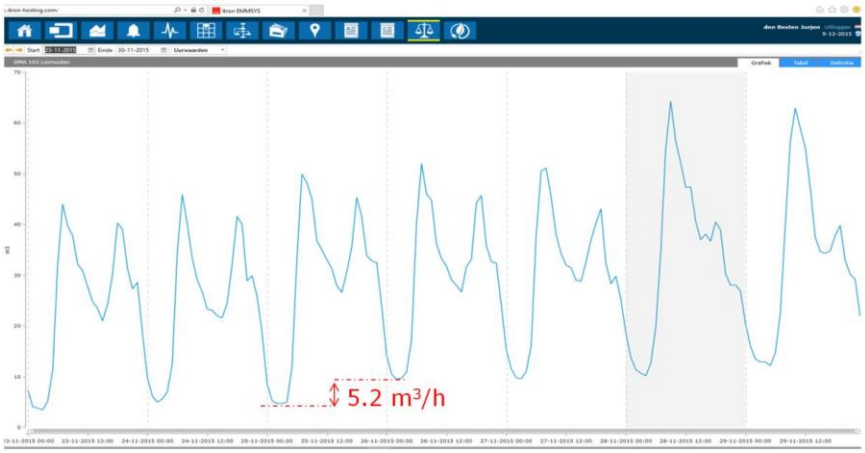

Figure 2.12 Increased MNF of 5.2 m^3/h in DMA Leimuiden

3. RESULTS AND DISCUSSION

3.1 SENSOR DEPLOYMENT

The sensor placement optimization of seven pressure loggers in addition to five already fixed smart meters is shown in Figure 3.1. The deployment of these sensors, at minimum night flow and a leak of 7.5 m^3/h, results in 56 groups with a unique leak signature that can be distinguished. The calibration result of the used model can be found in Appendix H (RMSE=0.64 mWc). The purple group b in the middle is the largest group, containing 1830 nodes that cannot be isolated separately because they cause the same leak response at the sensors (Figure 3.1).

The largest group contains more than half of the amount of total nodes (57%) and indicates a hard distinguishability of a leak in this looped area (purple group in Figure 3.1). A larger leak (15 m^3/h) results in 234 groups and the largest group consists of 392 nodes (Appendix I). This group is 4.5 times smaller than with a leak of 7.5 m^3/h.

The same deployment of sensors results in 54 groups during peak hour with a largest group of 557 nodes (Appendix I). This is more than 3 times smaller than during minimum flow with the same leak size (7.5 m^3/h). The smaller largest group confirms that, in theory, it is easier to distinguish leak signatures when there is more consumption flow in the network.

Figure 3.1 Result sensor placement optimization with a leak of 7.5 m^3/h during minimum flow. Top: groups with nodes that have the same leak response. Bottom: proposed location of measurement devices.

The actual deployment of sensors used during the artificial leaks is shown in Figure 3.2. This is nearly the same as the result from the sensor placement optimization except for smart meter 1027 (SM1027). This smart meter was

installed after the sensor placement optimization was executed. SM1027 was installed just before the artificial leaks and is taken into account for the leak localization.

The leak localization of the real leak is performed with the smart meters only since the pressure loggers were already removed when the leak occurred. The pressure logger at the inlet is reinstalled after the leak was detected.

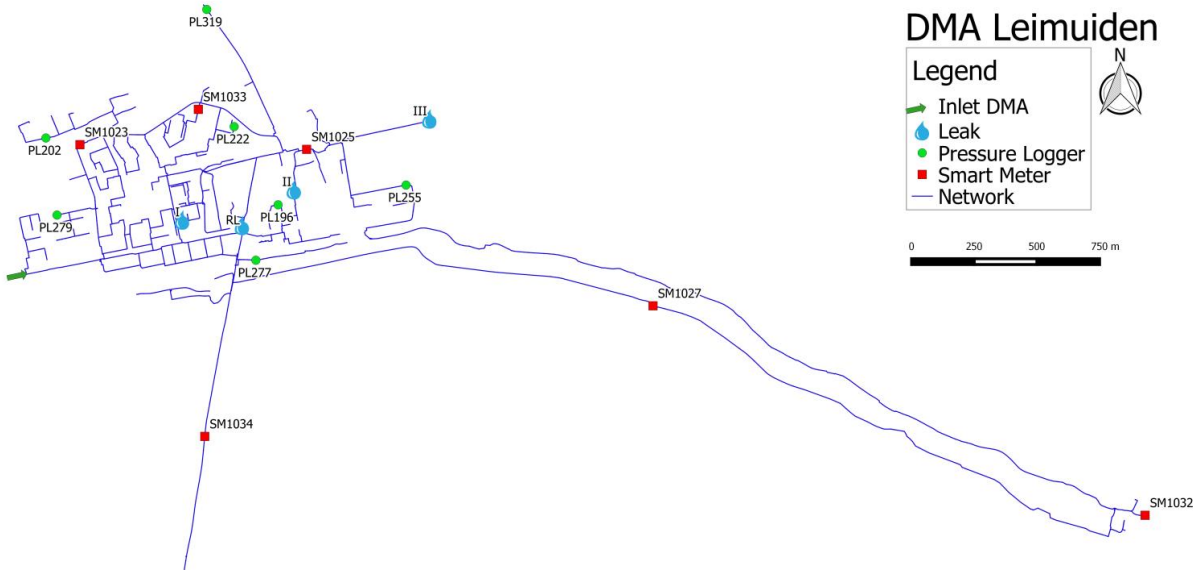

Figure 3.2 Sensor deployment. Pressure loggers and smart meters during artificial leaks. During real leak only smart meters are used.

3.2 PERFORMANCE LEAK LOCALIZATION ARTIFICIAL LEAKS

An overview of the leak localization performance executed with the uniform model and factorized model is shown in Table 3.1. The performance is the same for both models in 13 out of 16 artificial leaks (same class of percentage FP nodes: FP nodes<5% is good, 5%<FP nodes<15% is reasonable, FP nodes>15% is poor). For the remaining 3 leaks, the uniform model performs poorer than the factorized model. The calibration result of the factorized model was better than the calibration result of the uniform model (RMSE = 0.92 mWc, RMSE = 0.99 mWc for the fact., unif. model, respectively, Appendix J). Therefore, it is not surprising that the leak localization performance of the factorized model is better than the uniform model.

The leaks for which the factorized model scores poorer than the uniform model occurred only during the morning and daytime at location I (Table 3.1). There was no difference in performance for nighttime leaks. The morning leaks for which the models perform differently, are leaks at location I, with leak sizes of, 7.5 m3/h (performance is 14.7%, 5% for the unif. and fact. model, respectively) and 15 m3/h (performance is 41.0%, 14% for the unif. model and fact. model, respectively). During the daytime, this was with a leak size of 15 m3/h (performance is 22.1%, 7.8% for the unif. and fact. model, respectively).

It was expected that the performance of both models is the same during the night. At night, the flow of the leak is high compared to the consumption of the flow by customers (min 0.64 – max 0.96 Qleak/Qtotal flow ratios night, Table 3.1) while during the morning and daytime, the leak flow compared to the total flow is smaller (min 0.13- max 0.42 Qleak/Qtotal flow ratios morning and daytime, Table 3.1). Because the consumption by customers is small compared to the leak, the different consumption factors, essentially the distribution of the consumption, have a small impact on the pressure during the night (Appendix K).

Table 3.1 Leak localization performance artificial leaks. If the percentage of false positive (FP) nodes is underneath 5% it is classified as good (green), between 5 and 15% as reasonable (orange) and above 15% as poor (red). The blue rectangles indicate the leak localizations that have a different performance category for the consumption models.

Time	Location	Leak Size [m3/h]	Q Leak/Q Total Flow	Uniform Model [FP Nodes]	Factorized Model [FP Nodes]
Night	I	7.5	0.64	74.4%	72.2%
		15	0.72	41.7%	41.1%
	II	7.5	0.78	2.5%	4.0%
		15	0.96	1.0%	1.2%
	III	7.5	0.82	1.0%	1.3%
		15	0.91	0.8%	1.7%
Morning	I	7.5	0.13	14.7%	5.0%
		15	0.27	41.0%	14.0%
	III	7.5	0.17	2.2%	0.8%
		15	0.28	2.0%	1.9%
Daytime	I	7.5	0.20	26.1%	40.7%
		15	0.34	22.1%	7.8%
	II	7.5	0.20	20.1%	40.7%
		15	0.39	18.1%	19.8%
	III	7.5	0.28	2.4%	1.9%
		15	0.42	1.2%	0.8%

The leak localization result of the uniform model during the night is shown in Figure 3.3. The Figure shows the leak correlation for each network node. A high correlation with reference to the other nodes in the same network (red nodes) means the largest probability of the existence of a leak, while a low leak correlation with reference to the other nodes in the same network (black nodes) indicates the lowest probability of a leak at that location. The leak at location I is poor localized (top Figure in Figure 3.3).

One of the reasons for the poor localization of the leaks at location I during the night, can be the lower ratio of leak flow to total flow compared to ratio of leak flow to total flow at the other two locations. The leaks at location I were created earlier than the leaks at the other two locations (Loc I. start: 00.30h, Loc. 2 start 02:00h, Loc 3. Start 03:00h, see top left Figure in Figure 3.4). Compared to the other two locations, the ratio of leak flow to total flow was low (0.72, 0.96, 0.91 for location I, II, III, respectively, see Table 3.1 and Figure 3.4). A lower leak flow to total flow ratio means more inflow next to the leak flow that has to be distributed over the customers. Possibly, both models distributed the consumption flow next to the leaks at location I in an incorrect way. A wrong distribution of this consumption flow can result in pressure differences that hide the pressure difference caused by a leak (Pérez et al., 2011b).

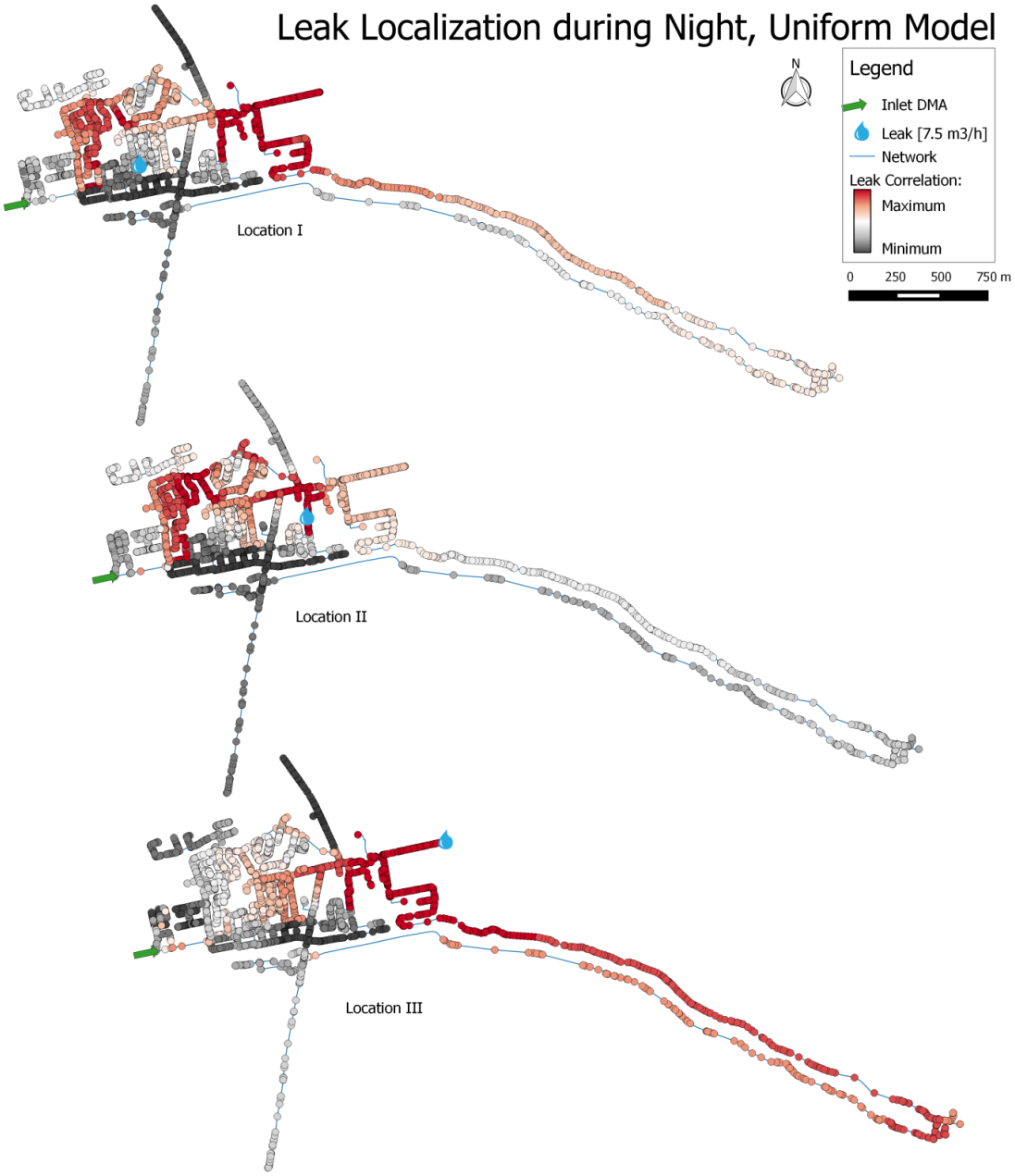

Figure 3.3 Leak localization result during night with the uniform model and a leak size of 7.5 m^3/h. Red nodes (high leak correlation with reference to other nodes in same network): high probability, black (low leak correlation with reference to other nodes in same network): low probability. Top: leak localization result location I, lc_{max}= 0.34. Middle: leak localization result location II, lc_{max}= 0.38. Bottom: leak localization result location III, lc_{max}= 0.79. Leak localization results of the factorized model look similar and are shown in Appendix L.

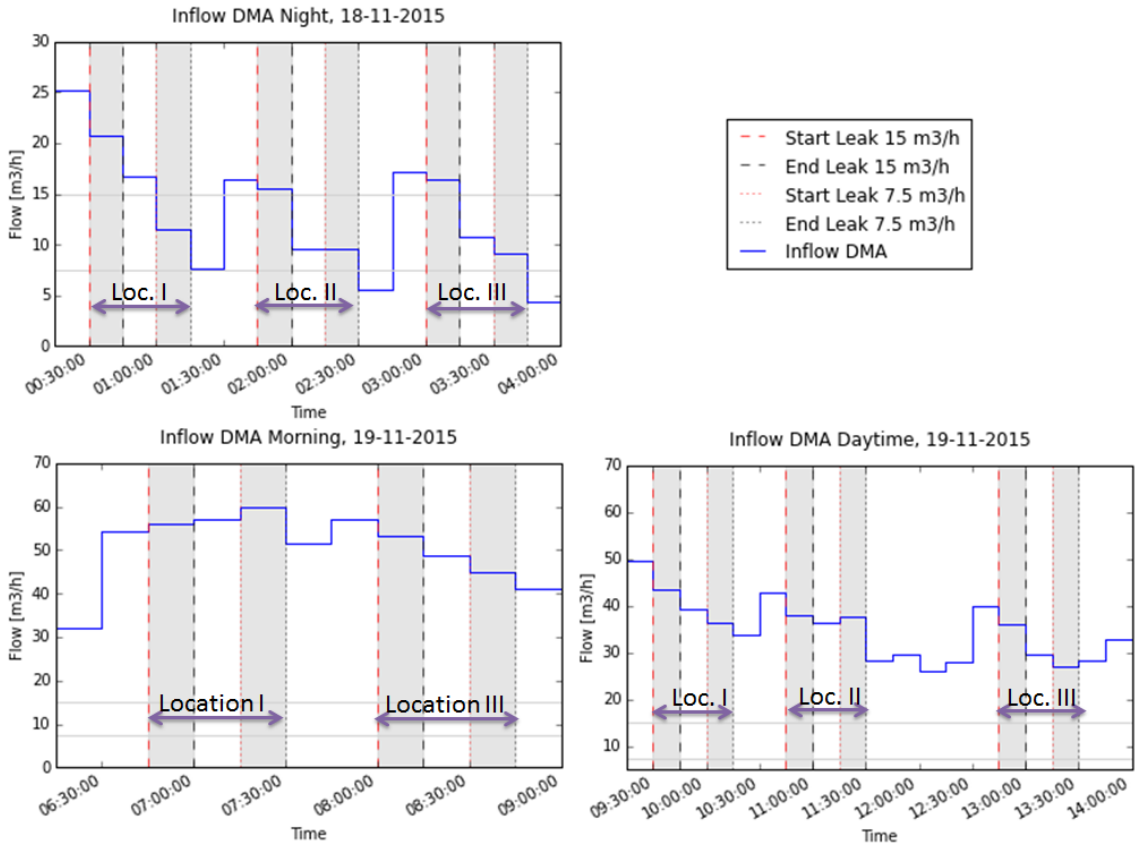

Figure 3.4 Inflow DMA during artificial leak testing. Top left: inflow and artificial leaks during night. Bottom left: inflow and artificial leaks during morning. Bottom right: inflow and artificial leaks during daytime

Table 3.2 Maximum simulated pressure response due to a leak at the sensor locations. All simulated pressure responses at the sensor locations can be found in Appendix M.

		Uniform Model			**Factorized Model**		
Location	*Leak Size [m3/h]*	Night *Maximal ΔP [mWc]*	Morning *Maximal ΔP [mWc]*	Daytime *Maximal ΔP [mWc]*	Night *Maximal ΔP [mWc]*	Morning *Maximal ΔP [mWc]*	Daytime *Maximal ΔP [mWc]*
I	7.5	-0.03	-0.16	-0.10	-0.02	-0.13	-0.08
II	7.5	-0.05	n/a	-0.15	-0.04	n/a	-0.14
III	7.5	-0.24	-0.51	-0.39	-0.21	-0.50	-0.35
I	15	-0.08	-0.29	-0.22	-0.07	-0.23	-0.18
II	15	-0.16	n/a	-0.33	-0.13	n/a	-0.28
III	15	-0.94	-1.51	-1.25	-0.77	-1.37	-1.11

Another reason why the leaks at location I are poorly localized can be the location of the leak in the network. Leak location I is located in the looped part of the network (Figure 3.2). 'The main handicap of the methodology is that in a highly looped network pressure drops due to a leak are not very significant (Pérez et al., 2011b).' Table 3.2 summarizes the maximum simulated pressure response in the model due to the leaks at the sensor locations (the pressure responses at all sensor locations are given in Appendix M). During the night the maximum simulated

pressure response due to a leak at location I is low with leak size 7.5 m^3/h (-0.03 mWc, -0.02 mWc for unif., fact. model, respectively, see Table 3.2). Uncertainty in the model and measurements can easily mask this small pressure response resulting in poor leak localization. The measured responses are shown in Appendix N.

Notable is that, despite the fact that the simulated maximal pressure response is lower for the leak of 7.5 m^3/h at location II than for the leak of 15 m^3/h at location I during the night, the leak at location II is localized good while the leak at location I is localized poor (unif. model: max. head loss is -0.05 mWc, -0.0.8 mWc and performance is 2.5%, 41.7% FP nodes for loc. II,I, respectively, fact. model: max. head loss is -0.04 mWc, -0.07 mWc and performance is 4.0%, 41.1% for loc. II,I, respectively, see Table 3.2 and Table 3.1). This may be due to larger errors in the model at the times the leak localization performance is performed. Errors in the model can be caused by misfitting of the consumption in the model and reality (Pérez et al., 2011b). The errors in the model, the day before the leaks are created, are shown in Figure 3.5 and are called the residuals (differences between measured pressures and modelled pressures). The larger the fluctuations, the larger the errors in the model. In the beginning of the night when leak location I, leak size 15 m^3/h is created (00:30 h) the fluctuations are larger than at the time of the creation of a leak of 7.5 m^3/h at location II (02:30) if the residuals are comparable with the night before. The model is less good in simulating the pressures right at the beginning of the night compared to the middle of the night. This can have a negative consequence on the performance of the leak localization, especially if the pressure losses due to the leak are small compared to the errors.

The simulated pressure differences for leaks at location III are the largest for all times which makes it easier to correctly localize a leak by the leak localization method (7.5 m^3/h, unif. model: -0.24 mWc, -0.51 mWc, -0.39 mWc during the night, morning, daytime, respectively and fact. model: -0.21 mWc, -0.50 mWc,-0.35 mWc during the night, morning, daytime, respectively, Table 3.2). The performances of the leak localizations of the leaks at location III are all good (Table 3.1).

During the morning, the maximum simulated pressure responses are higher for all leaks compared to the night (e.g. location I, fact. model, night: -0.02 mWc, morning: -0.13 mWc). The residuals, however, are also larger indicating a poorer performance of the model in predicting the correct pressures (larger fluctuations between 06:30h-08:00h in Figure 3.5). Residuals are caused by uncertainties in the model (e.g. diameter, length, roughness, consumption factor). Uncertainties in the model parameters have the largest effect on pressure differences in the morning (Appendix K). Larger head loss caused by a leak in the morning makes it easier to localize the leak but large uncertainties make the simulated pressure of the model poorer and this can have negative consequences for the leak localization result.

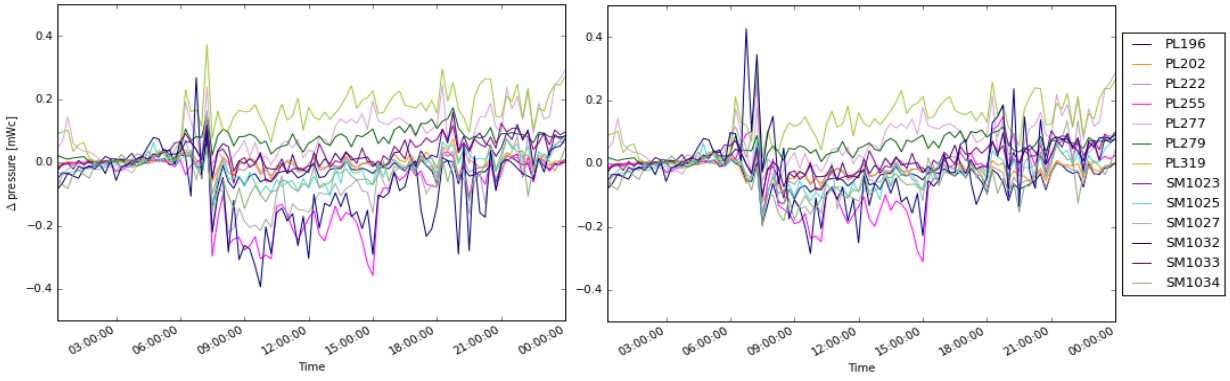

Figure 3.5 Residuals between measured values and simulated values after calibration the day before artificial leaks take place (2015-11-17). Left: residuals uniform model. Right: residuals factorized model.

During the morning, the uniform model shows the high leak correlations mainly in the loop on the right side of the DMA (red rectangle in Figure 3.6), while the leak is located at location I. The model detects a larger flow than expected in this part of the network resulting in high leak correlations. This larger flow is most likely caused by the fact that two of the three hotspots (consumer(s) group(s) with, according to the billing information, a significant higher consumption related to the rest of the area) are located here (red nodes in Figure 3.7). The uniform model distributes all consumption flow equally over all customers and only a leak causes a larger flow. A higher flow can be denoted as a leak and results in higher leak correlations at these locations. The result of the factorized model is better (bottom Figure 3.6). This indicates that the consumption factors in the factorized model have a positive effect on the performance of the leak localization of leaks at location I.

Leak location III is localized good with both models (%FP nodes <2.5, Table 3.1 and visualizations in Appendix L). This leak location is closer to the large consumers. Larger consumers cause more flow than the average flow that is equally distributed in the uniform model. This can positively influence the leak localization performance with the uniform model of leaks at location III since the leaks are faster denoted to the area where large consumers are. Large consumers in the uniform model cause a higher flow than expected which is the same effect with the occurrence of a leak.

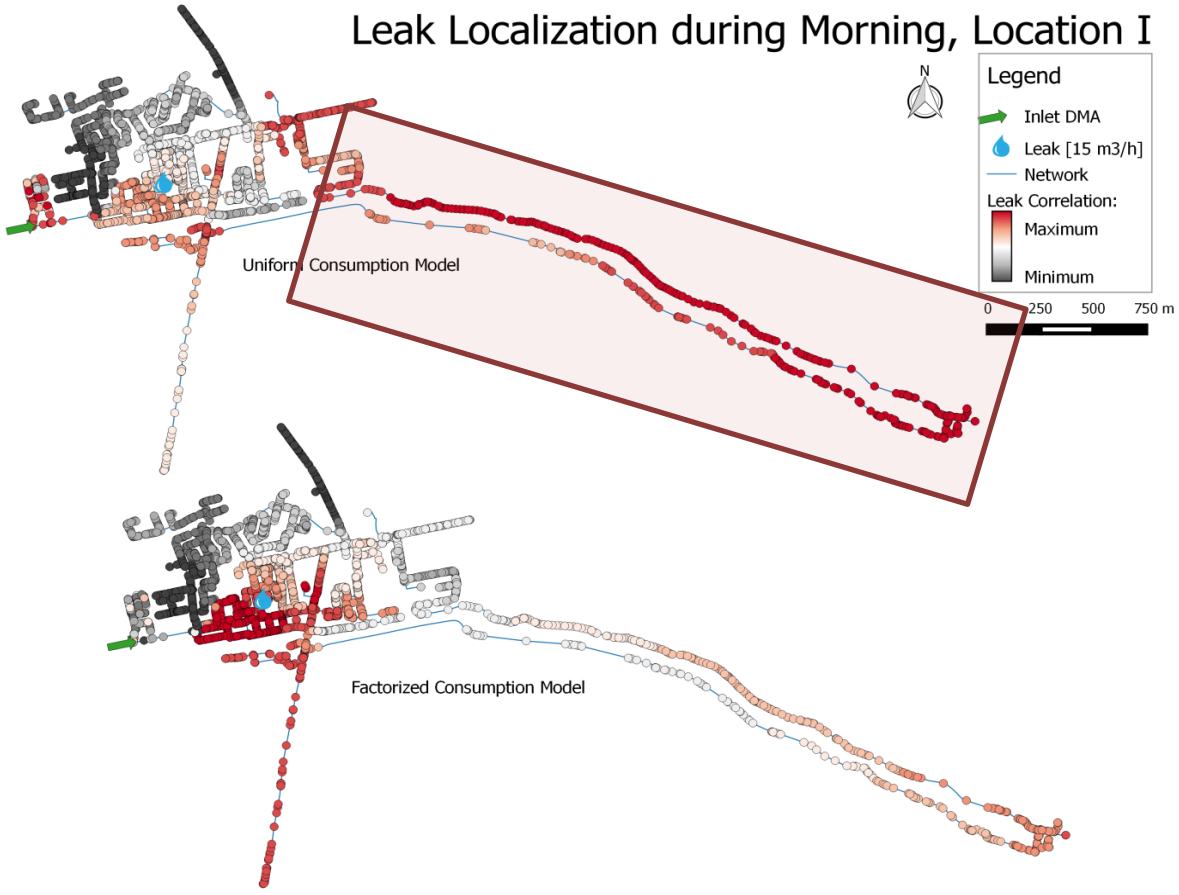

Figure 3.6 Leak localization result location I during the morning and a leak size of 15 m^3/h. Red nodes (high leak correlation with reference to other nodes in same network): high probability, black (low leak correlation with reference to other nodes in same network): low probability. Top: leak localization result performed with uniform model, lc_{max}= 0.85. Bottom: leak localization result performed with factorized model, lc_{max}= 0.54. Leak localization results of the other leak size (7.5 m^3/h) and for location III are shown in Appendix L. The red rectangle highlights the area where the uniform model indicates the highest probability of the leak.

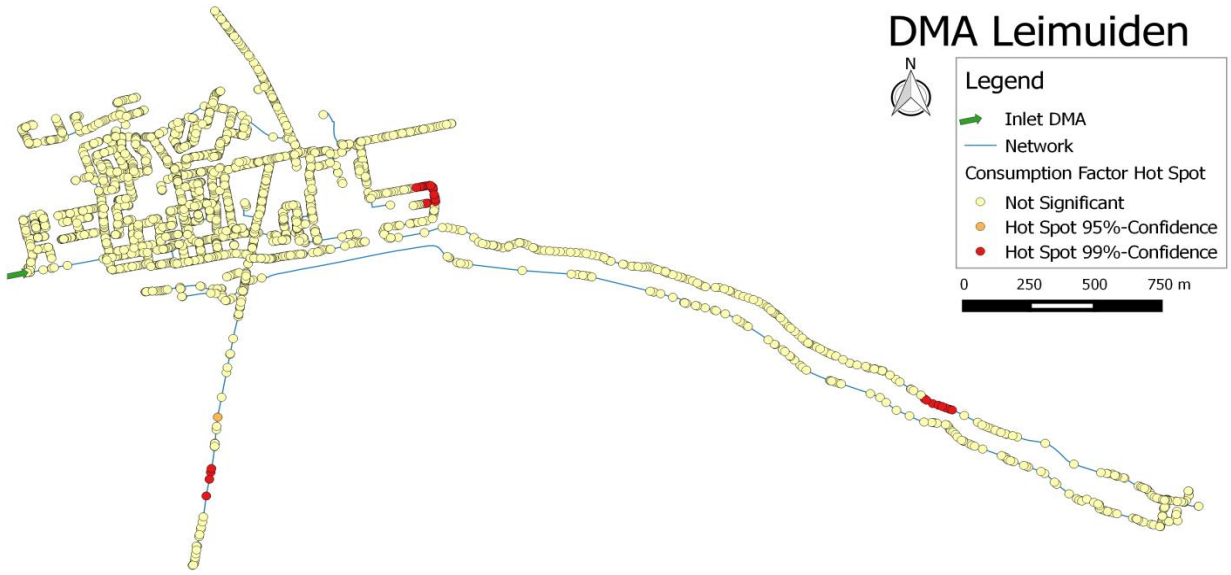

Figure 3.7 Consumption factors hot spot map. Red nodes represent a cluster of high consumption base factors compared to the rest of the area.

The leak localization method always gives an output but is not always reliable. 'There is not an actual yes/no determination if a leak has occurred. Instead, the methodology delivers the nodes with highest probability of having the leak (Pérez et al., 2014a)'. It is always possible to rank all leak correlations of the nodes in a descending order and start searching for the leak in the first group of nodes. According to (Quevedo et al., 2011), a high leak correlation value is needed in order to make the method reliable (maximum possible leak correlation is 1). It is not specified how high this leak correlation has to be have a reliable result.

In Figure 3.8, the maximum correlation and false positive nodes for all scenarios are shown. A high leak correlation does not always result in a good performance but there is a small agglomeration of points that have a high leak correlation and a poor performance (high percentage FP nodes). Despite the relative high maximum correlations during day time (uniform model: lc > 0.77, factorized model: lc>0.69, purple points in Figure 3.8), the performance is in most cases poor (%FP nodes>15%, Figure 3.8). A bad performance during day time indicates that the factors are not well in predicting the consumption in the network. During the day many people are to work and school and do not use water at home. Additionally, all consumers have the same pattern and it might be better if different patterns are used for different types of customers during the day. In Appendix O results of a third consumption distribution model are shown where the real consumption by the largest consumer in Leimuiden is added. This does not improve the performance of the factorized model.

Except for the daytime leaks, there is one more leak localization that has a high maximum leak correlation but a poor performance. This is the point, localized with the uniform model in the morning (max. leak corr.=0.85, percentage FP nodes=41%, see Figure 3.8: left). At this point the measurements have a high correlation with a simulated leak that is located directly in the branch after the inlet of the DMA. A simulated leak at this location does nearly affect the pressures at the inner sensors. The leak correlation, however, does not depend on the magnitude and only on the linear relation. A visual representation is given in Appendix O. An additional sensor in this branch can prevent the occurrence of having almost no effect at the sensor locations when a leak is simulated in this branch.

In order to apply the leak localization method in practice, it is important to know when the outcome of the leak localization is reliable. Based on these artificial leaks; a high correlation does not always result in a good performance. It is important to look in more detail to the node that has the highest correlation. It is possible that

this is one node, accidently having a high correlation with the FIV. In the morning and night there is a stronger relationship between high leak correlation and percentage FP nodes than during daytime.

In Appendix P a correlation matrix is given between the distance of the leak to the node with the maximum leak correlation and factors that can possibly indicate the performance of the leak localization.

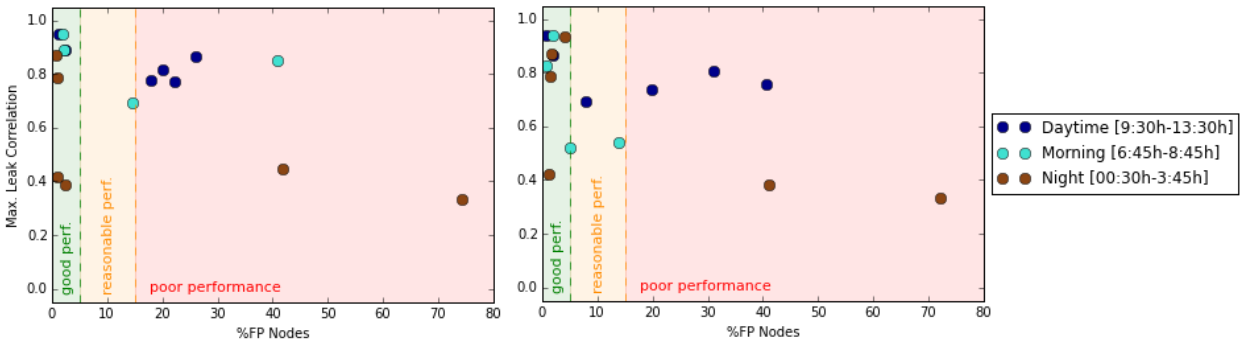

Figure 3.8 Performance leak localization in relation to the maximum leak correlations of the artificial leaks. Performance and max. leak correlation executed with the uniform model (left) and with the factorized model (right).

The artificial leak localizations show different performances for different locations. To the author's knowledge, research that discusses the case-studies and the proof of concept of the leak localization method, does not discuss the influence on performance of the leak localization depending on the location in the network (Pérez et al., 2014a; Mirats-Tur et al., 2014; Pérez et al., 2011b; Meseguer et al., 2014; Quevedo et al., 2011; Quevedo et al., 2012). With the exception of one (previous cited) sentence: 'The main handicap of the methodology is that in a highly looped network pressure drops due to a leak are not very significant (Pérez et al., 2011b).' DMA Nova Icària is often used to apply the leak localization method. This DMA is a highly looped system with less spatial differences than DMA Leimuiden. DMA Nova Icària is mainly looped (Figure S.1) while DMA Leimuiden is more a combination of a looped and branched network (Figure 1.1).

3.3 PERFORMANCE LEAK LOCALIZATION REAL LEAK

The performance of the accumulated leak correlation results is shown in Table 3.3. The results of 9 December are not used because every hour strange peaks occur in the inflow measurements and it is unknown how these peaks are caused. Since these peaks occur every last fifteen minutes of every hour of the day, the leak localization results are not considered further in this research. The results of 8 December are described in more detail below. Results of 7 and 10 December are presented in Appendix R.

Table 3.3 Performance leak localization accumulated correlations real leak. If the percentage of false positive (FP) nodes is underneath 5% it is classified as good (green), between 5 and 15% as reasonable (orange) and above 15% as poor (red).

	Uniform		Factorized	
	Performance accumulation lc of all T_a [FP nodes]	**Performance accumulation lc of all T_w [FP nodes]**	**Performance accumulation lc of all T_a [FP nodes]**	**Performance Accumulation lc of all T_w [FP nodes]**
7-Dec (15:00-24:00)	13.1%	31.3%	20.5%	8.8%
8-Dec (00:00-24:00)	0.5%	11.8%	0.2%	0.3%
~~**9-Dec (00:00-24:00)**~~	~~4.8%~~	~~14.2%~~	~~5.8%~~	~~7.0%~~
10-Dec (00:00-12:00)	9.4%	2.4%	10.8%	0.8%

The performance of the leak localization for the real leak during 8 December, every 15 minutes is shown in Figure 3.9 (green line). The inflow into the DMA is the blue line in Figure 3.9 and shows the daily pattern. The percentage of false positive nodes shows much fluctuation during one day. The minimum percentage FP nodes (best performance) of the leak localizations calculated with the uniform model is 0% FP nodes at 18:15h (the highest leak correlation is in this case the real location of the leak) and the maximum percentage FP nodes (poorest performance) is 41.05% FP nodes at 16:15h (Figure 3.9:left). The performance of the leak localizations calculated with the factorized model is 0% (9:45h and 12:45h), 57.49% (06:45h) FP nodes for the minimum and maximum performance, respectively (Figure 3.9: right).

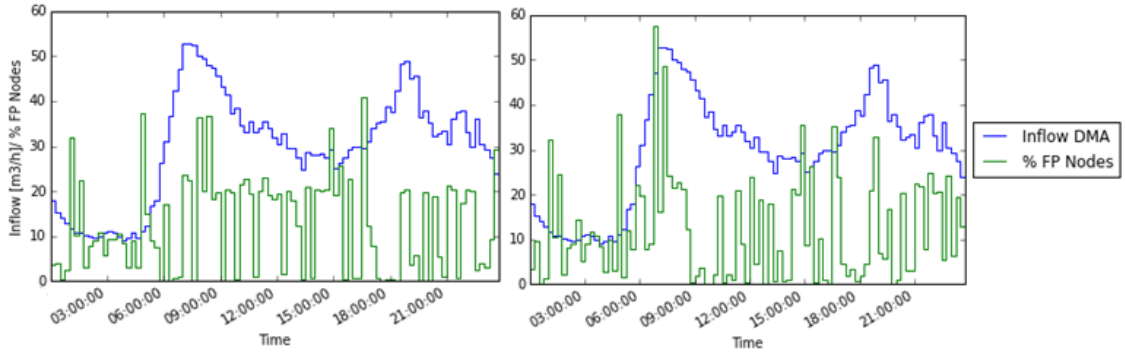

Figure 3.9 Performance leak localization real leak of one day with time analysis step of 15 minutes (2015-12-08). Leak localization performance executed with the uniform model (left) and with the factorized model (right).

Figure 3.10 shows the performance of the leak localization (% FP Nodes) and the maximum leak correlation. To observe the influence of the time of the day, the day is divided in 8 groups and the points are colored depending on the group (Figure 3.10). The leak localization executed with the uniform model shows two spots with a high agglomeration of points (Figure 3.10: left). There is one agglomeration of points in the left corner. These points have a high maximum leak correlation and a good performance (low percentage FP nodes). The other agglomeration of points is located more to the right. These still have a high maximum correlation (>0.8) but have a percentage of FP nodes around 20%. Leak localizations executed in the morning peak have high correlations but a poor performance (light-green points in Figure 3.10: left).

The points calculated with the factorized model show more spreading but the same agglomeration of points can be detected. There are points that despite the high maximum leak correlation, have a high percentage of FP nodes. These exceptions occur at all time intervals for both models.

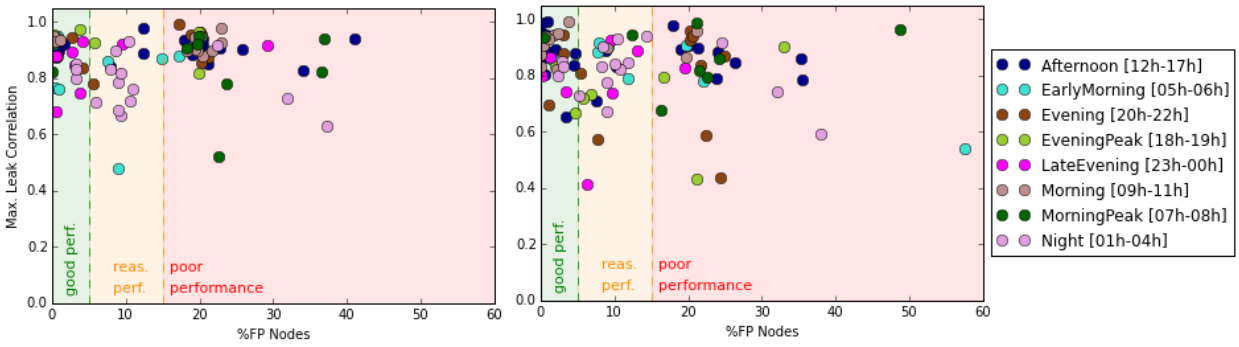

Figure 3.10 Performance leak localization in relation to the maximum leak correlations of the real leak executed every 15 minutes of one day (2015-12-08). Performance and max. leak correlation executed with the uniform model (left) and with the factorized model (right).

Figure 3.11 shows the result of the accumulated leak correlations above 0.5 for one day. Both (uniform and factorized) models have their maxima (red nodes) around the location of the leak. There is a difference, however, in the large loop in the right part of the network. The uniform model gives high accumulated correlations in this large loop in the network which indicates that there is a leak in this part of the network (light red points in Figure 3.11) while in fact there is no leak there. The factorized model does not (light grey points in Figure 3.11). This is the same phenomenon that is happening with the leak localization of the artificial leak location I (Figure 3.6). In the uniform model all consumers get the same consumption and red nodes indicate a larger flow than expected. This can be caused by a leak or consumption by customers. Larger customers are better represented in the factorized model and this can result in the difference of performance between the uniform model and factorized model in this right part of the network.

In the result of both models it looks like there are low and high accumulated leak correlations close to the real location of the leak. A zoom-in of the leak localization executed with the uniform model is also shown in Figure 3.11. It is clear that the nodes around the leak location are dark red and that the nodes that have a low accumulated leak correlation (grey nodes) are connected to a different part of the network. This holds true for the result of the uniform model.

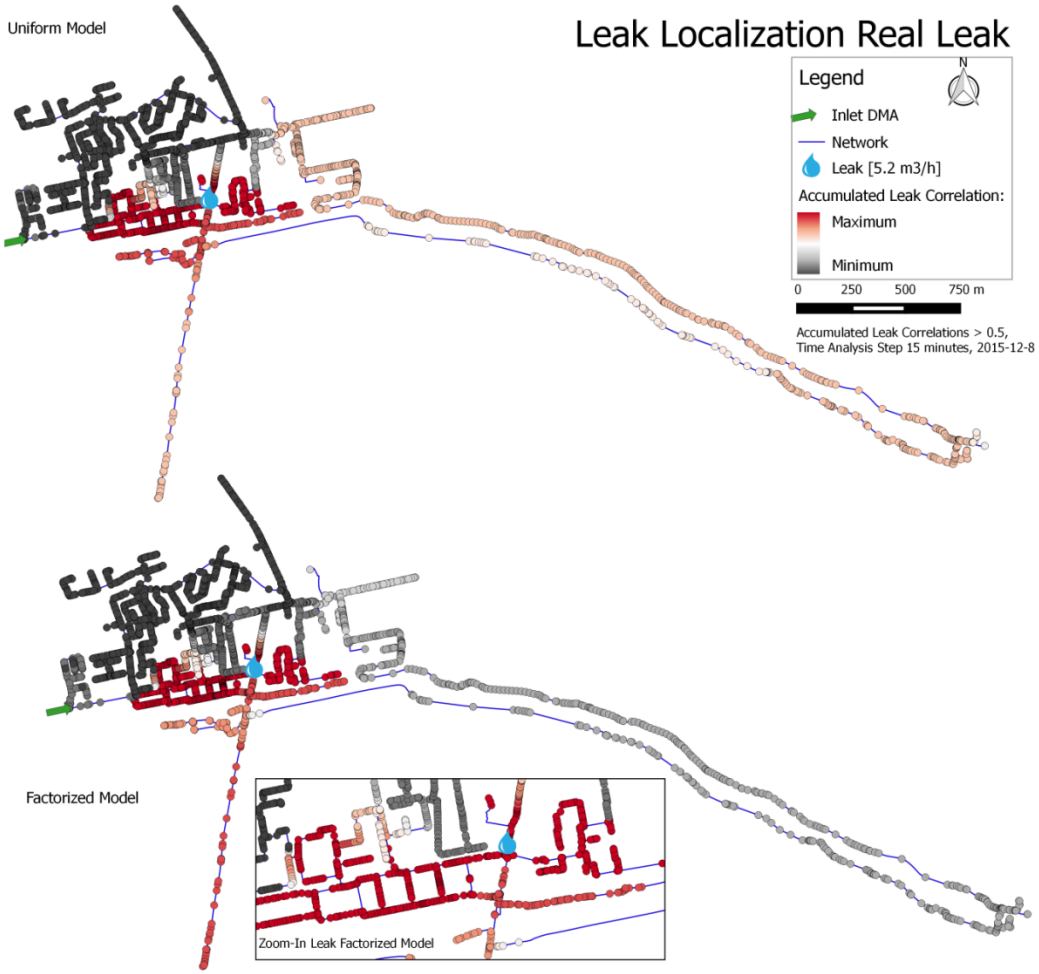

Figure 3.11 Leak localization result accumulated leak correlations with a time analysis step of 15 minutes. Red nodes (high leak correlation with reference to other nodes in same network): high probability, black (low leak correlation with reference to other nodes in same network): low probability. Top: leak localization result performed with uniform model. Bottom: leak localization result performed with factorized model.

The performance of the leak localization for the same day but with a time window of 1 hour (combination of four time analysis steps of 15 minutes) is shown in Figure 3.12. The performance of the uniform model fluctuates between 0.19% and 31.67% FP nodes (min. 17:00h and max. 08:00h in Figure 3.12: left). The performance is mainly poor during daytime. The factorized model fluctuates between 0.03% and 18.77% FP nodes (min. 09:00h and max. 14:00h in Figure 3.12: right). The maximum percentage of FP nodes are lower than when the leak localization is performed directly after every time analysis step of 15 minutes (41.05%, 57,49% FP nodes for the uniform, factorized model respectively, Figure 3.9). This is consistent with the statement that the use of multiple time analysis steps adds robustness to the performance of the leak localization (Meseguer et al., 2014).

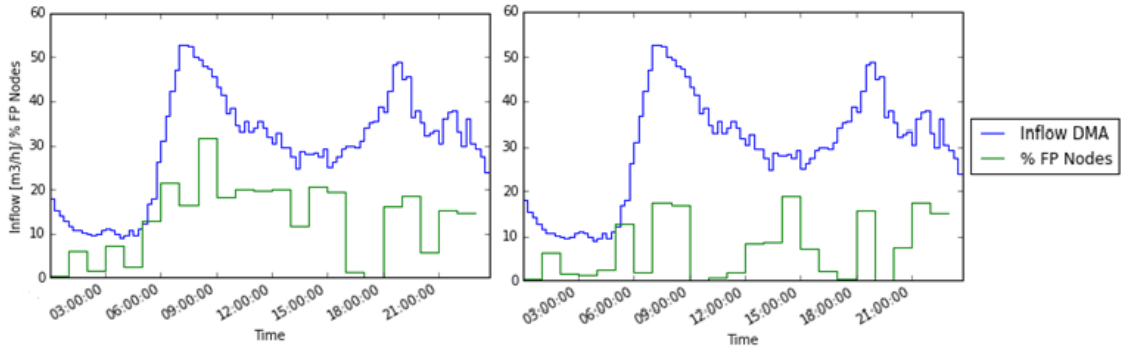

Figure 3.12 Performance leak localization real leak of one day with a time window of 60 minutes (2015-12-08). Leak localization performance executed with the uniform model (left) and with the factorized model (right).

Figure 3.13 shows the maximum leak correlation that result from the leak localization and the performance of this result (%FP nodes). The points that are calculated with the uniform model are horizontally spread around a maximum leak correlation of 0.8 and a high leak correlation does not indicate a bad performance (Figure 3.13: left). In the factorized model, an agglomeration of high maximum leak correlations with a good performance is present. A decreasing maximum leak correlation results in a reasonable and poor performance. For the factorized model, the maximum leak correlation is a good indicator if the result is reliable because the higher leak correlations results in lower percentages of FP nodes. This can be used in practice to judge the result of the leak localization.

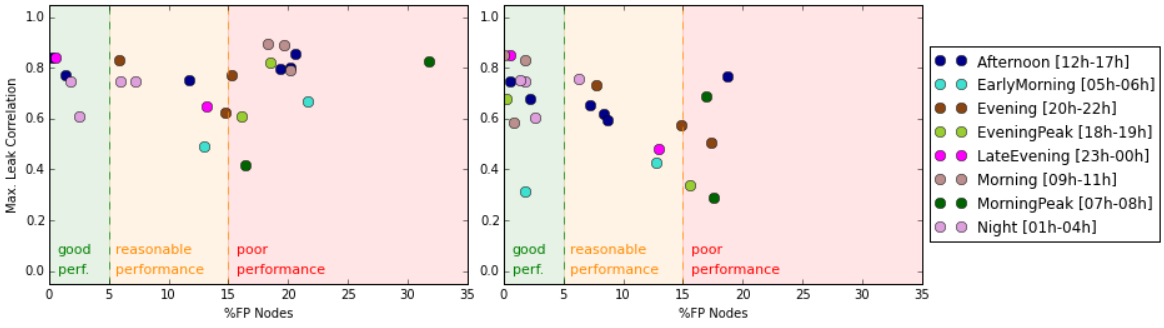

Figure 3.13 Performance leak localization in relation to the maximum leak correlations of the real leak with a time window of 60 minutes during one day (2015-12-08). Performance and max. leak correlation executed with the uniform model (left) and with the factorized model (right).

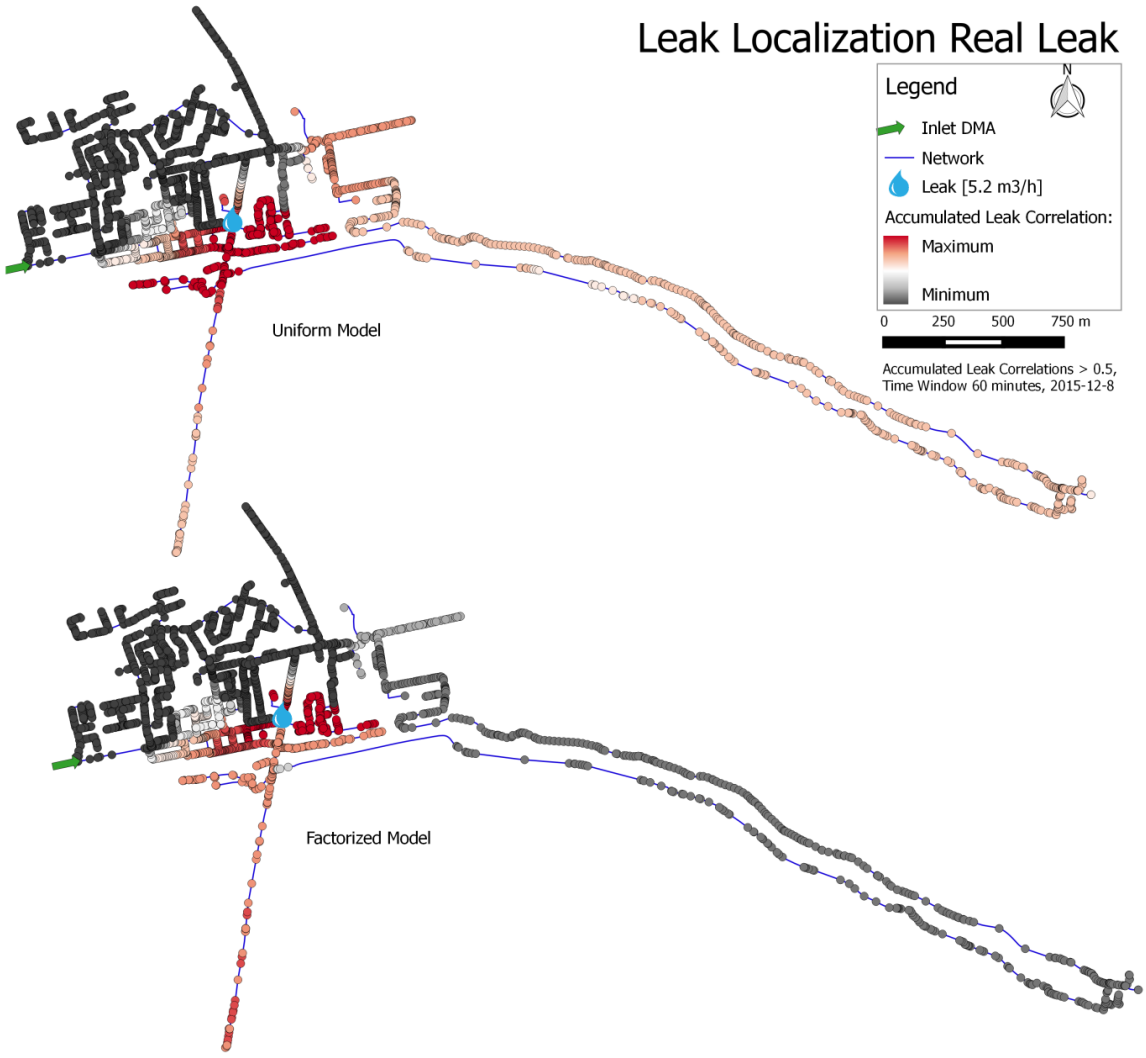

Figure 3.14 Leak localization result accumulated leak correlations with a time window of 60 minutes. Red nodes (high leak correlation with reference to other nodes in same network): high probability, black (low leak correlation with reference to other nodes in same network): low probability. Top: leak localization result performed with uniform model. Bottom: leak localization result performed with factorized model.

Figure 3.14 shows the leak localization after accumulation of c>0.5 for 24 time steps (24 hours). Both models have the real leak location in the area where the highest accumulated leak correlations occur (dark red nodes in Figure 3.14). The difference of the performance of the model is again in the large loop in the right part of the network and caused by the different consumption factors in the two models. Compared to the accumulation of the 15 minutes time step, the most probable search area is reduced for the factorized model (dark red nodes in Figure 3.11 and Figure 3.14).The performance for the uniform model got poorer (0.5%, 11.8% FP nodes with the accumulation of all T_a, T_w respectively).

Results of the real leak are compared with a leak localization results in DMA Nova Icària described in Appendix S. The leak in DMA Nova Icària is localized within a circle of 150 meter radius (Figure S.2 A graphical representation (black stars) of the most probable localization of the leak according to the accumulation of the resulting correlation vectors related to every node through a 48-h time window (December 20 and 21).Figure S.2). This is the area around the nodes that have the highest correlation values. The definition of highest correlation

values is not given.

Figure 3.15, shows the result of the hourly accumulated leak correlations of the real leak in Leimuiden. The location of the highest leak correlation (maximum accumulated c-value: C_{max_acc}) is represented with a big star. Smaller stars represent nodes that have c-values larger than $0.99*C_{max\_acc.}$. Both models have their maximum near the real leak (unif. model: 35 m, fact. model: 11 m). Hereby it is important to take into account that there is an uncertainty in the exact location of the leak in the model of approximately 10 meters.

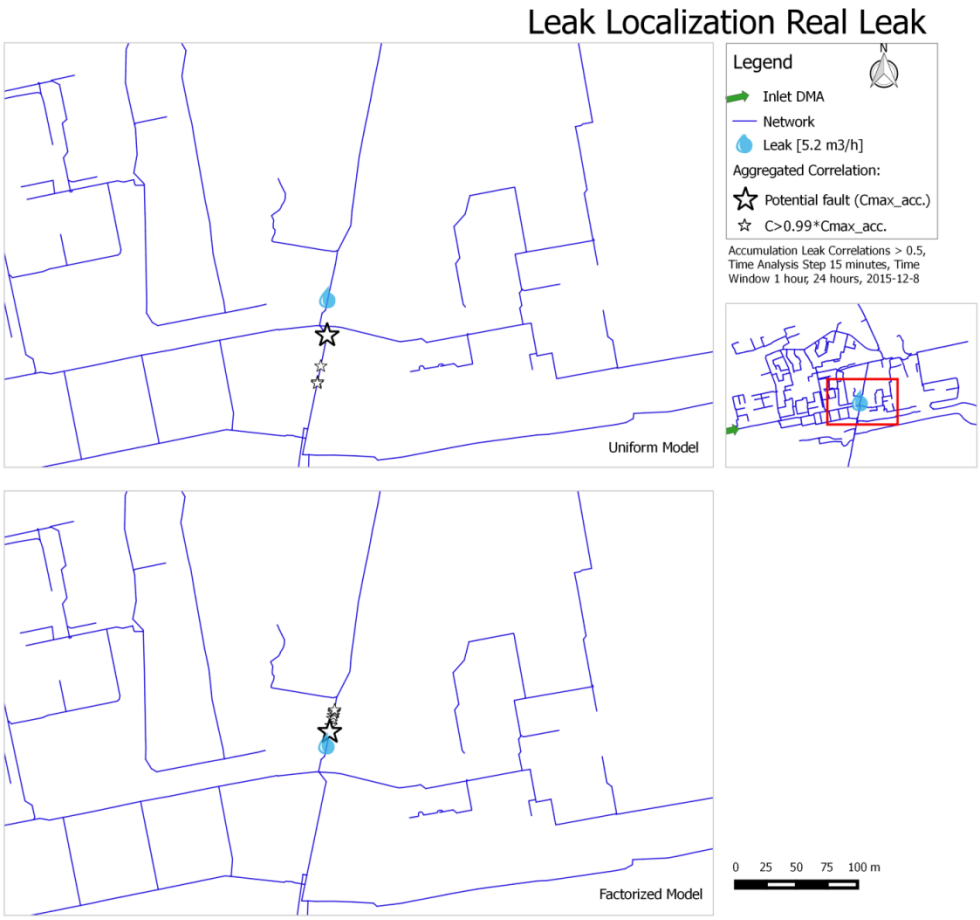

Figure 3.15 High leak correlations of leak localization result accumulated leak correlations with a time window of 60 minutes. The large star represents the node with the highest accumulated leak correlation (C_{max_acc}). Smaller stars are larger than $0.99*C_{max\_acc}$. Top: high leak correlations of leak localization result performed with uniform model. Bottom: high leak correlations of leak localization result performed with factorized model.

4. CONCLUSION AND RECOMMENDATIONS

The leak localization method of Quevedo et al. (2011) was validated for eight artificial leaks of 15 minutes and one real leak. For all leaks the method gave satisfying results (the leak was present in the top 15% of ranked leak locations). This shows that the leak localization method can be used in practice to reduce the search area of a small leak in a DMA with a small MNF. Based on this research several conclusions are drawn:

- The factorized model (consumption was based on billing information) yields better results than the uniform model (consumption was distributed equally over all customers). The method is influenced by the ability of a model to predict the consumption by every customer because this affects the pressures at the location of the sensors. It was found that during the night, the influence of the choice of consumption distribution method can be neglected. The results were similar for both the uniform model and the factorized model. This is due to the low consumption during the night, which causes the leak size (7.5 m^3/h and 15 m^3/h) to be large in comparison to the consumption flow (4 m^3/h). In the morning peak, however, the leaks are small compared to the consumption flow (50 m^3/h) and the performance of the factorized model was better.
- The leak localization performance depends on the location of the leak in the network. A leak located in the looped part of the network was not localized during the night. In the looped part of the network the pressure drops at the locations of the sensors are too low to detect and they are easily disturbed by uncertainties in the model.
- The leak localization performance depends on the time of the day the leak localization is executed. The head loss caused by a leak is larger when there is more flow. It is shown that the factorized model was able to localize the leak in the looped part of the network during the morning peak, while it was unable to do so during other times.
- The leak localization performance improves when the number of leak localization time steps is increased and combined. The artificial leaks that were created lasted for 15 minutes during which one leak localization time step was possible. In this way, the performance is sensitive to abnormalities in the consumption in the network. The advantage of a real leak is that it continues the whole day. The performance of the leak localization of the real leak was fluctuating highly every time analysis step of 15 minutes. The outliers of bad performance got less when four time analysis steps were combined.
- The accumulation of time steps increases the robustness of the leak localization. The leak was localized within 35 meter when the nodes that have a high result are accumulated for one day (leak correlation >0.5). Despite fluctuations during the day, the end-result of both consumption distribution models (factorized and uniform) was good. The result is best when the factorized model is used that accumulates the hourly values. The leak is located 11 meter away from the real leak.

Recommendations to apply the leak localization in practice:

1. Start with the set-up of the model, based on a geographical database. It is important to update this database as good as possible when faults are discovered. Errors in especially diameter can have a large influence on the performance of the leak localization because these influence the pressure at the sensors. In this way, a leak can be masked by errors.
2. Apply the sensor deployment optimization.
3. In order to make fast automatic leak localization possible, fixed sensitive measurement devices at the inlet and inside a DMA are needed. Preferably meters that can send pressure and flow measurements digitally. When smart meters are used as inner pressure sensors, start with investigating who are interested.

4. The uncalibrated model can be used to set the elevation of the sensors in the model. In this way, it is not necessary to know the elevations of the pipes and sensors locations. The prerequisite is that the sensors are installed at least one night before the leak occurs.
5. Accumulate the leak correlations above 0.5 of multiple diagnosis sliding windows.
6. With the result (small search area), acoustic techniques can be used to pinpoint the leak.

The performance of the leak localization method was good in DMA Leimuiden. To get more insight in general applicability of the method, more research is needed on:

- The performance of the leak localization method of other locations that endure a whole day in DMA Leimuiden.
- The performance of leak localization of different locations in other DMAs with different pipe network structures.
- The minimum leak size that can be localized with the leak localization method.
- The need of the accumulation of one whole day when leak sizes are larger.
- The performance of the leak localization on days the consumption pattern differs from weekdays (e.g. weekend, holidays, vacation). The leak localization performance highly depends on the ability of a model to predict the consumption distribution.
- A combination with other models. Pipes are more than only lines in a model. When two locations have the same probability of having a leak, there can be a preference for the location that has a higher chance on leaks based on for example pipe age and subsidence.
- The performance when two or more inlets are used.

In case the predicted leak location does not match well with the actual leak location, the leak localization performance is possibly improved by improving the model and changing the time analysis steps of the leak localization method. The model might improve by:

- A more detailed consumption distribution. Different patterns can be used for different types of customers (domestic, industrial, agricultural etc.). Another possible improvement is to use SIMDEUM. This uses different statistics to predict flow (Blokker, 2010).
- The parameter optimization. In this research only roughness in pipes are used while a combination with local losses might be better.
- Using a pressure-dependent leak in the model instead of a constant leak size over the day.

The leak localization method might improve by:

- A smaller time analysis step. In this research the smallest time step was bounded by the flow meter. The consumption varies much and a smaller time step might be better.
- A different number of time analysis steps that are accumulated. Maybe longer or shorter results in a better performance.
- Leave out certain times that have a bad performance. This can be times that the model is not able to predict the flows well.

A trustful smaller search area is helpful to localize leaks faster. The leak can be found before it possibly enlarges and causing supply interruptions and damage. This ultimately saves water and money.

REFERENCES

Abdel-Monim YK, Ead SA and Shabayek SA (2005) Effect of time on pipe roughness. In: *17th Canadian Hydrotechnical Conference; Edmonton, Alberta, Canada*.

Arunkumar M and Nethaji Mariappan VE (2011) Water demand analysis of municipal water supply using epanet software. *International Journal on Applied Bioengineering*. 5 (1), 9–19.

Bakker M (2014) *Optimised control and pipe burst detection by water demand forecasting*. (Doctoral dissertation). TU Delft, Delft, Netherlands.

Blokker EJM (2010) *Stochastic water demand modelling for a better understanding of hydraulics in water distribution networks*. (Doctoral dissertation). TU Delft, Delft, Netherlands.

Bonada E, Meseguer J and Mirats-Tur JM (2014a) On The Structure Of The Objective Function For A Pressure Sensor Placement Optimizing Methodology Based On Genetic Algorithms Applied To Model-Based Leakage Localization In Distribution Water Networks. *11th International Conference on Hydroinformatics (2014).*

Bonada E, Meseguer J and Mirats-Tur JM (2014b) Practical-Oriented Pressure Sensor Placement For Model-Based Leakage Location In Water Distribution Networks. *11th International Conference on Hydroinformatics (2014).*

Pipelife Nederland B.V. (2001) PVC drukleidingen: verwerkingsrichtlijnen voor PVC drukleidingen.

Casillas MV, Garza-Castañón LE, Puig V and Vargas-Martinez A (2015) Leak signature space: an original representation for robust leak location in water distribution networks. *Water*. 7 (3), 1129–1148.

CBS (2015) *Kerncijfers wijken en buurten 2014*. [Online] Available at: http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=82931NED&D1=0,3,22-26,45-61,99,121&D2=0,802-840,14224-14227&HDR=T&STB=G1&VW=T (accessed 10/01/16).

Chadwick A, Morfett J and Borthwick M (2013) *Hydraulics in civil and environmental engineering*. Crc Press.

Clemens FHLR (2001) *Hydrodynamic models in urban drainage application and calibration*. (Doctoral dissertation). TU Delft, Delft, Netherlands.

EPA (2005) *Water Distribution System Analysis: Field Studies, Modeling and Management*. U.S. Environmental Protection Agency. EPA/600/R-06/028. United States.

Farina G, Creaco E and Franchini M (2014) Using EPANET for modelling water distribution systems with users along the pipes. *Civil Engineering and Environmental Systems*. 31 (1), 36–50.

Farley M, Wyeth G, Ghazali ZM, Istandar A, Singh S, Dijk N, Raksakulthai V and Kirkwood E (2008) *The manager's non-revenue water handbook: a guide to understanding water losses*. Ranhill Utilities Berhad and the United States Agency for International Developement, Bangkok, Thailand.

Fortin F-A, Rainville D, Gardner M-AG, Parizeau M and Gagné C (2012) DEAP: Evolutionary algorithms made easy. *The Journal of Machine Learning Research*. 13 (1), 2171–2175.

Gao Y, Brennan M, Joseph P, Muggleton J and Hunaidi O (2005) On the selection of acoustic/vibration sensors for leak detection in plastic water pipes. *Journal of Sound and Vibration*. 283 (3), 927–941.

Gertler J (1998) *Fault detection and diagnosis in engineering systems*. Marcel Dekker, Inc. New York.

Hatchett S, Boccelli D, Haxton T, Janke R, Kramer A, Matracia A, Panguluri S and Uber J (2010) How accurate is a hydraulic model. In: *Proceedings of the 2010 Water Distribution Systems Analysis Symposium, Tucson, Ariz*.

Kadaster (2015) *Wat is de BAG?* [Online] Available at: http://www.kadaster.nl/web/Themas/Registraties/BAG/BAGartikelen/Wat-is-de-BAG.htm.

Kleiner Y and Rajani B (2001) Comprehensive review of structural deterioration of water mains: statistical models. *Urban water*. 3 (3), 131–150.

Lambert A and Hirner W (2000) Losses from water supply systems: standard terminology and recommended performance measures. *IWA Blue Pages, International Water Association*.

Li R, Huang H, Xin K and Tao T (2015) A review of methods for burst/leakage detection and location in water distribution systems. *Water Science and Technology: Water Supply*. 15 (3), 429–441.

Mense P (2015) Lekverliezen opsporen in Nederlandse waterleidingen. *H20*. 6-3-2015.

Meseguer J, Mirats-Tur JM, Cembrano G and Puig V (2015) Model-based Monitoring Techniques for Leakage Localization in Distribution Water Networks. *Procedia Engineering*. 119, 1399–1408.

Meseguer J, Mirats-Tur JM, Cembrano G, Puig V, Quevedo J, Pérez R, Sanz G and Ibarra D (2014) A decision support system for on-line leakage localization. *Environmental Modelling \& Software*. 60, 331–345.

Mirats-Tur JM, Jarrige P-A, Meseguer J and Cembrano G (2014) Leak detection and localization using models: field results. *Procedia Engineering*. 70, 1157–1165.

Oasen (2015a) *Eerste successen met DMA's in voorzieningsgebied Oasen behaald*. [Online] Available at: https://www.oasen.nl/nieuws/Paginas/Eerste-successen-met-DMA-s-in-voorzieningsgebied-Oasen-behaald.aspx (accessed 23/05/15).

Oasen (2015b) *Onderzoek met slimme watermeters bij Oasen*. [Online] Available at: https://www.oasen.nl/drinkwater-thuis/Paginas/slimme-watermeter.aspx (accessed 15/09/15).

Ofwat (2007) *International comparison of water and sewerage service 2007 report.* www.ofwat.gov.uk.

Pérez R, Cugueró M-A, Cugueró J and Sanz G (2014) Accuracy assessment of leak localisation method depending on available measurements. *Procedia Engineering*. 70, 1304–1313.

Pérez R, Nejjari F, Puig V, Quevedo J, Sanz G, Cugueró M and Peralta A (2011a) Study of the isolability of leaks in a network depending on calibration of demands. In: *11th International Conference on Computing and Control for the Water Industry*.

Pérez R, Puig V, Pascual J, Peralta A, Landeros E and Jordanas L (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. *Water science and technology: water supply*.

Pérez R, Puig V, Pascual J, Quevedo J, Landeros E and Peralta A (2011b) Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*. 19 (10), 1157–1167.

Pérez R, Quevedo J, Puig V, Nejjari F, Cugueró M, Sanz G and Mirats J (2011c) Leakage isolation in water distribution networks: A comparative study of two methodologies on a real case study. In: *Control & Automation (MED), 2011 19th Mediterranean Conference on*. 138–143.

Pérez R, Sanz G, Puig V, Quevedo J, Cuguero Escofet MA, Nejjari F, Meseguer J, Cembrano G, Mirats Tur JM and Sarrate R (2014a) Leak localization in water networks: A model-based methodology using pressure sensors applied to a real

network in barcelona [applications of control]. *Control Systems, IEEE*. 34 (4), 24–36.

Pudar RS and Liggett JA (1992) Leaks in pipe networks. *Journal of Hydraulic Engineering*. 118 (7), 1031–1046.

Puust R, Kapelan Z, Savic D and Koppel T (2010) A review of methods for leakage management in pipe networks. *Urban Water Journal*. 7 (1), 25–45.

Python (2016) *What is Python? Executive summary*. [Online] Available at: https://www.python.org/doc/essays/blurb/ (accessed 04/01/16).

Quevedo J, Cugueró M, Pérez R, Nejjari F, Puig V and Mirats J (2011) Leakage location in water distribution networks based on correlation measurement of pressure sensors. In: *IWA Symposium on Systems Analysis and Integrated Assessment, San Sebastián*.

Quevedo J, Pérez R, Pascual J, Puig V, Cembrano G, Peralta A and others (2012) Methodology to detect and isolate water losses in water hydraulic networks: Application to Barcelona water network.

Romano M, Kapelan Z and Savi'c DA (2012) Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management*. 140 (4), 457–467.

Rosario-Ortiz F, Rose J, Speight V, von Gunten U and Schnoor J (2016) How do you like your tap water? *Science*. 351 (6276), 912–914.

Rossman LA (2000) EPANET 2 Users Manual, US Environmental Protection Agency. *Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH*. 45268.

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M and Tarantola S (2008) Introduction to sensitivity analysis. *Global Sensitivity Analysis. The Primer*, 1–51.

Saltelli A, Tarantola S, Campolongo F and Ratto M (2004) *Sensitivity analysis in practice: a guide to assessing scientific models*. John Wiley & Sons.

Sanz G and Pérez R (2014) Parameterization and sampling design for water networks demand calibration using the singular value decomposition: Application to a real network. In: *Proc. 11th International Conference on Hydroinformatics, New York*.

Steffelbauer D, Neumayer M, Günther M and Fuchs-Hanusch D (2014) Sensor Placement and Leakage Localization Considering Demand Uncertainties. *Procedia Engineering*. 89, 1160–1167.

Tao T, Huang H, Li F and Xin K (2013) Burst detection using an artificial immune network in water-distribution systems. *Journal of Water Resources Planning and Management*. 140 (10).

Todlini E and Pilati S (1987) A gradient method for the analysis of pipe networks. In: *Proceedings of the international conference on computer applications for water supply and distribution, Leicester Polytechnic, UK*.

Tyleenslang.nl (2015) *ZPE buis Tyleen slang*. [Online] Available at: https://www.tyleenslang.nl/tyleen/ (accessed 05/07/15).

Vewin (2013) *Water in zicht 2012, Bedrijfsvergelijking drinkwatersector*. Accenture.

Van der Zon N (2011) Kwaliteitsdocument ahn-2. *Rijkswaterstaat & Waterschappen, Tech. Rep*. 1.

LIST OF TABLES

LIST OF FIGURES

APPENDIX A PICTURE ERODED LEAK

Figure A.1 Example of enlarged leak due to moving ground particles

APPENDIX B DARCY-WEISBACH ROUGHNESS CONVERSION

The Darcy-Weisbach roughness coefficient is set based on material and pipe age. The year of construction, to determine the pipe age, is not always precisely known and sometimes a range up to 10 years is given in DiaGIS. In the conversion to InfoWorks, automatically the first year in range is taken.

An overview of the pipe materials (Asbestos Cement, Cast Iron (Gietijzer), Hard Polyethyleen (High-Density Polyethylene, HDPE), PVC, Zacht Polyethyleen (low-density polyethylene, LDPE) is given in Table B.1. The length of pipe that is constructed every year is shown in Table B.2.

Table B.1 Overview pipe materials

Material	Length (m)	% of Total
PVC	14037.64	54.17%
HPE	8310.91	32.07%
ZPE	1969.74	7.60%
AC	1457.81	5.63%
ST	136.96	0.53%
GIJ	0.3	0.00%
Total	25913.36	100.00%

Table B.2 Year of construction of pipes

Year of construction	Pipe Length [m]
1960	2224
1970	6299
1975	2308
1980	6630
1981	32
1983	22
1985	1655
1988	54
1990	2082
1992	141
1993	70
1995	527
1996	797
2001	89
2002	517
2004	26
2009	133
2010	891
2011	186
2012	244
2013	798
2015	190
Total	*25913*

In InfoWorks an automatic lookup function is used to set the roughness coefficient. For every material the k-values for new pipes, 10-year old pipes and 40-year old pipes are set (Table B.3). Typical values from Chadwick et al.

(2013) are used to set the k-values of new pipes. The effect of time is different for different materials. The effect is higher if pipes are likely to rust and can result in forty times its original value. This in contrary to plastic pipes that show a minimum response (1.5 times its original value) (Abdel-Monim et al., 2005). The initial estimate that is used in the model is presented in Table B.4.

Table B.3 Roughness lookup table

Material	Age [year]	Darcy-Weisbach roughness coefficient [mm]
AC	0-10	0.10
	10-40	0.15
	>40	1.20
All plastics: HPE, PVC, ZPE	0-10	0.05
	10-40	0.10
	>40	0.15
ST	0-10	0.30
	10-40	1.00
	>40	2.00

Table B.4 Initial estimate roughness coefficient

Darcy-Weisbach roughness coefficient [mm]	Pipe length [m]	% of Total
0.03	0.3	0.00%
0.05	2440.02	9.42%
0.1	11115.09	42.89%
0.15	10271.14	39.64%
0.2	492.04	1.90%
1	49.21	0.19%
1.2	1457.81	5.63%
2	87.75	0.34%
Total	25913.36	100.00%

APPENDIX C DIAMETER CONVERSION

The diameter assigned to a pipe in DiaGIS represented sometimes the inner diameter, the outer diameter, internal diameter times wall thickness or outer diameter times wall thickness. All diameters are converted to internal diameters. For Asbestos Cement (AC) pipes it is assumed that the first value is the inner diameter and the second the wall thickness. For all other materials this is assumed the outer diameter with the second value as wall thickness. Even for pipes that are probably constructed in the same time, the diameter is sometimes represented as 150 and sometimes 150x13. The second value is either forgotten or unknown.

For AC there are some values with one and some with two values. All the single and first values are denoted as inner diameter. For cast iron only one value is given. There is only 0.3 meter of cast iron and the single value is used as inner diameter. For HPE almost all outer diameters and wall thicknesses are known. For a small amount the wall thickness of 50 mm pipes are missing. In other pipes of HPE the pipes with 50 mm have a value of 4.6 mm so this is assumed to be the same for the unknown ones.

For PVC only the 32 mm pipes have one value. The wall thickness is taken from a manufacturer (Pipelife Nederland B.V., 2001). For Steel there were some missing value for 168.3 and these are taken from the existing steel pipes of 168.3 mm. For ZPE a wall thickness of 3.5mm is set for pipes of 32 mm, 4.3 mm for 40 mm and 5.4 for 50 mm pipes (Tyleenslang.nl, 2015). The wall thickness for 63 mm pipes is taken from the pipe materials book of Oasen.

Table C.1 Pipe length sorted by materials and inner diameter

Inner Diameter [mm]	Length [m]	Inner Diameter [mm]	Length [m]
PVC	***14037.64***	***ZPE***	***1969.74***
32	6.36	32	155.14
70.6	64.92	40	362.2
103.6	11190.29	50	1089.25
104.6	154.51	63	363.15
110	1.55	***AC***	***1457.81***
150.6	1537.23	80	33.54
152	12.61	100	580.17
188.2	913.78	150	844.1
296.6	156.39	***ST***	***136.96***
HPE	***8310.91***	108	119.19
26	121.72	154.1	17.77
32.6	172.82	***GIJ***	***0.3***
40.8	718.15	100	0.3
42.8	206.72		
50	218.71		
51.4	113.66		
72.8	117.82		
73.6	306.55		
90	4526.57		
97.4	354.99		
141.8	121.4		
163.6	921.65		
177.2	253.57		
257.6	156.58		

Figure C.1 Inner diameters DMA Leimuiden

APPENDIX D NETWORK CALCULATIONS

A simple drinking water network without pumps, consists of n_p pipes, n_j ($<n_p$) junction nodes and n_f fixed-head nodes. The heads are known at the fixed-head nodes. To calculate the other heads at the nodes and flow in all pipes, three basic relations are used (EPA, 2005):

- Conservation of Flow
- Pipe Friction Head loss
- Conservation of Energy

The conservation of flow results in the flow continuity equation for every node:

$\sum_j Q_{ij} - D_i = 0$ for i = 1,….N

D_i is the flow demand at node I (flow into a node is positive).

The head loss over a pipe between nodes *i* and *j,* can be calculated with:

$$H_i - H_j = h_{ij} = rQ_{ij}^2 + mQ_{ij}^2$$

H=nodal head
h=head loss
r=resistance coefficient
Q=flow rate
m=minor loss coefficient

With the fixed grade nodes a solution is sought that satisfies the head loss relation and flow continuity equation. In Epanet, Todini's approach is used to solve the flow continuity and head loss equations. This approach is also called the 'Gradient Method' and more information can be found in Rossman (2000) and Todlini & Pilati (1987).

For this research the Darcy-Weisbach formula is used in Epanet to calculate the head loss:

$$h_l = 0.0252 \frac{f(\varepsilon, d, q)}{d^5} Lq^2$$

With:

h_l = head loss

f = friction factor (dependent on ε, d and q)

ε = Darcy-Weisbach roughness coefficient

d = pipe diameter

q = flow rate

L = pipe length

Epanet uses different methods to compute the friction factor f. This depends on the flow regime:

- The Hagen-Poiseuille formula is used for laminar flow (Re<2000):

$$f = \frac{64}{Re}$$

- A cubic interpolation from the Moody Diamgram is used for transitional flow (2000<Re<4000):

$$f = \Big(X1 + R\big(X2 + R(X3 + X4)\big)\Big)$$

$$R = \frac{Re}{2000}$$

$$X1 = 7FA - FB$$
$$X2 = 0.128 - 17FA + 2.5FB$$
$$X3 = -0.128 + 13FA - 2FB$$
$$X4 = R(0.032 - 3FA + 0.5FB)$$
$$FA = (Y3)^{-2}$$
$$FB = FA\left(2 - \frac{0.00514215}{(Y2)(Y3)}\right)$$
$$Y2 = \frac{\varepsilon}{3.7d} + \frac{5.74}{Re^{0.9}}$$
$$Y3 = -0.86859\, Ln\left(\frac{\varepsilon}{3.7d} + \frac{5.74}{4000^{0.9}}\right)$$

- The Swamee and Jain approximation to the Colebrook-White equation is used for fully turbulent flow (Re>4000)

$$f = \frac{0.25}{\left[Ln\left(\frac{\varepsilon}{3.7d} + \frac{5.74}{Re^{0.9}}\right)\right]^2}$$

Reynolds number:

$$Re = \frac{vD}{\vartheta}$$

With v = velocity, D = diameter and ϑ = kinematic viscosity. Kinematic viscosity is kept constant at $1.004*10^{-6}$.

Example head loss night and morning peak (with only resistance coefficient):

<u>Night:</u>
Normal flow = 4 m^3/h, $h_{ij}=r4^2=r16$
Normal flow + Leak (5 m^3/h) = 9 m^3/h, $h_{ij}=r9^2=r81$
$\Delta h_{ij\_night}=r65$

<u>Morning peak:</u>
Normal flow = 45 m^3/h, $h_{ij}=r45^2=r2025$
Normal flow + Leak (5 m^3/h) = 50 m^3/h, $h_{ij}=r50^2=r2500$
$\Delta h_{ij\_peak}=r475$

Difference head loss night and morning peak with same leak size: 410r

APPENDIX E NODE DENSITIY HEAT MAP

Figure E.1 Node density heat map. Darker colors indicate a higher density of nodes.

APPENDIX F FIRE HYDRANT LOCATIONS

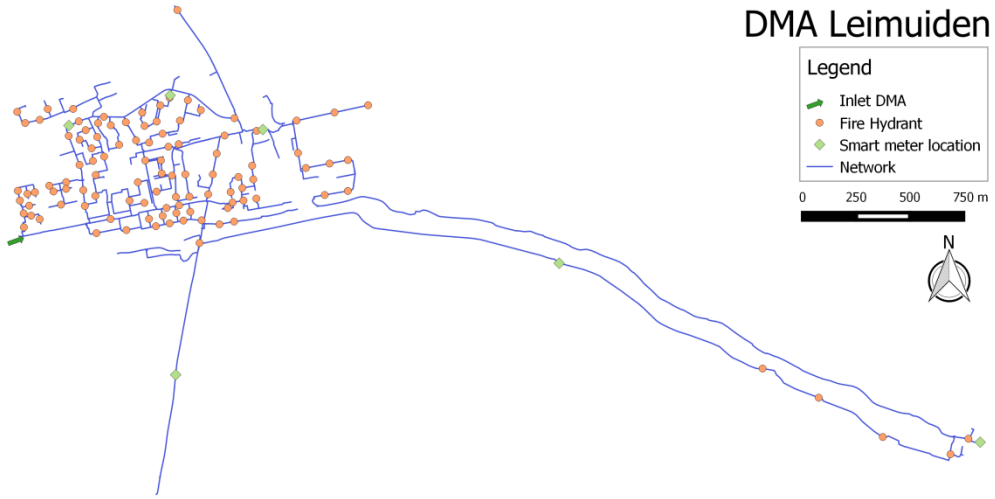

Figure F.1 Fire hydrants (possible sensor locations for pressure loggers) in DMA Leimuiden

APPENDIX G PICTURES ARTIFICIAL LEAK LOCATIONS

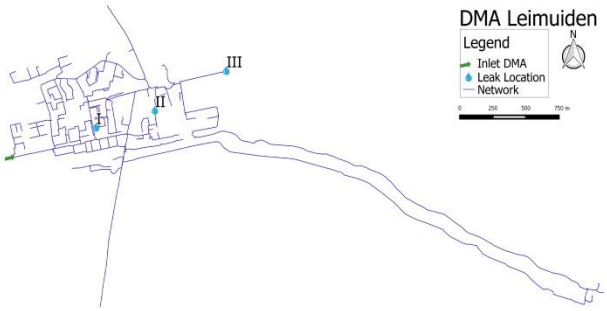

Figure G.1 Pictures artificial leak locations

APPENDIX H CALIBRATION MODEL FOR SENSOR DEPLOYMENT OPTIMIZATION

An overview of the locations of the smart meters and pressure loggers is given in Figure H.1.

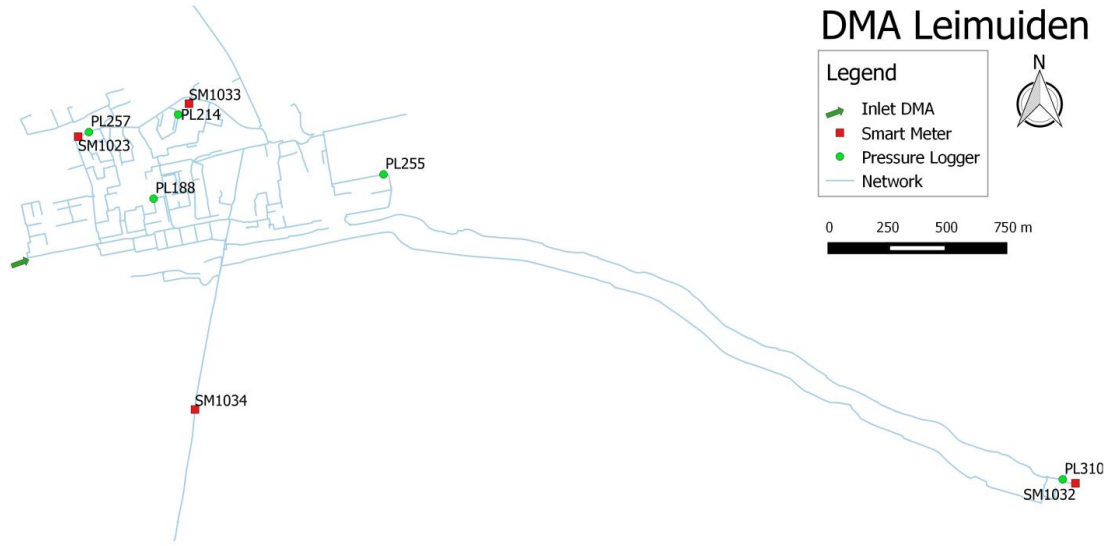

Figure H.1 Measurement locations during calibration

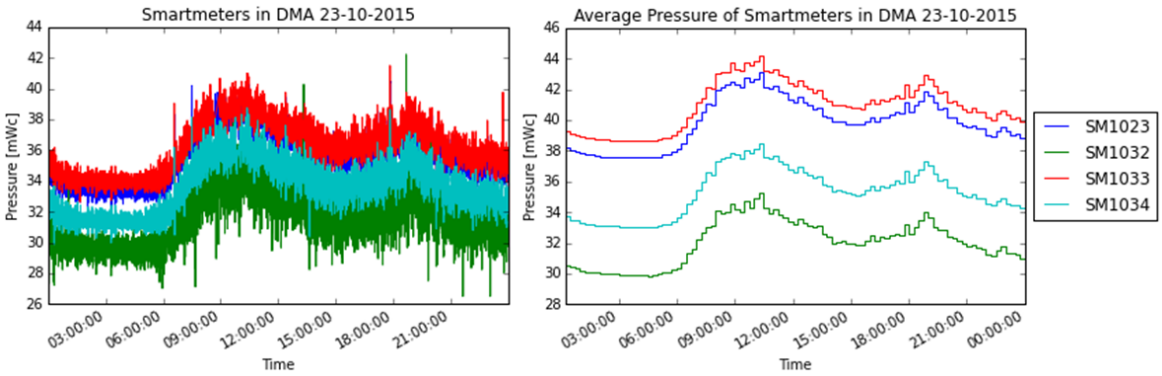

Figure H.2 Pressure measurement smart meters. Left: 5 second average. Right: 15 minutes average.

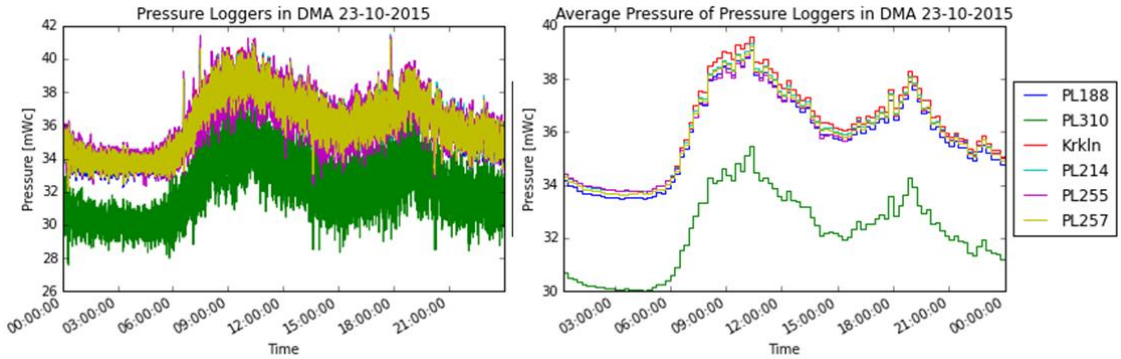

Figure H.3 Pressure measurements pressure loggers. Left: 1 second average Right: 15 minutes average.

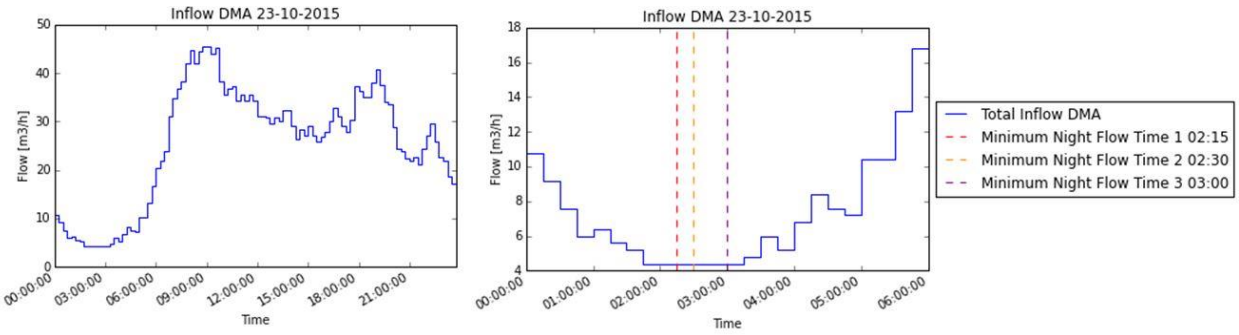

Figure H.4 Inflow DMA (2015-10-23). Left: Inflow DMA 2015-10-23. Right: MNF 2015-10-23.

The measurements of the smart meters (Figure H.2) and pressure loggers (Figure H.3) are used to calibrate the model. The elevation is corrected for three low flow times (02:15,02:30 and 03:00, see Figure H.4). There was not much deviation between these times and the average is taken (Table H.1).

Table H.1 Elevation corrections

Meter	Elev. corr. Time 1 [m]	Elev. corr. Time 2 [m]	Elev. corr. Time 3 [m]	Avg. elev. corr.	Standard deviation
PL188	4.29	4.29	4.29	4.29	0.00
PL310	4.06	4.08	4.08	4.07	0.01
PL214	4.31	4.32	4.32	4.32	0.00
PL255	4.39	4.40	4.40	4.40	0.01
PL257	4.25	4.26	4.25	4.25	0.00
SM1023	-3.78	-3.78	-3.78	-3.78	0.00
SM1032	3.82	3.84	3.86	3.84	0.02
SM1033	-4.89	-4.87	-4.87	-4.88	0.01
SM1034	0.71	0.74	0.73	0.72	0.01

To create the objective function, a model is used that continued if the system was unbalanced and divided the leak size in the leak free situation as consumption. The sum of least squares is lowest when a multiplication factor of 80.4 is used. The difference in least squares with the two manners of roughness optimization is only 0.13 Δp^2 (5.6 Δp^2, 5.73 Δp^2 with using the multiplication factor, system roughness respectively, see Figure H.5).

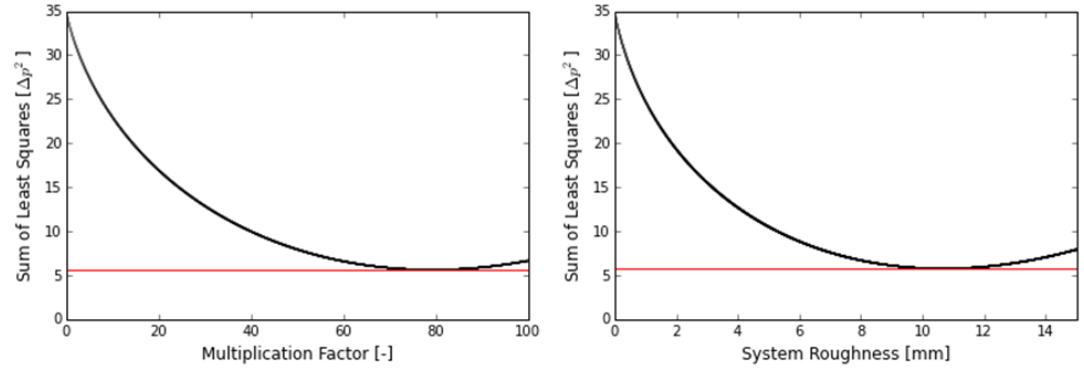

Figure H.5 Objective function roughness optimization. Left: multiplication factor. Right: system roughness.

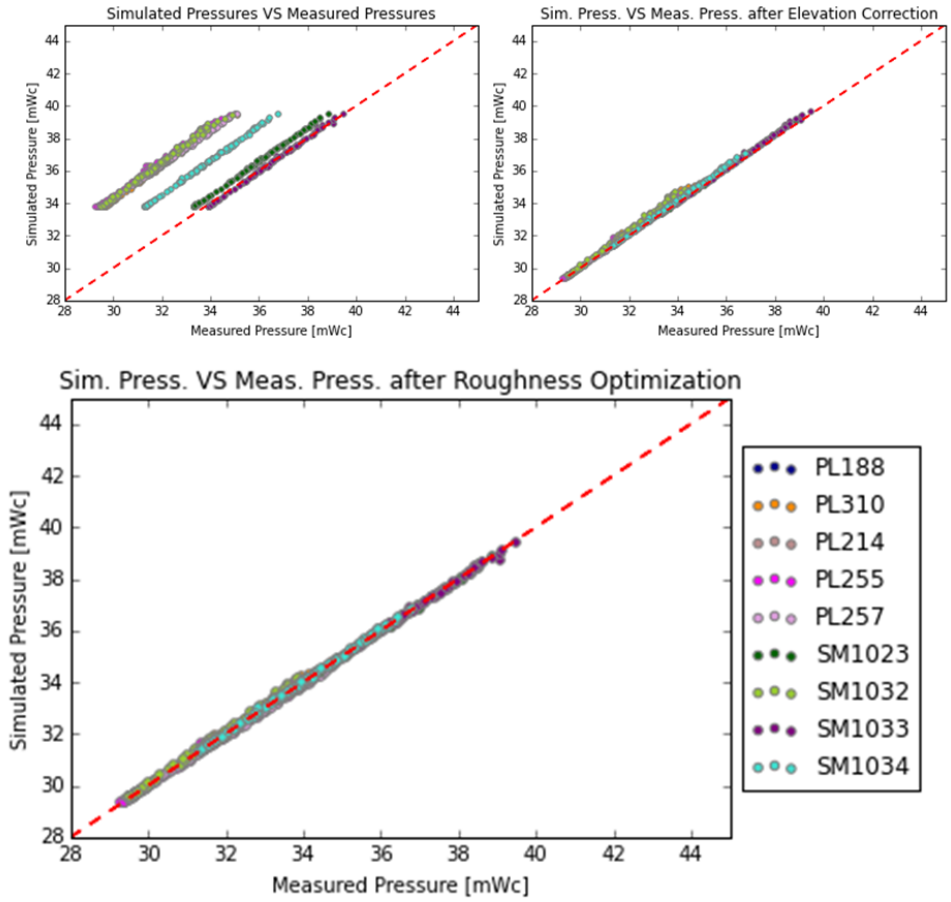

Figure H.6 Simulated pressures versus measured pressures (top left) after elevation correction (top right) and roughness optimization (bottom)

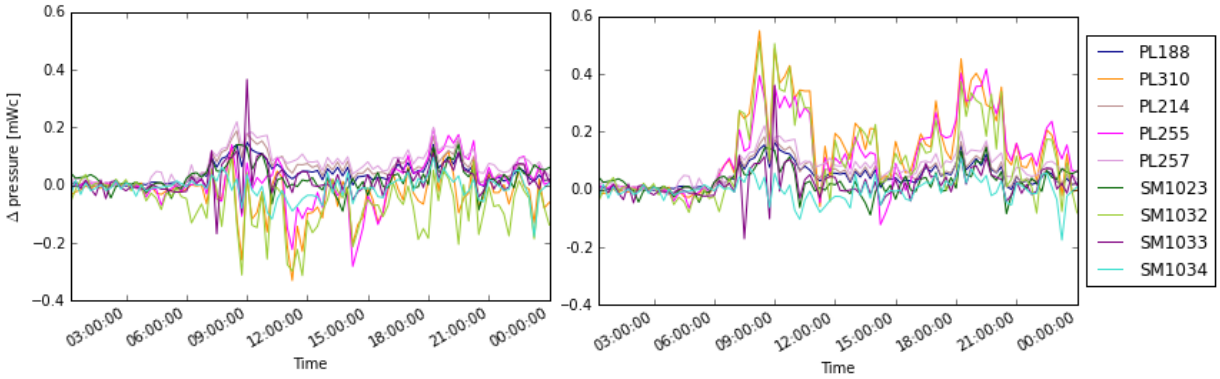

Figure H.7 Residuals between measured values and simulated values after calibration (2015-10-23). Left: residuals uniform model. Right: residuals factorized model.

Table H.2 RMSE uniform model used for sensor deployment

Meters	RMS Error	RMS Error after Elevation set	RMS Error after Roughness optimization
PL188	4.36	0.09	0.06
PL310	4.29	0.28	0.08
PL214	4.40	0.10	0.07
PL255	4.58	0.24	0.08
PL257	4.30	0.07	0.09
SM1023	0.53	0.12	0.05
SM1032	4.46	0.31	0.11
SM1033	0.11	0.15	0.06
SM1034	2.58	0.17	0.04
Total	*29.61*	*1.53*	*0.64*

APPENDIX I SENSOR DEPLOYMENT OPTIMIZATION AND GROUPS

In this research the optimal pressure sensor placement was performed in the situation where the largest groups are expected (during minimum night flow with a leak size of 7.5 m^3/h and the uniform model). The result is different for different operating points in the network. To take this into account in the future, the FSM_{ext} can be constructed by appending FSMs of different scenarios (i.e. day-night, seasonality) [meseguer2015model].

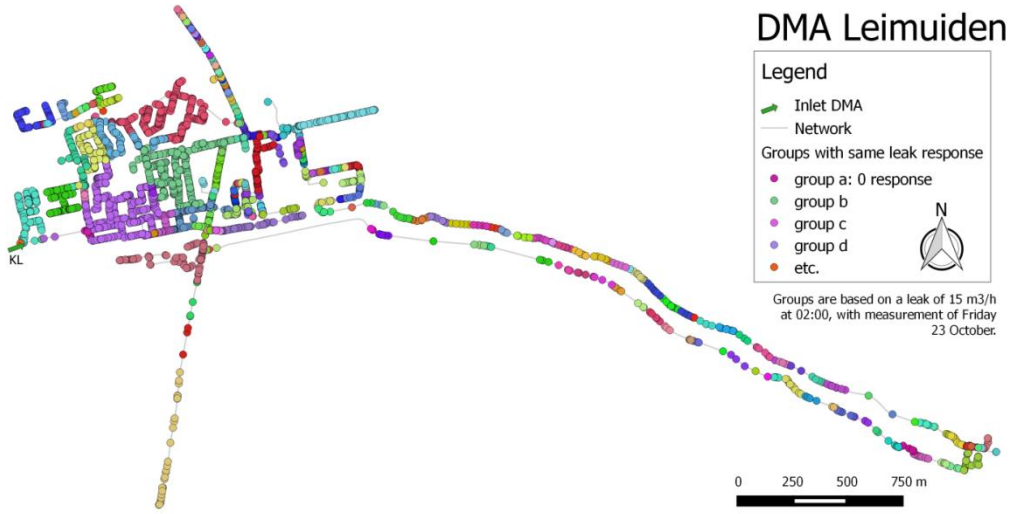

Figure I.1 Result groups with a leak of 15 m^3/h during minimum flow. Groups when the sensor placement optimization with a leak of 7.5 m^3/h during minimum flow is used.

Figure I.2 Result sensor placement optimization with a leak of 7.5 m^3/h during maximum flow. Top: groups with nodes that have the same leak response. Bottom: proposed location of measurement devices.

APPENDIX J CALIBRATION MODELS FOR LEAK LOCALIZATION

Table J.1 RMSE uniform model and factorized model

	Uniform Model			Factorized Model		
Meter	Initial RMSE [mWc]	RMSE after Elevation Correction [mWc]	RMSE after Roughness Set [mWc]	Initial RMSE [mWc]	RMSE after Elevation Correction [mWc]	RMSE after Roughness Set [mWc]
PL196	0.26	0.10	0.03	0.26	0.10	0.05
PL202	3.80	0.08	0.02	3.80	0.08	0.03
PL222	0.18	0.09	0.02	0.18	0.09	0.04
PL255	0.17	0.21	0.13	0.13	0.16	0.11
PL277	3.29	0.09	0.11	3.29	0.09	0.10
PL279	0.04	0.04	0.08	0.04	0.04	0.06
PL319	3.74	0.11	0.16	3.74	0.11	0.14
SM1023	3.91	0.06	0.05	3.91	0.06	0.04
SM1025	2.89	0.11	0.05	2.88	0.10	0.05
SM1027	1.71	0.17	0.09	1.67	0.13	0.08
SM1032	3.85	0.22	0.14	3.78	0.16	0.11
SM1033	4.96	0.07	0.05	4.97	0.06	0.04
SM1034	0.61	0.12	0.06	0.61	0.12	0.07
Model Total	*29.41*	*1.47*	*0.99*	*29.26*	*1.30*	*0.92*

Uniform model

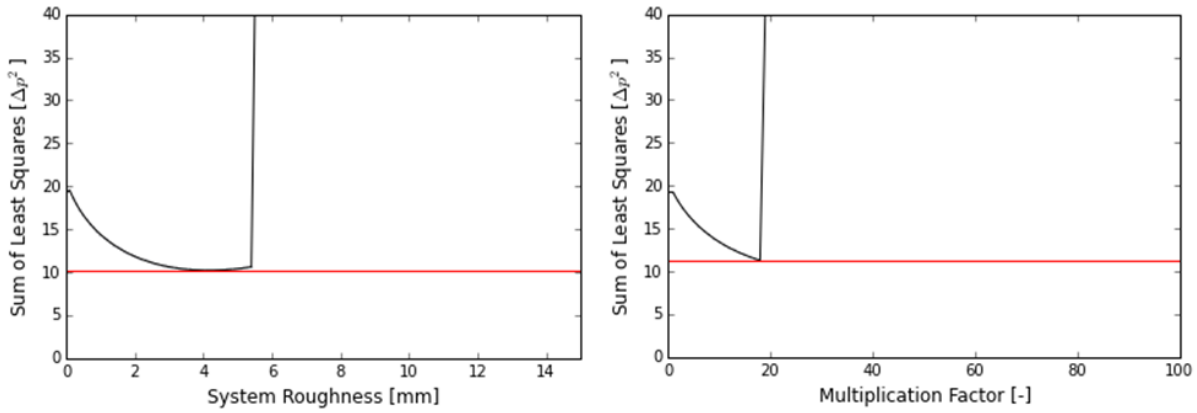

Figure J.1 Objective function roughness optimization uniform model. Left: multiplication factor. Right: system roughness.

For the uniform model the roughness is set to 4.14 mm (sum of least squares: 10.21 Δp^2). The sum of least squares with the multiplication factor (18) is higher (11.28 Δp^2).

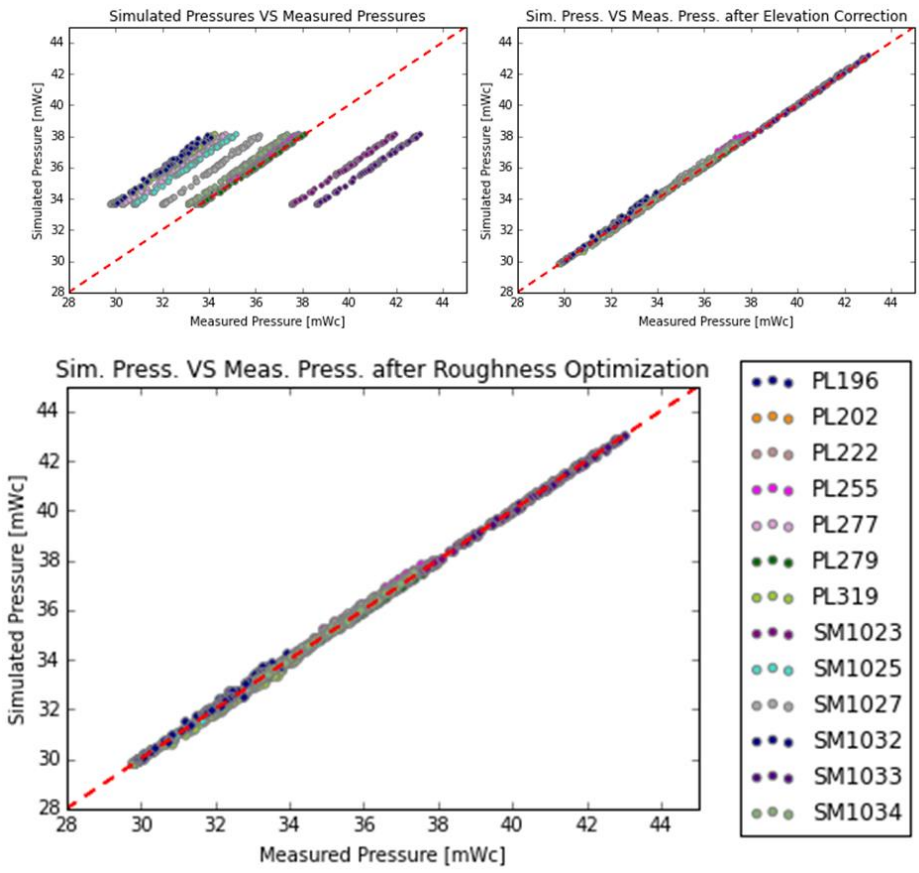

Figure J.2 Simulated pressures versus measured pressures (top left) after elevation correction (top right) and roughness optimization (bottom), uniform model

Factorized model

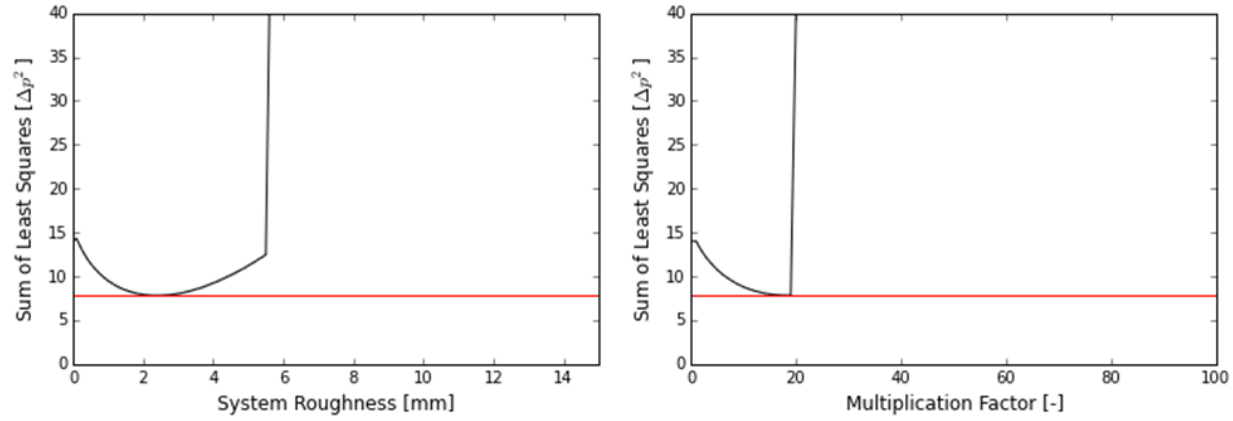

Figure J.3 Objective function roughness optimization factorized model. Left: multiplication factor. Right: system roughness.

The roughness in the factorized model is set to 2.39 (sum of least squares is 7.81 Δp^2) with multiplication factor 18 this is 7.83 Δp^2.

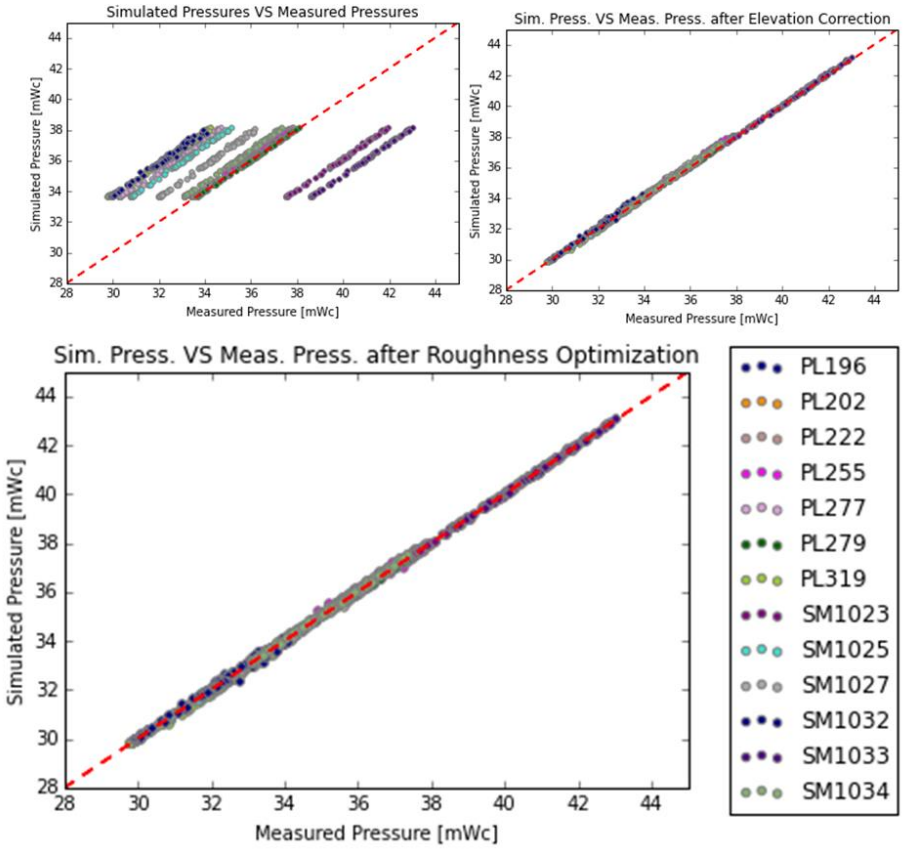

Figure J.4 Simulated pressures versus measured pressures (top left) after elevation correction (top right) and roughness optimization (bottom), factorized model

It is recommended to perform the leak localizaiton during the night because the pressures are more stable when there is less consumption (Pérez et al., 2009). Figure J.5 (left), however, shows the pressure measurement (fluctuations) during leak testing he night. Fluctuations in pressure are caused by changing pressures coming from the pumping station outside the DMA (Figure J.5: right), the leak and changes by different consumptions within the DMA.

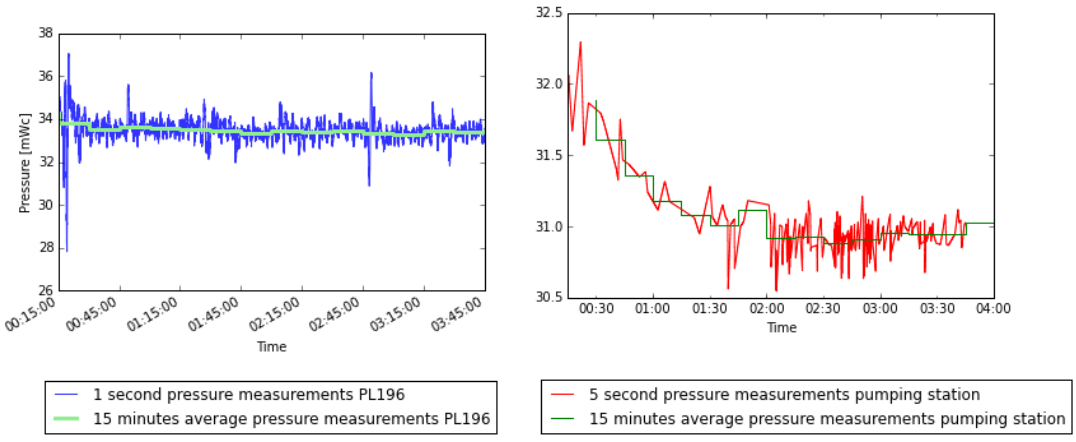

Figure J.5 Example pressure measurements during leak testing. Left: pressure measurement from a pressure logger. Right: pressure measurements from the pumping station.

APPENDIX K UNCERTAINTY AND SENSITIVITY ANALYSIS CONSUMPTION MODELS

The quality of the outcome of a model is directly related to the quality of the input of the model and can have an influence on the performance of the leak localization method. During calibration only the model-parameter roughness is estimated. All other input parameters (diameter, length, consumption factor) are considered to be true and are not changed during the calibration process. In the end of last century, Oasen digitalized the technical drawings of the pipes and stored them in DiaGIS. Errors in translating values, regarding placement year, material and diameter, can easily be made. Furthermore, are there values that are guessed when the data was missing. Additionally there can be errors in the drawings in for example the location. Currently GPS devices are used to locate the pipes, however, it used to be manual tape measurements. Lastly, are extern contractors that are not checked or did not get exact instructions of what to install. It is unsure if the pipes are actual installed at the places indicated on drawings.

In order to look at the effect of the uncertainties in diameter, length, roughness and consumption factor, these parameters are taken into account in the uncertainty and sensitivity analysis. Many different definitions and approaches are used for sensitivity analysis but in this research the definition of Saltelli et al. (2004) is used. Sensitivity analysis is: '*The study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input* (Saltelli et al., 2004).' 'The model' is the hydraulic model with the purpose to predict the behavior of the network during a leak. The behavior is translated to the pressure throughout the system. The output, are the pressures at the location of the pressure devices. Input is defined as everything that can vary the output of the model (Saltelli et al., 2008). The influence of total uncertainty due to all parameters is studied in the uncertainty analysis. For the uncertainty analysis the total uncertainty of the output is calculated based on the uncertainty of the following parameters: diameter, length, roughness and consumption factor. In the sensitivity analysis the influence of the separate input parameters is studied (one-at-a-time analysis).

The influence on pressure is calculated at the nodes where pressure measurement devices were located for calibration (Figure K.1). The influence of the error is explored on minimum, maximum and average flow of one day. Changes in flow regime over the day result in different friction factors.

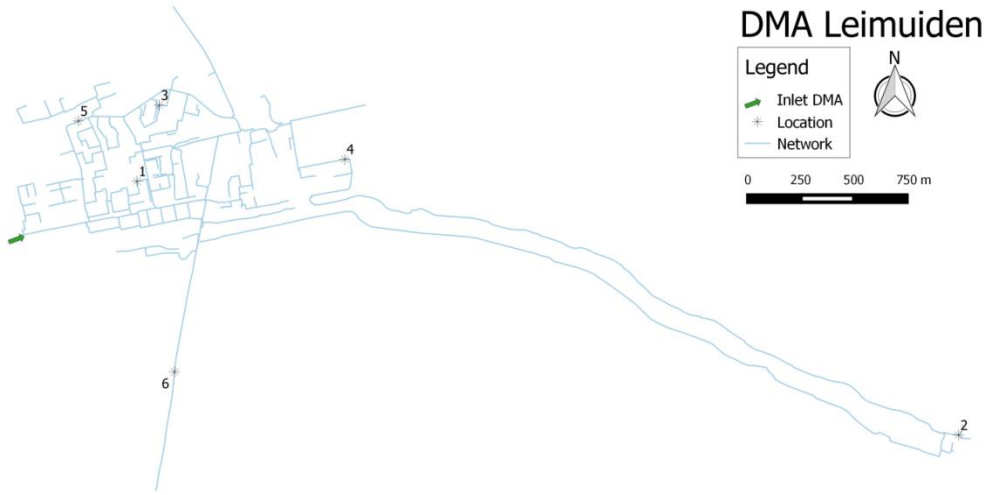

Figure K.1 Uncertainty and sensitivity analysis locations

The formula for head loss is given in Equation 14 (page 14). The influence of an error in a single pipe is rather straightforward and the inputs to calculate the head loss in a pipe are length, diameter, roughness and flow. In a

looped system, however, these errors are different depending on the location in the network. The model input parameters are still the same, except that it is for all pipes (3243) a different (combination of) diameter, length and roughness and for all nodes (1891) a different consumption factor. The consumption factor is directly related to the flow.

The uncertainty analysis is executed by performing a Monte Carlo analysis (Saltelli et al., 2008). During a Monte Carlo analysis, model simulations are performed with different input parameters. It is assumed that the input parameters come from a normal distribution with their current estimation being the mean.

For the Monte Carlo analysis, 1000 simulations are performed with different input parameters. It is assumed that the current value is with a 95% certainty within +/- 10% of the current value. Every original value of every parameter for all pipes and nodes is multiplied by a factor taken from the normal distribution with mean one and standard deviation of 0.05. The output, the pressure at the measurement devices, is saved for all outputs. An example output at one location is shown in Figure K.2. From all outputs the standard deviations are computed. The standard deviation is a measure for the total uncertainty of the output. Next to the standard deviation the total pressure range is given. This is the difference between the maximum and minimum pressure value. This is not the total pressure range possible since this is can easily vary when a new Monte Carlo analysis is performed.

The sensitivity analysis again contains a Monte Carlo analysis. During, the uncertainty analysis all factors where changed simultaneously. For the sensitivity analysis only one factor at a time is varied with a factor from the same normal distribution as in the uncertainty analysis.

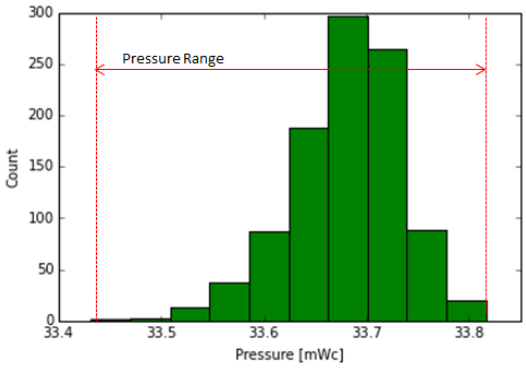

Figure K.2 Example output pressures after Monte Carlo analysis with 1000 simulations at one location

Result uncertainty analysis

Table K.1 shows the standard deviations and Figure K.3 the pressure ranges of the uncertainty analysis. The standard deviation is the highest during maximum flow. During minimum flow the standard deviations are almost zero at all locations. Uncertainties in the model have the least influence during minimum flow. Only location 2 with the factorized model shows a standard deviation of 0.001. Location 2 is located the farthest from the inlet. Figure K.4 shows the simulated pressure at the inlet and at location 2. During the night there is almost no difference between the pressure at the inlet and at the end of the system. The maximum head difference is 0.91 meter in the morning.

The uncertainty per location varies more in the factorized model compared to the uniform model. This is due to the consumption factors. The consumption factors in the factorized model vary between 0.008085 and 47.26457. The variation of large consumers influences the flow at the different locations. This effect is less in the uniform model where all consumption factors are initially 1. This is also visible in the result of the sensitivity analysis.

Table K.1 Standard deviations uncertainty analysis

	Uniform Model			Factorized Model		
Locations	**Std. Min. Flow [mWc]**	**Std. Avg. Flow [mWc]**	**Std. Max. Flow [mWc]**	**Std. Min. Flow [mWc]**	**Std. Avg. Flow [mWc]**	**Std. Max. Flow [mWc]**
Loc. 1	0.000	0.015	0.047	0.000	0.014	0.043
Loc. 2	0.000	0.018	0.056	0.001	0.028	0.086
Loc. 3	0.000	0.016	0.048	0.000	0.014	0.043
Loc. 4	0.000	0.018	0.054	0.000	0.021	0.066
Loc. 5	0.000	0.016	0.049	0.000	0.014	0.044
Loc. 6	0.000	0.015	0.047	0.000	0.014	0.043

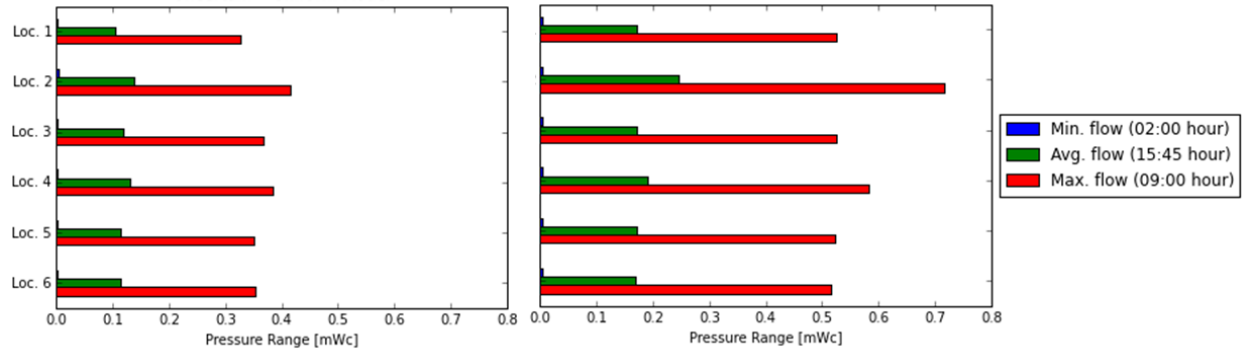

Figure K.3 Pressure range uncertainty analysis. Left: uniform model. Right: factorized model.

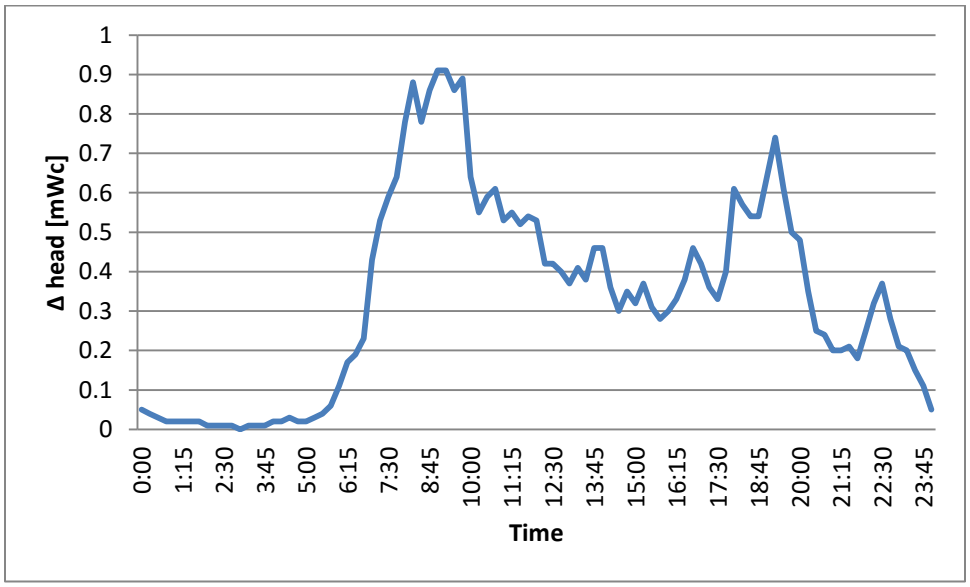

Figure K.4 Simulated head difference between inlet DMA and location 2 (end of network)

Result sensitivity analysis

For both models the parameter diameter has the highest standard deviation (Table K.2). This is followed by length and the roughness and consumption factor have the lowest standard deviation. Figure K.6 shows the flow regime of the two models. Especially during the night, the Reynolds numbers are below 2000. When Reynolds numbers are below 2000 the friction is not dependent on the Darcy-Weisbach coefficient (Appendix D). Varying the Darcy-Weisbach roughness coefficient has almost no influence on the pressure when most pipes have laminar flow.

Table K.2 Standard deviations sensitivity analysis. The standard deviation in this table is the average standard deviation of all six locations.

	Uniform Model			Factorized Model		
Parameter	**Std. Min. Flow [mWc]**	**Std. Avg. Flow [mWc]**	**Std. Max. Flow [mWc]**	**Std. Min. Flow [mWc]**	**Std. Avg. Flow [mWc]**	**Std. Max. Flow [mWc]**
Consumption Factor	0.000	0.000	0.001	0.000	0.002	0.007
Roughness	0.000	0.002	0.007	0.000	0.001	0.003
Length	0.000	0.003	0.010	0.000	0.002	0.007
Diameter	0.001	0.020	0.062	0.000	0.017	0.052

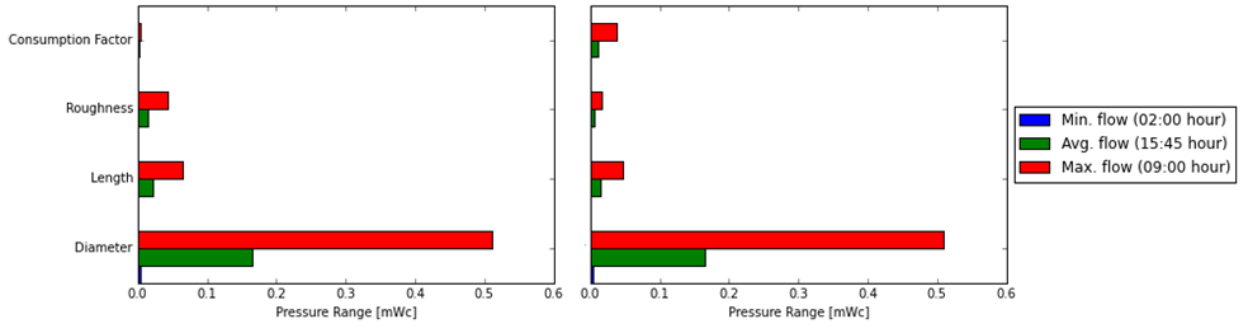

Figure K.5 Pressure range uncertainty analysis. The pressure range is the average standard deviation of all six locations. Left: uniform model. Right: factorized model.

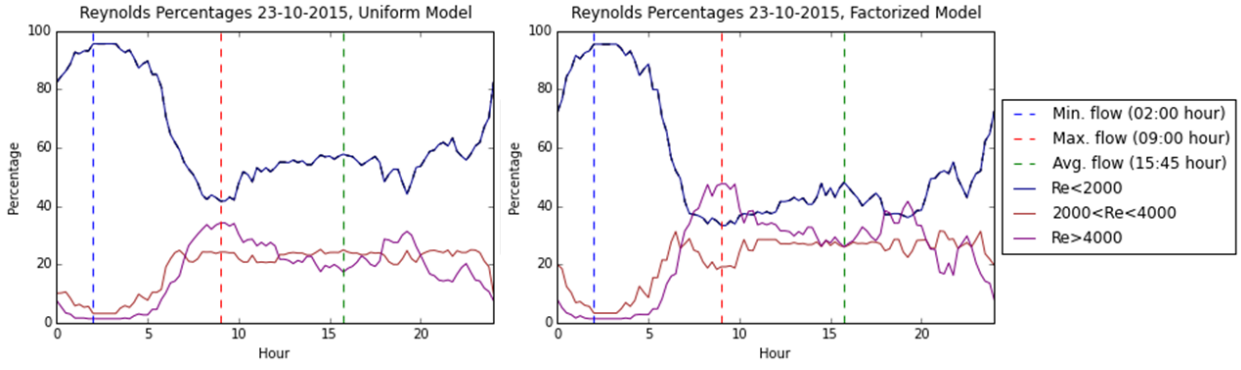

Figure K.6 Flow regimes during one day (2015-10-23). Left: uniform model. Right: factorized model.

APPENDIX L ARTIFICIAL LEAK LOCALIZATION VISUALIZATIONS

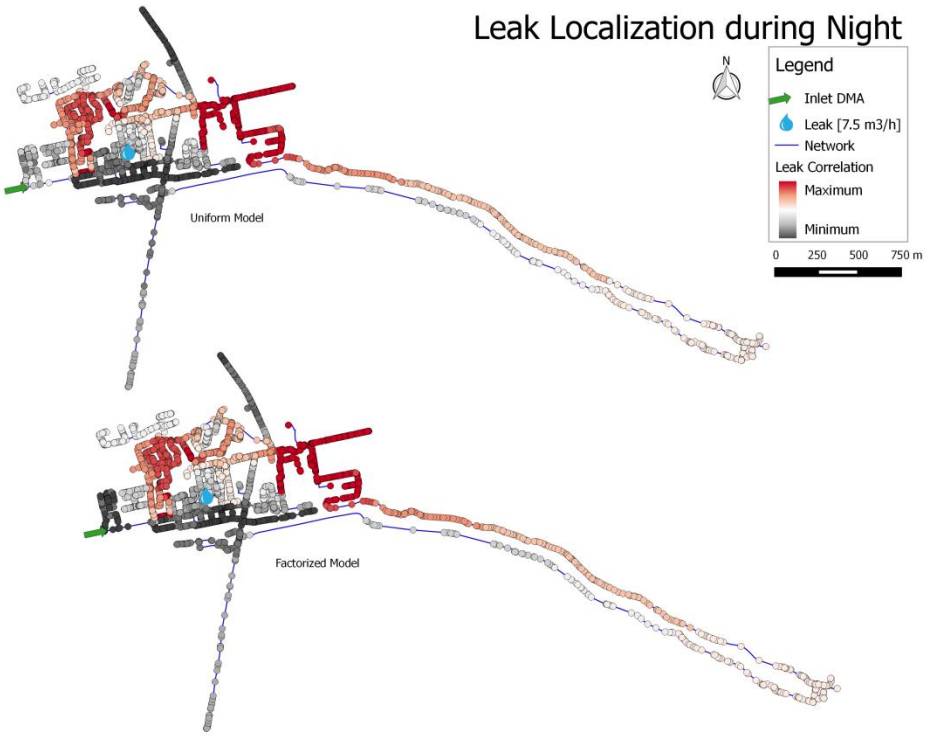

Figure L.1 Leak localizaiton result location I during the night and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.33. Bottom: fact. model lc_{max}=0.33

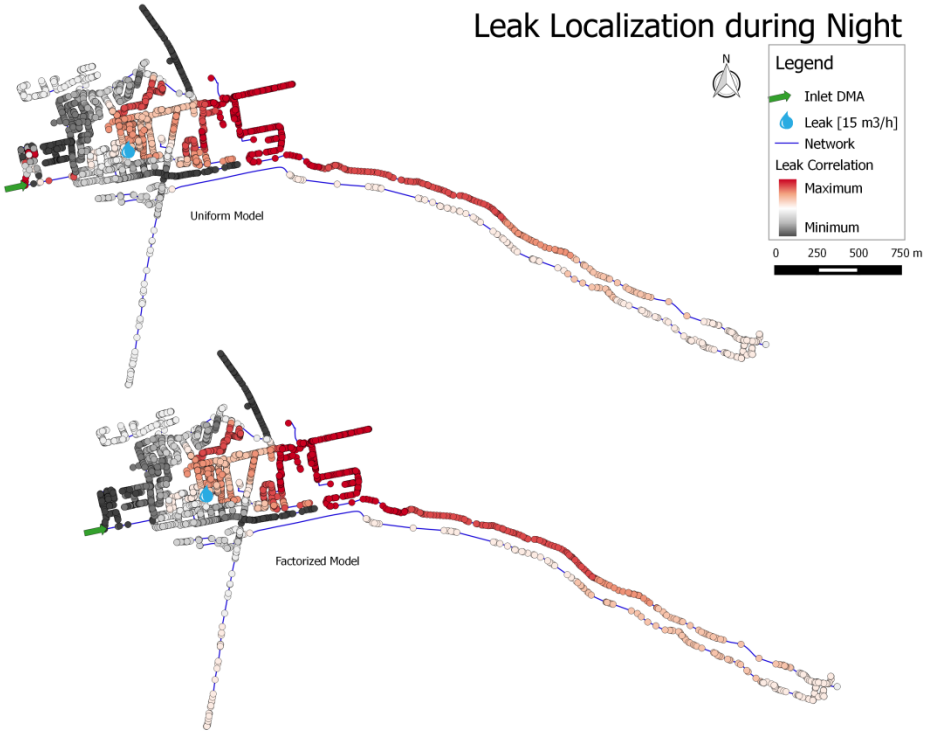

Figure L.2 Leak localizaiton result location I during the night and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.45. Bottom: fact. model lc_{max}=0.38

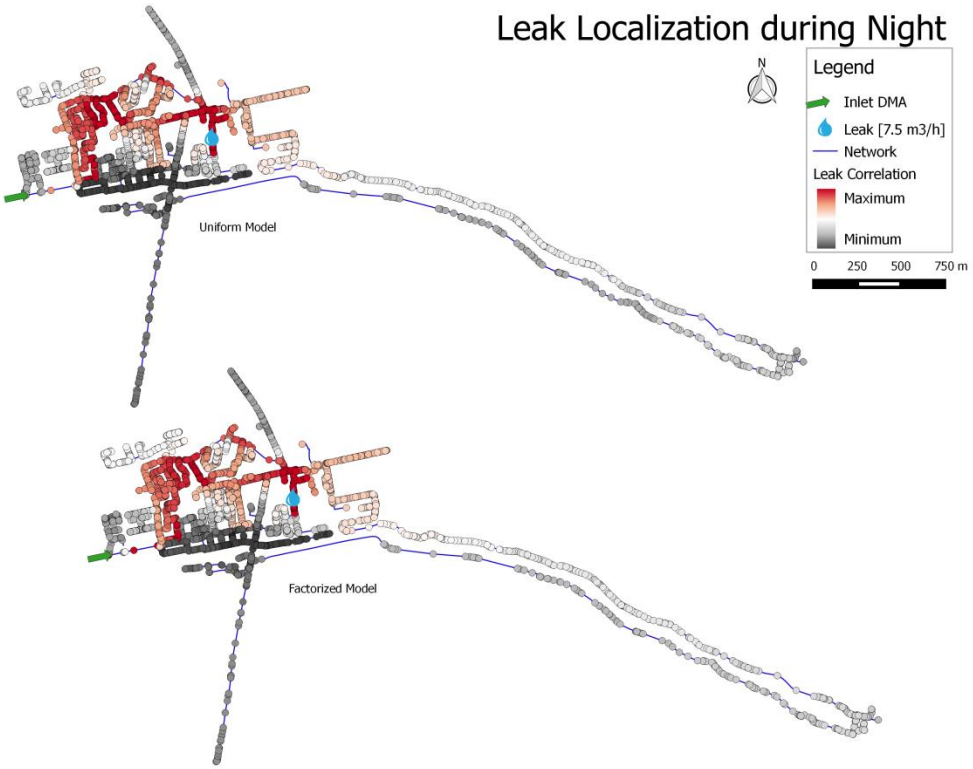

Figure L.3 Leak localizaiton result location II during the night and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.78. Bottom: fact. model lc_{max}=0.78

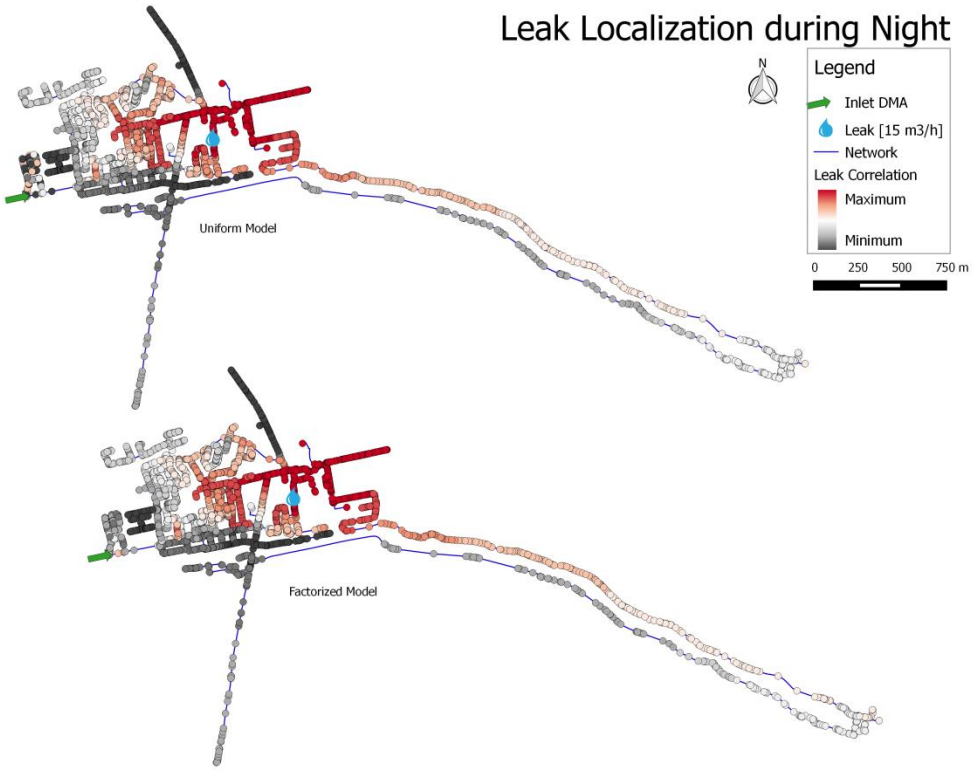

Figure L.4 Leak localizaiton result location II during the night and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.42. Bottom: fact. model lc_{max}=0.87

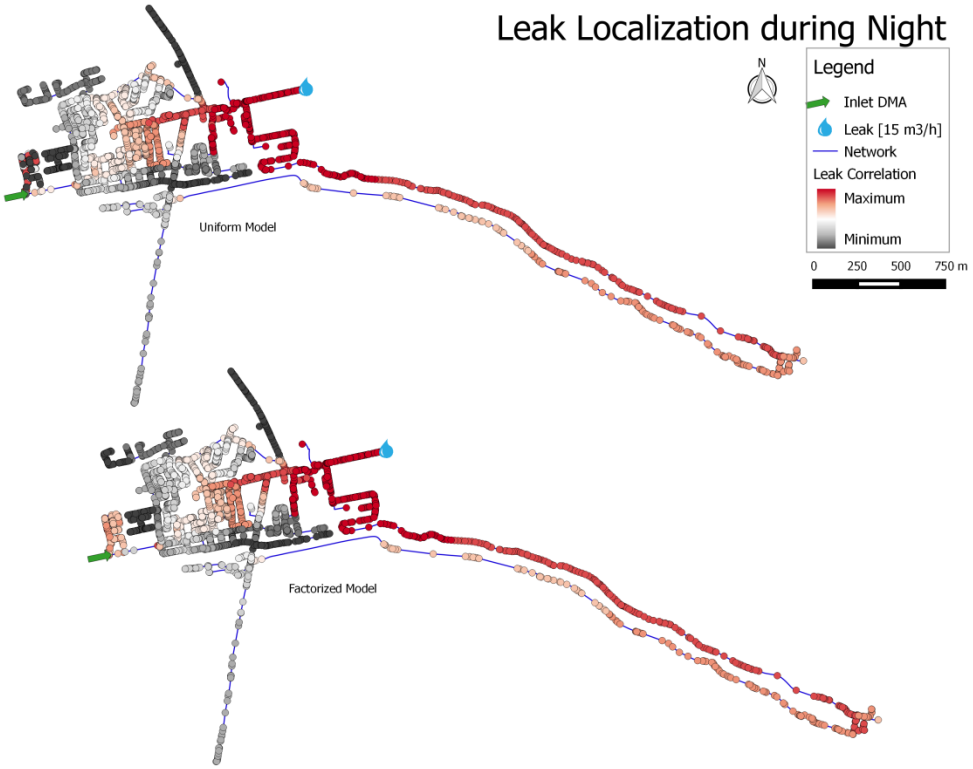

Figure L.5 Leak localizaiton result location III during the night and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.39. Bottom: fact. model lc_{max}=0.94

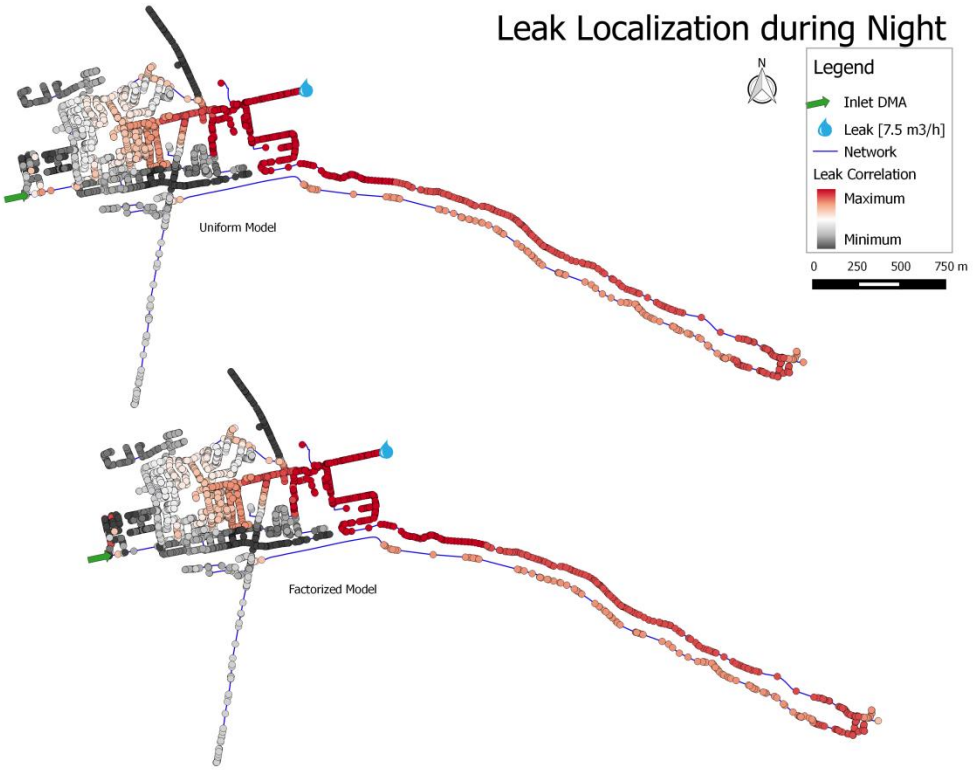

Figure L.6 Leak localizaiton result location III during the night and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.39. Bottom: fact. model lc_{max}=0.94

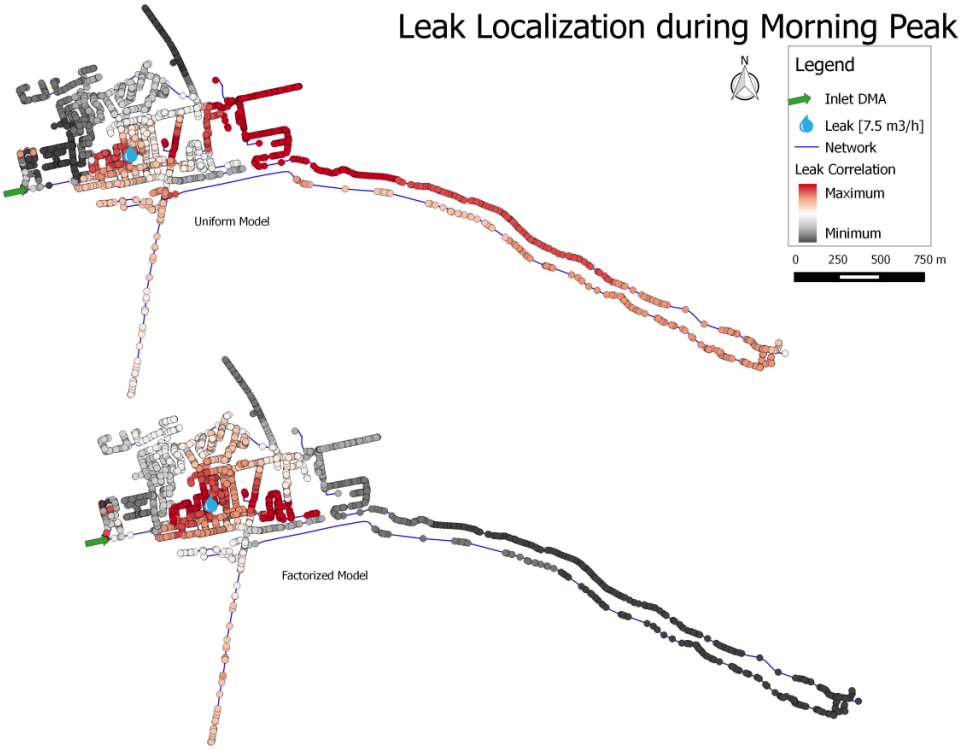

Figure L.7 Leak localizaiton result location I during the morning and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.69. Bottom: fact. model lc_{max}=0.52

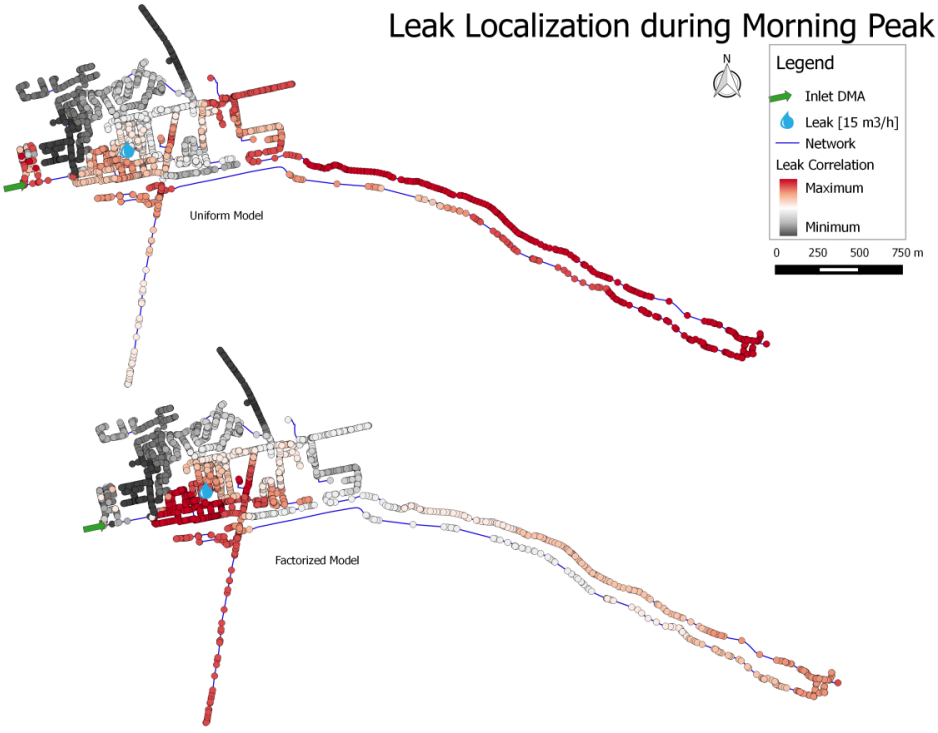

Figure L.8 Leak localizaiton result location I during the morning and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.85. Bottom: fact. model lc_{max}=0.54

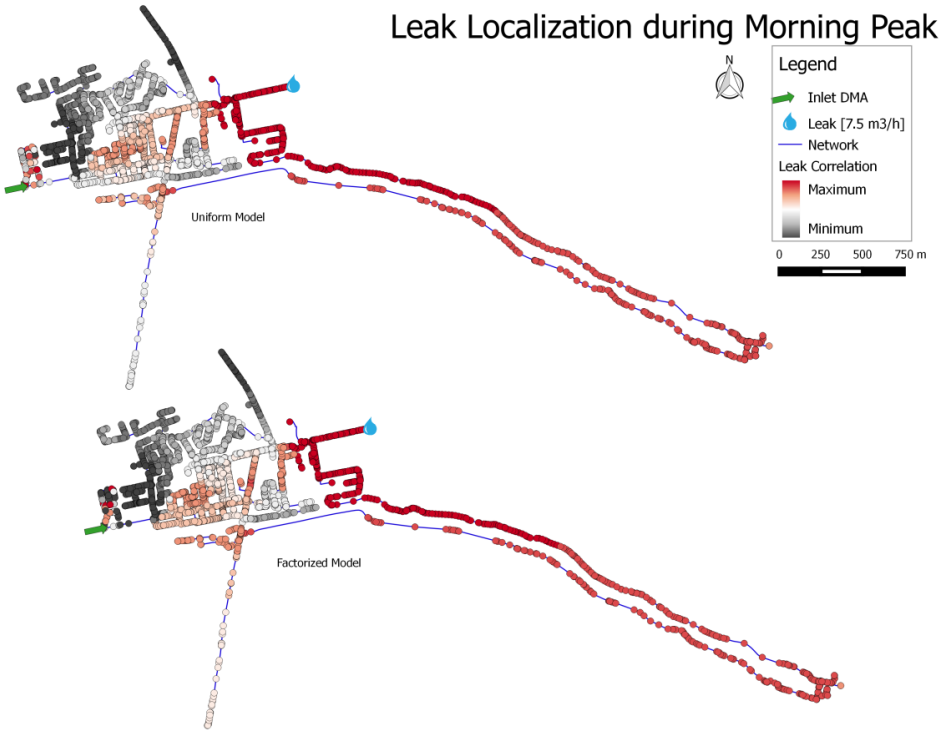

Figure L.9 Leak localizaiton result location III during the morning and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.88. Bottom: fact. model lc_{max}=0.83

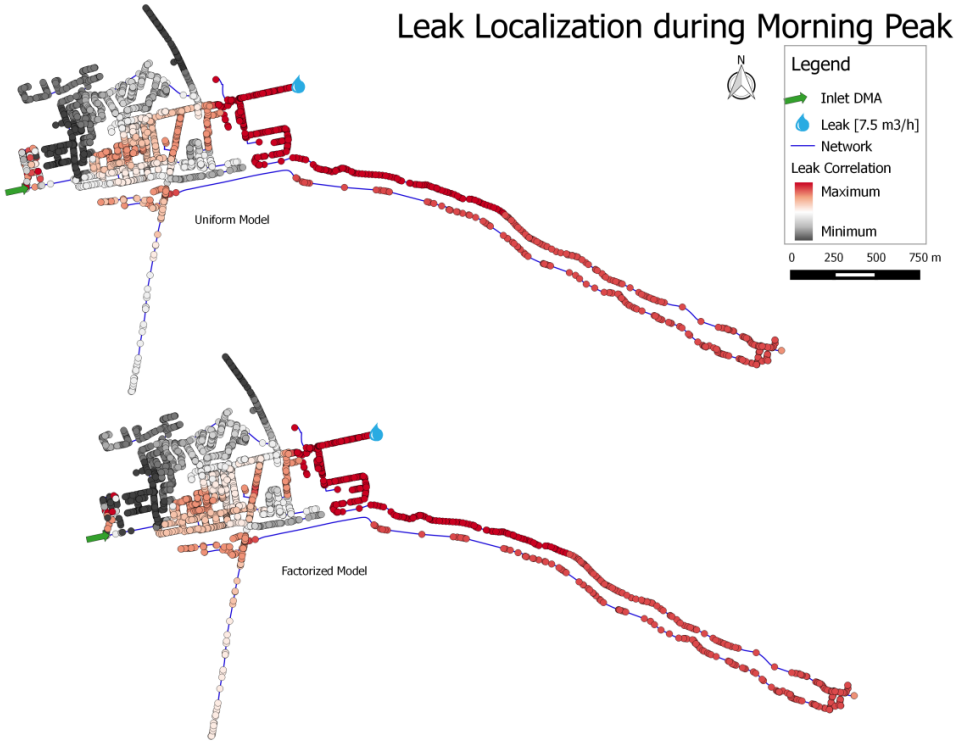

Figure L.10 Leak localizaiton result location III during the morning and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.95. Bottom: fact. model lc_{max}=0.94

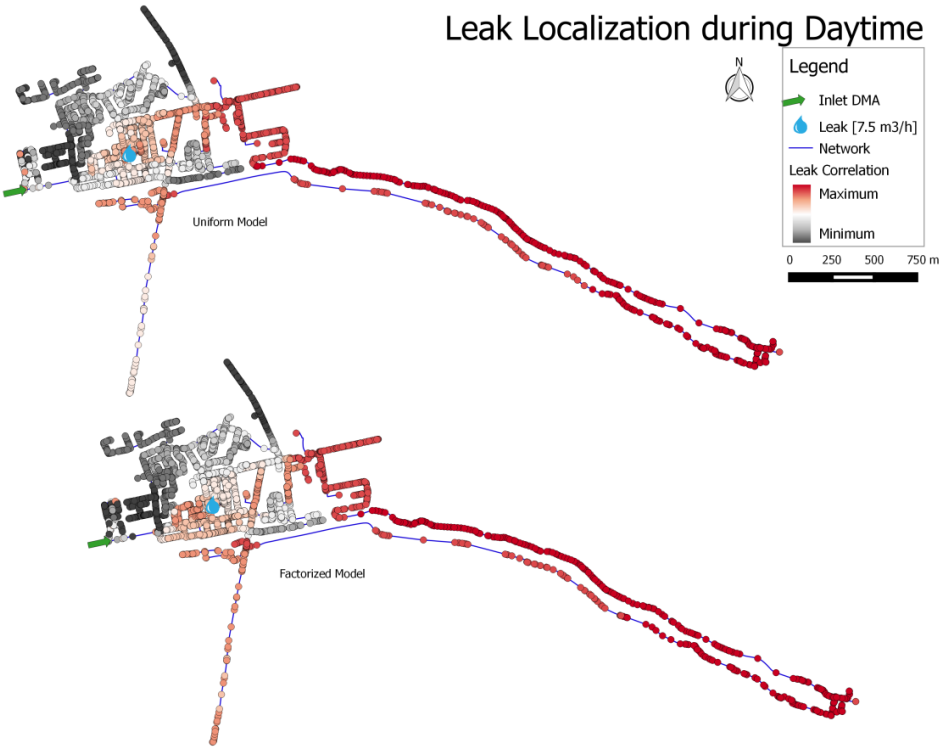

Figure L.11 Leak localizaiton result location I during the daytime and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.86. Bottom: fact. model lc_{max}=0.75

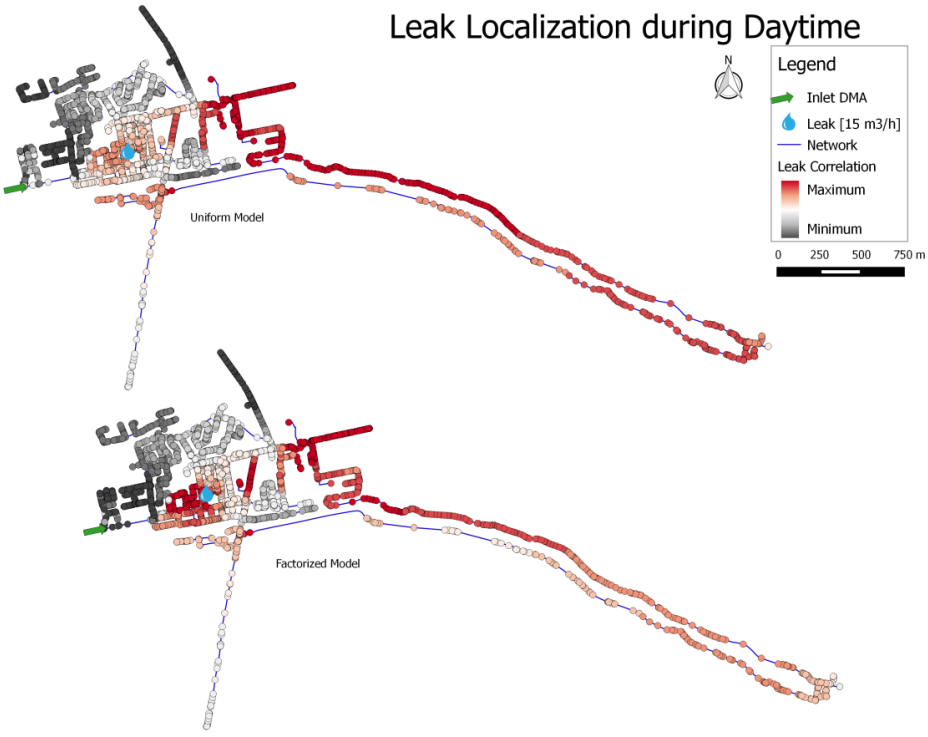

Figure L.12 Leak localizaiton result location I during the daytime and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.77. Bottom: fact. model lc_{max}=0.74

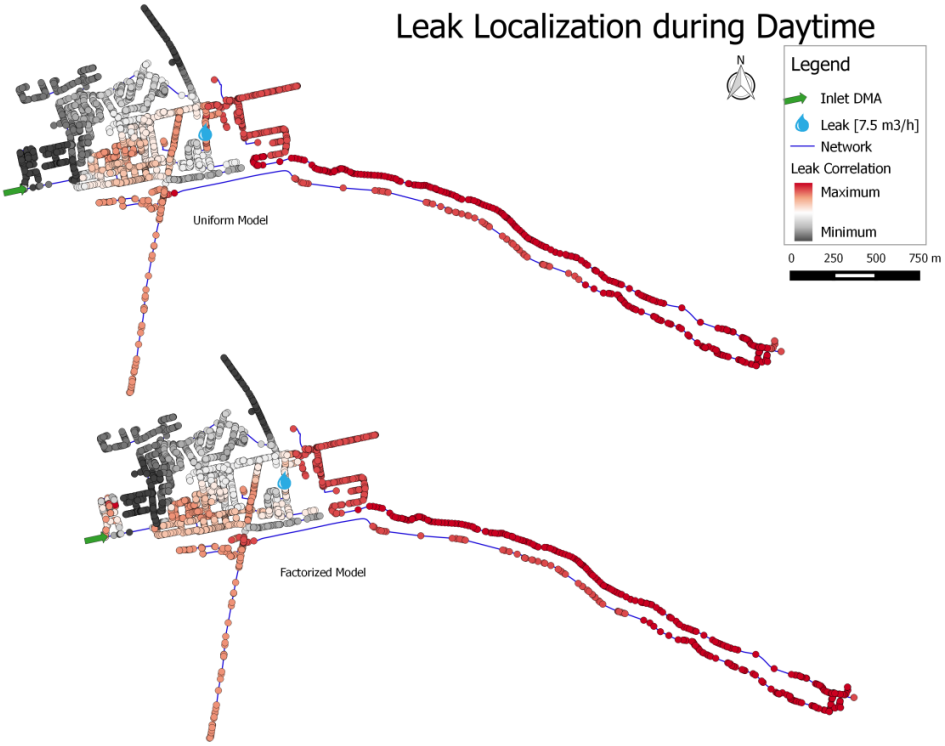

Figure L.13 Leak localizaiton result location II during the daytime and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.82. Bottom: fact. model lc_{max}=0.81

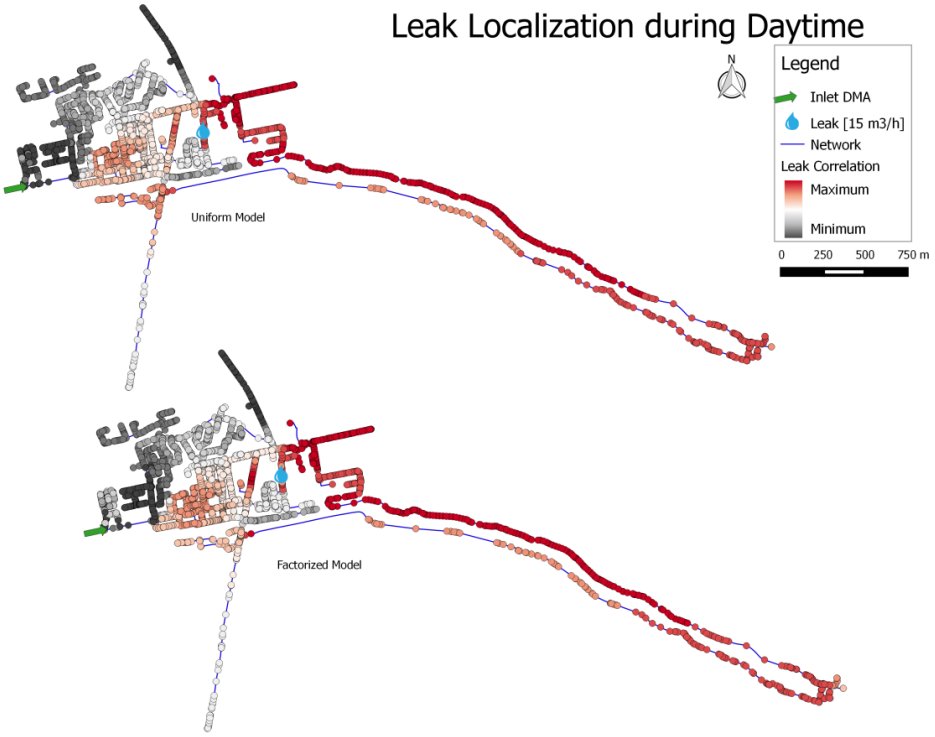

Figure L.14 Leak localizaiton result location II during the daytime and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.78. Bottom: fact. model lc_{max}=0.69

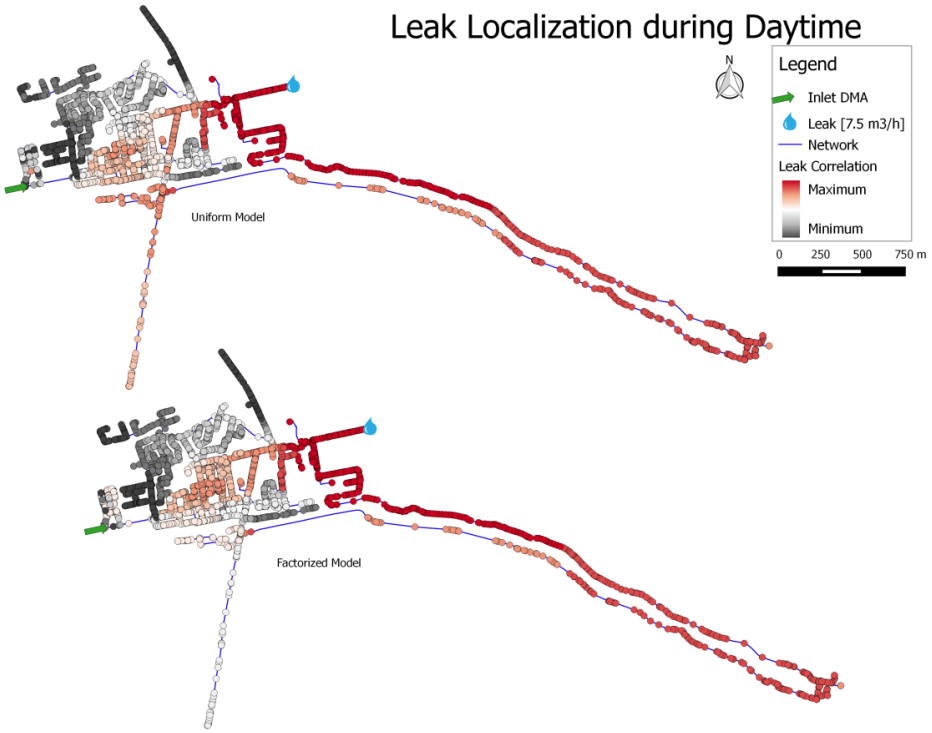

Figure L.15 Leak localizaiton result location III during the daytime and a leak size of 7.5 m^3/h. Top: unif. lc_{max}= 0.88. Bottom: fact. model lc_{max}=0.86

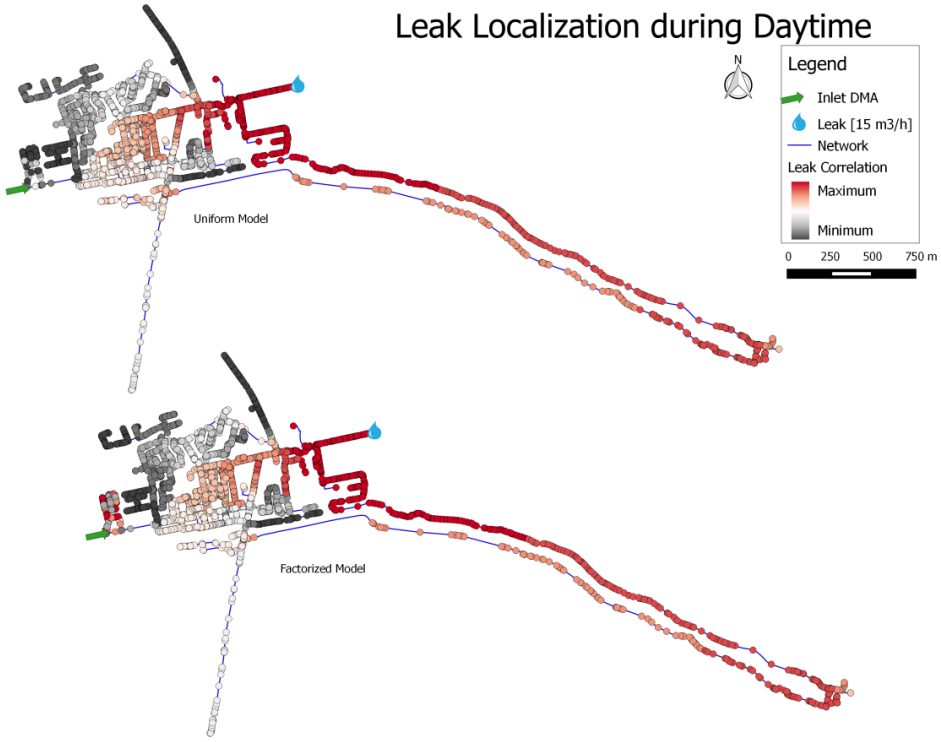

Figure L.16 Leak localizaiton result location III during the daytime and a leak size of 15 m^3/h. Top: unif. lc_{max}= 0.95. Bottom: fact. model lc_{max}=0.94

APPENDIX M SIMULATED HEAD LOSS DUE TO ARTIFICIAL LEAKS

Table M.1 Simulated pressure differences at sensor node locations after a leak with the uniform consumption model during the night

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.02	-0.08	-0.02	-0.05	-0.02	-0.05
PL202	-0.02	-0.07	-0.02	-0.06	-0.02	-0.06
PL222	-0.02	-0.08	-0.02	-0.07	-0.02	-0.07
PL255	-0.03	-0.08	-0.04	-0.15	-0.24	-0.94
PL277	-0.02	-0.08	-0.02	-0.05	-0.02	-0.05
PL279	-0.02	-0.06	-0.02	-0.04	-0.01	-0.04
PL319	-0.02	-0.08	-0.03	-0.09	-0.03	-0.09
SM1023	-0.02	-0.07	-0.02	-0.05	-0.02	-0.05
SM1025	-0.03	-0.08	-0.05	-0.16	-0.07	-0.28
SM1027	-0.02	-0.08	-0.02	-0.07	-0.07	-0.26
SM1032	-0.02	-0.08	-0.03	-0.10	-0.14	-0.52
SM1033	-0.02	-0.08	-0.02	-0.07	-0.02	-0.07
SM1034	-0.02	-0.08	-0.02	-0.05	-0.02	-0.06
Maximum Head loss	*-0.03*	*-0.08*	*-0.05*	*-0.16*	*-0.24*	*-0.94*

Table M.2 Simulated pressure differences at sensor node locations after a leak with the factorized consumption model during the night

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.02	-0.07	-0.01	-0.04	-0.01	-0.05
PL202	-0.02	-0.06	-0.02	-0.05	-0.01	-0.05
PL222	-0.02	-0.07	-0.02	-0.06	-0.02	-0.06
PL255	-0.02	-0.07	-0.04	-0.13	-0.21	-0.77
PL277	-0.02	-0.07	-0.01	-0.04	-0.01	-0.04
PL279	-0.01	-0.05	-0.01	-0.04	-0.01	-0.04
PL319	-0.02	-0.07	-0.03	-0.08	-0.02	-0.08
SM1023	-0.02	-0.05	-0.02	-0.04	-0.01	-0.04
SM1025	-0.02	-0.07	-0.04	-0.13	-0.06	-0.23
SM1027	-0.02	-0.07	-0.02	-0.06	-0.07	-0.22
SM1032	-0.02	-0.07	-0.03	-0.09	-0.13	-0.43
SM1033	-0.02	-0.07	-0.02	-0.06	-0.02	-0.06
SM1034	-0.02	-0.07	-0.01	-0.04	-0.01	-0.05
Maximum Head loss	*-0.02*	*-0.07*	*-0.04*	*-0.13*	*-0.21*	*-0.77*

Table M.3 Simulated pressure differences at sensor node locations after a leak with the uniform consumption model during the morning

Meter:	Location I Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location III Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.15	-0.26	-0.11	-0.25
PL202	-0.14	-0.24	-0.11	-0.25
PL222	-0.16	-0.28	-0.13	-0.30
PL255	-0.16	-0.28	-0.51	-1.51
PL277	-0.15	-0.26	-0.11	-0.25
PL279	-0.12	-0.20	-0.09	-0.21
PL319	-0.16	-0.28	-0.14	-0.33
SM1023	-0.13	-0.23	-0.11	-0.24
SM1025	-0.16	-0.29	-0.23	-0.59
SM1027	-0.15	-0.27	-0.29	-0.72
SM1032	-0.16	-0.28	-0.47	-1.20
SM1033	-0.15	-0.27	-0.13	-0.30
SM1034	-0.15	-0.26	-0.11	-0.25
Maximum Head loss	*-0.16*	*-0.29*	*-0.51*	*-1.51*

Table M.4 Simulated pressure differences at sensor node locations after a leak with the factorized consumption model during the morning

Meter:	Location I Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location III Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.12	-0.22	-0.09	-0.20
PL202	-0.11	-0.20	-0.09	-0.21
PL222	-0.13	-0.23	-0.11	-0.25
PL255	-0.13	-0.23	-0.50	-1.37
PL277	-0.12	-0.21	-0.09	-0.20
PL279	-0.09	-0.16	-0.08	-0.17
PL319	-0.13	-0.23	-0.12	-0.28
SM1023	-0.11	-0.19	-0.09	-0.20
SM1025	-0.13	-0.23	-0.21	-0.53
SM1027	-0.13	-0.22	-0.25	-0.67
SM1032	-0.13	-0.23	-0.40	-1.12
SM1033	-0.13	-0.22	-0.11	-0.25
SM1034	-0.12	-0.21	-0.09	-0.21
Maximum Head loss	*-0.13*	*-0.23*	*-0.50*	*-1.37*

Table M.5 Simulated pressure differences at sensor node locations after a leak with the uniform consumption model during daytime

Meter:	Location I Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location II Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location III Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.09	-0.20	-0.09	-0.16	-0.06	-0.16
PL202	-0.08	-0.18	-0.10	-0.18	-0.07	-0.16
PL222	-0.09	-0.21	-0.12	-0.22	-0.08	-0.20
PL255	-0.10	-0.21	-0.14	-0.32	-0.39	-1.25
PL277	-0.09	-0.20	-0.09	-0.16	-0.06	-0.16
PL279	-0.07	-0.15	-0.08	-0.14	-0.05	-0.13
PL319	-0.10	-0.21	-0.12	-0.24	-0.09	-0.22
SM1023	-0.08	-0.17	-0.09	-0.17	-0.06	-0.15
SM1025	-0.10	-0.22	-0.15	-0.33	-0.15	-0.44
SM1027	-0.09	-0.21	-0.12	-0.24	-0.20	-0.52
SM1032	-0.09	-0.21	-0.14	-0.30	-0.34	-0.89
SM1033	-0.09	-0.21	-0.11	-0.21	-0.08	-0.19
SM1034	-0.09	-0.20	-0.09	-0.16	-0.07	-0.16
Maximum Head loss	*-0.10*	*-0.22*	*-0.15*	*-0.33*	*-0.39*	*-1.25*

Table M.6 Simulated pressure differences at sensor node locations after a leak with the factorized consumption model during daytime

Meter:	Location I Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location II Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Location III Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.07	-0.16	-0.07	-0.13	-0.05	-0.13
PL202	-0.07	-0.15	-0.08	-0.15	-0.06	-0.13
PL222	-0.08	-0.18	-0.10	-0.18	-0.07	-0.16
PL255	-0.08	-0.18	-0.13	-0.26	-0.35	-1.11
PL277	-0.07	-0.16	-0.07	-0.13	-0.05	-0.13
PL279	-0.06	-0.12	-0.06	-0.12	-0.04	-0.11
PL319	-0.08	-0.18	-0.11	-0.21	-0.07	-0.18
SM1023	-0.06	-0.14	-0.07	-0.14	-0.05	-0.13
SM1025	-0.08	-0.18	-0.14	-0.28	-0.14	-0.39
SM1027	-0.07	-0.17	-0.09	-0.18	-0.17	-0.48
SM1032	-0.08	-0.17	-0.11	-0.24	-0.30	-0.85
SM1033	-0.08	-0.17	-0.09	-0.18	-0.06	-0.16
SM1034	-0.07	-0.16	-0.07	-0.13	-0.05	-0.13
Maximum Head loss	*-0.08*	*-0.18*	*-0.14*	*-0.28*	*-0.35*	*-1.11*

APPENDIX N MEASURED HEAD LOSS AFTER ARTIFICIAL LEAKS

Table N.1 Pressure differences at measurement locations with the uniform model during the night. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.04	-0.14	-0.02	-0.12	-0.01	-0.11
PL202	-0.02	-0.11	0.00	-0.06	-0.01	-0.06
PL222	-0.04	-0.14	-0.02	-0.08	-0.02	-0.10
PL255	-0.04	-0.15	-0.02	-0.10	-0.26	-0.77
PL277	0.34	0.13	0.21	0.13	-0.01	0.33
PL279	0.02	-0.04	0.02	-0.01	-0.01	-0.03
PL319	0.10	0.05	0.03	0.01	-0.02	0.20
SM1023	0.04	-0.01	0.02	-0.02	-0.01	-0.05
SM1025	-0.02	-0.10	-0.03	-0.09	-0.08	-0.27
SM1027	0.03	-0.10	0.02	-0.02	-0.07	-0.19
SM1032	-0.03	-0.10	0.02	-0.07	-0.14	-0.38
SM1033	0.05	-0.05	0.04	-0.02	-0.02	-0.07
SM1034	-0.01	-0.12	-0.01	-0.07	-0.01	-0.10
Maximum Head loss	*-0.04*	*-0.15*	*-0.03*	*-0.12*	*-0.26*	*-0.77*

Table N.2 Pressure differences at measurement locations with the factorized model during the night. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.04	-0.14	-0.02	-0.12	-0.02	-0.11
PL202	-0.02	-0.11	0.00	-0.06	-0.01	-0.06
PL222	-0.04	-0.14	-0.02	-0.08	-0.03	-0.10
PL255	-0.04	-0.15	-0.02	-0.10	-0.27	-0.78
PL277	0.34	0.12	0.21	0.13	0.18	0.33
PL279	0.02	-0.04	0.02	-0.01	0.00	-0.03
PL319	0.10	0.05	0.03	0.01	0.14	0.20
SM1023	0.04	-0.01	0.02	-0.02	0.00	-0.05
SM1025	-0.02	-0.11	-0.03	-0.09	-0.09	-0.27
SM1027	0.03	-0.10	0.02	-0.02	-0.06	-0.19
SM1032	-0.03	-0.10	0.02	-0.07	-0.12	-0.38
SM1033	0.05	-0.05	0.04	-0.02	-0.02	-0.07
SM1034	-0.01	-0.12	-0.01	-0.07	-0.04	-0.10
Maximum Head loss	*-0.04*	*-0.15*	*-0.03*	*-0.12*	*-0.27*	*-0.78*

Table N.3 Pressure differences at measurement locations with the uniform model during the morning. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.18	-0.32	-0.26	-0.47
PL202	-0.08	-0.24	-0.14	-0.30
PL222	-0.14	-0.29	-0.20	-0.39
PL255	-0.28	-0.33	-0.77	-1.46
PL277	0.00	-0.29	-0.11	-0.31
PL279	0.06	-0.10	0.02	-0.09
PL319	0.06	-0.14	-0.01	-0.19
SM1023	-0.10	-0.28	-0.20	-0.39
SM1025	-0.30	-0.42	-0.49	-0.77
SM1027	-0.22	-0.33	-0.65	-1.00
SM1032	-0.28	-0.65	-0.82	-1.27
SM1033	-0.12	-0.31	-0.24	-0.39
SM1034	-0.23	-0.38	-0.42	-0.56
Maximum Head loss	*-0.30*	*-0.65*	*-0.82*	*-1.46*

Table N.4 Pressure differences at measurement locations with the factorized model during the morning. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.26	-0.37	-0.30	-0.51
PL202	-0.16	-0.30	-0.18	-0.35
PL222	-0.22	-0.34	-0.24	-0.43
PL255	-0.13	-0.24	-0.68	-1.38
PL277	-0.07	-0.33	-0.14	-0.34
PL279	-0.01	-0.14	-0.02	-0.13
PL319	-0.02	-0.19	-0.05	-0.23
SM1023	-0.18	-0.32	-0.24	-0.42
SM1025	-0.35	-0.45	-0.52	-0.80
SM1027	-0.09	-0.25	-0.58	-0.93
SM1032	-0.01	-0.48	-0.67	-1.12
SM1033	-0.20	-0.35	-0.28	-0.43
SM1034	-0.30	-0.42	-0.46	-0.59
Maximum Head loss	*-0.35*	*-0.48*	*-0.68*	*-1.38*

Table N.5 Pressure differences at measurement locations with the uniform model during daytime. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.20	-0.39	-0.20	-0.39	-0.18	-0.33
PL202	-0.11	-0.26	-0.09	-0.23	-0.11	-0.22
PL222	-0.17	-0.36	-0.15	-0.29	-0.16	-0.28
PL255	-0.42	-0.60	-0.37	-0.53	-0.52	-1.30
PL277	-0.02	-0.23	-0.03	-0.20	-0.08	-0.06
PL279	0.00	-0.09	0.02	-0.09	0.01	-0.09
PL319	0.02	-0.19	0.06	-0.14	-0.03	-0.04
SM1023	-0.25	-0.33	-0.17	-0.30	-0.21	-0.27
SM1025	-0.30	-0.53	-0.28	-0.50	-0.34	-0.60
SM1027	-0.46	-0.57	-0.39	-0.51	-0.35	-0.66
SM1032	-0.73	-0.67	-0.59	-0.65	-0.52	-0.91
SM1033	-0.25	-0.41	-0.21	-0.33	-0.14	-0.30
SM1034	-0.31	-0.49	-0.32	-0.41	-0.28	-0.40
Maximum Head loss	*-0.73*	*-0.67*	*-0.59*	*-0.65*	*-0.52*	*-1.30*

Table N.6 Pressure differences at measurement locations with the factorized model during daytime. The pressure differences are the measurements subtracted by the simulated leak-free pressures.

Meter:	Location I		Location II		Location III	
	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]	Leak:7.5m3/h [mWc]	Leak:15m3/h [mWc]
PL196	-0.23	-0.42	-0.22	-0.41	-0.19	-0.34
PL202	-0.13	-0.29	-0.12	-0.24	-0.12	-0.23
PL222	-0.19	-0.38	-0.18	-0.31	-0.17	-0.29
PL255	-0.36	-0.54	-0.31	-0.49	-0.49	-1.26
PL277	-0.04	-0.25	-0.05	-0.21	-0.09	-0.07
PL279	-0.02	-0.11	-0.01	-0.10	0.00	-0.10
PL319	0.00	-0.21	0.04	-0.16	-0.04	-0.05
SM1023	-0.27	-0.36	-0.19	-0.31	-0.22	-0.28
SM1025	-0.32	-0.55	-0.30	-0.51	-0.34	-0.61
SM1027	-0.41	-0.52	-0.34	-0.47	-0.32	-0.62
SM1032	-0.65	-0.59	-0.50	-0.60	-0.48	-0.87
SM1033	-0.27	-0.43	-0.23	-0.34	-0.15	-0.31
SM1034	-0.33	-0.51	-0.34	-0.42	-0.28	-0.41
Maximum Head loss	*-0.65*	*-0.59*	*-0.50*	*-0.60*	*-0.49*	*-1.26*

APPENDIX O RESULTS SPECIFC MODEL

One large consumer in DMA Leimuiden uses 74 times more than the average customer consumption. This customer has two connections and both connections have a smart meter. The pressures of this smart meter are not used because the customer uses water (nearly) the whole day. The smart meter also measures the inflow and this is included in the specific model. Figure O.1 shows the consumption by this largest consumer during the artificial leaks.

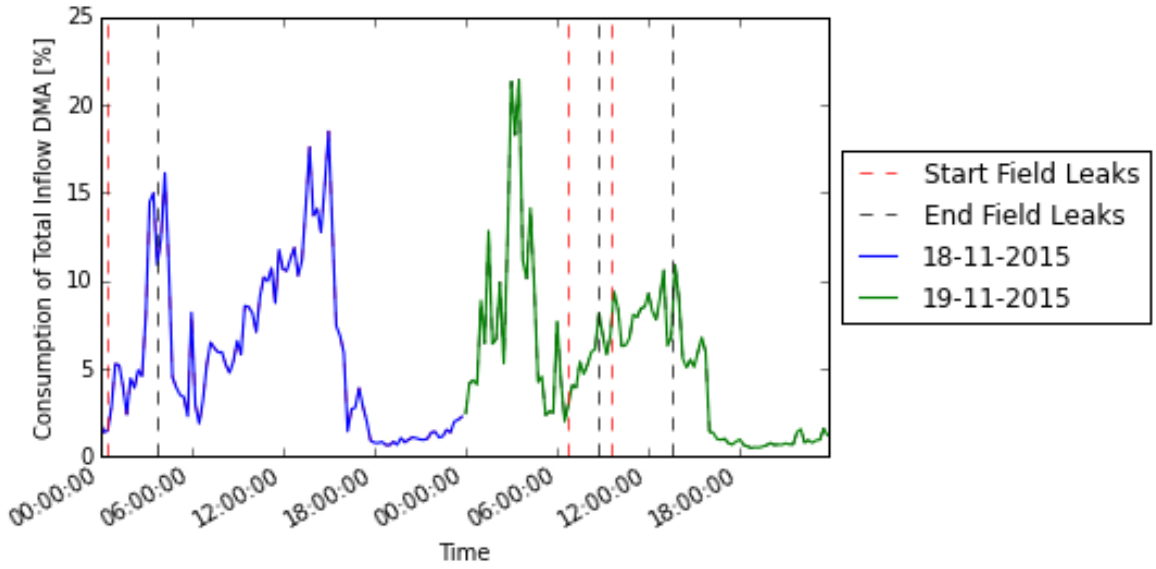

Figure O.1 Consumption by largest consumer during artificial leaks. This is included in the specific model.

Table O.1 RMSE uniform model, factorized model and specific model

	Uniform Model			**Factorized Model**			**Specific Model**		
Meter	Initial RMSE	RMSE after Elevation Correction	RMSE after Roughness Set	Initial RMSE	RMSE after Elevation Correction	RMSE after Roughness Set	Initial RMSE	RMSE after Elevation Correction	RMSE after Roughness Set
PL196	0.26	0.10	0.03	0.26	0.10	0.05	0.26	0.10	0.06
PL202	3.80	0.08	0.02	3.80	0.08	0.03	3.80	0.08	0.03
PL222	0.18	0.09	0.02	0.18	0.09	0.04	0.18	0.09	0.04
PL255	0.17	0.21	0.13	0.13	0.16	0.11	0.10	0.13	0.04
PL277	3.29	0.09	0.11	3.29	0.09	0.10	3.29	0.09	0.10
PL279	0.04	0.04	0.08	0.04	0.04	0.06	0.04	0.04	0.06
PL319	3.74	0.11	0.16	3.74	0.11	0.14	3.74	0.11	0.14
SM1023	3.91	0.06	0.05	3.91	0.06	0.04	3.91	0.06	0.04
SM1025	2.89	0.11	0.05	2.88	0.10	0.05	2.88	0.10	0.05
SM1027	1.71	0.17	0.09	1.67	0.13	0.08	1.66	0.12	0.06
SM1032	3.85	0.22	0.14	3.78	0.16	0.11	3.77	0.14	0.09
SM1033	4.96	0.07	0.05	4.97	0.06	0.04	4.97	0.06	0.04
SM1034	0.61	0.12	0.06	0.61	0.12	0.07	0.61	0.12	0.07
Total	*29.41*	*1.47*	*0.99*	*29.26*	*1.30*	*0.92*	*29.21*	*1.24*	*0.82*

The RMSE is smaller for the specific model than the factorized model (0.82, 0.92 for specific model, fact. model respectively, see Table O.1). The leak localization performance does not improve with the specific model (Table O.2). During the morning, leak location I with a leak size of 7.5 m^3/h, the performance got poorer compared to the factorized model (good, reasonable performance for the fact., specific model respectively, see Table O.2).

Table O.2 Leak Localization Performance Artificial Leaks including Specific Model. If the percentage of false positive (FP) nodes is underneath 5% it is classified as good (green), between 5 and 15% as reasonable (orange) and above 15% as poor (red). The blue rectangle indicates the leak localizations that have a different performance category for the factorized and specific model.

Time	Location	Leak Size [m3/h]	Uniform Model [% FP Nodes]	Factorized Model [% FP Nodes]	Specific Model [% FP Nodes]
Night	I	7.5	74.4%	72.2%	72.0%
		15	41.7%	41.1%	41.9%
	II	7.5	2.5%	4.0%	4.3%
		15	1.0%	1.2%	1.0%
	III	7.5	1.0%	1.3%	2.2%
		15	0.8%	1.7%	2.2%
Morning	I	7.5	14.7%	5.0%	6.3%
		15	41.0%	14.0%	13.9%
	III	7.5	2.2%	0.8%	0.9%
		15	2.0%	1.9%	2.3%
Daytime	I	7.5	26.1%	40.7%	43.9%
		15	22.1%	7.8%	5.1%
	II	7.5	20.1%	40.7%	43.9%
		15	18.1%	19.8%	26.1%
	III	7.5	2.4%	1.9%	2.4%
		15	1.2%	0.8%	1.4%

APPENDIX P PARAMETER CORRELATION MATRIX

The leak localization has always a result. The maximum leak correlation is often an indication if the leak localization has a good performance. In this example the distance between the leak and node with maximum leak correlation are taken. Besides the leak correlation other values are correlated with the distance to leak. These are the mean, min. and max. values of the FIV and FSMHC. No high correlations were found for these values (Table P.1). This is, however, with old results. The model improved but the correlation matrix is not again created.

The result of the leak localizations after the artificial leaks are not always the same. Table P.2 shows characteristics of the different leak locations. These factors might have an influence on the performance of the leak localization. Other factors are the time the leak localization is performed (night, morning or daytime) or the leak/flow rate.

Table P.1 Example correlation matrix

Index	MaxCor	MinCor	DistToLeakMaxCo
MaxCor	1	0.13	0.0586
MinCor	0.13	1	0.302
DistToLeakMaxCor	0.0586	0.302	1
DistToLeakMinCor	0.571	0.0192	-0.327
CorrLengthLeakC…	0.0855	-0.141	0.441
count_FIV	nan	nan	nan
mean_FIV	-0.731	0.258	0.0525
std_FIV	0.798	-0.239	-0.0816
min_FIV	-0.821	0.181	0.0168
25%_FIV	-0.7	0.292	0.108
50%_FIV	-0.704	0.174	-0.0114
75%_FIV	-0.56	0.22	0.0144
max_FIV	-0.391	0.0567	-0.0869
count_FSMHC	nan	nan	nan
mean_FSMHC	-0.0804	-0.314	-0.82
std_FSMHC	0.0444	0.327	0.815
min_FSMHC	-0.04	-0.327	-0.814
25%_FSMHC	-0.583	0.122	-0.0265
50%_FSMHC	-0.428	-0.0652	-0.228
75%_FSMHC	-0.446	-0.0173	-0.482
max_FSMHC	-0.0657	0.221	-0.0471

Table P.2 Characteristics leak location

	Characteristic Network	Dist to Inlet [m]	Pipe Diameter [mm]
I	'Densily looped'	762	110
II	Looped	1285	160
III	Branched	1958	110

In the correlation matrix (Table P.1) the 'CorrLenghtLeakCorr' is used. For all leak localization results the leak correlation is related to the distance to leak. With a good leak localization result the nodes close to the leak have a large leak correlation and nodes further from the leak have an increased distance. The left Figure of Figure P.1 shows a poor result. The right Figure of Figure P.1 shows a good result where the high leak correlations have a small distance to the leak. The 'CorrLenghtLeakCorr' is the correlation between leak correlation and distance to leak for one output. The values for all outputs are again correlated with the distance to leak of the highest leak correlation. The 'CorrLenghtLeakCorr' is a linear relation while this is probably not the case and dependent on the network structure.

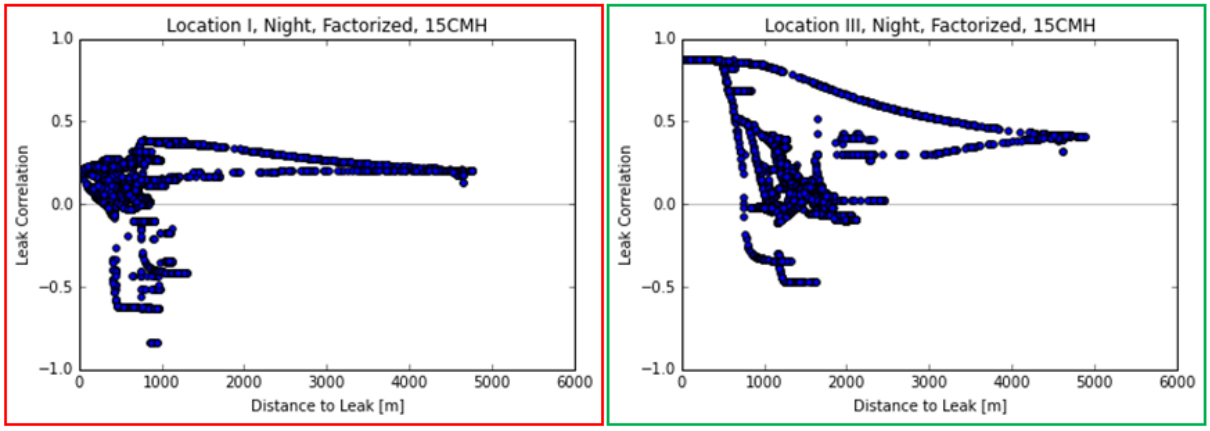

Figure P.1 Relation leak correlation and distance to leak. Left: leak localization for location I during night with the factorized model and a leak size of 15 m^3/h. Right: leak localization for location III during night with the factorized model and a leak size of 15 m^3/h.

APPENDIX Q HIGH LEAK CORRELATION WITH POOR PERFORMANCE

The leak at location I (15 m^3/h) calculated with the Uniform model in the morning peak results in a high leak correlation (0.85) but a poor performance (40.96% FP nodes).

Figure Q.1 Highest leak correlation (star) of leak localization result with the uniform model for location I during the morning with a leak size of 15 m^3/h.

In this leak localization is the node that has the highest correlation, located in the left branch directly after the entrance of the DMA (star in Figure Q.1). A leak in this area, results in very small pressure differences at the sensor locations between the leak free situation and a leak at this location (-0.00188 mWc and -0.00189 mWc, FSMHC column in Table Q.1). This is because there is no sensor located in this branch. The modelled pressure differences are presented in Table Q.1 and this vector is abbreviated to the FSMHC. The FSMCHC is a column from the Fault Signature Matrix (FSM) that results in the Highest Correlation with the FIV. The measured pressures subtracted with the leak free situation are also shown in Table Q.1 (FIV).

Table Q.1 FIV and FSMHC of leak localization with the uniform model for location I during the morning with a leak size of 15 m^3/h

Pressure sensor	Meter name		FIV	FSMHC
0	**PL196**	2709	-0.31946	-0.00188
1	**PL202**	2618	-0.24454	-0.00188
2	**PL222**	2459	-0.29242	-0.00188
3	**PL255**	233	-0.33415	-0.00188
4	**PL277**	2048	-0.29277	-0.00188
5	**PL279**	2137	-0.09527	-0.00188
6	**PL319**	2274	-0.13823	-0.00188
7	**SM1023**	950	-0.27829	-0.00188
8	**SM1025**	1812	-0.41933	-0.00188
9	**SM1027**	2123	-0.32692	-0.00188
10	**SM1032**	2319	-0.6542	-0.00189
11	**SM1033**	2529	-0.30696	-0.00188
12	**SM1034**	2285	-0.38313	-0.00188

The FIV and FSMHC are used to calculate the Pearson correlation. The values at the sensor location for the FIV and FSMHC from Table Q.1 are visualized in Figure Q.2. A line is drawn between the values at the sensor locations. The FSMHC looks horizontal in Figure Q.2. In Figure Q.3, however, a zoom-in of this line is shown. The values of the FMSHC are extremely small but the computer calculates these small differences. The correlation coefficient is only dependent on the linear relation and not on the order of magnitude. The linear relation between the FIV and FSMHC is high while the order of magnitude is completely different.

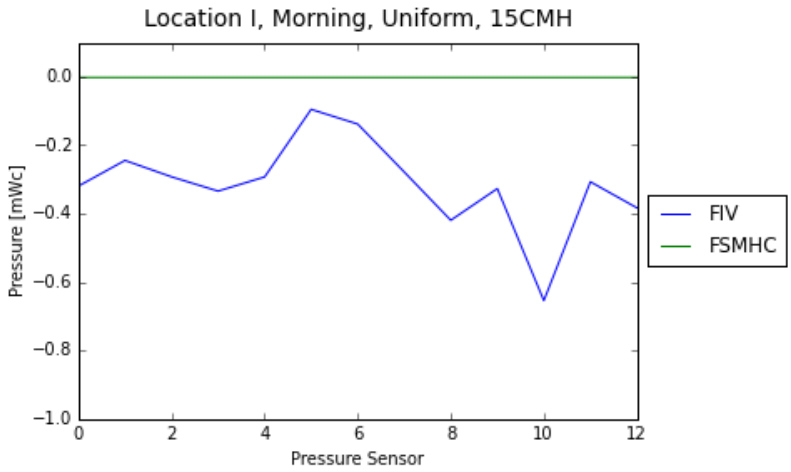

Figure Q.2 Continuous line between FIV and FSMHC values at the sensors

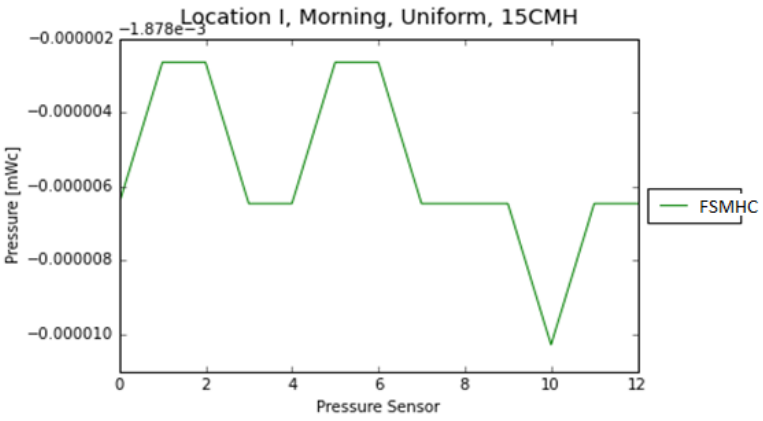

Figure Q.3 Zoom-in of the FSMHC

APPENDIX R DETAILED PERFORMANCE LEAK LOCALIZATION REAL LEAK

The leak localization results can be used between 7 December 15:00 and 10 December 12:00. Before and after this time, DMA Leimuiden was operated in a different way and fed by two inlets. Figure R.3 and Figure R.4 show the results of the leak localization performance on 9 December. The blue line represents the DMA inflow. This inflow shows a clear peak every hour. The origin of this peak is unclear and can influence the performance in a negative way.

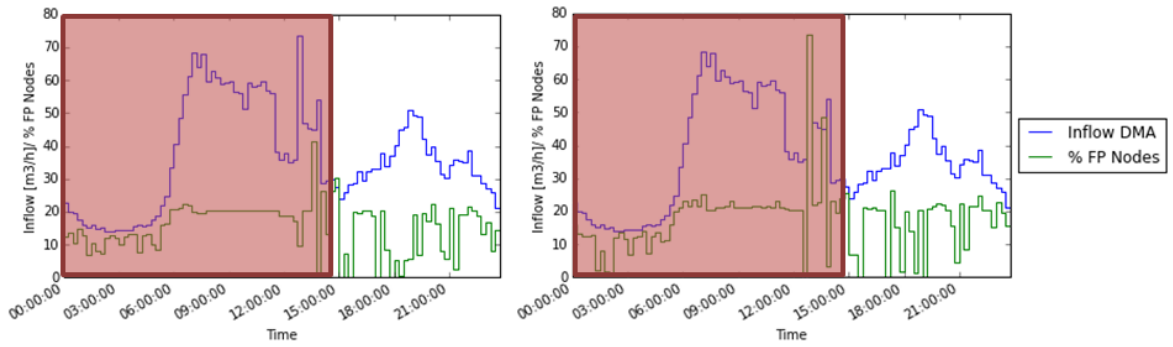

Figure R.1 Performance leak localization real leak of one day with a time analysis step of 15 minutes (2015-12-07). Left: uniform model. Right: factorized model.

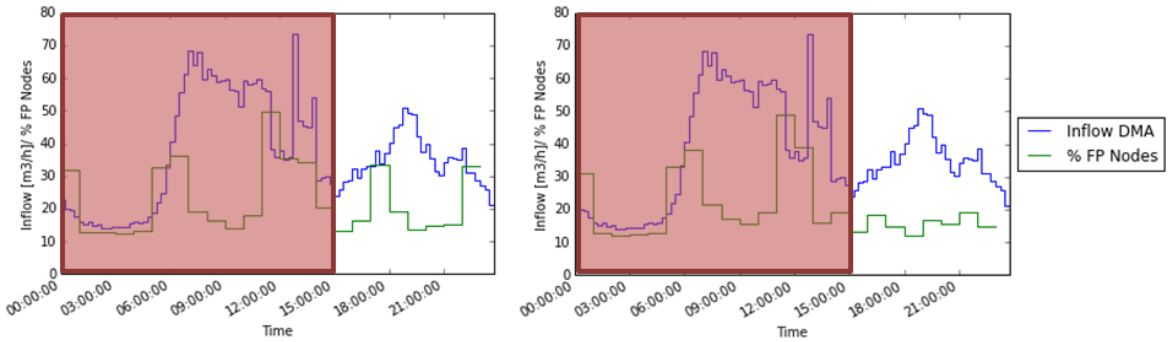

Figure R.2 Performance leak localization real leak of one day with a time window of 60 minutes (2015-12-07). Left: uniform model. Right: factorized model.

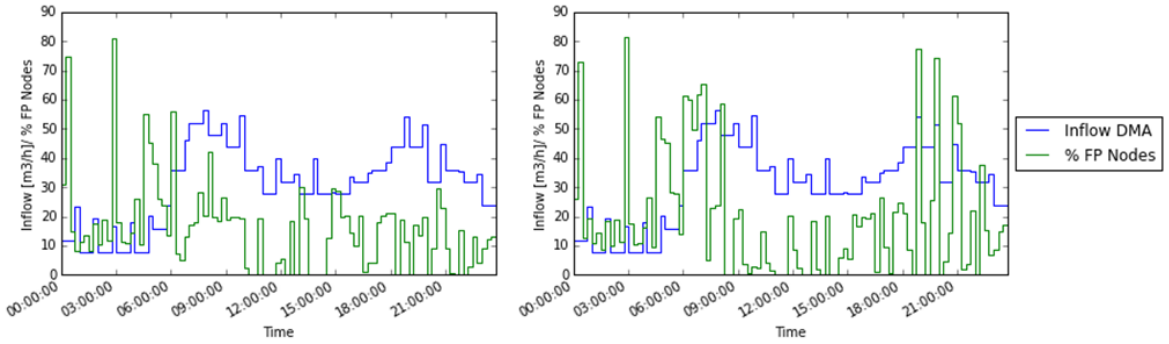

Figure R.3 Performance leak localization real leak of one day with a time analysis step of 15 minutes (2015-12-09). Left: uniform model. Right: factorized model.

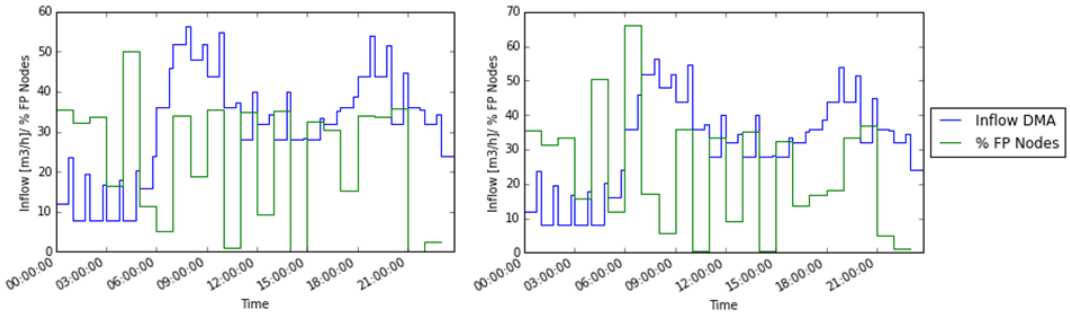

Figure R.4 Performance leak localization real leak of one day with a time window of 60 minutes (2015-12-09). Left: uniform model. Right: factorized model.

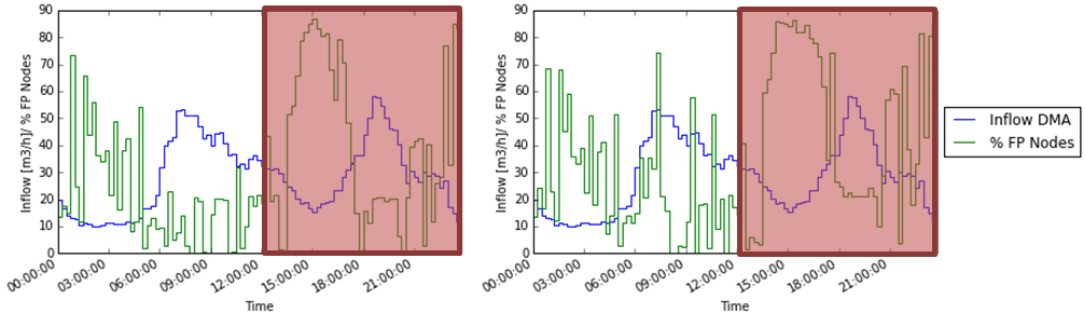

Figure R.5 Performance leak localization real leak of one day with a time analysis step of 15 minutes (2015-12-10). Left: uniform model. Right: factorized model.

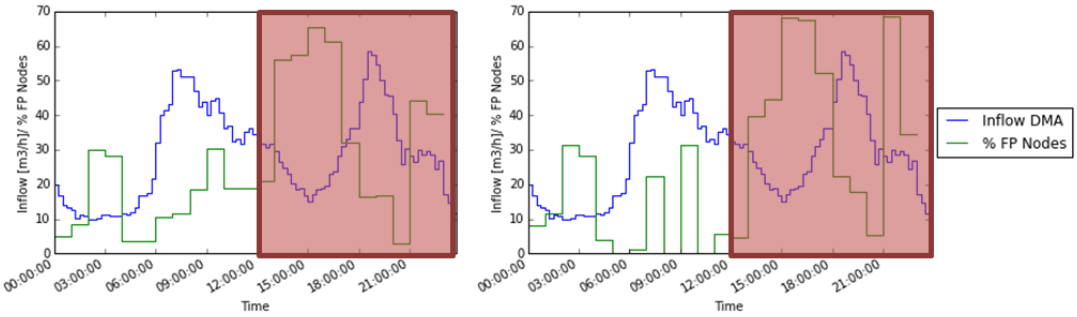

Figure R.6 Performance leak localization real leak of one day with a time window of 60 minutes (2015-12-10). Left: uniform model. Right: factorized model.

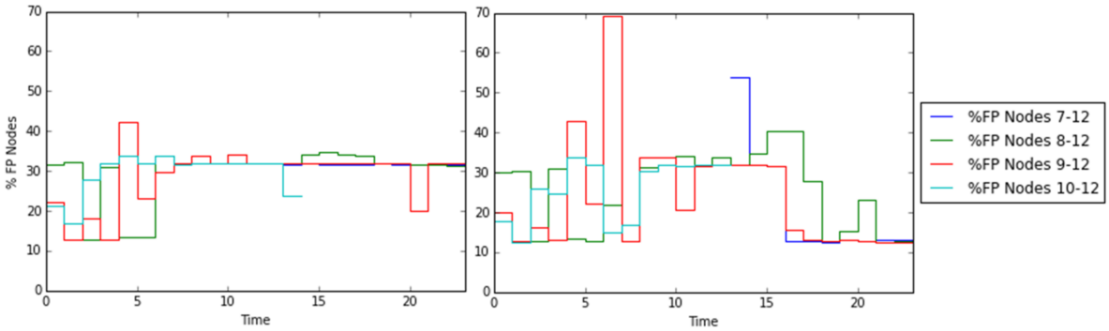

Figure R.7 Performance leak localization real leak of all days with a time window of 60 minutes. Left: uniform model. Right: factorized model.

APPENDIX S COMPARISON LEAK LOCALIZATION IN DMA NOVA ICÀRIA

The accumulated results of the real leak in DMA Leimuiden are compared with the leak localization result of a forced leak of 20.2 m^3/h in DMA Nova Icària (Figure S.1) (Pérez et al., 2014a). The differences between these two cases are shown in Table S.1. Main differences between these two cases are the leak size (20.2 m^3/h, 5.2 m^3/h in DMA Nova Icària, DMA Leimuiden respectively, see Table S.1) and the delivered monthly volume (170,000 m^3, 20,000 m^3 in DMA Nova Icària, DMA Leimuiden respectively, see Table S.1). The leak localizations in DMA Nova Icària use a larger time analysis step (1 hour, 15 minutes for DMA Nova Icària, DMA Leimuiden respectively, see Table S.1). The consumption distribution model in DMA Nova Icaria is a factorized model (based on billing information) but uses different patterns for different groups of customers (i.e. domestic, industry). The time analysis step is smaller for DMA Leimuiden compared to DMA Nova Icària (1 h, 0.25 h for DMA Nova Icària, DMA Leimuiden respectively, see Table S.1).

Table S.1 Differences between leak in DMA Nova Icària and leak in DMA Leimuiden

	DMA Nova Icària	DMA Leimuiden
Leak size [m^3/h]	20.2	5.2
Pipe network length [km]	43	26
Area [km^2]	2.27	1.46
DMA inlet(s)	2	1
Inner pressure sensors	5	6
Network Characteristic	Mainly looped	Looped and branched
Delivered monthly volume [m^3]	170,000	20,000
Time analysis step [hour]	1	0.25
Accumulation time window [hour]	48	24

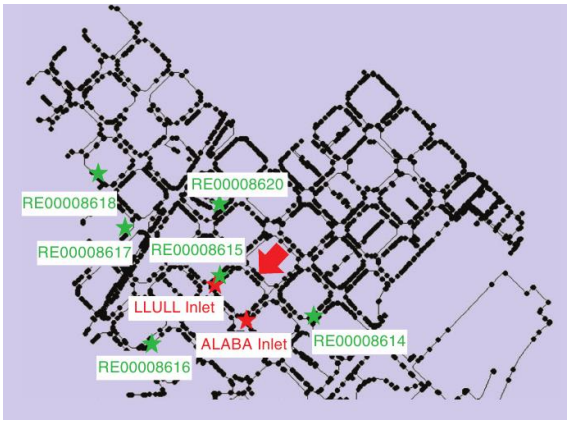

Figure S.1 DMA Nova Icària with inner pressure sensors (green stars) , two inlets (red stars) and the exact location of the leak (red arrow). RE000008615 was faulty during leak localization of 20, 21 December 2012 and is not used.

Figure S.2, shows the accumulated result of correlations larger than 0.5 for 48 hours: the larger the star size, the larger the corresponding accumulated correlations. The red star represents the leak and a red outline ellipsoid is drawn around the nodes with the highest correlations (Pérez et al., 2014a). It is unclear what the exact definition of 'high correlations' is in Pérez et al. (2014a).

In Pérez et al. (2014a) the result is not quantified and considered as acceptable. The same result is shown in Mirats-Tur et al. (2014) where two different leak localization methods are compared. It is stated that the leaks shown in Mirats-Tur et al. (2014) are localized within a circle of 150 meter radius. The result in Barcelona is more

spread than the result in DMA Leimuiden. This can be caused by the highly looped structure and two inlets in DMA Nova Icària. A leak in the looped structure is harder to detect. This is, however, also influenced by the flow regimes in the pipes and therefore of the diameters and consumption in the DMA.

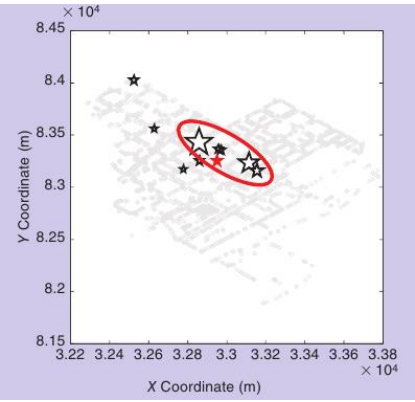

Figure S.2 A graphical representation (black stars) of the most probable localization of the leak according to the accumulation of the resulting correlation vectors related to every node through a 48-h time window (December 20 and 21). The bigger the star size, the bigger the corresponding accumulated correlations. Additionally, the real location of the leak (red star) and the area containing those nodes with the highest correlation values (red outlined ellipse). Figure 11 in Pérez et al. (2014a).

Similar to this research, different outcomes for different hourly leak localizations are found (Figure S.3). The maxima correlations range between 0.54 and 0.75. These values are similar to the hourly results of DMA Leimuiden in the afternoon and evening (Figure 3.14). This in contrast to papers where only leaks in the model were created and 'noise' was added to make it more realistic. Leak correlations were above 0.99 (Quevedo et al., 2011).

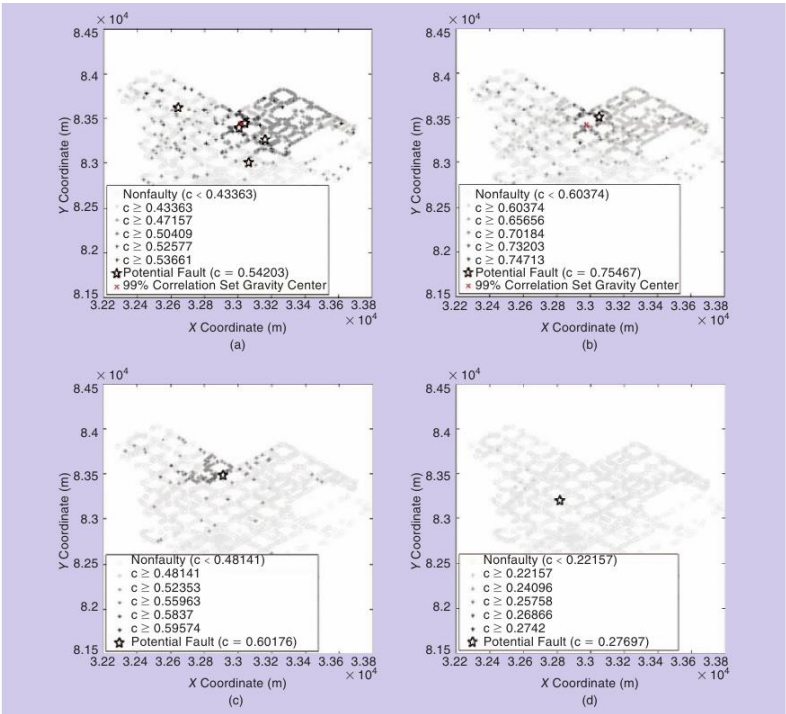

Figure S.3 Hourly leak localization results DMA Nova Icària and leak size (20.2 m^3/h). a: 15 h, b: 15 h, c:22 h and d:20h after the leak was repaired. D shows a decrease of the maximum leak correlation. Figure 10 in Pérez et al. (2014a).

Figure S.4, shows another leak localization result in DMA Nova Icària (Mirats-Tur et al., 2014; Meseguer et al., 2014). In this case there was a single leak of 21.6 m^3/h at the location of the red star and in addition, multiple leak were located that added up to 14.4 m^3/h in the red rectangle. Before the single leak was located, the leak localization method indicated a leak that was in between the two leaks (green rectangle). After the reparation of the single leak (red star in Figure S.4), the leak localization of the multiple leaks was in the area where the multiple leaks were. Meseguer et al. (2014) do not describe how the size of this rectangle is defined except that this is the most rated location proposed by the leak localization (Meseguer et al., 2014).

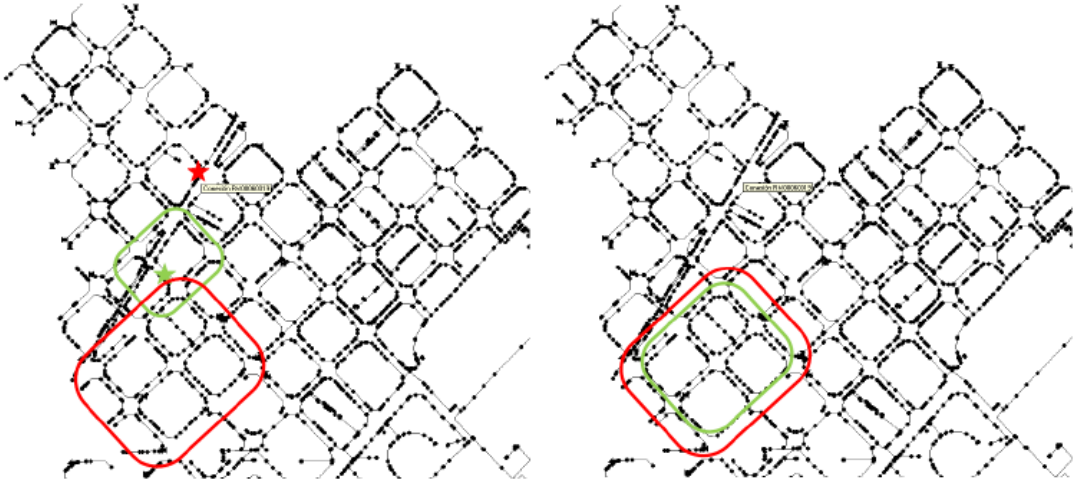

Figure S.4 Leak localization result in DMA Nova Icària of leak size 21.6 m^3/h and multiple leak that add up to 14.4 m^3/h. Left: Green rectangle represents the leak localization of leak (21.6 m^3/h) (red star) and multiple leak that add up to 14.4 m^3/h (red rectangle). Right: Result leak localization (green rectangle) after main leak was repaired. Figure 2 in Mirats-Tur et al. (2014).