A Comparison Between Machine Learning Techniques to Find Leaks in Pipe Networks

By

Joseph Cornelius van der Walt

A dissertation submitted in partial fulfillment of the requirements for the degree

Master of Engineering

in the Department of Mechanical and Aeronautical Engineering

Faculty of Engineering, the Built Environment and Information Technology

University of Pretoria

2017

A Comparison Between Machine Learning Techniques to Find Leaks in Pipe Networks

By

Joseph Cornelius van der Walt

Supervisor: Prof. P.S. Heyns, Dr. D.N. Wilke
Department: Mechanical and Aeronautical Engineering
Degree: Masters of Engineering

Abstract

In 2012, the National Non-Revenue Water assessment revealed that South Africa has 37% of non-revenue water. With the steadily growing demand for this scarce resource, the detection of leaks in pipe networks is becoming more important. Currently, in South Africa the primary method of detecting leaks is to install pressure management systems and monitoring minimum night time flows [1].

The pressure-flow deviation method, can be used to formulate an inverse analysis model based leak detection problem. This problem can then be solved using Artificial Neural Networks, Support Vector Machines and other optimization methods.

With EPANET, different networks were tested to compare these methods to finding leaks, using an inverse analysis formulated problem. Four different numerical networks were modeled and tested, a simple single pipe network, a small agricultural site, a distribution network proposed and investigated by Poulakis et al. [2] and the simulated model of the experimental network that was designed and commissioned during the study in our laboratory.

From the numerical investigation, it was found that the optimization methods struggled to find solutions for simple networks with infinite number of solutions for the problem. For more complex numerical networks, it was seen that the Support Vector machine and the Artificial Neural Networks trained to the averages of their respective data sets.

Errors to ensure an accurate solution found by these algorithms were calculated as 2.6% for the numerical experimental network. The experimental network consisted of six possible leaking pipes, each having a length of $3m$ and a diameter of $10mm$. Three leak cases were tested with diameters of $3mm$ and $2mm$. Overall, the Support Vector machine could locate the leaking pipe with the best accuracy, while the minimizing of non-regularized error could calculate the size and location of the leak the most accurately.

Multiple leak cases were measured with the experimental network. The Support Vector machine was tested on these measurements, where it was found that two of the three leak cases could be solved with relative accuracies. Sensor usage optimization was completed on

the measurements for the experimental network, where it was found that the leaks could be classified correctly with probabilities higher than 98% if only two sensors were used in the training of the SVM instead of all twelve.

Overall this method of leak detection shows promise for certain applications in the future. With practical applications on water distribution, transportation, and agricultural networks.

Contents

Nomenclature

Symbols

c	The speed of light	m/s
C	Emitter discharge coefficient	
D	Nodal demands	l/s
F, Q or q	Flows in pipes or links	l/s
H or h	Pressure or head at junctions or nodes	m
K	Minor loss coefficient	
M	Mass flow rate in pipes or links	kg/s
P	Pipe number or location	
σ	Uncertainty in the measurements	
τ	Time lag between two signals	s
θ	Set of unknown parameters	
ω	Rotational speed	rev/s
$\mathbf{x}$	Set of observed parameters	
γ	Emitter pressure exponent	

Subscripts

i and j	Node numbers of pipe ends
in	Input into a network or system
L	Some loss of a parameter
max	Maximum of some parameter
out	Output of a network or system

Abbreviations

ANN	Artificial Neural Network
DMA	District Metered Area
MSE	Mean Squared Error
NRW	Non-revenue Water
PIG	Pipeline Inspection Gauge
SVM	Support Vector Machine

Chapter 1

Introduction

In this work various strategies to formulate and solve the inverse problem are compared to find the best algorithm to detect leaks in water distribution networks. This chapter contains the background to the problem, the problem statement and the literature study.

1.1 Background

In South Africa, water is a scarce and important resource with a steadily growing demand. In $2011 - 2012$ the National Non-Revenue Water assessment [1] found the average national non-revenue water to be 37%. While the world average water loss was found to be 36.6%. Some municipalities in South Africa were found to have up to 68% non-revenue water, which shows a clear scope for improvement. South Africa could however be compared to Australia, which has similar water resources and a total average water loss of only 10%.

The Draft Second National Water Resources Strategy set a target to reduce the non-revenue water in municipalities to 15% by 2014 [3]. The assessment also shows that on average, every person uses about $273l$ of water every day. This indicates the average person is not realizing the scarcity of water.

Currently, to reduce non-revenue water in South Africa, pressure management systems are installed, minimum night flows are logged and water balances are completed. The methods used reduce the amount of leaked water or makes the municipalities aware of the leaks. The primary method of locating leaks comes from reports from the consumers. External contractors are used to find leaks when they are aware of them. These contractors use listening sticks, geophones, ground penetrating radar and noise loggers. More advanced techniques such as the Smartball, the Saharah system and JD7 systems are also used[1].

In recent years, research to find leaks using on-line machine learning techniques with real time data has intensified. Installation of pressure sensors with machine learning techniques to find leaks was shown to be popular. Researchers were found to test their own algorithms on unique water networks. Various theoretical and practical networks have been tested, with machine learning techniques such as Neural Networks, Support Vector Machines and other statistical approaches. A generalization between these techniques are required to find

the best technique for varying applications on different water distribution networks. This is inherently an inverse problem and this study investigates different ways to formulate and solve this inverse problem.

1.2 Literature Study

In this literature study, different leak detection methods and techniques were investigated to find the most applicable techniques. This was followed by the software used to simulate water networks and finally, relevant work done within the leak detection field.

1.2.1 Leak Detection Methods

Leak detection overall is an inference problem, where certain aspects of the network can be measured to estimate the leak size and location. The leak detection methods most commonly used can be seen in the following paragraphs. These methods are subdivided into three groups, steady state methods, transient state methods and methods that use specific equipment. These methods were presented by Wang et al. [4] in a review on leak detection methods using hydraulic models.

1.2.1.1 Steady State Methods

Steady state methods analyse pipe lines and networks when the supply and demand acting on the network stays constant. Some steady state methods have been extended to be used on transient state networks [4].

Hydrostatic leak detection methods
The hydrostatic leak detection methods examine the pipeline when there is no flow in it [4]. This is usually achieved by closing of sections of pipe, and measuring the pressure. If there are any leaks in the pipeline, it can easily be identified by pressure drop over time.

This method of leak detection is most commonly used by pipe manufacturers to test the strength and quality of the pipe before operation. This method is also used by pressure vessel manufacturers, and by electrical engineers to find leaks in high pressure oil filled electrical cables.

This method can not find the exact location of the leaks, but the location can be approximated by adding flow meters into the system. The flow direction can be used to approximate the leak location.

1.2.1.2 Transient State Methods

Transient state methods analyze pipelines and networks when the supply or demand vary.

Transient Damping Analysis
This method of leak detection measures the transient pressure waves in a pipeline. The transients decay as they move through the pipe due to pipe friction and leaks. The leak can be detected by comparing the leak induced damping to the transient damping of a pipeline

without leaks.

With this method laboratory experiments have been completed where a leak could be located and quantified with accuracies up to 0.2% of the cross sectional area. It is found that this method can detect leaks in a pipeline, but it is unlikely for this method to find the position of the leaks in a pipeline [4].

Reflected Pressure Waves

This method sends a transient pressure wave through a pipeline. When it reaches a leak, a portion of the wave is reflected. The location of the leak in the pipeline can be calculated by multiplying half the reflection time with the wave speed. The size of the leak can be calculated from the reduction in reflected wave amplitude.

This method has been able to find leaks of 0.5% the cross sectional area. Pipeline elements such as elbows, change in pipe diameter and partially closed valves can also cause transient reflections. This make these methods difficult to implement in pipe networks [4].

Transient Frequency Analysis

In any transient state pipeline, transient pressure waves can be measured. When a leak exists in a pipeline, the measured transients change compared to a no-leak situation. The leak information is hidden within the measured transients. Using a Fourier transform on the measurements, and comparing the dominant frequencies of the leak and no-leak case, the location of the leak can be calculated [4].

This method of leak detection is difficult and accurate results not only depend on the leak, but on the shape of the transient waves and the measurement locations. This method can be improved by instead of measuring the freely damped transient waves, a steady-oscillatory wave can be generated by using a periodic forcing function at the boundary of the pipeline. The response of the pipeline system to different frequency forcing functions determines the leak location and size.

1.2.1.3 Methods that can be used in Steady and Transient States

Mass Balance

Mass balance [4] is used to complete water audits. This simple technique measures the difference between the inlet volume, the outlet volumes and the change of mass in the network due to compressibility. This can be written as:

$$M_L = M_{in} - M_{out} - \Delta M. \tag{1.1}$$

In this equation, M_L is the mass of water leaked, M_{in} is the mass flow into the network, M_{out} is the mass flow out of the system, and ΔM is the change in mass inventory of the system.

The accuracy of this method depends on the accuracy of the flow meters and the calculation of the network mass inventory. In water audits the network mass inventory is usually ignored. This method is mostly used to find the integrity of a pipeline in real time. The

response time depends on the size of the leak, the complexity of the pipeline system and the accuracy of the flow meters. This method cannot detect the exact location of the leaks, but can give estimates of the leak locations within sectors of the network.

Pressure-Flow Deviation

The pressure-flow deviation method [4] is a model-based leak detection method and is sometimes called inverse analysis method. This method consists of pressure and flow sensors installed at strategic points in the network. If a leak occurs, the measured pressures and flow rates will deviate from the expected pressures and flow rates. The leak location can be estimated by using a model of the network. This approach requires an inverse analysis to identify model parameters to simulate model measurements and match them to the actual measurements of the network. This is done by optimizing the roughness coefficient, minor loss coefficients and emitter coefficients.

Various investigations into this method has been completed with reasonable results. In Japan, the pressure-flow deviation method was installed on the Niigata-Sendai gas pipeline. Based on the tests, a minimum detectable leak flow rate of 1.1% was achieved.

1.2.2 Alternative Leak Detection Techniques and Equipment

Alternative techniques and equipment that are used to find leaks in pipe networks inlcude acoustic, vibrational, visual, electromagnetic, radio frequency and ultrasound techniques. Zangenhmadar et al. [5] investigated all the leak detection methods used in water distribution networks. These methods are given below.

1.2.2.1 Acoustic and Vibration Techniques

These techniques use the acoustical vibrations generated by the leak or by generating its own sound to detect the leak location.

Sonar Profiling System

The sonar profiling system consists of a scanner unit, a skid and a sonar siphon float. The device travels through the pipeline while sending out sound waves. The waves travel through the cross section of the pipe, the device listens for the reflected waves due to possible leaks. An advantage of this device is its capability of working under water. Due to the speed of sound in water and air being different, the device can only work in one fluid medium at a time.

The frequency of the sound waves can be adjusted. Higher frequencies will give better resolution, while low frequencies results in higher penetration. This device can estimate corrosion loss and volume of debris in a pipe [5].

LeakfinderRT

The LeakfinderRT [5] system consists of acoustic sensors such as accelerometers or hydrophones. Two of these sensors are installed onto a pipeline at two different access points. The system uses a cross-correlation function to find the leak. The cross-correlation function

finds the time lag between two similar wave forms. The leak location can be calculated with:

$$L_1 = (D - c\tau_{max})/2 \tag{1.2}$$

$$L_2 = D - L_1 \tag{1.3}$$

where L_1 and L_2 indicate the lengths from each sensor to the leak, D is the total length of the pipe, c is the speed of sound in the fluid medium and τ_{max} is the time lag between the two signals.

Noise in the measured signals requires additional processing to calculate the leak location. This device cannot find the size of the leak and does not work if there are discontinuities, such as bends, pumps, and valves, in the pipeline.

Sahara System

The Sahara system [5] consists of a single listening device. The device is inserted into a pipeline and travels through it with the help of a parachute. The noise in the pipe is recorded and by using a locator above the surface, the position of the leaks can be located.

Environmental noise such as rivers or municipal constructions influence the system. Recent Sahara systems have video and lighting sensors to inspect defects more accurately.

Smartball

The Smartball [5] consists of various sensors, including a magnetometer, an ultrasonic transmitter, accelerometers and a temperature sensor. These sensors are encased in a ball, surrounded with an outer layer to shield from traveling noise. The diameter of the ball should be less than one-third of the pipe diameter. The size of the ball is selected depending on the size restrictions of valves and access points to the pipeline.

The device is inserted into the pipeline and can collect information for up to 12 hours inside pipelines. This device can find leaks within a one meter accuracy by analyzing the recorded data.

1.2.2.2 Visual Techniques

These techniques include all techniques where visual aids are required.

Closed-Circuit Television

Closed-circuit television [5] is a cheaper and safer alternative to direct human entry into pipelines. This technique consists of a camera being inserted into a pipeline which is connected to some winch system. The camera records the internal surface of the pipeline. The footage is reviewed to try and find leaks. A problem with this method is the interpretation of the footage by the operator. This allows for misinterpretation of the footage, depending on the experience of the operator. A solution to this is to use image processing techniques to assess the footage automatically.

Laser Scan

The laser scan [5] method consist of a carrier inside a pipe, with a spinning laser mounted to it. This device scans the interior of the pipe wall over its length to find defects. This method can estimate, the roughness of the pipe and the corrosion loss.

Factors affecting this method is the carriers velocity, the speed of the spinning laser, sampling rate, roughness and the color of the pipe wall. Diffraction of the laser under water seems to affect the accuracy of this method as well.

1.2.2.3 Electromagnetic and Radio Frequency Techniques

Magnetic Flux Leakage

This method uses magnets to initiate a magnetic field in metallic pipes [5]. Magnetic sensors measure the changes in the magnetic field due to defects. The system is mounted on a pipeline inspection gauge (PIG), which is moved through the pipeline.

This technique has been improved by exciting the magnetic flux with a pulse. This allows for estimation of the depth of the defects in the pipeline wall. Other sensors have also been added to the PIG systems to help increase the performance.

Hydroscope Technology

This method is based on the principles of eddy current technology [5]. It transmits electromagnetic signals through the pipe wall. A detector measures the changes in the electromagnetic signal, indicating defects in the pipe. The system is inserted into the pipeline in the form of a PIG. This technology can inspect up to 1000m a day and can work in water.

Rapid Magnetic Permeability Scan

This method uses a couple of strong magnets which are installed inside a pipeline [5]. The magnetic flux induced by the magnets can be measured. The magnetic flux density is related to the cross section of the pipe. Changes in the flux indicate defects in the pipe cross section.

Sharp changes in the flux indicate a crack while gradual changes indicate wall thickness reduction due to corrosion. This method of leak detection is used on metallic pipelines with a diameter of 100mm or greater.

Ground Penetrating Radar

Ground penetrating radar [5] works by sending electromagnetic waves into the ground. A reflection is received based on the density of the soil below the radar. Defects in the pipeline can be found as voids in the ground created by the leaks. This technique works well up to a depth of 5m. The operator of this device requires skill to interpret the signals correctly.

Two different types of ground penetrating radar exists, traditional and in-pipe. The traditional ground penetrating radar can be categorized in three groups, time domain, frequency domain and spatial domain. In-pipe ground penetrating radar have two modes of inspection, look-through and look-out. With look-through mode both the transmitter and receiver is placed inside the pipe. With look-out mode the transmitter is placed inside the pipe while

the receiver is placed on the surface of the ground.

A new method called ground penetrating imaging radar has been developed which generate 3D images of the pipeline underground. The image has a resolution of 50mm and can detect leaks in all pipe materials. The survey velocity is roughly 0.36km/h.

1.2.2.4 Other Techniques

Other techniques used to find leak in water distribution and supply networks. These techniques are external techniques which requires access to the pipe.

Guided Wave

This method of leak detection is based on the propagation of waves through pipeline walls [5]. It consists of transducers mounted on the outside of pipes which sends and receive waves. Two types of waves can be excited, torsional and longitudinal. This technique can be used on pipelines with a diameter above 50mm and wall thickness up to 40mm.

The propagating wave through the wall will partially be reflected when it hits a defect in the pipe. The location of this leak can be calculated by taking half the wave travel time. The size of the leak can be estimated from the amplitude of the reflected wave. Plastic pipes damp the sound waves and access to the pipes is required.

Infrared Thermography

Infrared thermography [5] consists of a scanner and a camera. The scanner measures temperature changes and produces thermographic images. At the position of a leak, a temperature difference occurs in the soil. This method of leak detection is sensitive, reliable and requires minimum instrumentation.

Infrared thermography cannot detect the depth of the leak or the size. Ground cover, moisture content and wind speed influence the results.

Continuous Wave Doppler Sensing

The system uses a 2.45GHz continuous wave Doppler sensing unit, transmitter and a receiver [5]. The sensing unit sends a signal and receives reflected waves back. The reflected waves are shifted if there is a leak in the pipe. This method of leak detection works in all environmental conditions, soil types and pipeline materials.

Acoustic Fiber Optics

This method of leak detection consist of a fiber optic cable along the outside length of a pipeline. A laser shines through the fiber. An optical sensor is placed at the same end. The acoustic noise from a leak causes reflections of the light through the fiber. This system is susceptible to physical damages.

1.2.2.5 Summary

Various methods of leak detection were investigated, where different common problems could be seen. Table 1.1 shows a summary of the leak detection methods. It can be seen that only the pressure-flow deviation method fit all the criteria.

Table 1.1: Summary of Leak Detection methods

Method	On-line usage	Material limits	Pipe access	Location estimation
Hydrostatic				x
Transient Damping		x		
Reflected waves		x		x
Transient Frequencies		x		x
Mass balance	x			
Pressure-flow Deviation	x			x

A summary of leak detection equipment can be seen in Table 1.2. It can be seen that most methods require pipe access, which results in the location being estimated accurately. Some of these methods can be seen to have material restrictions, with a few of them having on-line usage capabilities.

Table 1.2: Summary of Leak Detection equipment

Method	On-line usage	Material limits	Pipe access	Location estimation
Sonar Profiling			x	x
LeakfinderRT		x	x	x
Sahara System			x	x
Smartball			x	x
CCTV	x		x	x
Laser Scan		x	x	x
Magnetic Flux		x	x	x
Hydroscope		x	x	x
Magnetic Permeability		x	x	x
Ground Penetrating Radar				x
Guided Wave		x	x	x
Infrared Thermography				x
Doppler Sensing			x	x
Fiber Optics	x		x	

The best methods to investigate are therefore the pressure-flow deviation method. This method is a model based inverse analysis problem, which act as a limitation of this method, although this allows for application on any network of which a model is available.

1.2.3 Pipe Modeling Software

EPANET is a program commonly used for modelling and research purposes of water networks. The software was developed by the United States Environmental Protection Agency. The software was created to model water distribution piping systems and to simulate the movement of water and the quality thereof [6].

The software uses pipes, nodes, pumps, valves and storage tanks to build and solve models. The software can calculate the flow in every pipe, the pressure at each node, the height of water in each tank, the type of chemical concentration throughout the network during a simulated period, water age, water source and water tracing.

The software has a programmers toolkit allowing it to be used in C/C++, Delphi, Pascal, Visual Basic or any other language that can call functions from a Windows DLL.

EPANET has five main input parameters for pipes, start and end nodes, pipe diameter, pipe length, roughness coefficients and a status. The status either opens or closes a the pipe to flow. the main computed output parameters for pipes are flow rate, velocity and headloss. Nodes have two input parameters, elevation and water demand. Calculated output parameters for nodes are hydraulic head and pressure.

1.2.3.1 EPANET Algorithm to Solve a Network

EPANET [6] solves the flow continuity and headloss equations to find solutions for different water networks. This algorithm is called the hybrid node-loop approach.

The first equation is the relationship between the headloss and flow. If the pipe network has N junctions and NF tanks and reservoirs, and i and j represent the two nodes at the end of a pipe, we can write [6]:

$$H_i - H_j = H_{ij} = rQ_{ij}^n + mQ_{ij}^2, \tag{1.4}$$

where H is the nodal head, r is the resistance coefficient, Q is the flow rate, n is the flow exponent and m is the minor loss coefficient. The resistance coefficient depends on the friction headloss formula selected. The second equation to be solved is the continuity equation. This is given by:

$$\sum_j Q_{ij} - D_i = 0 \quad \text{for} \quad i = 1...N, \tag{1.5}$$

where D_i is the flow demand at node i. If we know all the heads at the tanks and reservoirs, we want a solution for H_i and Q_{ij} so that the residual equal zero. Newton's method requires an initial estimate for the flows in each pipe which is updated in subsequent by solving:

$$\mathbf{AH} = \mathbf{F}, \tag{1.6}$$

where $\mathbf{A}$ is a $N \times N$ Jacobian matrix, $\mathbf{H}$ is a vector of unknown nodal heads, and $\mathbf{F}$ is a vector with the flows in each pipe. The diagonal elements of the Jacobian matrix are given by:

$$A_{ij} = \sum_{j=1} P_{ij}, \tag{1.7}$$

and the off-diagonal terms are:

$$A_{ij} = -p_{ij}, \tag{1.8}$$

where p_{ij} is the inverse derivative of the headloss in the link between the two nodes i and j. This term for pipes are:

$$p_{ij} = \frac{1}{nr|Q_{ij}|^{n-1} + 2m|Q_{ij}|}, \tag{1.9}$$

The flow imbalance at a node plus the flow correction factor is calculated as:

$$F_i = (\sum_j Q_{ij} - D_i) + \sum_j y_{ij} + \sum_f p_{if} H_f)_f, \tag{1.10}$$

where the last term applies to any links connected to tanks or reservoirs and the flow correction factor y_{ij} is:

$$y_{ij} = p_{ij}(r|Q_{ij}|^n + m|Q_{ij}|^2) sign(Q_{ij}), \tag{1.11}$$

for pipes. The new heads of the next iteration are calculated by:

$$Q^*_{ij} = Q_{ij} - (y_{ij} - p_{ij}(H_i - H_j)). \tag{1.12}$$

This is repeated until the flow change relative to the total flow in all links is smaller than some tolerance.

1.2.3.2 Calculation of the Friction Coefficient

Hydraulic headloss due to friction can be calculated by three different formulas namely, Hazen-Williams, Darcy-Weisbach and Chezy-Manning [6]. To calculate the headloss all three formulas follow the same equation [6]:

$$h_L = Aq^B, \tag{1.13}$$

where h_L is the headloss, q is the flow rate, A is the resistance coefficient and B is the flow exponent. The resistance coefficient and flow exponent are calculated individually depending on which formula is used. The default formula used is the Hazen-Williams formula, which is given by:

$$A = 4.727C^{-1.852}d^{-4.871}L \quad \text{and} \quad B = 1.852, \tag{1.14}$$

where C is the Hazen-Williams roughness coefficient, d is the pipe diameter and L is the pipe length. Table 1.3 shows the Hazen-Williams roughness coefficients for different materials. It can be seen that steel and plastic pipes have the highest friction coefficients, with clay pipes having the lowest.

Table 1.3: Hazen-Williams Roughness coefficients for different materials

Material	Hazen-Williams Roughness C
Cast Iron	130 – 140
Concrete	120 – 140
Galvanized Iron	120
Plastic	140 – 150
Steel	140 – 150
Vitrified Clay	110

Minor Losses

Turbulence created by fittings in a network creates minor losses. Depending on the accuracy required, these losses can be added into a pipe network. The headloss due to minor losses can be calculated by [6]:

$$h_L = K(\frac{v^2}{2g}), \tag{1.15}$$

where K is the minor loss coefficient, v is the flow velocity and g is the acceleration due to gravity. Table 1.4 shows minor loss coefficients for a few different types of fittings. It can be seen that a glove valve fully opened has the highest loss coefficient, with a gate valve fully open having the smallest loss coefficient.

Table 1.4: Minor Loss Coefficient for some Fittings

Fitting	Loss Coefficient
Glove Valve, Fully Open	10
Angle Valve, Fully Open	5.0
Swing Check Valve, Fully Open	2.5
Gate Valve, Fully Open	0.2
Short Radius Elbow	0.9
Medium Radius Elbow	0.8
Long Radius Elbow	0.6
45 Degree Elbow	0.4
Closed Return Bend	2.2
Standard Tee, Flow Through Run	0.6
Standard Tee, Flow Through Branch	1.8
Square Entrance	0.5
Exit	1.0

1.2.3.3 Approaches to Model Leaks

Three methods of modeling leaks can be used. These techniques are using emitter coefficients, demands and adding reservoirs to allow flow out of the network.

Emitter Coefficients Emitters can be added to nodes. These emitters simulate flow through a nozzle or orifice that discharges to the atmosphere. The calculation of the flow rate through the emitter can be calculated with [6]:

$$q = Cp^{\gamma}, \tag{1.16}$$

where q is the flow rate through the emitter, p is the pressure at the node, C is the discharge coefficient and γ is the pressure exponent. The value of γ is usually taken as 0.5 for nozzles and sprinklers. van Zyl et al. [7] completed research into the effect of leakage on a water distribution network and found that in most cases the emitter coefficient is between 0.5 and 1.0. Table 1.5 shows more details about their findings.

Table 1.5: Leakage Exponents for Different Types of Failure

Failure Type	uPVC	Asbestos Cement	Mild Steel
Round hole	0.52	-	0.52
Longitudinal crack	1.38-1.85	0.79-1.04	-
Circumferential crack	0.41-0.53	-	-
Corrosion cluster	-	-	0.67-2.30

Adding Demands Modeling a leak as a demand on a node gives precise control over the size of the leak. A problem with this method is that the leak will always draw the amount specified, resulting in negative pressures in the system when the specified size of leak is to large.

Adding Reservoirs By inserting an empty reservoir at the position of the leak, giving the pipe connected to the reservoir a small length, similar to the wall thickness of the pipe, and the actual diameter of the leak. This allows for accurate quantification of the leak size when experimental investigations are completed. Minor losses can be added to the small length of pipe to simulate the losses found within orifices.

1.2.4 Research on the Inverse Analysis Method

A focus to research on the inverse analysis or pressure-flow deviation method were done since these methods showed the most promise for leak detection as found in previous sections. The Pressure-flow deviation method measured pressures and flows, which generates an inverse problem to try and find the leak size and location. Various machine learning techniques can be used with this method to find leaks in water networks. The main methods were Artificial Neural Networks, Support Vector Machines, error calculation optimization methods and Kalman Filters. Other leak detection techniques currently researched was also found, with additional research on sensor placement optimization.

1.2.4.1 Artificial Neural Networks

Caputo et al. [8] proposed a method of using artificial neural networks to estimate the leak location in piping network. They performed tests on a network, where they generated input

data for leaking and non leaking states of the network. Two neural networks were used in their proposal, the first identifying the leaking branch and the second estimating the leakage amount and location for the selected branch. Applying the neural networks to a case study they found that the leaking branch could be correctly identified, with the leak size estimated to between $2 - 10\%$ of the actual value. The location of the leak could be estimated within $50 - 100m$ of the actual leak location.

Mounce et al. [9] performed tests on a actual water supply network in the UK. Bursts in the network were simulated by opening fire hydrants. Two sensor locations were used, one at the input of the network and one at the output going to the neighboring DMA. The sensors measured both pressure and flow. Five different burst locations were simulated. They tested different neural network layouts, with the best being a (4,2,1,1) multi layered network. They found that they could locate the bursts with an accuracy of 98.33%.

Salam et al. [10] investigated an on-line monitoring system to detect leakages in pipe networks. They used a network from Makassar in Indonesia. They used pressure measurements at each junction as input data. The input data were generated by simulating leaks in the network by changing emitter coefficients between 0.005 and and 0.3, which results in leak sizes between $0.01l/s$ and $0.6l/s$. They used a Radial Basis Function Neural Network which could detect the leak location and sizes with an accuracy of 98%.

1.2.4.2 Support Vector Machines

De Silva et al. [11] investigated support vector machines to act as pattern recognisers to detect leaks in pipe networks. They started with a SVM as a regressor to try and predict emitter coefficients. Data were generated using a single leaking node, with emitter coefficients between 0 and 0.3. Six monitoring nodes were used to act as sensor locations. After training the SVM, a mean squared error of $4.47 \times 10^{-}5$ was found on the testing data set. They further used a SVM as a classifier to find the leaking node in the pipe network. They selected 10 candidate nodes, and generated a data set with varying emitter coefficients. The SVM could, after training, achieve a testing accuracy of 76.8%. Optimization of the training parameters were needed.

They went further and used 40 candidate leaking nodes, and created a data set. After training the SVM and optimizing the training parameters, a testing accuracy of 57.2% was achieved. Although the result might not seem that good, they found that the predicted leak location was within 500m of the actual leak location in all cases for a network that could fit into a 1000 by 1100m square box.

They went on to investigate if the SVM could detect leaks when small leaks occur in the network. The smallest leak registered by EPANET to generate a pressure difference was a leakage of $90l/hour$. A new data set was created to which the SVM was trained. A testing accuracy of 35% was found.

1.2.4.3 Bayesian Probabilistic Framework

Poulakis et al. [2] investigated a Bayesian probabilistic framework to detect leaks in a water pipe network. This method of leak detection handles uncertainties, measurement errors and model errors. The derivation of their optimization function starts by assigning θ as the parameters to be optimized. These parameters include the leaking pipe, location and size of the leak. It can also be written as $x(\theta)$ to indicate the measured values such as pressure and flow for a given set of leak parameters.
A model error can now be written as:

$$e_{ij} = \bar{x}_{ij} - x_{ij}(\theta), \tag{1.17}$$

where $\bar{x}_{ij}$ are the actual measurements from the system. Another parameter can be added to the parameters to be optimized namely σ. This parameter represents the uncertainty within the error. The unknown parameters are $[\theta, \sigma]$.

Using Bayes theorem, and applying the uncertainty and parameter set to be quantified by a probability density function, $\pi(\theta, \sigma)$, it follows that:

$$P(\theta, \sigma|\bar{x}) = c_1 P(\bar{x}|\theta, \sigma)\pi(\theta, \sigma), \tag{1.18}$$

where c_1 is a normalization constant so that $P(\theta, \sigma|\bar{x})$ integrates to one. Assuming the model error e_{ij} is independent and normally distributed with a zero mean and standard deviation of σ, the likelihood of $P(\bar{x}|\theta, \sigma)$ can be written as:

$$P(\bar{x}|\theta, \sigma) = \prod_{i=1}^{L}\prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_{ij}-\bar{x}_{ij})^2}{2\sigma^2}}, \tag{1.19}$$

where L is the total number of monitoring locations and N is the total number of flow tests. This equation can be simplified by assuming the initial probability density function $\pi(\theta, \sigma)$ is constant and substituting equation 1.19 into equation 1.18. The simplification can be written as:

$$P(\theta, \sigma|\bar{x}) = c_2 \frac{1}{(\sqrt{2\pi\sigma})^{LN}} e^{\left(-\frac{\sum_{j=1}^{N} ||x_{ij}-\bar{x}_{ij}||^2}{2\sigma^2}\right)}. \tag{1.20}$$

By maximizing this function the most probable leak location can be found within a network. Poulakis considered a network that consists of 50 pipes, 31 nodes and 20 loops. The network forms a grid network supplied by one reservoir. The horizontal and vertical pipe sections are 1000 and $2000m$ long. The pipe diameters vary between 300 to $600mm$. Each node has a demand of $50l/s$. They tested this network with a single leak with a $22.8l/s$ demand and found that the leak location could be estimated correctly.

They went on by introducing variation in pipe roughness coefficients, variation in the assumed demands and a variation in the model measurements. They found that when the model measurements had a uncertainty of 2% the location could be calculated. If the uncertainty in the model measurements were increased to 5% the model was unsure about the actual leak location.

1.2.4.4 Other Techniques using fitness functions

Vitovsky et al. [12] investigated leak detection and calibration of pipe roughness coefficients in water distribution networks. They used a genetic algorithm (GA) to maximize the following function:

$$\text{fitness} = -\sum_{i=1}^{M} |H_i^* - H_i|, \tag{1.21}$$

where M is the total number of data points, H_i^* is the measured pressures and H_i is the modeled pressures. They completed a case study with a network consisting of 11 pipes and 7 nodes. The network is supplied from a reservoir and a constant inflow into a network. They optimized the roughness coefficient and emitter coefficient for each pipe and node. The roughness coefficient error was found to be under 3.42% of the average values, and the emitter coefficient could be found within 0.5% of the actual value.

Nasirian et al. [13] created an algorithm combining a step-by-step elimination method with a genetic algorithm. They used this algorithm to minimize the fitness, to which the fewest probable leaking nodes were removed from the optimization procedure. The fitness function they used was:

$$f(x) = \min \left[\frac{\sum_{nh=1}^{N_H} w_{nh} \left(\frac{H_{sim_{nh}} - H_{obs_{nh}}}{H_{pnt}} \right)^2}{N_H} \right], \tag{1.22}$$

where nh is the number of pressure measurements, N_H is the total number of observation nodes, w_{nh} is a weight index, H_{obs} is the measured pressures, H_{sim} are pressure values from the model and H_{pnt} are adjusted parameters at each node.

They investigated three network cases. The first consisted of 9 pipes and 6 junctions. The first system could be solved after three elimination stages. The second network consisted of 30 nodes, 50 pipes and 4 observation nodes. They found that the fitness stayed constant until 11 non-leaking nodes were removed. At the 20th iteration the correct solution could be found.

1.2.4.5 Techniques using Kalman Filters

Choi et al. [14] created a self adjusting sampling time Kalman Filter to detect bursts in water distribution networks. They illustrated the difference between different sampling frequencies for the Kalman Filter when a leak is simulated by calculating the normalized residual. The adjusted sampling frequency reduced the number of false triggers by the filter and could accurately detect the burst on a tested DMA.

Okeya et al. [15] presented a burst detection methodology that utilized the distribution of real time sensor data in a DMA using data assimilation and hydraulic modelling. They created four leaking cases, with random noise of 5%, 10%, systematic errors of +10% and systematic errors of −10%. They tested eight different metrics with the Kalman filter, and found that the corrected flow residuals performed better than pressure residuals. They went on and used Data Assimilation methods to improve the realtime estimation of water demand

and hydraulic system states [16]. They found that the Ensemble Kalman filter outperformed the Kalman filter.

Jung et al. [17] investigated the use of an Extended Kalman filter to detect bursts in water distribution networks. They used the Austin network to which five bursts were simulated, with a size of as small as 2.6% of the total system demand. These bursts could all be detected accurately.

1.2.5 Other Leak Detection Techniques

Asgari et al. [18] investigated a method of detecting leaks by calculating a leak index. The leak index is a non-dimensional value calculated as:

$$LI^k_{j/i} = \frac{\Delta H^k_i}{\Delta H^k_j} \tag{1.23}$$

where i and j represent the node number of observed nodes and k indicates the tested leaking node number. This calculation could be completed for every possible leak location in the network. They added a 5% margin to the results and compared it between observed sensors. Using three observed nodes, the leak could be located between three possible locations. Using four observed nodes, the leak could be located at the correct node.

Caguero-Escofet et.al [19] investigated a method of detecting leaks on networks with demand uncertainty. The algorithm they used calculated the difference between demands and inflow into the network to classify a leaking or non leaking network. The network they used was a small section of a network to demonstrate their work. In the network the boundary conditions, head and outflow of the network is known, with a 5% uncertainty to the network demands. With three possible leak locations, they could locate the first and third leak accurately, with the second leak being classified to the first and third leak.

Zan et al. [20] investigated a method of leak detection using joint time-frequency analysis. They measured a network in Singapore by simulating a leak by opening fire hydrants. The sensor they used measured pressure, flow and water quality (pH and ORP). Measurements was sampled and collected at $2kHz$. A clear pressure drop could be seen as the leak occurred in the time domain. Applying a Gabor transform with conjunction with the Blackman window, they could plot a spectrogram which clearly showed the leak.

Pèrez et al. [21] applied an algorithm which calculates fault signatures and pressure residuals between fault measurements and a model. The network tested consisted of five observed pressure sensor locations, with two inlet reservoirs. After a day of measuring the leaking network, the most probable leak location was found within $500m$. The network was the actual network supplying water to the Nova Icaria DMA.

1.2.6 Work Done with Sensor Placement Optimization

Gong et al. [22] proposed a mobile sensor that travels freely inside the water network. By optimizing the minimum number of sensors and beacons for the tested networks, they could optimize the leak localization problems. They compared two approaches, greedy and a linear,

which they combined to find the best combination of sensor and beacon placements. They tested two networks, a simple EPANET model and three parts of the Micropolis Virtual city.

Farley et al. [23] presented a method of identifying optimum locations for pressure sensors to detects leaks within a DMA. They tested five different leak events by flushing the system ad different locations. Eight locations were used to measure the pressures of the network. They calculated the Chi-Squared value over the observed nodes between leak and normal conditions and used it to populate a Jacobian matrix. They found that four of the eight sensors had no real contribution to the problem.

Soldevila et al. [24] presented a sensor placement method for classifier based leak localization in water distribution networks. They used a Genetic Algorithm(GA) which decided which sensors to be used by a classifier. The classifier they used was based on the k-Nearest Neighbor approach. The network they did their research on was on the Hanoi DMA and compared the results to the Exhaustive Search Algorithm(ES). They found that the GA could find the same results as the ES algorithm using a population size of 20, in less than a third of the time.

1.2.7 Discussion of Methods

From all the work done with the different methods on the pressure-flow deviation method, it was seen that each researcher performed an investigation with a certain technique on a certain problem. No comparison have been completed between the different techniques on the same networks.

1.3 Scope of the Research

In this chapter it was seen that there is an urge to reduce the amount of water lost due to leaks in South Africa's water distribution and supply networks. From the literature study, it could be seen that various methods of leak detection exist. Each of these techniques have their own advantages and disadvantages. The inverse analysis and pressure-flow deviation methods are model based leak detection methods showing promise for future application.

This study investigates the inverse analysis method of leak detection by comparing different approaches to solve the inverse problem. Error calculation methods between measurements is a basic approach to solve the problem. Poulakis et al.[2] derived a method using a Bayesian probabilistic framework to allow error methods to handle uncertainties in measurements and model errors better. Machine learning techniques such as Neural Networks and Support Vector Machines are also commonly used to solve the inverse analysis problem. EPANET was found to be popular in this research field and will therefore be used for all the models of networks.

This study investigates and compares the following techniques:

- A Mean Squared Error optimization problem as an error calculation problem to set a baseline solution for the research.

- The Bayesian Probabilistic Analysis investigated by Z.Poulakis et al. [2] as an improvement on the error calculation method.
- A Support Vector Machine with a simple linear kernel function which will be trained to classify and locate the leak locations.
- A FF-Artificial Neural Network which will be trained to the network to classify and locate the leak locations.

In past research, these techniques were applied to different networks. In this study three different numerical networks are tested using the same strategies to generalize between applications of the inverse analysis methods. This research will solely focus on round holes as leaks. The numerical networks tested are:

- A simple transportation pipeline to introduce the methodology used with the minimum number of observables.
- A simple agricultural site which simulates a water supply network on a smaller scale to test branching pipelines with few observed nodes.
- A water distribution network introduced by Poulakis et al. [2] to test the effectiveness of the algorithms on loops with two cases where:
  - only pressure measurements are used.
  - only flow measurements are used.

An experimental investigation is also completed. To do this a numerical investigation into a network was performed. The chosen network simulate a water distribution network and a water supply or transportation network. The numerical investigation on the experimental network then evaluates the size of the network to allow smaller lengths of pipe, test the effect of a pump on the solutions, test if the leak locations can be calculated accurately on the small scale experiment and find allowable errors between the measurement and model for sensor accuracy identification.

The experimental investigation is completed in accordance with the conclusions found from the numerical experimental investigation. The experimental investigation includes the construction of the experimental network, the testing of the pump to ensure it fits the data used for the model, calibration of the model to actual leak and no-leak measurements, finding single leaks in the experimental and finding multiple leaks in the experimental setup.

Enough sensors are used in the experimental investigation to allow for sensor placement optimization between the experimental networks. The sensor placement optimization is then performed to find patterns for accurate leak detection strategies.

1.4 Chapter Layout

The first chapter contains the introduction to the research. This consists of the background, literature study and scope of the research. Different leak detection methods, pipe modeling software, work done in leak detection and the methods used are investigated.

Chapter 2 explains the algorithms and techniques used to solve the different leaking networks. Three different numerical networks are tested, starting with a simple network and concluding with a simulation of a water distribution network.

The third chapter tests a simulated model of the experimental network to estimate the allowed errors within the measurements to successfully find leaks in the experimental investigation, leading to the experimental investigation in chapter 4. The experimental investigation contains the design and building of the network tests, the calibration of the EPANET model and the results found for the different leak cases. This chapter additionally investigates a method of detecting multiple leaks with a Support Vector machine, which is followed by optimal sensor placement for the experimental network in chapter 5.

This is followed with the findings, conclusion, recommendations and bibliography. Three appendices follow the bibliography that contain results for the second numerical experimental network referred to in chapter 3, figures for the EPANET model calibration to the experimental measurements and the results for additional leak cases from the experimental testing referred to in chapter 4.

Chapter 2

Comparison between Algorithms on Simulated Networks

This chapter contains a comparison between four solution strategies to find leaks in numerical EPANET water distribution models. The algorithms are outlined, followed by the results of all four solution strategies on three simulated leaking water distribution networks. This chapter aims to identify weaknesses in the different algorithms for different network configurations. Initial investigation allows us to compare the performance between algorithms for a simplified scenario in which there is no model errors between the observed measurements.

2.1 The Algorithms

Four algorithms considered are, a Mean Squared Error (MSE) optimization, a Bayesian Probabilistic Framework optimization, the use of a Support Vector Machine (SVM) and an Artificial Neural Networks (ANN).

2.1.1 Strategy 1: Mean Squared Error Minimization

The mean square error between the observed and estimated pressures and flows are computed separately. this allows for independent weighting of associated errors. The pressure MSE is calculated by:

$$MSE_H = \sum_{i=1}^{N} \left(H_i^* - H_i\right)^2, \tag{2.1}$$

where i is the observation node, N is the total number of observation nodes, H_i^* is the observed pressures and H_i is the simulated pressures. The MSE for the flow can be calculated as:

$$MSE_F = c_1 \sum_{j=1}^{L} \left(F_j^* - F_j\right)^2, \tag{2.2}$$

where j is the observed pipe, L is the total number of observed pipes, F_j^* is the observed flows, F_j is the simulated flows and c_1 is a weighting factor. This factor enables the flow to carry the same, more or less weight than the pressures. This factor is set to 10 since to allow the flow rate, which is 10 times smaller than the pressures in most cases, to carry the same weight as the pressure measurements. The final MSE to be optimized is calculated as:

$$\text{MSE} = MSE_H + MSE_F \tag{2.3}$$

The optimization algorithm used is a differential evolution algorithm which is a built in function into Python. It is found within the Scipy package. This algorithm allows upper and lower limit constraints for the optimized parameters.

2.1.2 Strategy 2: Bayesian Probabilistic Analysis

This method used was proposed by Poulakis et al. [2]. It uses the same MSE calculation as the previous algorithm, to which a probability is calculated. An extra optimization parameter is introduced, namely σ, which is the uncertainty in the MSE calculation. The formulation is written as:

$$P(\theta, \sigma|\bar{x}) = c_2 \frac{1}{(\sqrt{2\pi}\sigma)^{LN}} e^{\left(-\frac{MSE(\theta)}{2\sigma^2}\right)}, \tag{2.4}$$

where θ is the optimized leak parameters such as location and size. The optimization of this algorithm requires the equation to be maximized for the largest probability. The optimization algorithm used requires a function to be minimized. Therefore the log-likelihood of this function can be calculated as:

$$g(\theta, \sigma) = -\ln(P(\theta, \sigma|\bar{x}) = \frac{MSE(\theta)}{2\sigma^2} + \frac{LN}{2}\ln(\sigma^2). \tag{2.5}$$

The log-likelihood given by $g(\theta, \sigma)$ can now be minimized. The optimization algorithm used is the differential evolution algorithm found within the Scipy package.

2.1.3 Strategy 3: Support Vector Machine

For the support vector machine, a data set needs to be generated for training. SVMs aim to solve the problem as a classification or inverse regression problem. The input parameters for the SVM are the pressure and flow measurements from a simulated model and the output parameters are the leak location and size. Figure 2.1 shows a model of the input and output parameters for the classification SVM. The output is an integer value depending on the number of pipes in the network, shown as P_n. In this diagram H indicates the input pressures, N is the number of observed nodes, F is the input flows and L is the number of observed pipes. Figure 2.2 shows a diagram of the of the regression SVM used to find the leak size and location. The number of SVMs used equals the number of possible leaking pipes. In this diagram D_n indicates the size of the leak and L_n indicates the location of the leak.

The SVMs are generated using scikit-learn [25] which is a Python package used for machine learning. The kernel functions are left to the default which is a linear kernel function.

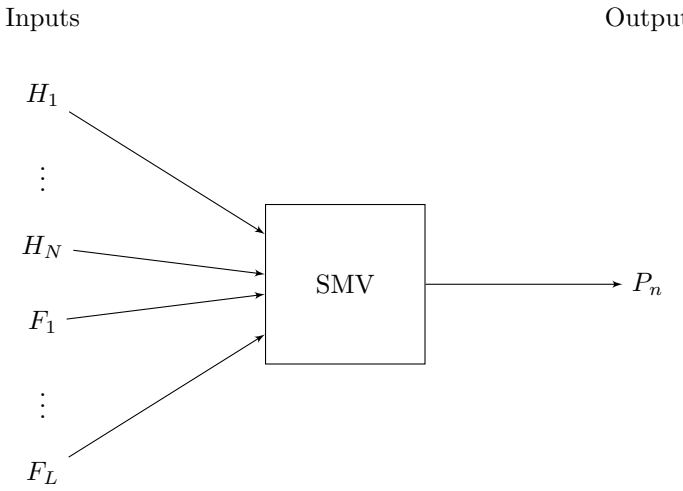


Figure 2.1: Diagram of Classification SVM

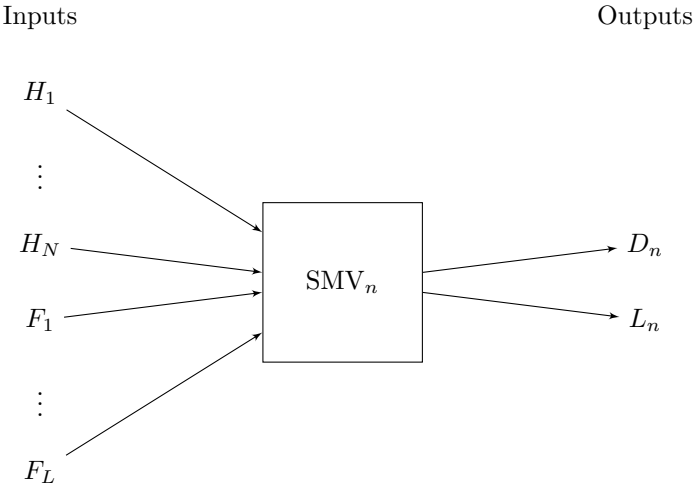


Figure 2.2: Diagram of Regression SVM for Diameter

2.1.4 Strategy 4: Artificial Neural Network

The Artificial Neural Networks used use the same input and output parameters as the SVMs. Figure 2.3 shows a diagram of the ANN to be used for classification.

In the diagram the circular nodes indicate the input layer, having the same formulation as shown for the SVM. The hidden layers are shown with square boxes and the number of hidden layers are indicated with X. A total of five hidden layers were chosen. The values of $a, b, ..x$ indicate the number of nodes in each hidden layer. These values were selected to be 50, since this allows for enough flexibility without deteriorating the training. The output layer is shown with a diagonal square box. The value of n varies depending on the problem. This value equals the number of pipes with possible leaks in the network. Each arrow in the diagram represents a weight which will be optimized during training. A bias is added to the network which is not shown in the diagram. All the hidden and output layers have a sigmoid activation function.

Each pipe in the network has its own regression ANN which calculates the leak location

and size. Figure 2.4 shows the diagram of this ANN. In this diagram D_n indicates the size of the leak for a certain pipe while L_n indicates the location of the leak for a certain pipe. The number of hidden layers and nodes in a layer is identical to that of the classification ANN. The output layer uses a linear activation function, with sigmoidal activation functions for the hidden layers.

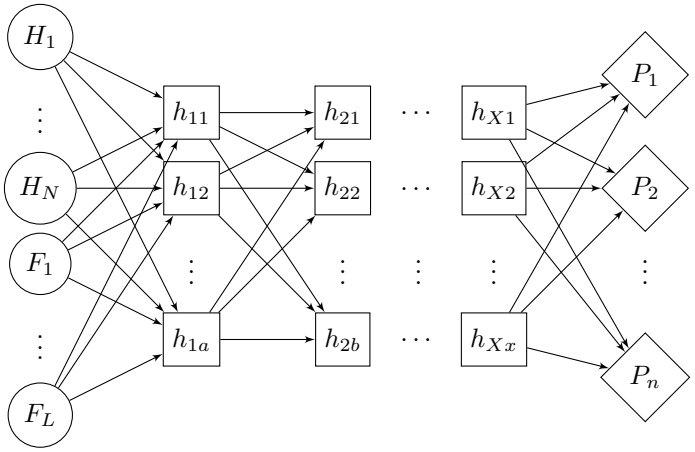


Figure 2.3: Diagram of Classification ANN

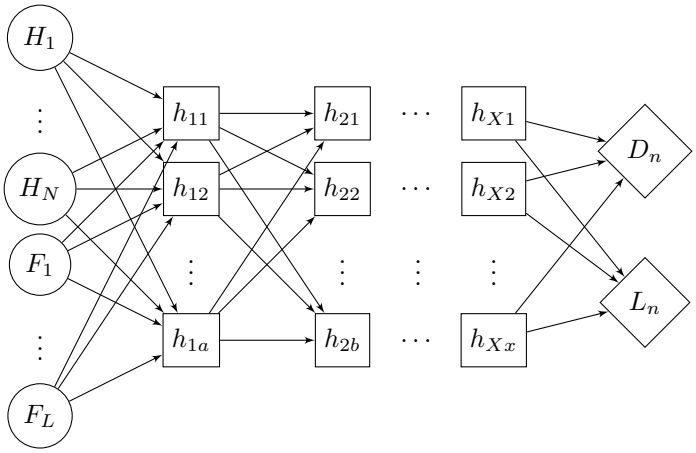


Figure 2.4: Diagram of Regression ANN

In Python the ANNs are created using the PyBrain [26] package. This package allows for easy usage of ANN. The trainer used for both ANNs is the RProp algorithm.

2.2 Investigated Problems

The algorithms were tested on three different water networks. These networks include a simple single pipe network, a practical network of a small agricultural site to simulate a water supply network and a complex network found in the literature to simulate a water distribution network. The leaks will be modeled by empty reservoirs and there will be assumed that all the leaks are round.

2.2.1 Problem 1: Simple Single Pipe System

This network consists of two reservoirs connected to each other. Water flows from the one reservoir to the other. Figure 2.5 shows a diagram of the network layout. R_1 and R_2 indicate the two reservoirs, N_1 and N_2 indicate the two nodes at which pressure can be measured and P_1 indicates the pipe with a leak on it. The flow in this pipe can also be measured.

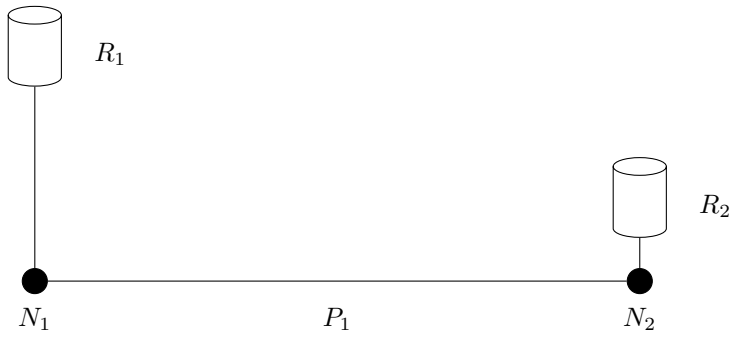


Figure 2.5: Diagram of of the Simple pipe network

The height of R_1 and R_2 is chosen to be $50m$ and $20m$ respectively. The elevation of N_1 and N_2 is chosen to be the same, at ground level. The length of P_1 is chosen as $100m$ with a diameter of $32mm$. The location to the leak is measured from N_1. The leak is placed in the center of the pipe with a size of $3mm$.

2.2.1.1 Mean Squared Error Optimization

The solution found for this network by minimizing the MSE using differential evolution are presented in Figure 2.6.

It is clear that for any location of the leak along the pipe, a leak size can be found which results in an MSE of zero. This indicates there are an infinite number of solutions for this problem. The dot on the diameter graph indicates the actual leak location.

2.2.1.2 Bayesian Probabilistic Analysis

Similarly the solution found for this network by the Bayesian Probabilistic Analysis can be seen in Figure 2.7.

A similar solution can be seen except for a leak probability of 100% and an uncertainty in the solution of zero for every location in the pipe. The dot on the diameter graph indicates the actual leak location.

2.2.1.3 Support Vector Machine

To solve this problem with a SVM, a data set of 10000 samples was generated. Random locations and diameters between $0 - 100m$ and $0 - 5mm$ were generated. Pressure and flow measurements could be generated by simulating the EPANET model with the randomly generated leak parameters. The input data for the SVM are the pressures, flows and the leak location. The leak location was added to the inputs since the two dimensional problem can be reduced to one to allow easy visualization of the results. The output data for the SVM is simply the diameter of the leak.

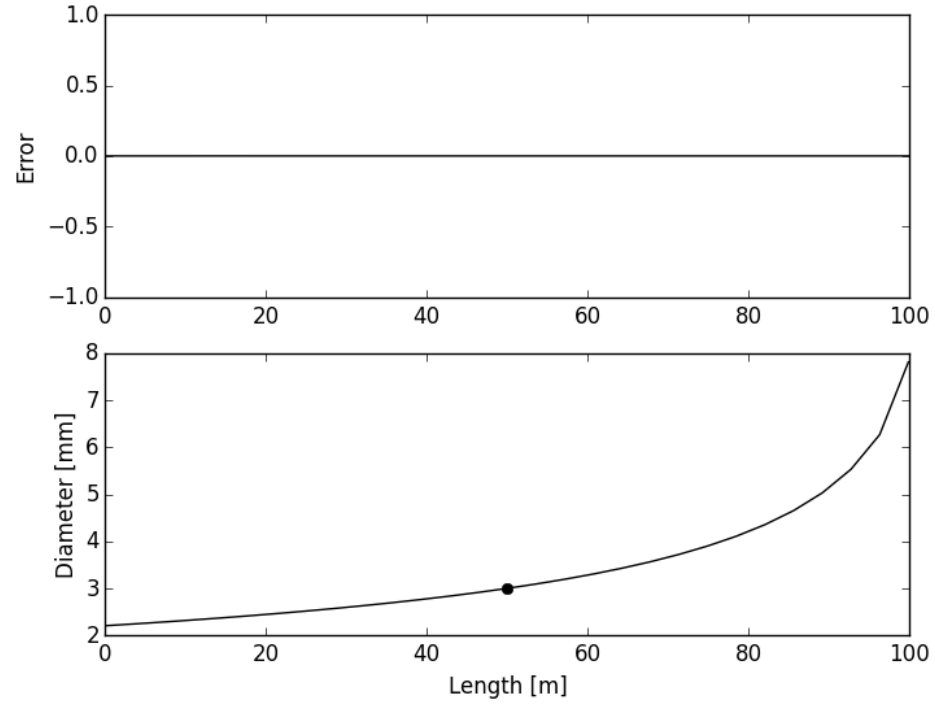


Figure 2.6: Problem 1: Mean squared error minimization result. The dot indicates the actual leak size and location.

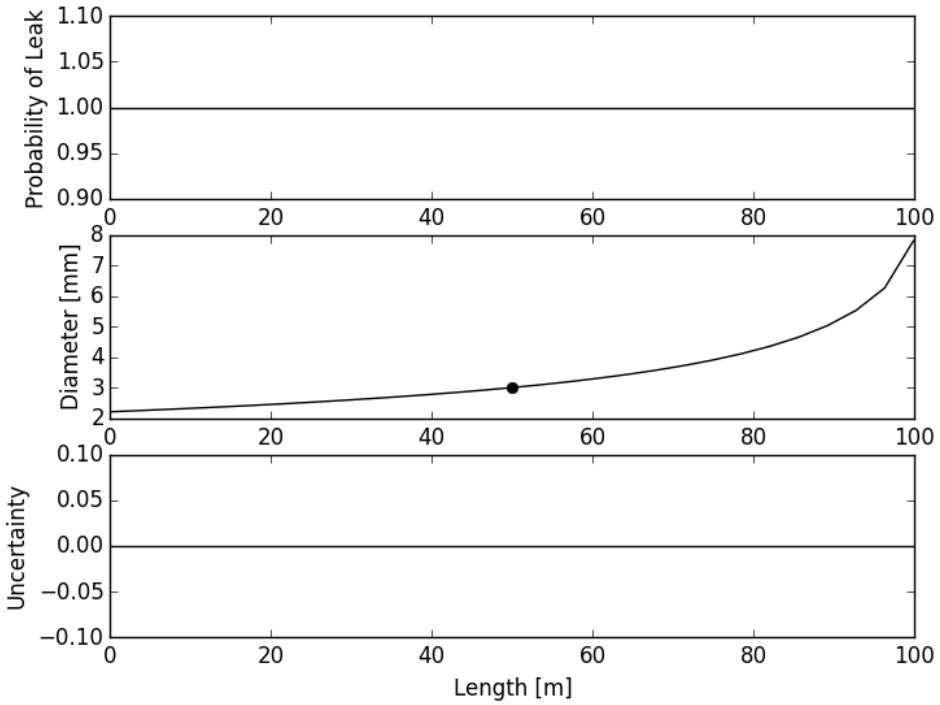


Figure 2.7: Problem 1: Bayesian Probabilistic Analysis result. The dot indicates the actual leak size and location.

Training the SVM and estimating the leak size over the pipe lengths results in Figure 2.8. The prediction error can also be seen. This error is calculated using the MSE function with the predicted diameters. This error is calculated since the SVM can not find the probability of the leak diameter at the different locations. The minimum error location can be chosen as the most probable leak location.

From the solution it can be seen that the SVM gives a unique solution for this problem which is when the error is calculated to be 0.003 at its minimum. The dot on the graph indicates the actual leak location, with a length of $49.9m$ and a diameter of $2.99mm$.

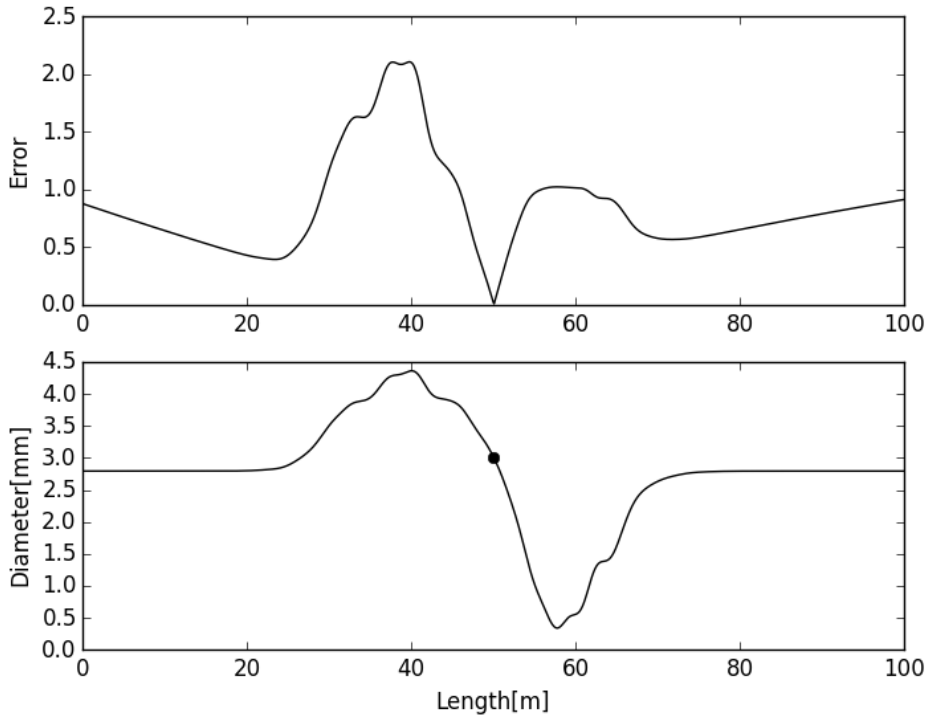


Figure 2.8: Problem 1: Support Vector Machine result. The dot indicates the actual leak size and location.

2.2.1.4 Neural Network

To solve this problem with an ANN, similar data sets were generated as used by the SVM. The same inputs and outputs were selected. After training the ANN, leak diameter estimations could be found for different leak locations along the length of the pipe. The MSE error was minimized as a function of the leak location. Figure 2.9 shows the solution found by the ANN.

From this figure it can be seen that a unique solution was found for this system. The error for the solution was calculated using the MSE function. It can also be seen that the ANNs do not have any boundary condition restrictions, resulting in the diameter being calculated as negative after the length reaches about $55m$.

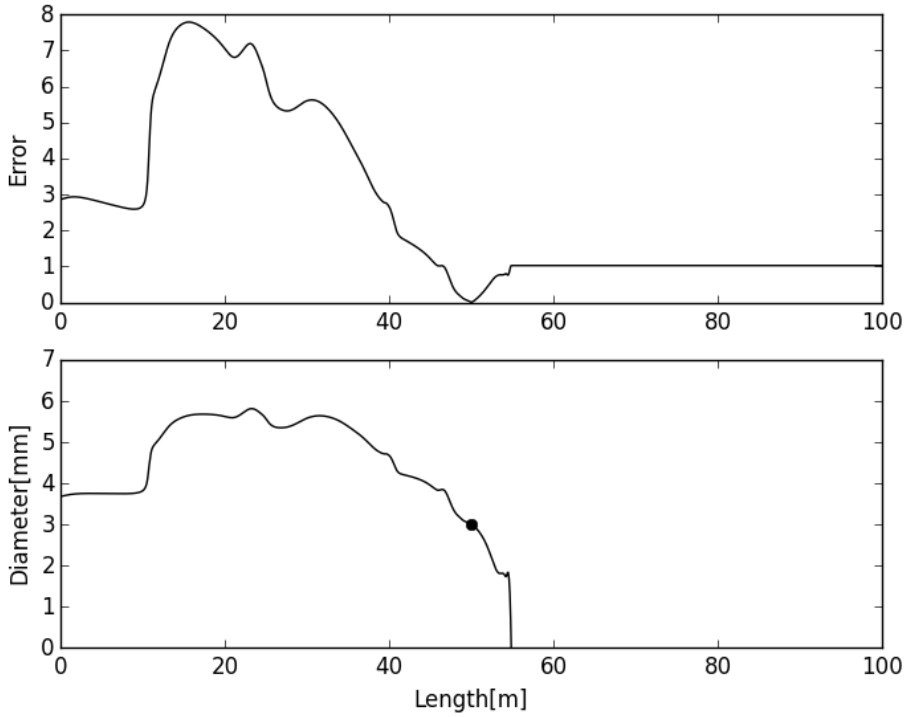


Figure 2.9: Problem 1: Artificial Neural Network result. The dot indicates the actual leak size and location.

The minimum error of the ANN solution is found to be 0.007. This results in a leak location of 50.1m and a leak size of 2.99mm.

2.2.1.5 Additional Leak Cases

Additional leak cases were tested to ensure the algorithms solve the problem correctly. Table 2.1 shows the results found for three additional leak cases.

Table 2.1: Results for Additional Leak Cases

Algorithm	Location	Size	Location	Size	Location	Size
Actual Location	20	2	60	2	80	1
Mean Squared Error	-	-	-	-	-	-
Bayesian Probabilistic Analysis	-	-	-	-	-	-
Support Vector Machine	19.8	1.99	59.1	2.09	80.9	1.01
Artificial Neural Network	18.6	1.98	59.5	1.96	79.9	0.99

In this table the three leak cases are at 20m, 60m and 80m respectively with sizes of 2mm and 1mm. It was found that the MSE and the Bayesian Probabilistic Analysis found an

infinite number of solutions for all three cases. The SVM and ANN found solutions for each of the leak cases with the largest error in location of $1.4m$ by the ANN and the largest error in size of $0.09mm$ by the SVM.

2.2.1.6 Discussion of Results

From the solutions found by the different algorithms it could be seen that the error minimization approaches namely, Mean Squared Error and Bayesian Probabilistic Analysis algorithms faced an ill-posed problem with an infinite number of solutions. The algorithms could find a possible leak size at every leak location. This is due to the simplicity of the network, not having enough observation nodes and pipes. Another factor affecting the solution is the fact that the algorithms ignore all network parameters, focusing on the observations only.

The SVM and ANN could find accurate solutions for this network since they train parameters to simulate the network in in different states for the given data sets. This allows for the usage of the actual pressure and flow measurements, instead of the difference between them as used with the error minimization approaches.

2.2.2 Problem 2: Realistic Numerical Network

This network is based on a small agricultural network. The locations of nodes and elevations were estimated based on geographical data from the satellite imagery [27]. The network can be seen in Figure 2.10.

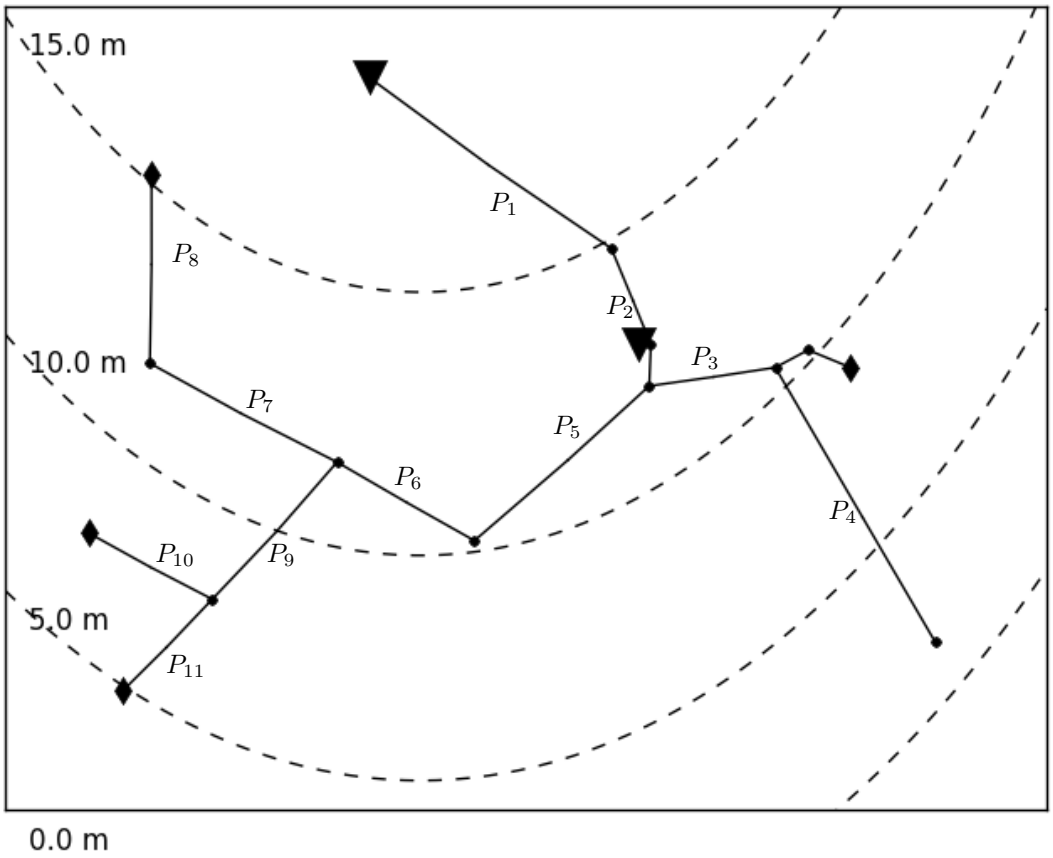


Figure 2.10: Map of the Realistic Numerical Network

This network consists of two reservoirs marked with ▼. Both these reservoirs are at the same height and supply the whole system with $20m$ of head. Nodes in this network are represented with ▪. Nodes with demands on them are marked with ♦. These demands are set to $0.2l/s$.

The observed nodes, at which pressures are measured, are marked as N_3, N_4, N_7, N_8 and N_{12}. The observed pipes, at which flows are measured, are marked as P_4, P_7 and P_9. A total of 8 observed locations are chosen in this network. Leaks can occur on pipes numbered from $P1 - P11$. The diameters of the pipes vary between $40mm$ to $32mm$ depending on the flows in the pipes. Pipes P_1, P_4, P_7, P_8 and P_{12} have diameters of $32mm$, with the rest of the pipes having a diameter of $40mm$. Elevation for each node changes over the system between $0 - 15m$. Table 2.2 shows the connecting pipes for each node as well as the coordinates and elevations for each node.

Two leak cases are considered. The first case is a leak on $P5$, with a size of $3mm$ and the location at $55m$, which is in the center of the pipe. The second leak case is a leak on $P10$, with a size of $4mm$ and a location of $42.2m$, which is $\frac{2}{3}$ into the total pipe length.

Table 2.2: Realistic Numerical Network Node and Pipe Information

Pipe Number	Node 1	Node 2	Node Number	X-coord	Y-Coord	Elevation
P_1	R_1	N_1	N_1	262	325	15
P_2	N_1	N_2	N_2	280	275	13
P_3	N_3	N_4	N_3	278	254	13
P_4	N_4	N_5	N_4	332	264	11
P_5	N_3	N_6	N_5	402	122	0
P_6	N_6	N_7	N_6	203	174	11
P_7	N_7	N_8	N_7	143	215	10
P_8	N_8	N_9	N_8	63	266	12
P_9	N_7	N_{10}	N_9	63	363	15
P_{10}	N_{10}	N_{11}	N_{10}	90	144	8
P_{11}	N_{10}	N_{12}	N_{11}	36	178	8
			N_{12}	51	97	6
			R_1	157	413	20

2.2.2.1 Mean Squared Error Optimization

The solution found by the MSE optimization can be seen in Figure 2.11. The first column of figures refer to case A and the second column to case B. The minimum error solution for Case A corresponds to a leak size of $3mm$ and leak location of $54.97m$.

From the graph showing the leak location it can be seen that, for all the pipes with the exception of $P5$ and $P10$ the locations of the leaks are placed as close to the actual leak as possible. This is indicated with either the leak location as zero or as the maximum pipe length.

The minimum error solution for Case B corresponds to three possible pipes, $P9 - P11$. The leak location for these pipes are within $8.5m$ of each other, with the leak size within $0.09mm$ of each other. The leak location and size estimation corresponding to the lowest error was found to be for $P10$ and are $55m$ and $3.99mm$.

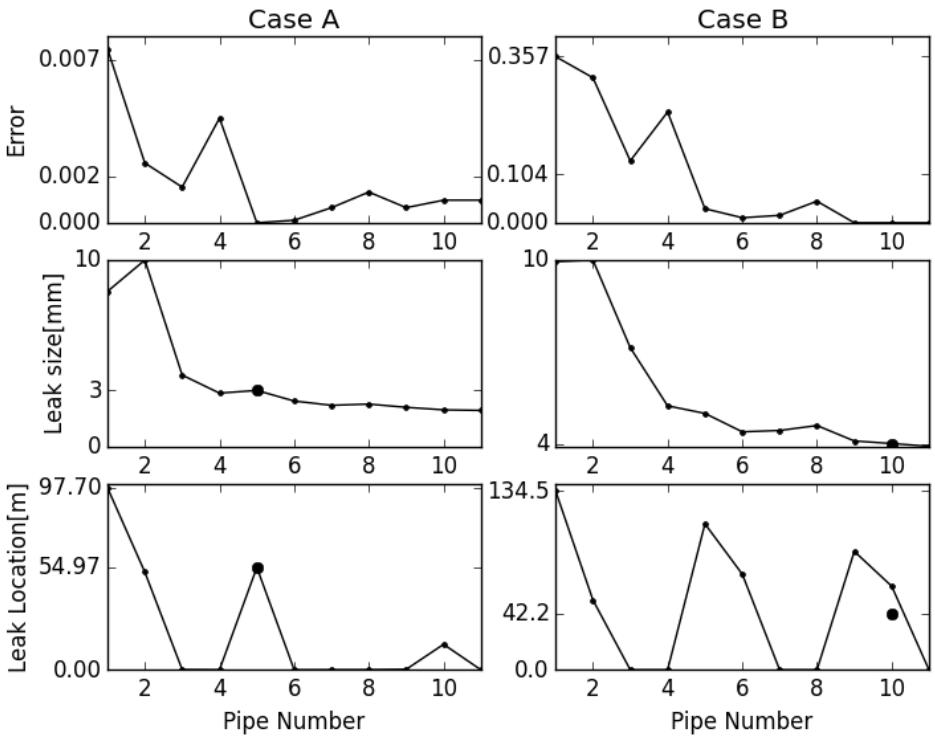


Figure 2.11: Problem 2: Mean squared error minimization result. The dot indicates the actual leak size and location.

2.2.2.2 Bayesian Probabilistic Analysis

The solution for the Bayesian Probabilistic Analysis can be seen in Figure 2.12. The first column shows the results for Case A and the second column for Case B. The calculated probabilities are normalized so that $\sum_{i=1}^{N} P_i(\theta, \sigma|\bar{x}) = 1$.

It is clear that, for case A, similar leak sizes and locations are found compared to the MSE method. A leak probability of 70% is found for $P5$, with a size of $3mm$ and the location at $54.97m$, resulting in an error of $0.03m$ for the location. If two additional closest leak probabilities are considered, the probability of finding the leak within a $50m$ length of pipe is about 78%.

For case B, a leak probability at $P10$ of 31% can be calculated. The size of the leak was found to be $4.0mm$ and the location of the leak at $2m$. The results are similar to the MSE method where the leak locations for all other pipes are placed as close to the actual leak as possible. If two additional closest leak probabilities are taken into consideration, the probability of finding the leak in about $50m$ will be about 91%.

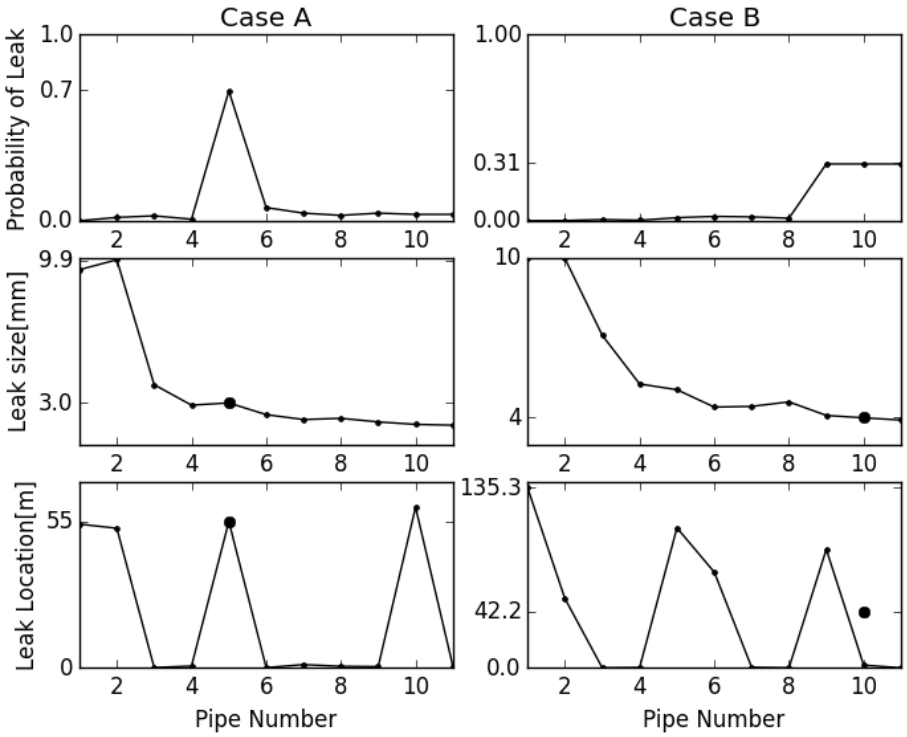


Figure 2.12: Problem 2: Bayesian Probabilistic Analysis result. The dot indicates the actual leak size and location.

2.2.2.3 Support Vector Machine

The solution from the SVM can be seen in Figure 2.13. It can be seen that the SVM generated a probability based on the classification of the leaking pipe and an estimate to the size and location of the leak on the respective pipes.

For case A it can be seen that a 78% probability of a leak at $P5$ was found, with a leak size of $3.69mm$ and a location of $46.6m$. This results in an size error of $0.69mm$ and a location error of $8.4m$. The second best leak probability can be seen as about 13%.

For case B it can be seen that a 70.2% probability for a leak at $P10$ with a leak size of $3.9mm$ and a location of $29.7m$.

Looking at these results shows similarity between the leak size and location for the two cases. To ensure unique solutions are found for different cases, more tests were completed. It was found that the location of the leaks were trained to a fixed solution while the leak size could be estimated correctly for the different cases.

2.2.2.4 Neural Network

The solution from the ANN can be seen in Figure 2.14. It can be seen that the ANN can clearly identify the leaking pipe.

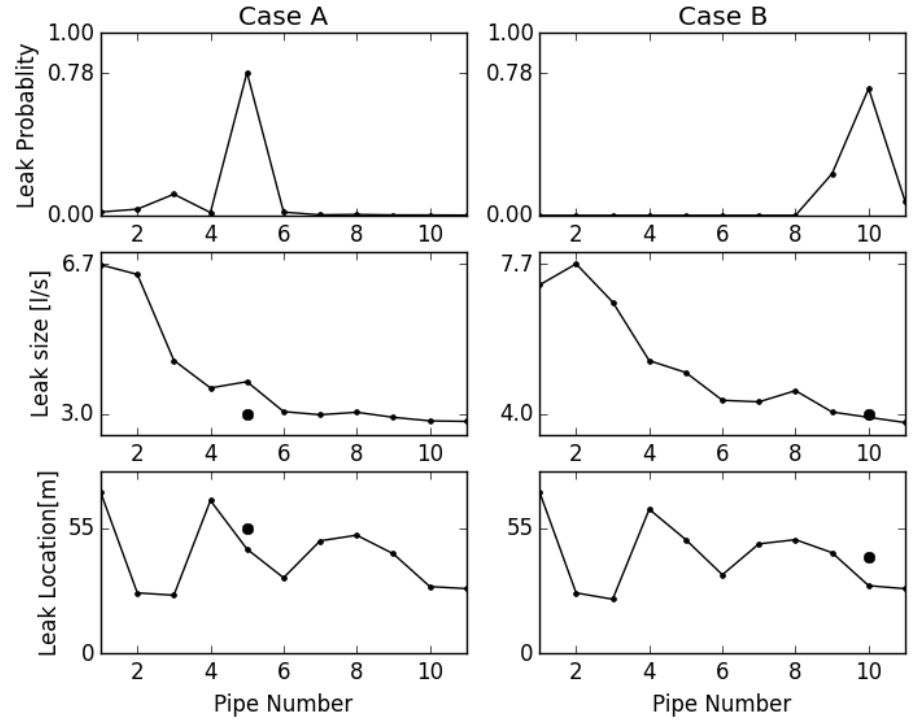


Figure 2.13: Problem 2: Support Vector Machine result. The dot indicates the actual leak size and location.

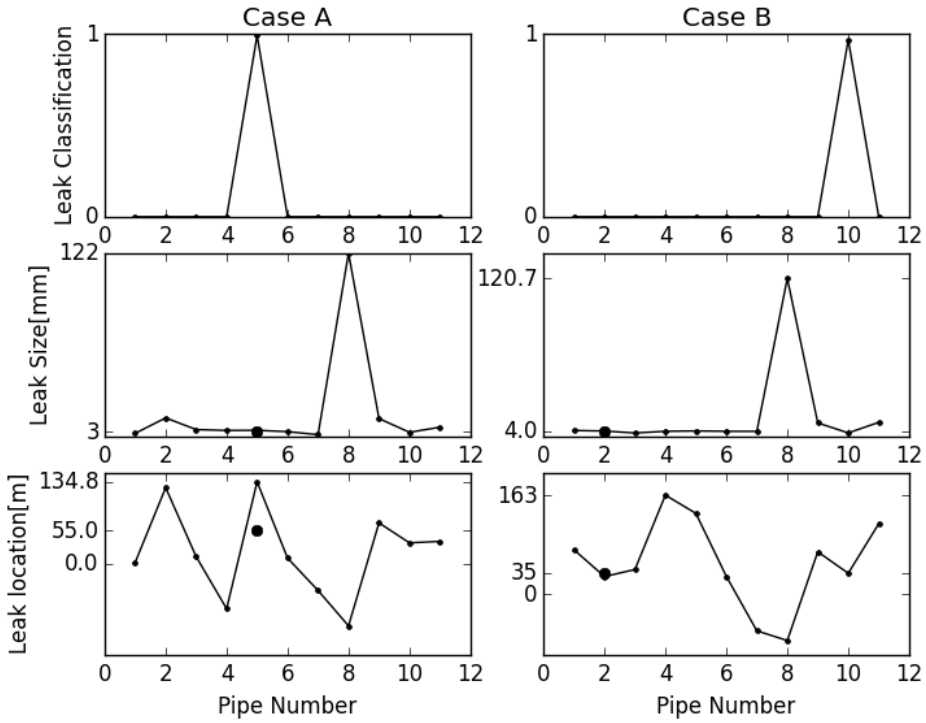


Figure 2.14: Problem 2: Artificial Neural Network result. The dot indicates the actual leak size and location.

For case A an estimated leak size of $3.83mm$ was found and the leak location was estimated at $135m$. This location is overestimated by the ANN since the pipe length is only $109m$, this is due to no boundary condition restrictions.

For case B an estimated leak size of $3.59mm$ was found and the leak location was estimated to be $33.5m$. This results in a location error of about $10m$.

Testing additional leak cases shows that the solution by the ANN generates unique solutions for each case.

2.2.2.5 Discussion of Results

From the four solutions found by the different algorithms it could be seen that all the algorithms could estimate the leak size correctly. The leak location could be estimated within $50m$ with a high probability by the Bayesian Probabilistic Analysis.

The SVM could locate the leaking pipe with a high probability. The leak location on each pipe was found to be a fixed solution at the center of each pipe length.

The ANN could find the leaking pipe with high accuracy. The leak location estimations were found to be non-reliable due to the solutions exceeding the maximum pipe lengths. A summary of the estimation errors can be seen in Table 2.3.

Table 2.3: Summary of Errors in the Estimations for the Realistic Network

	Case A		Case B	
	Size Error	Location Error	Size Error	Location error
Mean Squared Error	$0mm$	$0.03m$	$0.01mm$	$12.8m$
Bayesian Probabilistic Analysis	$0mm$	$0.03m$	$0mm$	$40.2m$
Support Vector Machine	$0.69mm$	$8.4m$	$0.1mm$	$12.5m$
Artificial Neural Network	$0.83mm$	$80m$	$0.41mm$	$8.7m$

2.2.3 Problem 3: Benchmark Network

This network represents a theoretical distribution network. The network consists of 20 loops, 30 nodes with demands, 1 reservoir supplying the network and 50 pipes.

This network was introduced by Poulakis et al. [2] and they applied the Bayesian Probabilistic framework on it. The same network was used by Nasirian et al. [13] to benchmark their new heuristic genetic algorithm model to find leaks. In 2016 the same network was used by Asgari et al. [18] where they investigated a new method of locating a leak by calculating a leak index.

The network is depicted in Figure 2.15. The supply to the network comes from reservoir $R1$ which has a static head of $52m$. Each junction has a demand of $50l/s$. The horizontal and vertical pipe lengths are $1000m$ and $2000m$ respectively. The pipe diameters vary from

$600mm$, to $450mm$ and finally $300mm$ as the flow decreases through the network.

Two cases are considered, in the one case only pressure is measured, while in the other case only flow rate is measured. For case A the pressure observation nodes are $N5$, $N8$, $N12$, $N17$, $N20$, $N24$, and $N29$. For case B the flow rate observation pipes are $P3$, $P17$, $P29$, $P55$, $P61$, $P80$ and $P97$. The red squares indicate the observational node location for case A and the blue squares indicate the observed pipes for case B.

A leak of $22.8l/s$ is added to the network modeled as a demand. The leak is modeled at the center of $P25$, at node $N55$.

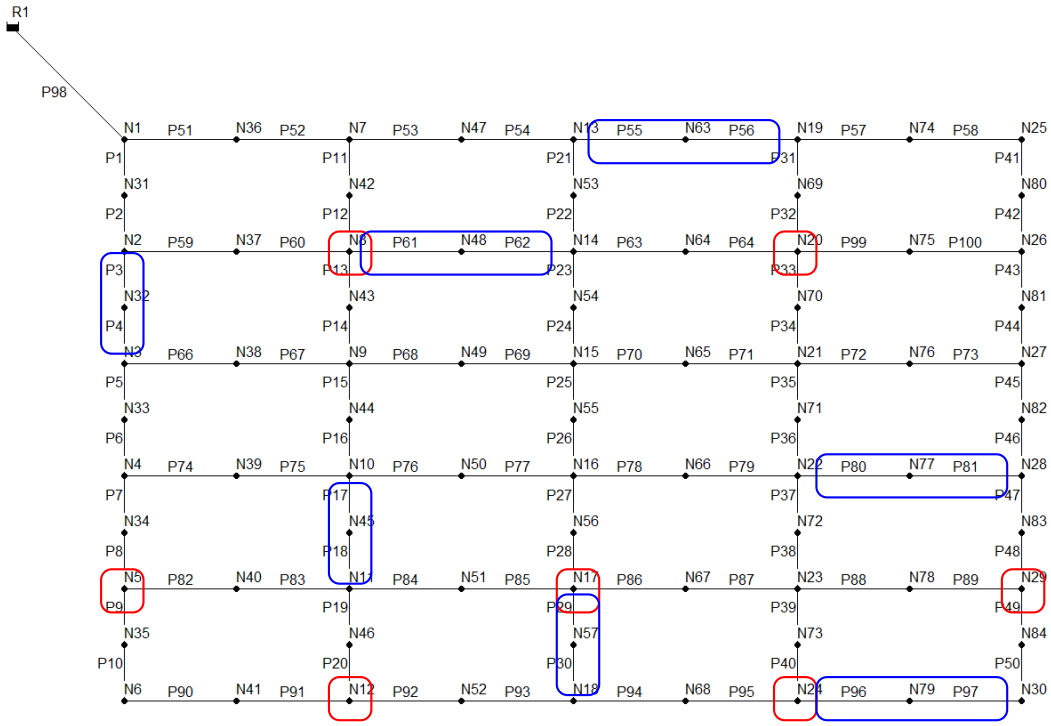


Figure 2.15: Diagram of the Benchmark Network

2.2.3.1 Mean Squared Error Optimization

The solution for this network by the MSE method can be seen in Figure 2.16.

For case A it can be seen that the correct solution was found, although the error ranges over the network between $0 - 1$. The next closest leak locations are $P20$ and $P30$, these locations correspond to a leak at $N50$ and $N60$, which are adjacent nodes.

For case B similar results can be seen, with the error ranging between $0 - 180$. The next most probable leak locations can be seen to be the same as case A. This indicates that the predicted location on pressure and flow rate measurements allow for both the location and leak size to be estimated with a fair amount of confidence.

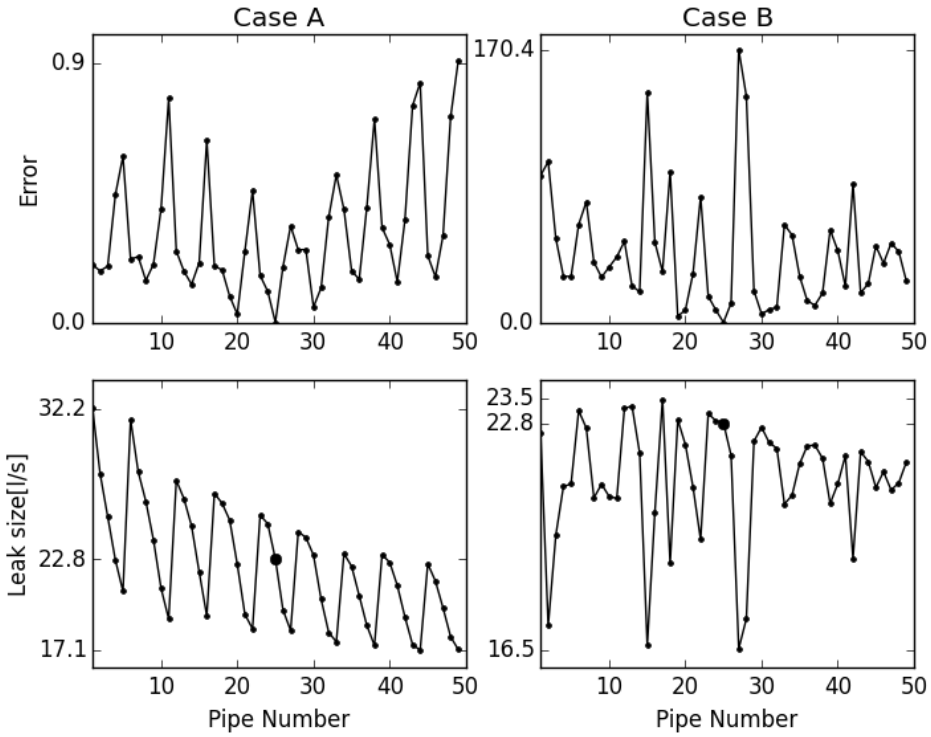


Figure 2.16: Problem 3: Mean squared error optimization result. The dot indicates the actual leak size and location.

2.2.3.2 Bayesian Probabilistic Analysis

The solution found for this network by the Bayesian probabilistic analysis can be seen in Figure 2.17. In this figure it can be seen that for case A, the most probable leak location can be found at $P25$, with a probability of 43.1%. The leak size was estimated to be $22.8l/s$

For case B, the most probable leak location was calculated at pipe $P25$, with a probability of 39%. The leak size was estimated as $22.8l/s$.

2.2.3.3 Support Vector Machine

For the SVM 10000 data samples were generated by selecting random leaking pipes with leak sizes randomly selected between $10l/s$ and $50l/s$. The results after training the SVM to the generated data set can be seen in Figure 2.18.

For case A it can be seen that a probability of a leak at $P25$ was calculated as 56.6% with the adjacent pipes with leak probabilities of 17% and 8%.

For case B it can be seen that the actual leak location was given a probability of 1.8%, with other pipes having higher probabilities of leaks. The highest probability of a leak was calculated as 3% at the first pipe. The leak size estimation of the SVM was calculated between $29l/s$ and $31l/s$, while the actual leak size was $22.8l/s$.

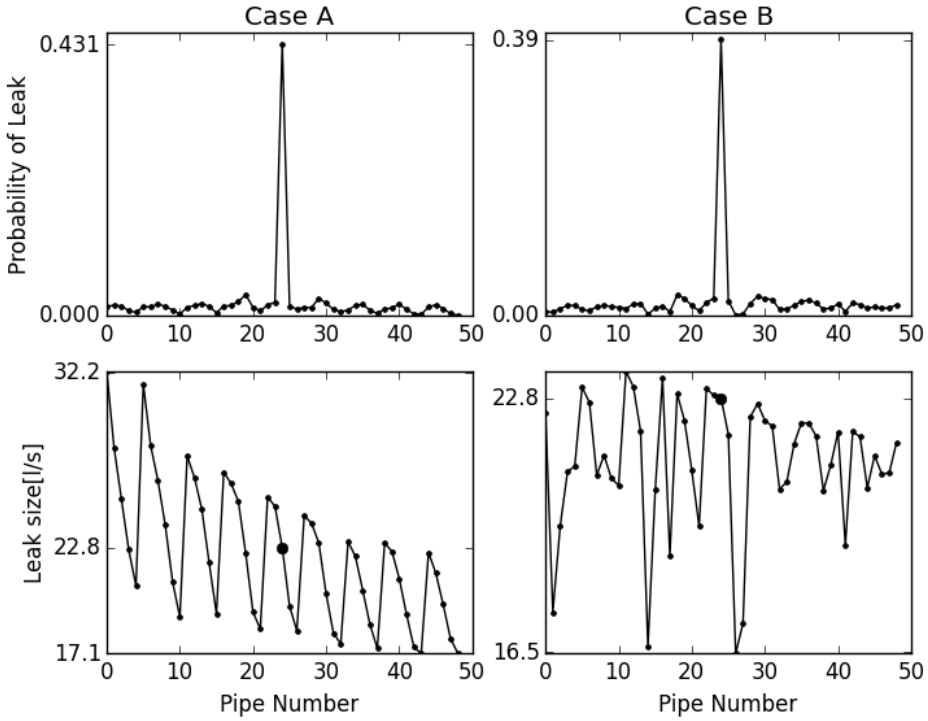


Figure 2.17: Problem 3: Bayesian Probabilistic Analysis result. The dot indicates the actual leak size and location.

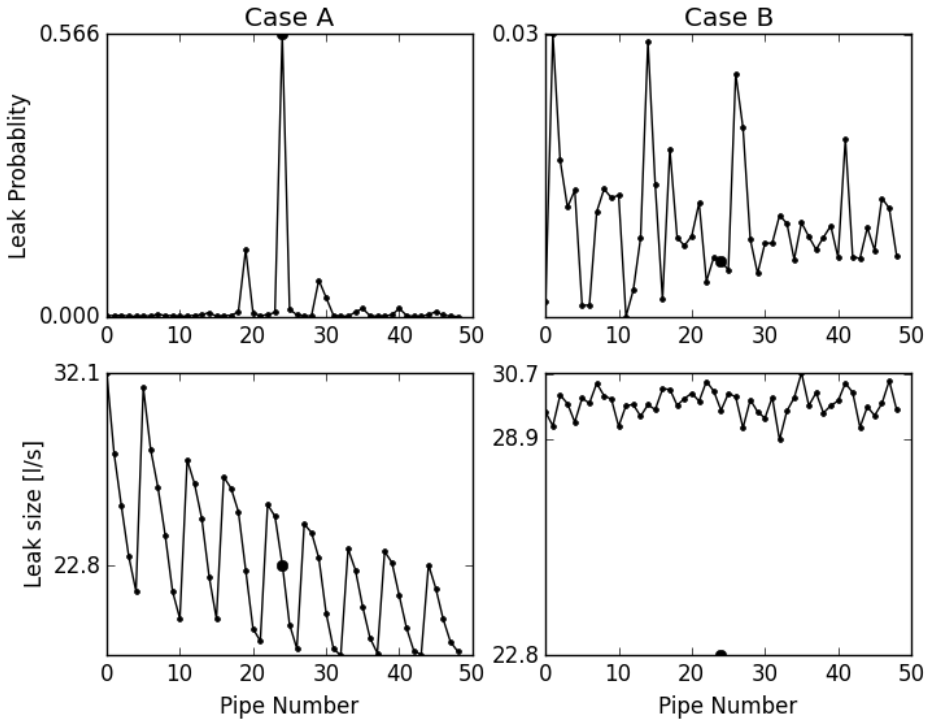


Figure 2.18: Problem 3: Support Vector Machine result. The dot indicates the actual leak size and location.

2.2.3.4 Neural Network

Similar to the SVM, data were generated for different leak cases. The leaking cases were randomly selected with leak sizes between $10l/s$ and $50l/s$. The results found by the ANN after training can be seen in Figure 2.19.
It can be seen that for case A, a perfect classification solution was found as well as a leak estimation of $22.8l/s$. For case B it can be seen that an incorrect classification solution was found. The leak size estimation for this case was found to be $22.7/s$.

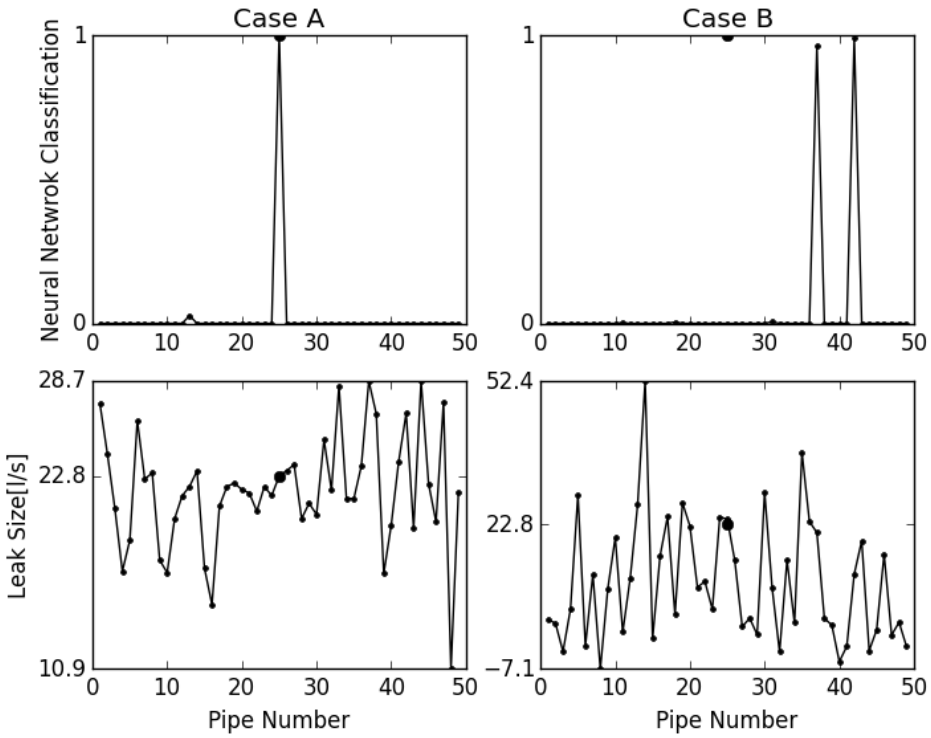


Figure 2.19: Problem 3: Artificial Neural Network result. The dot indicates the actual leak size and location.

2.2.3.5 Discussion of Results

For case A, it was found that all four algorithms could find the correct solution. The Bayesian Probabilistic Analysis and SVM could find the correct solution with a probability between 43% and 57%. The leak size estimation for all these algorithms were found to be approximately $22.8l/s$.
For case B, only the MSE and the Bayesian Probabilistic Analysis could find the correct leak location and size. The SVM could not find the correct leak location or the correct leak size. The ANN was able to find the correct leak size, but with the incorrect leak location. Table 2.4 shows a summary of the results found for the benchmark network by the different solution strategies.

Table 2.4: Summary of Errors in the Estimations for the Benchmark Network

	Case A		Case B	
	Size Error	Probability	Size Error	Probability
Mean Squared Error	$0l/s$	-	$0l/s$	-
Bayesian Probabilistic Analysis	$0l/s$	43.1%	$0l/s$	0.39%
Support Vector Machine	$0l/s$	56.6%	$7.2l/s$	1.8%
Artificial Neural Network	$0l/s$	100%	$0l/s$	0%

2.3 Conclusion

In this chapter, four different algorithms were designed to find leaks in networks. These algorithms were tested on three different networks. The first network was a simple single pipe network which simulates a water transportation pipe. It was found that the first two algorithms, the MSE and the Bayesian Probabilistic Analysis [2] algorithms found a solution for every possible location in the pipe. The SVM and ANN found unique solutions for the system for any leak location. It was found that the ANN had boundary condition restriction problems, resulting in negative size estimations at certain locations.

The second investigated network was based on a simple agricultural site which simulate a water supply network. Two leak cases were added to the problem. For the MSE and the Bayesian Probabilistic analysis algorithms it was found that the solution could easily be estimated within $50m$, although the leaking pipe was not identified accurately. The SVM and ANN could easily identify the leaking pipe, with accurate leak sizes. The location estimations on these pipe were inaccurate, with the SVM training to the average of the data sets and the ANN exceeding the maximum pipe lengths.

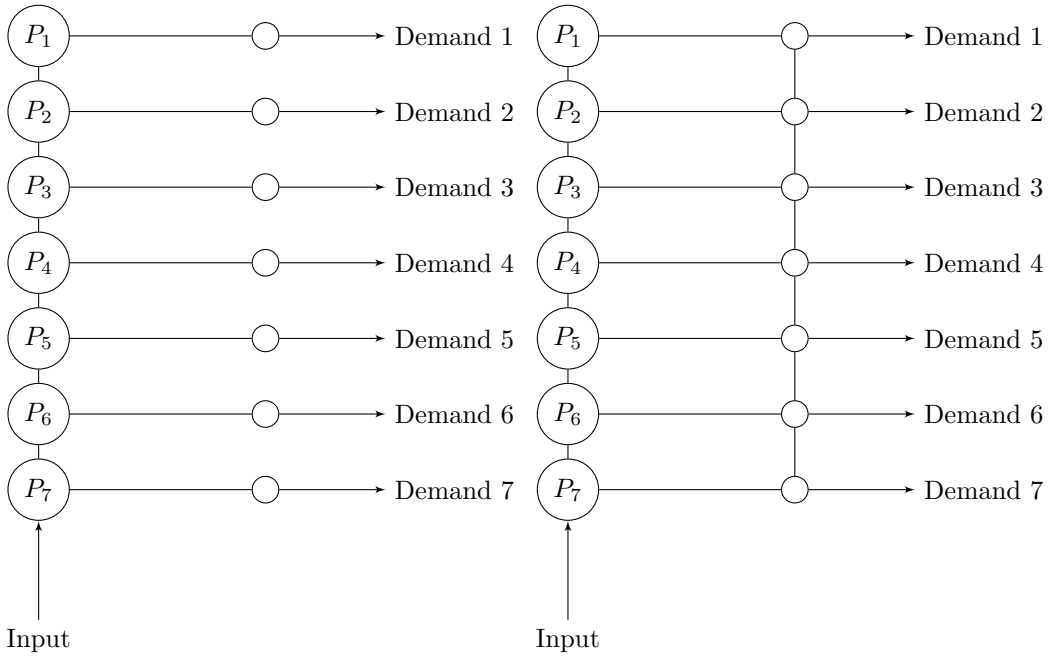
The third problem was based on a theoretical water distribution network. The MSE method could correctly identify the leak location and size. It could be seen that adjacent leak locations could easily be mistaken for the actual leak. The Bayesian Probabilistic Analysis algorithm could easily find the leak location and size in both cases. The SVM and ANN could easily find the leak location for the first case in the problem, where only pressure measurements were used. For the second case, where flow measurements were used, the SVM could not find the leak location or the leak size. The ANN could find the leak size for both cases.

From this investigation it can be concluded that the MSE and Bayesian Probabilistic Analysis algorithms have difficulties solving leaks with few observable nodes. It can also be seen that the SVM and the ANN train to the average of the given data set if the penalty of the error function is small.

Chapter 3

Numerical Investigation for the Design of an Experimental Network

Before we proceed to the experimental study for two numerical experimental network configurations depicted in Figure 3.1, a numerical investigation of these networks is first conducted. The numerical investigation allows us to quantify the expected limiting factors for the experimental networks. In this figure the nodes marked P_1 to P_7 indicate pressure measurement locations, with a single flow measurement at the input to the network.



(a) First Theoretical Experimental Network (b) Second Theoretical Experimental Network

Figure 3.1: Diagrams of the Two Theoretical Experimental Networks

Both the networks will be fed by a pump, with leaks modeled by empty reservoirs. The flow will be restricted by changing the connecting pipe diameters. The outflow diameters

to the empty reservoirs were selected to be $3mm$. All other pipes in this network have a diameter of $10mm$. The seven main pipes are spaced $30cm$ apart, while each pipe is $3m$ long. Seven leaking pipes are added to both networks.

The leak locations are set to be at the third pipe from the top at a location of $2m$ measured from the observation nodes and a leak size of $3mm$. The second leak case is at the sixth pipe at a location of $1m$ and a size of $3mm$.

The pump used is based on the data of a SALFLO SAM 100 ECO pump [28]. Reservoirs are added to the center of each main pipe which simulate possible leak locations. The diameters of the connector pipes, connecting the reservoirs to the main pipes, is set to either the size of the leak or zero to model no leak depending on the leak scenario to be modeled.

3.1 Numerical Network 1

Figure 3.2 shows a diagram of this network. It can be seen that the demands for each pipe acts individually based on the pressure in the pipe. In this figure the pump curve used can also be seen.

3.1.1 Mean Squared Error Optimization

Figure 3.3 shows the results found by the MSE method for the first numerical experimental network. It can be seen that for both leak cases the leak was correctly identified on the actual leaking pipe with an error of zero. For case A the size of the leak was found within $0.06mm$ and the location within $1.5m$ of the known solution. For case B the leak size was found within $0.05mm$ and the location within $1.2m$ of the known solution.

3.1.2 Bayesian Probabilistic Analysis

Figure 3.4 shows the results found by the Bayesian Probabilistic Analysis. It can be seen that, for case A a leak probability of 80% and for case B a leak probability of 90% was found. In both leak cases the leak size was estimated within $0.08mm$ from the known solution. The leak locations were found to be within $1m$ for case A and $1.2m$ for case B from the known solution.

3.1.3 Support Vector Machine

Figure 3.5 shows the results found by the Support Vector Machine. It can be seen that for case A the correct main pipe could be identified with a probability of 90%, while for case B the leaking main pipe could be found with a 100% probability. The leak size for case A was estimated with an error of $0.05mm$ to the actual leak size, while for case B the leak size was estimated with an error of $0.01mm$. The leak location was found to be within $0.4m$ of the actual leak location for case A and $0.5m$ for case B. It can also be seen that the SVM trained the location of the leak to the average of the given data set at about $1.5m$. This makes the leak location estimation inaccurate for the SVM.

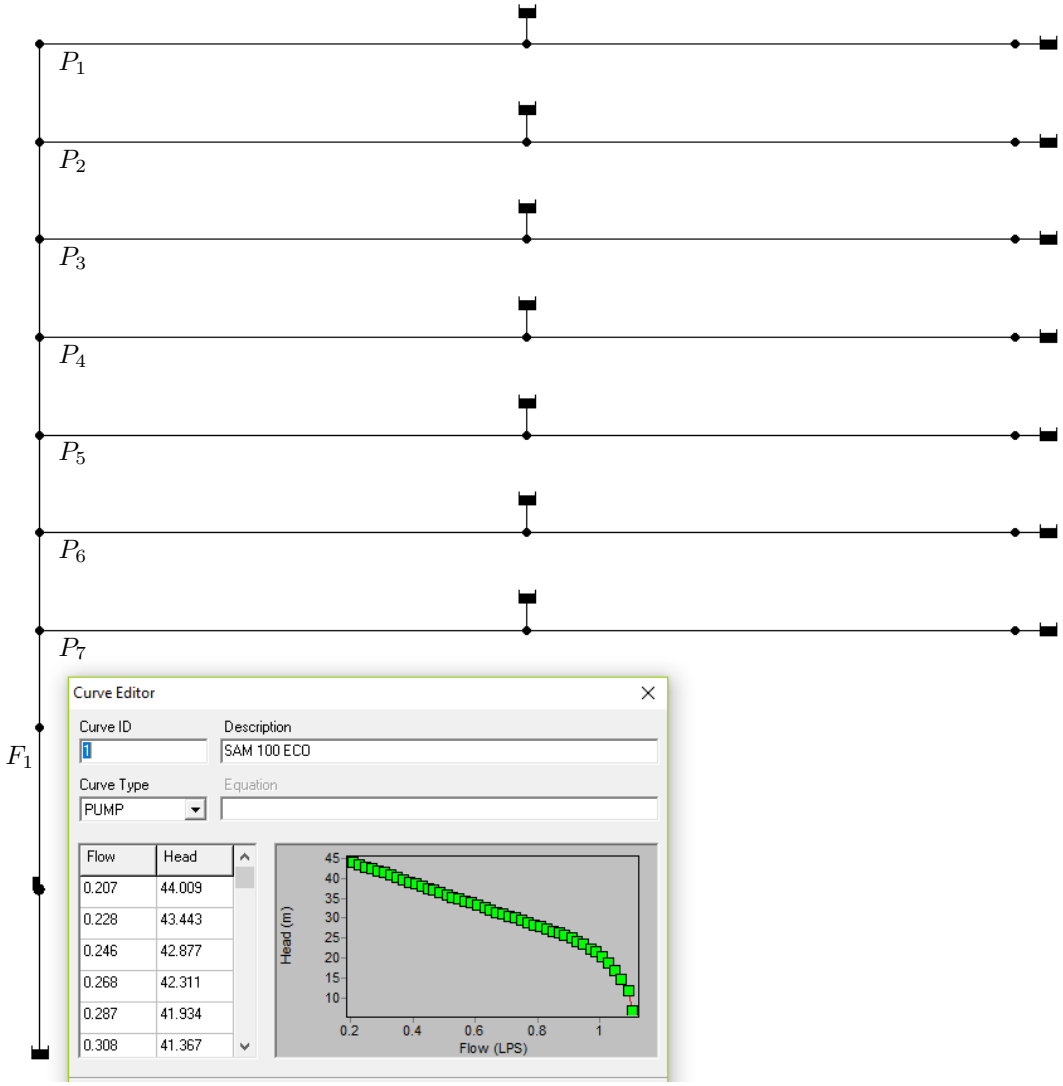


Figure 3.2: Network diagram for the Numerical Network 1

3.1.4 Neural Network

Figure 3.6 shows the results found by the Artificial Neural Network. It can be seen that the leaking pipe was correctly classified for both cases, while the leak size was estimated within $0.4mm$ for case A and $0.05mm$ for case B. The leak location can be seen to be averaged over the length of the pipe. This indicates that the Neural Network trained to the average of the data set for the leak location.

3.2 Numerical Network 2

The second numerical experimental network can be seen in Figure 3.7. It can be seen that all the pipes are now connected together to create loops between them. The added connecting pipes also have a diameter of $10mm$. The new network configuration allows for more flow through the seventh pipe, with less flow through the first pipe. All the previously defined parameters for this network stays the same. The changes to this network should increase the difficulty to find the correct solution.

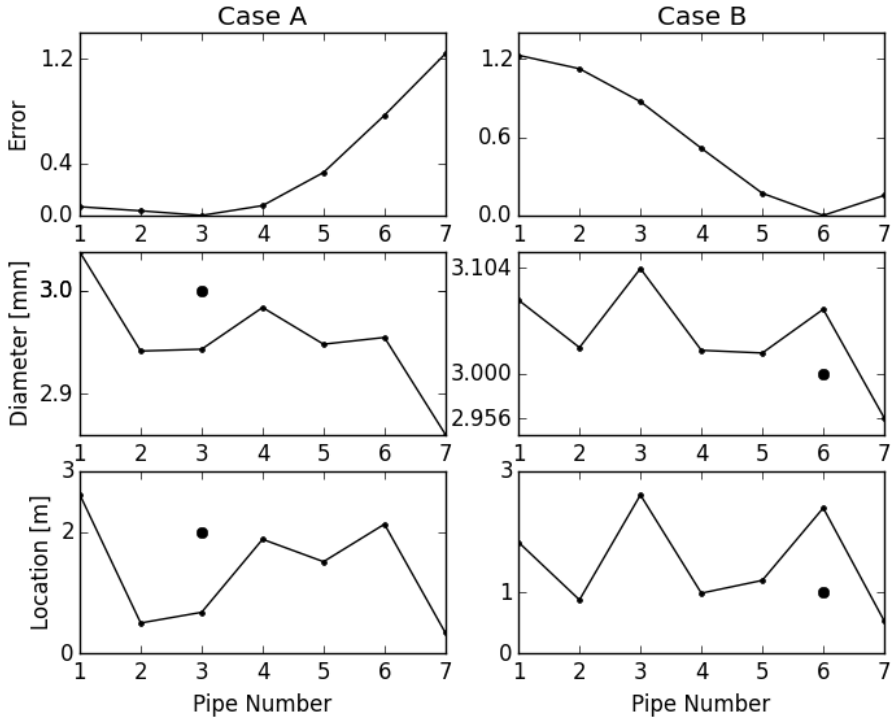


Figure 3.3: Numerical Network 1: Mean Squared Error Solution. The black dots indicate the known solution.

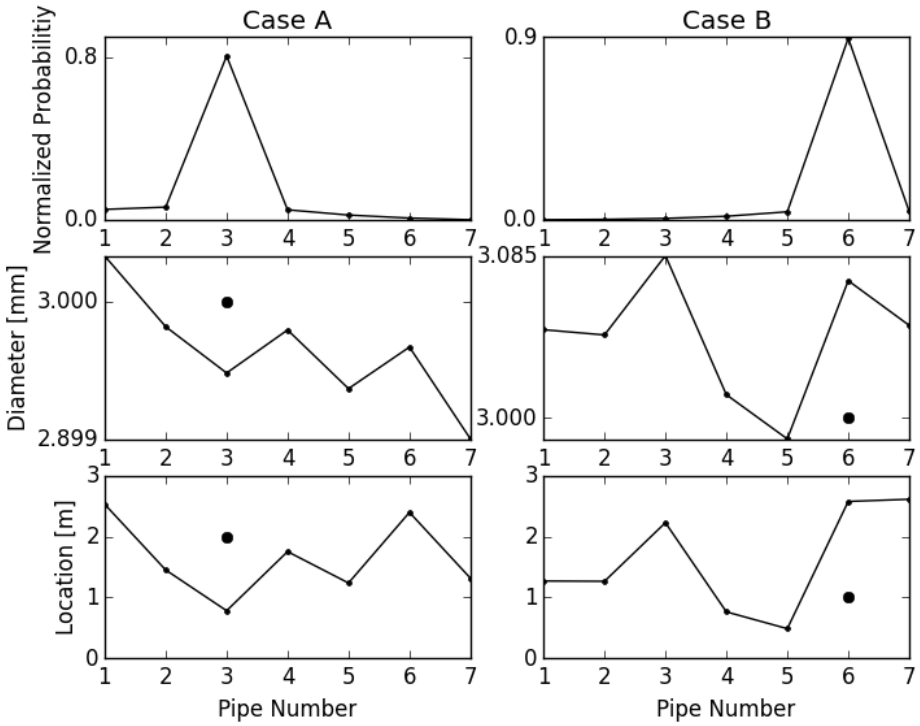


Figure 3.4: Numerical Network 1: Bayesian Probabilistic Analysis. The black dots indicate the known solution.

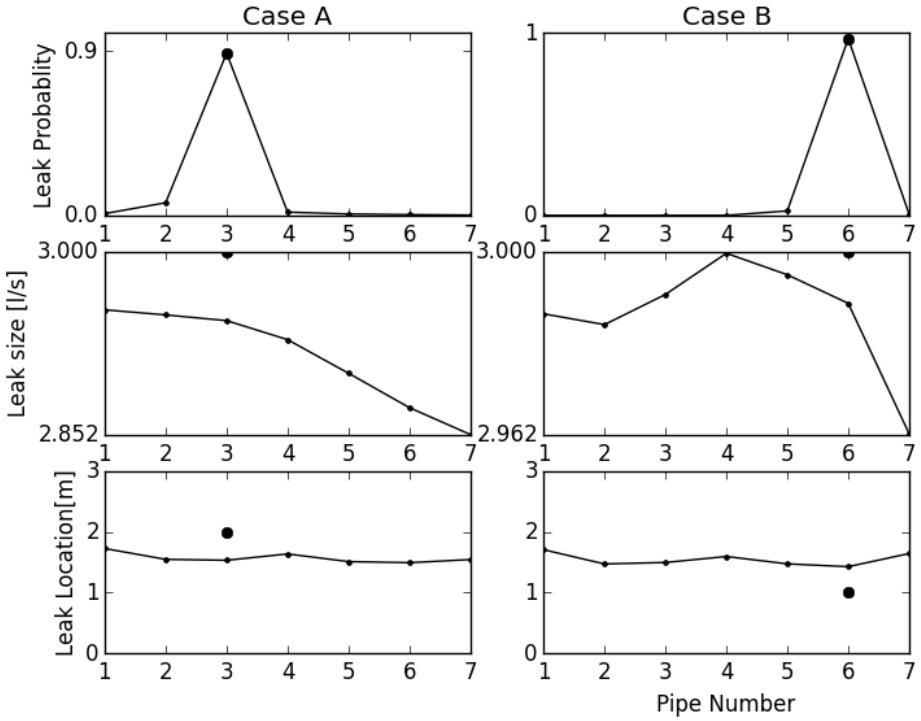


Figure 3.5: Numerical Network 1: Support Vector Machine. The black dots indicate the known solution.

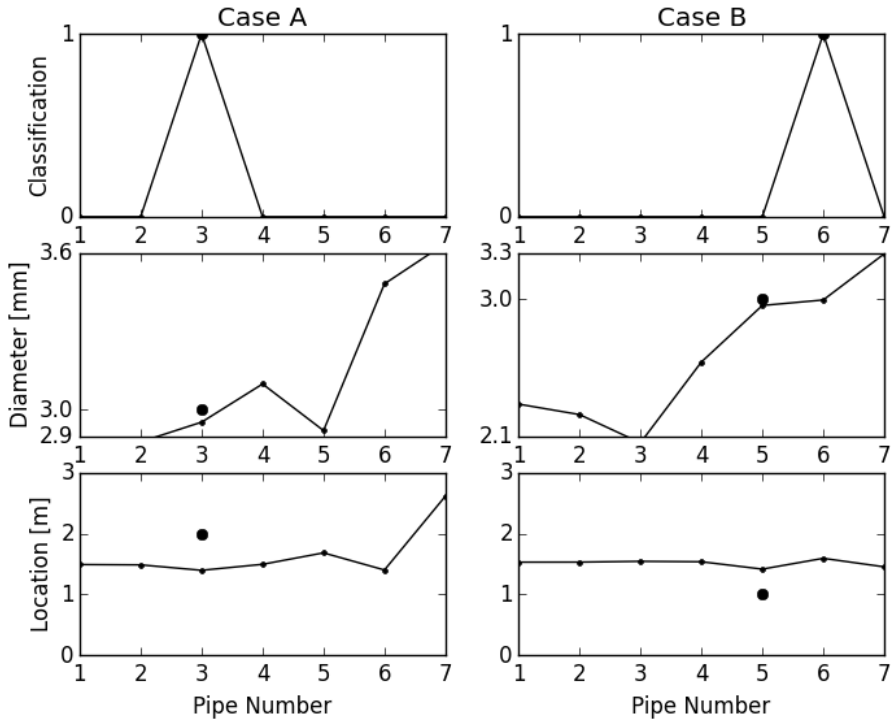


Figure 3.6: Numerical Network 2: Neural Network. The black dots indicate the known solution.

The solutions for this network can be seen in Figure A.1 to A.4 shown in Appendix

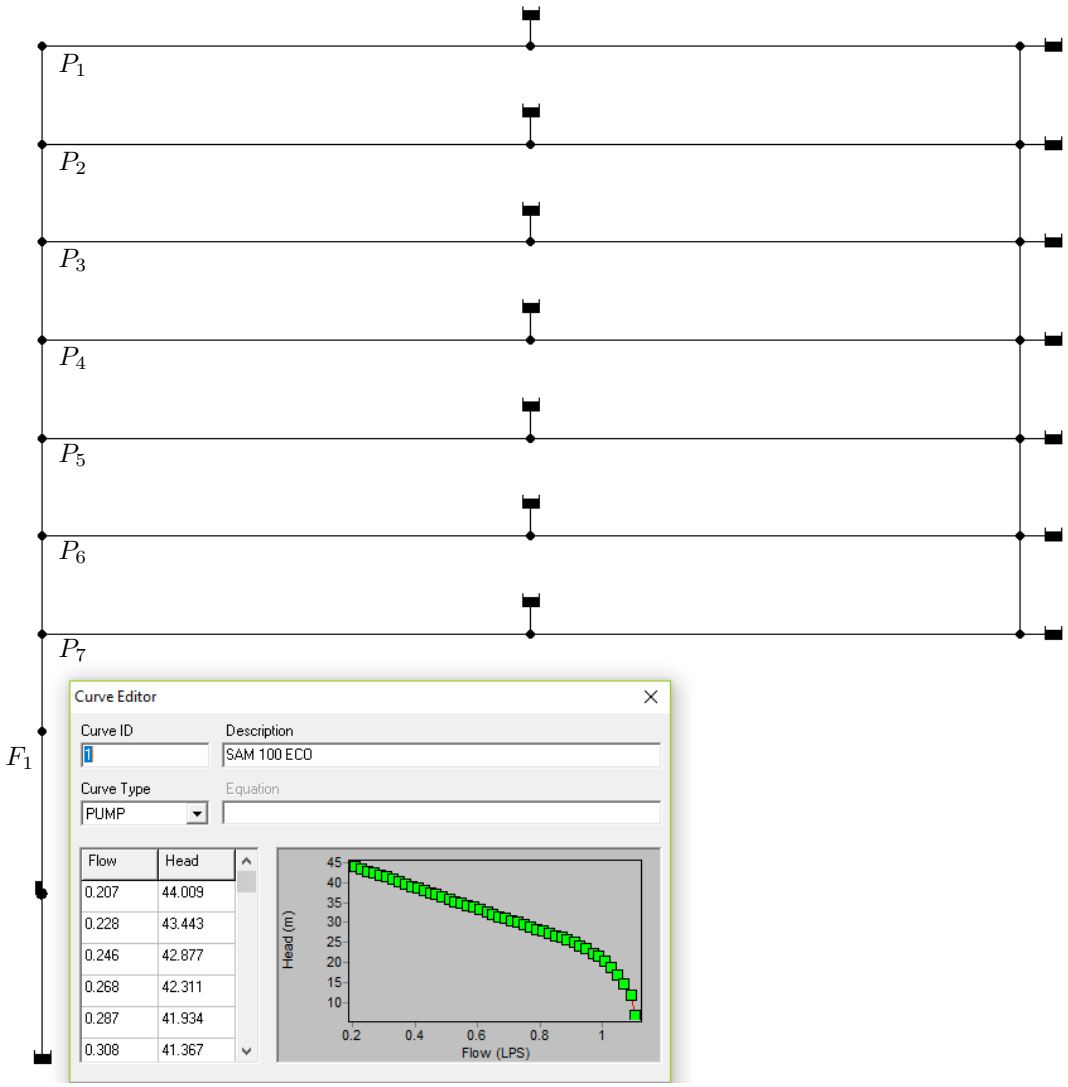


Figure 3.7: Network diagram for the Numerical Network 2

A. From the results it can be seen that for the MSE method the error curves stay the same compared to the first numerical network, with changes in the estimated leak size and location. The leak size was estimated to be $3.02mm$ and a location of $2.6m$ for case A. For case B the leak size was estimated as $2.9mm$ at a leak location of $0.8m$. The actual leak location was at $2m$ for case A and had a leak size of $3mm$.

For the Bayesian Probabilistic analysis it can be seen that the same leak probabilities was found for the two cases. For case A the leak size was estimated as $2.96mm$ and the leak location as $0.9m$. For case B the leak size was estimated as $3.02mm$ and the leak location as $2.1m$.

The solution found by the SVM can be seen to be better for case A, with a 95% leak probability, a leak size of $2.96mm$ and a leak location of $1.5m$. The leak size and location estimation by the SVM is similar to that found for the first network. For case B the leak

probability was found to be 98%, a leak size of 2.9mm and a leak location of 1.5m. The solution found by the ANN can also be seen to be the same as the solution for the first numerical experimental network. A perfect leaking pipe classification was found, with the leak size estimated as 3mm and the location at 1.5m for case A. For case B the leak could be classified correctly to the leaking pipe, with a leak size estimation of 3.6mm and a location of 2m.

Discussion of Results

For the first numerical experimental network, all the algorithms could detect the correct leaking pipe, and estimate the leak size correctly. The leak location was found hard to solve. This is due to the small length of pipe and the low friction coefficients in the pipe. The MSE and Bayesian Probabilistic Analysis methods solved the leak location within $0.5-1.5m$. Table 3.1 shows a summary of the results found for the first numerical network.

Table 3.1: Summary of Results found for the First Experimental Network

	Case A		Case B	
	Leak Size	Leak Location	Leak Size	Leak Location
Actual Leak	3mm	2m	3mm	1m
Mean Squared Error	2.94mm	0.5m	3.05mm	2.2m
Bayesian Probabilistic Analysis	2.92mm	0.95m	3.08mm	2.6m
Support Vector Machine	2.95mm	1.8m	2.98mm	1.8m
Artificial Neural Network	2.98mm	1.5m	2.95mm	1.5m

The SVM and ANN could not find a solution for the leak location, since they trained a solution to the average location of the leak in the given data set. The best solution was found by the SVM and ANN, although these results were merely the averages of the data sets.

For the second numerical experimental network, it could be seen that the same classifications for the leaking pipes could be found. The SVM found a better probability for the leaking case. Overall the leak location estimation seemed better for this network configuration, with the MSE optimization performing the best. The range of the solution for the leak size ranged between $0.04-0.6mm$ and the leak location between $0.1-1.1m$. Table 3.2 shows a summary of the results found by the algorithms for the second numerical network.

Table 3.2: Summary of Results found for the Second Experimental Network

	Case A		Case B	
	Leak Size	Leak Location	Leak Size	Leak Location
Actual Leak	3mm	2m	3mm	1m
Mean Squared Error	3.02mm	2.6m	2.9mm	0.8m
Bayesian Probabilistic Analysis	2.96mm	0.9m	3.02mm	2.1m
Support Vector Machine	2.96mm	1.5m	2.9mm	1.5m
Artificial Neural Network	3.0mm	1.5m	3.6mm	1.5m

3.3 Error Analysis on the Numerical Network 1

The error analysis is performed to find the sensitivity of the network, and to find an estimation of allowable model errors. This estimation will also indicate the accuracy required by the sensors. The error analysis can be performed by calculating the error between two leaking points on the network. These two points are the measurement at the actual leaking location, with a simulation of a leak at another location. The error between these two measurements can be calculated by:

$$\%Error = \sum_{i=1}^{N} \frac{(H_i - H_i^*) \times 100}{H_i^*} + \sum_{j=1}^{L} \frac{(F_j - F_j^*) \times 100}{F_j^*}, \tag{3.1}$$

where H and F indicate the observed pressure and flow values, i and j indicate the observed node and pipe numbers, and N and L indicate the number of observed nodes and pipes. The $*$ indicates the measurements for the known leak location for the associated leak scenario. This calculation is performed over the length of each pipe with an exhaustive search algorithm to quantify the error over the network. The error calculated gives an indication of an applied error to the network with the worst possible division between the observational nodes.

The error analysis is performed on the same network used in Section 3.1, with the same leak locations. The numerical experimental network used can be seen in Figure 3.2 with the first leaking case on pipe three with a leak size of $3mm$ and at a location of $2m$. The second leaking case was chosen on pipe six with a size of $3mm$ at a location of $1m$. For the error analysis, the leak size is left at a constant of $3mm$ since it was found in the previous section that the leak size can be solved accurately with the leak location causing the most uncertainty.

Results

Calculating the error for each pipe over its length, the results can be plotted. Only the pipes with errors less than 5% were plotted.

Network 1: Case A

The error analysis for the case A on the first numerical network can be seen in Figure 3.8. The actual leak case can be seen in the third plot, marked with a dot. In this graph a 2% error line can be seen. This indicates the location of the leak if a 2% error exist on the actual measurements. It can be seen that for a 2% error the leak location can be estimated between $1.5m$ and $2.4m$.

If the error on the measurements increased to 5%, it can be seen that a solution can be found up to three pipes away. It can also be seen that, to ensure the solution is on the actual leaking pipe, a measurement error of 2.6% or less needs to be achieved.

Network 1: Case B

The error analysis for leaking scenario B can be seen in Figure 3.9. The actual leak location can be seen in the second plot in this figure. The actual leak location is marked with a dot.

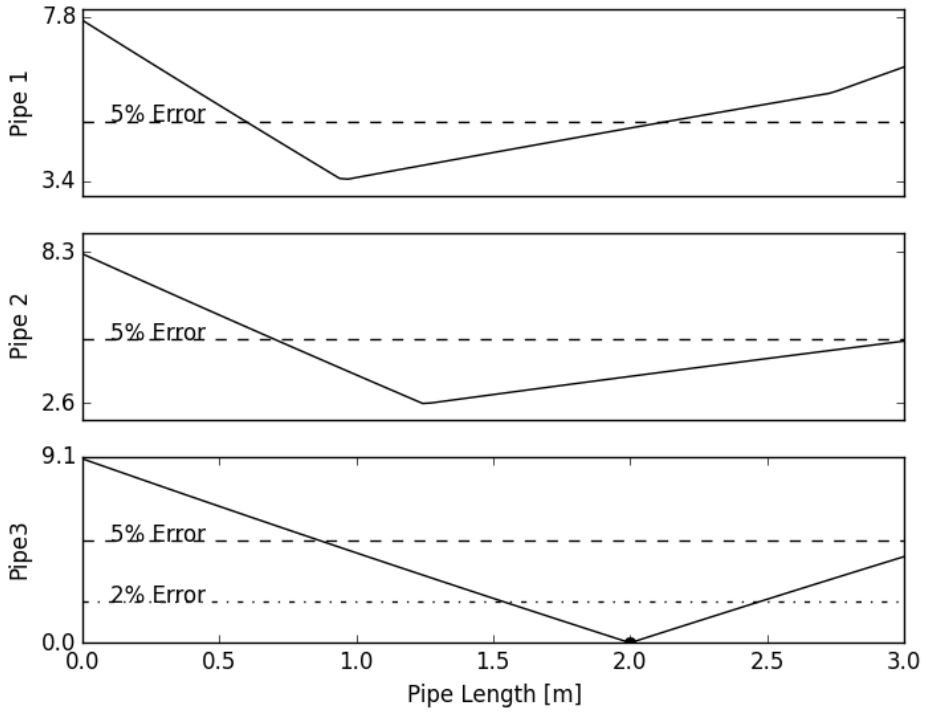


Figure 3.8: Numerical Network 1 Leak Case A: All errors (computed using Eq:3.1) that are less than 5%, when considering all location for all seven main pipes.

For this leak case, it can be seen that, if a 2% error is found in the measurements, the location of the leak can be estimated between $0.6m$ to $1.3m$. This gives a maximum location error of $0.4m$. If there is a 5% error in the measurements, it can be seen that the leak location can be potentially identified in three possible pipes. The maximum allowable error to ensure the leak is uniquely identified on the correct main pipe is 3.1%.

Discussion of Results

In this section an error analysis was performed on the numerical experimental network where two leak cases were tested. For the first leak case, a maximum total error of 2.6% is required to ensure the solution is found on the actual leaking main pipe. For a total error of 2.6% the leak location can be estimated within $0.6m$.

For the second leak case, it was found that a maximum total error of 3.1% would ensure the leak to be found on the actual leaking main pipe. For a total error of 3.1% the leak location should be estimated within $0.75m$ of the actual leak location.

3.4 Conclusion

In this initial numerical investigation and the studies conducted in the previous chapter on the different networks, it was found that all the algorithms could find the leaking pipe and

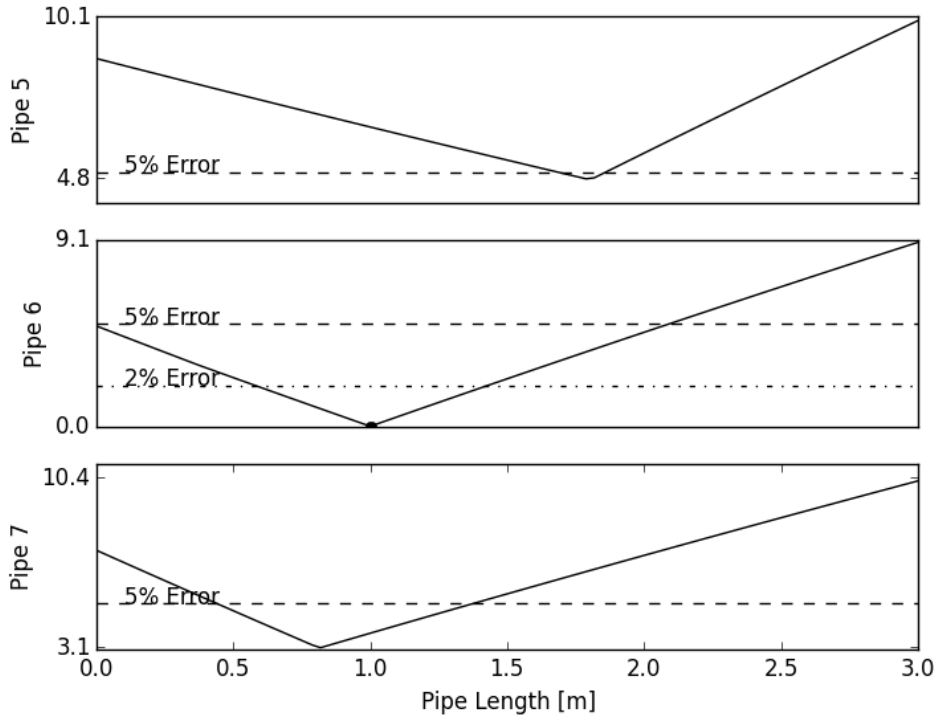


Figure 3.9: Numerical Network 1 Leak Case B: All errors (computed using Eq:3.1) that are less than 5%, when considering all locations for all seven pipes.

the size of the leak with good accuracy. The leak location on the specific pipe could not be estimated accurately, with the MSE and the Bayesian Probabilistic Analysis algorithms terminating before the correct solution was found. This is due to the EPANET model not having enough sensitivity to allow for further optimization with the differential evolution algorithm. It was also found that the SVM and the ANN trained to the average location of the given data sets.

From the error analysis on the numerical network 1 it could be seen that the largest total error that could be accommodated to still uniquely identify the correct main pipe for the leak is 2.6%. This total error includes sensor measurement, noise and model errors.

This numerical investigation showed that all algorithms had difficulties to estimate the exact locations of the leaks, while the leak size and leaking main pipe could be robustly and accurately estimated. Since the leak can be located on the main leaking pipe accurately, it is not necessary to test seven pipes in one network. Additionally, the addition of the pump to supply the network does not impact the solution of the network, since a leak has such a small effect on the network, the pump can nearly be seen as a constant pressure supply. This indicates that the pump selection does not require a significant amount of attention.

The EPANET model should be calibrated for the actual experimental setup that requires the roughness coefficients, pump efficiencies and minor loss coefficients at strategic positions to be estimated.

Chapter 4

Experimental Investigation of Two Flow Networks

This chapter outlines the experimental setup and results for an experimental investigation. The focus of this experimental study is to generate real data and apply the different machine learning techniques to actual data to try and locate the leaks when actual model and measurement errors are present in the measurement signal.

In the numerical investigation in Chapter 3, merely the maximum allowable total error that could accommodate the identification of a unique leak location was quantified. In this chapter we investigate whether a lab experiment allows for the model and errors to be sufficiently low to still uniquely define the leak location along a main pipe. The model errors is a direct result of differences between the simulate network and the physical network, while measurement errors are a direct result of the quantity of the pump and sensors used in experimental setup.

4.1 The Design

The experimental network used is based on the networks seen in Figures 3.2 and 3.7. The planned experimental networks in these figures used six pressure sensors at the start and end of each pipe, with a single flow meter measuring the flow into the network. Figure 4.1 shows the layout for the experimental network.

In this figure it can be seen that six possible leaking pipes were selected. These pipes are fed by a pump from a reservoir. The demands on the network will be modeled with $3mm$ holes, which feed back to the reservoir. Valves are inserted next to the demand locations to change between experimental network 1 and 2. The network was simplified from the planned experimental network investigated in the previous chapter. The layout of the network was chosen as shown to allow for the water to be returned to the pump reservoir.

The 12 pressure sensor locations at the start of each pipe, and at the end of it can be seen. Six flow sensors were added at the end of the network, measuring the demands from each pipe. An additional flow sensor has been added to measure the flow rate into the

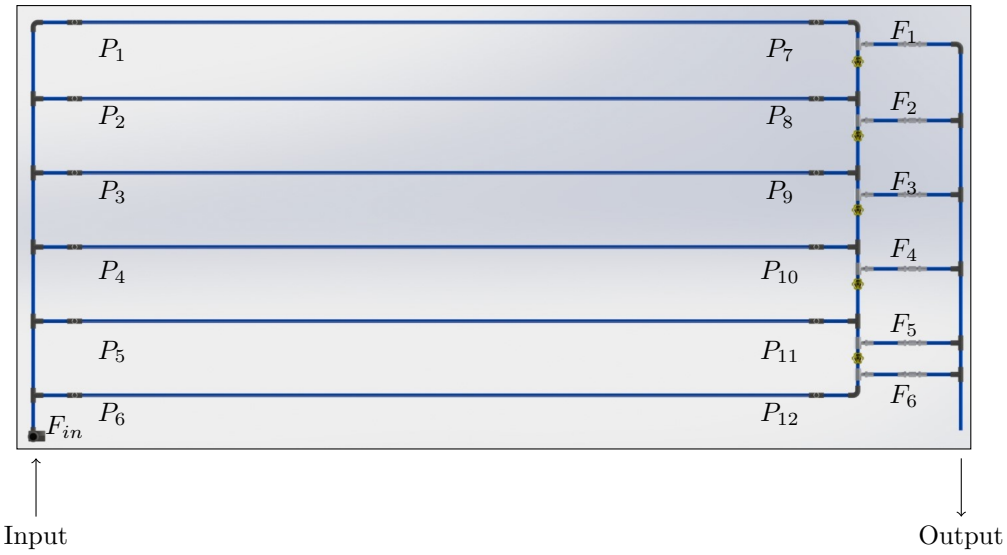


Figure 4.1: Schematic of the Experimental Network

network.

The Pentax CM 210 pump was used, it was seen that a flow of $1l/s$ can be expected through the network, with this pump delivering a maximum pressure of about $0.44MPa$. The lengths of each pipe is $3m$ with a diameter of $10mm$.

4.2 Physical Experimental Setup

The experimental setup was completed following the dimensions and layout of the experimental design.

4.2.1 Construction

Figure 4.2 shows the built experimental network. In this network the six pipes can be seen, with all the sensors installed as well as the pump. The reservoir used is a simple container holding $50l$ of water and can not be seen in this photo.

Pipe referred to as Polycop $15mm$ diameter pipe was used for the experiment. It is made from polypropylene and has an inner diameter of $10mm$.

4.2.2 Leak locations

Three leak locations were added to the network. The first leak is located at the first pipe, at a length of $2m$ and a diameter $3mm$. The second leak is located at the third pipe, at a length of $1m$ and a diameter of $2mm$. The third leak is located at the fifth pipe at a length of $1.5m$ with diameter of $2mm$. The leaks are added or removed to the network by opening and closing valves. This makes changing between leak cases and multiple leak cases to be tested easy.

Figure 4.3 shows a $3mm$ leak applied to the network. The leak is applied by clamping a saddle over the pipe, and drilling the correct size hole through the pipe. All the leaks are

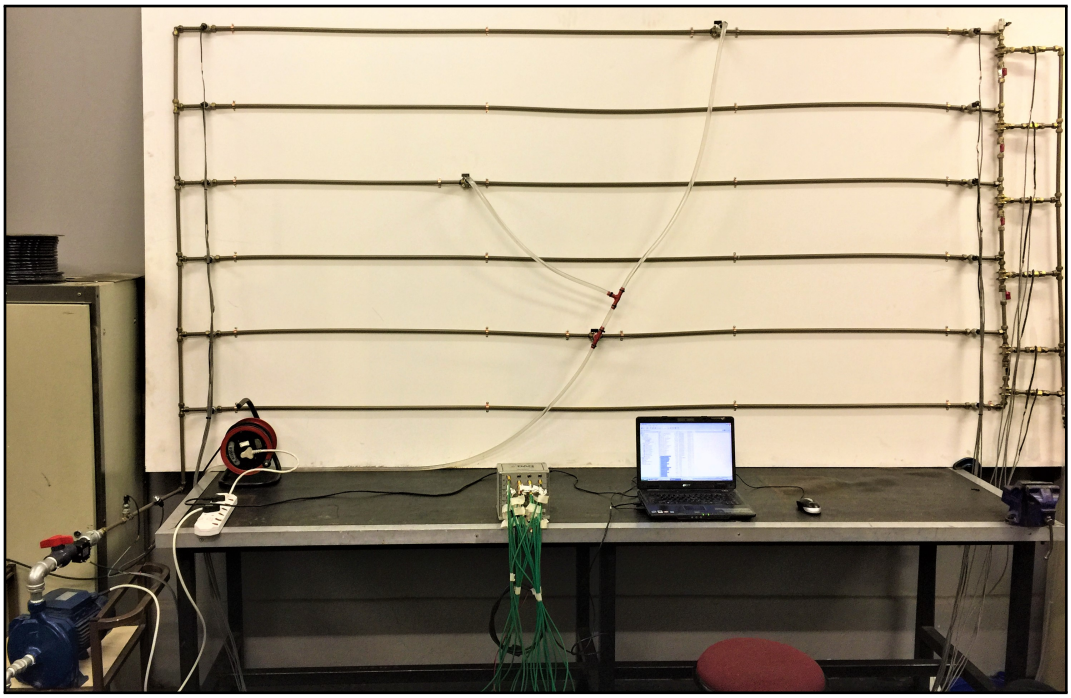


Figure 4.2: Physical Experimental Setup Built

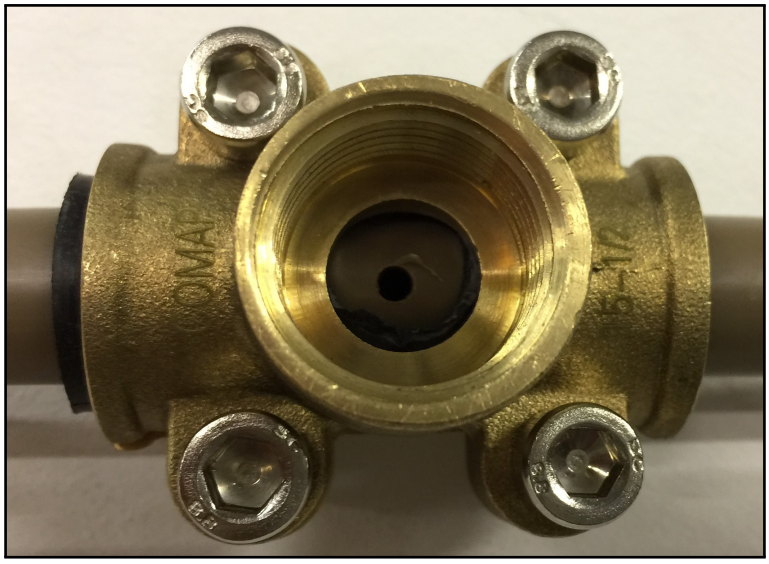


Figure 4.3: Actual Leak of $3mm$ Applied to the Experimental Network

connected together and fed back to the reservoir. The demands applied to each pipe was created in the same way. All the holes used were created in the same way, and therefore can be considered to be round.

4.2.3 Pressure and Flow Sensors Installed

The pressure and flow sensors used can be seen in Figure 4.4 $a - c$. The pressure sensor used can measure pressure between $0 - 0.5MPa$. The sensor uses a $5V$ input voltage and returns a voltage between $0.5 - 4.5V$. From the measured voltage the pressure range can be computed. These pressure sensors have a measurement accuracy of 1%.

The flow sensor used to measure demands from each pipe can be seen in Figure 4.4 *b*. This sensor can measure flow between $1 - 25l/min$ with a measurement accuracy of 3%. The flow sensors returns $5V$ pulses, from which the frequency can be calculated. The flow rate for this sensor can be calculated with:

$$f = F_{PC} \times Q, \tag{4.1}$$

where f is the frequency of the pulses, Q is the flow rate in l/min and F_{PC} is the flow pulse characteristic for the sensor. This value is given as 11 for this sensor in the data sheet.

The second flow sensor shown in Figure 4.4 *c* is used to measure the input flow into the system. This flow sensor can measure a flow of size up to $120l/min$. This flow sensor works similar to the previous sensor, with a flow pulse characteristic of 0.45. This sensor has a measurement accuracy of 2%.

To record the measurements by these sensors, a SOMAT eDAQlite data acquisition system was used. The system can take up to 24 input channels and was set to sample at a frequency of $1000Hz$. The data was collected by a laptop computer where it could be analyzed.

(a) $0.5MPa$ Pressure Sensor

(b) $1 - 25l/min$ Flow Sensor

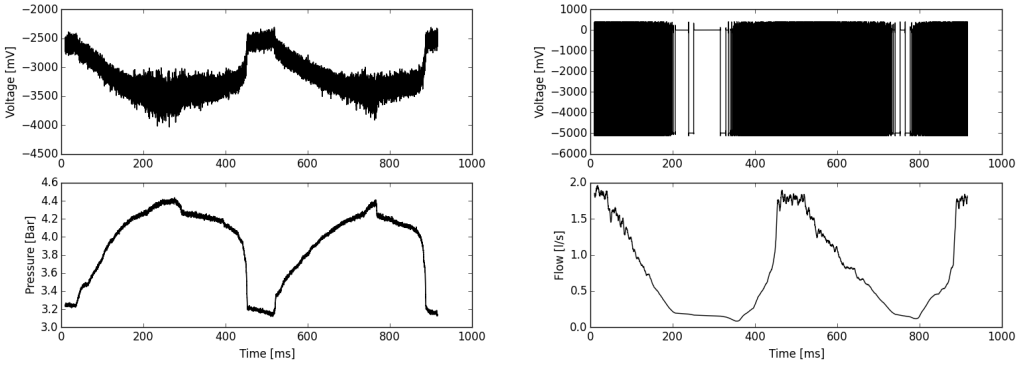
(c) $1 - 120l/min$ Flow Sensor

Figure 4.4: Sensors used to Measure the Response of the Experimental Network

4.2.4 Pump Test

The pump used for the experiment was the Pentax CM210. The pump is capable of supplying a maximum pressure of 4.4 bar and a maximum flow rate of $2l/s$ at a pressure of 3.2 bar. A test was performed with the pump to verify the pump curve supplied by the supplier. To do this a flow and pressure sensor was added to the pump before a valve. The pressure sensors used were calibrated to ensure they measure constant atmospheric pressure. Changing the valve setting induced load to the pump. Figure 4.5 shows the raw pressure and flow sensor voltage measurements and the converted pressure and flow rate values for this test. It can be seen that the valve was opened and closed two times within about 10 seconds. The

measurement does not reach zero due to the filter applied to the raw measurements.



(a) Pressure Sensor Measurement (b) Flow Sensor Measurement

Figure 4.5: Measurements taken and converted for the pump test

Combining these measurements onto one graph gives the pump curve, which can be seen in Figure 4.6. The dashed line in this figure shows the given pump curve from the pump data sheet. It can be seen that the pump does not deliver the flow as it should, indicating some pump efficiency will be required. An IIR filter was used on all the data to reduce the amount of noise.

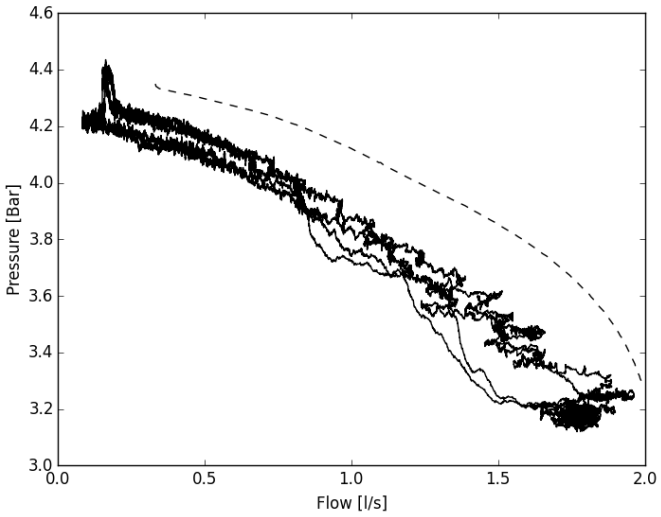


Figure 4.6: Tested pump curve for Pentax CM210 pump

4.3 Measurements for Three Leak Cases

The measurements for three leak cases were completed by measuring for about 30 seconds with no leak, opening the leak for 1 minute and then closing the leak while measuring for another 30 seconds. Figure 4.7 shows the pressure measurements for experimental network 1 and leak case A.

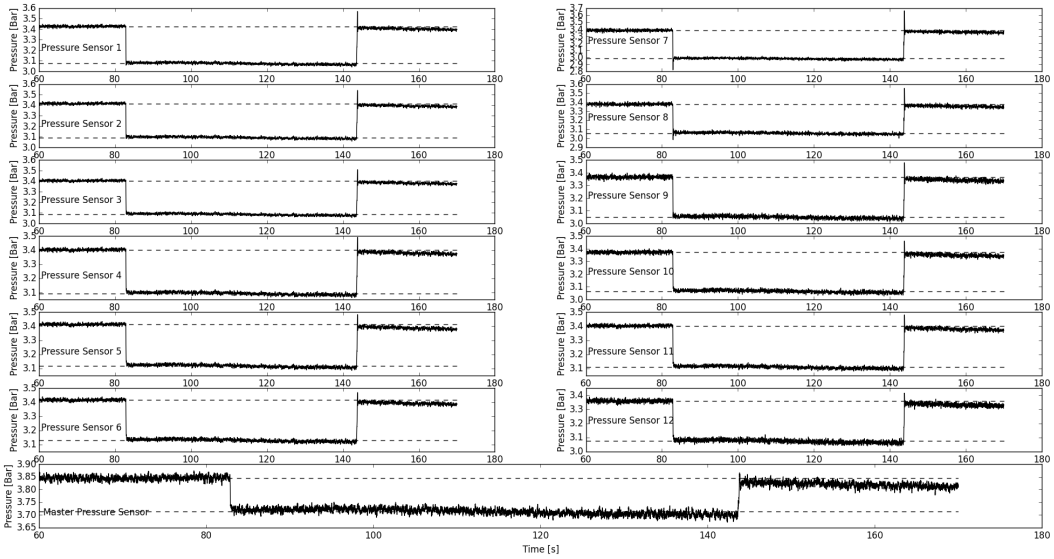


Figure 4.7: Network 1 Leak Case A: Pressure Measurement for the twelve pressure sensors and an additional master pressure measurement at the inlet of the system.

In this figure a clear drop in pressure can be seen for every sensor. The sensors in this figure are labeled 1 to 12, indicating the locations of the pressure sensors as outlined in Figure 4.1. A $13th$ pressure was measured at the outlet of the pump. Dashed lines are shown in each measurement graph. These lines indicate the average values for the leak and no leak cases that was used to find the leak locations.

Figure 4.8 shows the flow measurements for the first experimental network and the first leak case. In this figure the different demands can be seen for each pipe. The drop in flow can clearly be seen in the first six graphs, indicating the leak was applied. The seventh graph shows the inlet flow to the network. It can be seen that this flow increases when a leak occurs in the system. The final graph in this figure shows the sum of the inlet and outlet flows. It can be seen that this sums to zero when no leak is applied to the network, and when a leak is in the system, a leak size of $0.2l/s$ can be calculated.

In total, fourteen different leak cases were tested. They were a single leak in the network for the three leak cases, two leaks in the network, and three leaks in the network. These leak measurements were completed for network 1 and 2..

4.4 Model Calibration

The modeled networks of the anticipated experimental networks tested in previous chapters were adapted to more accurately represent the current experimental networks. The updated numerical network can be seen in Figure 4.9. In this figure the location of the minor losses to be calibrated can be seen, marked with red rectangles. These minor losses will be calibrated to simulate inlet and outlet losses due to turbulence at these locations by solving an inverse problem. An additional loss was added to the leak return pipe and one to the demand return pipe. Two additional parameters will be calibrated, the roughness coefficients of the

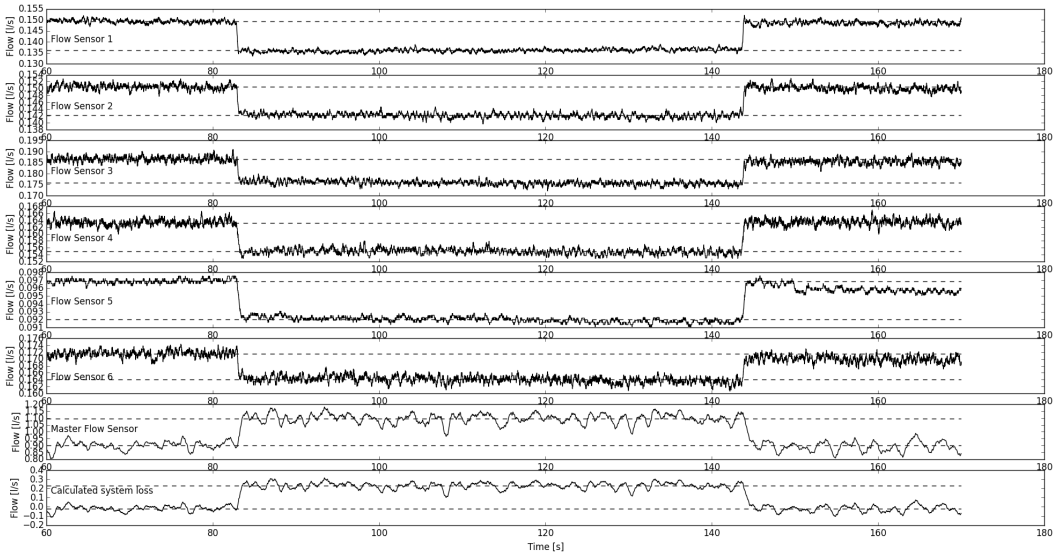


Figure 4.8: Network 1 Leak Case A: Flow measurements for the six outlets, the inlet and a calculated system loss.

network and the pump efficiency.

The calibration of these parameters will be done by comparing the pressure and flow values between the simulated network and the actual measurements. Four cases will be used to determine the optimum parameters, a case where no leak exist, and the three single leak cases. The optimization of the parameters will be completed for both Network 1 and 2 resulting in the eight measurement sets used to calibrate the two models. The algorithm used to calibrate the parameters is a differential evolution algorithm. The calibration function is a Mean squared error between the twelve pressure and six out flow measurements. The calibration is completed for Network 1 and 2 since the outflow loss coefficients change.

4.4.1 Numerical and Experimental Network 1

Figure 4.10 *a* shows the measurements with no leak before calibration of the parameters were completed. The dashed line shows the measurements from the EPANET model and the solid lines show the three independent measurements with no leaks applied to the experimental network. Figure 4.10 *b* shows the measurements with no leak after calibration of the parameters. It can be seen that the calibrated pressures vary with very little head between the EPANET model and the experimental measurements. These figures can also be seen with increased size in Appendix B, Figures B.1 and B.2.

Figure 4.11 shows the measurements of the first leak case after calibration. Appendix B contains the figures for the additional leak cases, in Figures B.3 to B.5. It can be seen that the pressure measurements and the model simulated values match quite accurately. The flow measurements for these leak cases could not fit the leak cases, additional parameters and changes to the network might be required. Only pressure measurements will be used to solve the leak locations, therefore the optimized parameters should be sufficient.

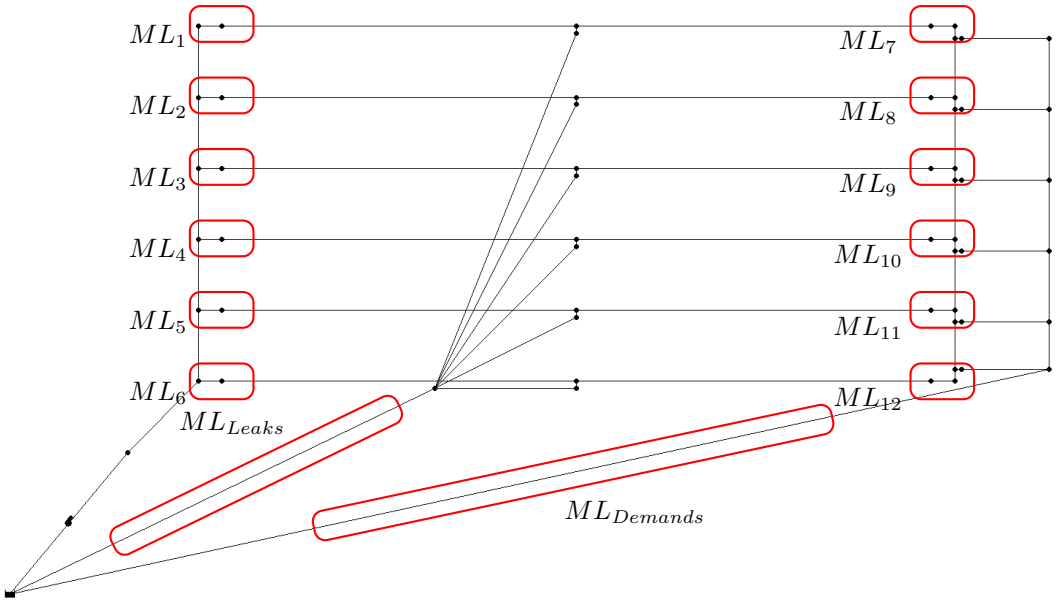
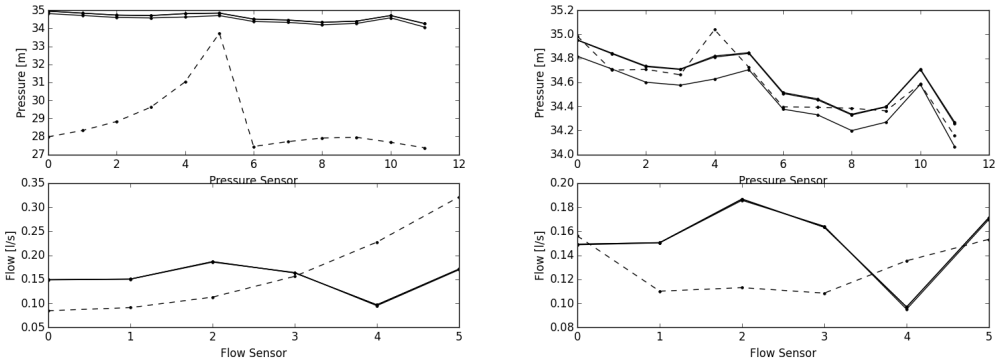


Figure 4.9: EPANET model used to calibrate measurements of the updated numerical network to measure on the experimental network by varying the surface roughness, minor losses and pump efficiency.



(a) Values before calibration

(b) Values after calibration

Figure 4.10: Network 1: Numerical (dashed lines) and experimental (solid lines) pressure and flow measurements (a) before and (b) after calibration with no leak, at the 12 pressure and 6 flow sensor locations.

Before calibration, the maximum and minimum pressure difference between the numerical model and the experimental measurements with no leak was found to be $7m$ and $1.5m$. The difference in flow rate before and after calibration of the numerical model and the experimental measurements showed to be between $0.05l/s$ and $0.15l/s$. After calibration of the parameters, the difference range of the pressure measurements was found to be between $0m - 0.1m$ with the difference in flow rate being between $0.08l/s - 0.7l/s$. Table 4.1 shows a summary of the minimum and maximum differences between the pressure and flow rate measurements of the numerical model and the experimental measurements after calibration for the three leak cases.

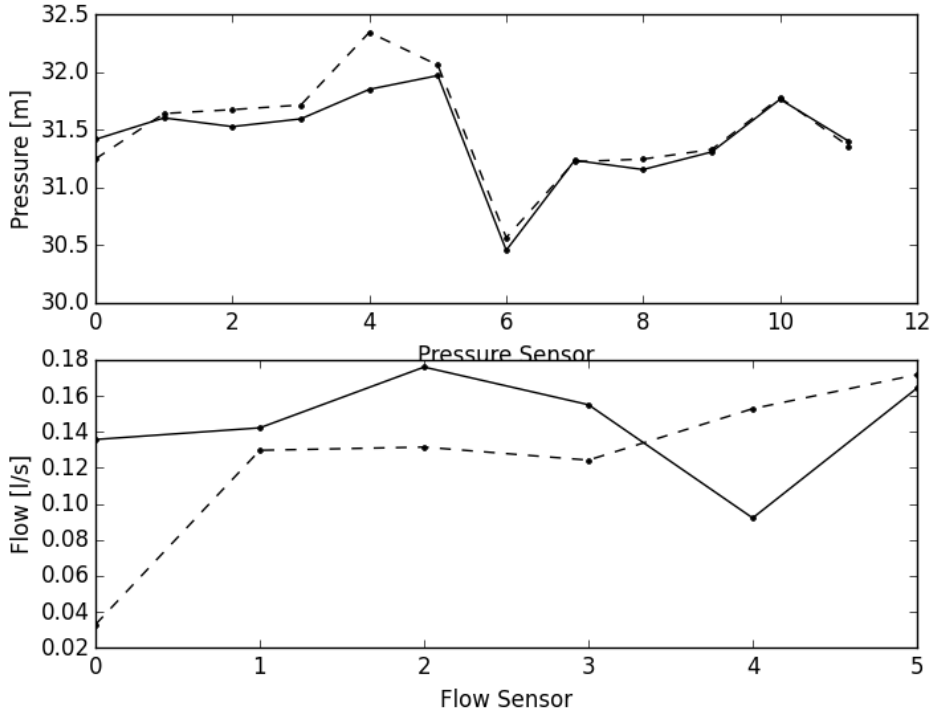


Figure 4.11: Network 1 Leak Case A: Numerical (dashed lines) and experimental (solid lines) pressure and flow measurements after calibration with a leak, at the 12 pressure and 6 flow sensor locations.

Table 4.1: Summary if the differences between the minimum and maximum differences for the three leak cases after parameter calibration.

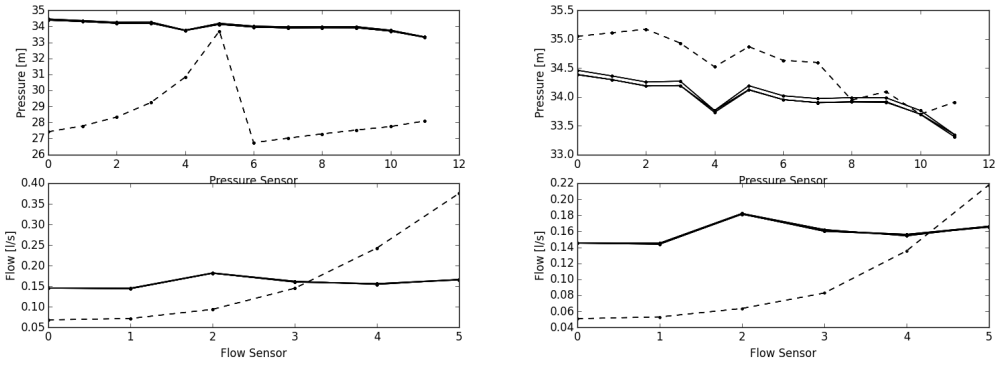
	Pressure Measurements		Flow Rate Measurements	
	Minimum	Maximum	Minimum	Maximum
Leak Case A	$0.01m$	$0.2m$	$0.01l/s$	$0.7l/s$
Leak Case B	$0m$	$0.6m$	$0.01l/s$	$0.12l/s$
Leak Case C	$0.65m$	$1.65m$	$0.1l/s$	$0.18l/s$

4.4.2 Numerical and Experimental Network 2

Figure 4.12 shows the measurements before and after calibration with no leaks. Within the second graph it can be seen that a good fit between the pressure measurements was found, with a maximum error of about $0.5m$. Appendix B Figure B.6 and B.7 shows the same images a bit larger.

Figure 4.13 shows the calibrated pressure and flow measurements for the first leak case. Appendix B figure B.7 to B.9 contains the figures for all three the leak cases after calibration. It can be seen that the flow measurements do not match the measured values. Changing the parameters can be seen to have no effect on the demand flow, this is due to the demands being connected. The pressure values can be seen to have a good match.

Before calibration, the maximum and minimum pressure difference between the numerical model and the experimental measurements with no leak was found to be $6m$ and $0.8m$. The difference in flow rate before and after calibration of the numerical model and the experimental measurements showed to be between $0.02l/s$ and $0.2l/s$. After calibration of the parameters, the difference range of the pressure measurements was found to be between $0m - 0.6m$ with the difference in flow rate being between $0.03l/s - 0.8l/s$. Table 4.2 shows a summary of the minimum and maximum differences between the pressure and flow rate measurements of the numerical model and the experimental measurements after calibration for the three leak cases.



(a) Values before calibration

(b) Values after calibration

Figure 4.12: Network 2: Numerical (dashed lines) and experimental (solid lines) pressure and flow measurements (a) before and (b) after calibration with no leak, at the 12 pressure and 6 flow sensor locations.

Table 4.2: Summary if the differences between the minimum and maximum differences for the three leak cases after parameter calibration.

	Pressure Measurements		Flow Rate Measurements	
	Minimum	Maximum	Minimum	Maximum
Leak Case A	$0.01m$	$1m$	$0.015l/s$	$0.1l/s$
Leak Case B	$0.1m$	$1.4m$	$0.015l/s$	$0.09/s$
Leak Case C	$0.01m$	$1.1m$	$0.3l/s$	$0.12/s$

4.4.3 Calibrated Network Parameters

Table 4.3 shows the optimized parameters for the two networks. It can be seen that the roughness coefficients changes significantly between the two networks. This is due to the sharp changes in flow rate within the pipes.

In this table, the first minor loss coefficient was found to be smaller than the other loss coefficient since this coefficient is applied to an L-bow, where the other coefficients are applied to T-pieces. Additionally it can be seen that, for Network 1, the minor loss coefficients $2 - 6$ have similar values. It can once again be seen that, for Network 1, the smallest minor

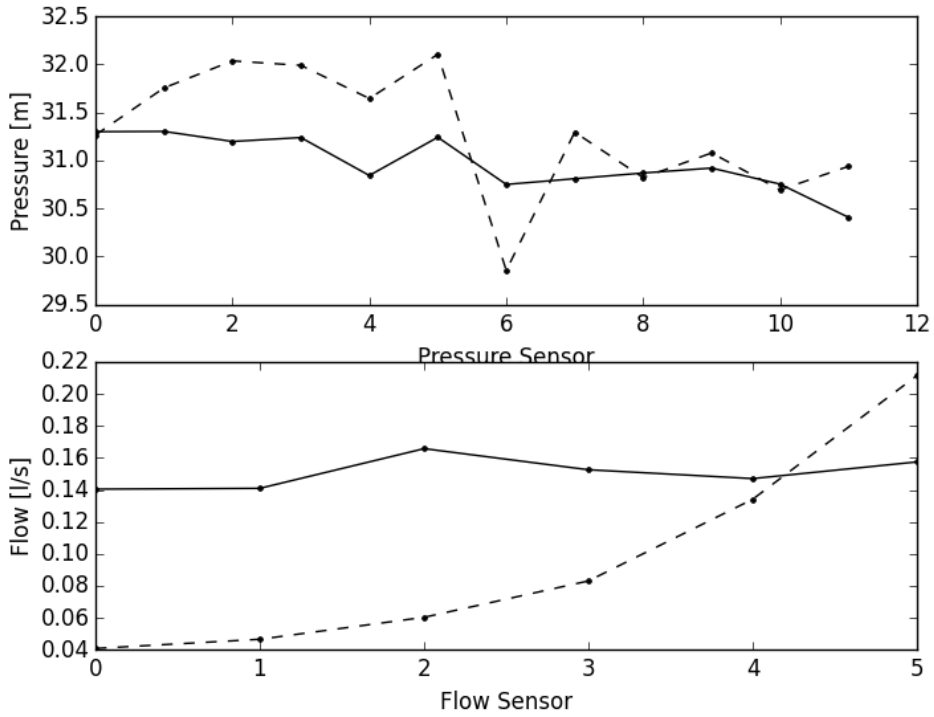


Figure 4.13: Network 2 Leak Case A: Numerical (dashed lines) and experimental (solid lines) pressure and flow measurements after calibration with a leak, at the 12 pressure and 6 flow sensor locations.

loss is applied to the L-bow at the exit of the first pipe. In Network 2, this value can be seen to be one of the largest coefficients. This is due to the low flows within this pipe that makes this parameter insensitive to the pressure and flow rate measurements. Minor loss coefficients 1 to 6 for Network 2 can be seen to increase in size as the flow increases through the network.

The minor losses in the demand return pipe can be seen to be quite high. This is due to the amount of flow in this pipe, coupled with the length of the pipe. The minor losses within the return pipe of the leaks were found to be between 10 and 20, although the pipe is similar to that of the demand return pipe.

The network errors can be calculated using the same method used in chapter 3. The calculated errors can be seen in Table 4.4. It can be seen that for network 1, small errors were calculated for the first and no leak cases, with large errors for the second and third leak case. High errors were calculated for network 2 on all leak cases.

4.5 Inverse Strategies Results

The average of the measured pressures on the experimental network 1 and 2 is calculated. The strategies to solve for the leak location and leak sizes could be used to find results for the two experimental networks. For this section, the dashed lines indicate the numerical network solution for the leak, with the solid line indicating the solution from the averaged

Table 4.3: Optimized parameters for Network 1 and 2

Optimized Parameter	Network 1	Network 2
Roughness Coefficient	245.7	136.6
Pump Efficiency	80.1%	82.2%
Minor Loss 1	11.0	0.8
Minor Loss 2	28.6	5.9
Minor Loss 3	31.4	5.0
Minor Loss 4	40.8	19.4
Minor Loss 5	28.9	38.4
Minor Loss 6	29.6	45.7
Minor Loss 7	1.6	42.6
Minor Loss 8	1.7	27.9
Minor Loss 9	0.2	4.0
Minor Loss 10	0.1	5.7
Minor Loss 11	2.4	0.0
Minor Loss 12	0.5	0.1
Minor Loss Demands	287.0	348.3
Minor Loss Leaks	14.2	18.6

Table 4.4: Percentage error in the measurements after calibration

	No Leak	Leak A	Leak B	Leak C
Error in Network 1 [%]	2.52	4.3	44.7	28.8
Error in Network 2 [%]	19.2	18.9	15.2	21.0

experimental measurements. This section only shows the results found for leak Case A on Network 1 and 2. Additional leak case results can be seen in Appendix C.

4.5.1 Experimental Network 1

The solution for Network 1 leak case A is shown in Figures 4.14 -4.17 for each respective solution strategy. For leak case A, one leak is located on pipe number 0 at $2m$ with a leak size of $3mm$. For leak case B, one leak is located on pipe number 2 at $1m$ with a leak size of $2mm$. For leak case C, one leak is located on pipe number 4 at $1.5m$ with a leak size of $2mm$.

Mean Squared Error

The solutions for leak case A by using the mean squared error optimization can be seen in Figure 4.14. The results for all the leak cases can be seen in Appendix C, Figures C.1 to C.3. For leak case A, it can be seen that the numerical solution was found with an error of effectively zero. The error found for the solution with the experimental measurements was about 0.2, which is the minimum error in the solution. The size of the leak was estimated to be $3.3mm$ instead of $3mm$, and the location of the leak was found with an error of $1m$.

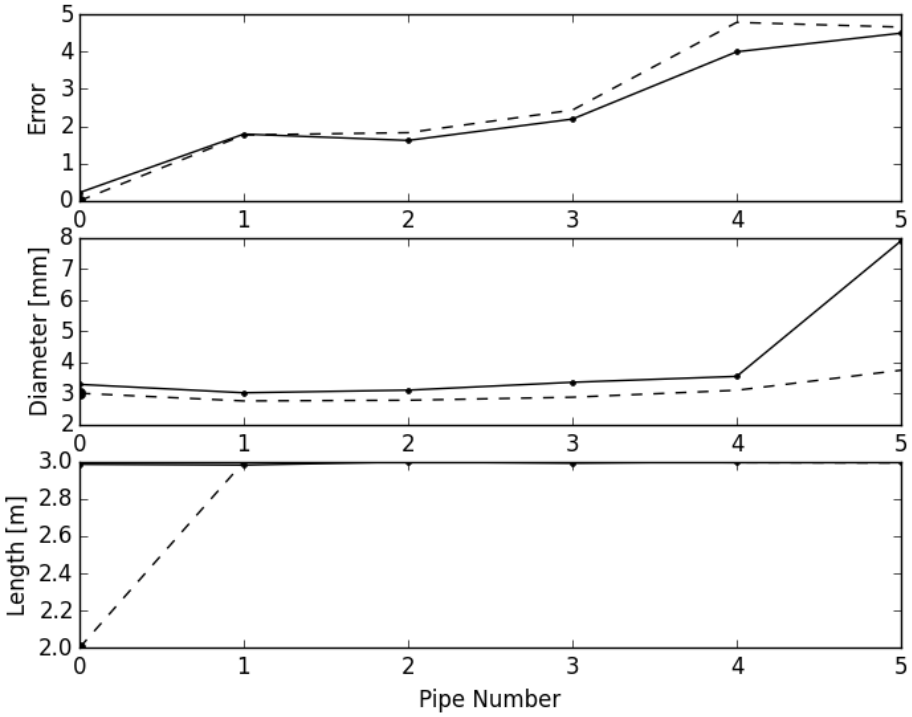


Figure 4.14: Network 1 Case A: Mean Squared Error Optimization solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

For case B the leak was found with an error of about 0.4, which is the minimum error in the solution. The leak size was found to be $1.35mm$ instead of $2mm$ and the location of the leak was found as $3m$.

For leak case C, the solution found was with an error of 0.4, which was the minimum error, followed by an error of 0.6 for the sixth pipe. The size of the leak was calculated as $1.35mm$ instead of the actual $2mm$. The leak location was found to be at $1.1m$, with an error of $0.4m$.

Bayesian Probabilistic Analysis

The solution found by the Bayesian Probabilistic Analysis for leak case A can be seen in Figure 4.15. The results for all the leak cases can be seen in Appendix C, Figures C.4 to C.6. For leak case A, it can be seen that the numerical leak could be found with a probability of 98%. The experimental solution was given a probability of 52%, with the next closest solution being the second pipe with a probability of 20%. The leak size and locations solutions are similar to the Mean Squared Error solution.

The numerical case leak probability found by this algorithm can be seen as 90% for leak case B. The experimental leak probability was found to be 50% with the fourth pipe having a probability of 36%. The third leak case could be found with a 98% probability for the numerical leak data. The algorithm found the leak with a probability of 55% for the exper-

imental leak data. The next closest solution to the actual leak case was found with a leak on the sixth pipe with a probability of 21%.

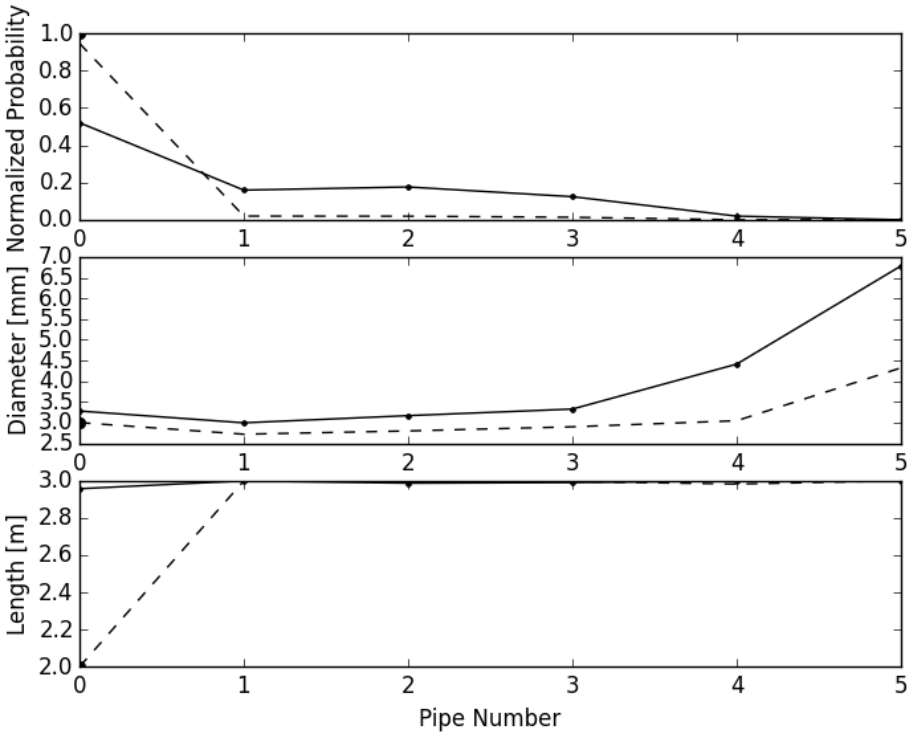


Figure 4.15: Network 1 Case A: Bayesian Probabilistic Analysis solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

Support Vector Machine

The solution found by the Support Vector Machine can be seen in Figure 4.16. The results for all the leak cases can be seen in Appendix C, Figures C.7 to C.9. For all the leak cases, the leak could be located with a probability of 100% for the numerical case and the experimental measurements. For leak case A the leak size could be estimated as $3.38mm$, resulting in an error of $0.38mm$. The leak location was found with an error of $1m$.

For leak case B the leak size was estimated to be $1.4mm$, and the leak location was estimated to be at $2.2m$, resulting in an error of $1.2m$. For leak case C he leak size was found to be $1.4mm$ instead of $2mm$, and the location of the leak was found with an error of $0.2m$.

Artificial Neural Network

The solution found by the Artificial Neural Network can be seen in 4.17. The results for all the leak cases can be seen in Appendix C, Figures C.10 to C.12. The solutions was classified correctly for all the numerical leak cases. For leak case A with the experimental measurements, the leak classification was found with an accuracy of 100% and the size of the leak was estimated to be $2.6mm$. The leak location was found to be $2m$ for the numerical

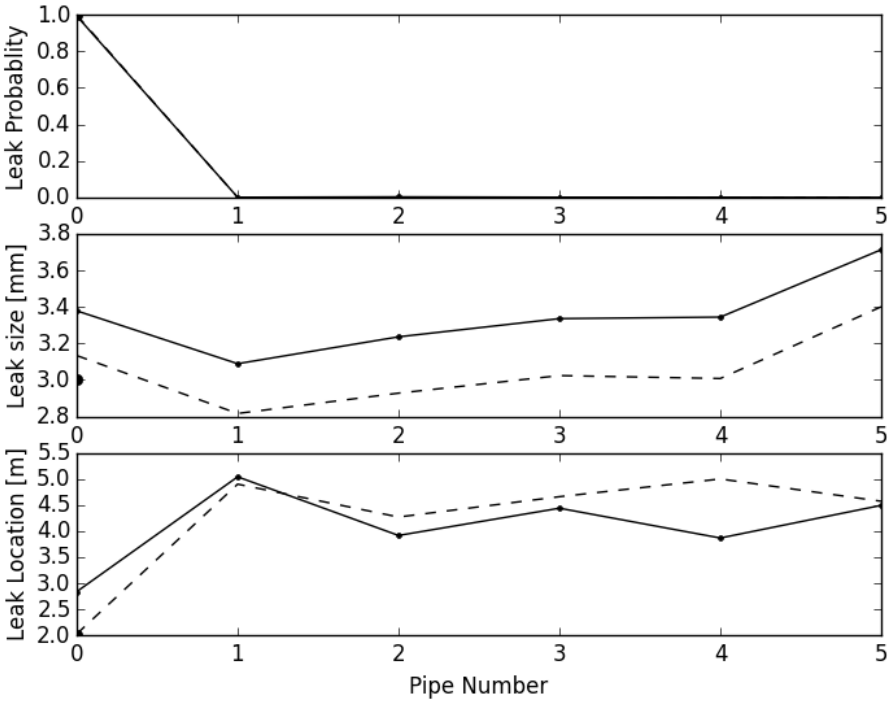


Figure 4.16: Network 1 Case A: Support Vector Machine solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

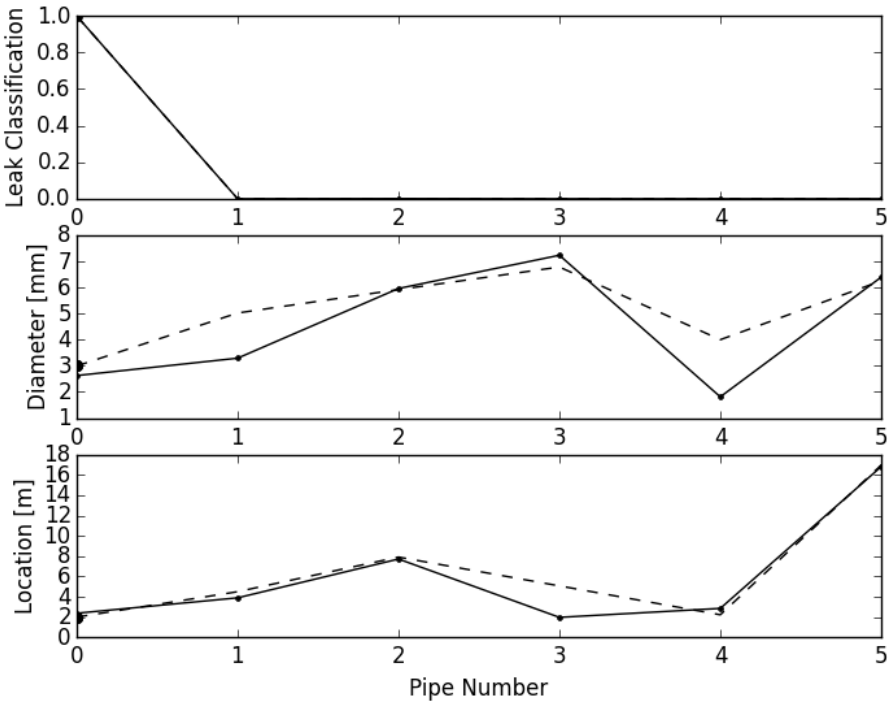


Figure 4.17: Network 1 Case A: Artificial Neural Network solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

leak case, and 2.4m for the experimental case.

The solution for the experimental measurements of leak case B and C were classified incorrectly. The size for leak case B was estimated to be 1.3mm, with a location of 2.1m. For case C the no leak was classified. The leak size was estimated to be 4.2mm, and the location of the leak was estimated at $-0.3m$.

4.5.2 Experimental Network 2

The solution for Network 2 leak case A is shown in Figures 4.18 -4.21 for each respective solution strategy. For leak case A, one leak is located on pipe number 0 at 2m with a leak size of 3mm. For leak case B, one leak is located on pipe number 2 at 1m with a leak size of 2mm. For leak case C, one leak is located on pipe number 4 at 1.5m with a leak size of 2mm.

Mean Squared Error

The solution found for network 2 by the mean squared error optimization can be seen in Figure 4.18. The results for all the leak cases can be seen in Appendix C, Figures C.13 to C.15. For leak case A an error of effectively zero was found for the numerical sensor data. With the experimental measurements, an error of 2.5 was achieved, with the second smallest error being 3.5 for the second pipe. The leak size was estimated to be 9mm, and the location was estimated with an error of 2m.

For leak case B, an error of effectively zero was achieved for the numerical case and an error of 1.5 for the experimental measurements. The leak size was estimated to be 1.78mm with an error in the leak location of 1m.

For leak case C, the location of the leak was estimated to be at the second and third pipe with errors of 1.5 and 1.7. The error for the actual leak case was calculated as third most likely with a value of 2.1. The leak size was calculated as 1.65mm and the location was calculated accurately with no error.

Bayesian Probabilistic Analysis

The solution for leak case A on network 2 by the Bayesian Probabilistic Analysis can be seen in Figure 4.19. The results for all the leak cases can be seen in Appendix C, Figures C.16 to C.18. For leak case A, a probability of 43% for the leak at the actual location was found using the experimental data. A probability of 35% was found for the leak being at the third pipe. The leak size and locations estimations are similar to the solutions obtained by the Mean Squared Error optimization.

For leak case B a leak probability of 37% was found for the actual leaking location. Probabilities of 29% and 25% were calculated for the first and second pipes to be leaking. For leak case C, a probability of 38% and 29% was found for the third and second pipe. The actual leaking pipe only had a probability of 15% to be leaking.

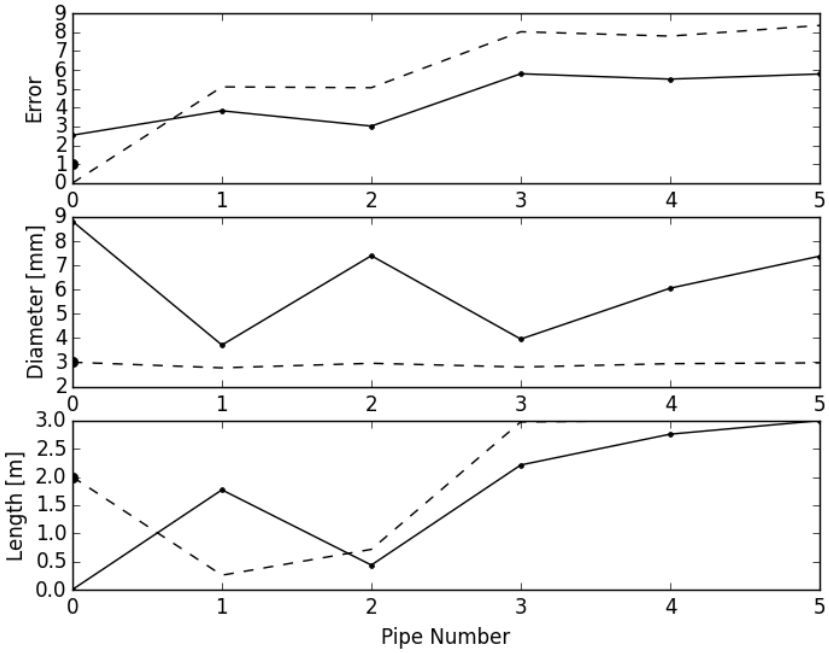


Figure 4.18: Network 2 Case A: Mean Squared Error optimization solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

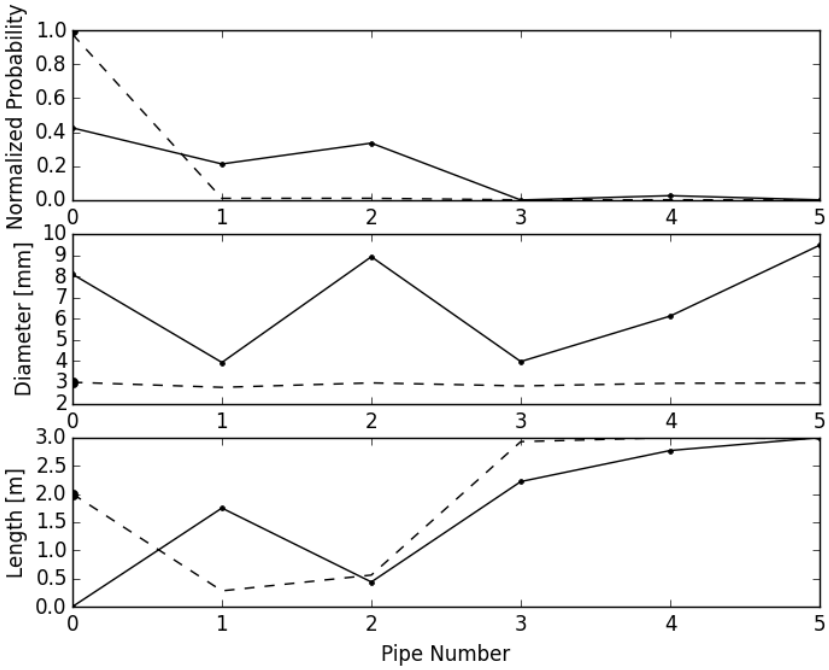


Figure 4.19: Network 2 Case A: Bayesian Probabilistic Analysis solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

Support Vector Machine

For case A of network 2 the solution by the support vector machine can be seen in Figure 4.20. The results for all the leak cases can be seen in Appendix C, Figures C.19 to C.21. For leaking case A, a probability of 35% was calculated for the leak being at the third pipe, while the actual leak location was given a probability of 25% as the second most likely leak location. The leak size was estimated to be 3.9mm and the location was found with an error of 1.25m.

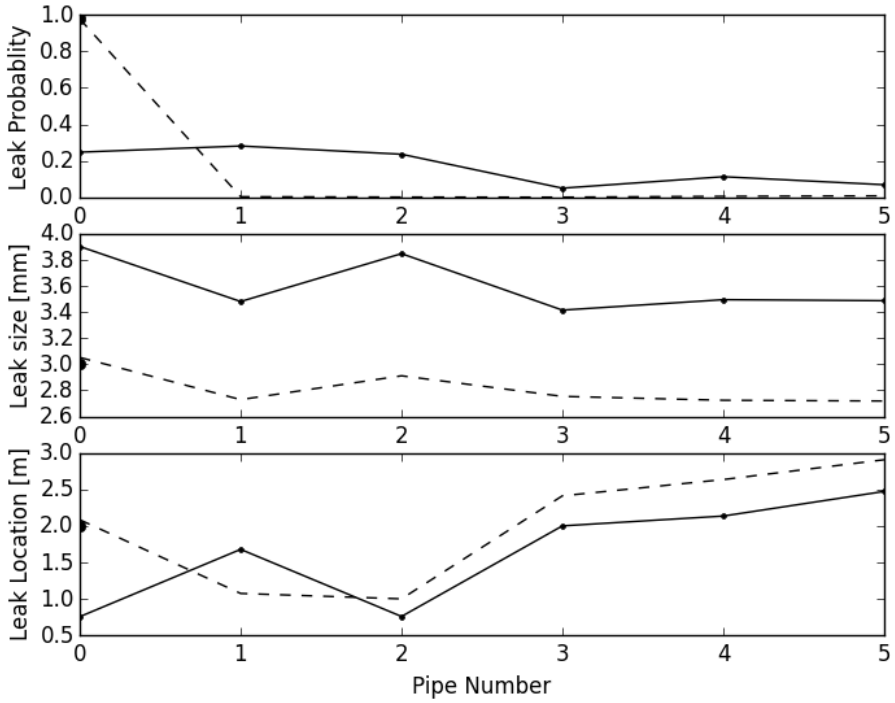


Figure 4.20: Network 2 Case A: Support Vector Machine solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

For leak case B, a leak probability of 28% was found for the first and second pipe, with a leak probability of 26% for the actual leaking pipe. The leak size was estimated to be 1.9mm, with the leak location at $-0.5m$.

The probability calculated for leak case C gave the second pipe a leak probability of 35%, with the actual leaking pipe a probability of 22%. The leak size was estimated as 1.76mm and the location of the leak at 1.5m.

Artificial Neural Network

The solutions for leak case A on network 2 by the Artificial Neural Network can be seen in Figure 4.21. The results for all the leak cases can be seen in Appendix C, Figures C.22 to C.24. For all the leak cases, it can be seen that the correct classification could be given for the numerical data. For leak case A, the classification was identical to that of the numerical case. The leak size was estimated to be 5mm with the location at $-3.7m$.

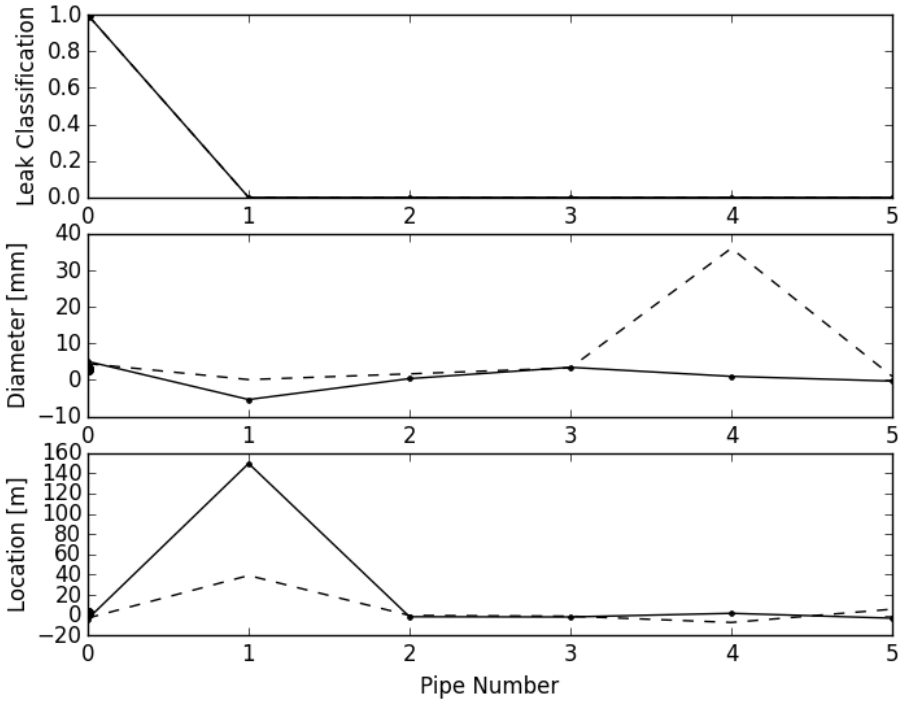


Figure 4.21: Network 2 Case A: Artificial Neural Network solution obtained when estimating the numerical sensor data (dashed lines) and the experimental sensor data (solid lines).

For leak case B, the classification of the leak was calculated to be at the sixth pipe. The leak size was estimated to be $1.92mm$, with the leak location at $0.16m$, resulting in an error of $1.34m$. For leak case C no classification of the leak was given. The leak size was estimated to be $1.3mm$ and the leak location was estimated to be at $0.3m$.

Discussion of Results

For network 1, all the leaking pipes and sizes could be estimated accurately by the algorithms except for the Artificial Neural Network. The Support Vector Machine could locate the leaking pipe with a probability of 100% for all three the leaks. The exact locations of the leaking pipes could be estimated anywhere with an error between $0.5m$ to $2m$ for all the leaking cases by all the algorithms.

For network 2, all the algorithms could find the solution for leaking case A with relative high probability. The Bayesian Probabilistic Analysis could locate this leak with a probability of 40%. The SVM found three possible locations with a probability of about 25% each. Similar leak probabilities were found by the Bayesian Probabilistic Analysis and the SVM for leak case B. For leak case C, only the SVM could find a relative solution with a probability of 35% for a leak on two different pipes. The MSE optimization, Bayesian Probabilistic Analysis and SVM could find the location of the leak perfectly for leak case C.

From all these results, it could be seen that the MSE optimization, Bayesian Probabilistic Analysis and the SVM could find the leaking pipes with good accuracy, with the SVM

preforming the best to locate the leaking pipe. The best leak location estimations on each pipe was found by the MSE optimization and the Bayesian Probabilistic Analysis. Table 4.5 shows a summary of the results found for the different tested networks with their individual leak cases.

Table 4.5: Summary of the results for the three leak locations on network 1 and 2.

	Leak Case A on Network 1 (size of $3mm$ and location of $2m$)		
Solution Strategy	Probability[%]	Leak Size [mm]	Leak Location [m]
Mean Squared Error Optimization	–	3.2	3
Bayesian Probabilistic Analysis	52	3.2	2.95
Support Vector Machine	100	3.4	3
Artificial Neural Network	–	2.6	2.4
	Leak Case B on Network 1 (size of $2mm$ and location of $1m$)		
Mean Squared Error Optimization	–	1.35	3
Bayesian Probabilistic Analysis	50	1.35	2.95
Support Vector Machine	100	1.38	2.7
Artificial Neural Network	–	1.3	2.1
	Leak Case C on Network 1 (size of $2mm$ and location of $1.5m$)		
Mean Squared Error Optimization	–	1.3	1.1
Bayesian Probabilistic Analysis	55	1.32	1.1
Support Vector Machine	100	1.3	2
Artificial Neural Network	–	4.2	-0.3
	Leak Case A on Network 2 (size of $3mm$ and location of $2m$)		
Mean Squared Error Optimization	–	8.8	0
Bayesian Probabilistic Analysis	43	8	0
Support Vector Machine	25	3.9	0.7
Artificial Neural Network	–	5	-3.7
	Leak Case B on Network 2 (size of $2mm$ and location of $1m$)		
Mean Squared Error Optimization	–	1.77	0
Bayesian Probabilistic Analysis	37	1.77	0
Support Vector Machine	26	1.9	-0.5
Artificial Neural Network	–	1.92	0.16
	Leak Case C on Network 2 (size of $2mm$ and location of $1.5m$)		
Mean Squared Error Optimization	–	1.7	1.55
Bayesian Probabilistic Analysis	15	1.7	1.55
Support Vector Machine	22	1.76	1.5
Artificial Neural Network	–	1.3	0.3

4.6 Multiple Leak Detection

This section uses the Support Vector Machine to classify multiple leak cases for the first experimental network. The SVM was chosen since it performed well in the previous section, and it can easily be applied to multiple leak cases. Both ideal and actual measurements are tested. The same parameters for the EPANET model were used as optimized in this chapter.

Data Set Generation

A data set was generated containing all the possible leak cases. The different leak cases were numbered from 1 to 6 for a single leaking pipe, for two leaking pipes, Table 4.6 shows the case numbering:

Table 4.6: Leak Class Numbers for Two Applied Leaks

		First Leak Location					
		P_1	$P2$	P_3	P_4	P_5	P_6
Second Leak case	P_1		7	8	9	10	11
	P_2	7		12	13	14	15
	P_3	8	12		16	17	18
	P_4	9	13	16		19	20
	P_5	10	14	17	19		21
	P_6	11	15	18	20	21	

The data set is generated by randomly selecting two leaking pipes, with a unique diameter and location for each. The simulated values are classified according to the given leak case number. Leak size and location estimations were left out for this section. The leak size could be estimated easily, while the leak location along a pipe proved too challenging. Only the probability for which pipe was estimated as the size could easily be estimated by the SVM in the previous sections. The leak location was left out since it was seen that the total length of $3m$ proved to be to large of a challenge for the algorithms to solve.

Results

The results for three leak cases after training the SVM can be seen in Figure 4.22. The figure consists of three graphs, with the first graph having a leak at the first and third pipe (leak class 8), the second having a leak at the first and fifth pipe (leak class 10) and the third having a leak at the third and fifth pipe (leak class 17).

In the first graph, the numerical solution can be seen with the dashed line. It can be seen that a 96% probability of a leak for the numerical leak case at the two leak locations was found. A leak probability of 20% was found for the experimental measurements for this case. Adjacent pipes to the third pipes gave leak probabilities of 17% and 27%. This shows a probability of 64% that there is a leak at the first pipe, which had a larger leak size than the other pipes.

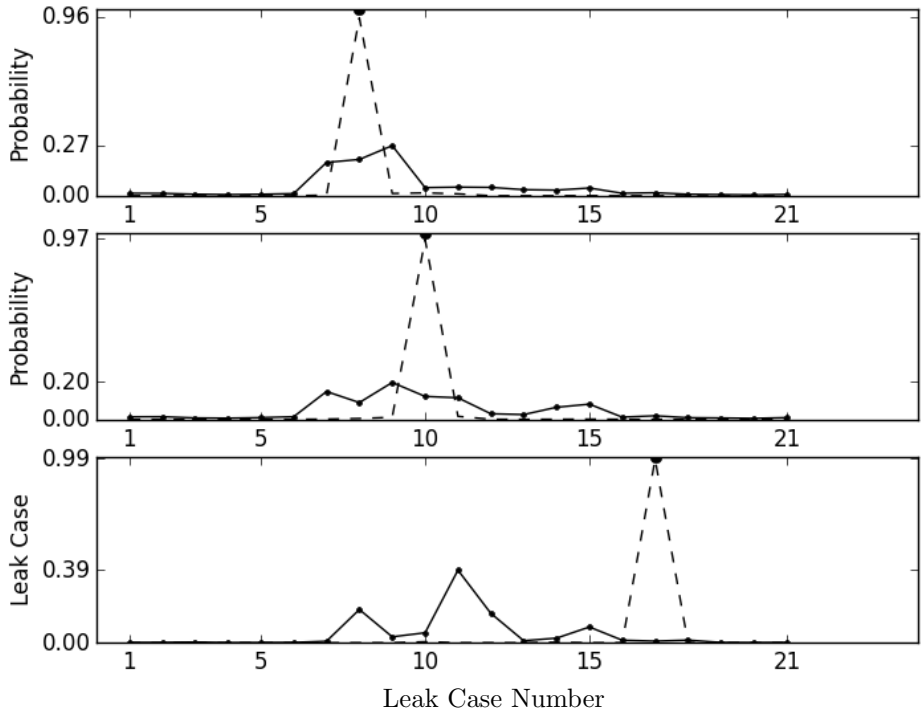


Figure 4.22: Solution found by the Support Vector Machine for Multiple leak cases

The second graph shows the solution for a leak at the first and fifth pipe. For the numerical case solution it could be seen that the correct leak locations were found with a probability of 97%. For the experimental data, the most probable leak was found at the first and fourth location, with a probability of 20%. The actual leak locations was given a probability of about 12%. This gives a leak probability for the actual locations and adjacent locations a total probability of 45%. With an 75% probability of a leak at the first pipe.

The third graphs shows the solution for a leak at the third and fifth pipe. The highest leak probability was found for the first and sixth pipe (leak class 11) to be leaking. The closest leak estimation was found for pipe two and six to be leaking (leak class 15), with a probability 8%. The correct leak location could be found with the numerical measurements with a 99% possibility.

Table 4.7: Summary of the results for the three multi-leak cases tested.

Leak Case Number	Probability Found
8	20%
10	12%
17	3%

4.7 Conclusion

In this chapter, the experimental network was planned and build. The network consisted of six pipes of $3m$ lengths with possible leaks. Twelve pressure sensors and seven flow sensors were added to the network, to which the EPANET model was calibrated for different leak cases.

The network could be changed between two configurations and had three possible leak locations with $3mm$ and $2mm$ diameter holes. It was found that the minor loss coefficients depend on the flow pattern and flow in the pipe. The pump efficiency was optimized and found to operate at an 82% efficiency. The error on the measurements and model was calculated between 2.6% to 44%.

For experimental network 1, it was found that all the algorithms could detect the correct leaking pipe except for the ANN. The SVM could find the leaking pipe with a 100% probability. The leak sizes could be estimated within $0.7mm$ for all the leak cases and algorithms. The leak locations could be found within $0.5m$ to $2m$ from the actual leak position.

For the second experimental network, only the Bayesian Probabilistic Analysis, the Mean Squared Error Optimization and Support Vector machine could find solutions for the first two leak cases. The ANN could find the leaking pipe for leak case A. The leak size could be estimated within $0.4mm$ for the second and third leak case, for the first leak case the size was estimated with an error up to $6mm$. The location of the leak on the length of pipe was estimated with an error between $0.5m$-$2m$.

From the experiment it could be seen that the Support Vector Machine and the Bayesian Probabilistic Analysis out performed the Artificial Neural Network. Taking the optimization and training time into account, the SVM outperformed all other algorithms. In this experiment, both the SVM and ANN found unique solutions for the leak location for each leak case.

Three multiple leak cases were tested for the first experimental network. The various leak cases were given classification numbers for the SVM to train on. With the actual measurements, a leak at the first and third pipe could be estimated with a probability of 20%, with adjacent pipes to the third pipe having higher probabilities. A leak at the first and fifth pipe, with adjacent pipes to pipe five, resulted in a leak probability of 45%. The third tested case of multiple leaks, at the third and fifth pipe, failed to result in accurate leak predictions.

To improve the results found in this chapter, additional minor loss coefficients can be added to be optimized. This will reduce the error between the simulated and measured pressure values. Additionally, adding more leak cases to be optimized will help create a more robust model of the experimental network. This chapter shows the possibility of estimating multiple leak locations. It can be seen that, when looking at the highest probabilities for these leak cases, the leaking pipe in common could be fixed first, to adjust the measurements for better estimations for the second leak.

Chapter 5

Optimum Sensor Locations for the Experimental Network

In this chapter, the support vector machine is trained on different combinations of sensors to find the best sensor combination for the experimental network measurements.

5.1 Optimization Algorithm

To find the best sensor placement locations, all the sensor locations are tested up to a maximum of four sensors used. This is based on an Exhaustive Search algorithm. Only four sensor locations are chosen due to the computational time. The algorithm consists of 4 nested loops, picking sensor locations. The SVM is trained to the numerical sensor data for the chosen locations, from which the best solution is chosen when applied to the experimental sensor measurements.

5.2 Optimization Results

The optimization was completed for experimental network 1 and 2. These sub-sections contain the best sensor placement locations for each leak of the three single leak cases tested.

Experimental Network 1

The best sensor placement locations for leak case A on experimental network 1 can be seen in Table 5.1. In this table it can be seen that, for a single sensor to be used, the best placement location will be the seventh pressure sensor. This sensor is located closest to the actual leak, at the end of the leaking pipe. It was calculated, that the SVM could locate the leak with a probability of 98.94% if only this sensor is used.

The optimum location for two sensors to be used was calculated to be the first and the seventh sensor. These sensors are the sensors before and after the leak, and will result in a leak probability of 99.72%. If three sensors are used, it was found that sensor 1, 3 and 11 must be used. The best locations if four sensors must be used was found to be sensor 1, 7,

Table 5.1: Sensor optimization for leak case A on experimental network 1

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_7	P_1	P_1	P_1
Sensor 2		P_7	P_3	P_7
Sensor 3			P_{11}	P_{10}
Sensor 4				P_{11}
Probability	98.94	99.72	99.93	99.97

10 and 11. The leak probability for these two optimizations was found to be above 99.9%.

Table 5.2 shows the best sensor locations for the second leak case. It can be seen that sensor used in all the optimization cases is the third pressure sensor. This sensor is located at the start of the leaking pipe. When two sensors are used, the eighth sensor is chosen, resulting in a leak probability of 99.04%. The addition of the 11th and 10th sensor comes next when three and four sensors are used. Table 5.3 shows the results for the third leak

Table 5.2: Sensor optimization for leak case B on experimental network 1

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_2	P_3	P_3	P_3
Sensor 2		P_8	P_8	P_8
Sensor 3			P_{11}	P_{10}
Sensor 4				P_{11}
Probability	25.75	99.04	99.99	99.86

case on the first experimental network. It can be seen that the fifth sensor is used for all four cases. This sensor is located at the start of the leaking pipe. The next optimal sensor to be added is the 11th sensor, which results in a leak probability of 99.41%. The 11th sensor is located at the end of the leaking pipe. Furthermore, it can be seen that the 9th, 10th and 12th sensors are also used. With this leak case, the addition of the sensors beyond 2 decreases the leak probability. This is due to additional uncertainty being added to the measurement signals for the SVM to classify on.

Table 5.3: Sensor optimization for leak case C on experimental network 1

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_5	P_5	P_5	P_5
Sensor 2		P_{11}	P_{11}	P_6
Sensor 2			P_{12}	P_9
Sensor 2				P_{10}
Probability	27.94	99.41	71.06	22.37

Second Experimental Network

The optimum sensor placements for first leak case on the second experimental network can be seen Table 5.4. For one sensor to be used, the 7th pressure sensor can be seen to result in a leak probability of 98.15%. When multiple sensors are allowed, sensor 7 is replaced by sensor 1 and sensor 11. Sensor 9,10 and 11 are used as additional sensors in these cases.

Table 5.4: Sensor optimization for leak case A on experimental network 2

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_7	P_1	P_1	P_1
Sensor 2		P_{11}	P_9	P_9
Sensor 3			P_{11}	P_{10}
Sensor 4				P_{11}
Probability	98.15	99.95	99.90	99.65

Table 5.5 shows the results for the second leak case. It can be seen that the best results for a single sensor to be used was the 12th sensor, resulting in a leak probability of 26.55%. The third sensor is used for two and more allowed sensors, which is the sensor before the leaking pipe. Sensor number 10, 11 and 12 is used for additional optimum sensors.

Table 5.5: Sensor optimization for leak case B on experimental network 2

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_{12}	P_3	P_3	P_3
Sensor 2		P_{10}	P_{10}	P_{10}
Sensor 3			P_{11}	P_{11}
Sensor 4				P_{12}
Probability	26.55	99.86	99.72	99.17

Table 5.6 shows the optimum sensors to be used for the third leak case. These sensor placement results shows similar results as seen in all previous leak cases, with the sensor at the start of the leaking pipe, followed by additional sensors 9, 10 and 12.

Table 5.6: Sensor optimization for leak case C on experimental network 2

Sensor	One Sensor	Two Sensor	Three Sensor	Four Sensor
Sensor 1	P_9	P_5	P_5	P_5
Sensor 2		P_{10}	P_9	P_9
Sensor 3			P_{10}	P_{10}
Sensor 4				P_{12}
Probability	20.17	99.99	99.99	99.99

5.3 Conclusion

In this chapter optimum sensor placements were investigated for the experimental networks. In previous chapters, the SVM could detect leaks for the first experimental network with high accuracies, and struggled to find leaks for the second experimental network.

For the first experimental network, it was seen that sensors placed before and after the leaking pipe results in high accuracies. Additional sensors include the 9th, 10th, 11th and 12th sensors which are placed behind the possible leaking nodes. This unfortunately results in the total of 12 sensors required for this network for the best possible results.

Similar results were seen for the second experimental network. Higher leak probabilities were found when only two sensors were used instead of the total of twelve as completed in the previous chapters. Similar to the first experimental network, it was found that the sensors before a leaking pipe and sensors 9, 10, 11 and 12 gives good results. Adding more sensor measurements for the second network added more uncertainty to the trained parameters, resulting in lower leak probabilities. With the observed optimum results, it can be concluded that for this network 7 sensors, one at the entry of each pipe and a sensor at the 10th location might be optimum. Additionally, measuring all the sensor locations and only training the SVM on two sensor measurements for each pipe could provide accurate results.

Chapter 6

Findings, Conclusion and Recommendations

6.1 Findings

From the 2011-2012 National Non-Revenue Water assessment it was seen that South Africa has a relatively high percentage of non-revenue water. The current techniques used proved to be inefficient due to pipe material restrictions and the requirement for pipe access.

The pressure-flow deviation method showed the most promise to be applied to water distribution and supply networks. This method is an inverse solution strategy, with multiple benefits such as on-line usage capabilities, capabilities to estimate leak size and location, no material restrictions and only uses a few pressure and flow sensors, which makes this method cost effective. Some disadvantages of this method is that it is a model bases solution strategy, resulting in new challenges such as model calibration. The Research into different pipe network modeling software showed EPANET as a candidate to use as a basis to model pipe networks.

Various authors have used EPANET with the pressure-flow deviation method with Support Vector Machines, Artificial Neural Networks and error minimization methods to find leaks in networks. A comparison between different solution strategies to solve leaking networks was investigated. The different solution strategies investigated are a mean squeared error minimization strategy, a Bayesian Probabalistic Analysis [2], a Support Vector Machine and a Artificial Neural Network to train to the network.

A comparison between these strategies showed that the tested error minimization method and the Bayesian Probabilistic Analysis can not find leaks in simple pipe networks, while the SVM and ANN could solve the simple pipe network accurately. With more complex networks it was seen that the SVM and ANN trained to the average location of the leaks in the data sets. The benchmark network which simulates a water distribution network showed that pressure sensors result in a more accurate leak identification, although flow measurements indicate the leak size more accurately.

The numerical investigation for the planned experimental network showed perfect classification of the leaking pipe for most of the algorithms, with problems finding the exact leak location on the main pipe. An allowable error estimation calculation showed that a worst case error larger than 2.6% on the experimental network could result in the leak being located on the wrong pipe. It was also seen that, reducing the number of possible leaking pipes from seven to six will not change the complicity of the solution.

The Experimental network was built, with three leak cases with diameters of $3mm$ and $2mm$. Sixteen calibration factors were used which included the roughness coefficient for all the pipes, the pump efficiency and minor loss coefficients. The minor losses were found to be much higher than the values in the literature, with larger values depending on the flow in the pipe. Model calibration indicated that the pump had an efficiency of $80 - 82\%$. The model calibration showed an error between the measurements and the model of $2.6 - 44\%$. The solution strategies could be applied to the leak case measurements with the calibrated EPANET model.

From the results found by the different strategies, the SVM showed the best accuracy to find the leaking pipe, with the error minimization methods having the best solution to the leak size and location on the pipe. The Neural Network performed the worst between all the tested algorithms.

The SVM was used with the same leak cases added together to find multiple leaks in the experimental network. For the two cases where the first leak location was combined, one leak could be located with a probability higher than 80%.

Sensor usage location optimization was completed on the twelve pressure sensors in the experimental network. It was seen that the leak could be located on a pipe with a probability of 98% or higher using only two sensors, before and after the leaking pipe. Additional sensors included the 9th, 10th and 11th pressure sensor in most cases. This showed choosing sensors to be used could result in better results than using all the sensors.

6.2 Conclusion

The pressure-flow deviation method was used to find leaks in various water distribution and supply networks. This method showed to have more advantages above other techniques, such as online monitoring of networks and cost effectiveness. The major disadvantage of this method includes the calibration of a numerical model. The four solution strategies used could find leaks in all the networks tested. The algorithms used were the Mean Squared Error minimization, the Bayesian Probabilistic Analysis, a Support Vector Machine and an Artificial Neural network.

Overall the Bayesian Probabilistic Analysis and the Mean Squared Error optimization resulted with the same results, with the Bayesian Probabilistic Analysis giving clearer distributions of leak probabilities between possible leaking pipes. The Support Vector Machine performed similar with estimating leak size and location, with clearer probabilities to find

the leaking main pipe. The Artificial Neural network could give perfect solutions for the leaking pipe if there was no noise in the measurements, with pipe boundary restriction problems. Problems with the error minimization methods were found when a few observable were available. The SVM and ANN had problems training to the averages of the data sets when complexity of the problems increased.

The pressure-flow deviation method shows significant application in leak detection. Practical implementations could consist of water transportation networks, distribution networks and agricultural networks. Combining this method with on-line monitoring systems could allow quick and accurate detection of leak size and location. Limitations of the pressure-flow deviation method include the calibration of the numerical model and estimation of fluctuating demands and supplies.

6.3 Recommendations for Future Work

In chapter 3 a calculation to find the minimum allowable error was completed. This error was calculated as 2.6%, although a leak could be located with an error of 44%. The calculation of this error gives a good baseline of the allowed uncertainties within the model and the network and could prove important for the application of this method in actual networks.

In chapter 4 the numerical model was calibrated to fit to the experimental measurements. With sixteen parameters optimized an error between the model and measurements between 2.6-44% was found. This can be improved by adding additional measurements, or fixing some parameters like the roughness coefficients and pump efficiencies.

Multiple leak cases were tested with the SVM within this research. The results found can be greatly improved by re-calibrating all the model parameters with the additional leak cases. Combining all the solutions with optimum sensor placements and usages as found in chapter 5, can increase leakage detection probabilities. It was seen that better results could be found when only two sensors were used instead of all twelve.

References

[1] R. Mckenzie, ZN. Siqalaba, and WA. Wegelin. The state of non-revenue water in south africa. Technical report, Water Research Commission, August 2012.

[2] Z. Poulakis, D. Valougeorgis, and C. Papadimitriou. Leakage detection in water pipe networks using a bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 18(4):315–327, July 2003.

[3] R. Mckenzie, ZN. Siqalaba, and WA. Wegelin. The state of non-revenue water in south africa. *The Water Wheel*, 12(1):15–18, January 2013.

[4] Wang X.J., Simpson A.R., Lambert M.F., and Vitovsky J.P. Leak detection in pipeline systems using hydraulic methods: a review. *Conference on Hydraulics in Civil Engineering*, (23-30):391–400, November 2001.

[5] Z. Zangenhmadar and O. Moselhi. Study of leak detection technologies in water distribution networks. *CSCE 2014 General Conference*, May 2014.

[6] L.A. Rossman. *EPANET 2 Users Manual*. United States Environmental Protection Agency, September 2000.

[7] J.E. van Zyl and C.R Clayton. The effect of pressure on leakage in water distribution systems. *Water Management*, 160(WM2):109–114, 1 2007.

[8] A.C. Caputo and P.M. Palegagge. Using neural networks to monitor piping systems. *Process Safety Progress*, 22(2):119–127, June 2003.

[9] S.R. Mounce and J. Machell. Burst detection using hydraulic data from water distribution systems with artificial neural networks. *Urban Water Journal*, 3(1):21–31, February 2007.

[10] A.E.U Salam, M. Tola, M. Selintung, and F. Maricar. On-line monitoring system of water leakgage detection in pipe networs with artificial intelligence. *ARPN Journal of Engineering and Applied Science*, 9(10):1817–1822, October 2014.

[11] D. De Silva, J. Mashford, and S. Burn. Computer aided leak location and sizing in pipe networks. Technical Report 17, Urban Water Security Research Alliance, April 2011.

[12] J.P. Vitovsky, A.R. Simpson, and M.F. Lambert. Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management*, pages 262–265, July 2000.

[13] A. Nasirian, M.F. Maghrebi, and S. Yazdani. Leakage detection in water distribution network based on a new heuristic genetic algorithm model. *Journal of Water Resource and Protection*, 5:294–303, March 2013.

[14] D.Y. Choi, S.W. Kim, M.A. Choi, and Z. W. Geem. Adaptive kalman filter based on adjustable sampling interval in burst detection for water distribution system. *Water*, 8(4):142, 2016.

[15] I. Okeya, Z. Kapelan, C. Hutton, and D. Naga. Online burst detection in water distribution systems using the kalman filter and hydraulic modelling. *Procedia Engineering*, 89(16):418–427, July 2014.

[16] I. Okeya, Z. Kapelan, C. Hutton, and D. Naga. Online modelling of water distribution system using data assimilation. *Procedia Engineering*, 70(12):1261–1270, September 2014.

[17] D. Jung and K Lansey. Burst detection in water distribution system using the extended kalman filter. *Procedia Engineering*, 70(12):902–906, September 2014.

[18] H.R. Asgari and M.F. Maghrebi. Application of nodal pressure measurements in leak detection. *Flow Measurement and Instrumentation*, 50:128–134, June 2016.

[19] P. Cuguero-Escofet, J. Blesa, R. Perez, M.A. Cuguero-Escofet, and G. Sanz. Assessment of a leak localization algorithm in water networks under demand uncertainty. *IFAC-PapersOnLine*, 48(21):226–231, 2015.

[20] T.T.T. Zan, H.B. Lim, K.J. Wong, A.J. Wittle, and B.S. Lee. Event detection and localization in urban water distribution network. *IEEE Sensors Journal*, 14(12):4134–4142, December 2015.

[21] R. Perez, G. Sanz, V. Puig, J. Quevendo, M.A. Cugeuro Escofet, F. Nejjari, J. Meseguer, G. Cembrano, J.M. Mirats Tur, and R. Serrate. A model-based methodology using pressure sensors applied to a real network in barcelona. *IEEE Control Systems Magazine*, pages 24–36, August 2014.

[22] W. Gong, M.A. Suresh, L. Smith, A. Ostfeld, R. Stoleru, A. Rasekh, and M.K. Banks. Mobile sensor networks for optimal leak and backflow detection and localization in municipal water networks. *Environmental Modelling & Software*, 80:306–321, March 2016.

[23] B. Farley, S.R. Mounce, and J.B. Boxall. Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal*, 7(6):345–356, December 2010.

[24] A. Soldevila, S. Tornil-Sin, R.M. Fernandez-Canti, J. Blesa, and V.Puig. Optimal sensor placement for classifier-based leak localization in drinking water networks. *Conference on Control and Fault-Tolerant Systems*, 2016(3):325–330, November 2016.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[26] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Schnke, T. Rückstieß, and J. Schmidhuber. PyBrain. *Journal of Machine Learning Research*, 11:743–746, 2010.

[27] Google Maps, DigitalGlobe, Agricultural Site 23.8698866N, 29.5340912W, and elevation 1300m. https://www.google.co.za/maps/@-23.8698866,29.5340912,365m/data=!3m1!1e3?hl=en, viewed 28 November 2017.

[28] SALFLO. *SALFLO Water Pumps Product Catalogue and SALFLO and Edition 3 and May 2015*, 3 edition, 5 2015.

Appendix A

Second Experimental Network Numerical Results

This appendix contains figures for the numerical solutions for the second experimental network as calculated in Chapter 3. Figure A.1 shows the results for the two leak cases by the MSE optimization. Figure A.2 shows the solution for the two leak cases by the Bayesian Probabilistic Analysis. Figure A.3 and A.4 shows the results for the SVM and ANN for the two leak cases.

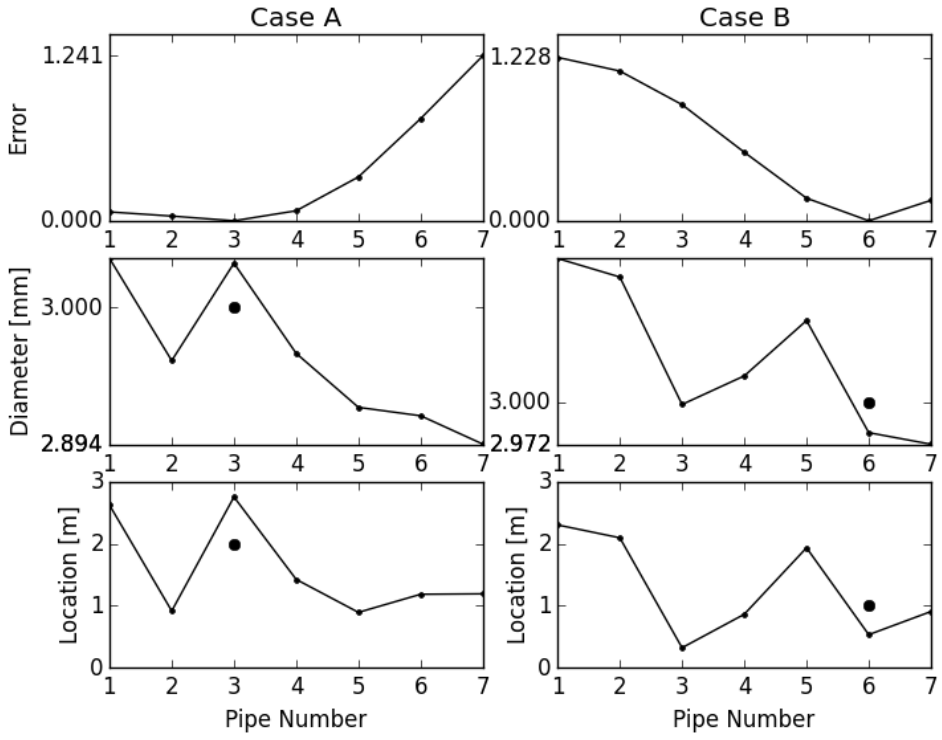


Figure A.1: Solution found by the Mean Squared Error Optimization for the Second Experimental Network

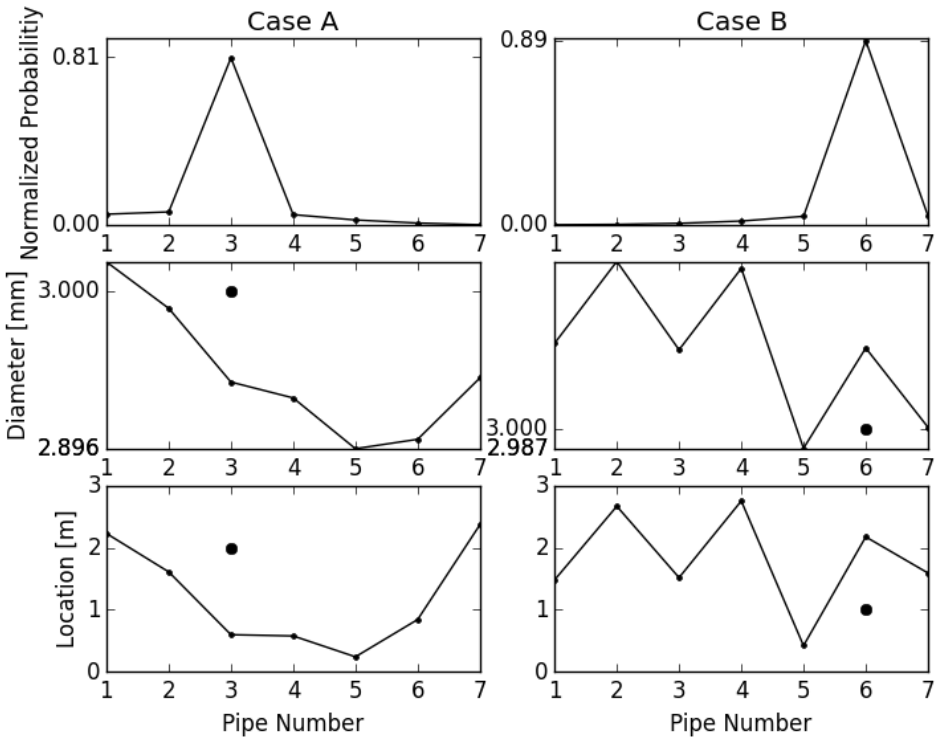


Figure A.2: Solution found by the Bayesian Probabilistic Analysis for the Second Experimental Network

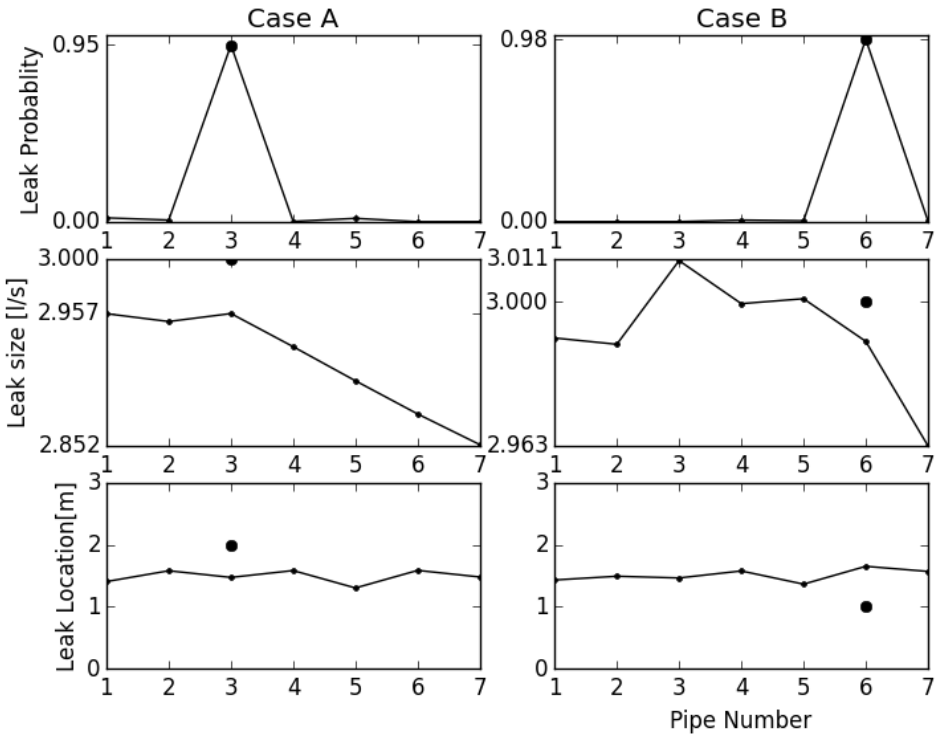


Figure A.3: Solution found by the Support Vector Machine for the Second Experimental Network

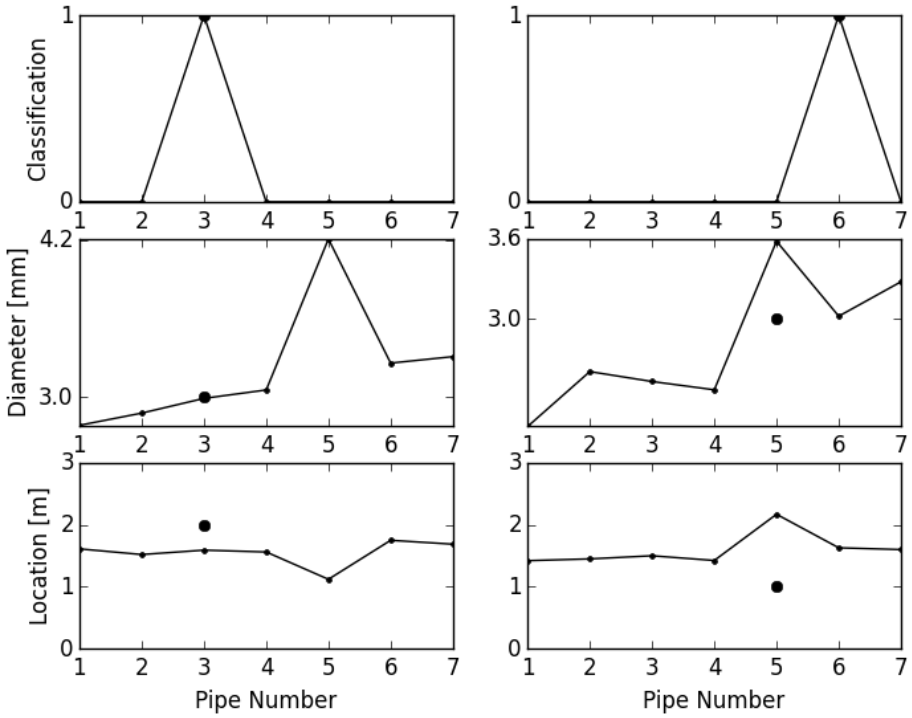


Figure A.4: Solution found by the Artificial Neural Network for the Second Experimental Network

Appendix B

Model Measurements After Calibration

This appendix contains the figures showing the experimental model calibration measurements. This appendix is referenced to from Chapter 4.

First Experimental Network

Figure B.1 shows the measurements and the simulated values before calibration with no leak in the network. Figure B.2 shows the measurements and simulated values after calibration with no leak in the network.
Figure B.3 to B.5 shows the measured and simulated values for the pressure and flows after calibration of the model was completed

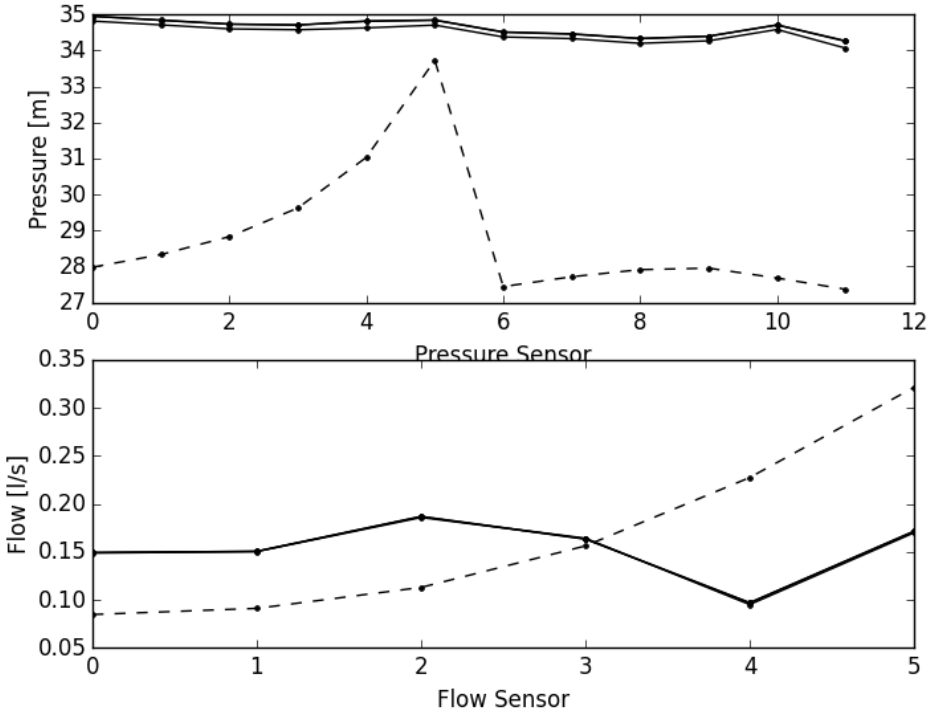


Figure B.1: Values Before Calibration of the First Experimental Network with no Leaks

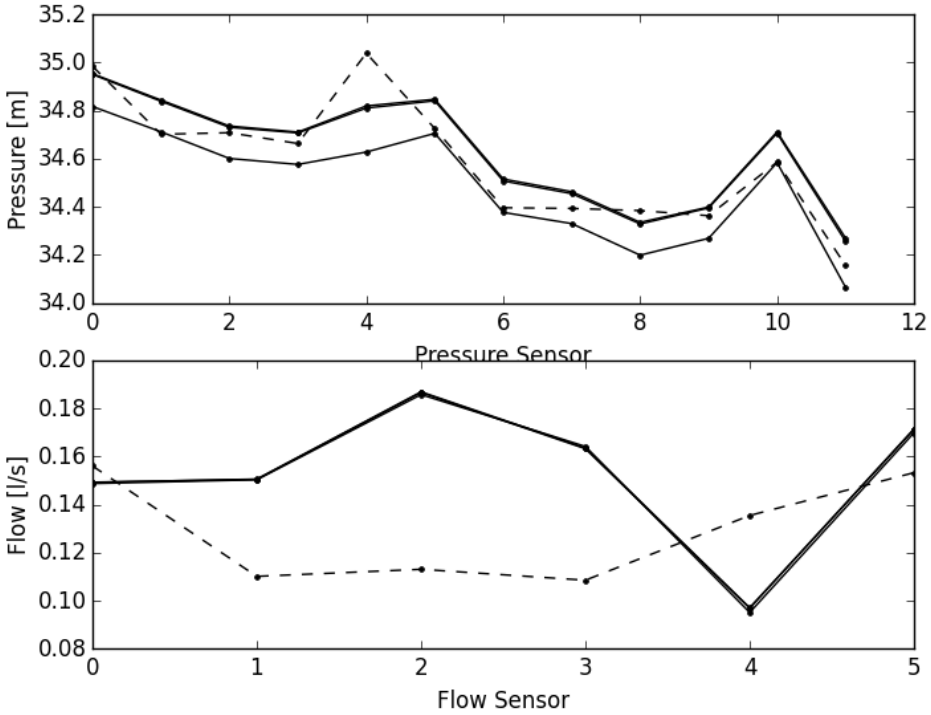


Figure B.2: Values After Calibration of the First Experimental Network with no Leaks

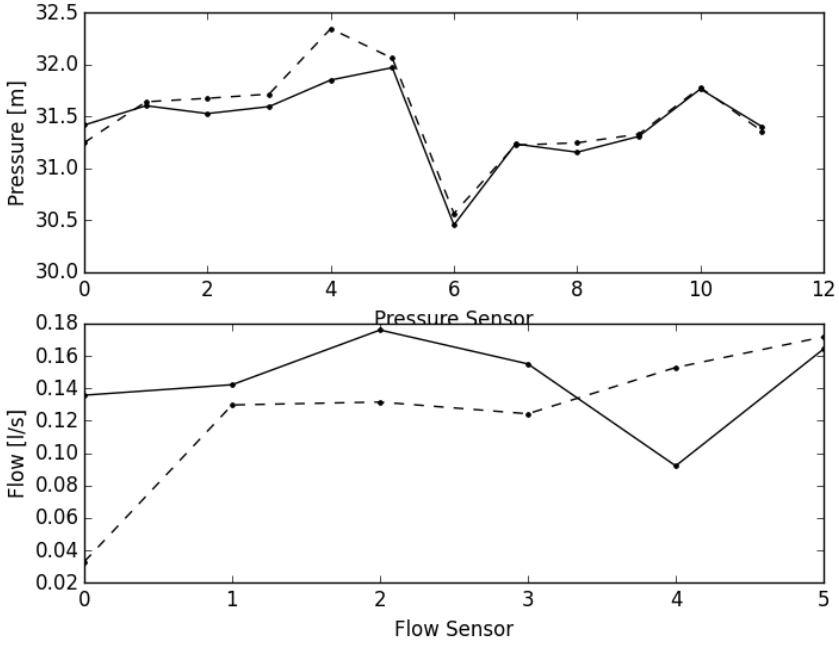


Figure B.3: Simulated and Measured parameters for the First Experimental Network for the First Leak Case After Calibration

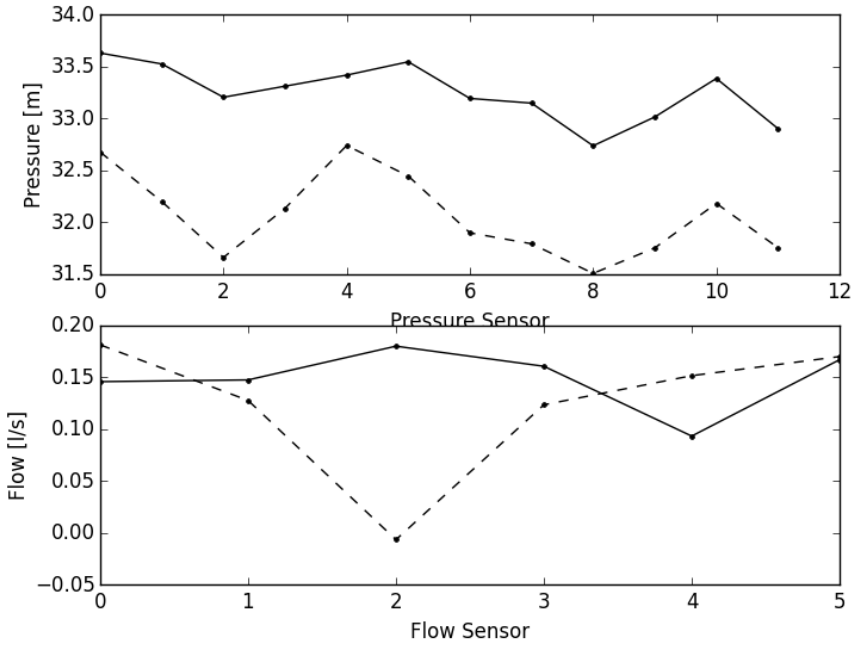


Figure B.4: Simulated and Measured parameters for the First Experimental Network for the Second Leak Case After Calibration

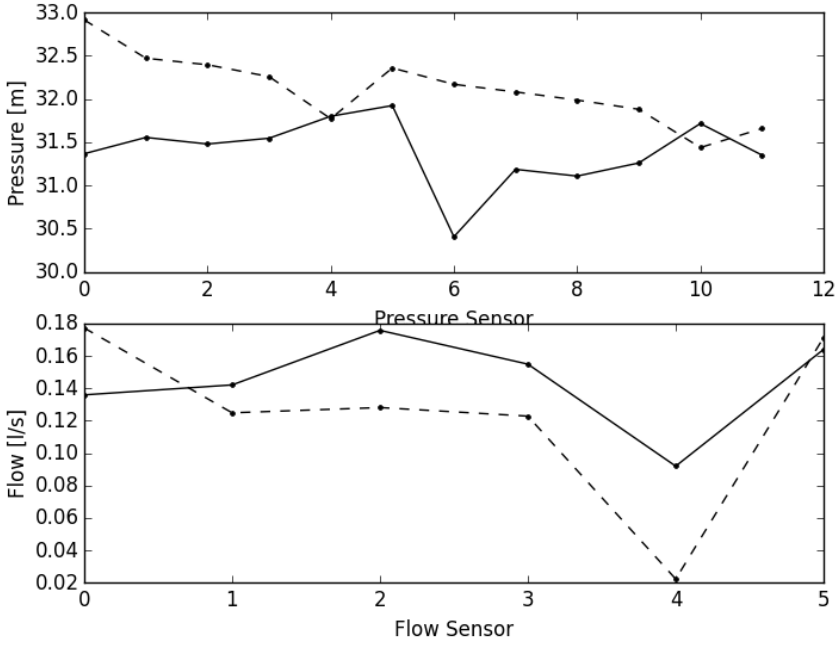


Figure B.5: Simulated and Measured parameters for the First Experimental Network for the Third Leak Case After Calibration

Second Experimental Network

Figure B.6 shows the measurements and the simulated values before calibration with no leak in the network. Figure B.7 shows the measurements and simulated values after calibration with no leak in the network.

Figure B.8 to B.10 shows the measured and simulated values for the pressure and flows after calibration of the model was completed.

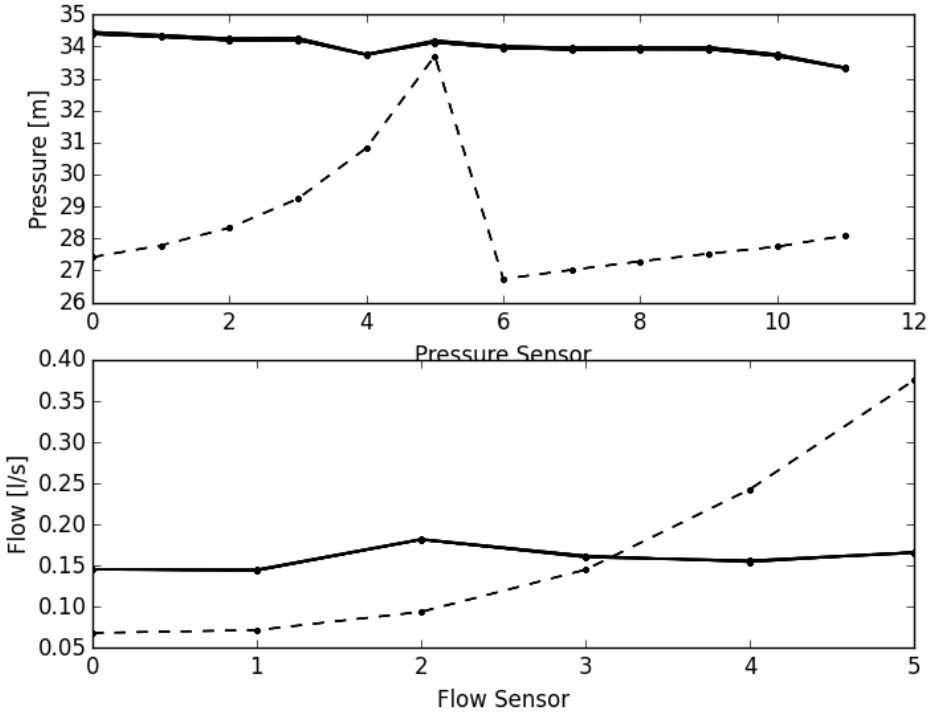


Figure B.6: Values Before Calibration of the First Experimental Network with no Leaks

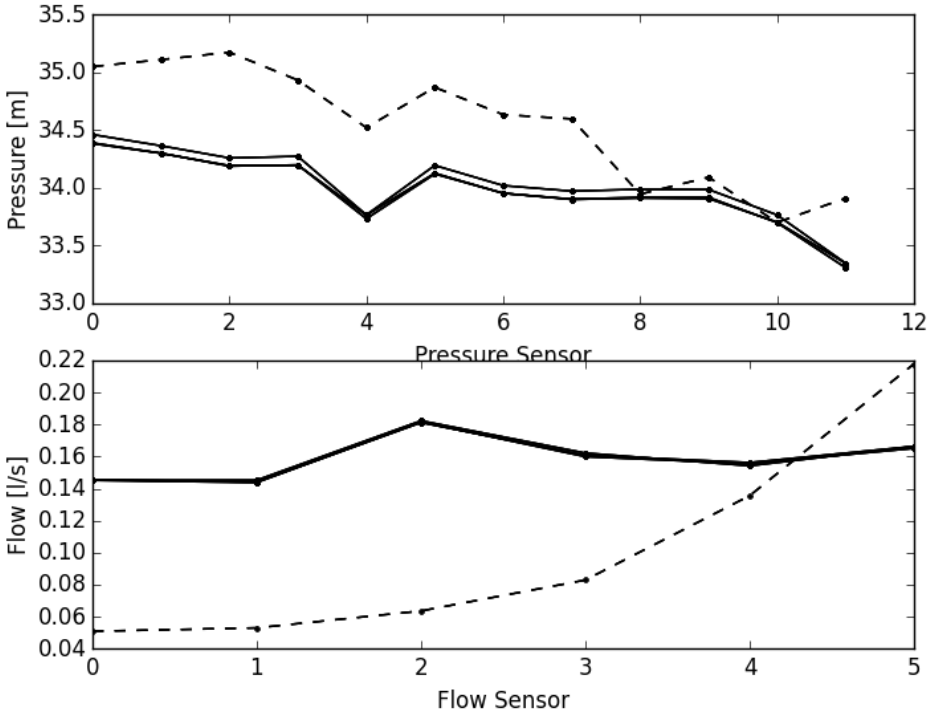


Figure B.7: Values After Calibration of the First Experimental Network with no Leaks

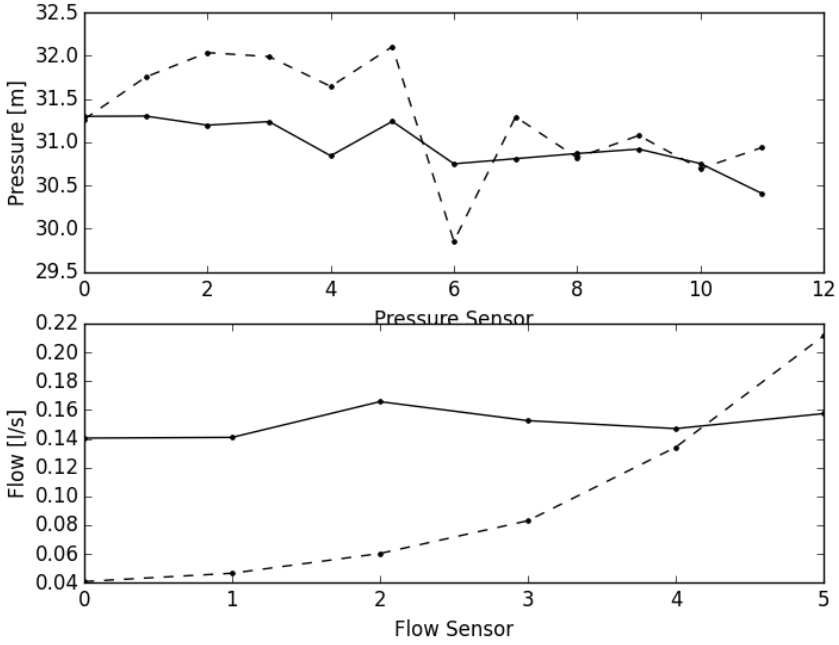


Figure B.8: Simulated and Measured parameters for the First Experimental Network for the First Leak Case After Calibration

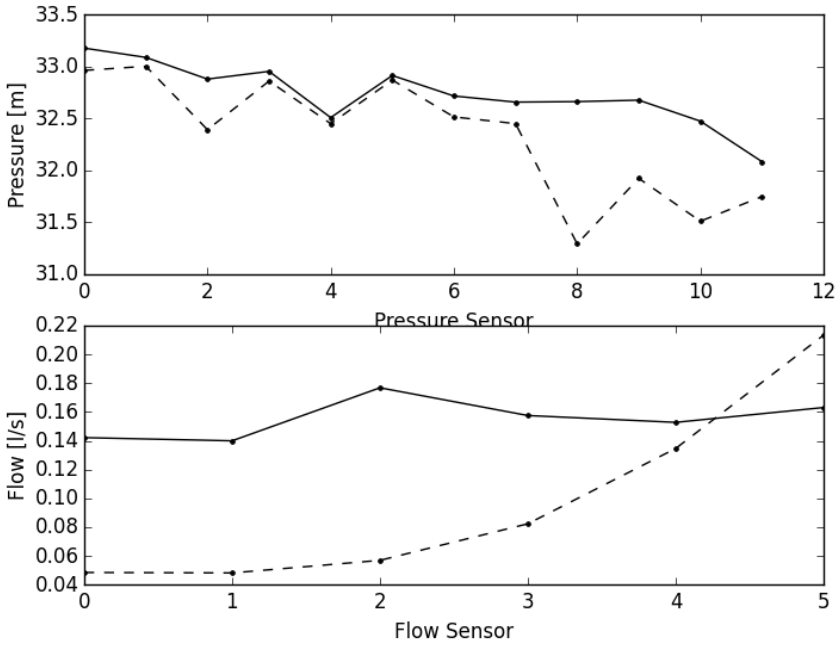


Figure B.9: Simulated and Measured parameters for the First Experimental Network for the Second Leak Case After Calibration

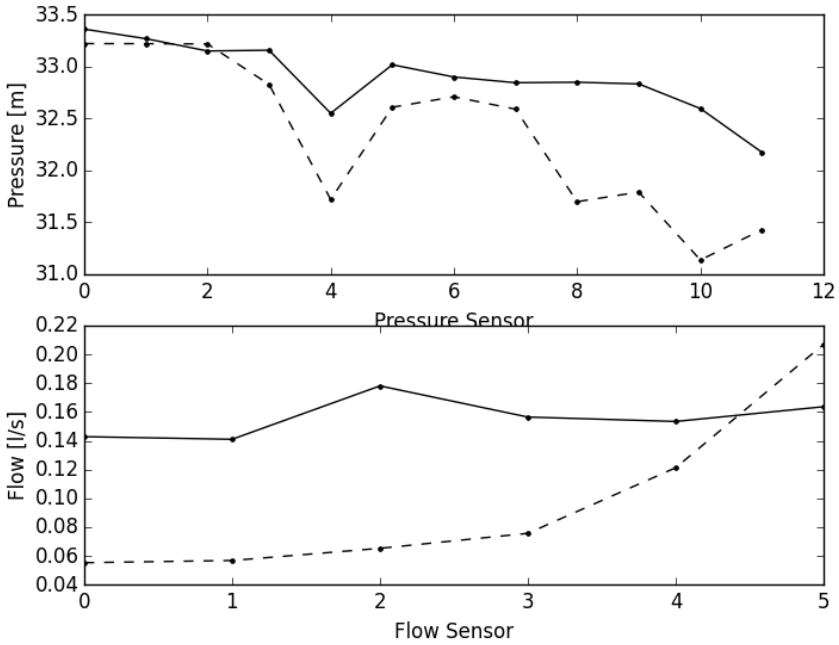


Figure B.10: Simulated and Measured parameters for the First Experimental Network for the Third Leak Case After Calibration

Appendix C

Experimental Results

This appendix contains large figure of the leak location results for the experimental investigation. The results are divided between the first and second experimental networks.

First Experimental Network

The results for the first experimental network are subdivided between each algorithm which contains the results for the three leak cases.

Mean Squared Error Optimization

Figure C.1 to C.3 shows the results found for the three leak cases by the mean squared error optimization.

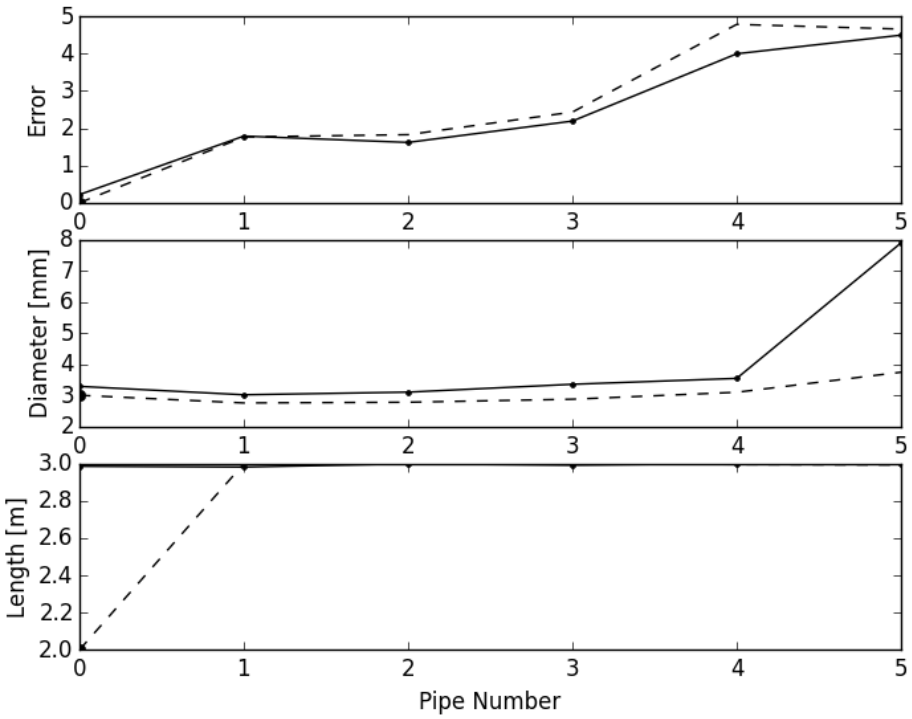


Figure C.1: Solution for the First Leak on the First Experimental Network by the Mean Squared Error Optimization

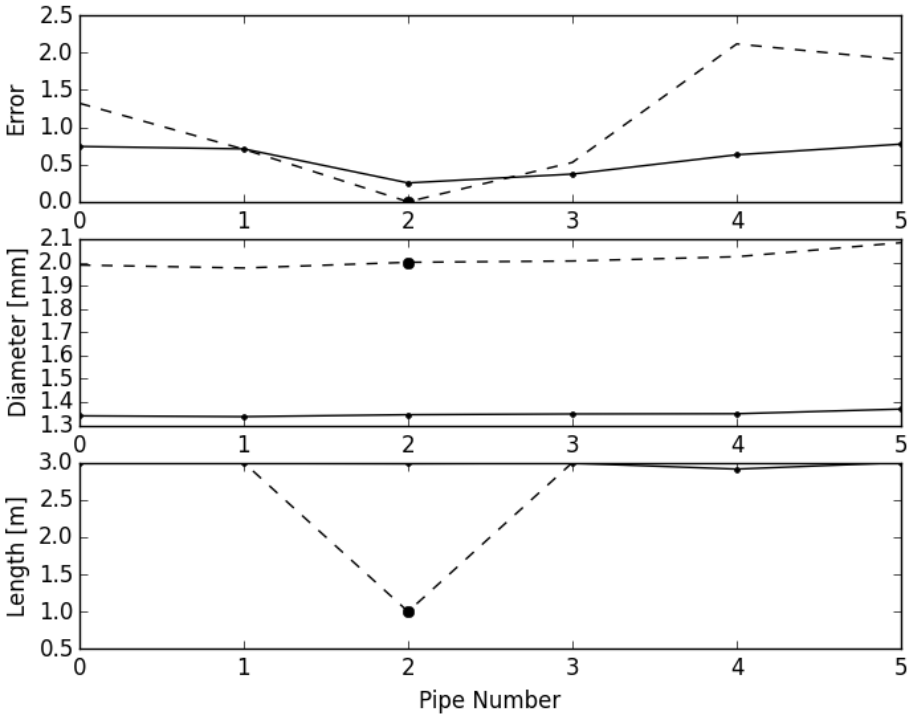


Figure C.2: Solution for the First Second on the First Experimental Network by the Mean Squared Error Optimization

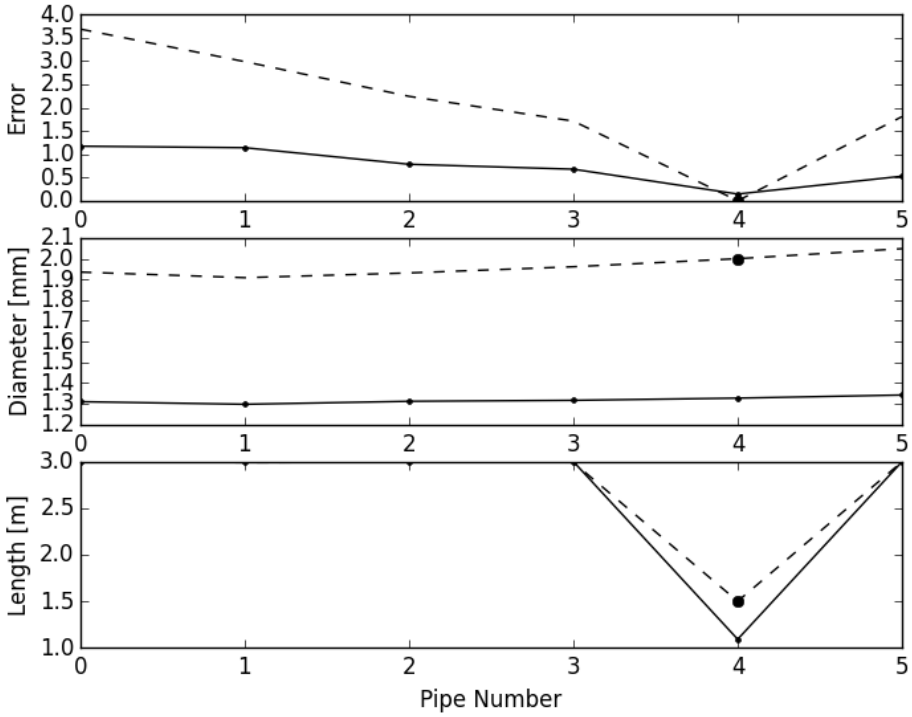


Figure C.3: Solution for the Third Leak on the First Experimental Network by the Mean Squared Error Optimization

Bayesian Probabilistic Analysis

Figure C.4 to C.6 shows the results found for the three leak cases by the Bayesian Probabilistic analysis.

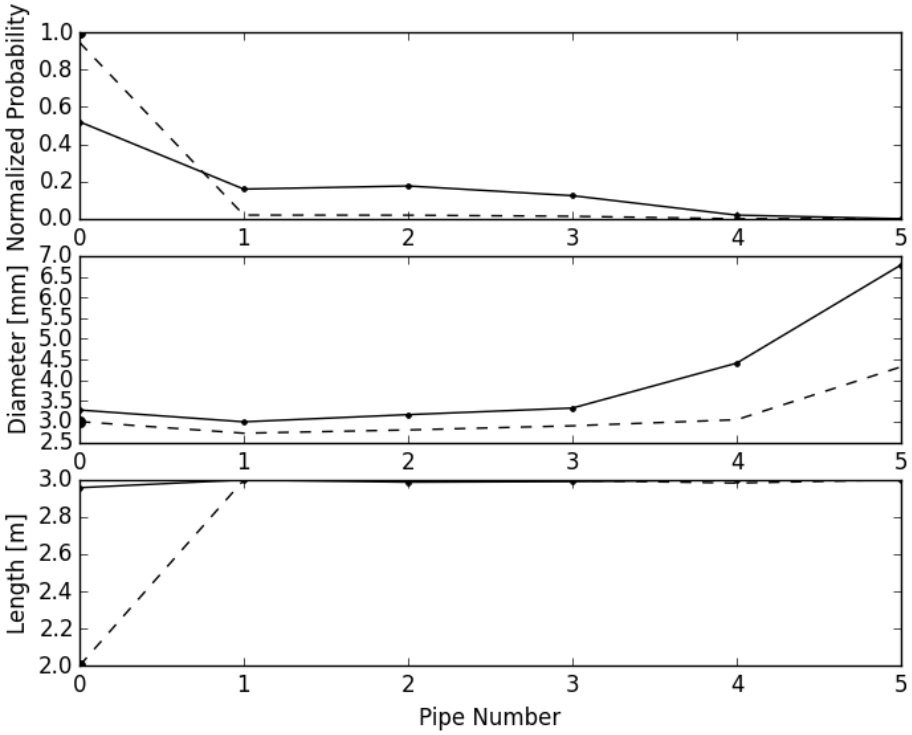


Figure C.4: Solution for the First Leak on the First Experimental Network by the Bayesian Probabilistic Analysis

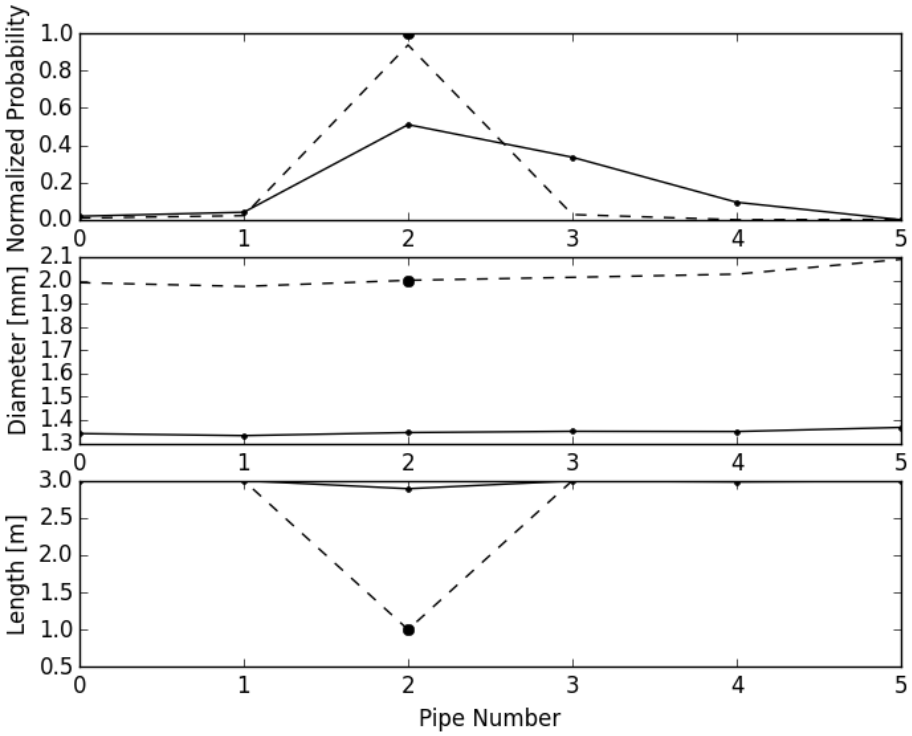


Figure C.5: Solution for the Second Leak on the First Experimental Network by the Bayesian Probabilistic Analysis

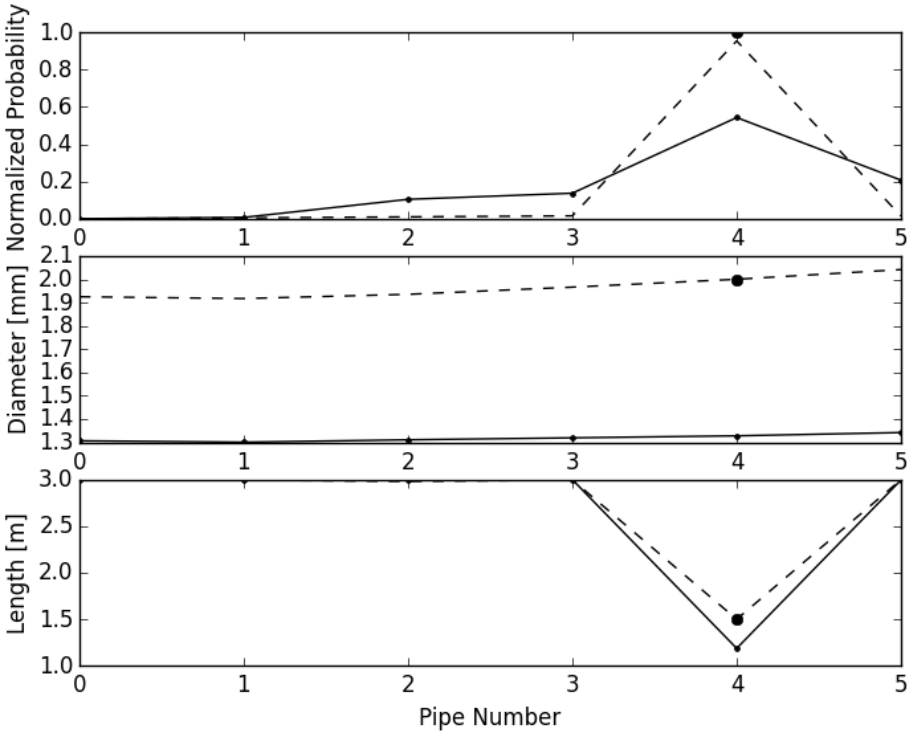


Figure C.6: Solution for the Third Leak on the First Experimental Network by the Bayesian Probabilistic Analysis

Support Vector Machine

Figure C.7 to C.9 shows the results found for the three leak cases by the Support Vector Machine.

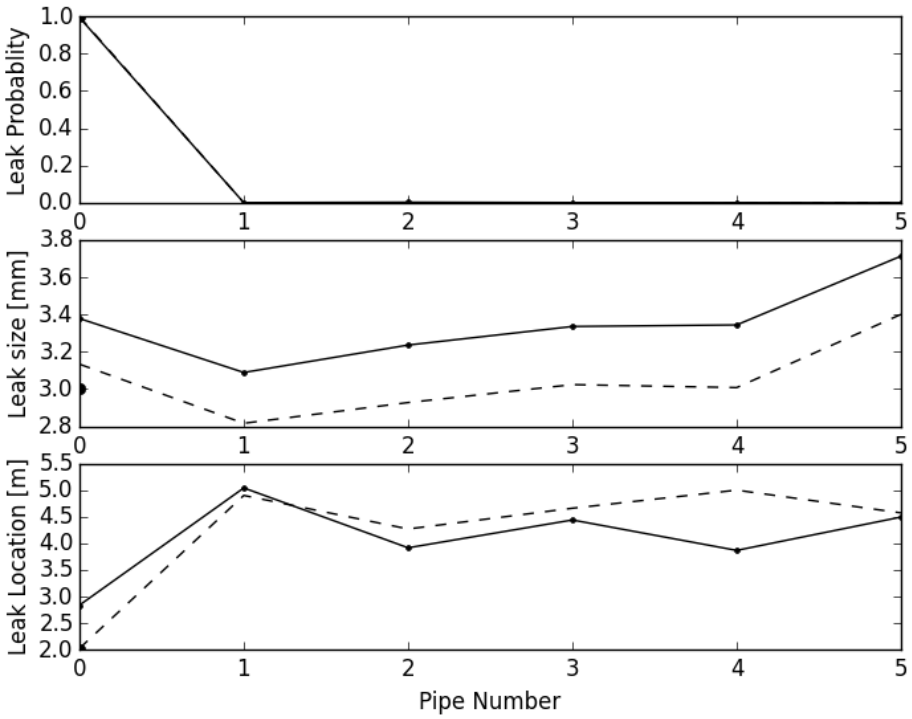


Figure C.7: Solution for the First Leak on the First Experimental Network by the Support Vector Machine

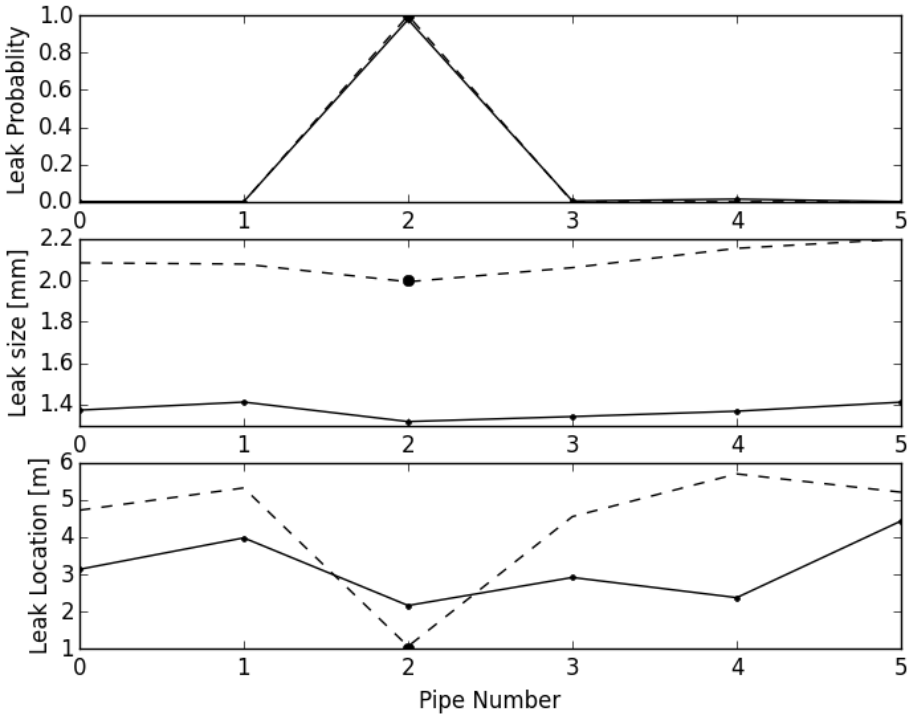


Figure C.8: Solution for the Second Leak on the First Experimental Network by the Support Vector Machine

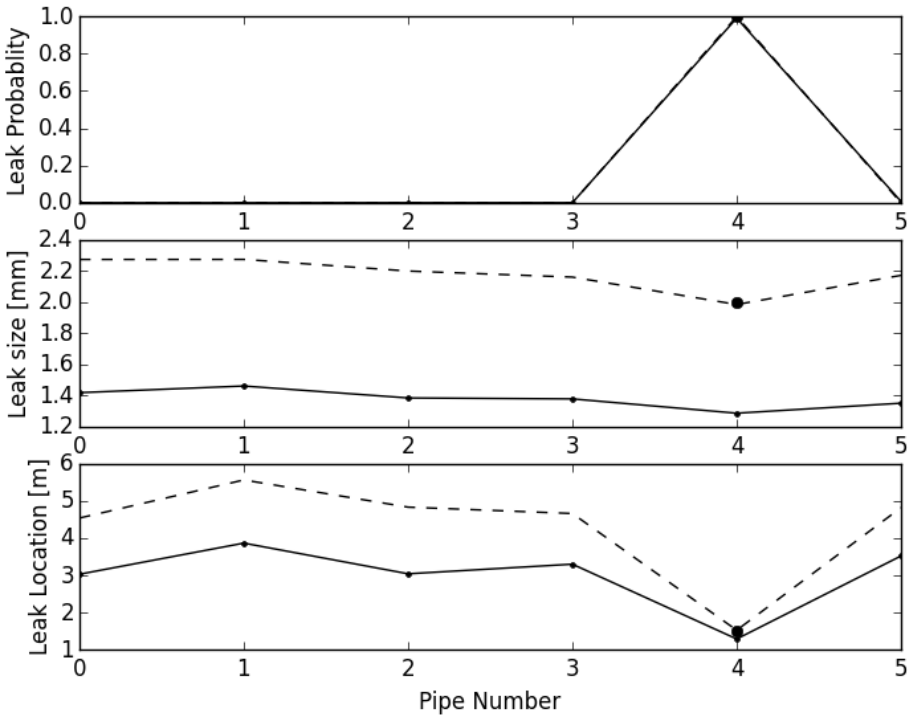


Figure C.9: Solution for the Third Leak on the First Experimental Network by the Support Vector Machine

Artificial Neural Network

Figure C.10 to C.12 shows the results found for the three leak cases by the Artificial Neural Network.

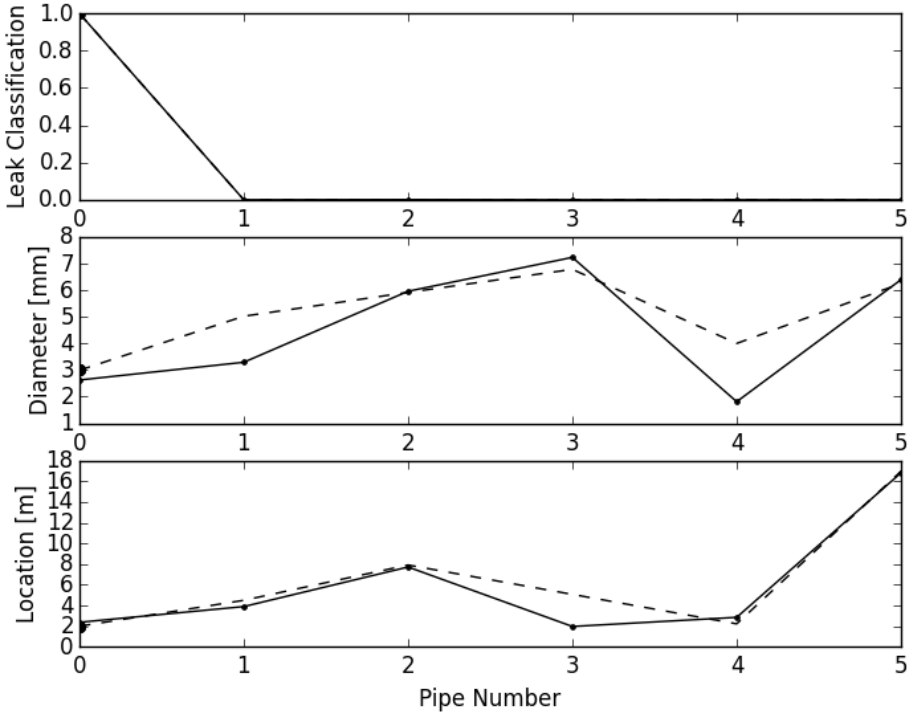


Figure C.10: Solution for the First Leak on the First Experimental Network by the Artificial Neural Network

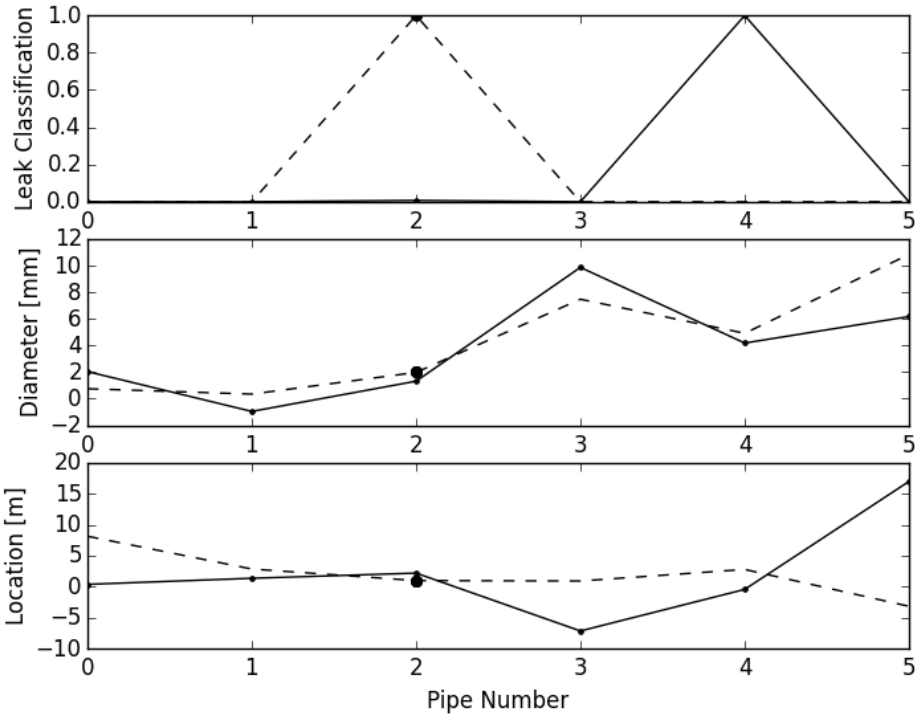


Figure C.11: Solution for the Second Leak on the First Experimental Network by the Artificial Neural Network

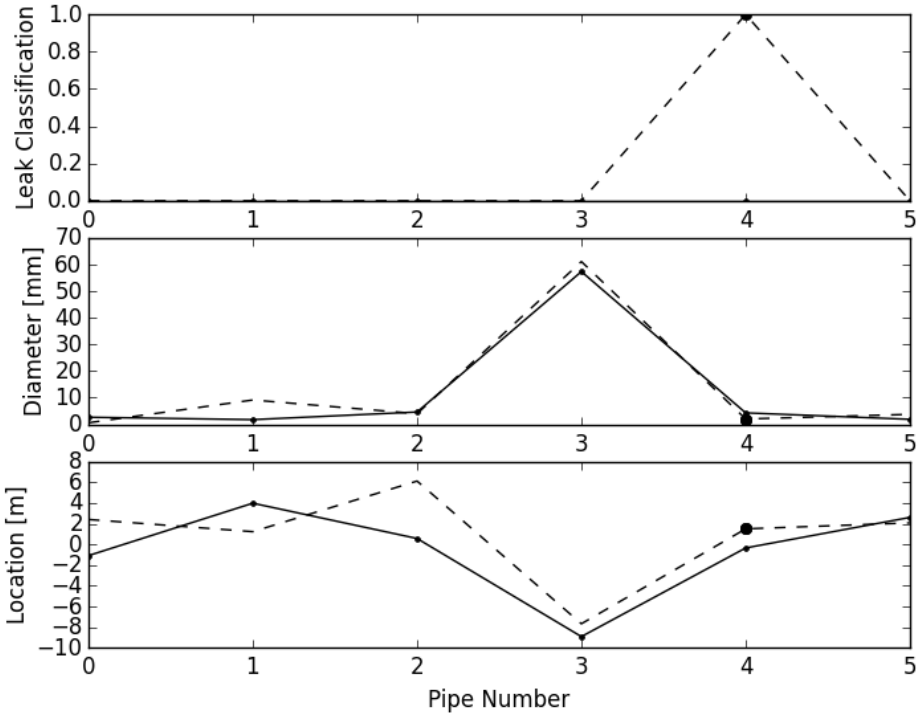


Figure C.12: Solution for the Third Leak on the First Experimental Network by the Artificial Neural Network

Second Experimental Network

The results for the Second experimental network are subdivided between each algorithm which contains the results for the three leak cases.

Mean Squared Error Optimization

Figure C.13 to C.15 shows the results found for the three leak cases by the mean squared error optimization.

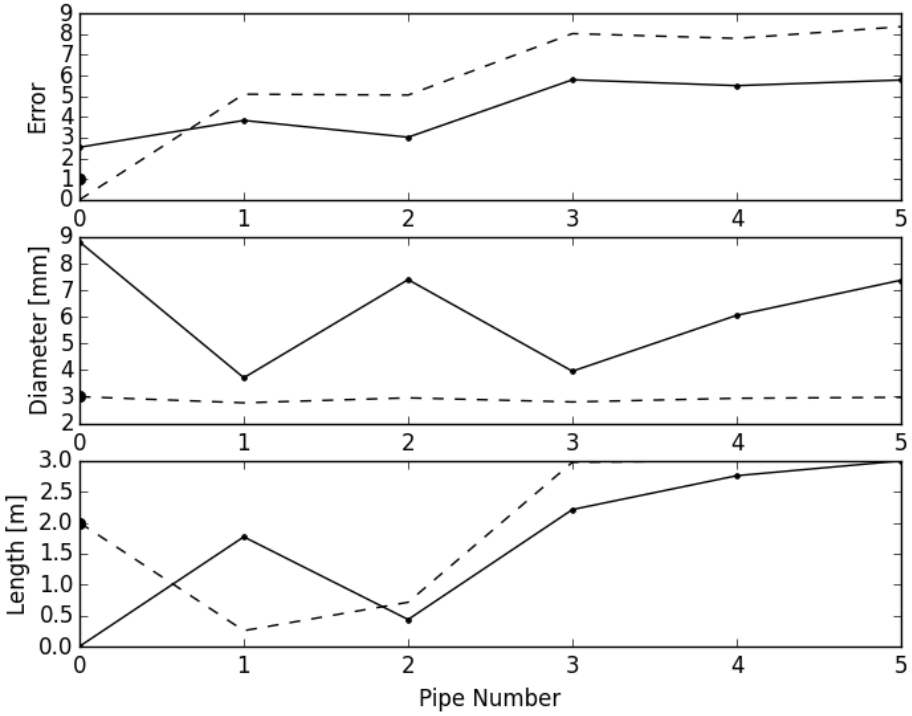


Figure C.13: Solution for the First Leak on the Second Experimental Network by the Mean Squared Error Optimization

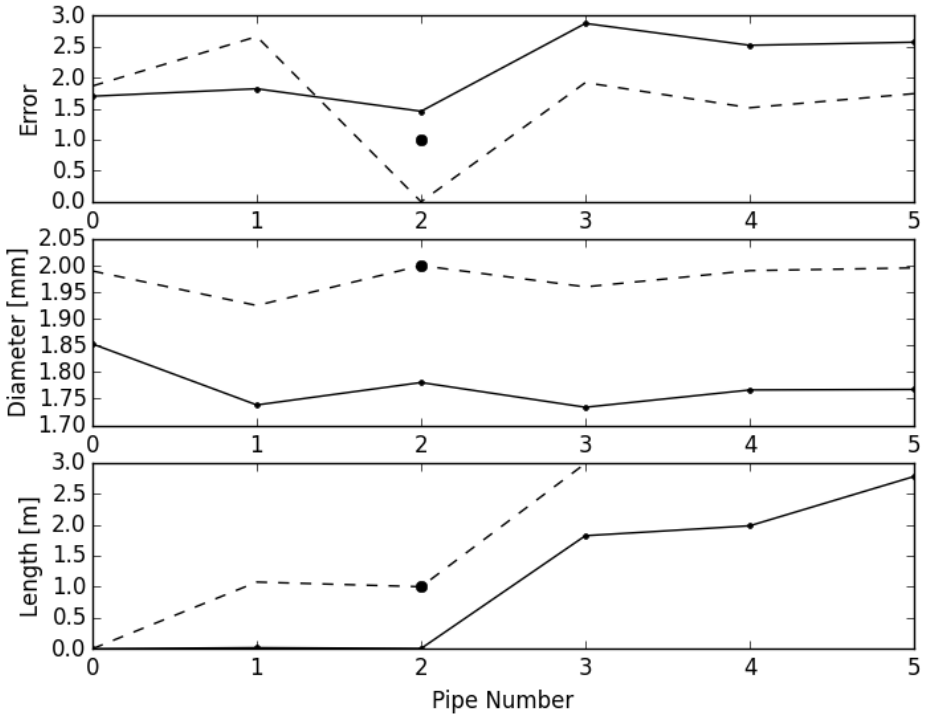


Figure C.14: Solution for the First Second on the Second Experimental Network by the Mean Squared Error Optimization

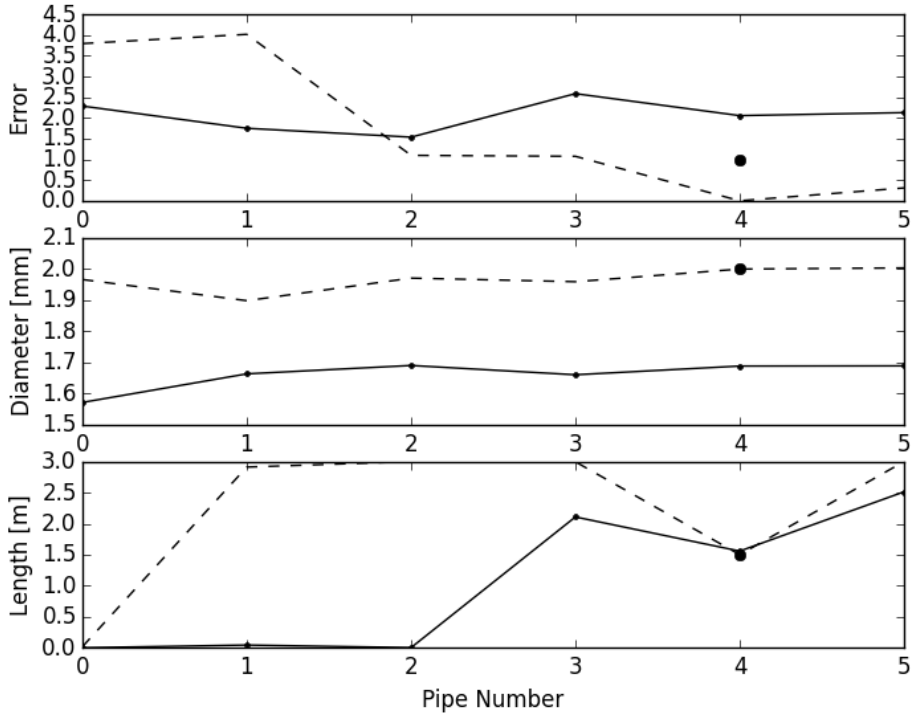


Figure C.15: Solution for the Third Leak on the Second Experimental Network by the Mean Squared Error Optimization

Bayesian Probabilistic Analysis

Figure C.16 to C.18 shows the results found for the three leak cases by the Bayesian Probabilistic analysis.

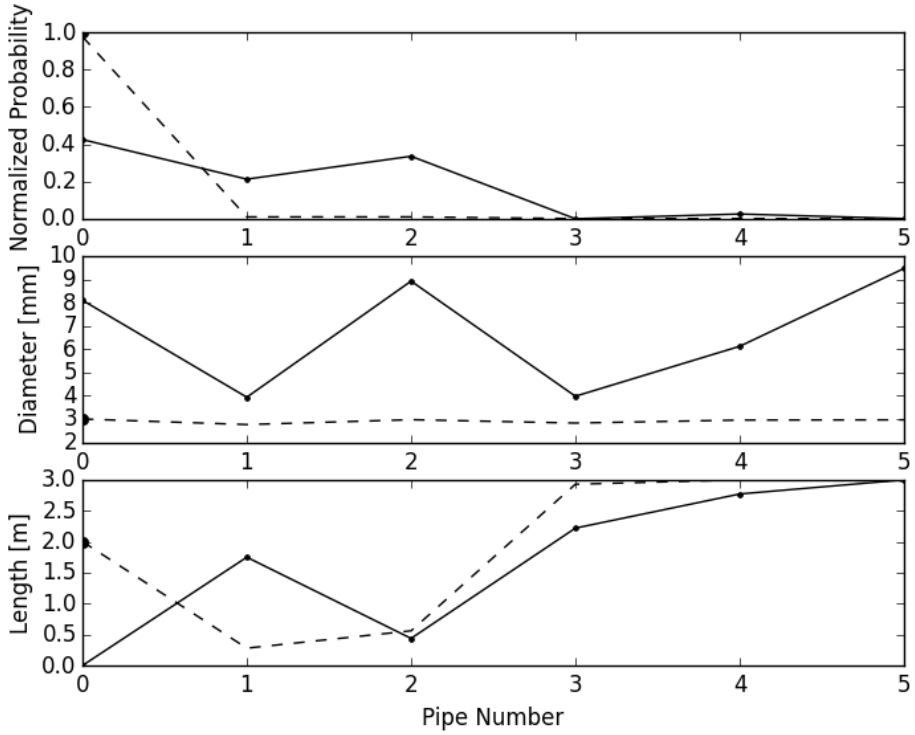


Figure C.16: Solution for the First Leak on the Second Experimental Network by the Bayesian Probabilistic Analysis

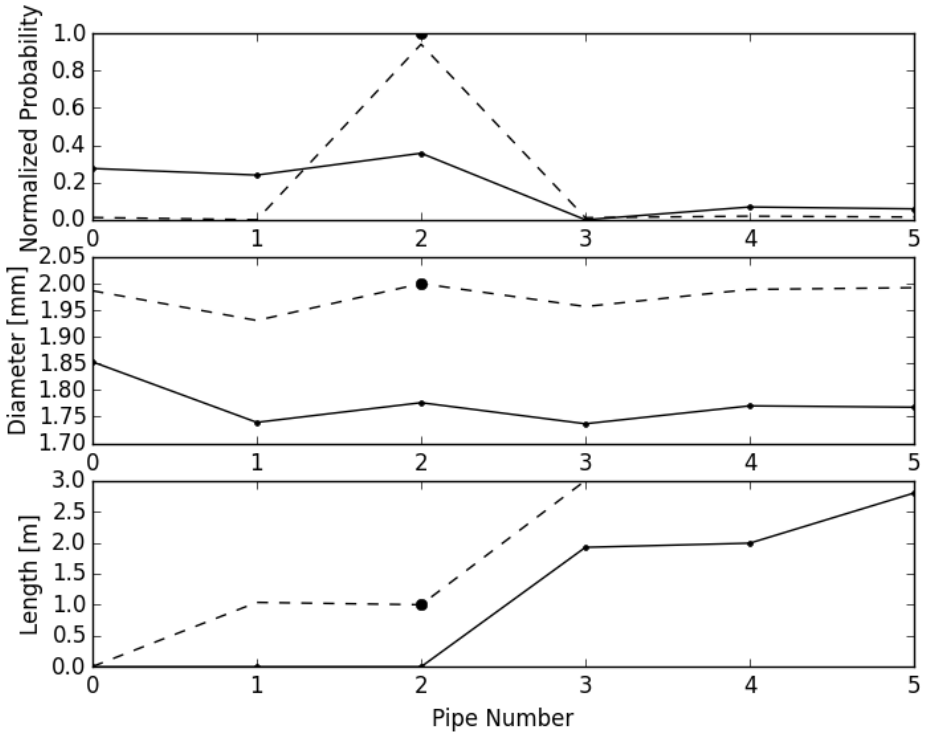


Figure C.17: Solution for the Second Leak on the Second Experimental Network by the Bayesian Probabilistic Analysis

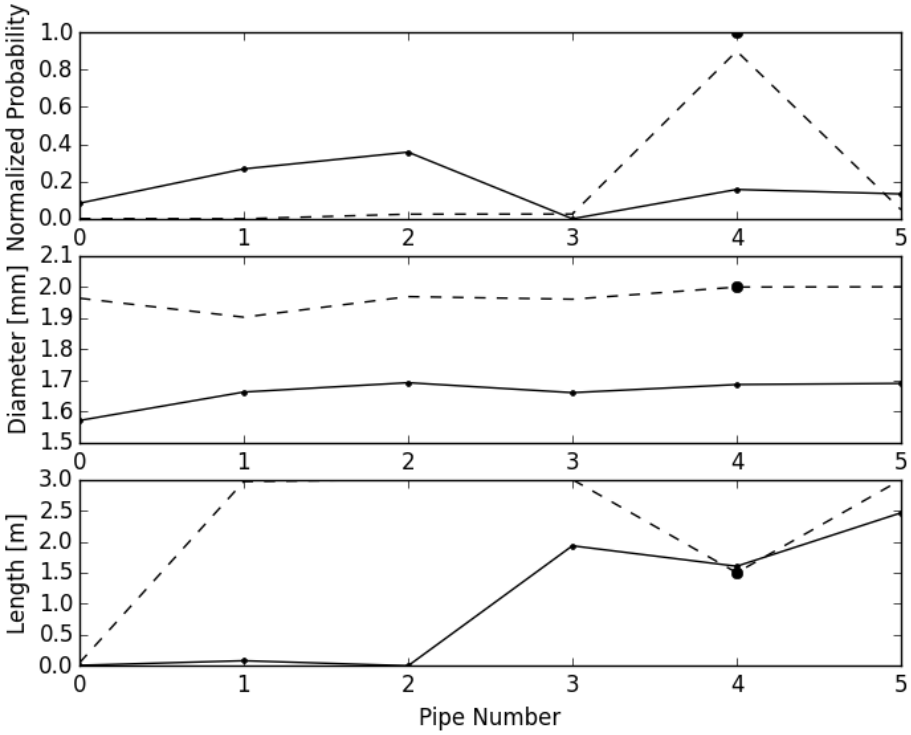


Figure C.18: Solution for the Third Leak on the Second Experimental Network by the Bayesian Probabilistic Analysis

Support Vector Machine

Figure C.19 to C.21 shows the results found for the three leak cases by the Support Vector Machine.

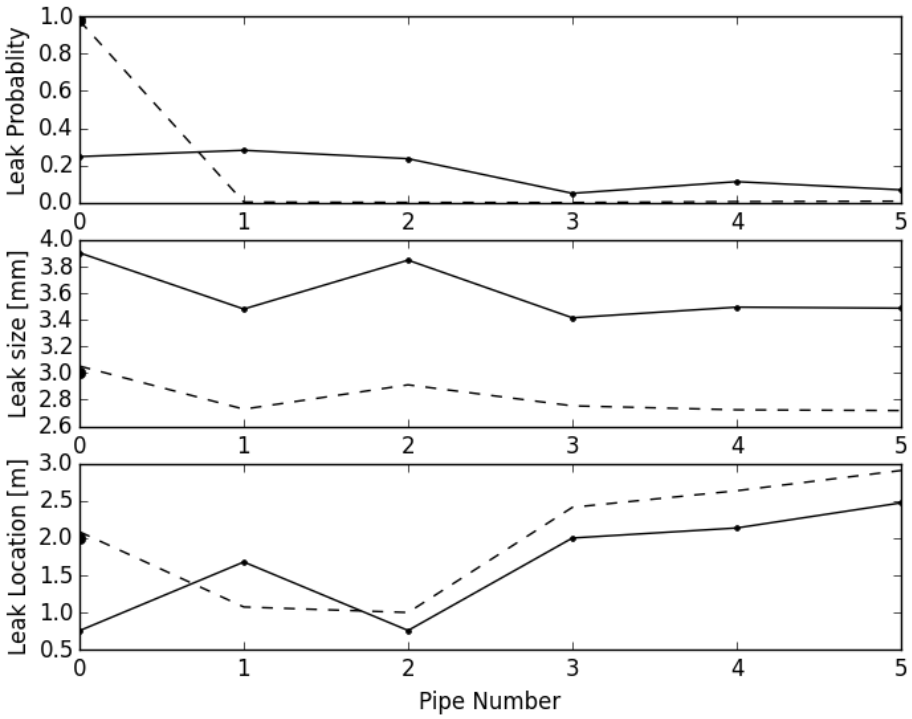


Figure C.19: Solution for the First Leak on the Second Experimental Network by the Support Vector Machine

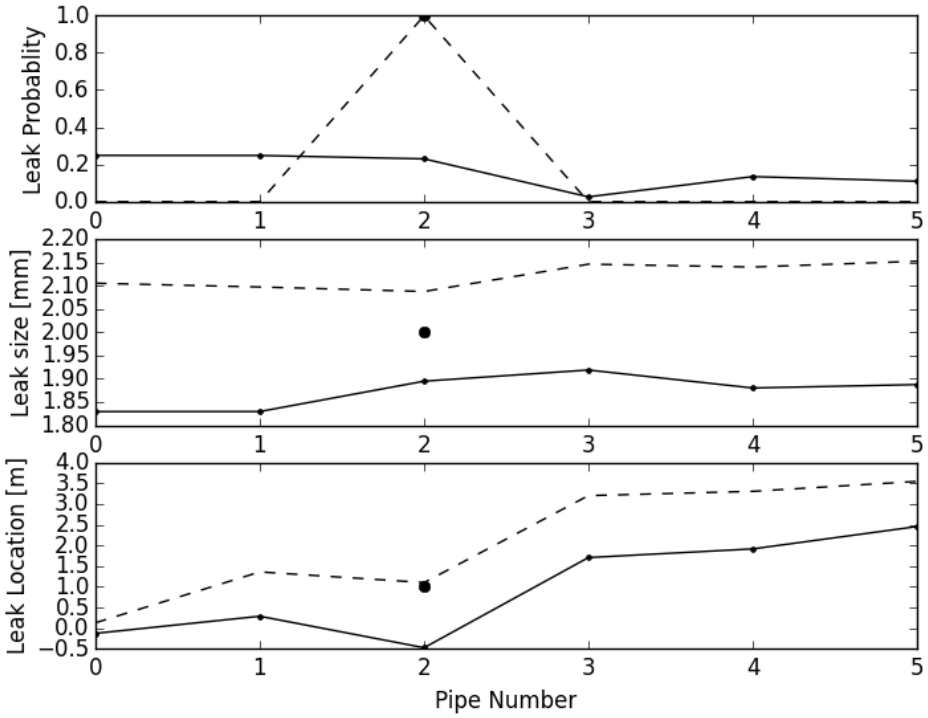


Figure C.20: Solution for the Second Leak on the Second Experimental Network by the Support Vector Machine

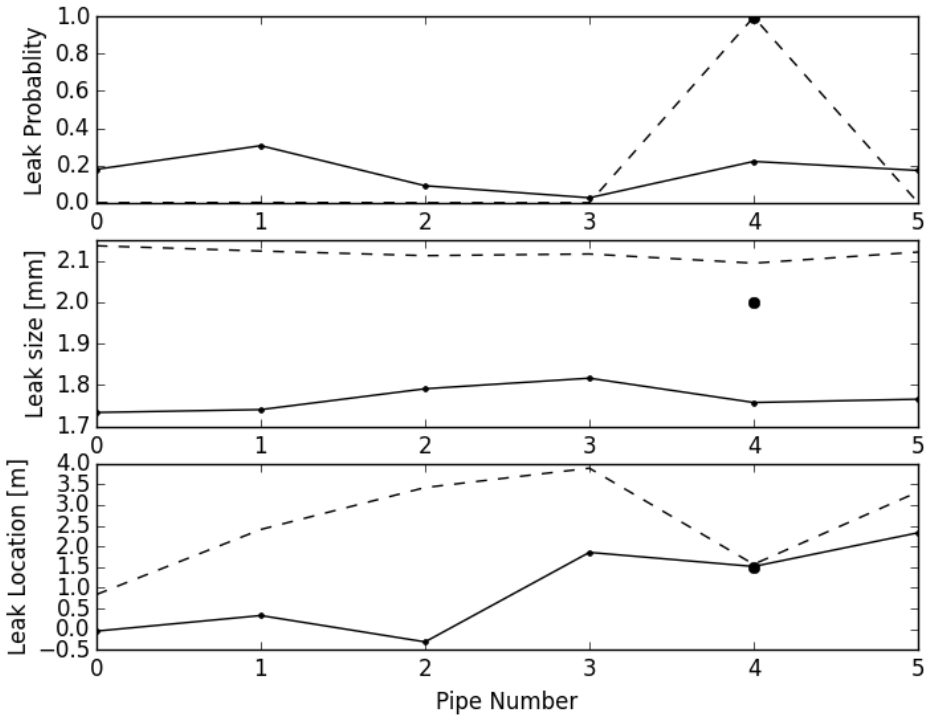


Figure C.21: Solution for the Third Leak on the Second Experimental Network by the Support Vector Machine

Artificial Neural Network

Figure C.22 to C.24 shows the results found for the three leak cases by the Artificial Neural Network.

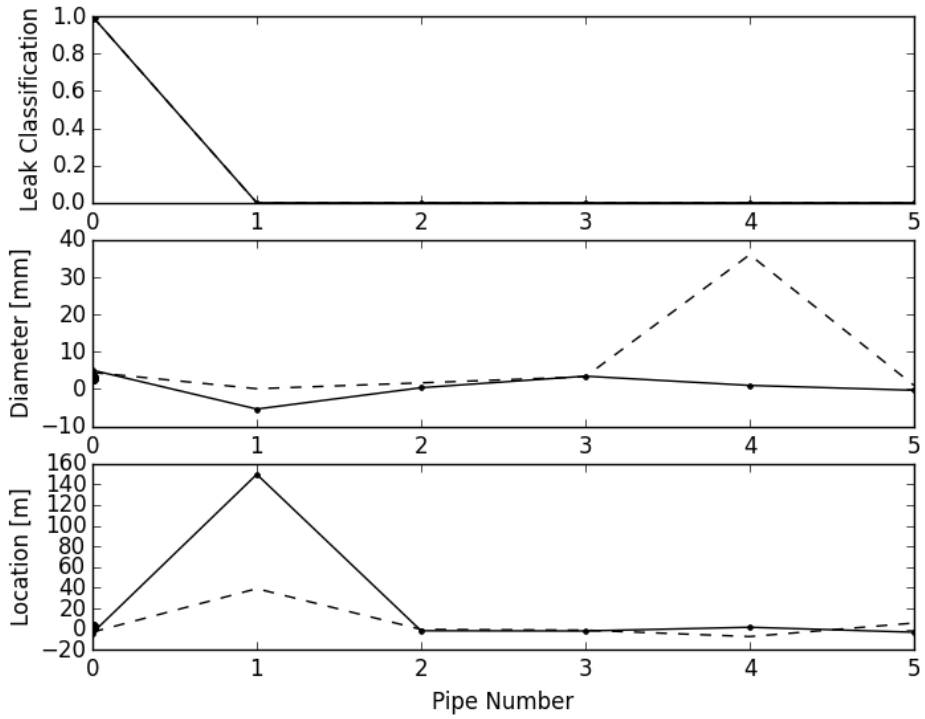


Figure C.22: Solution for the First Leak on the Second Experimental Network by the Artificial Neural Network

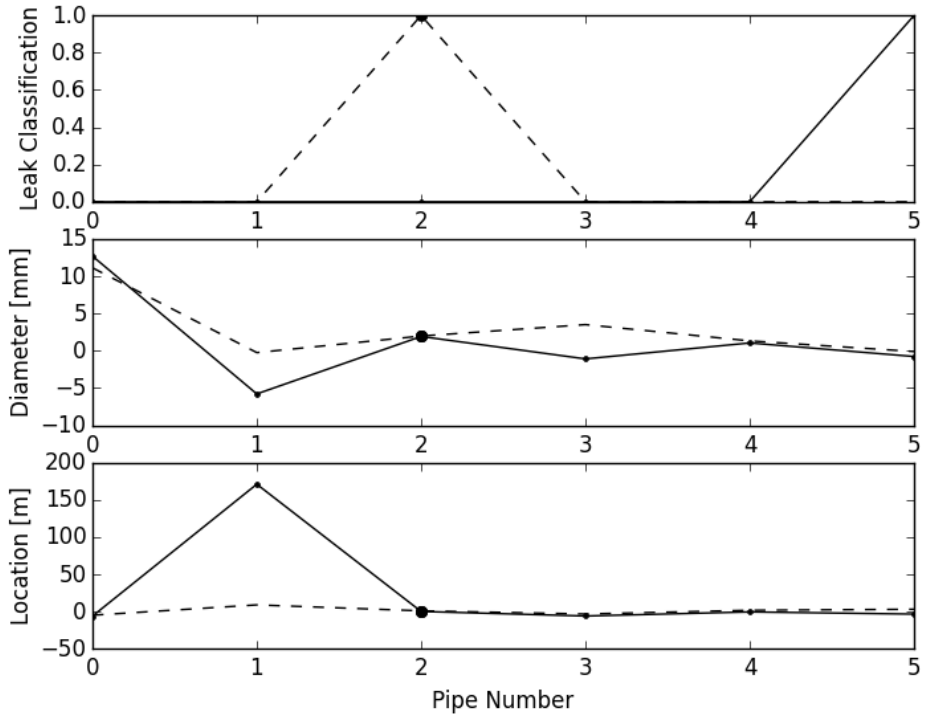


Figure C.23: Solution for the Second Leak on the Second Experimental Network by the Artificial Neural Network

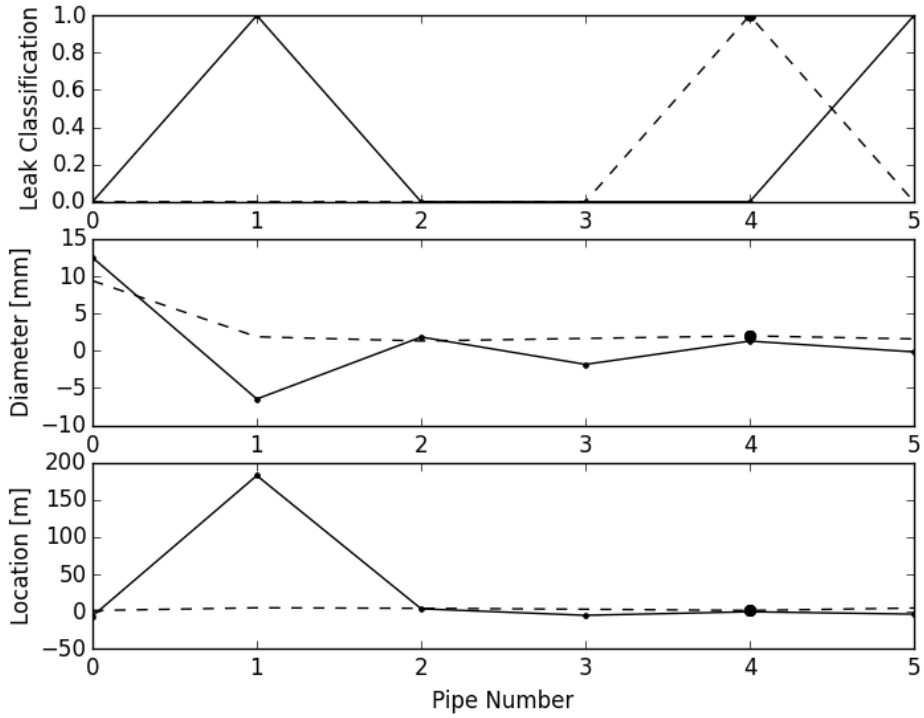


Figure C.24: Solution for the Third Leak on the Second Experimental Network by the Artificial Neural Network