Pipe network leak detection: Comparison between statistical and machine learning techniques

JC van der Walt, PS Heyns, DN Wilke

February 2018

Center of Asset Integrity Management

Department of Mechanical and Aeronautical Engineering
University of Pretoria

Keywords: Leakage, Artificial intelligence, Integrated water management

Abstract

This paper investigates an inverse analysis technique to find leaks in water networks and compares different solution strategies. Although a number of strategies have been proposed by different authors to identify leaks on a vast selection of pipe networks, limited research has been done to compare strategies and point out their weakness. Three strategies, a Bayesian Probabilistic Analysis, a Support Vector machine and, an Artificial Neural Network were combined with the inverse analysis technique on different numerical and experimental networks to point out each strategies weakness. Two numerical networks are investigated and one experimental network. It is shown that the Bayesian Probabilistic Analysis struggles to find unique solutions when a few observations are available, while the Support Vector Machine and the Artificial Neural Network struggle when only flow measurements are available. Additionally it is shown that the Artificial Neural Network struggles to estimate unique solutions for leak size and location.

1 Introduction

In the 2011 − 2012 National Non-Revenue Water assessment [1] in South Africa the average national non-revenue water was found to be 37%. While the world average in that period was estimated to be 36.6% [1]. 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 only reduce the amount of leaked water, or raise the awareness. Contractors use listening sticks, geophones, ground penetrating radar, and noise loggers to help assist in finding the leaks [1].

In recent years research to find leaks using on-line machine learning techniques with real time data has intensified. Pèrez et al. [2] applied this the model-based methodology to a real network in Barcelona. Their case study focused on the Nova Icària DMA (District Metered Area), where a real leak occurred. Soldevila et al. [3] investigated Bayesian Reasoning with the model-based methodology on the Hanoi DMA. They investigated four cases, leaks varying between 25 − 75 l/s, 5% noise on pressure, 5% uncertainty in the demands and a case where all three these cases were combined. They compared their results with previously found results using a k-Nearest Neighbors approach. Additonally they investigated using Bayesian Classifiers [4] with the model-based methodology. In this work they investigated two case studies, the Hanoi and Nova Icària DMA.

Additional to the techniques used to find the leaks, sensor placement for this method of leak detection is highly important. Fuchs-Hanusch et al.[5] compared six different sensor placement algorithms on a real-world network. In their work they opened fire hydrants to simulate leakages within the network ranging between 0.25 − 1 l/s. The six algorithms compare are shortest path 1, shortest path 2, Shannon entropy [6], a binarized sensitivity matrix [7], a non binarized sensitivity matrix [8] and SPuDU [9].

The following sub-sections show work done by different authors to compare several techniques, as well as work done in this field by using Artificial Neural Networks, Support Vector Machines and the Bayesian Probabilistic analysis. These three techniques were chosen as a focus due to their and popularity.

1.1 Comparisons Between Different Strategies

A. Nowicki et al.[10] investigated data-driven models for fault detection in water distribution networks. They investigated Kernel PCA (Principle Component Analysis) and applied the method for fault detection to the water distribution network of Chojnice, located in northern Poland. They compared the results of the Kernel PCA with a normal PCA and a simple Control Chart. M. Romano et al.[11] investigated geostatistical techniques for burst detection in water distribution networks. They compared four different techniques, namely inverse distance weighted interpolation, local polynomial interpolation, ordinary Kriging and ordinary CoKriging. They compared these techniques on a case study with a rural water network consisting of 17.8km of pipes. They measured 13 pressure measurements throughout the network and simulated bursts by opening fire hydrants.

1.2 Artificial Neural Networks

Caputo et al. [12] proposed a method of using artificial neural networks to estimate the leak location in piping networks. They performed tests on a network where they generated input data for leaking and non leaking states. Two neural networks were used in their proposal, the first identifying the leaking branch and the second estimating the leakage amount and location. Applying the neural networks, 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. [13] performed tests on an 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 found that they could locate the bursts with an accuracy of 98.33%.

Salam et al. [14] 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. They used a Radial Basis Function Neural Network which could detect the leak location and sizes with an accuracy of 98%.

1.3 Support Vector Machines

De Silva et al. [15] investigated support vector machines to act as pattern recognisers to detect leaks in pipe networks. They started with a SVM (Support Vector Machine) as a regressor to try and predict emitter coefficients. Six monitoring nodes were used to act as sensor locations. They selected 10 candidate leaking nodes and generated a data set with varying emitter coefficients. The SVM could, after training, achieve a testing accuracy of 76.8%.

They then used 40 candidate leaking nodes and created a data set, for which a testing accuracy of 57.2% was achieved. 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 whether the SVM could detect small leaks 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.

1.4 Bayesian Probabilistic Framework

Poulakis et al. [16] investigated a Bayesian probabilistic framework to detect leaks in a water pipe network. The derivation starts by assigning θ as the parameter to be optimized. This parameter includes 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 parameter.
A model error can now be written as:

$$e_{ij} = \bar{x}_{ij} - x_{ij}(\theta), \tag{1}$$

where $\bar{x}_{ij}$ are the actual measurements from the system. Another parameter was added to be optimized, namely σ. This parameter represents the uncertainty within the error.

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

$$P(\theta, \sigma|\bar{x}) = c_1 P(\bar{x}|\theta, \sigma)\pi(\theta, \sigma), \tag{2}$$

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, normally distributed with a zero mean, and standard deviation of σ, the likelihood of $P(\bar{x}|\theta, \sigma)$ can be written as:

$$P(\bar{x}|\theta, \sigma) = \prod_{i=1}^{L}\prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_{ij}-\bar{x}_{ij})^2}{2\sigma^2}}, \tag{3}$$

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 3 into equation 2. The simplification can be written as:

$$P(\theta, \sigma|\bar{x}) = c_2 \frac{1}{(\sqrt{2\pi\sigma})^{LN}} e^{\left(-\frac{\sum_{j=1}^{N} ||x_{ij}-\bar{x}_{ij}||^2}{2\sigma^2}\right)}. \tag{4}$$

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 with one leak.

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 an uncertainty of 2%, the location could be calculated. If the uncertainty in the model measurements was increased to 5% the model was unsure about the actual leak location.

1.5 This Research

There was found that various theoretical and practical networks have been tested with machine learning techniques such as Neural Networks, Support Vector Machines, and other statistical approaches. This paper investigates the pressure-flow deviation method with different solution strategies to find the application of each strategy for specific networks and an overall comparison between the strategies.

Three strategies are investigated: the Bayesian Probabilistic analysis, a Support Vector Machine, and an Artificial Neural Network. The three investigated strategies are then applied to three water networks, of which two are numerical and the other an experimental network, to gain a deeper understanding of the networks and strategies.

The first two tested networks simulate a numerical transportation and distribution network. The solutions for these networks are idealized, with no model or measurement error. This is to find limitations within the solution strategies. The experiential network contains calibration of the model and therefore contains model and measurement errors. The experimental model is based on a simple distribution network to offer some complexity to the problem.

2 The Algorithms

Three strategies investigated in this research are discussed in this section.

2.1 Strategy 1: Bayesian Probabilistic Analysis

This method was proposed by Poulakis et al. [16]. It uses the Mean Squared Error (MSE) calculation 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:

$$P(\theta, \sigma|\bar{x}) = c_2 \frac{1}{(\sqrt{2\pi\sigma})^{LN}} e^{\left(-\frac{MSE(\theta)}{2\sigma^2}\right)}, \tag{5}$$

where θ is the optimized leak parameter. The optimization of this algorithm requires the equation to be maximized for the largest probability. Therefore the log-likelihood of this function can be calculated as:

$$g(\theta,\sigma) = -\ln(P(\theta,\sigma|\bar{x}) = \frac{MSE(\theta)}{2\sigma^2} + \frac{LN}{2}\ln(\sigma^2), \tag{6}$$

where L is the total number of monitoring locations, N is the total number of flow tests and σ is the uncertainty within the error. The log-likelihood given by $g(\theta,\sigma)$ can now be minimized. For this strategy no data set is generated. Within the optimization algorithm the EPANET model is simulated with the estimated parameters which results in the pressure and flow measurements to calculate the error. This is repeated until the error is minimized.

2.2 Strategy 2: Support Vector Machine

For the Support Vector Machine a data set is generated. 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. Two SVM types are used, the first estimating on which pipe a possible leak occurred, and the second to find the possible leak size and location on all the pipes in the network.

For the classification SVM the outputs are an integer value depending on the number of pipes in the network. For the regression SVM, the outputs are the leak size and leak location on the length of the pipe. All the SVMs use a RBF kernel function and they are generated using scikit-learn [17] which is a Python package used for machine learning.

2.3 Strategy 3: Artificial Neural Network

The Artificial Neural Networks considered the same input and output parameters as the SVMs. The sizes of the ANNs (Artificial Neural Networks) vary between problems with the sizes being chosen for each problem to result in the optimum results. This was completed by increasing the size of the ANN until the accuracy of the prediction stopped increasing or before over fitting occurred. The output for the classification ANN returns a string of numbers which suggest pipes with possible leaks.

Each pipe in the network has its own regression ANN which calculates the leak location and size. The number of hidden layers and nodes in a layer is identical to that of the classification ANN. The output for the regression ANN is the leak size and location of the leak. In Python the ANNs are created using the scikit-learn [17] package similar to the SVMs.

For both the SVMs and ANNs the training and testing data are split with a 70:30 ratio. Within the training algorithm 10% of the training data is used as validation to ensure over fitting does not occur, which is then validated using the testing data set.

3 Numerically Simulated Networks

The strategies were tested on two different numerical water networks. These networks include a simple single pipe network and a distribution network found in the literature. The simple problem is chosen to offer an in depth understanding of the performance of the solution strategies for the most simple problem. The second numerically investigated problem is a complex network that is reproducible from the literature. Both these networks are idealized with no model or measurement errors. This is to ensure this method of leak detection have the ability to solve the leak size and location from the measured data.

3.1 Problem 1: Simple Single Pipe System

This network consists of two reservoirs connected together. Figure 1 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 will be measured and P_1 indicates the pipe with a leak on it.
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 the same and at $0m$. The length of P_1 is $100m$ with a diameter of $32mm$. The location to the leak is measured from N_1. The leak is modeled with an emitter coefficient of 0.3. The flow though this network

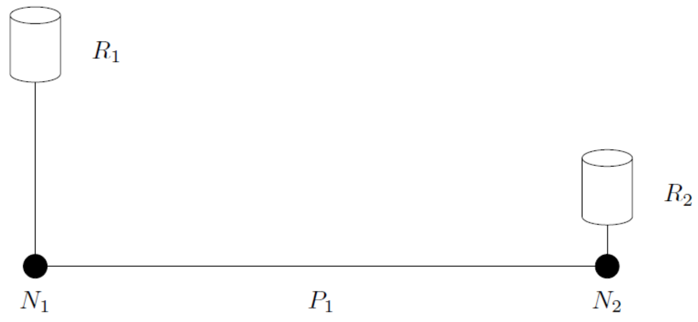


Figure 1: Diagram of Problem 1

was calculated by EPANET as 2.68m/s, giving a Reynolds Number of 85760 resulting in turbulent flow in the pipe. The Hazen-Williams model was used to solve the EPANET model.

To solve this problem with the SVM and ANN, a data set of 1000 samples were generated with random leaks at locations between $0 - 100m$ and leak sizes with emitter coefficients between $0 - 2$. Pressure and flow measurements could be generated by simulating the EPANET model with the randomly generated leak parameters. For this problem the input data for the strategies is the pressures, flows and the leak location. The leak location was added to the input to help with the presentation of the results. The output data for the SVM and ANN is simply the diameter of the leak.

3.1.1 Solutions

The solution found by the three solution strategies can be seen in Figure 2. The dot in the figures shows the actual modeled leak. From the figure the Bayesian Probabilistic analysis error can be seen as 0 throughout the length of the pipe. This indicates there is an infinite number of solutions for this strategy. This is due to the Bayesian Probabilistic Analysis optimizing to find the minimum error for the problem, which is the difference between the two measured pressures. For this problem a leak can be added anywhere on the pipe which will result in the error being zero, giving an infinite amount of solutions. Adding the flow within the pipe to the input of the algorithms could solve this problem since the problem is no longer ill-posed. The dot in the figure indicates the actual leak location.

From the solution for the SVM it can be seen that the SVM gives a unique solution for this problem which is where the error is calculated to be 0.0005, at its minimum. The estimated solution was found at a length of $58.4m$ and a emitter coefficient of 0.413. For the ANN it can be seen that a unique solution was found for this problem. The minimum error of the ANN solution is found to be 0.02. This results in a leak length of $42.9m$ and an emitter coefficient of 0.244. The difference between the Bayesian Probabilistic analysis and the SVM or ANN is that the SVM and ANN uses the absolute values of the pressures while the Bayesian Probabilistic analysis uses the pressure difference over the pipe.

3.2 Problem 2: Benchmark 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. [16] and they applied the Bayesian Probabilistic framework to it. The same network was used by Nasirian et al. [19] to benchmark their new heuristic genetic algorithm methodology to find leaks. In 2016 the same network was used by Asgari et al. [20] where they investigated a new method of locating a leak by calculating a leak index.

The network is depicted in Figure 3. The supply to the network comes from reservoir $R1$, which has a static head of $52m$. Each junction has a demand of 50 l/s, which are numbered from $N1 - N30$. The pipe lengths are $1000m$ and $2000m$ respectively, as depicted in the figure. The pipe diameters vary from $600mm$, to $450mm$, and finally $300mm$ as the flow decreases through the network.

Two cases are considered: the first only pressure is measured, and the other only flow rate is measured. For case A the pressure observation nodes are marked with circles around the nodes. For case

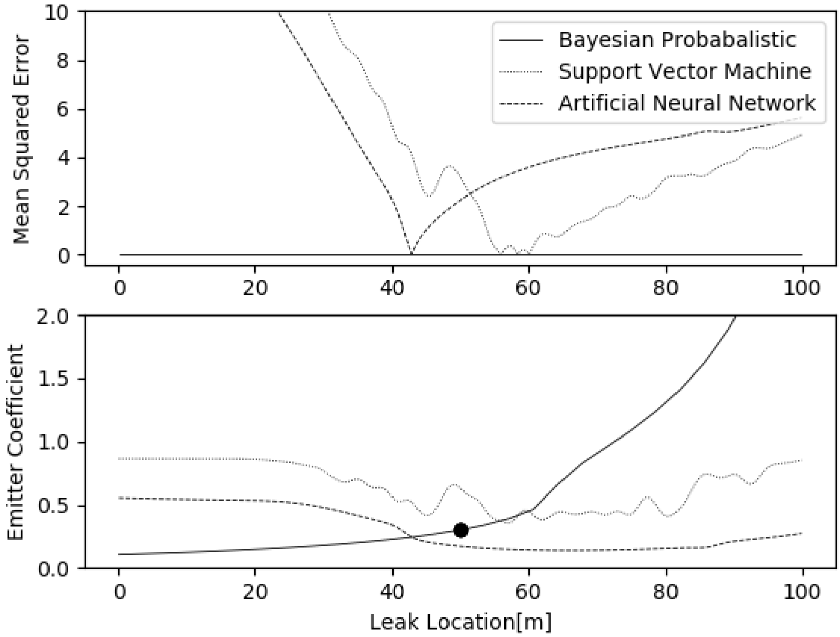


Figure 2: Problem 1: Solutions found by the three strategies

B the flow rate observation pipes are marked with rounded rectangles. A leak of $22.8l/s$ is added to the network. The leak is modeled at the center of $P25$, at node $N55$, which is marked with a square. This leak is modeled as a demand. The same sensor placement was used as introduced by Poulakis et al. [16]. The flow regime within this network was calculated and found to be turbulent, therefore the Hazen-Williams formula is used within EPANET.

Similar to the previous problem, a data set was generated for different leak cases. The data set was used for both the SVM and the ANN, which contained 10000 data samples. The data samples were randomly generated with leak sizes between $10l/s$ and $50l/s$. This data set contained two outputs, the leaking pipe and size. The inputs for this data set were simulated with the EPANET model depending on the tested case.

3.2.1 Solutions

The solution found by the solution strategies can be seen in Figure 4. This figure shows the solution for the Bayesian Probabilistic analysis for case A, where the most probable leak location was 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 39.9%. The leak size was estimated as $22.8l/s$.

The solution found after training the SVM to the generated data set can also be seen in this figure. For case A it can be seen that a probability of a leak at $P25$ was calculated as 54.3%, with the adjacent pipes with leak probabilities of 17% and 8%. For case B it was found that the actual leak location was given a probability of 1.8%, with other pipes having higher probabilities of leaks. The leak size estimation of the SVM was calculated between $29l/s$ and $31l/s$, while the actual leak size was $22.8l/s$.

The solution found by the ANN after training shows a perfect classification solution as well as a leak estimation of $22.8l/s$ for case A. For case B it can be seen that an incorrect classification solution for the location was found. The leak size estimation for this case was found to be $22.7/s$.

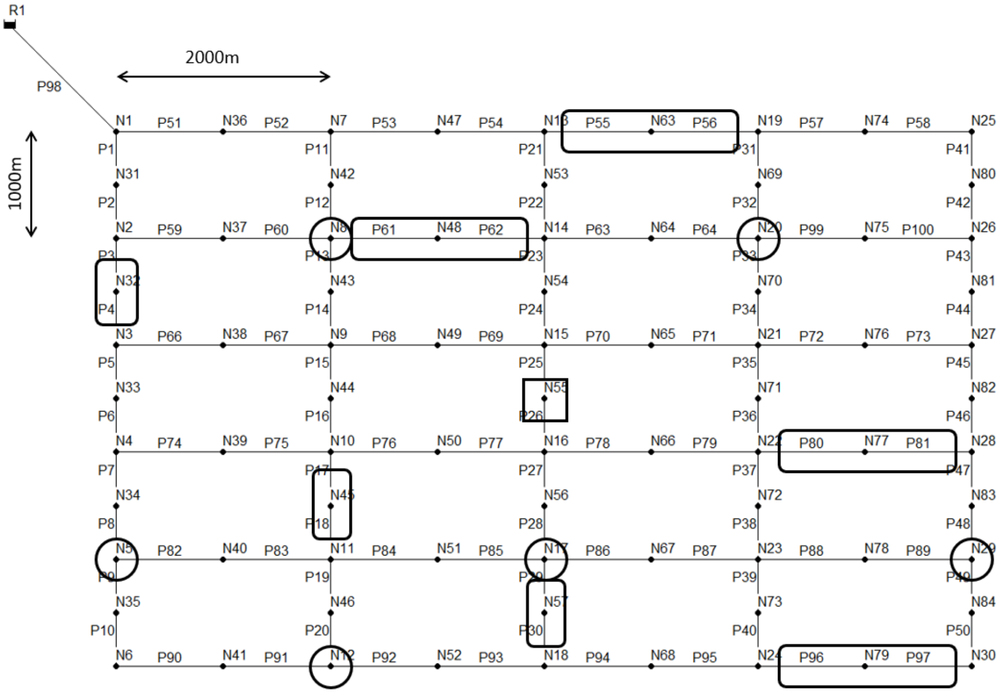


Figure 3: Diagram of Problem 2

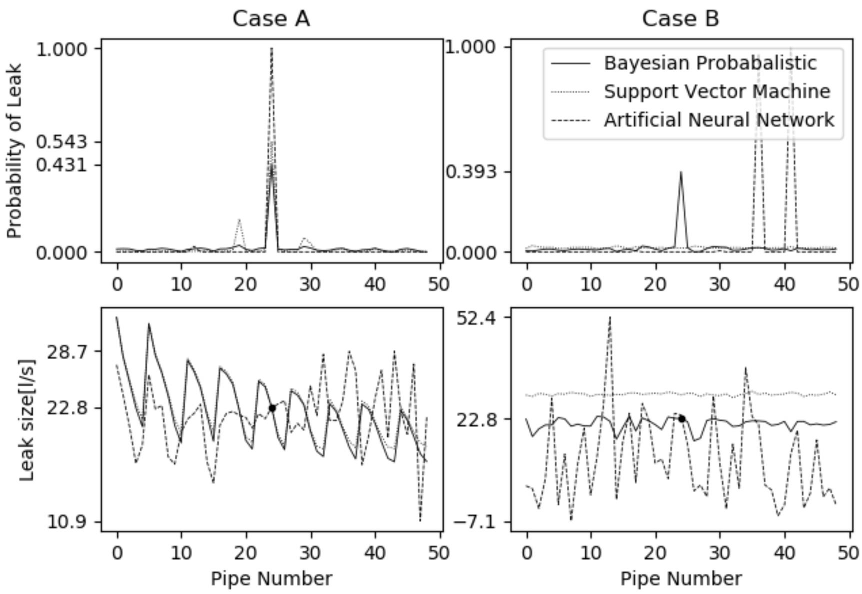


Figure 4: Problem 2: Solutions found by the three strategies

3.3 Discussion of Results

A comparison between these strategies showed that the Bayesian Probabilistic Analysis could not find leaks in simple pipe networks due to the ill-posed nature of the problem, while the SVM and ANN could solve the simple pipe network accurately by algorithmic regularization. This indicates that the Bayesian Probabilistic Analysis have difficulties solving the problem when too little information is known about the network.

The solutions for Problem 2 showed that the the Bayesian Probabilistic Analysis can solve the problem accurately using either pressure or flow measurements. This is not the case for the SVM and ANN as both of these strategies could not find the leaking pipe or its size when only flow measurements are used. Indicating that the error optimization methods have an advantage when using flow measurements.

4 Experimental Network

The solution strategies were tested on an experimentally built network. The experimentally measured values are calibrated and tested to find the actual leak for two different cases.

4.1 Setup

The experimental network built can be seen in Figure 5. The experimental networks used twelve pressure sensors, one at the start and end of each pipe. Additionally, seven flow meters are used: one measuring the input of the network while six measure the output flows of each pipe. In this figure it can be seen that six possible leaking pipes were used. These pipes are fed by a pump from a reservoir. The demands in the network was modeled with $3mm$ holes, which fed back to the reservoir.

The reservoir used is a simple container holding $50l$ of water and can not be seen in this photo. A Pentax CM 210 pump was used which is capable of supplying the network with the necessary pressure and flow. The lengths of each pipe is $3m$ with a diameter of $10mm$. The pressure in this network with the specified supply achieved a pressure in the network ranging between $2-3$ bar, with an average flow of about $0.1-0.2$ l/s through each pipe, resulting in a Reynolds Number ranging between $100-200$. This results in the flow to be Laminar in this network, therefore the Darcy-Weisbach formula was used within EPANET.

Three leak locations were added to the network of which only the first two leak locations were considered. The leaks was located on the first and third pipe with diameters of $3mm$ and $2mm$ respectively. The leaks was located at a location of $2m$ and $1m$. The valves between the pipes to change the network layout are closed.

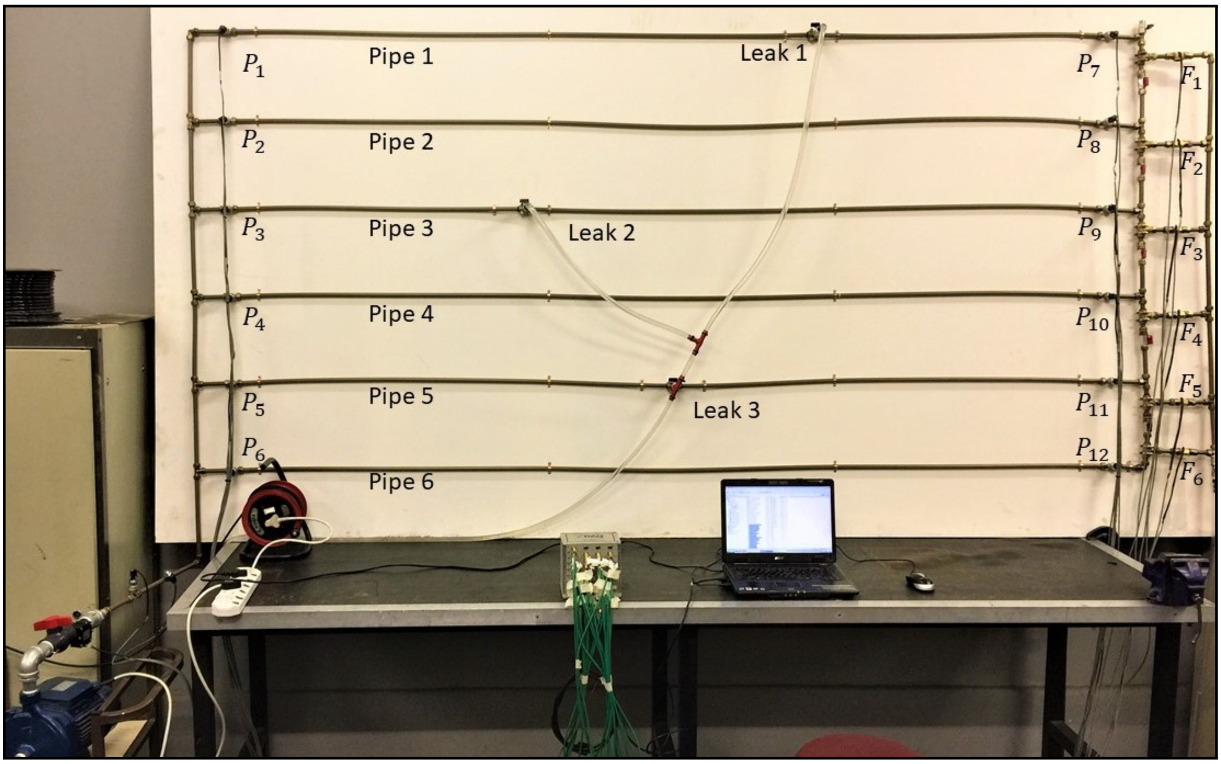


Figure 5: Photo of the Built Experimental Network

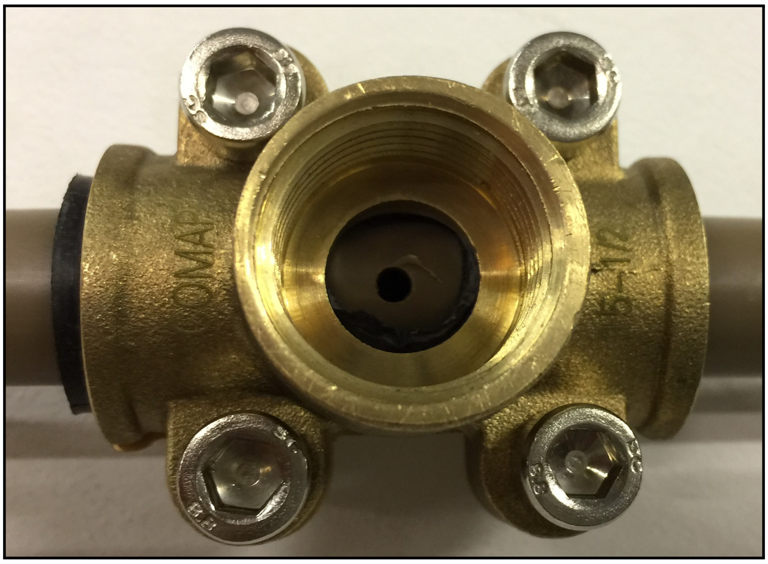


Figure 6: Actual Leak of $3mm$ Applied to the Experimental Network

Figure 6 shows the $3mm$ leak applied to the network. The leak is applied by clamping a saddle over the pipe, and drilling the correct diameter hole through the pipe. The demands applied to each pipe were created in the same way, resulting in the network being pressure driven. The pressure sensors had a full scale error of 1.5% and could measure pressure from $0-5bar$. The flow sensors used could measure between $1-25l/min$ and had an error rating of 3%.

Figure 7 shows the EPANET model used to calibrate the experimental measurements. In this figure $P1-P12$ indicate the locations of the pressure sensors while $F1-F7$ indicate the locations of flow sensors. $L1-L6$ indicate the leaks on the pipe. The calibration process consists of calculating an error between the experimental measurements and the model measurements when there was no leaks. These measurements include all 12 pressure and 7 flow measurements. The parameters optimized include the roughness coefficient, a loss coefficient on each pipe, an emitter coefficient simulating the demand in the network and the pump efficiency. Calibrating these parameters resulted in a total error of 2.56% when there was no leak.

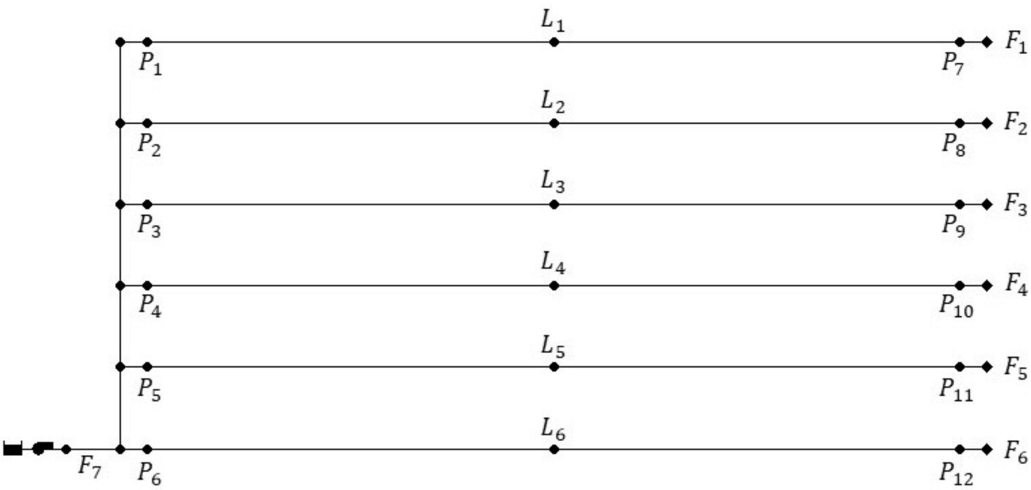


Figure 7: EPANET model used to calibrate experimental measurements

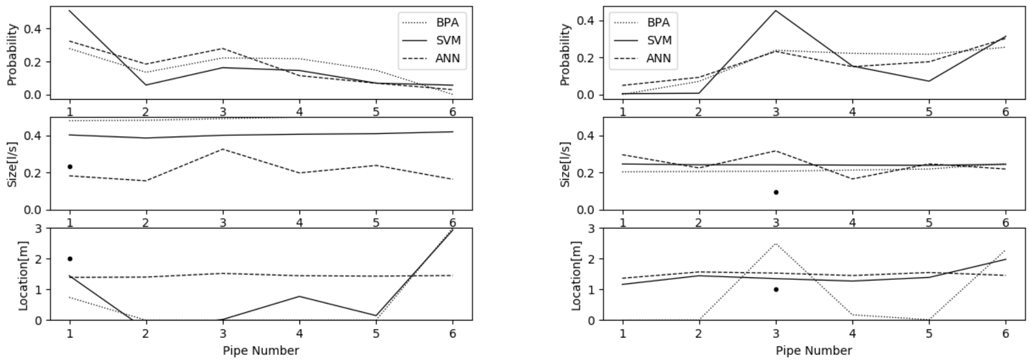
To solve the experimental problem limits to the optimization of the Bayesian Probabilistic analysis and the data sets for the SVM and ANN had to be set. For the three strategies, the leak size was limited to an emitter coefficient between $0-0.1$, while the leak location was limited between $0-3m$. For the SVM and ANN the data set contained 1500 data samples per pipe, with a total of 9000 samples.

4.2 Results of Experimental Measurements

The results found by the three strategies for the two leak cases can be seen in Figure 8. For the two leak cases, the actual leak size and location was marked with a black dot. The leak size for the two leaks were calculated from the flow measurements as $0.23l/s$ and $0.0935l/s$. The leak locations were at $2m$ and $1m$.

In these figures it can be seen that the Bayesian Probabilistic analysis could find the leaking pipes with probabilities of 27.9% and 23.4%. For the second leak case the leak was identified on the sixth pipe with a probability of 25.4%. The Bayesian Probabilistic analysis calculated the leak size to be $0.48l/s$ and $0.21l/s$ for the two leak cases, with the location calculated at $0.74m$ and $2.49m$.

The SVM could identify the leaking pipe accurately for both cases with a probability of 51.0% and 45.1%. The leak size for the two cases was estimated to be $0.40l/s$ and $0.24l/s$, with the location of the leaks estimated at $1.44m$ and $1.35m$.



(a) Leak on first pipe (b) Leak on third pipe

Figure 8: Experimental Network: Solutions of the two leak cases

The ANN classified the two leaking pipes with probabilities of 32.4% and 23.2%. For the second leak case the sixth pipe was estimated to be leaking with a probability of 30.2%. The leak size for the two leak cases was estimated as $0.18l/s$ and $0.31l/s$. The location of the leaks was estimated to be at $1.39m$ and $1.53m$.

4.3 Discussion of Results

For the experiment the model calibration could be completed by optimizing 39 coefficients. These coefficients included the roughness coefficient, minor losses, emitter coefficients and the pump efficiency which resulted in the model error of 2.56%.

The three strategies could all identify the first leak case correctly, with only the SVM finding the correct leaking pipe for the second leak case. Both the Bayesian Probabilistic analysis and the SVM estimated the leak to be nearly double of its actual size, with the ANN training to constant values for the leak size and location in both cases.

5 Conclusion

The pressure-flow deviation method was used to find leaks in various water distribution and supply networks. The three solution strategies used could find leaks in all the networks tested. The algorithms used were the Bayesian Probabilistic Analysis, a Support Vector Machine, and an Artificial Neural Network.

In the first numerical network it was seen that the ANN and SVM outperform the Bayesian Probabilistic Analysis since the Bayesian Probabilistic Analysis make use of error calculations, removing the absolute pressure and flow values from its estimation. In the second numerical problem it was seen that the ANN and SVM struggled to find the leaking pipe as well as its size when only flow measurements were used, although the Bayesian Probabilistic analysis could find the leak in both cases.

In the experimental network the SVM could identify the leaking pipes for both cases, while the Bayesian Probabilistic analysis and the ANN only identified the leaking pipe for the first case. Furthermore it was seen that the Bayesian Probabilistic analysis and SVM estimated the leak to have nearly double its size. The ANN was seen to train to constant estimations for the leak size and location.

From this work it can be seen that no strategy consistently outperform any of the other. This begs the question whether these strategies should not be combined to work together to find better predictions. This leak detection technique can also be combined with on-line monitoring systems allowing for quick and accurate detection of leaks. Although this technique requires further investigation to find accurate leak sizes and exact location estimations, it can accurately find leaks in networks and identify the pipes they are on. In addition, this work shows more research needs to be completed on model calibration techniques to help with the detection of the leaks, their sizes and locations.

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] 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.

[3] A.Soldevila, R.M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig. Leak localization in water distribution networks using model-based bayesian reasoning. *European Control Conference*, pages 1758–1763, June 2016.

[4] A.Soldevila, R.M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig. Leak localization in water distribution networks using bayesian classifiers. *Journal of Process Control*, pages 1–9, April 2017.

[5] D. Fuchs-Hanusch and D. Steffelbauer. Real-world comparison of sensor placement algorithms for leakage localization. *XVIII International Conference on Water Distribuion Systems Analysis*, pages 499–505, 2017.

[6] W.B.F de Schaentzen, G.A. Walters, and D.A. Savic. Optimal sampling desing for model calibraion using shortest path, genetic and entropy algorithms. *Urban Water*, pages 114–152, June 2000.

[7] R. Perez, J. Quevedo, V. Puig, F. Nejjari, M.A. Cuguero, G. Sanz, and J.M. Mirats. Leakage isolation in water distribuion networks: A comparative study of thwo methodologies on a real case study. *19th Mediteranean Conference on Control Automation*, pages 138–143, 2011.

[8] M.V Casillas, V. Puig, L.E Garza-Castanon, and A Rosich. Optimal sensor placement for leak location in water distribuion networks using genetic algorithms. *Conference on Control and Fault-Tolerant Systems*, pages 61–66, June 2013.

[9] D. Steffelbauer, M. Neumayer, M. Günther, and D. Fushs-Hanusch. Sensor placement and leakage localization considering demand uncertanties. *Procedia Engineering*, pages 1160–1167, 2014.

[10] A. Nowicki, M. Grochowiski, and K. Dusinkiewichz. Data-driven models for fault detection using kernal pca: A water distribution system case study. *Applied Mathematics and Computer Science*, 22:939–949, 2026.

[11] M. Romano, Z. Kapelan, and D.A. Savić. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15.3:634–651, 2013.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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