An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine

John Mashford, Dhammika De Silva, Donavan Marney and Stewart Burn
Commonwealth Scientific and Industrial Research Organisation
PO Box 56, Highett, Vic. 3190, Australia
{John.Mashford,Dhammika.Desilva,Donavan.Marney,Stewart.Burn}@csiro.au

Abstract

***Abstract*—This paper presents a method of mining the data obtained by a collection of pressure sensors monitoring a pipe network to obtain information about the location and size of leaks in the network. This inverse engineering problem is effected by support vector machines (SVMs) which act as pattern recognisers. In this study the SVMs are trained and tested on data obtained from the EPANET hydraulic modelling system. Performance assessment of the SVM showed that leak size and location are both predicted with a reasonable degree of accuracy. The information obtained from this SVM analysis would be invaluable to water authorities in overcoming the ongoing problem of leak detection.**

Keywords – leak detection; pipe networks; pattern recognition; support vector machines

I. INTRODUCTION

Leaks in pipe networks represent an important problem costing many millions of dollars annually. The difficulty in leak location is compounded by their hidden nature (Figure 1). The benefit of leak detection and rehabilitation is largely determined by the efficacy of leak detection and location techniques. The primary method of detecting and locating leaks at this time is an acoustic method, in which acoustic signal emissions of pipes are monitored in order to detect

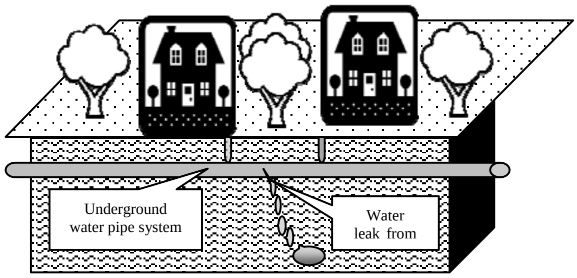

Figure 1. Diagrammatic representation of a leak from an underground water pipe

the sound made by leaks. The disadvantages of the basic acoustic listening techniques are well documented in an overview by Fuchs and Riehle [1].

The non-acoustic techniques of leak detection in pipe systems include methods based on the following:

1. Injection of tracing substances into the fluid stream;
2. Electro-magnetic inspection of the pipe from the inside;
3. Analysis of quasi-static signals detected by sensors built into the pipe system such as pressure, flow rate and temperature sensors;
4. Analysis of transient signals detected by sensors built into the pipe system such as pressure wave analysis;

5. Analysis of leak generated temperature variations using infrared thermography by sensors located outside the pipe system;

6. Identification of radar or radio frequencies emitted from transmitters located inside the pipes and permeating through pipe cracks.

Some techniques including #'s 2, 5 and 6 are relatively expensive while others including #'s 1, 2 and 6 intrude into the fluid stream thereby increasing the risk of contamination. Although considerable research has been carried out on the development of transient methods (*i.e*., #4), examples, such as the inverse transient method of Liggett and Chen [2] are not straightforward in their application beyond trivial pipe networks.

The detection methods based on technique #3 involve interpretation of induced signals. They seek to determine the pipe system state from the measured parameters of pressure, flow rate and sometimes temperature at various points and times within the pipeline system. This is an inverse engineering problem for which data mining and pattern recognition techniques are applicable. Computational pattern recognisers such as artificial neural networks (ANNs) and support vector machines (SVMs) are particularly useful for such types of problems.

Neural networks operating on quasi-static pressure and flow readings have been used for leak detection in pipe systems. Caputo and Pelagagge describe an approach to detecting spills and leakages from pipeline networks using a multilayer perceptron back-propagation artificial neural network (ANN) [3]. The system analyses data from pressure and flow rate information in order to determine the location and size of leaks in the pipe network. A similar approach utilising only pressure readings is described by Shinozuka et al. [4]. Another application of ANNs operating on steady state process parameters for leak detection in pipe systems was described by Belsito et al. [5]. Similarly, neurofuzzy techniques have also been applied to the problem by Feng and Zhang [6], as well as Izquierdo *et al* [7].

SVMs are statistical pattern recognisers which perform similar functions to ANNs. They operate by seeking separating hyperplanes in a high dimension feature space obtained by an, in general, non-linear transformation of their input space. They are implemented by solving a quadratic optimisation problem. They have a number of properties which make them superior to ANNs such as having better generalization ability and requiring smaller training sets [8]. In the present paper we consider the use of SVMs for analysing pressure values collected from a pipe network for the purpose of leak location and sizing.

II. Leak Location and Sizing Using Support Vector Machine

A method using SVMs was tested for leak size and location prediction. SVMs can be used for regression or classification. When acting as a classifier the output is a predicted class associated with an input pattern and when acting as a regressor the output is a real number associated with an input pattern. An SVM acting as a regressor behaves as a function approximator. SVMs are trained on a training set consisting of a number of input patterns and the associated output values or categories. They can then be tested on a testing set to determine the relevant performance metric; *i.e.*, classification accuracy in the case where they are acting as a classifier or the mean squared error (MSE) and correlation coefficient (R^2) when they are acting as a regressor.

The method used for leak size and location prediction was to monitor the pressure at a number of nodes in the pipe network under consideration and to feed these pressure values into SVMs trained to predict leak size and leak location. The SVMs are trained on a number of cases representing the presence of leaks of various sizes and locations in the pipe network.

The SVMs require 100's or possibly 1000's of cases in their training sets. Therefore it is not feasible to generate the training sets by introducing actual leaks into the real pipe network. The training sets can be obtained by simulation of the pipe network under consideration. The simulation tool EPANET [9] can be used to achieve this. EPANET is a computerized simulation model produced by the Environmental Protection Agency of the USA that predicts the dynamic hydraulic and water quality behaviour within a drinking water distribution system operating over an extended period of time. Leaks of various sizes can be simulated in EPANET and the resulting pressures and flows in the network can be calculated.

In order to generate the large number of cases required for the SVM training set, the implementation of EPANET can be automated by developing a program which calls EPANET many times with various values of the leak size.

III. Modelling Leakage in EPANET

Although EPANET is primarily designed for modelling network supply and water quality issues, the emitter property in EPANET, which is designed to model fire hydrants/sprinklers can be adapted to model leaks.

The definition of the emitter is based on the classical Torricelli equation for flow through an orifice;

$$Flowrate: Q = C * A * P^{P\exp} \quad (1)$$

Where: C is a coefficient, A is the orifice aperture area, P is the fluid pressure and $Pexp$ is the pressure exponent.

The pressure exponent is typically 0.5 for circular apertures.

Based on this equation, EPANET applies a simple definition for the emitter function.

$$EC = Q / P^{P\exp} \qquad (2)$$

Where *EC* is the emitter coefficient, *Q* is the flow rate, *P* is the fluid pressure and *Pexp* is the pressure exponent.

The units associated with the emitter coefficient are flow rate per unit pressure, *i.e.* litres per second per meter of pressure ($ls^{-1}m^{-1}$).

EPANET permits the value of the Emitter Coefficient to be specified for individual leak sites, but the pressure exponent can be only specified for the entire network.

Eq. 2 indicates that the desired low leakage rates of 50-100 lhr^{-1} can only be achieved by specifying emitter coefficients in the order of 0.001 in combination with a pressure exponent value of 0.5.

Hydraulic simulations through EPANET are affected by very low values of emitter coefficient therefore, convergence (of the numerical simulation) could only be attained by increasing the system accuracy to 0.00075.

Investigation by Greyvenstein and Van Zyl, on the effects of leakage in water distribution systems have shown that the pressure exponent is dependant on the geometry of the orifice [9]. Table 1 (from van Zyl and Clayton, [10]) shows that corrosion areas in metal pipe correspond to the highest values of the exponent, and, that in plastic pipe, depending on the orientation of the leak defect, the exponent can take different values.

TABLE I. LEAKAGE EXPONENTS FOUND IN EXPERIMENTAL STUDY

Failure Type	Leakage Exponent* for pipe materials		
	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

*These values are for turbulent flows and with the leaks to the atmosphere.

The soil will exert a back pressure at the leak site. This soil induced back- pressure directly influences the leakage rate by lowering the available head; *i.e.* the higher the soil pressure, the lower the leakage rate. However, soils can't withstand large pressures, and the behaviour outside a leak seems to be very complex. It is likely that the soil will not affect the exponent, but might affect the coefficient.

In the absence of experimental evidence to the contrary, the EPANET simulations for this study were carried out with the exponent value set at 0.5.

IV. DEVELOPMENT OF THE SVM ANALYSIS METHOD

The EPANET model was applied to simulate the pipe network in an area in South Eastern Melbourne. The site was selected with a view to validating simulation outputs with *in-situ* measurements to be collected later in the project.

The first experiment that was carried out was to determine how effectively an SVM regressor predicts emitter coefficient values when a given fixed node is leaking. Leakage from zero to a high rate of approximately 3.0 ls^{-1} was modelled in these experiments. The EPANET driving program was used to generate a data set of 300 cases with the pressure exponent set at 0.5. In each case the node with label Nd73 was leaking and its emitter coefficient value was varied from 0.000 to 0.300 in steps of 0.001. From these, 200 and 100 cases were randomly selected to form a training set and a testing set respectively. The SVM was trained using the radial basis function kernel. The training converged very rapidly and when the trained SVM was applied to the testing set the mean squared error (MSE) and the squared correlation coefficient (R^2) were:

MSE = 4.47569e-005,
R^2 = 0.994289.

Thus the testing results are considered to be good because of the relative values of MSE and R^2. The SVM is acting as a regressor (or function approximator) and its accuracy is measured by the MSE and R^2 values because its outputs are floating point numbers (leak sizes, *i.e.*, emitter coefficient values) rather than discrete classes.

The next experiment that was carried out, was to determine the effectiveness of using an SVM classifier for leak location. 10 nodes were selected from the pipe network under consideration as candidate leak nodes. The EPANET driving program was used to generate data representing the pressures at the 6 monitoring nodes for the case when exactly one of the candidate leak nodes is leaking for emitter coefficient values increasing from 0.000 to 0.300 in steps of 0.002. 10 data sets of 150 cases each were generated. The data sets were then amalgamated and 1000 cases randomly selected to form a training set. The remaining 500 cases were used as a testing set. The cases in the training and testing sets contain representatives of a variety of leak locations and sizes.

When the SVM was trained with its default parameter settings it achieved a testing accuracy of 19.0%; this is not quite as bad as it might first seem, because if it were guessing randomly, it would achieve the correct answer 1 time in 10; *i.e.*, this corresponds to an accuracy of 10%.

The training parameters were then adjusted in order to optimise the testing accuracy. The best testing accuracy achieved was 76.8%. The training parameters are the cost '*C*', and the kernel function parameter 'γ'. The testing accuracy for various values of these parameters are shown in Table 2.

In addition to the above mentioned leaking nodes, a further 30 nodes were used as candidate nodes for leaking making a total of 40 candidate leak nodes. 6000 cases were generated by applying the EPANET driving program 40 times with emitter values going from 0.00 to 0.30 and a training set of 4000 cases was randomly selected. The remaining 2000 cases were used for testing. After manually optimising the training parameters the testing accuracy was 57.2%. The accuracies obtained for various values of the training

TABLE II. TESTING ACCURACIES WITH 10 CANDIDATE LEAK NODES

C	γ	% accuracy
1000	0.17	60.2
10000	0.17	67.6
100000	0.17	72.2
1000000	0.17	75.2
10000000	0.17	76.8
100000000	0.17	74.8
10000000	0.001	73.4
10000000	0.0001	68.4
10000000	1	76
10000000	10	74.4

parameters are shown in Table 3. If the SVM was guessing randomly then it would achieve the correct answer 1 time in 40 (or 2.5% of the time). Therefore an accuracy of 57.25% for 40 candidate leak nodes is considered to be satisfactory.

While the system is not correctly predicting the actual leaking node in 42.75% of the cases it is of interest to know if the predicted leaking node is near to the actual leaking node. A histogram of the distance between the predicted leaking node and the actual leaking node for 40 candidate leaking nodes with the testing set of 2000 cases is shown in Figure 2. The interval between 0 and 1500 was divided up into 15 histogram bins (the diameter of the study region is 1243.4 m).

The predicted node equals the actual node in 57.25% of the cases. However Figure 3 shows that the

TABLE III. TESTING ACCURACIES WITH 40 CANDIDATE LEAK NODES

C	γ	% accuracy
10000000	0.17	57.25
1000000	0.17	55.2
1000	0.17	32.85
100000000	0.17	56.75
9000000	0.17	57.1
11000000	0.17	56.75
10000000	0.001	53.45
10000000	0.01	55.15
10000000	0.1	56.15
10000000	1.0	56.4

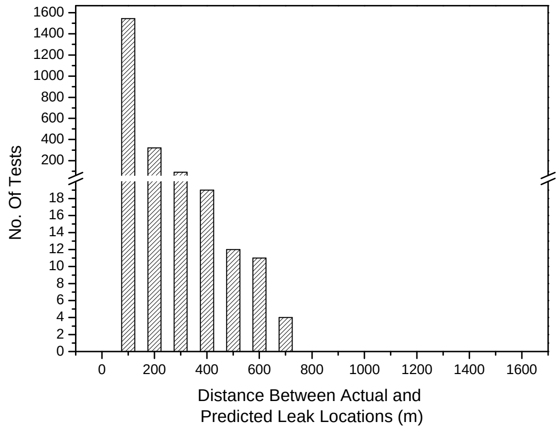

Figure 2. Histogram of distance between predicted leaking node and actual leaking node.

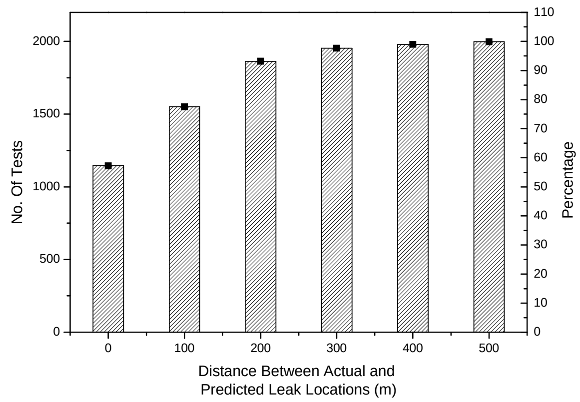

Figure 3 Histogram of percent accuracy within a nominated distance between actual and predicted node

predicted node is within 100 metres of the actual node in 77.5% of the cases, within 200 metres of the actual node in 93.2% of the cases and within 300 metres of the actual node in 97.7% of the cases. Thus the prediction would provide useful information to the water authority to direct their search for the leaking node

V. Applying the SVM analysis method for low leak rates

Having established the procedure to generate the required data through EPANET and apply the data to train the SVM to predict the position and size of leaks, the next step was to establish the limitations of the procedure in detecting small (100 lhr^{-1} or less) leaks.

The combination of an emitter coefficient of 0.3 and emitter exponent of 0.5 correspond to a leakage level of 2.5 ls^{-1}. As this represents leak volumes up to 10,000 lhr^{-1}, additional experiments were conducted in order to determine if the technique could be applied for lower leak rates.

Operation of EPANET with a series of low emitter coefficients (using an emitter exponent fixed at 0.5) indicated the lowest leakage successfully processed was at a coefficient of 0.0001; which is equivalent to a leakage rate of 3.45 lhr^{-1}.

On that basis the EPANET driving program was used to generate data representing the pressures at the 6 monitoring nodes in the case when exactly one of the candidate leak nodes is leaking for emitter values going from 0 to 0.0001 in steps of 0.00001. 10 data sets of 150 cases each were generated.

The results showed there was no difference in pressure from the 'no' leak scenario in any of the simulations. In other words the EPANET sensitivity was not sufficient to register the small difference in pressure.

The experiment was repeated while progressively increasing the upper limit of the coefficient from 0.0001 until a pressure difference registered on the EPANET output. This exercise showed the lowest coefficient that produces a pressure difference in EPANET to be 0.0025. This was equivalent to a leakage of 90 lhr^{-1}.

The EPANET driving program was then applied to generate data while changing emitter coefficient from 0 to 0.0025 in steps of 0.0001. The output data was used as the training data set for SVM analysis. The training set was limited to 10 data sets. The analysis showed that prediction of the exact location had a 35% success rate. As the distance between actual and predicted leak location increased, the success rate increased. A 100% success rate was predicted by 500 m (Figure 4).

Although the accuracy at shorter distances could be improved, this will be at the expense of increased leakage volume. Experiments with an emitter coefficient of 0.005 produced a success rate of 56% for prediction of exact location; this represents a 60% increase in accuracy from the analysis using 0.0025 coefficient. As shown in Figure 4 the prediction accuracy at other distances also improved

VI. Practical aspects of applying the method in field measurements

The practicality of designing a leak detection method using SVMs depends on monitoring the pressure changes which are traceable to leaks. Table 4 shows the output from the EPANET driving program for coefficients from zero to 0.0025. The pressure changes registered at the nodes are in the order of 0.00001 m (0.1 Pa). Therefore pressure changes of that order are required to be monitored in order to detect leaks in the range of 90 lhr^{-1}. (Emitter coefficient 0.0025).

The viability of a detection method based on these measurements depends on instrumentation with the required sensitivity. In addition the need to filter pressure

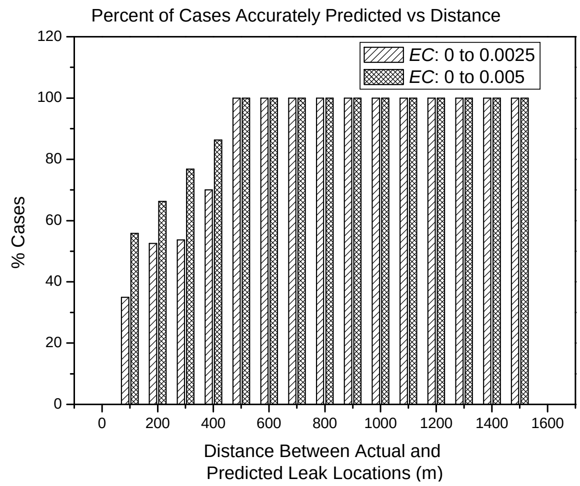

Figure 4 Prediction accuracy with Emitter Coefficient 0 to 0.0025 (max leak size 90 lhr^{-1}) and Emitter Coefficient 0.005, (leak size 180 lhr^{-1}), exponent 0.5 [EC=Emitter coefficient]

TABLE IV. Output from EPANET driving program for emitter coefficient 0 to 0.0025

Node[Nd17] .Emitter	Node[Nd46]. Pressure	Node[Nd16]. Pressure	Node[Nd23]. Pressure	Node[Nd59]. Pressure	Node[Nd70]. Pressure	Node[Nd36]. Pressure
0	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0001	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0002	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0003	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0004	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0005	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0006	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0007	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0008	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0009	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.001	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0011	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0012	112.5000000	109.9000015	111.5000000	107.0000000	99.5000000	105.7999954
0.0013	112.5000000	109.8999939	111.5000000	107.0000000	99.5000000	105.7999954
0.0014	112.5000000	109.8999939	111.5000000	107.0000000	99.5000000	105.7999954
0.0015	112.5000000	109.8999939	111.5000000	107.0000000	99.5000000	105.7999954
0.0016	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0017	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0018	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0019	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.002	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0021	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0022	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0023	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0024	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954
0.0025	112.5000000	109.8999939	111.4999924	107.0000000	99.5000000	105.7999954

differences from background flow variations needs to be addressed. While most of the background effects could be minimised by collecting data between 2:00 and 3:00 am, the total elimination cannot be guaranteed. Therefore a method of identifying pressure variation due to non-leak events is required.

Hydraulic data simulated through EPANET and analysed using the SVM has demonstrated the feasibility of detecting and locating leaks in the order of 90 lhr^{-1}. The detection limit could be improved with more sensitive simulation of the hydraulic parameters. While the limits of EPANET have probably been reached in this study, other hydraulic simulation tools may be available that provide data at lower leakage levels suitable as training sets for SVM analysis.

VII. CONCLUSION

This paper has described a method for the location and sizing of leaks in a pipe network by processing pressure values obtained at a number of points in the network using SVMs. The work is based on the concept that the required information can be obtained by sufficiently sophisticated processing of the pressure data in a network and that SVMs are suitable for this processing task.

The training data for the SVMs has been obtained by using the hydraulic modelling system EPANET.

The results of our experiments with a modelled pipe network is that leak size can be accurately predicted and that leak location can be predicted to within specified spatial resolution with an accuracy that depends on the resolution. The location prediction limits the set of locations that need to be considered when searching for a leak, thereby providing useful information for a water authority.

The applicability of the technique in practice depends on the ability of pressure sensors to detect small changes in pressure and the accuracy to which EPANET models real pipe networks.

ACKNOWLEDGMENT

Funding for this research was provided by the SEQ Urban Water Security Research Alliance. The authors would like to thank Mike Rahilly for programming work on this project and for valuable discussions.

REFERENCES

[1] H.V. Fuchs and R. Riehle, “Ten years of experience with leak detection by acoustic signal analysis”. App. Acoust. 33(1), 1991, pp. 1-19.

[2] J.A Ligget and L.C. Chen, “Inverse transient analysis in pipe networks”, J. Hyd. Eng. 120(8), 1994, pp. 934-955.

[3] Caputo, A.C. and Pelagagge, P.M. (2003), “Using neural networks to monitor piping systems”, Process Safety Progress, Vol. 22, No. 2, 2003, pp.119-127.

[4] M. Shinozuka, J. Liang and M.Q. Feng, “Use of supervisory control and data acquisition for damage location of water delivery systems”, Journal of Engineering Mechanics, 2005, pp. 225-230.

[5] S. Belsito, P. Lombardi, P. Andreussi, and S. Banerjee, “Leak detection in liquefied gas pipelines by artificial neural networks”, AIChE Journal, Vol. 44, No. 12, 1998, pp. 2675-2688.

[6] J. Feng and H. Zhang, “Algorithm of pipeline leak detection based on discrete incremental clustering method”, in ICIC 2006, LNAI 4114, Huang, D.-S., Li, K and Irwin, G.W. (eds.), Springer-Verlag, 2006, pp. 602-607.

[7] J. Izquierdo, P.A. López and R.P. Martínez, “Fault detection in water supply systems using hybrid (theory and data-driven) modelling”, Mathematical and Computer Modelling 46, 2007, pp. 341-350.

[8] Kwok, T.J., Support vector mixture for classification and regression problems, Proc. Of the International Conference on Pattern Recognition (ICPR), Brisbane, Australia, 1998, pp. 255-258.

[9] Rossman, L.A., EPANET 2 User’s Manual, U.S. Environmental Protection Agency, Cincinatti, Ohio, U.S.A., 2000.

[9] B. Greyvenstein and J.E. Van Zyl, “An experimental investigation into the pressure leakage relationship of some failed water pipes”, Proc. Leakage 2005, Halifax, Canada, 2005.

[10] Van Zyl, J.E. and Clayton, C.R.I., “The effect of pressure on leakage in water distribution systems”, Water Management 160, 2007, pp. 109-114