Pipeline Failure Prediction in Water Distribution Networks using weather conditions as explanatory factors

Konstantinos Kakoudakis[a], Raziyeh Farmani[b] and David Butler[c]

[a,b,c] College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF

Corresponding author: kk337@exeter.ac.uk

Abstract

This paper examines the impact of weather conditions on pipe failure in water distribution networks using Artificial Neural Network (ANN) and Evolutionary Polynomial Regression (EPR). A number of weather related factors over four consecutive days are the input of the binary ANN model while the output is the occurrence or not of at least a failure during the following two days. The model is able to correctly distinguish the majority (87%) of the days with failure(s). The EPR is employed to predict the annual number of failures. Initially, the examined network is divided into six clusters based on pipe diameter and age. The data from the last year of monitoring data is used for testing while the remaining years since the beginning of the monitoring period are retained for model development. A distinctive EPR model is developed for each cluster based on the relevant training data. The obtained results indicate a strong relationship between the annual number of failures and frequency and intensity of low temperatures. The outputs from the EPR models are used to calculate the failures of the homogenous groups within each cluster proportionally to their length.

Keywords: Artificial Neural Network, Data-driven modelling, Evolutionary Polynomial Regression, Pipe failure predictions, Rehabilitation, Weather factors, Water distribution networks

1. Introduction

The optimal management strategy for a WDN balances issues of water safety, reliability, quality and quantity while exploiting the full extent of the useful life of pipes and reducing long-term costs through proactive management (Kleiner and Rajani 2001, Clair and Sinsha 2012). In order to enhance this strategy, the use of predictive models is fundamental since they provide insights into the relationships between pipe failure and all the factors influencing it. These factors can be split into pipe-intrinsic, operational and environmental. Environmental and pipe-intrinsic factors can be further divided into static and dynamic (time-dependent), while the operational factors are inherently dynamic. The pipe-intrinsic factors such as the pipe material, diameter, length, age and the operational factors such as pressure, previous number of failures have been examined in several studies (e.g. Kleiner and Rajani 2001, Clair and Sinsha 2012, Nishiyama. and Filion 2013).

A few approaches have examined the impact of environmental factors on pipe failure trend in Canada and northern USA (Kleiner and Rajani 2002; Rajani et al. 2012; Laucelli et al. 2014), Australia (Gould et al. 2011), Netherlands (Wols and Thienen 2013) and Austria (Fuchs-Hanusch et al. 2013).

Gould et al. (2011) conducted a statistical analysis to examine the impact of weather factors on the pipe failure of various material, diameter and soil type groups. The focus of the analysis was to relate the variation in the monthly failure rate with the dynamic weather factors. Wols and Thienen (2013) used a linear regression analysis to ascertain the relationship between weather data and pipe failure. The analysis was conducted separately for different cohorts, depending on the type of pipe material, year of installation, and diameter for a two months interval. Fuchs-Hanusch et al. (2013) examined the correlation between failure frequencies and climatic indicators. The winter and summer failure frequencies were examined separately. Rajani et al. (2012) used a non-homogenous Poisson-based pipe deterioration model to examine the impact of air temperature-based and water temperature-based covariates on breaks of homogenous groups of pipes with respect to pipe material, age and diameter. They examined a number of (non-overlapping) time steps lasting from 5 up to 90 days concluding that the best time step for data aggregation is 30 days. The proposed model was not validated on a test dataset since the analysis merely aimed at ascertaining the impact of temperature-based covariates on failure trends rather than using them for predictions. Laucelli et al. (2014) investigated the relationship between climate data and pipe bursts of 150mm cast iron pipes using the Evolutionary Polynomial Regression. They examined three non-overlapping time steps (5, 15 and 30 days) and concluded that the 30 days' time step provides the most accurate results. The analysis was conducted separately for the warm and cold season.

The failure frequency in a WDN is not constant due to the inherent nature of some of the factors affecting it. Hence, this paper examines the relationship between the annual number of pipe failures and time-dependent weather factors. The proposed approach does not require the distinction between cold and warm seasons to be made. The approach for making annual predictions is complementary to a previous one (Kakoudakis et al. 2017) which calculated the average failure rate for a specific period using pipe-related factors as explanatory variables. The failure frequency is the cumulative effects of several factors on the pipes, therefore the results of the two approaches are combined. Furthermore, a method is proposed to identify more vulnerable regions of the network and visualize them on a map.

The occurrence of pipe failures requires the fast response of the network's operators. The water companies aim to respond as soon as possible after a burst is reported to minimize the amount of lost water and the customer dissatisfaction that might result in need for compensation. The response time depends on several factors including, amongst other, the availability of human resources and the ability to predict time intervals with an above the normal failure frequency. Previously developed approaches have resulted in relationships with low accuracy (i.e. Rajani et al. 2012) for short-term predictions. This paper proposes a method to predict the occurrence of pipe failure(s) on a short-period without requiring knowledge of the weather). In addition, the weather-related factors are ranked based on their importance for predictions.

It should be noted here that the annual predictions can be used in conjunction with long-term predictions for pipe maintenance/rehabilitation/replacement scheduling while the short-term predictions are strictly for operational use. Furthermore, the results for the annual predictive models are on a cluster level while the short-term predictions refer to the entire examined network (due to the small number of failures in some clusters) and do not associate the failure occurrence with specific pipes.

The remainder of the paper is organized as follows: First the proposed methodologies are explained. Section 3 provides description of the software used. Then the process to evaluate the accuracy of the proposed methodology is explained. The main features of the case study are provided in section 5. This is followed by discussion of the results and the evaluation of their accuracy. Finally, the last section highlights the most important conclusions.

2. Methodology

2.1 Annual pipe failure prediction

The annual failure rate is an important performance indicator for assessing the overall structural condition of a water distribution network (Fuchs-Hanusch et al. 2013) and therefore models that predict it are of significant interest. This paper presents a method for predicting the next year's number of failures considering weather conditions as explanatory variables. Furthermore, outputs from the models are used to calculate the failure rates of individual pipes in order to identify regions of the network that are more prone to failure. The methodology consists of the following steps:

1. First, individual pipes are *aggregated* into homogenous groups based on their diameter, the age and the soil type assuming that pipes that share the same characteristics are expected to have a similar failure rate (Kleiner and Rajani 2012). Soil type is used as an *aggregation* criterion because soil properties have been associated with the corrosion of the metallic pipes (Sadiq et al. 2004) which has been identified as a dominant factor contributing to their failure (Makar 2000). The original dataset containing a large number of individual pipes is converted to a new dataset containing homogenous groups of pipes.
2. The created homogenous groups of pipes are allocated into six clusters using their attributes of diameter and age based on the findings of a previous analysis (Kakoudakis et al. 2017) that demonstrated how splitting the network into six clusters using the K-means clustering method could result in more accurate predictions. Instead of using a single model for making predictions for all the homogenous groups, six separate predictive models are developed.
3. For each cluster the annual number of failures is equal to the sum of failures of the homogenous groups within. The candidate weather related explanatory variables are: average minimum air temperature (Eq. 1), average maximum air temperature (Eq. 2), average soil temperature (Eq. 3) and freezing index which is calculated only for the days below a predefined threshold (Eq. 4).

$$\mathrm{AveT_{min}}=\frac{\sum_{j=1}^{m} \mathrm{T_{minimum}^{j}}}{\mathrm{m}} \quad (1)$$

$$\mathrm{AveT_{max}}=\frac{\sum_{j=1}^{m} \mathrm{T_{maximum}^{j}}}{\mathrm{m}} \quad (2)$$

$$\text{AveST}=\frac{\sum_{j=1}^{m} ST^{j}}{m} \tag{3}$$

$$\text{FI}=\sum_{j=1}^{m}(\theta - T_{minimum}^{j}) \tag{4}$$

Where m is the number of days in the time step (i.e. 365 days), $T_{minimum}^{j}$ is the minimum daily temperature of day j, $T_{maximum}^{j}$ is the maximum daily temperature of day j , ST^{j} is the average daily soil temperature of day j and θ is the predefined air temperature threshold

4. The freezing index (Eq 4) is defined as the cumulative minimum daily temperature below a specified air temperature threshold and is considered as a surrogate for the severity of extreme air temperatures within a time step (Kleiner and Rajani 2002). The cross-correlation function in MATLAB (® R2014b) is applied to measure the similarity between the candidate thresholds and the number of failures. The thresholds examined range between -2^{0}C and 4^{0}C with a step of 1^{0}C. This process is repeated separately for each cluster. The threshold that provided the highest similarity (highest values of cross-correlation) was selected for data aggregation.

5. The Evolutionary Polynomial Regression models are selected with respect to their goodness of fit, the minimization of model's polynomial terms and the possibility to describe the physical phenomenon. The predicted numbers are used to calculate the number of failures for the homogenous groups within each cluster proportionally to their length.

6. Then, it is assumed that all the pipes within those homogenous groups have the same failure rate for this specific year. These values are used in combination with the complementary approach (Kakoudakis et al. 2017) to calculate the final failure rate for this year.

2.2 Daily prediction of the occurrence of pipe failures

The proposed method aims to predict the occurrence of failure and consists of the following steps:

1. Define the inputs and the output of the model. The inputs of the ANN model are: the minimum air temperature, the maximum air temperature, the mean air temperature, the soil temperature, the freezing index and the number of failures for a number of consecutive days while the targeted output of the model is 1 if there is at least a pipe failure the following few days and 0 if not. The temperature

variation can occur relatively quickly whereas the potential pipe failure because of that might take longer (Rajani and Kleiner 2012). Therefore, different combinations of number of days are examined in order to obtain the models with the highest accuracy. Exhaustive trials were conducted leading to the conclusion that the use of four consecutive days as input and the following two days as output results in the highest accuracy. The first input is the set of variables for the first four days and the output is the occurrence of failure(s) in the fifth and sixth day. Respectively the second input is the set of variables from the second up to the fifth day, while the output is the failure in the sixth and seventh day.

2. The inputs and the outputs are divided into two parts, for training (70%) and test (30%). The ANN model is built relying only on the training data.

3. The actual output of the model is not an integer number; therefore the optimal threshold for converging to 1 (failure) or 0 (non-failure) has to be identified. The selection of the optimal threshold entails three steps:

3a Initially a set of candidate thresholds covering the entire range between the model's minimum and maximum responses for the test data is defined. Then, the model's actual outputs are rounded (to 1 or 0 respectively) for all the values of candidate thresholds.

3b The True Positive Rate (TPR) and the False Positive Rate (FPR) are calculated for all the candidate thresholds. This iterative process provides a set of TPR/FPR pairs which are used to plot the Receiver Operating Characteristic (ROC) curve. Each point on the ROC plot (Figure 1) represents a specific TPR/FPR pair. A model with perfect discrimination has a ROC curve that passes through the upper left corner (optimal point) (Zweig and Campbell 1993). On the contrary, the closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the model is. Therefore, the most accurate curve is the C and the least accurate the A.

3c The Euclidian distance (distance between each point on the curve and the optimal point) is calculated as follows:

$$\text{Euclidian distance}=\sqrt{(1-\mathrm{TPR})^2+(\mathrm{FPR})^2} \quad (5)$$

The threshold with the minimum Euclidian distance is selected since it provides the most accurate results by minimising the false positive rate and maximising the true positive rate.

4. At the last stage, the influence of the inputs on the model's response is assessed. The analysis is performed using the following equation (Duncan et al. 2013):

$$W_{io} = W_1 * W_2 \quad (6)$$

Where: W_{io}= input-to-output influence vector; W_1= ANN hidden layer weight matrix; W_2= ANN output layer weight vector. Thus W_{io} has dimensions of $N_{in}*N_{out}$ where N_{in} is the number of inputs and N_{out}=1 is the number of output neurons

3. Evolutionary Polynomial Regression

Evolutionary Polynomial Regression (Giustolisi and Savic 2006) is a data-driven method which combines numerical and symbolic regression. The implementation of EPR returns a predefined number of models on a Pareto optimal front which is a trade-off between accuracy and parsimony. The accuracy criterion aims to maximise the model fit to the observed data and the parsimony to minimise the number of explanatory variables and/or polynomial terms in the model. The role of the parsimony rule is to prevent over-fitting of the model to data and thus capture the true underlying general phenomena (Laucelli et al. 2014). The user selects the generalised model structure and EPR employs a multi-objective search strategy to estimate the unknown parameters. The model structure selected here for analysis of pipe failure is (Giustolisi and Savic 2006):

$$Y = \sum_{j=1}^{K} a_j \left((X_1)^{E_{1j}} \ldots (X_i)^{E_{ij}}\right) + a_0 \quad (7)$$

Where: Y=predicted number of pipe failures; a_j and a_0= the constant coefficients; X_i =is the explanatory variable i, E_{ij} =the matrix of unknown exponents and k is the maximum number of polynomial terms

The candidate exponent values (E_{ij}) in Equation (7) were -2, -1, -0.5, 0, 0.5, 1 and 2 describing potential square, linear or square root exponents for explanatory variables of the EPR model. The positive and negative values were considered to describe potential direct and inverse relationship between the inputs and the output of the model while the value 0 was chosen to deselect input candidates without impact on the output. The maximum number of polynomial terms (k) was set to 1 excluding the constant term (a_0) to ensure the best fit without unnecessary complexity as the addition of new terms that fit mostly random noise in the raw data rather than explain the underlying phenomenon (Savic et al. 2009). The Least Square (LS) parameter was constrained to search for positive polynomial coefficient values only (i.e. a_j >0) because negative polynomial coefficients usually try to balance positive terms providing a better description of the noise (Giustolisi et al. 2007).

4. Model performance assessment

The performance indicator used to evaluate the accuracy of the EPR models is the Coefficient of Determination (R^2) as a measure for correlation between predictions and observations. The mathematical relationship is expressed as follows (Moriasi et al. 2007):

$$R^2=\frac{\left(\sum_{i=1}^{n}(y_{p,i}-\bar{y}_p)(y_{o,i}-\bar{y}_o)\right)^2}{\sum_{i=1}^{n}(y_{p,i}-\bar{y}_p)^2\sum_{i=1}^{n}(y_{o,i}-\bar{y}_o)^2} \tag{8}$$

Where $y_{p,i}$ = prediction value for test sample i; $y_{o,i}$ = measurement value for test sample i, $\bar{y}_p$ = mean value of predictions, $\bar{y}_o$= mean value of measurements and n = the number of test data samples

The performance of the binary model is assessed using the True Positive Rate (Eq. 9), and True Negative Rate (Eq. 10). TPR measures the proportion of correctly identified positives while TNR measures the proportion of correctly identified negatives respectively. The mathematical expressions of TPR and TNR are defined as (Kohavi and Provost 1998):

$$\text{True Positive Rate}=\frac{\text{True Positives}}{\text{True positives}+\text{False negatives}} \tag{9}$$

$$\text{True Negative Rate}=\frac{\text{True Negatives}}{\text{True negatives}+\text{False positives}} \tag{10}$$

5. Case study

The proposed methodology is demonstrated in a case study which is part of a water distribution network (WDN) of a UK city. The database contains pipe failure data between the 1st of January 2003 and the

31th of December 2013.Preliminary analysis showed that Cast Iron (CI) pipes which constitute 78% of the network's total length have the highest pipe failure rate (expressed in number of failures/km/year) which is 0.264 compared to other pipe material types which are 0.194 for Asbestos Cement (AC) pipes, 0.071 for Ductile Iron (DI) pipes, 0.030 for Polyethylene (PE) pipes and 0.113 for Polyvinyl chloride (PVC) pipes. Hence, only CI pipes are considered in this paper for construction of the predictive models. Table 1 shows their main features of the examined dataset.

Daily climate data for the case study were obtained from the British Atmospheric Data Centre and consisted of the minimum air temperature, the maximum air temperature and the soil temperature on a daily basis in 0C. To avoid negative values, all temperatures are converted to Fahrenheit.

Preliminary analysis of the data showed (Figure 2) that the majority of the failures occur during the coldest months. Therefore, in the development of the models, a particular emphasis is given to the factors that describe the severity of the cold period (i.e. freezing index).

6. Results and discussion

6.1 Results of the annual predictions approach

Following the procedure described above for the data preparation, grouping of individual pipe failure data resulted in 148 homogenous groups for developing the EPR models. Those homogenous groups were split into six clusters as shown in Figure 3. The dataset created was split into two parts for model development and validation respectively. The last year (i.e. 2013) of the monitoring period was used for validation purposes.

The implementation of the proposed methodology resulted in six EPR models each corresponding to the training data of the relevant cluster. The selected threshold for the freezing index is 0^0C degrees because it provided the highest correlation in the preliminary analysis. Table 2 lists the associated models and the coefficient of determination for the train dataset.

The relationship between the number of failures and the freezing index is a direct indication that lower temperatures and consequently higher values of the freezing index cause an increase in the number of failures. The addition of more candidate explanatory variables (e.g. minimum air temperature, maximum air temperature, soil temperature) did not increase the model's accuracy and therefore they were not selected.

Figure 4 shows the predictions vs the observations for all the clusters with the test dataset (2013). The developed models are very accurate in predicting the number of failures for clusters 1 and 6 which have the lowest number of failures. The absolute difference between observations and predictions for clusters 1 and 6 tends to zero whereas it varies between 3 and 6.5 for the rest of the clusters. The lowest error is achieved for clusters 1 and 6 which have the lowest failure rate.

The predicted number of failures was used to calculate the number of failures for all the homogenous groups (i.e. diameter, age, soil type) within the clusters proportionally to their length. Then it was assumed that all the individual pipes within the homogenous groups share the same failure rate. The individual pipe failure rates were classified using the Jenks Natural Breaks (Jenks 1963) method into five ranges as 'very low' [0-0.091], 'low' (0.091-0.236], 'medium' (0.236-0.472], 'high' (0.472-0.75] and 'very high' [greater than 0.751] as shown in Figures 5 and 6 (observations and predictions respectively).

The accuracy obtained in allocating the individual pipes in ranges is 46%, 73%, 78%, 87% and 76 % for the 'very low', 'low', 'medium', 'high' and 'very high' failure rates respectively when only weather-related factors are used. The predictions have a high accuracy for the majority of the failure ranges ('low' to 'very high'). The lowest accuracy is achieved for the pipes with a 'very low' observed failure rate. This low accuracy can be attributed to the fact that a number of homogenous groups of pipes have experienced zero number of failures. The predicted failures for each cluster are distributed to the homogenous groups proportionally to their length value leading to a slight overestimation for those groups.

6.2 Results of Combined weather and pipe-intrinsic factors based approach

The outputs of the proposed method are used in conjunction with the results of an approach which calculated the average failure rate for the entire monitoring period using pipe-intrinsic factors as explanatory variables (Kakoudakis et al. 2017) as shown in Figure 7. The final failure rate of the individual pipes is the combination of the two values.

The examined WDN consists of a big number of individual pipes and the improvement achievement is highlighted in Figure 8. Figure 8 compares the accuracy of the predictions when only environmental variables are used and when they are combined with the physical variables. The inclusion of the physical factors increased the accuracy of the predictions for the majority of the ranges. The highest improvement is observed for the 'very low' range for which shifted to 69%.

6.3 Results for the short-term predictions

Following the approach described in the methodology section, the data preparation resulted in 3653 data samples, 56.20% of which correspond to cases without failure(s) and the remaining 43.80% to failure(s). The model's responses were compared to a set of threshold values and the generated pairs of TPR/FPR were used to plot the ROC curve (Figure 9). The selected threshold with the lowest Euclidean distance from the optimal point is 0.538. As shown in Figure 9 the majority of the non-failures are correctly identified (the FPR is 0.87) as such while a similar conclusion can be derived for the failures despite the lower accuracy (the TPR is 0.72). The value of AUC which is used as a measurement of model's performance is 0.814 indicating that the model has a good accuracy.

The influence of the inputs on the model's response is assessed and they are ranked to identify the most influential. The result of the analysis input's influence on the model's responses is a column matrix with the weight of all the inputs (Table 3). The FI is shown to be the most influential factor. This observation is linked to the fact that the majority of the failures occur in the coldest months (also shown in Figure 2) when pipes are subject to frost actions which is a cause for axial causes on them (Rajani et al. 1996). The frost imposes additional load on the buried pipes and is influenced by frost penetration, trench width, soil type, soil stiffness, frost heave of trench fill and side fill as well as the interaction at the trench backfill–side fill interface (Rajani and Zhan 1996). The negative values indicate a reverse relationship between these variables and the occurrence of pipe failure(s).

7. Conclusions

This paper presents a method to predict the occurrence of pipe failure and its annual variation due to weather factors. Only CI pipes were considered due to their highest failure rate in the network. However, it can be applied to other pipe materials as well. For the annual predictions the individual pipes were allocated into homogenous groups. The created homogenous groups were then split into a predefined number of clusters and an individual EPR model was developed for each cluster using weather conditions as explanatory variables. The FI was selected as the most influential variable by the models. The mathematical relationship obtained between the number of failures and the FI is a direct indication that lower temperatures and consequently higher values of the FI cause an increase in the number of failures. The outputs of the proposed method were used in conjunction with a previous approach which calculated the average failure rate of the entire monitoring period using pipe-intrinsic explanatory variables in order to improve the quality of predictions. The final failure rate was calculated as the average failure rate of the two approaches which resulted in more accurate predictions. The highest improvement was achieved for the pipes with a ''very low'' observed failure rate.

The method for predicting the occurrence of failure(s) was implemented using an ANN binary model and was shown to be able to distinguish between the days with and without failure(s). The influence of inputs on the models' output responses was assessed showing that low temperatures have a strong influence. This approach can be used operationally to alert water utilities to manage pipe failures reducing potential water loss, associated costs and service disruption to consumers. Further research should be done to associate the short-term prediction of failure(s) with specific pipes.

Acknowledgments

The work reported is supported by the UK Engineering & Physical Sciences Research Council (EPSRC) project Safe &SuRe (EP/K006924/1).

References

Clair St, A. M. and Sinha, S., 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models. *Urban Water Journal*, **9**(2), 85-112.

Duncan, A., Tyrrell, D., Smart, N., Keedwell, E., Djordjevic, S., and Savic, D. 2013. Comparison of machine learning classifier models for bathing water quality exceedances in UK. *Proceedings of the*

35th International Association for Hydro-Environment Engineering and Research (IAHR 2013) Congress, Chengdu, China

Fuchs-Hanusch, D., Friedl F., Scheucher R., Kogseder B. and Muschalla D., 2013. Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions. *Water Science & Technology: Water Supply,* **13**(2), 435-446.

Giustolisi, O. and Savic, D. A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics,* **8**(3), 207-222.

Giustolisi, O., Doglioni, A., Savic, D. A., and Webb, B. W., 2007. A multi-model approach to analysis of environmental phenomena. *Environmental Modelling and Software*, **22**(5), 674-682.

Giustolisi, O. and Savic, D. A., 2009. Advances in data-driven analyses and modelling using EPR-MOGA. Special Issue on Advances in Hydroinformatics. *Journal of Hydroinformatics***, 11**(3-4), 225–236.

Giustolisi, O., Savic, D. and Laucelli, D. 2009. Asset deterioration analysis using multi-utility data and multi-objective data mining. *Journal of Hydroinformatics,* **11**(3-4), 211–224.

Gould, S. J. F., Boulaire F. A., Burn S., Zhao X. L. and Kodikara J. K., 2011. Seasonal factors influencing the failure of buried water reticulation pipes. *Water Science & Technology,* **63**(11), 2692–2699.

Jenks, G. F. 1963. Generalization in statistical mapping. *Annals of the Association of American Geographers,* **53**(1), 15-26.

Kakoudakis K., Behzadian K., Farmani R. and Butler D., 2017. Pipeline Failure Prediction in Water Distribution Networks using Evolutionary Polynomial Regression combined with K-means clustering. *Urban Water Journal,* **14**(7), 737-742

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

Kleiner, Y. and Rajani B., 2002. Forecasting variations and trends in water-main breaks. *Journal of Infrastructure Systems*, **8**(4), 122-131.

Kleiner, Y. and Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *Journal of Hydroinformatics,* **14**(3), 659-681.

Kohavi, R. and Provost, F. 1998. Glossary of terms. *Machine Learning,* **30**(2-3), 271-274.

Laucelli, D., Rajani, B., Kleiner, Y. and Giustolisi, O., 2014. Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. *Journal of Hydroinformatics*, **16** (4), 743-757.

Makar, J. M., 2000. A preliminary analysis of failures in grey cast iron water pipes. *Engineering Failure Analysis,* **7**(1), 43-53.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the Asabe,* **50**(3), 885-900.

Nishiyama, M. and Y. Filion (2013). Review of Statistical Water Main Break Prediction Models: 2002-2012. *Canadian Journal of Civil Engineering,* **40**(10), 972-979.

Rajani, B. and C. Zhan (1996). On the estimation of frost loads. *Canadian geotechnical journal*, **33**(4), 629-641.

Rajani, B., C. Zhan and S. Kuraoka. (1996). Pipe–soil interaction analysis of jointed water mains. *Canadian Geotechnical Journal*, **33**(3), 393–404.

Rajani, B., Kleiner Y. and Sink J., 2012. Exploration of the relationship between water main breaks and temperature covariates. *Urban Water Journal,* **9**(2), 67-84.

Sadiq, R., Rajani, B. and Kleiner, Y., 2004. Fuzzy-based method to evaluate soil corrosivity for prediction of water main deterioration. *Journal of Infrastructure Systems*, **10**(4), 149–156

Wols, B. A. and van Thienen P., 2013. Impact of weather conditions on pipe failure: a statistical analysis. *Journal of Water Supply: Research and Technology,* **63**(3), 212-223.

Zweig, Mark H. and Gregory Campbell, 1993. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry,* **39**(4), 561-577.

List of Figures

List of Tables