  Journal of Hydroinformatics | 11.1 | 2009

Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling

M. Tabesh, J. Soltani, R. Farmani and D. Savic

ABSTRACT

In this paper two models are presented based on Data-Driven Modeling (DDM) techniques (Artificial Neural Network and neuro-fuzzy systems) for more comprehensive and more accurate prediction of the pipe failure rate and an improved assessment of the reliability of pipes. Furthermore, a multivariate regression approach has been developed to enable comparison with the DDM-based methods. Unlike the existing simple regression models for prediction of pipe failure rates in which only few factors of diameter, age and length of pipes are considered, in this paper other parameters such as pressure and pipe depth, are also included. Furthermore, an investigation is carried out on most commonly used mechanical reliability relationships and the results of incorporation of the proposed pipe failure models in the reliability index are compared. The proposed models are applied to a real case study involving a large water distribution network in Iran and the results of model predictions are compared with measured pipe failure data. Compared with the results of neuro-fuzzy and multivariate regression models, the outcomes of the artificial neural network model are more realistic and accurate in the prediction of pipe failure rates and evaluation of mechanical reliability in water distribution networks.

Key words | artificial neural network, mechanical reliability, multivariate regression, neuro-fuzzy system, pipe failure rate, water distribution networks

M. Tabesh (corresponding author)
Center of Excellence for Engineering and Management of Infrastructures,
Department of Civil Engineering,
University of Tehran,
PO Box 11155-4563,
Tehran,
Iran
Tel.: +98 21 6111 2258
Fax: +98 21 6646 1024
E-mail: *mtabesh@ut.ac.ir*

J. Soltani
Department of Civil Engineering,
University of Tehran, Tehran,
Iran

R. Farmani
D. Savic
Department of Engineering, School of Engineering,
Computer Science and Mathematics,
University of Exeter,
Exeter EX4 4QF
UK

NOMENCLATURE

$\lambda\ (t)$	the failure rate per year t (based on number of failures/yr/km (or mile))
a	the growth rate (1/yr)
λ	the pipe failure rate
D_l	the pipe diameter in (inches) or (mm)
N_p	the number of existing pipes in the network
(A_l)	the availability parameter of component l in the network (including pipes)
β_1	number of pipe failures per unit of time that is obtained as $\beta_l = L_l{}^* \lambda$
L_1	the length of pipe l in miles or km
α_1	number of expected repairs for the lth pipe per unit of time
MTBF	mean time between failures (duration of connectivity)
MTTR	mean time to repair (duration of disconnection and repair)
y_{actual}	the actual (observed) data
$y_{prediction}$	the predicted data
$y_{average}$	the average of data
RMSE	root of mean squared error
IOA	the index of agreement
n	the number of observations (a real number)
P_l	the hydraulic pressure of pipe l
H_l	the depth of installation of pipe l
Ag_l	the age of pipe l
D_{max}, L_{max}, P_{max}, H_{max}, Ag_{max}, λ_{max}	the maximum values of diameter, length, pressure, depth, age and break rate for asbestos cement pipes in the district, respectively.

doi: 10.2166/hydro.2009.008

w_1, w_2, w_3 the weight matrices

b_1, b_2, b_3 bias values obtained from the neural network for different layers of the selected model

gaussmf, gauss2mf, gbellmf, trapmf and trimf hybrid and different types of studied membership functions

ALPNLRFR availability (AL) from Poisson formula (P) using nonlinear multivariate regression method (NLR) for failure rate (FR)

ALPANNFR availability (AL) from Poisson formula (P) using artificial neural network model (ANN) for failure rate (FR)

ALPANFISFR availability (AL) from Poisson formula (P) using neuro-fuzzy model (ANFIS) for failure rate (FR)

ALPSFR availability (AL) from Poisson formula (P) using Su formula (S) for failure rate (FR)

ALKANNFR availability values from Khomsi formula with the failure rates extracted from the relation and model of ANN

ALKANFISFR availability values from Khomsi formula with the failure rates extracted from the relation and model of ANFIS

ALKSFR availability values from Khomsi formula with the failure rates extracted from the relation and model of Su formula (S)

ALKNLRFR availability values from Khomsi formula with the failure rates extracted from the relation and model of NLR

ALKANNFR availability values from Fujiwara & Tung formula with the failure rates extracted from the relation and model of ANN

ALKANFISFR availability values from Fujiwara & Tung formula with the failure rates extracted from the relation and model of ANFIS

ALKSFR availability values from Fujiwara & Tung formula with the failure rates extracted from the relation and model of Su formula (S)

ALKNLRFR availability values from Fujiwara & Tung formula with the failure rates extracted from the relation and model of NLR

INTRODUCTION

The main task of water distribution networks is to supply water to consumers (domestic, commercial and industrial) in the required quantity, quality and pressure. Reliability indicators are used to evaluate the efficiency of water distribution networks in providing water with standard quality, sufficient quantity and within the appropriate pressure range to consumers under different operational (normal and abnormal) conditions such as component failure and hydraulic changes (Farmani *et al.* 2005). Reliability of water distribution networks relates to two types of failure, mechanical failure of system components and hydraulic failure caused by changes in demand and pressure head (Tabesh 1998). Accidents, and especially failures of pipes in urban water distribution networks, lead to financial and capital losses for repair and restoration of the network. Failures reduce the reliability of the network due to lowering of the pressure or due to interruption of the water supply in parts of the distribution network, which ultimately leads to dissatisfaction of customers. Sensitive customers such as industrial centers, governmental buildings, hospitals, etc., are most likely to be affected.

To evaluate the mechanical reliability of a water distribution system, a relationship should be established between pipe failures and other parameters of the system. In general, numerous factors such as age, diameter, material, corrosion, quality of pipe material, installation conditions, operational conditions and traffic contribute to accidents and mechanical failure of pipes. Among the parameters that affect pipe failure, only some of them are measurable such as age, length, diameter, depth and pressure. The most often applied formulae for estimating the pipe failure rate have been obtained using simple regression models on the available pipe failure data from a limited time period. These relationships include only a number of influential parameters that affect pipe failure, e.g. age (Shamir & Howard 1979), age and diameter (Kettler & Goulter 1985; Giustolisi *et al.* 2006), diameter (Kettler & Goulter 1983; Su *et al.* 1987; Goulter & Kazemi 1988, 1989; Mays 1989; Cullinane *et al.* 1992; Goulter *et al.* 1993; Tabesh & Abedini 2005), climatic conditions (Harada 1988; Sacluti 1999; Welter 2001; Ahn *et al.* 2005). Consideration of all available parameters should lead to more realistic failure

Journal of Hydroinformatics | 11.1 | 2009

rate predictions. In view of the characteristics and capabilities of data-driven methods and their ability to include a large number of parameters involved in complex phenomena, there has been a great deal of interest among researchers and practitioners to use this type of model. Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) are among the DDM techniques which have the ability to capture the complex and nonlinear relationship between different variables of a system by learning from data (Demuth & Beale 2002). Since there have been numerous applications of these techniques in various fields of engineering, use of DDM such as ANN and ANFIS would be useful for prediction of pipe failure rates. There are very few publications on the applications of data-driven models for pipe failure prediction (e.g. Sacluti 1999; Ahn *et al.* 2005). On the other hand, a few relationships have been developed to calculate mechanical reliability of pipe systems (Su *et al.* 1987; Fujiwara & Tung 1991; Cullinane *et al.* 1992; Khomsi *et al.* 1996). All these formulae involve pipe failure rate but, as will be seen in this paper, they produce different results.

The main objective of this paper is to investigate the potential of neural networks and neuro-fuzzy systems to predict pipe failure rates using a range of measurable parameters of the system such as pipe age, diameter, depth, length and pressure. A multivariate regression model is also constructed with these parameters. To evaluate different mechanical reliability relationships the outputs of three above-mentioned models for pipe failure prediction are incorporated in a number of commonly used mechanical reliability relationships and the results are compared. The outcomes are used to classify the available reliability measures and propose a criterion to use these indices more appropriately.

PIPE FAILURE INDICATORS

Pipe breaks are a type of mechanical failure of the system and are considered as one of the significant factors contributing to water losses. Pipe failure imposes huge direct and indirect economic losses and requires human capital for the restoration and repair of the networks (Dandy & Engelhardt 2001). Failure rate (the number accidents per year and per pipe length unit) can be used as a performance indicator (Tabesh & Abedini 2005). Various researchers have carried out investigations into the analysis of mechanical failure and prediction of pipe failure rate based on a limited number of parameters involved. These approaches usually result in a set of formulae derived using statistical methods or regression models (Lei & Saegrov 1998). This has led to considerable differences between the results of these models.

Shamir & Howard (1979) presented an exponential model for prediction of the pipe failure rate based on time:

$$\lambda(t) = \lambda(t_0)e^{a(t-t_0)} \tag{1}$$

where $\lambda(t)$ is the failure rate per year t (based on number of failures/yr/km (or mile)), t_0 is the year of analysis, $\lambda(t_0)$ is the failure rate at t_0 and a is the growth rate (1/yr). The authors suggested that a coefficient varies from 0.05 to 0.15 based on the material and diameter of pipes. Kettler & Goulter (1985) carried out studies to express the variations in the number of annual breaks as a function of age and diameter of cast iron and asbestos pipes in Winnipeg (Canada). According to this research, the trend of age variations in both cases is linear and shows the same increase in number of failures at each year. Su *et al.* (1987) proposed the following relation for the pipe failure rate (number of annual failures per mile):

$$\lambda = \frac{0.6858}{D_l^{3.28/}} + \frac{2.7158}{D_l^{1.3131}} + \frac{2.7658}{D_l^{3.5792}} + 0.42 \quad \forall l = 1, \ldots, N_p \tag{2}$$

where λ is the pipe failure rate, D_l is the pipe diameter in (inches) and N_p is the number of existing pipes in the network.

Ahn *et al.* (2005) presented a procedure based on ANN to predict the pipe failure rate in the water distribution network in Seoul City, South Korea, considering the variation of failures in pipes, against soil, water and air temperatures. According to their model, in autumn and spring when the temperature of water and soil changes, the number of failures in pipes increases. Tabesh & Abedini (2005) studied and analyzed pipe failure rates of water supply networks in several cities in Iran and discovered some relationships between the number of breaks per year and the diameter and age of different pipe types. All the above-mentioned studies were based on limited data and very few important parameters were available and were incorporated in the models obtained.

PIPE AVAILABILITY INDICATORS

According to research carried out so far, the mechanical reliability factor in a water network represents the availability parameter (A_l) of the component l in the network (including pipes). Some of the conventional methods are described below.

Su et al. (1987) proposed the following relation for calculating the pipe availability in the water distribution network based on Poisson's probability distribution:

$$A_l = e^{-\beta_l} \quad \forall l = 1, \ldots, N_p \tag{3}$$

where β_1 is the number of pipe failures per unit of time that is obtained as $\beta_l = L_l{}^{*}\lambda$ and L_1 is the length of pipe l in miles.

Fujiwara & Tung (1991) presented the following relationship for calculating pipe availability in the water distribution network:

$$A_l = \frac{\alpha_l}{\alpha_l + \beta_l} \quad \forall l = 1, \ldots, N_p \tag{4}$$

in which α_1 = number of expected repairs for the lth pipe per unit of time. The value of α_l is obtained by dividing the number of annual breaks in pipes of special material by the number of days in a year (365).

Finally they proposed the following relationship for pipe reliability:

$$A_l = \frac{0.64}{[0.64 + L_l(0.005485 - 0.0000175 D_l)]}$$
$$\forall\, l = 1, \ldots, N_p \tag{5}$$

where L_l and D_l are in (km) and (mm), respectively, and day is used as the unit of time.

Following the definition of Ang & Tang (1984), the probability (A_l) of the operational state of link (pipe) l can be represented as

$$A_l = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \tag{6}$$

in which MTBF = mean time between failures (duration of connectivity) and MTTR = mean time to repair, i.e. duration of disconnection and repair.

Using the datasets of Mays (1989) and Walski & Pelliccia (1982), the following relationship for pipe availability was obtained by Cullinane et al. (1992):

$$A_l = \frac{0.21218 D_l^{1.462131}}{(0.00074 D_l^{0.285} + 0.21218 D_l^{1.462131})}$$
$$\forall l = 1, \ldots, N_p \tag{7}$$

where D_1 is the diameter of pipe in inches. Since the pipe failure rate is not directly incorporated into this formula it cannot demonstrate variations of availability with pipe break rate.

Khomsi et al. (1996) presented the following relation for calculating the pipe availability in the water distribution network:

$$A_l = 1 - \frac{\lambda^{*} L_l}{365} \quad \forall l = 1, \ldots, N_p \tag{8}$$

where L_1 is the length of the pipe (in km).

APPLICATION OF DATA-DRIVEN TECHNIQUES IN PIPE FAILURE ANALYSIS

The purpose of analyzing accidents and breaks is to find the relationship between relevant indices and characteristics of the pipes and to use this relationship to compute mechanical reliability. In order to carry out a comprehensive analysis, information such as material, diameter, length, age, installation depth of pipes and operation conditions (e.g. hydraulic pressure) should be collected and stored in a database.

Artificial neural network model

Artificial neural networks (ANNs) are essentially parametric regression estimators and are well suited for the purpose of this research, as they can approximate virtually any (measurable) function up to any arbitrary degree of accuracy (Hornik *et al.* 1989). A neural network model with the perceptron structure is composed of a number of layers (usually three layers) and each layer includes a number of processing units called neurons. A neuron can be a nonlinear mathematical function. As a result, a neural network composed of an aggregation of these neurons can also be a fully complicated nonlinear system. In a neural network, each neuron acts independently and the overall behavior of the network is the outcome of local behaviors of numerous neurons. This makes local errors less influential in the output. In other words, the neurons correct one another in a cooperation process. This property increases the durability of the ANN. One neuron generates a special output quantity based on a number of different inputs with the use of an activation function to produce the outgoing signal of the node (Karamouz *et al.* 2007).

Neural networks are dynamic systems which have the ability to capture the relationship between input and output parameters of a system by learning from experimental data. They learn general rules based on the numerical data. In this research, a backpropagation artificial neural network with a multilayer perceptron structure is used.

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) model

A neuro-fuzzy system is a combination of the logical functions of fuzzy systems and neural networks. Neuro-fuzzy systems have the potential to combine the benefits of these two fields, i.e. their hybrid training methodology, in a single framework (Abdel-Hamid *et al.* 2007). This method eliminates the basic problem in fuzzy system design (obtaining a set of fuzzy if–then rules) by effectively using the learning capability of an ANN for automatic fuzzy if–then rule generation and parameter optimization. Different structures have been proposed for implementing a fuzzy system by neural networks. One of the most powerful structures developed by Jang & Guley (1996) is a neuro-fuzzy network system known as ANFIS. The basic idea behind these neuro-adaptive learning techniques is very simple. These techniques provide a method for the fuzzy modeling procedure to learn information about a dataset, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. This learning method works similarly to that of neural networks (Jang & Guley 1996). One of the main applications of ANFIS is its use in modeling and control of complicated systems. In general, any phenomenon which allows for recording of a set of behavioral observations can be modeled by this method. The very important characteristic of this type of system is that they do not need any mathematical formula or model for design. Therefore, they seem to be useful and appropriate for the design of systems whose functions cannot be expressed explicitly in the form of mathematical models. In order to develop a fuzzy model the number of membership functions, inputs and outputs, values of condition parameters and result parameters should be specified first.

Models developed in this research were coded in MATLAB (Ver. 7.04). The developed ANN and ANFIS models take as input five parameters including pipe diameter, length, age, depth of installation and hydraulic pressure. The output of the models is the pipe failure rate. Assessment of the quality of data-driven models is one of the major procedures in this type of modeling, as many models might be generated and trained. The assessment of a model is meant to show to what level the model is capable of providing an acceptable response to the new inputs in regard to the training it has received. The root of mean squared error (RMSE) and the index of agreement (IOA) are used as assessment criteria of the constructed data-driven models in this paper:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{\mathrm{actual},\,i} - y_{\mathrm{prediction},\,i})^2} \tag{9}$$

$$\mathrm{IOA} = 1 - \frac{\left[\sum_{i=1}^{n}|y_{\mathrm{prediction},i} - y_{\mathrm{actual},i}|^2\right]}{\left[\sum_{i=1}^{n}(|y_{\mathrm{prediction},i} - y_{\mathrm{avarege},i}| + |y_{\mathrm{actual},i} - y_{\mathrm{avarege},i}|)^2\right]} \tag{10}$$

where y_{actual} is the actual (observed) data, $y_{\mathrm{prediction}}$ is the predicted data, y_{average} is the average of data and n is the number of observations (Demuth & Beale 2002).

CASE STUDY

To evaluate the proposed methodology a part of a water distribution network of a city in Iran is considered as the study area (see Figure 1). The area of this district is 2,418 ha and covers 93,719 properties with 579,860 m of distribution pipes, including steel pipes 800, 700 and 600 mm in diameter and asbestos cement and cast iron pipes 400, 300, 250, 200, 150, 100 and 80 mm in diameter. The installation and execution of the network pipelines in this area generally started in 1981.

At the moment, due to considerable topographic differences (1,021–1,214 m above sea level), the highest rate of pipe failures is recorded in this district. It should be pointed out that, according to the existing reports and statistics prepared by the local water and waste water company, the highest rate of events and accidents in main pipelines are recorded on pipes (especially asbestos cement pipes) with diameters less than 300 mm and the available data for cast iron and steel pipes is not enough to be used by data-driven models (Aghayee 2006). A large number of parameters that contribute to pipe failure, e.g. diameter, length, age, depth of installation and average hydraulic pressure of asbestos pipes with diameters of 80–300 mm, were considered in development of the relationship for pipe failure rate. The pipe data collected for asbestos pipes in the range of 80–300 mm includes 337 cases for a period of one year. Eighty percent of the data were used for training of the network, 15% for testing and 5% for verification of the results.

Modeling of pipe failure rate

Artificial neural network model

In this research, a number of neural network structures were prepared and tested by varying the number of layers, neurons, activation functions and epochs (500–10,000). Table 1 shows details of some of the main structures categorized into 9 groups. Figures 2 and 3 represent different assessment criteria (Equations (9) and (10)) for each case. Finally, considering some criteria such as lower error indicator values, robustness of multiple-layer networks (Demuth & Beale 2002) and evaluation of epochs with the lowest error values during verification stage for different groups of Table 1, one of the structures of case 5, including two hidden layers with 5 and 10 neurons, was selected as the most appropriate one. The methodology presented in this paper is seen as the first phase of an investigation, with the second phase incorporating the obtained pipe break rates into an optimization program to evaluate the influence of changes in pipe parameters on the pipe failure rate. Therefore, a combination of several criteria

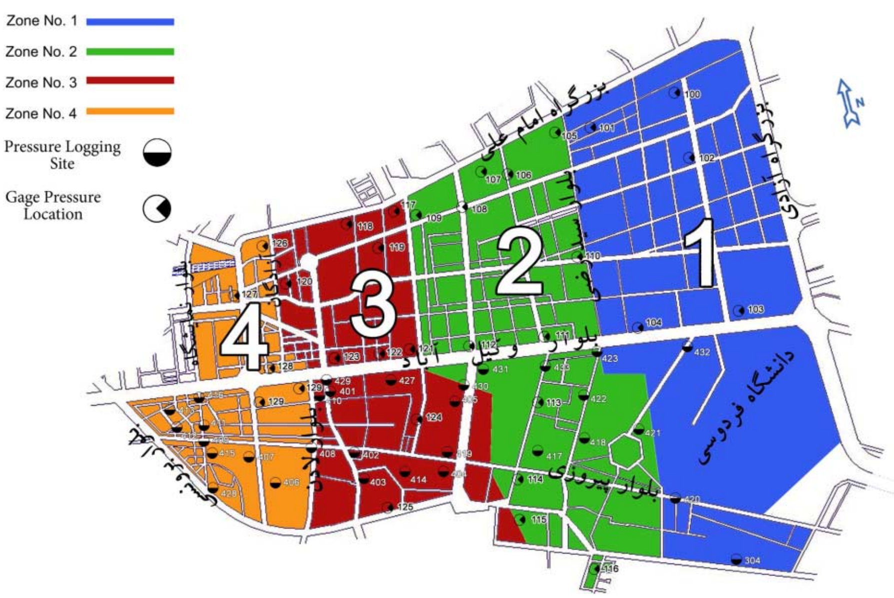

Figure 1 | Schematic of study area and pressure measurement points.

Table 1 | Specifications of different constructed ANN models

	Type of activation functions of each hidden layer			Number of neurons in each hidden layer				
Case no.	1st layer	2nd layer	Output layer	1st layer	2nd layer	Output layers	Number of epochs	Run time (s)
1	Tangsig	–	Pureline	5	–	1	500	16.29
							1,000	39.69
							2,000	148.2
							5,000	908.88
							10,000	3535.03
2	Tangsig	–	Pureline	10	–	1	500	14.8
							1,000	52.14
							2,000	158.13
							5,000	1006.63
							10,000	3609.75
3	Tangsig	–	Pureline	15	–	1	500	13.79
							1,000	46.91
							2,000	162.67
							5,000	922.96
							10,000	4099.82
4	Tangsig	Tangsig	Pureline	5	5	1	500	16.03
							1,000	49.83
							2,000	181.91
							5,000	1132.84
							10,000	3213.92
5	Tangsig	Tangsig	Pureline	5	10	1	500	10.33
							1,000	35.72
							2,000	129.20
							5,000	807.98
							10,000	3124.95
6	Tangsig	Tangsig	Pureline	5	15	1	500	10.53
							1,000	37.78
							2,000	166.88
							5,000	899.72
							10,000	3159.42
7	Tangsig	Tangsig	Pureline	10	5	1	500	37.22
							1,000	60.25
							2,000	204.66
							5,000	851.89
							10,000	4429.76
8	Tangsig	Tangsig	Pureline	10	10	1	500	16.10
							1,000	50.87
							2,000	185.84
							5,000	827.29
							10,000	4009.17

Table 1 | (continued)

	Type of activation functions of each hidden layer			Number of neurons in each hidden layer				
Case no.	**1st layer**	**2nd layer**	**Output layer**	**1st layer**	**2nd layer**	**Output layers**	**Number of epochs**	**Run time (s)**
9	Tangsig	Tangsig	Pureline	10	15	1	500	48.82
							1,000	46.88
							2,000	199.53
							5,000	952.50
							10,000	3682.10

was used to choose the appropriate structure for the ANN model. The results show that case 6 with 5000 epochs obtained the lowest error values for all three stages (training, testing and verification), but with a run time of 900 s. However, case 5 with 500 epochs, which is second-best with respect to error values, requires just about 10 s. The other parameters of the chosen structure include tangent sigmoid and linear activation functions, five input parameters and one output. The tangent sigmoid and linear activation functions which were most often used for ANN are expressed as follows (Demuth & Beale 2002):

$$y = \tan\text{sig}(n) = 2/(1 + e^{-2n}) - 1 \tag{11}$$

$$y = \text{pureline}(n) = n \tag{12}$$

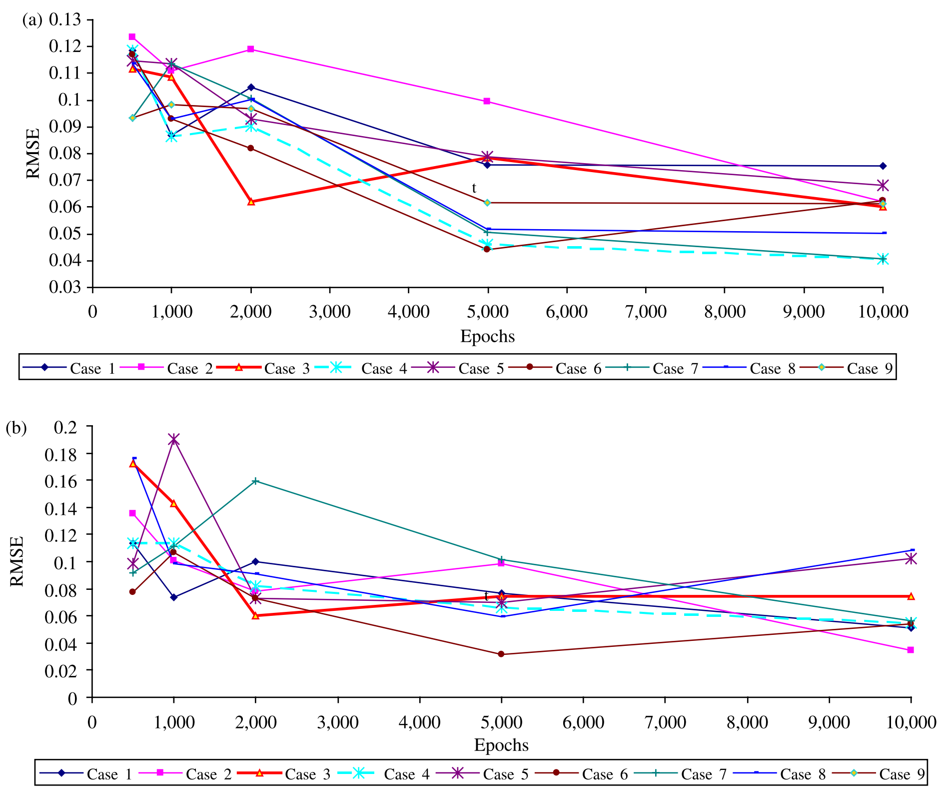

Figure 2 | (a) Variations of the RMSE values of ANN models based on different values of epochs at training stage. (b) Variations of the RMSE values of ANN models based on different values of epochs at verification stage.

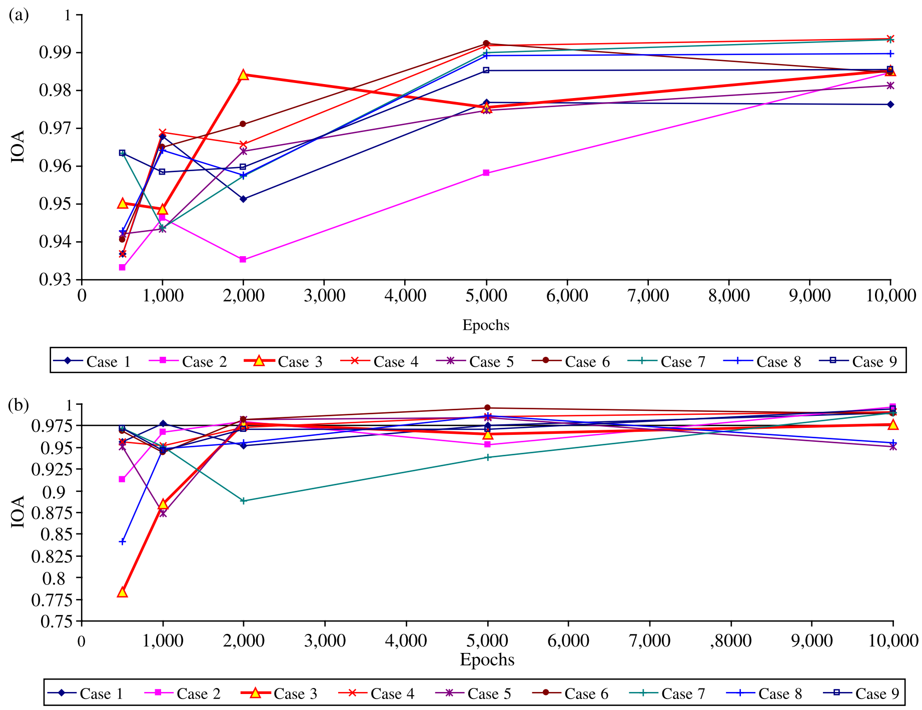

Figure 3 | (a) Variations of the IOA values of ANN models based on different values of epochs at training stage. (b) Variations of the IOA values of ANN models based on different values of epochs at verification stage.

where n is a real number, but y is bounded between -1 and 1 (for tansig) and $-\infty$ and $+\infty$ (for pureline).

After selecting the most appropriate structure for the artificial neural network, its applicability and efficiency should be tested. The results of the ANN predictions as well as actual values in the three stages of training, testing and verification are presented in Figure 4. Comparing the observed and simulated values in all these stages indicate the appropriateness of the selected model. As can be seen, the predicted results are very close to the observed data. In order to compare the simulated and observed results, the simulated failure rates for the selected neural network and the observed data are plotted in Figure 5. This figure shows that the ANN model was able to capture and learn the existing trends and behavior of the data with very good accuracy and to generalize the training to different unseen cases.

Finally, the failure rate equation of asbestos cement pipes for the study area is obtained from the following relationship based on the pipe diameter, age, length, depth of installation and hydraulic pressure:

$$\lambda = \text{pureline}(w_3(\text{tangsig}(w_2(\text{tansig}(w_1 \times (\text{Test DataIn}) + b_1) + b_2) + b_3) \times \lambda_{\max} \tag{13}$$

With substitution of the test parameters the formula below is obtained:

$$\lambda = purline\left(w_3 \left(tan\,sig \left(w_2 \left(tan\,sig \left(w_1 \begin{bmatrix} D_l/D_{max} \\ L_l/L_{max} \\ P_l/P_{max} \\ H_l/H_{max} \\ Ag_l/Ag_{max} \end{bmatrix} + b_1 \right) \right) + b_2 \right) \right) + b_3 \right) \lambda_{\max} \tag{14}$$

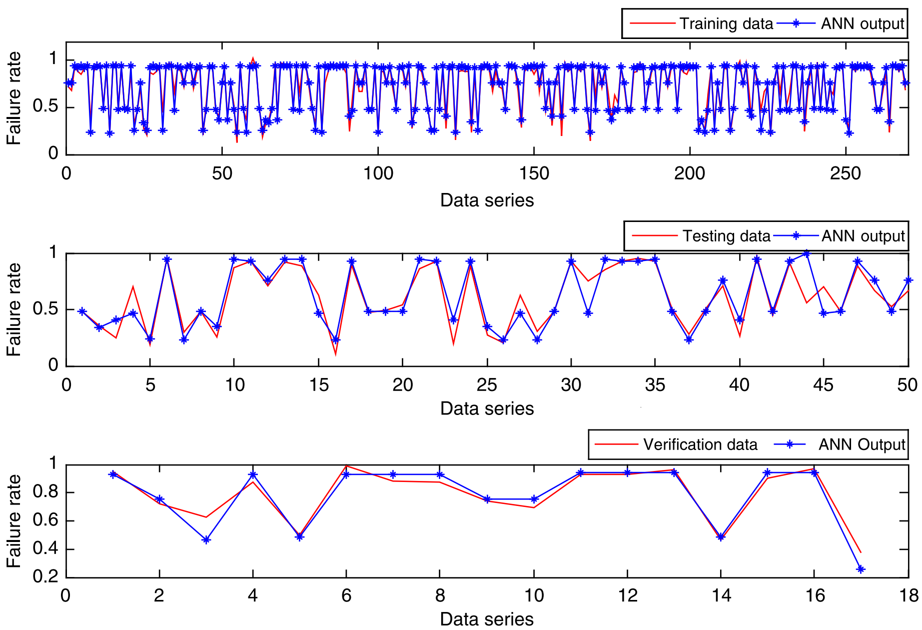

Figure 4 | Results of the ANN model for simulated and actual values in three stages of training, test and verification.

where λ is the failure rate, D_l is the diameter, L_l is the length, P_l is the hydraulic pressure, H_l is the depth of installation, Ag_l is the age of pipe l and $D_{\max}$, $L_{\max}$, $P_{\max}$, $H_{\max}$, $Ag_{\max}$ and $\lambda_{\max}$ are the maximum values for asbestos cement pipes in the district, respectively. w_1, w_2 and w_3 are the weight matrices and b_1, b_2 and b_3 are the bias values obtained from the neural network for different layers of the selected model, the values of which are presented in Table 2.

Neuro-fuzzy model

In this study, in order to choose the appropriate neuro-fuzzy structure, the ANFIS model is implemented for a series of membership functions with different epochs

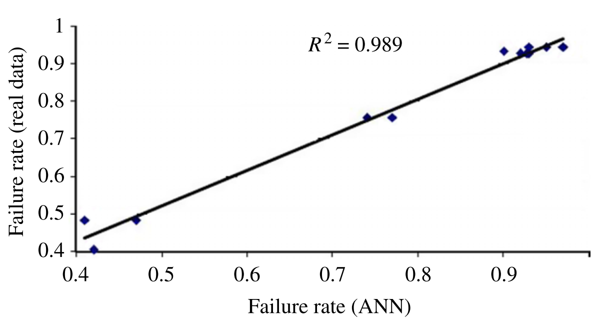

Figure 5 | Comparison of failure rate predicted by the ANN model and the actual data.

ranging from 10–150 and the results are compared by the error indicators. The outcomes are presented in Table 3. The final specifications of the ANFIS model structure considered for this research are as follows:

Number of inputs: 5 parameters, number of outputs: 1, number of input data pairs for training: 270, number of input data pairs for testing: 50, number of input data pairs for verification: 17, size of epochs: 10–150, type of optimization method of membership function: hybrid and different types of studied membership functions include gauss2mf, gaussmf, gbellmf, trapmf and trimf. More detailed information related to the membership functions can be seen in Jang & Guley (1996).

As this table shows, the gaussmf membership function provides the best result with most error indicators. In further stages of the work, this model was subjected to test and verification for the accuracy of its application and efficiency. The outcomes of all three stages are presented in Figure 6.

Comparison between the observed and simulated values in all three stages of training, testing and verification of the network indicates the appropriateness of the selected model for the concerned study area. As can be seen, the predicted results are very close to the observed data, which indicates the proper training and capability of the network. The results of application of the ANFIS model in predicting

Table 2 | Weight matrices and bias values for different layers of the proposed ANN model

Type of variable	Input parameters	Number of neurons in the hidden layers 1	2	3	4	5	6	7	8	9	10
W_1 [5 * 5]	1	0.1821	0.3492	5.02	2.7446	1.0163	–	–	–	–	–
	2	−2.7328	−1.1073	−3.0463	3.214	−2.4476	–	–	–	–	–
	3	−3.3495	0.5907	4.5567	0.5471	−2.0669	–	–	–	–	–
	4	−2.3838	0.3738	−1.4264	−3.387	4.9999	–	–	–	–	–
	5	−2.7966	2.4204	−0.7034	0.1411	4.2703	–	–	–	–	–
W_2 [10 * 5]	1	−1.0711	−1.9987	−0.021	0.1251	0.771	–	–	–	–	–
	2	0.7703	−1.1975	0.9115	0.9512	1.0315	–	–	–	–	–
	3	−0.8422	−0.0554	−0.0637	−1.3451	−1.5443	–	–	–	–	–
	4	1.2972	0.0562	1.1455	−1.2958	0.4394	–	–	–	–	–
	5	−0.3427	1.7843	0.478	−0.5062	−1.069	–	–	–	–	–
	6	0.3397	0.7869	−1.3985	0.2817	1.3806	–	–	–	–	–
	7	0.2811	−1.1855	−1.291	1.3663	−0.1556	–	–	–	–	–
	8	0.3111	0.2936	−0.1149	−2.3224	0.2974	–	–	–	–	–
	9	−1.5557	0.7187	−0.5576	−0.6852	0.1246	–	–	–	–	–
	10	−0.3242	−0.7931	1.7768	−1.3798	0.1656	–	–	–	–	–
W_3 [10 * 1]	1	−0.6567	−0.7492	0.0345	−0.0391	0.0664	0.0555	0.045	0.1319	−0.7154	0.342
b_1[5 * 1]		−8.9721	5.6689	−1.2129	0.4981	−4.8612	–	–	–	–	–
b_2 [10 * 1]		2.0939	−1.8358	1.275	−0.8041	0.2468	0.2216	0.5657	0.9569	−1.6516	−2.1249
b_3 [1]		0.4895	–	–	–	–	–	–	–	–	–

failure rate of water supply network pipes are compared with each other using different error indicators and presented in Figure 7. This figure clearly shows that the results are very close to the observed data and have very minor errors.

Nonlinear regression model (NLR)

To check the performance of regression methods in comparison with DDM approaches in reliability calculations, a multivariate nonlinear regression model was built to predict failure rate of the water distribution network pipes in the study area. The failure rate equation for asbestos pipes was obtained as a function of diameter, length, age, depth of installation and average hydraulic pressure as:

$$\lambda = -0.4197(D_l^{0.3762}) + 0.4168(L_l^{0.0872}) + 0.2813(P_l^{0.5668}) + 0.0903(H_l^{-1}) + 0.7408(Ag_l^{0.4281}) \quad (15)$$

Table 3 | Error values for training, testing and verification stages of the ANFIS model with different types of membership functions

Type of membership function	Training stage RMSE	IOA	Testing stage RMSE	IOA	Verification stage RMSE	IOA
Gaussmf	0.024642	0.99682	0.023051	0.99799	0.050288	0.988
Gauss2mf	0.029297	0.99764	0.024181	0.99704	0.032148	0.99131
Gbellmf	0.02635	0.99732	0.036533	0.99491	0.046858	0.98719
Trimf	0.025673	0.99755	0.020785	0.99802	0.057833	0.99551

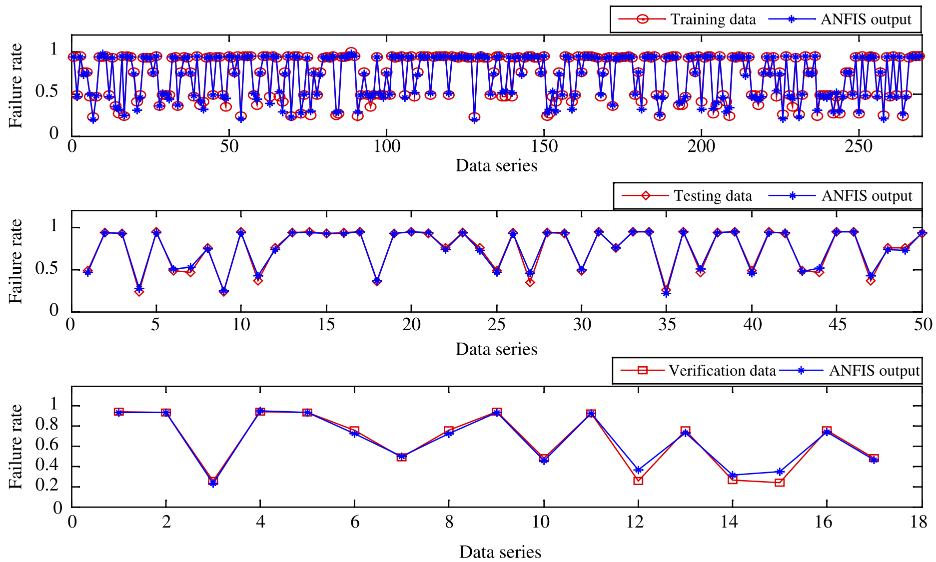

Figure 6 | Results of the ANFIS model for simulated and actual values in training, test and verification stages.

Sensitivity analysis of the proposed methodologies

To assess the performance of the proposed methodologies in predicting pipe failure rate, a series of analyses are performed by ANN, ANFIS and NLR models and Su *et al.*'s (1987) formulation (Equation (2)) for a pipe with basic data of $D = 80\,\text{mm}$, $L = 100\,\text{m}$, $P = 4\,\text{atm}$, $H = 1.2\,\text{m}$ and $Ag = 20\,\text{yr}$. In the analysis, each of these parameters was varied within its range in the database while the remaining parameters were kept constant. The selected results are presented in Table 4.

It is seen that the following trends can be identified with variations of different parameters. Pipe break rates are increased when pipe diameter and depth are decreased and pressure, length and age are increased. It is observed that the Su *et al.* (1987) formula, which only considers the effect of pipe diameter, is not sensitive to variation of the other parameters. Therefore, the existing formulations for pipe break rates, which consider very few parameters such as diameters and/or age or length, are not able to predict the break rate properly. These results highlight the necessity of consideration of as many parameters as possible.

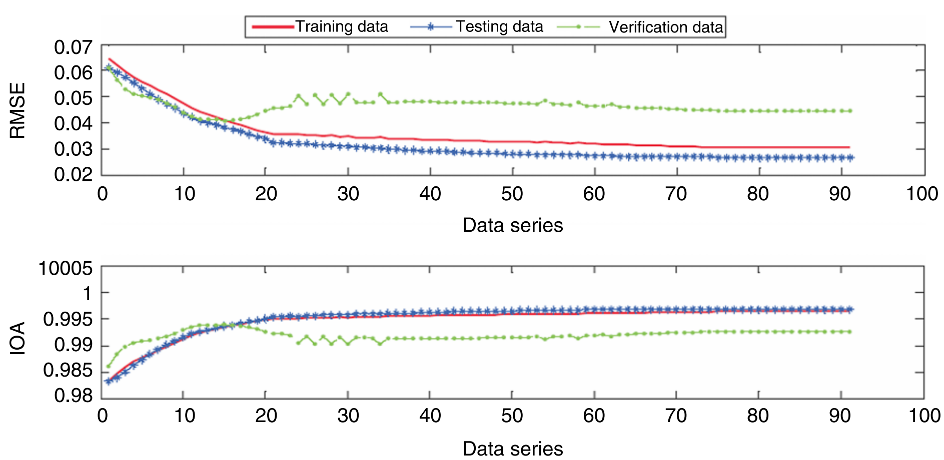

Figure 7 | Results of the ANFIS model with different methods of error calculation.

Table 4 | Sensitivity analysis of the proposed methods

Parameters		Break rate (breaks/km/yr) ANN	ANFIS	NLR	Su
D (mm)	80	1.56	5.33	1.81	0.85
	100	1.16	4.14	1.61	0.59
	150	0.61	1.77	1.22	0.33
	200	0.66	0.11	0.91	0.23
L (m)	100	1.56	2.18	1.81	0.85
	200	1.58	2.21	1.84	0.85
	300	1.61	2.21	1.87	0.85
	400	1.64	2.21	1.89	0.85
	500	1.66	2.21	1.90	0.85
P (atm)	4	1.56	2.19	1.81	0.85
	5	1.54	2.25	1.89	0.85
	6	1.99	2.28	1.96	0.85
	7	2.07	2.16	2.04	0.85
H (m)	0.5	1.59	2.22	1.91	0.85
	1	1.55	2.18	1.82	0.85
	1.5	1.51	2.18	1.79	0.85
	2	1.51	2.14	1.78	0.85
Ag (yr)	14	1.35	1.95	1.43	0.85
	18	1.44	2.13	1.69	0.85
	22	1.71	2.23	1.92	0.85
	26	2.22	2.31	2.12	0.85

The result of the ANFIS model show unrealistic values and very sharp variations for break rates when diameters are increased from 80 to 200 mm. For variations of pipe lengths between 100–500 m the ANFIS model produces almost the same break rates. The ANN and NLR methods produce lower break rates, respectively, that are more realistic than the ANFIS results but still vary in close range. It can be concluded that, because of the existence of about 95 km of pipes with 80 mm diameter, this small range of variations is reasonable. When pressure is changed from 4 to 7 atm the break rates resulting from the ANN and NLR models show an increase of up to 0.5 and 0.2, respectively. However, again variations of the ANFIS outputs are less than 0.1. The change of pipe depth from 0.5 to 2 m causes a very smooth decrease of break rate by ANN, ANFIS and NLR (about 0.1). Finally all three models show the same increasing trend for variation of age from 14 to 26 yr, although ANN and NLR produce a larger range of pipe break rates.

The other findings are that generally the trend of variations in the ANN and NLR models for most parameters are almost the same and each model is more sensitive to some specific parameters. Because in the ANN and ANFIS models input data are trained to learn the existing relations between the parameters, if the number of data are very few for specific values or ranges of some parameters, it may lead to unrealistic results. Generally in comparison with the ANFIS model, the ANN model produces more reliable results and can be introduced for pipe break rate prediction. Since mechanical reliability measures are affected by pipe failure rates, the advantages of each type of above-mentioned model are assessed in the next stage.

Mechanical reliability (availability) results

In this part of the research, the conventional relations for calculating mechanical reliability (availability) such as Poisson (Su *et al.* 1987, Equation (3)), Fujiwara & Tung (1991) (Equation (4)) and Khomsi *et al.* (1996) (Equation (8)) were selected and the failure rate parameter in these equations was substituted by the pipe failure models developed in this work (i.e. ANN, ANFIS and NLR) and also the Su *et al.* (1987) formulation (Equation (2)) and results were compared.

Poisson exponential distribution function (Su *et al.* 1987)

Based on the Poisson exponential distribution function (Su *et al.* 1987) and four pipe failure rate relationships the availability of pipes in the water distribution network are calculated and the results are illustrated in Figure 8. It is seen that the trend of variations of availability index for pipe failure rates from Su (ALPSFR) and NLR (ALPNLRFR) are the same. Also the trend of variation of availability index for pipe failure rates from ANN (ALPANNFR) and ANFIS (ALPANFISFR) methods are similar. On the other hand, availability values resulted from the Su *et al.* (1987) and the ANN pipe failure relations are close together and show an upper limit. Furthermore, availability values resulted from the ANFIS and NLR pipe failure methods have a similar trend and show a lower limit. Besides, it is observed that variations of reliability results from the ANN and ANFIS models are higher than the Su *et al.* (1987) and NLR models.

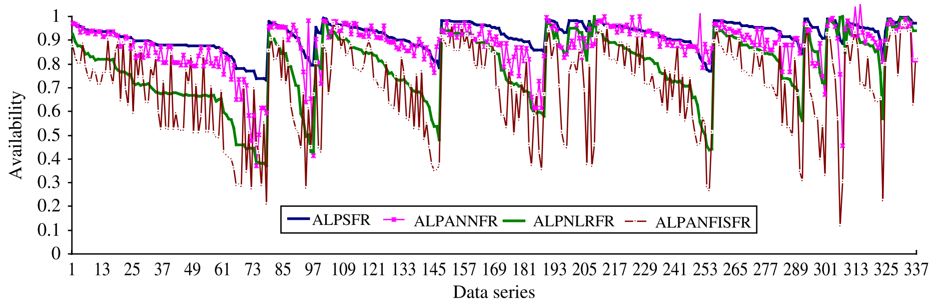

Figure 8 | Comparison of the results of availability from Poisson formula with the failure rates extracted from relationship and models of Su, ANN, ANFIS and NLR. ALPNLRFR = availability (AL) from Poisson formula (P) using nonlinear multivariate regression method (NLR) for failure rate (FR), ALPANNFR = artificial neural network model, ALPANFISFR = neuro-fuzzy model and ALPSFR = Su formula.

Also these two methods can learn the nature of the failures better than the other ones by training the available datasets to predict the failure rates. The ANFIS results represent the higher magnitude of variations than the ANN results.

Khomsi *et al.* formula (1996)

By substituting the developed three pipe failure rate models and the Su *et al.* (1987) relationships in Khomsi *et al.*'s (1996) formula the availability of the pipes in the water distribution network was calculated and the results are shown in Figure 9. Again the same conclusions as Figure 8 can be obtained.

Fujiwara & Tung formula (1991)

Considering the failure rate function of Fujiwara & Tung (1991) as the base and using the value of $\alpha_l = 0.923$ from the field data, the following relationship is obtained for

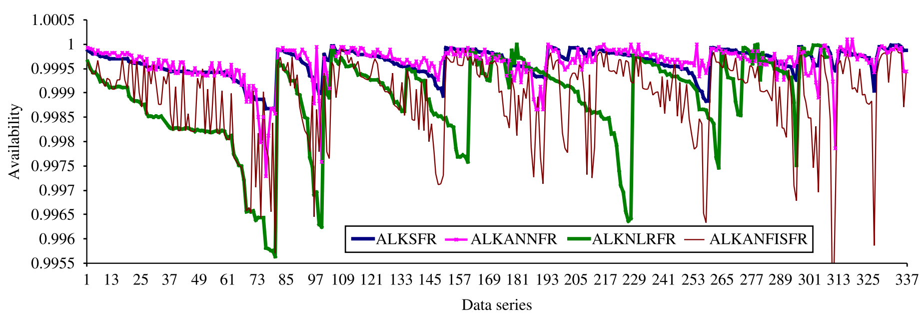

Figure 9 | Availability values from Khomsi formula with the failure rates extracted from relations and models of Su, ANN, ANFIS and NLR.

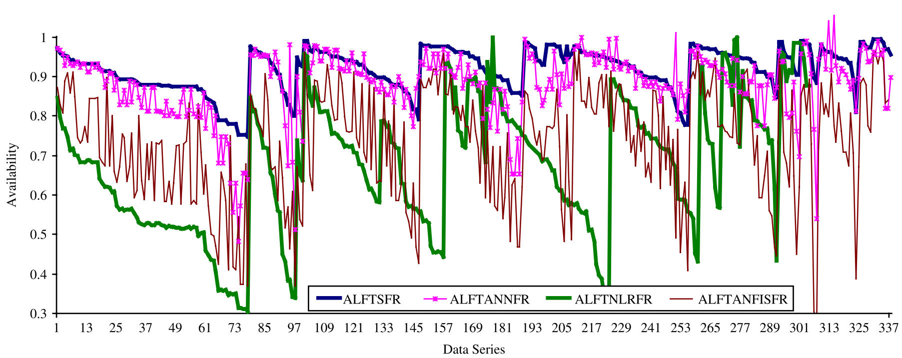

Figure 10 | Availability values from Fujiwara & Tung formula with the failure rates extracted from relations and models of Su, ANN, ANFIS and NLR.

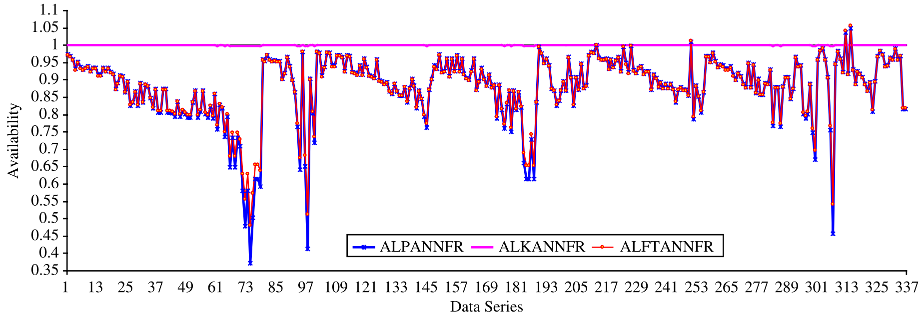

Figure 11 | Comparison of availability values involve ANN results for failure rates.

availability:

$$A_l = \frac{0.923}{[0.923 + (L_l{}^{*}\lambda)]}. \qquad (16)$$

Using the above formula and incorporating the relations and models developed for predicting pipe failure rate, the availability values of pipes are presented in Figure 10. Again the same conclusions as for the two previous figures can be drawn.

Finally, the presented methods are compared in order to study the differences of the three availability formulae of Poisson (Su et al. 1987), Fujiwara & Tung (1991) and Khomsi et al. (1996) and the results are illustrated in Figures 11 and 12. It can be seen that in both figures the Khomsi et al. (1996) equation produces very high values for pipe availability and can be considered as an upper limit. On the other hand, in both figures availability formulae of Poisson (Su et al. 1987) and Fujiwara & Tung (1991) show very high variations. However, the range of these variations is about 10% lower when using the ANN model. With the ANN model availability values from Poisson (Su et al. 1987) and Fujiwara & Tung (1991) formulae are close to each other (Figure 11). However, with the ANFIS model, availability values from the Fujiwara & Tung (1991) formula are higher in comparison with the Poisson formula (Figure 12). In general, the results obtained from the basic Khomsi et al. (1996) equation with applying failure rate from a artificial neural network model, i.e. ALKANNFR, produce the upper limit of mechanical reliability. The results calculated from the Poisson (Su et al. 1987) relation and applying failure rate from a neuro-fuzzy model, i.e. ALPANFISFR, present the lower limit of mechanical reliability.

As the managers and decision-makers in the water industry are always interested in minimizing the consequences of failures and improving the reliability of water distribution networks, the evaluation of the existing network is necessary in order to assess the current state of the system and predict future. To do this, having some practical indices to evaluate pipe failure and availability is necessary and important. The precision of these indicators directly influences managers' decisions. Among several available pipe failure and availability indices Figures 8–12 produce a

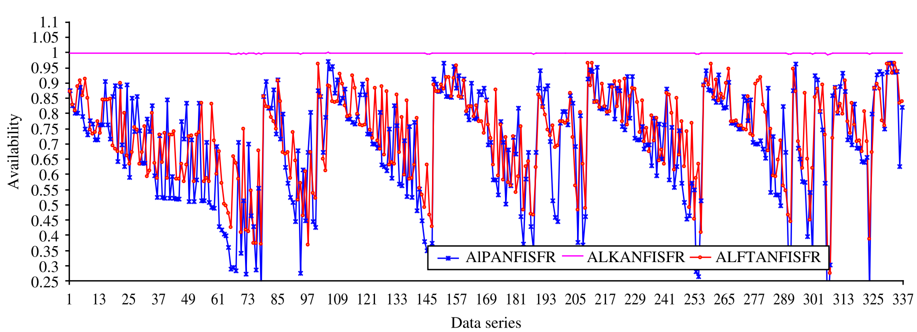

Figure 12 | Comparison of availability values involve ANFIS results for failure rates.

good framework to identify the appropriateness of the most popular and commonly used indices. It is observed that application of ANN for prediction of pipe failure rate leads to higher values for pipe availability. Therefore, any decision made based on the results of an ANN approach produces lower costs in comparison with the outputs of ANFIS and NLR models. On the other hand, application of Poisson and Fujiwara & Tung availability relationships produces lower availability results and any decision based on these results leads to higher costs. All the costs are related to repair or replacement programs, leak detection and pressure management schemes.

CONCLUSIONS

The mechanical failure of pipes in the water distribution network has been studied by numerous statistical models in the past. However, each of these models includes only a small number of contributing parameters. In this paper a new modeling approach is introduced to predict pipe failure rates and mechanical reliability of pipes using data-driven models. The pipe failure data have been collected from a real water distribution network. During the study several parameters which affect the failure rate were collected in the field. These include pipe diameter, length, age, depth and average hydraulic pressure. Then the pipe failure rates were obtained by three different methods. The results indicate that, in the prediction of pipe failure rates by the ANN and ANFIS models, the trend of variations in the observed data and in the simulated data has shown a reasonable behavior and they are able to predict failure rates with a high accuracy. Based on sensitivity analysis it was found that sensitivity of the ANN model is higher to variations of pipe diameter, pressure and age in comparison with pipe length and depth.

To evaluate different available relationships proposed for pipe availability calculations, the conventional availability relations of Poisson (Su *et al.* 1987), Fujiwara & Tung (1991) and Khomsi *et al.* (1996) were selected and combined with the relations and models developed in this research for predicting failure rate of pipes and the results were compared with each other as well as with the method of Su *et al.* (1987). It can be concluded that the results of the ANN model in all three availability relations, i.e. ALPANNFR, ALKANNFR and ALFTANNFR, produce the upper limit. The ANFIS models of ALPANFISFR, ALKANFISFR and ALFTANFISFR introduce the lower limit of mechanical reliability (availability) of water distribution networks. The results from the NLR failure rate prediction produced lower limit availability values but with very smooth variations in comparison with the ANFIS results. Furthermore, it was concluded that the availability formula of Khomsi *et al.* (1996) produces very high values (about 1) and Poisson (Su *et al.* 1987) formulation of availability represents the lowest values. Finally, because of good precision in predicting failure rate of pipes, comprehensiveness, flexibility and the possibility of connecting to hydraulic models, the ANN pipe failure rate model appears to be more appropriate in evaluating mechanical reliability (availability) values.

ACKNOWLEDGEMENTS

The authors appreciate the financial support of the Center of Excellence for Engineering and Management of Infrastructures, Department of Civil Engineering, University of Tehran.

REFERENCES

Abdel-Hamid, W., Noureldin, A. & El-Sheimy, N. 2007 Adaptive fuzzy prediction of low-cost inertial-based positioning errors. *IEEE Trans. Fuzzy Syst.* **15** (3), 519–529.

Aghayee, A. 2006 *Investigation of Burst Prediction Approach for Water Distribution Systems by Evolutionary Computing*. MSc Thesis, Civil Engineering Department, Faculty of Engineering, University of Ferdowsi, Mashad, Iran (in Farsi).

Ahn, J. C., Lee, S. W., Lee, G. S. & Koo, J. Y. 2005 Predicting water pipe breaks using neural network. *Water Sci. Technol.* **51** (3–4), 159–172.

Ang, A. H.-S. & Tang, W. H. 1984 *Probability Concepts in Engineering Planning and Design. Volume II Decision, Risk, and Reliability*. John Wiley & Sons, New York.

Cullinane, M. J., Lansey, K. E. & Mayes, L. W. 1992 Optimization-availability-based design of water distribution networks. *J. Hydraul. Eng.* **118** (3), 420–441.

Dandy, G. C. & Engelhardt, M. 2001 Optimal scheduling of water pipe replacement using genetic algorithms. *J. Water Res. Plan. Manage.* **127** (4), 214–223.

Demuth, H. & Beale, M. 2002 *Neural Network Toolbox User's Guide for Use with MATLAB*, Math Works, Inc. Natick, MA.

Farmani, R., Walters, G. A. & Savic, D. 2005 Trade off between total cost and reliability for Anytown water distribution network. *J. Water Res. Plan. Manage.* **131** (3), 161–171.

Fujiwara, O. & Tung, H. 1991 Reliability improvement for water distribution networks through increasing pipe size. *Water Resour. Res.* **27** (7), 1395–1402.

Giustolisi, O., Laucelli, D. & Savic, D. A. 2006 Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment. *Civil Eng. Environ. Syst.* **23** (3), 175–190.

Goulter, I. C., Davidson, J. & Jacobs, P. 1993 Predicting water-main breakage rates. *J. Water Res. Plan. Manage.* **119** (4), 419–436.

Goulter, I. C. & Kazemi, A. 1988 Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can. J. Civil Eng.* **15** (1), 91–97.

Goulter, I. C. & Kazemi, A. 1989 Analysis of water distribution pipe failure types in Winnipeg Canada. *J. Trans. Eng.* **115** (2), 95–111.

Harada, H. 1988 Statistics of the cold wave in the temperate region and prediction of the number of damaged service pipes. *J. AWWA* **57** (8), 12–15.

Hornik, K., Stinchcombe, M. & White, H. 1989 Multilayer feed forward networks are universal approximators. *Neural Networks* **2**, 359–366.

Jang, J. R. & Guley, N. 1996 *The Fuzzy Logic Toolbox for Use with MATLAB*. Math Works Inc. Natick, MA.

Karamouz, M., Rezapour Tabari, M. & Kerachian, R. 2007 Application of genetic algorithms and artificial neural networks in conjunctive use of surface and groundwater resources. *J. Water Int.* **32** (1), 163–176.

Kettler, A. J. & Goulter, I. C. 1983 Reliability consideration in the least cost design of looped water distribution systems. In: *Proc. Int. Symp. on Urban Hydrology, Hydraulics and Sediment Control. University of Kentucky, Lexington (EEUU), KY, 25–28 July*.

Kettler, A. J. & Goulter, L. C. 1985 An analysis of pipe breakage in urban water distribution networks. *Can. J. Civil Eng.* **12** (2), 286–293.

Khomsi, D., Walters, G. A., Thorley, A. R. D. & Ouazar, D. 1996 Reliability tester for water distribution networks. *J. Comput. Civil Eng.* **10** (1), 10–19.

Lei, J. & Saegrov, S. 1998 Statistical approach for describing failures and lifetime of water mains. *Water Sci. Technol.* **38** (6), 209–217.

Mays, L. W. 1989 *Reliability Analysis of Water Distribution Systems*. ASCE, New York.

Sacluti, F. R. 1999 *Modeling Water Distribution Pipe Failures Using Artificial Neural Networks*. MSc Thesis, Department of Civil and Environmental Engineering, University of Alberta, Canada.

Shamir, U. & Howard, C. D. D. 1979 An analytical approach to scheduling pipe replacement. *J. AWWA* **71**, 248–258.

Su, Y. C., Mays, L. W., Duan, N. & Lansey, K. E. 1987 Reliability based optimization model for water distribution systems. *J. Hydraul. Eng.* **114** (12), 1539–1556.

Tabesh, M. 1998 *Implications of the Pressure Dependency of Outflows on Data Management, Mathematical Modeling and Reliability Assessment of Water Distribution Systems*. PhD Thesis, Department of Civil Engineering, University of Liverpool, UK.

Tabesh, M. & Abedini, A. A. 2005 Analysis of pipe failure in water distribution networks. *Iran-Water Resour. Res.* **1** (1), 78–89. (in Farsi).

Walski, T. M. & Pelliccia, A. 1982 Economic analysis of water main breaks. *J. AWWA* **74** (3), 140–147.

Welter, G. J. 2001 Predicting water main breaks in winter. *AWWA Proc., AWWA Annual Conference, Washington, DC, 17–21 June*. American Water Works Association.

First received 18 December 2007; accepted in revised form 31 August 2008