

Prediction of Availability Indicator of Water Pipes Using Artificial Intelligence

Małgorzata Kutylowska^{1,*}

¹Wrocław University of Science and Technology, Faculty of Environmental Engineering, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland

Abstract. The paper presents the results of artificial neural networks application to the availability indicator prediction. The forecasted results indicate that artificial networks may be used to model the reliability level of the water supply systems. The network was trained using 147 and 173 operational data from one Polish medium-sized city (distribution pipes and house connections, respectively). 50% of all data was chosen for learning, 25% for testing and 25% for validation. In prognosis phase, the best created network used 100% of 114 and 133 values for testing. Following functions were used to activate neurons in hidden and output layers: linear, logistic, hyperbolic tangent, exponential. The learning of the artificial network was performed using following input parameters: material, total length, diameter. In the optimal models hyperbolic tangent was chosen to activate the hidden and output neurons in modeling the availability indicator of house connections during 68 epochs of training. Hidden and output neurons were activated (20 epochs of learning) respectively by hyperbolic tangent and linear function during the prediction of availability indicator of distribution pipes. The maximum relative errors in learning and prognosis step were equal to 0.10% and 1.20% as well as 0.27% and 1.15% for distribution pipes and house connections, respectively.

1 Introduction

The suitable management of water supply systems seems to be nowadays one of the crucial aspect influencing the proper maintenance of all parts of the distribution network. F. Ijjas [1] noticed that all problems related to water are connected with human being. Also the technical condition of water-pipe networks should be considered as important factor determining the way of management and maintenance of water utilities. Consequently, the reliability level of water-pipe networks should be estimated very precisely and should have an influence on the water utility management. There are a lot of investigations and studies carried out in Poland and abroad concerning the failure, safety and reliability analysis of water supply system which could be treated as a part of critical buried infrastructure [2–5].

Mathematical modeling of the reliability indicators of water-pipe network is a promising tool which seems to be used more often nowadays [6, 7]. There are a lot of mathematical approaches which are used to estimate the technical condition and

* Corresponding author: malgorzata.kutylowska@pwr.edu.pl

deterioration as well as the risk of failure of water pipes, e.g. statistical and physically based models or (very popular nowadays) fuzzy logic [8, 9]. Recently, also the model of water outflow from fault conduit was proposed [10].

In the literature there are a lot of artificial neural network modeling results related to failure rate, number of damages, water distribution profiles or time to failure [11–15], but there is lack of investigations concerning availability indicator. Till now this parameter was predicted only with applying typical mathematical models.

The aim of the presented paper was continuation, complement and development of investigations results of failure rate prediction using artificial networks [16]. Among other things, two parameters (failure rate and availability indicator) describe the reliability level of water-pipe network and should be estimated inclusively.

Availability indicator of house connections and distribution pipes in one selected medium-sized Polish city was forecasted using typical network structure – multilayer perceptron (MLP). The aim of this work was to construct relatively simple model (with basic information about the pipes given as the input signals) which would be used by water utility to predict in the future reliability level of water pipes without necessity of collecting more complicated data. The modeling results were compared with experimental data. Implementation of the model is simple which seems to be the main advantage.

2 Methodology and range of studies

2.1 Artificial neural networks

Artificial neural networks (ANN) are used to predict, classify, recognize, associate and analyze data, to filter signals and for optimization purposes. Neural networks enable to model non-linear and complex problems. The basis of ANN were described in literature [16-18]. In this paper only the main important information will be reminded. Today the most popular and best theoretically described are MLP networks [19]. This type of the network has one input layer, one or more hidden layers and one output layer.

To achieve good convergence between experimental and predicted values as well as to obtain generalization purposes, the ANN model should be learnt properly using learning methods [16, 18]. In the problem of availability indicator prediction following algorithms were used to learn the network: conjugated gradient algorithm, steepest descent method and quasi-Newton approach (BFGS algorithm). BFGS algorithm described in [16, 18, 20] seems to be the best for solving the problem of availability indicator prediction.

The proper choice of activation functions, which are responsible for transferring the information between neurons and layers, is also very crucial problem during the ANN modeling. In the problem of availability indicator prediction following functions were used to activate neurons in hidden and output layers: linear (eq. (1)), logistic (eq. (2)), hyperbolic tangent (eq. (3)), exponential (eq. (4)).

$$f(x) = x \quad (1)$$

$$f(x) = \frac{1}{1 - e^{-x}} \quad (2)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

$$f(x) = e^{-x} \quad (4)$$

2.2 Artificial neural network structure for availability indicator prediction

The availability indicator (*AI*) prediction of distribution pipes and house connections was carried out on the basis of the operating data received from water utility in one selected Polish city [21]. The reliability level of the water-pipe network could be established during the analysis of selected reliability indicators, e.g. failure rate, average time of work and availability indicator [3, 4, 22].

$$AI = \frac{1}{1 + \lambda \cdot L \cdot T_o} \quad (5)$$

where

λ – failure rate in the time t , fail./($\text{km} \cdot \text{year}$);

L – average length of the pipes, km;

T_o – average time of pipes restoration, h.

The stationary availability indicator calculated by relationship (1) was treated in this paper as experimental one, in the contrary to predicted by ANN values of *AI*. The experimental values of failure rate (presented in [16]) were calculated according to the, described in the literature, equation (2) [3, 22–24]. Generally, the availability indicator means the probability that in every time t the facility is efficient (is working) [3, 22]. In practice, this indicator determines the probabilistic assessment of facility disposal. This is reliability indicator of renewal object which is damaged and then repaired. The average time of pipe restoration was assumed at the level of 8 hours. The detailed information about T_o were not registered in the operational documentation prepared by water utility.

$$\lambda = \frac{N(t)}{L \cdot \Delta t} \quad (6)$$

where:

$N(t)$ – the number of failures in the time t , unit;

Δt – the analyzed period of time, year.

The learning of the network was performed using the operational data from the time span 2001–2006. The realization of artificial neural network approach was carried out in the program Statistica 10.0. The table 1 shows the most important assumptions of ANN architecture for availability indicator prediction.

Table 1. Main assumptions of ANN architecture.

Type of conduit	Number of data (total/learning/prognosis)			Max./min. number of hidden neurons	Number of hidden layers	Number of learning epochs
House connections	306	173	133	14/1	1	5–886
Distribution pipes	261	147	114	12/1	1	5–442

The optimal ANN model, which predicted the availability indicator in the best way, was chosen in several steps. Firstly, tens ANN architectures were checked (created) using assumptions displayed in the table 1 and also “trial and error method” (because ANN

modeling has features similar to a “black box”). Then the absolute and relative root mean squared errors were calculated for each model. Several models (characterized by the lowest errors) were chosen. The prognosis (prediction) step was carried out on the basis of these several models and then the optimal ANN model was selected. The prediction results (in learning and prognosis step) displayed in the next section were achieved using this one chosen model.

The learning of the artificial network was performed using following input parameters: material, total length, diameter. The availability indicator was an output signal predicted by ANN. Length, diameter and availability indicator were quantitative parameters. Each of mentioned above input signals was represented by one neuron. On the other hand, material was qualitative parameter and several neurons represented this signal (see table 2). The prognosis of availability indicator was performed using different data than in learning phase. This assumption is the main in ANN modeling. Firstly, the network should be learnt properly (show the proper patterns and learn the suitable generalization) and then it is necessary to check the model using new data which has not been known by the network previously.

3 Results and discussion

As it was mentioned above, the whole data set (306 and 261 data for house connections and distribution pipes, respectively) was divided into two subsets: one used for learning the artificial network, the second to make a prediction (prognosis) of availability indicator using selected before one optimal ANN model. The values of input and output parameters in learning and prediction phase are shown in the table 2.

Table 2. Input and output parameters in learning and prognosis step.

Type of conduit	Input parameters			Output parameter	
	Material	Length, km	Diameter, mm	Availability indicator	
				Experimental	Predicted by ANN
Learning step					
House connections	Grey cast iron, steel, galvanized steel, PE, PVC	23.4–29.9	25–100	0.962139–0.97864	0.962148–0.978639
Distribution pipes	Grey cast iron, steel, PE, PVC	57.3–59.7	80–200	0.970938–0.98183	0.970864–0.981838
Prognosis step					
House connections	Grey cast iron, steel, galvanized steel, PE, PVC	31.4–50.2	20–100	0.976071–0.989957	0.976926–0.978614
Distribution pipes	Grey cast iron, steel, PE, PVC	69.9–88.7	90–200	0.979353–0.992046	0.988310–0.991265

The ranges of the input parameters (material and diameter) were similar in learning and prognosis step because the water-pipe network structure has not changed significantly. The total length of house connections and distribution pipes was higher in prediction phase than in learning step because the water network has been extended year by year. In analyzed Polish city house connections were made of grey cast iron, steel, galvanized steel, PE and PVC. The distribution pipes were made of grey cast iron, steel, PE and PVC. Material is

qualitative input signal and was represented by 5 and 4 neurons during availability indicator prediction of house connections and distribution pipes, respectively. The main features of the optimal ANN models are shown in the table 3.

Table 3. Optimal ANN model.

Type of conduit	Type of MLP	Activation function (hidden layer)	Activation function (output layer)	Relative mean squared errors, %
House connections	7-14-1	Hyperbolic tangent	Hyperbolic tangent	0.00–1.15
Distribution pipes	6-10-1	Hyperbolic tangent	Linear	0.00–1.20

It was necessary to use 68 and 20 epochs of learning to train the network properly and to obtain reasonable predictions of availability indicators of house connections and distribution pipes, respectively. In comparison with failure rate prediction (that was also modeled using ANN and BFGS learning algorithm) [16] the number of learning epochs was almost 1.7 times higher for house connections and six times lower for distribution pipes. We should remember that the simple comparison between ANN architecture used for failure rate and availability indicator prediction cannot be carried out because ANN structure is directly connected with the solved problem. ANN is like a “black box” modeling (we have not knowledge what is going on inside the network) and it is not possible to obtain one uniform (universal) architecture for every kind of problem. It is necessary to seek the proper ANN structure for specific problem which is currently solved.

The decision concerning the number of hidden layers and neurons is fundamental creating the ANN structure. It has been proven that one hidden layer is sufficient for modeling engineering problems [11, 12]. The proper assignation of number of hidden neurons also makes some difficulties because we should avoid the risk of overtraining. On the other hand, it is necessary to show the artificial network proper patten and to learn in sufficient way. In the problem of availability indicator prediction we have 7 or 6 input neurons (concerning house connections and distribution pipes). The optimal ANN structures consist of 14 or 10 hidden neurons, respectively.

Similar as in failure rate prediction [16], the bias value was very important relating to the training process. Hyperbolic tangent was the optimal function in hidden and output layer for availability indicator prediction of house connections. Availability indicator of distribution pipes was predicted with the lowest relative error using hyperbolic tangent and linear function to activate the neurons in hidden and output layer, respectively.

The results of ANN modeling of availability indicator are displayed in the figures 1 and 2. The predictions of availability indicator of distribution pipes were, from engineering point of view, satisfactory. In the learning step, Fig. 1a, (2001–2006) the relative errors varied between 0.00% and 0.10%. The correlation between experimental and predicted data was characterized by coefficients $R=0.9918$ and $R^2=0.9836$. Such convergence is very good. The prognosis step, Fig. 1b, was characterized by higher discrepancies between experimental and predicted data ($R=0.3105$). The values of predicted availability indicator were relatively higher than experimental what could be caused by the fact that the ANN model did not know (in prognosis phase) the input data previously. It is obvious that in prognosis the discrepancies between experimental and predicted values are higher than in learning step. Only in 2011 the predictions were almost exactly the same as experimental values of availability indicators.

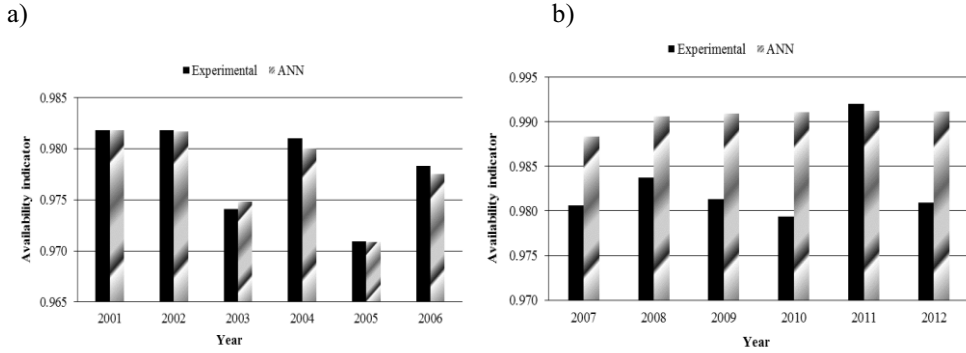


Fig. 1. Experimental and predicted availability indicator of distribution pipes, a) learning step, b) prognosis step.

The modeling results of availability indicator of house connections seem to be quite promising (Fig. 2). The predictions in learning phase (Fig. 2a) were characterized by coefficients $R=0.9870$ and $R^2=0.9742$. It means that the correlation between experimental values and prognosis is almost ideal. The relative errors were in the range 0.00 and 0.27%. Worse results were obtained in years 2007–2012 (prognosis step, Fig. 2b). Pearson correlation coefficient was equal to 0.3258, which is slightly better than in distribution pipes modeling. But still the relative error was not higher than 1.15%. In the prognosis phase one input signal (total length) was significantly higher than in learning step which could cause lower convergence. From engineering point of view the model MLP 7-14-1 could be recommended and implemented by water utility to forecast in the future the availability indicator of house connections because learning phase was characterized by relatively good correlation.

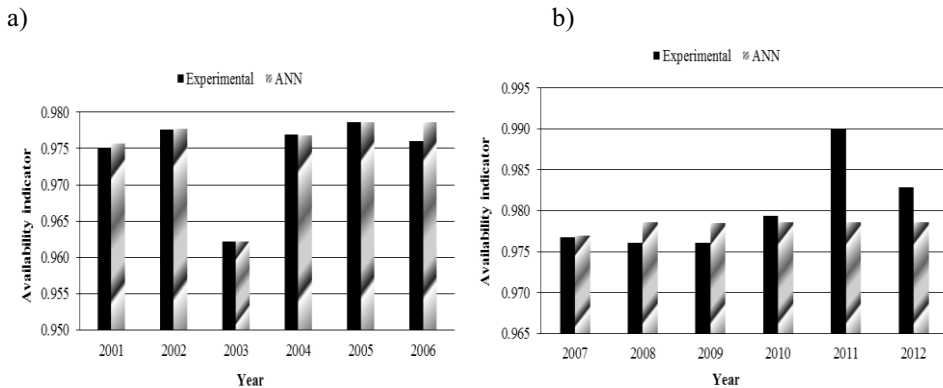


Fig. 2. Experimental and predicted availability indicator of house connections, a) learning step, b) prognosis step.

The results of availability indicator prediction of distribution pipes and house connections confirm that ANN modeling could be useful forecasting tool. Unfortunately, the above mentioned results show quite high discrepancies between experimental and predicted values in the prognosis step. This could be improved if bigger input vector is used. Generally, we should remember that the size of water-pipe network is increasing and also the diameters and materials are changing even year by year. In consequence of this, the input parameters could be slightly different in learning and prognosis step. The idea of, proposed in this paper, modeling is to adapt created models to changing conditions of water pipes. Next step could be training the network using input signals not from 6 but 12 years

and the validation (prognosis) step will be done on next years (e.g. 2013, 2014). In such case the pattern will be recognized more precisely. The adaptation (changes in the network structure and in the ranges of input parameters) could not be performed to infinity and should be stopped when the reasonable level of the correlation between experimental and predicted values is achieved. Nevertheless, ANN modeling seems to be promising tool for prediction of the reliability level of water-pipe network. The main advantage of such modeling is relatively short time of calculation and possibilities of adaptation as well as the simple implementation by water utility for its own purposes. Moreover, it is necessary to remember that assessing the failure frequency of water pipes should be carried out together with the estimation of proper operation of sewerage systems, e.g. storm water system [25, 26]. These two municipal systems are very important for proper functioning of the whole buried infrastructure and should be considered together.

4 Conclusions

The results of availability indicator prediction in one Polish medium-sized city might be concluded as follows:

- The generalization phenomenon of ANN allows to describe the relationships (due to weight adaptation during the network training) between variables without having exact knowledge about these relations.
- It is needed to create the artificial network for each problem separately. The solved problem extorts the number of neurons, the kind of activation function and the training methods.
- There are some limitations of using ANN for prediction purposes. It is necessary to collect huge data base. This requirement is sometimes difficult to realize by water utilities. Some important, from forecasting point of view, variables are not registered. The situation is still improving because more and more water utilities attach importance to connect information about damages with GIS.
- It is also necessary to analyse deeply the data which are used as the training parameters in the process of neural network modelling. In some cases it is necessary to eliminate the outliers. If there is too much eliminated outlier data we should answer the question whether the modelling is rational. Unfortunately, real operational data are burden sometimes by huge mistakes due to e.g. improper or incomplete registration or lack of collecting all information in GIS database. In such cases it is necessary to eliminate or remain outliers reasonably. Before modelling approach it is required to cooperate with exploiters and to explain all inaccuracy as well as to complete data.
- The learning of the artificial neural network was performed using the operational data from the time span 2001–2006. The next years (2007–2012) were used to make a prognosis of availability indicator.
- BFGS method of learning was established as the best one in comparison with other training algorithms.
- MLP network structures (6-10-1 and 7-14-1) seem to be efficient for availability indicator predictions of distribution pipes and house connections, respectively.
- In the optimal models: hyperbolic tangent was chosen to activate the hidden and output neurons in modeling the availability indicator of house connections during 68 epochs of training. Hidden and output neurons were activated (20 epochs of learning) respectively by hyperbolic tangent and linear function during the prediction of availability indicator of distribution pipes.
- The maximum relative error in learning and prognosis step were equal to 0.10% and 1.20% as well as 0.27% and 1.15% for distribution pipes and house connections, respectively.

- The convergence between experimental and predicted values of availability indicator of distribution pipes and house connections was characterized by $R^2=0.9836$ and $R^2=0.9742$ (learning step), respectively. The prognosis phase gave worse results of modeling what could be improved by adaptation the model to the changing operational conditions.

The work was realized within the allocation No. 0401/0069/16 awarded for Faculty of Environmental Engineering Wrocław University of Science and Technology by Ministry of Science and Higher Education in years 2016–2017.

References

1. F. Ijjas, *Period. Polytech-Civ.* **58**, 2 (2014)
2. I. Zimoch, E. Łobos, *Environ. Prot. Eng.* **38**, 3 (2012)
3. H. Hotłoś, Quantitative assessment of the effect of some factors on the parameters and operating costs of water-pipe network (Wrocław University of Technology Publishing house, Wrocław, 2007) (in Polish)
4. M. Kwietniewski, J. Rak, Reliability of water supply and wastewater infrastructure in Poland (Polish Academy of Science, Warszawa, 2010) (in Polish)
5. J. M. Makar, *Eng. Fail. Anal.* **7**, 1 (2000)
6. M. Kutylowska, *Period. Polytech-Civ.* **59**, 1 (2015)
7. D. Laucelli, B. Rajani, Y. Kleiner, O. Giustolisi, *J. Hydroinform.* **16**, 4 (2014)
8. Y. Kleiner, B. Rajani, *Urban Water.* **3**, 3 (2001)
9. B. Tchórzewska-Cieślak, *Environ. Prot. Eng.* **37**, 3 (2011)
10. M. Iwanek, D. Kowalski, M. Kwietniewski, *Ochr. Sr.* **37**, 4 (2015)
11. A. Shirzad, M. Tabesh, R. Farmani, *KSCE J. Civ. Eng.* **18**, 4 (2014)
12. R. Jafar, I. Shahrour, I. Juran, *Math. Comput. Model.* **51**, 9-10 (2010)
13. M. Tabesh, J. Soltani, R. Farmani, D. Savic, *J. Hydroinform.* **11**, 1 (2009)
14. W. Cieżak, J. Cieżak, *Environ. Prot. Eng.* **41**, 2 (2015)
15. R. Harwey, E. A. McBean, B. Gharabahgi, *J. Water Res. Pl-ASCE.* **140**, 4 (2014)
16. M. Kutylowska, *Eng. Fail. Anal.* **47**, (2015)
17. M. Kabsch-Korbutowicz, M. Kutylowska, *Desalination Water Treat.* **14**, 1-3 (2010)
18. R. Tadeusiewicz, *Neural networks* (Academic Publishing House RM, Warszawa, 1993) (in Polish)
19. X. Zeng, D. Yeung, *Neurocomputing* **69**, 7-9 (2006)
20. S. Osowski, *Neural networks for information transferring* (Warsaw University of Technology Publishing House, Warszawa, 2000) (in Polish)
21. Operational data given by Water Utility in one Polish city, 2013
22. M. Kwietniewski, M. Roman, H. Kłoss-Trębaczkiwicz, Reliability of water pipes and sewers (Arkady, Warszawa, 1993) (in Polish)
23. M. Kutylowska, H. Hotłoś, *Eng. Fail. Anal.* **41**, (2014)
24. D. Kowalski, K. Miszta-Kruk, *Eng. Fail. Anal.* **35**, (2013)
25. B. Kaźmierczak, M. Wdowikowski, *Period. Polytech-Civ.* **60**, 2 (2016)
26. A. Kotowski, K. Wartalska, M. Nowakowska, *Ochr. Sr.* **38**, 1 (2016)