Periodica Polytechnica
Civil Engineering

61(3), pp. 548–553, 2017
https://doi.org/10.3311/PPci.9997

RESEARCH ARTICLE

Prediction of Failure Frequency of Water-Pipe Network in the Selected City

Małgorzata Kutyłowska[1*]

Received 8 September 2016; Revised 04 November 2016; Accepted 9 December 2016

Abstract
The paper presents the modelling results of failure rate of water mains, distribution pipes and house connections in one Polish city. The prediction of failure frequency was performed using artificial neural networks. Multilayer perceptron was chosen as the most suitable for modelling purposes. Neural network architecture contained 11 input signals (sale, production, consumption and losses of water, number of water-meters, length and number of failures of water mains, distribution pipes and house connections). Three neurons (failure rates of three conduits types) were put to the output layer. One hidden layer, with hidden neurons in the range 1-22, was used. Operating data from years 2005-2011 were used for training the network. Optimal model was verified using operational data from 2012. Model MLP 11-10-3 was chosen as the best one for failure rate prediction. In this model hidden and output neurons were activated by exponential function and the learning was done using quasi-Newton approach. During the learning process the correlation (R) and determination (R^2) coefficients for water mains, distribution pipes and house connections equaled to 0.9921, 0.9842; 0.8685, 0.7543 and 0.9945, 0.9891, respectively. The convergences between real and predicted values seem to be, from engineering point of view, satisfactory.

Keywords
artificial neural network, failure rate, modelling, water conduits

[1]Faculty of Environmental Engineering
Wrocław University of Science and Technology
Wybrzeże St. Wyspiańskiego 27, 50-370 Wrocław, Poland
[*]Corresponding author, e-mail: malgorzata.kutylowska@pwr.edu.pl

1 Introduction

Water-pipe networks belong to the critical infrastructure and the proper management of the whole water supply systems should be established at the high level of importance [1]. Due to this fact the failure frequency seems to be one of the most important indicators considered during the management process as well as during estimation of the reliability level of water distribution systems [2, 3]. Nowadays, failure rate of water pipes should be calculated not only on the basis of operational data, but also using the best available mathematical techniques and models [4, 5]. Mathematical modelling must follow the collection of operational data which are of course the information base of considered water-pipe network. There are a lot of typical mathematical models and solutions [6-8] which allow us: to predict the failure frequency, to enable the reliability and risk analysis and to have an influence on the quick reaction when the serious damage occurs.

On the other hand, artificial intelligence is nowadays used more often as the alternative for typical statistical or physically based models. The most popular method seems to be forecasting using artificial neural networks (ANN). Artificial neural networks 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 information between neurons laid in layers (input, hidden and output) is transferred in one direction and neurons calculate the weighted sum of signals. The weight values are obtained during the learning process. Depending on their structure and way of transmitting signals between neurons, the following types of artificial neural networks are distinguished: linear networks, networks with radial base functions, recursive networks and multilayer perceptron networks (MLP). The additional feature of ANN it is ability of adaptation to unstable and dynamic conditions and changes the values of parameters. This property enables efficient usage especially in engineering where a lot of parameters are changing dynamically. Artificial neural networks are an alternative (instead of traditional modelling) because they enable prediction of unstable parameters without necessity of knowing exact relation between variables.

The disadvantage of prediction using ANN is the necessity of collecting huge experimental data base which is used for learning and then for making the prognosis. Generally, it should be remembered that ANN modelling is like "black box" approach and that is why it is impossible to penetrate deeply inside the way of forming the network structure.

Neural networks are used e.g. to predict the water distribution profiles [9], to forecast the water level in reservoir dam [10], to establish the time of damage occurrence [11], to predict number of failure [12]. Neural network will function correctly when the proper pattern is shown. On the basis of this pattern the network training will be carried out [13]. The aim of the network learning is to obtain the correct output signals as well as to establish all weights which are modified to reach the most suitable values. The neural network, which is trained on one data set, should generate correct results when the input vector contains different data (not included before to learning set). This feature is one of the most important because shows the generalization abilities of artificial neural network modelling.

The main aim of this paper was to present the possibilities of using ANN for prediction of failure rate of water mains, distribution pipes and house connections in one Polish city. Multilayer perceptron with one input, one hidden and one output layer was considered for prediction purposes. This network structure is nowadays the best described and the most suitable for engineering phenomena forecasting [14, 15]. The important advantage of neural networks, using for predicting unstable values, is the generalization ability between input and output data without necessity of knowing the relationships' nature. This aspect was the most important during the selection of some input parameters (theoretically not connected with the failure frequency) put to the neural network described in this paper.

2 Material and methods

Failure rate prediction of water pipes was performed using ANN, but the approach presented in this paper was completely different in comparison to earlier author's results [16]. Operational data (received from Water Utility) from the time span 2005-2011 were used for network training. The verification of the chosen optimal neural network model was carried out on the basis of data from 2012. Till now the input signals for failure rate prediction using ANN were connected with the character of water pipes, e.g. material, age, diameter [12, 16]. In this work general data describing the character of water-pipe network were used as the input neurons. According to this assumptions following data were selected as input parameters: water sale (SW), water production (PW), water consumption (BW), water losses (TW), number of water-meters (LW), length of water mains (D_m), length of distribution pipes (D_r), length of house connections (D_p), number of failures of water mains (L_m), number of failures of distribution pipes (L_r), number of failures of house connections (L_p). Such approach is innovative and shows the universality of using artificial intelligence, even in the case when the relationships between input and output signals are not known or are difficult to assess. Three failure rates of three conduit's types (water mains - λ_m, distribution pipes - λ_r and house connections - λ_p) were put to the output layer. On the basis of operating data (registered by Water Utility), the total annual water sale, production, consumption and losses of water as well as the number of damages of the water pipes were taken into consideration. The number of water-meters meant the total number of meters installed in individual and industrial customers who are under Water Utility's management. The values of experimental indicator λ (fail./km·a) were calculated on the basis of the simple relation (the number of damages per year divided by average pipe's length). In the next sections of this paper the values of experimental failure rate indicator (λ) are called real values as opposed to predicted by ANN values.

The ranges of input and output parameters is shown in the table 1. The values of water losses are not included in the table because they are like "top secret" from exploiters point of view and cannot be published.

Exemplary neural network architecture (MLP type) with eleven input neurons (I) and three output neurons (O) is displayed in the figure 1. The number of hidden neurons was varied between 1 and 22 according to the model. The whole data set was divided into subsets: 70% of all data from the training sample (operational data from 2005-2011) was used to learn the artificial neural network, 15% for testing and 15% for validation. The verification of the chosen ANN models was done using prognosis sample (operational data from 2012). Linear, logistic, exponential and hyperbolic tangent activation functions were used in the training process of ANN models. The weights reduction was used in hidden and output layers to achieve more stable solutions.

Table 1 Scope of input and output signals in the time span 2005-2012

SW, m^3	PW, m^3	BW, m^3	TW, m^3	LW	D_m, km	D_r, km
35698593	43808650	47773882	-	43057	201.7	1197.3
36678414	46400316	49429010	-	44256	212.4	1266.5
D_p, km	**L_m**	**L_r**	**L_p**	**λ_m, fail./(km·a)**	**λ_r, fail./(km·a)**	**λ_p, fail./(km·a)**
401,9	16	328	109	0.08	0.26	0.26
434,4	35	518	380	0.16	0.43	0.93

Table 2 Parameters of artificial neural network models

No.	Model	Learning quality	Learning error	Testing error	Validation error	Number of learning epochs	Activation function-hidden layer	Activation function -output layer
1.	**MLP 11-8-3**	**0.996815**	**0.000554**	**0.005477**	**0.001642**	**66**	**logistic**	**logistic**
2.	**MLP 11-5-3**	**0.989200**	**0.001029**	**0.007251**	**0.002911**	**11**	**linear**	**tanh**
3.	MLP 11-1-3	0.510212	0.017006	0.013971	0.001497	34	logistic	logistic
4.	MLP 11-1-3	0.570704	0.024691	0.003823	0.001284	9	linear	logistic
5.	**MLP 11-15-3**	**0.997235**	**0.000239**	**0.005919**	**0.002073**	**11**	**hyperbolic tangent**	**linear**
6.	**MLP 11-4-3**	**0.979028**	**0.002170**	**0.000568**	**0.005717**	**11**	**linear**	**tanh**
7.	MLP 11-3-3	0.589936	0.012236	0.006677	0.001812	7	logistic	exponential
8.	MLP 11-10-3	0.610937	0.026316	0.015773	0.000355	4	exponential	logistic
9.	MLP 11-11-3	0.999753	0.000042	0.033948	0.000063	38	exponential	exponential
10.	**MLP 11-10-3**	**0.999856**	**0.000021**	**0.003647**	**0.000337**	**112**	**exponential**	**exponential**
11.	MLP 11-5-3	0.962269	0.002016	0.010596	0.002122	13	linear	exponential
12.	MLP 11-12-3	0.971449	0.003112	0.000289	0.009297	9	hyperbolic tangent	tanh
13.	MLP 11-10-3	0.936228	0.008271	0.002407	0.002961	5	hyperbolic tangent	logistic
14.	MLP 11-3-3	0.645893	0.023620	0.002484	0.010894	16	exponential	logistic
15.	MLP 11-1-3	0.658002	0.018063	0.005177	0.000834	9	linear	linear
16.	MLP 11-10-3	0.967928	0.003303	0.004431	0.007306	14	logistic	tanh
17.	MLP 11-7-3	0.949071	0.005963	0.002727	0.001563	8	exponential	tanh
18.	**MLP 11-19-3**	**0.998526**	**0.000209**	**0.007996**	**0.000985**	**19**	**hyperbolic tangent**	**exponential**
19.	MLP 11-11-3	0.944676	0.004177	0.001033	0.005042	6	linear	tanh
20.	**MLP 11-4-3**	**0.991992**	**0.000671**	**0.002774**	**0.000723**	**17**	**hyperbolic tangent**	**logistic**

The methodology of choosing the optimal model was as follows: firstly, 20 of ANN models, characterized by learning quality higher than 0.5, were created (in this step the network was learnt using the data from the time span 2005-2011), the learning quality and the relative mean-squared errors and root mean-squared errors for each model were calculated. Secondly, the correctness of prediction was checked (prognosis step) using operational data from 2012. Optimal model was characterized by the lowest relative mean-squared error and root mean-squared error as well as the best convergence between real and forecasted by ANN data. The calculations were performed in the program Statistica 10.0.

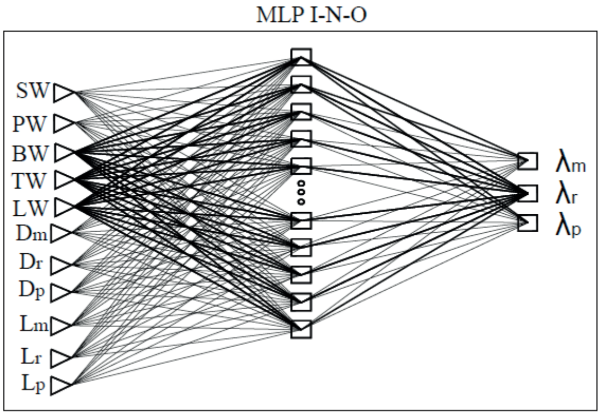

Fig. 1 Artificial neural network structure

3 Results and discussion

On average, in years 2005-2012 real failure rates (λ, fail./(km·a)) of water mains, distribution pipes and house connections were equal to 0.13, 0.35 and 0.69, respectively. On the basis of the methodology mentioned above, 20 artificial neural network models were created. The main parameters of the models are listed in the table 2. For the further analysis (model verification and making prognosis) 7 models were chosen (they are bolded in the table 2). The analysis of mean-squared errors and the convergences between real and predicted by ANN data proved that the model no. 10 was the optimal one.

The real and predicted by the model no. 10 values of failure rate are shown in the table 3.

Table 3 Failure rates (optimal model MLP 11-10-3)

Year/type of conduit	Water mains		Distribution pipes		House connections	
	Real	ANN	Real	ANN	Real	ANN
2005	0.08	0.08	0.32	0.32	0.71	0.72
2006	0.14	0.14	0.43	0.43	0.93	0.93
2007	0.13	0.14	0.34	0.33	0.72	0.70
2008	0.15	0.15	0.34	0.34	0.74	0.74
2009	0.10	0.10	0.37	0.37	0.26	0.27
2010	0.16	0.16	0.35	0.35	0.71	0.71
2011	0.16	0.16	0.26	0.33	0.63	0.58
2012	0.08	0.13	0.36	0.33	0.78	0.29

Chosen optimal model MLP 11-10-3 contained 10 hidden neurones activated by exponential function. In the output layer exponential function was also the most suitable. The number of learning epochs varied between 4 and 112 for twenty selected ANN models. From among of three learning methods (conjugated gradients, steepest descent and quasi-Newton) the last one (using quadratic approximation of objective function near the known solution) seemed to be the most suitable for prediction of failure rate of water pipes.

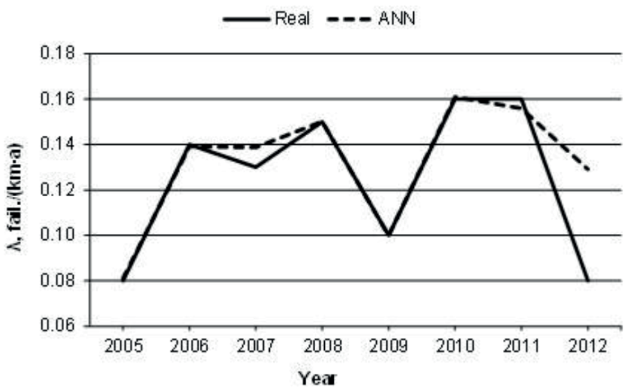

Fig. 2 Real and predicted values of failure rate of water mains

The analysis of the table 3 and figures 2-4 shows that the results obtained during the network learning (years 2005-2011) are ideally convergent with real data. Only outlier data (year 2011, distribution pipes and house connections) were not modelled properly by ANN. This fact testifies that it is 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 (Water Utilities) and to explain all inaccuracy as well as to complete data.

Concerning water mains (fig. 2) the results obtained using optimal ANN model (learning step) were the same as real values. The prognosis of failure rate (year 2012) was characterized by slight discrepancy between real and predicted data. But one should remember that the network training was performed only on seven years of exploitation that had probably the great influence on the verification quality. The verification was done using the data which were not known previously by the artificial neural network model. It is not recommended to overtrain the network because in such case the generalization ability would be lost. As it was mentioned above, the number of failures influences the values of failure rate. Especially, in the case of water mains the number of years taken into consideration during the learning process is essential because small amount of damages occur in each year in comparison to e.g. distribution pipes. It was proved [17] that the diameter has the great effect on the level of failure frequency. If the diameter is smaller, the number of damages is higher. That is the reason why more years of operation should be taken into consideration when the failure rate of water mains is considered. In this case more damages would be registered and the accuracy of prediction could increase.

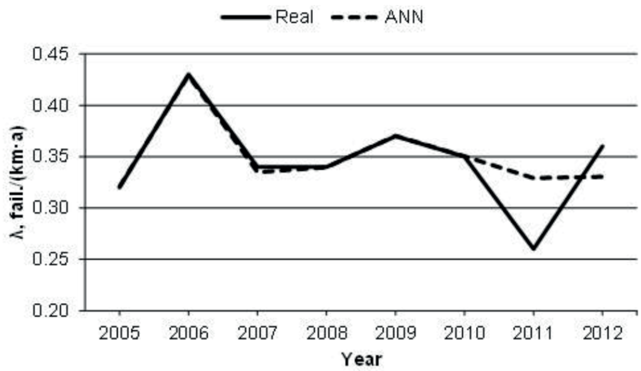

Fig. 3 Real and predicted values of failure rate of distribution pipes

The prognosis of failure rate of distribution pipes (fig. 3) is, from engineering point of view, acceptable. Root mean-squared errors in prognosis step (year 2012) for chosen optimal model were equalled to 7.8%, 2.9% and 7.9% respectively for water mains, distribution pipes and house connections that is satisfactory result.

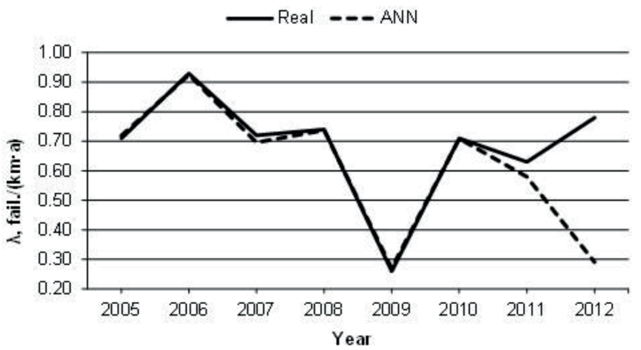

Fig. 4 Real and predicted values of failure rate of house connections

Quite big discrepancies between real and predicted values of failure frequency of house connections were observed (fig. 4). It is the hint for future researches. Maybe it would be necessary to change the neural network architecture or to increase the size of learning vector. It is required to collect more records of each single input parameter (instead of hundreds we should have for example thousands). In such case the training would be more efficient. The predicted by ANN value of failure rate λ of house connections was over 2.5 times lower in comparison to real one. Nevertheless, it was decided to choose the model MLP 11-10-3 because for other types of conduit (water mains and distribution pipes) the errors could be acceptable. During the process of choosing the optimal model we should take into consideration not only the agreement between predicted by ANN and real data, but also the role of conduit type and its influence on the reliable operation of the whole water-pipe network.

One failure of water mains or distribution pipes have higher impact than even tens of damaged house connections at the same time. Such situation was observed on Friday 3.07.2015 in Polish city (number of citizens is equalled to ca. 700 000) which is considered in this paper. In the early morning the huge damage of water main with the diameter 1200 mm has occurred. During several hours water was not delivered to the most of citizens. In some districts the water pressure in water-pipe network was significantly lower. It had the great impact on operation of hospitals and industry. In such case the reliability level of water-pipe network which deliver water to critical infrastructure (e.g. hospitals, industry) should be higher than in the rest of the city. That is the reason why other water resources for e.g. hospitals or duplication of water conduits should be planned and designed. The failure of water main had an influence not only on the quality of water delivering, but also on life quality. Near the place of damage the traffic jam has been occurring for long hours, public transport was forced to change the route and on the road surface the thick mud layer was observed.

Taking into consideration mentioned above situation, choosing the optimal neural network model should be connected not only with the agreement between predicted by ANN and real data, but also with the effects of improper estimation of failure rate. Incorrect prediction of failure rate indicator λ of house connections would not have such great consequences as improper forecasting failure frequency of water mains or distribution pipes. Other neural network models (tab. 2) were characterized by higher mean-squared errors during the prognosis of failure rate λ of significant types of conduit (water mains and distribution pipes) which are important for reliable operation of the whole water-pipe network. It was the reason why model MLP 11-10-3 (which was characterized at the prognosis step in 2012 by the best convergence between real and predicted values of failure rate indicator) was chosen as the optimal one.

4 Conclusions

The obtained results modelling using artificial neural networks might be concluded as follows:

- failure rate indicator λ, (fail./(km·a) of water mains, distribution pipes and house connections in one Polish city was predicted with the acceptable, from engineering point of view, error;
- multilayer perceptron with one hidden layer was chosen as the most suitable for modelling purposes;
- neural network architecture contained 11 input signals (sale, production, consumption and losses of water, number of water-meters, length and number of failures of water mains, distribution pipes and house connections);
- three neurons (failure rates of three conduits types) were put to the output layer;
- operating data from years 2005-2011 were used for training the network;
- verification was performed on the basis of operational data from 2012;
- model MLP 11-10-3 was chosen as optimal, hidden and output neurons were activated by exponential function and the learning was done using quasi-Newton approach;
- the learning process was characterized by the correlation (R) and determination (R^2) coefficients for water mains, distribution pipes and house connections which were equaled to 0.9921, 0.9842; 0.8685, 0.7543 and 0.9945, 0.9891, respectively.

The problem stated in this paper seems to be crucial because it is necessary to estimate the level of failure rate properly and relatively quickly. ANN are convenient method of quick parameters prediction. The methodology suggested in this paper had significant changes in comparison to earlier approaches proposed for failure rate of water pipes prediction. The input signals were different than previously. Input neurons described the general character of water-pipe network, e.g. sale, production, consumption and losses of water or number of water-meters. Such approach seems to be reasonable especially in the case when the operational data are not registered carefully or there is lack of exact information about age, pipe-laying depth, type of soil or temperature and other operational parameters. On the other hand such information as the amount of water pumped to the system or sold water are registered as a general rule. The agreement between real and predicted by ANN data could be acceptable and proved that it was possible to include different input parameters for learning purposes. The main feature (the generalization ability without knowing the relationships between input and output vector) of ANN was checked and proved.

Moreover, it is necessary to remember that assessing the failure frequency of water pipes should be carried out together with impact of water leakages on the soil (suffusion processes) [18] and with the estimation of reliability level of sewerage systems, e.g. storm water system [19, 20]. These two municipal systems are very important for proper functioning of the whole buried infrastructure and should be considered together.

References

[1] Pietrucha-Urbanik, K. "Assessment model application of water supply system management in crisis situations." *Global NEST Journal*. 16(5), pp. 893-900. 2014. URL: http://journal.gnest.org/sites/default/files/Submissions/gnest_01414/gnest_01414_published.pdf

[2] Kowalski, D., Miszta-Kruk, K. "Failure of water supply networks in selected Polish towns based on the field reliability tests." *Engineering Failure Analysis*. 35, pp. 736-742. 2013. DOI: 10.1016/j.engfailanal.2013.07.017

[3] Kwietniewski, M., Rak, J. „Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce." „Reliability of water supply and wastewater infrastructure in Poland." Polish Academy of Science, Warszawa, Poland. 2010. (in Polish)

[4] Bogárdi, I., Fülöp, R. "A spatial probabilistic model of pipeline failures." *Periodica Polytechnica Civil Engineering*. 55(2), pp. 161-168. 2011. DOI: 10.3311/pp.ci.2011-2.08

[5] Shin, H., Kobayashi, K., Koo, J., Do, M. "Estimating burst probability of water pipelines with a competing hazard model."*Journal of Hydroinformatics*. 18(1), pp. 126-135. 2016. DOI: 10.2166/hydro.2015.016

[6] Kleiner, Y., Rajani, B. "Comprehensive review of structural deterioration of water mains: statistical models."*Urban Water*. 3(3), pp. 131-150. 2001. DOI: 10.1016/S1462-0758(01)00033-4

[7] Scheidegger, A., Scholten, L., Maurer M., Reichert P. "Extension of pipe failure models to consider the absence of data from replaced pipes."*Water Research*. 47(11), pp. 3696-3705. 2013. DOI: 10.1016/j.watres.2013.04.017

[8] Seo, J., Koo, M., Kim, K., Koo, J. "A study on the probability of failure model based on the safety factor for risk assessment in a water supply network." *Procedia Engineering*. 119, pp. 206-215. 2015. DOI: 10.1016/j.proeng.2015.08.877

[9] Cieżak, W., Cieżak, J. "Routine forecasting of the daily profiles of hourly water distribution in cities. An effectiveness analysis." *Environment Protection Engineering*. 41(2), pp. 179-186. 2015. DOI: 10.5277/epe150215

[10] Ünes, F., Demirci, M., Kisi, Ö. "Prediction of Millers Ferry Dam reservoir level in USA using artificial neural network." *Periodica Polytechnica Civil Engineering*. 59(3), pp. 309-318. 2015. DOI: 10.3311/PPci.7379

[11] Harwey, R., McBean, E.A., Gharabaghi, B. "Predicting the timing of water main failure using artificial neural networks." *Journal of Water Planning and Management*. 140(4), pp. 425-434. 2014. DOI: 10.1061/(ASCE)WR.1943-5452.0000354

[12] Jafar, R., Shahrour, I., Juran, I. "Application of artificial neural networks (ANN) to model the failure of urban water mains." *Mathematical and Computer Modelling*. 51(9-10), pp. 1170-1180. 2010. DOI: 10.1016/j.mcm.2009.12.033

[13] Tadeusiewicz, R. "Neural networks." Academic Publishing House RM, Warszawa (in Polish).

[14] Zeng, X., Yeung, D. "Hidden neuron pruning of multilayer perceptrons using a quantified sensitivity measure." *Neurocomputing*. 69(7-9), pp. 825-837. 2006. DOI: 10.1016/j.neucom.2005.04.010

[15] Kutyłowska, M. "Comparison of two types of artificial neural networks for predicting failure frequency of water conduits." *Periodica Polytechnica Civil Engineering*. 61 (1), pp. 1–6. 2017. DOI: 10.3311/PPci.8737

[16] Kutyłowska, M. "Neural network approach for failure rate prediction." *Engineering Failure Analysis*. 47, Part A, pp. 41-48. 2015. DOI: 10.1016/j.engfailanal.2014.10.007

[17] Aydogdu, M., Firat, M. "Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM methods." *Water Resources Management*. 29(5), pp. 1575-1590. 2015. DOI: 10.1007/s11269-014-0895-5

[18] Iwanek, M., Kowalska, B., Hawryluk, E., Kondraciuk, K. "Distance and time of water effluence on soil surface after failure of buried water pipe. Laboratory investigations and statistical analysis." *Eksploatacja i Niezawodnosc-Maintenance and Reliability*. 18(2), pp. 278-284. 2016. DOI: 10.17531/ein.2016.2.16

[19] Kaźmierczak, B., Wdowikowski, M. "Maximum rainfall model based on archival pluviographic records – case study for Legnica (Poland)." *Periodica Polytechnica Civil Engineering*. 60(2), pp. 305-312. 2016. DOI: 10.3311/PPci.8341

[20] Kotowski, A., Wartalska, K., Nowakowska, M. „Generalized analytical method of overfall storm water retention reservoir sizing." Ochrona Środowiska. 38(1), pp. 45-52. 2016. (in Polish)