Periodica Polytechnica
Civil Engineering

61(1), pp. 1–6, 2017
DOI: 10.3311/PPci.8737

RESEARCH ARTICLE

Comparison of Two Types of Artificial Neural Networks for Predicting Failure Frequency of Water Conduits

Małgorzata Kutyłowska[1*]

Received 30-10-2015; accepted 17-05-2016

Abstract

This paper presents the results of a comparison between two artificial neural network structures, i.e. the multilayer perceptron and the ANN with radial basis functions, with regard to the prediction of the failure intensity (failure rate) indicator for water mains, distribution pipes and house connections. The artificial neural network architecture included seven input signals (the number of house connections, the length of water mains, distribution pipes and house connections and the number of their failures). There were three neurons (the failure frequency indicators for the three types of conduits) at the ANN's output. Operating data from the years 1999-2013 were used to train the ANNs while the optimal model was verified using data from the year 2014. Two models (MLP 7-14-3 and RBF 7-4-3), characterized by the best agreement between the predicted results and the experimental ones, were selected from a few tens of models. The RBF ANNs would generate results showing poorer agreement with the experimental failure frequency indicator values.

Keywords

artificial neural network, prediction, water-pipe network

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

1 Introduction

Artificial neural networks (ANNs) are increasingly often used to model many engineering phenomena. According to literature reports, the multilayer perceptron (MLP) is currently one of the best ANN structures used to model various variables in broadly understood environmental engineering, e.g. to forecast hourly water demand time series [1], determine the pressure in a water supply system [2], predict the frequency of failures of a water distribution network [3, 4] and predict the dam water level [5]. So far there have been no reports on the application of radial basis functions (RBF) ANNs to predict the failure intensity (failure rate) indicator for water conduits. Therefore the principal aim of this research was to find out, using a simple example (basic input variables), if an artificial neural network of the RBF type would predict (with an error acceptable for engineering purposes) the failure frequency indicator for water conduits and whether this kind of ANN is competitive with a typical multilayer perceptron. Moreover, the motivation of comparison between RBF and MLP neural networks was to check the behaviour and results of radial basis function which have been also used by author in other regression methods, e.g. support vector machines. Such comparison could be useful in the future investigations in relation to choose the optimal algorithm of failure rate prediction.

1.1 Frequency of failures of water supply networks

The condition and frequency of failures of water distribution networks are very important aspects which should be taken into account in any assessment of the reliability level of water supply systems and in their proper management. Research devoted to the determination of the water conduit failure intensity indicator and other factors having a bearing on the proper functioning of municipal water networks has been conducted in Poland and in the world for many years [6-12]. A widely known and described methodology [6, 13], based on the failure occurrence probability calculus, is used to determine the parameters characterizing the level of reliability of water supply networks. The problem mentioned in practically all the studies on this subject is the necessity to investigate each water distribution system separately since there is no single universal model which would describe failure frequency phenomena in a general way. This is naturally due to the fact that the topology, the materials, the age, the water conduit diameters and the operating and other conditions differ between water supply systems [14, 15]. Nowadays modelling is very popular in relation not only to water supply networks, but also to sewerage systems [16]. That is the reason it seems to be reasonable to model also the failure frequency of water pipelines. Currently many studies dealing with the broadly understood frequency of failures of water conduits and with the risk of such failures are based on mathematical modelling using, e.g., fuzzy logic and artificial neural networks [17, 18]. So far the failure intensity indicator for water conduits has been predicted using MLP ANNs. In this paper the possibility of using RBF ANNs to predict the failure frequency indicator for pipelines is investigated.

1.2 Artificial neural networks

In an artificial neural network of the MLP type, consisting of at least three layers (an input layer, a hidden layer and an output layer), each neuron calculates the weighted average of the signals which reach it. Using the transition function the result is converted and outputted from one layer to another. Such an ANN connects input signals with output signals in a relatively simple way, i.e. through unidirectional connections between neurons. The weights of the connections are adjusted and changed during the training of the ANN in order to invest the model with generalization properties. The advantage is that using an appropriate (for the problem being solved) number of layers and neurons, training method and activation function one can model practically any dependence between the input signals and the output signals [19].

Since the theory of artificial neural networks is described in detail in the literature on the subject [20-22] only essential information on RBF ANNs is presented here. Unlike the multilayer perceptron, RBF ANNs include radial neurons (performing a function radially changing around a given centre in the vicinity of which non-zero values are assumed). Each such neuron models a Gaussian response surface. The information from the inputs is transmitted to a basis function and each neuron calculates the Euclidean distance between the input vectors, the standard vectors and the output vectors. It is important that there is a sufficient number of radial neurons to accurately connect the function with the sought solution. At the ANN's output it is possible to obtain results convergent with the experimental ones when a larger number of neurons and a larger training vector are employed, which is a drawback in comparison with the MLP ANN. Solutions based on RBF ANNs are slow-running and require a considerable storage area, which sometimes constitutes a serious limitation.

2 Investigative methodology

The water conduit failure rate indicator (λ, fail./(km·year)) would be predicted by means of ANNs. The aim of the analysis presented in this paper is to show differences between pipeline failure rate prediction results depending on the type of ANN used (the multilayer perceptron versus RBF ANNs). Operating data from the years 1999-2013, obtained from one of the water companies in Poland, were used to train the ANNs. In the training stage 50 % of the data were used for learning, 25 % for testing and 25 % for validating. The selected model was verified using data from the year 2014. Until now data describing a given water conduit, such as the material, the diameter, the age, the conduit length, the pressure, the type of soil and the conduit laying depth, have been used as input signals to model indicator λ [3, 4, 17]. In this research the input neurons were such parameters as: the number of house connections (NH), the length of water mains (L_m), the length of distribution pipes (L_r), the length of house connections (L_p), the number of failures of water mains (N_m), the number of failures of distribution pipes (N_r) and the number of failures of house connections (N_p). In the output layer there were the failure rate indicators for: water mains (λ_m), distribution pipes (λ_r) and house connections (λ_p). The actual (determined on the basis of the operating data) failure rate indicator would be calculated from this relation:

$$\lambda = \frac{N(t)}{L \cdot \Delta t}, \quad \text{fail.}/(\text{km} \cdot \text{year}) \tag{1}$$

where:

$N(t)$ the number of failures over time t, pc.;
Δt the analyzed period, year;
L the average length of the conduit, km.

One should note that the ANN was trained on variables which have a direct bearing on the value of the failure rate indicator (relation (1)). Unlike in earlier analyses [3, 4, 23], such basic data as the conduit length and the number of failures were used intentionally in order to check for a simple case whether MLP artificial neural networks are really one of the best ANN structures for predicting engineering phenomena. Table 1 shows the ranges of variation of the data fed to the ANN's input and output.

Optimal models for MLP and RBF structures were selected through a two-stage process. For each of the ANN structures twenty models were built and trained on data from the years 1999-2013. Then relative mean-square errors were determined for each of the models. A few models characterized by the lowest errors (for water mains, distribution pipes and house connections) were selected for each of the ANN structures and verified (the forecast stage) using data from the year 2014. The optimal model was characterized by the lowest mean-square error of the forecast and the best agreement between the experimental results and the ones predicted by the ANN. The computations were performed in Statistica 10.0.

Table 1 Ranges of variation of input parameters and actual failure intensity indicator in years 1999-2014

Variable	NH	L_m, km	L_r, km	L_p, km	N_m	N_r	N_p	λ_m, fail./(km·year)	λ_r, fail./(km·year)	λ_p, fail./(km·year)
Min	1954	28.2	99.2	35.4	2	27	15	0.07	0.23	0.34
Max	2721	31.0	118.4	46.3	16	66	46	0.57	0.67	1.01

3 Results and discussion

The average experimental failure intensity (failure rate) indicator for water mains, distribution pipes and house connections in the years 1999-2014 amounted to respectively 0.25, 0.38 and 0.64 fail./(km·year). Using the methodology described above two optimal models (for both the MLP and RBF structure), characterized by the best agreement between the experimental results and the predicted ones, were selected. Table 2 shows the relative errors between the experimental and predicted results for ANN training (the years 1999-2013) and optimal model verification (the year 2014).

Table 2 Relative errors (optimal model MLP 7-14-3 and RBF 7-4-3)

Year	Water mains		Distribution pipes		House connections	
	Relative error, %					
	MLP	RBF	MLP	RBF	MLP	RBF
1999	0.00	2.86	1.49	5.97	1.79	1.79
2000	0.00	4.76	0.00	5.36	0.00	1.56
2001	0.00	0.00	0.00	3.77	1.41	2.82
2002	0.00	2.33	2.27	11.36	1.49	10.45
2003	0.00	1.75	0.00	4.55	0.00	2.94
2004	0.00	4.00	2.86	11.43	0.00	4.29
2005	0.00	5.56	5.41	5.41	1.69	11.86
2006	0.00	0.00	0.00	9.76	0.00	1.25
2007	0.00	6.25	0.00	15.38	1.32	6.58
2008	0.00	7.14	0.00	11.43	1.89	7.55
2009	0.00	5.56	3.03	12.12	2.86	22.86
2010	0.00	71.43	0.00	12.50	0.00	2.94
2011	0.00	0.00	0.00	4.17	0.00	0.00
2012	5.88	17.65	0.00	42.86	1.56	17.19
2013	0.00	0.00	0.00	0.00	0.00	0.99
2014	4.35	34.78	0.00	17.86	1.10	7.69

In the selected MLP model fourteen hidden neurons and three output neurons were activated by the linear function, whereas in the RBF model four hidden neurons and three output neurons were activated using respectively the Gaussian function and the linear function. One should note that in the Statistica program there is really no choice of activation function for modelling by means of RBF ANNs since the Gaussian function and the linear function are set by default [19]. This is a drawback in comparison with the multilayer perceptron to which the following activation functions can be applied: the linear function, the logistic function, the exponential function, the hyperbolic tangent (tanh) and the sine function. Therefore it should not be surprising that larger discrepancies (Table 2) between the experimental results and the predicted ones were obtained in the case of the RBF ANN.

The reason was the limited number of combinations of functions activating hidden and output neurons. Also the choice of a training method poses a difficulty. MLP ANNs can be trained by respectively the conjugate gradient method, the steepest descent method and the quasi-Newton method, which are described in more detail in the literature on the subject [20-22]. In the case considered here, the last algorithm would generate the most convergent results after 28 epochs of training. RBF ANNs are trained completely differently. Training is conducted in two stages [19]: first radial basis functions are arranged using input signals and then weights between the RBFs and the output neurons are determined. Consequently, no iteration process is required, which is evidence of the lack of typical training epochs. The number of hidden neurons in MLP ANNs does not depend on the number of input and output signals. Even though the number of neurons in the hidden layer is customarily determined to some degree in relation to the neurons in the other layers, the relations are not *a priori* known. They are experimentally determined, often by trial and error. When determining the maximum number of hidden neurons during modelling by means of RBF ANNs the following program message appeared: "the largest possible number of neurons is the sum of the training cases and the output cases". The compliance with this message seriously limits the choice of the number of signals in the hidden layer. The ANN should not be overtrained since this would contribute to the loss of its generalization abilities. This can happen if the hidden vector is too large. On the other hand, the imposition of too severe limitations already at the start results in lower learning quality and subsequently adversely affects model verification results, i.e. they may be less convergent with the experimental results, as shown in Figs. 1-3. For comparison, Table 3 shows the experimental values of the failure intensity indicator and the ones predicted by the ANN.

An analysis of Tables 2-3 and Figs. 1-3 shows that as regards water conduit failure rate indicator prediction larger discrepancies occur when the RBF ANN is used. In the training stage the

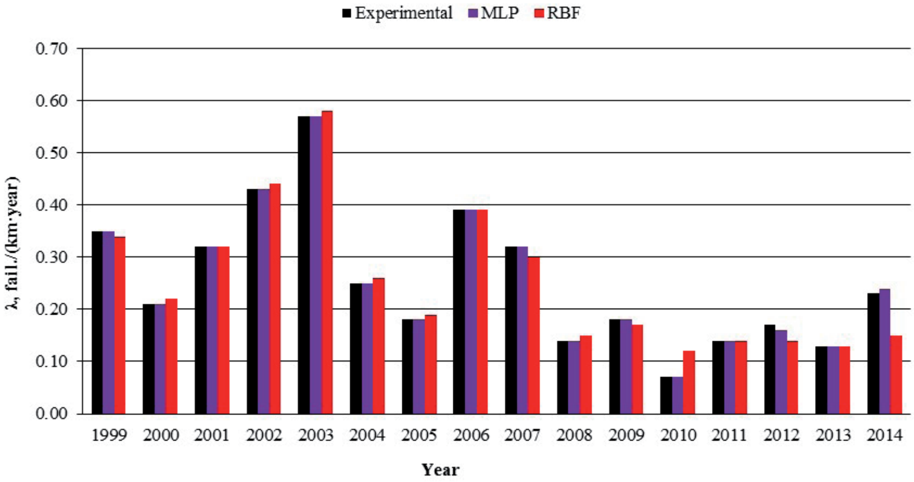

Fig. 1 Experimental and predicted failure rates of water mains

correlation between the experimental results and the ones predicted using the MLP 7-14-3 model was high, i.e. R^2 amounted to 0.999, 0.998 and 0.999 for water mains, distribution pipes and house connections, respectively. In the training process the RBF 7-4-3 model was characterized by a worse correlation, amounting to 0.981, 0.808 and 0.942.

Table 3 Output signals
(experimental values vs. optimal model MLP 7-14-3 and RBF 7-4-3)

Year	Water mains			Distribution pipes			House connections		
	λ, fail./(km·year)								
	Exp.	MLP	RBF	Exp.	MLP	RBF	Exp.	MLP	RBF
1999	0.35	0.35	0.34	0.67	0.66	0.63	0.56	0.57	0.55
2000	0.21	0.21	0.22	0.56	0.56	0.59	0.64	0.64	0.63
2001	0.32	0.32	0.32	0.53	0.53	0.55	0.71	0.70	0.73
2002	0.43	0.43	0.44	0.44	0.45	0.49	0.67	0.66	0.74
2003	0.57	0.57	0.58	0.44	0.44	0.46	0.68	0.68	0.70
2004	0.25	0.25	0.26	0.35	0.36	0.39	0.70	0.70	0.73
2005	0.18	0.18	0.19	0.37	0.39	0.39	0.59	0.60	0.66
2006	0.39	0.39	0.39	0.41	0.41	0.37	0.80	0.80	0.79
2007	0.32	0.32	0.30	0.26	0.26	0.30	0.76	0.77	0.71
2008	0.14	0.14	0.15	0.35	0.35	0.31	0.53	0.54	0.57
2009	0.18	0.18	0.17	0.33	0.34	0.29	0.35	0.36	0.43
2010	0.07	0.07	0.12	0.32	0.32	0.28	0.34	0.34	0.35
2011	0.14	0.14	0.14	0.24	0.24	0.25	0.36	0.36	0.36
2012	0.17	0.16	0.14	0.42	0.42	0.24	0.64	0.63	0.75
2013	0.13	0.13	0.13	0.23	0.23	0.23	1.01	1.01	1.02
2014	0.23	0.24	0.15	0.28	0.28	0.23	0.91	0.90	0.98

The value of indicator λ for water mains predicted by the MLP ANN is practically the same as the experimental value. This applies to both the training stage and the year 2014 (the forecast stage). In the training process the RBF ANN would generate discrepancies between the experimental results and the predicted ones, ranging from 0.00 to 71.43 %. The verification of the model in 2014 was characterized by an error of nearly 35 %. The above facts (Fig. 1) are evidence of the worse fit between the failure frequency indicator experimentally determined for water mains and the one predicted by the RBF ANN despite the use of basic variables (e.g. the conduit length and the number of failures) as the training vector. As opposed to proven MLP ANNs [3, 4, 17, 23], one can expect that if more detailed information about the pipeline (e.g. the material, the diameter, the conduit age and pressure) was included in the training data, the discrepancies would increase since the RBF ANN (under all the constraints described above) would not be able to correctly identify the dependences between the input vector and the output vector.

Also in the case of failure frequency indicator prediction for distribution pipes (Fig. 2) the multilayer perceptron performs better. The relative discrepancies between the experimental values and the predicted ones did not exceed 5.5 % while for optimal model verification they amounted to as little as 0.00 %. Whereas the RBF ANN was characterized by the largest error, amounting to almost 43 %. In the case of forecasting, the error exceeded 17 %. In the case of the RBF ANNs such factors as: the use of only the Gaussian function and the linear function to activate neurons, practically no influence on the size of the hidden vector and the different training algorithm, could have contributed to the relatively low quality of training and verification. In artificial neural networks with RBFs the latter are determined on the basis of the input vector and after totalling the weights the result is fed to the output. The basis function location and width and the weights connecting the basis functions with output signals are important in RBF ANNs [20]. In the future it should be checked whether the above aspect, i.e. the fact that the function value usually depends on solely the distance from a given point, is significant and perhaps responsible

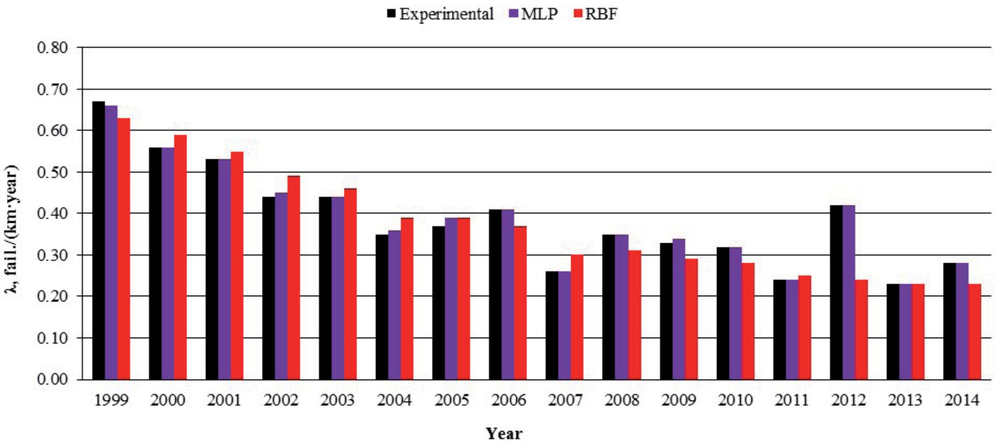

Fig. 2 Experimental and predicted failure rates of distribution pipes

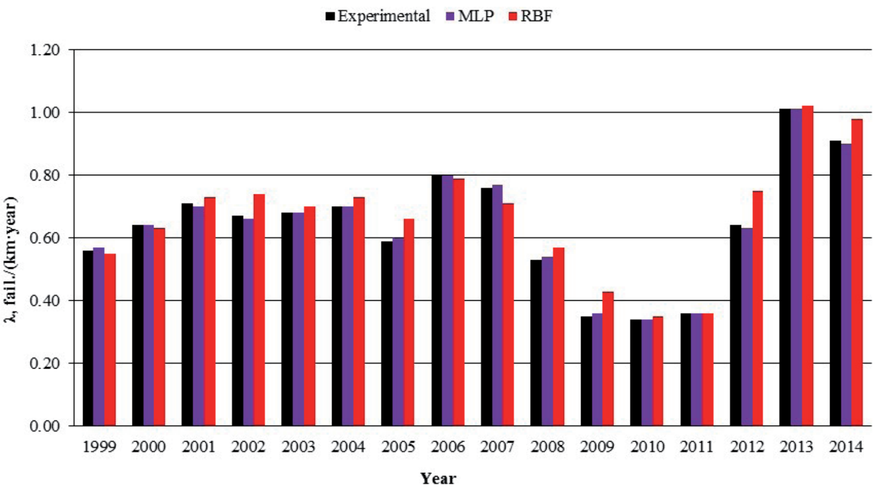

Fig. 3 Experimental and predicted failure rates of house connections

for the worse convergence than that of MLP ANNs in which the neuron activation and training mechanisms are based on different principles, consisting in single-stage (not two-stage as in the case of RBF ANNs) model construction.

The differences between the failure intensity indicator values for house connections (Fig. 3) obtained from the MLP and RBF ANNs are not so large as for the larger-diameter conduits, but also in this case the typical multilayer perceptron shows a better convergence. The maximum error (the year 2009) amounted to less than 3 % and more than 22 % for respectively the MLP ANN and the RBF ANN. When determining the suitability of a given ANN model one should take into consideration the consequences of an incorrect estimate of output data. If from the engineering point of view, the house connection failure frequency indicator prediction error (the RBF ANN) is acceptable, the same model generates distribution pipes failure rate indicator results showing much larger discrepancies from the experimental results. It is obvious that a failure of a distribution pipeline will have more serious consequences than a failure of a house connection. In further studies one should take into account also the type of failure, the place and time of its occurrence and the number of people affected by the failure. Since the aim is to select a single model predicting the failure frequency indicator for each of the three types of conduits (as opposed to the proposed modelling of indicator λ for separately distribution pipes and house connections [4, 23]) the model should be characterized by good agreement between the experimental results and the predicted ones for pipelines whose failure frequency should be more precisely estimated considering the quality and reliability of water supply to the much greater number of water customers.

4 Conclusions

A comparative analysis of two artificial neural network structures: the multilayer perceptron and the RBF ANN, which were used to predict the failure intensity indicator for water conduits, was presented. The analysis of the modelling results can be summarized as follows:

- from a few tens of artificial neural network models two optimal models (MLP 7-14-3 and RBF 7-4-3), characterized by the best convergence with the experimental results and by the smallest mean-square error, were selected;

- RBF ANNs show a lower convergence between the predicted results and the experimental ones than MLP ANNs;
- the maximum relative error between the experimental results and the ones predicted by the MLP ANN amounted to 5.88 %, 5.41 % and 2.86 % for respectively water mains, distribution pipes and house connections, whereas in the case of the RBF ANN the respective errors amounted to 71.43 %, 42.86 % and 22.86 %; the results indicate that unlike RBF ANNs which are rather not recommended for predicting the failure rate indicator, the multilayer perceptron can be used to model the failure frequency of water conduits;
- owing to the use of such basic parameters as the conduit length and the number of failures in the input vector the performance of the RBF ANN was checked in a simple way;
- the inclusion of other water conduit variables in the training data probably would not result in any significant improvement in the agreement between the experimental results and the ones predicted by the RBF ANN since the limitations in the structure of this model (e.g. the imposed activation function type or number of hidden neurons) would still remain;
- nevertheless the author intends to check whether more accurate input data will have a beneficial effect on the generalization and prediction abilities of RBF ANNs when the latter are applied to another water distribution system.

Acknowledgement

The work was realized within the allocation No. B50519 awarded for Faculty of Environmental Engineering Wroclaw University of Science and Technology by Ministry of Science and Higher Education. The grant was allocated for scientific researches of young scientists in years 2015-2016.

References

[1] 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

[2] Ridolfi, E., Servili, F., Magini, R., Napolitano, F., Russo, F., Alfonso, L. "Artificial neural networks and entropy-based methods to determine pressure distribution in water distribution systems." *Procedia Engineering*. 89, pp. 648-655. 2014. DOI: 10.1016/j.proeng.2014.11.490

[3] 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

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

[5] 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

[6] Hotloś, H. "Ilościowa ocean wpływu wybranych czynników na parametry i koszty eksploatacyjne sieci wodociągowych." (Quantitative assessment of the effect of some factors on the parameters and operating costs of water-pipe networks.) Wrocław University of Technology Publishing House, Wrocław, Poland, 2007. (in Polish)

[7] 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)

[8] Le Gat, Y., Kropp, I., Poulton, M. "Is the service life of water distribution pipelines linked to their failure rate?" *Water Science and Technology: Water Supply*. 13(2), pp. 386-393. 2013. DOI: 10.2166/ws.2013.089

[9] Dell'Orfano, F., Esposit, V., Gualtieri, P., Pulci Doria, G. "Mean values of water pipe breakage rates around the world and in geographical areas." *Water Science and Technology: Water Supply*. 14(5), pp. 766-777. 2014. DOI: 10.2166/ws.2014.031

[10] Kutyłowska, M. "Modelling of failure rate of water-pipe networks." *Periodica Polytechnica Civil Engineering*. 59(1), pp. 37-43. 2015. DOI: 10.3311/PPci.7541

[11] Gould, S. J. F., Boulaire, F. A., Burn, S., Zhao, X. L., Kodikara, J. K. "Seasonal factors influencing the failure of buried water reticulation pipes." *Water Science and Technology*. 63(11), pp. 2692-2699. 2011. DOI: 10.2166/wst.2011.507

[12] 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

[13] Kwietniewski, M., Roman, M., Kłoss-Trębaczkiewicz, H. "Niezawodność wodociągów i kanalizacji." (The reliability of water and sewer systems.) Arkady, Warszawa, Poland, 1993. (in Polish)

[14] Pietrucha-Urbanik, K. "Failure analysis and assessment on the exemplary water supply network." *Engineering Failure Analysis*. 57, pp. 137-142. 2015. DOI: 10.1016/j.engfailanal.2015.07.036

[15] 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

[16] 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

[17] Shirzad, A., Tabesh, M., Farmani, R. "A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks." *KSCE Journal of Civil Engineering*. 18(4), pp. 941-948. 2014. DOI: 10.1007/s12205-014-0537-8

[18] Boryczko, K., Tchórzewska-Cieślak, B. "Analysis of risk of failure in water main pipe network and of delivering poor quality water." *Environment Protection Engineering*. 40(4), pp. 77-92. 2014. DOI: 10.5277/epe140407

[19] Statistica 10.0, Electronic Manual

[20] Kutyłowska, M. "Prognozowanie parametrów procesu separacji membranowej z wykorzystaniem sztucznych sieci neuronowych." (Forecasting the parameters of membrane separation processes using artificial neural networks.) PhD Thesis, Wrocław, Poland, 2010. (in Polish)

[21] Osowski, S. "Sieci neuronowe." (Neural networks.) Warszawa University of Technology Publishing House, Warszawa, Poland, 1994. (in Polish)

[22] Hertz, J., Krogh, A., Palmer, R. "Introduction to the theory of neural computation." Addison-Wesley Publishing Company, Redwood City, 1991.

[23] Kutyłowska, M. "Prediction of water pipes failure rate using artificial networks." *Instal*. 1, pp. 44-46. 2016. (in Polish)