

Abstract

This paper presents Artificial Neural Network (ANN) and Multiple Linear Regression (MLR) methods for predicting future daily water consumption values based on three antecedent records of water consumption and humidity forecast for a given day, which are considered as independent variables. Mean Absolute Percentage Error (MAPE) is obtained for different configurations of the input sets and of the ANN model structure. Additionally, sets of explanatory variables are enhanced with dummy variables indicating typical days: working day, Saturday, Sunday/public holidays. The results indicated the superiority of the ANN approach over MLR, although the observed difference in performance was very limited.

Keywords

artificial neural networks, multiple-linear regression, water consumption

1 Introduction

The magnitude of water demand results from a number of factor types, such as: socio-economic (number of customers, water price, ecological awareness of society); infrastructure (length of distribution network, number of connections, water quality, water losses [1]); economic (number and size of industrial facilities); and climatic (rainfall, temperature, humidity, evaporation). The climatic factor is gaining importance with the advent of more severe climate variability [2].

A number of forecasting approaches have been developed and implemented towards the needs of one-day-ahead forecasting of water consumption. So far, the models and methods have been extensively presented in [3]. Current research is focused on hybrid methods which combine Artificial Neural Networks (ANNs) with Wavelet Transform (WT) [4, 5] or ANNs coupled with Fuzzy Logic [6]. In general, a combination of different forecasting approaches is capable of extracting more information available for prediction; Wang, X et al. [7] found that hybrid models outperformed a classic, single-method approach.

At this point, it is worth mentioning that, for different forecasting models, different sets of input variables are used. Only a limited number of authors has focused exclusively on historical water demand as an input to their models [8, 9]. Instead, most models utilize meteorological variables such as temperature, evaporation or humidity [10, 11], or socio-economic variables, as in two papers: Firat et al. [12] who has also compared different ANN models; and [13], where stress was put on the identification of prominent explanatory variables in ANN models.

The literature overview indicates that the current trend in water demand forecasting is concentrated on the development of novel methods which will enable more accurate forecasting. In many cases those new approaches are based on ANN. Therefore, the aim of this paper was not to create an utterly new method, but to compare results obtained from two popular models of ANN and MLR. The paper aims to verify the efficacy of those methods in the case of a relatively small urban unit. The city of Torun, which is representative of water and sewage management in Poland, was selected as the object of study.

1 Department of Economics, Finance, and Environmental Management,
Faculty of Management, AGH University
30 Adama Mickiewicza Ave., 30-059 Kraków, Poland

2 Department of Engineering Management,
Faculty of Management, AGH University
30 Adama Mickiewicza Ave., 30-059 Kraków, Poland

3 Faculty of Environmental Engineering,
Wrocław University of Science and Technology,
27 Wybrzeże Wyspiańskiego Street, 50-370 Wrocław, Poland

* Corresponding author, email: adm.piasecki@gmail.com

2 Object of study

Torun is one of the oldest cities in Poland, with a very long and rich history. It is situated in northern central Poland on the Vistula River. Its current population is about 200,000, inhabiting an area of 115.75 square kilometres.

The history of water and wastewater network in Torun is almost as long as the city itself. Wooden pipes providing water to the city were put in operation at the beginning of the 14th century. Their existence has been proven by archeological works conducted in the city center. Some parts of this wooden pipes network has been replaced by iron pipes as late as the beginning of the 19 century.

Despite an intensive development of the waterworks, since 1990s a systematic decrease in consumed water has been observed. In the year 1990 (the second year of political transformation in Poland) the water usage in Torun was close to 45.6 hm³, but nowadays it is as small as 23% of this value. To the main reasons of such drastic change one may account:

- a political and economic transformation which led to decommissioning of a multitude of industrial facilities in the whole country but also in Torun,
- the waterworks company has changed its charging method – which now will be based on reading from water meters,
- a systematic increase of water prices for individual and industrial consumers,
- ecological awareness-raising campaigns,
- dissemination of modern water saving households appliances in form of washing machines or dishwashers.

In recent years there has been a dynamic development of the length of the waterworks, which currently amounts to 549.2 kilometres (in the year 1990 it was a mere 189.5 kilometres) [14]. As indicated, the significant increase in the length of the water supply network, and hence also the number of new customers, did not translate into halting or reversing the trend of declining water consumption. The previously mentioned factors reducing the demand for water in the city were much more important and the emergence of new customers did not compensate for the decline caused.

3 Materials and Methods

The data used to water demand forecasting was obtained directly from water company Toruńskie Wodociągi LLC and covered the years 2011–2013, totalling 1,096 records. The mean daily water consumption over this period was 36,495 m³ with a standard deviation of 3,488 m³. Note the increase in water demand during the summer months (see Fig. 1) and the decreasing overall trend of water consumption over this three-year period. Because this study focusses on creating day-ahead forecasts, a detrending procedure was used on a raw water consumption data set prior to the model construction. Detrending was performed by means of Matlab software's inbuilt function $y = \text{detrend}(x)$. Additionally, as a part of data pre-processing, all values were normalized to a 0–1 range.

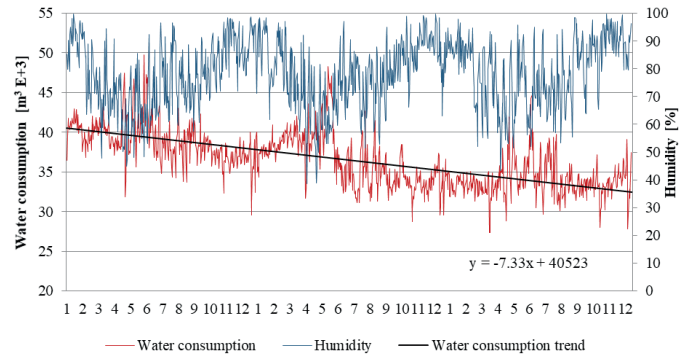


Fig. 1 Variability of water consumption and humidity over the period 2011–2013

Values for meteorological parameters were obtained from IMGW (the Institute of Hydrology and Water Management). Their impact on water consumption has been assessed on the basis of their correlation coefficient (Table 1). Only humidity has a theoretical impact on water consumption (Pearson correlation: -0.413). Thus, the remaining variables (temperature, rainfall and wind speed) were not considered as a part of the input sets for either ANN or MLR models. The value of the Pearson correlation for humidity and water consumption was then used as a threshold for assessing the significance of other explanatory variables. In practice it meant that only those antecedent volumes of water consumption whose correlation coefficient with water consumption exceeded -0.413 in absolute terms were used as part of the input set. Calculations have shown that $t-1$, $t-2$ and $t-3$ values ($t-1$ and so on are historical water consumption values, here $t-1$ refers to one day earlier) of water consumption exhibit, respectively, a 0.696, 0.544 and 0.440 correlation with water demand in the period t .

Table 1 Correlation coefficients

| | Temperature | | Rainfall | Wind speed | Humidity |
|-------------------|-------------|-------------|----------|------------|----------|
| | Mean | Min. Max. | | | |
| Water consumption | 0.242 | 0.119 0.284 | -0.040 | -0.012 | -0.413 |

A mean daily water consumption over the typical week in considered water system exhibits significant variability (Fig. 2). In Torun the lowest water demand can be observed during Sunday and public holidays (PH). This led to the conclusion that dummy variables should be included informing the model of what kind of day a given forecast will be performed for. Accordingly, working days (Monday–Friday) were tabbed with "1", Saturdays with "2" and others with "3". Analysis of the annual variability of monthly water consumption has shown that it remains at the same level from January to April, then reaches its peak in May and exhibits a decreasing tendency till November, before picking up again in December. A procedure of assigning dummy variables, similar to that used for days, was performed, i.e., January = 1, February = 2, and so on.

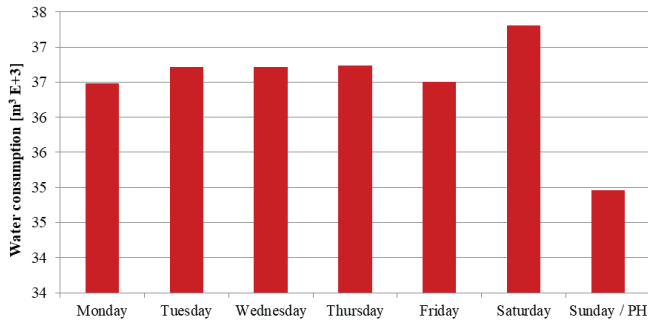


Fig. 2 Water demand variability over a typical week

To sum up, the ANNs and MLR input sets consisted of three antecedent values of water consumption, meteorological variable (humidity) and two subsets of dummy variables denoting recognized day pattern and month.

The subdivision of data into training, validation and testing (as required by the ANN creation approach) is an open research question [15, 16]. In this paper, an approach based on integer programming was applied (which is, to our knowledge, a novelty). The assumption was that the ratio of number of training, validation and testing subsets should be 70:15:15, and subsets' mean and standard deviation must be equal. This sort of model has no objective function, and is a so-called feasibility problem. The proposed mathematical model is formulated as follows:

variables: $a_{i,j}$ – binary variable (0–1) indicating membership to subset $i = 1..n$, $n = 3$, where 1 is training, 2 is validation and 3 is the testing subset of water consumption in the period $j = 1..m$, $m=1096$; W_j – water consumption in the period j ; μ_i – a subset i mean.

subject to:

$$\sum_{j=1}^m a_{1,j} = \frac{14}{3} \sum_{j=1}^m a_{2,j} \quad (1)$$

$$\sum_{j=1}^m a_{2,j} = \sum_{j=1}^m a_{3,j} \quad (2)$$

$$\frac{1}{0.7 * m} \sum_{j=1}^m a_{1,j} W_j = \frac{1}{0.15 * m} \sum_{j=1}^m a_{2,j} W_j \quad (3)$$

$$\frac{1}{0.15 * m} \sum_{j=1}^m a_{2,j} W_j = \frac{1}{0.15 * m} \sum_{j=1}^m a_{3,j} W_j \quad (4)$$

$$\sqrt{\frac{1}{0.7 * m} \sum_{j=1}^m a_{1,j} (W_j - \mu_1)^2} = \sqrt{\frac{1}{0.15 * m} \sum_{j=1}^m a_{2,j} (W_j - \mu_2)^2} \quad (5)$$

$$\sqrt{\frac{1}{0.15 * m} \sum_{j=1}^m a_{2,j} (W_j - \mu_1)^2} = \sqrt{\frac{1}{0.15 * m} \sum_{j=1}^m a_{3,j} (W_j - \mu_3)^2} \quad (6)$$

Equations (1) and (2) ensure that the numerousness of training, validation and testing subsets will remain at the desired ratio. Subsequent formulas guarantee that subsets will have the same mean (3–4) and standard deviation (5–6). Execution of

the presented model divided the set into a training set with 767 records, a validation set with 165 records and a testing set with 164 records. During the feasibility problem solving not all constraints (1–2) were satisfied, but the resulting sets were equal in terms of mean and standard deviation.

The MLR and ANNs models were constructed in MS Excel and Statistica Software, respectively. The procedure for MLR creation is a well-known concept and has been thoroughly described in [17]. MLR models are perceived as a complicated curve-fitting exercise.

A more recent concept of Artificial Neural Networks is presented in detail in [18]. One of the most controversial issues when it comes to the creation of ANN models is the selection of the number of neurons in the hidden layer [19, 20]. According to [21], a greater number of neurons in the hidden layer may lead to an ANN model which is capable of solving a problem faster, but not always more accurately. In general, the procedure for ANN model creation should start with a smaller number of neurons in the hidden layer which is gradually increased whilst monitoring the results. Shibata and Ikeda [22] proposed a simple formula for this problem; the minimal number of neurons in a hidden layer is equal to the root of the product of the number of neurons in the input and output layers, while the maximal number is equal to three times the minimal number plus one.

The second issue is the selection of transfer functions. Here, four different activation functions were considered, namely: linear, logistic, hyperbolic tangent and exponential. As a result, including the ANN model, where the input set consisted of humidity, three antecedent water consumption levels and dummy variables (19 neurons in input layer) and the models without humidity in the input set (18 neurons in the input layer), a total of 176 ANN architectural configurations were considered. In both cases, the number of neurons ranged from 4 to 14 in the hidden layer, while there was one in the output layer, and four different activation functions were considered for the input and output layers. According to combinatorics principles it gives: $11 \times 4 \times 4$ configurations of architecture. Based on the Statistica Software's inbuilt capabilities, ANNs were trained based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm and their performance during the creation process was assessed based on the sum of the square of errors. This study considered only the Multi-Layer Perceptron (MLP) type of ANNs models, due to its popularity [23, 24] and appreciation in the area of forecasting. As a stopping condition (criteria) during the ANN model creation procedure a minimal change in forecasting error was used. To continue the learning process the change in error should be greater than 0.000001% otherwise the process will be stopped.

In order to assess the forecasting performance of the models created, a mean absolute percentage error (MAPE) (7) criterion was used, where W_j refers to observed water consumption and W_j^* is predicted water consumption, both during day j . This criterion was applied only to testing subsets.

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{W_j - W_j^*}{W_j} \right| \quad (7)$$

Additionally, a commonly used in hydrological time series forecasting Nash-Sutcliffe [25] model efficiency coefficient has been calculated based on Equation 8.

$$E = 1 - \frac{\sum_{j=1}^n (Q_m^j - Q_o^j)^2}{\sum_{j=1}^n (Q_o^j - \bar{Q}_o)^2} \quad (8)$$

where: E – Nash-Sutcliffe model efficiency coefficient, Q_m^j – modelled discharge in day j (here water used in waterworks), Q_o^j – observed discharge, \bar{Q}_o – is the mean of the observed discharges.

Nash-Sutcliffe efficiency ranges from $-\infty$ to 1. An efficiency of $E = 1$ translates into a perfect match of modelled discharge to the observed one. An efficiency of $E = 0$ means that the model predictions are as accurate as the mean of the observed data, whereas an efficiency $E < 0$ occurs when the observed mean is a better predictor than the considered forecasting model.

Apart from MAPE and E criteria, the ANN and MLR models have been compared to the forecasting performance of two naïve methods. Namely, using previous day or the same day in the previous week as the forecasted values.

4 Results and discussion

Based on the chosen performance criterion (MAPE), the two ANN and MLR models with the lowest MAPE values were selected. Their characteristics/equations are presented in Table 2. Additionally, in Table 2 one can observe that the testing performance for ANN models, calculated as a Pearson coefficient, indicates that a model with an additional explanatory variable (humidity) performed better than one without it.

Calculated values of the MAPE criterion for all models based on testing subset values indicated that the performance of ANN and MLR approach was similar. The lowest value was obtained by the ANN_H model and amounted to 2.28%, whereas the highest was for an ANN model without humidity as an explanatory variable and was equal to 2.69%. The two remaining models performed as follows: MLR_H – 2.49% and MLR – 2.56%. These results indicate that, in general, models with the additional meteorological variable performed slightly better than those without it.

A more detailed results analysis focused on error values within a given type of day and month. As can be seen in Fig. 3, the ANN_H model is superior to the three others. Interestingly, the application of humidity as part of the explanatory variables set for working days and Saturdays significantly improved the results of ANN models, while this phenomenon is not observed for MLR models during Sundays and public holidays (PH). Similarly, (see Fig. 4 and 5), the ANN_H model

outperformed other models from the perspective of MAPE criterion within selected months, with the exception of January, where ANN and MLR proved to be more accurate.

Table 2 General description of forecasting models

| Model | Structure | Learning Performance | Testing Performance | Validation Performance | Algorithm | Input Fun. | Output Fun. |
|-------|--|----------------------|---------------------|------------------------|-----------|------------|-------------|
| ANN_H | 19-6-1 | 0.871 | 0.874 | 0.894 | BFGS 62 | Log | Exp. |
| ANN | 18-7-1 | 0.779 | 0.816 | 0.837 | BFGS 13 | Log | Linear |
| MLR_H | $W_j^* = 3320.6 - 190.8b_1 - 430.6b_2 - 1132b_3 - 1169.7b_4 - 323.3b_5 - 877.8b_6 - 1112.3b_7 - 1070.3b_8 - 884.5b_9 - 644.3b_{10} - 273.7b_{11} + 0b_{12} + 2544.6a_1 + 1503.6a_2 + 0a_3 - 61.8H_j + 0.072W_{j-3} + 0.032W_{j-2} + 0.57W_{j-1}$ | | | | | | |
| MLR | $W_j^* = -2084.4 - 220.4b_1 - 12.1b_2 - 118.3b_3 + 149.3b_4 + 966.7b_5 + 151.2b_6 - 347.9b_7 - 64.1b_8 - 287.3b_9 - 290.6b_{10} - 370.8b_{11} + 0b_{12} + 2485a_1 + 1464.6a_2 + 0a_3 + 0.057W_{j-3} + 0.012W_{j-2} + 0.63W_{j-1}$ | | | | | | |

Where: W_{j-1} , W_{j-2} , W_{j-3} – water consumption from three antecedent days, H_j – humidity at day j , $b_{1,12}$, $a_{1,3}$ – dummy variables. ANN_H and MLR_H stand for models which used humidity as an additional variable in input set, BFGS 62 stands for-Broyden-Fletcher-Goldfarb-Shanno which solved the weights vector optimization problem after the 62nd iteration.

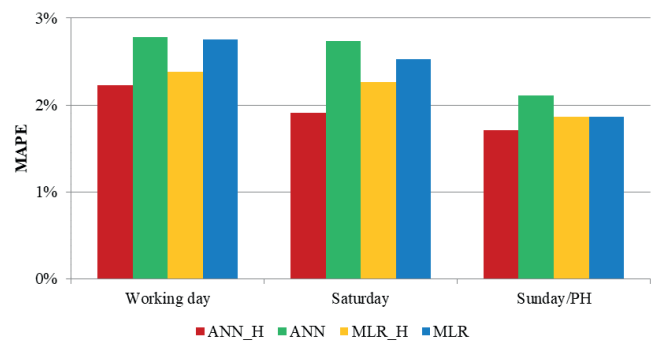


Fig. 3 Values of MAPE within a given day

The conducted research indicates that an ANN approach is marginally superior to MLR models. However, all ANN and MLR models are superior to naïve method based on the previous day or the same day from the previous week. Respectively, calculated MAPE values were slightly above 4% and close to 5.5%. In the case of the Nash-Sutcliffe statistic the ANN model using humidity as an additional explanatory variable has a highest E value of 0.753. It does not indicate a perfect match between observed and modelled phenomena and shows that there may exist explanatory some variables which may improve forecasting accuracy. The second highest values of the E criterion was observed in case of a MLR considering humidity ($E = 0.702$). Models which did not use humidity exhibited similar forecasting performance in terms of Nash-Sutcliffe statistic. Respectively, 0.660 for an ANN model, 0.678 for

a MLR model and 0.658 for naïve method considering the previous day. By considering only this statistic we can see that the performance of more sophisticated models (ANN and MLR) is not significantly greater than that from a naïve method. But in terms of the MAPE criterion they are superior. At the end it is worth to add that the naïve method considering the same day from the previous week exhibited very weak performance in terms of Nash-Sutcliffe statistic and its E value was 0.351.

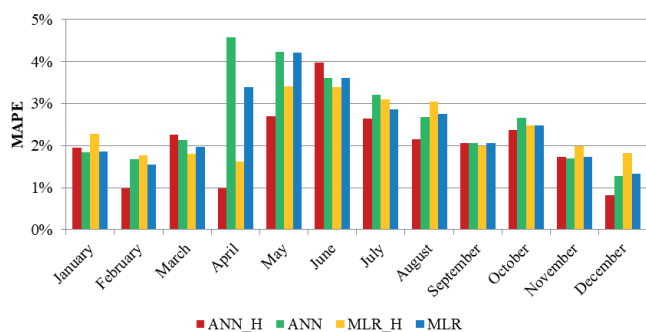


Fig. 4 Values of MAPE within a given month

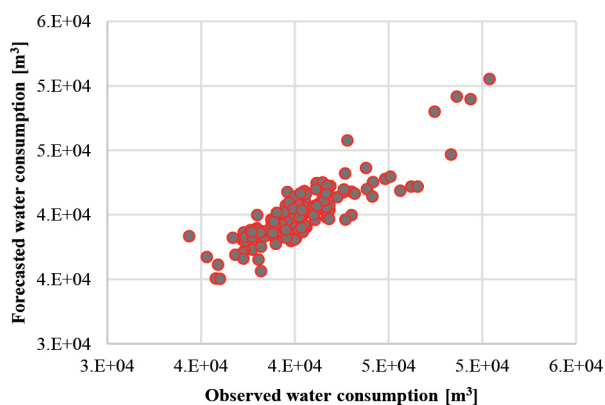


Fig. 5 Scatter plot for ANN_H model for testing subset

What is more, in the case of universal models (i.e., those created for a whole year) the application of the additional meteorological variable slightly improves model performance. The usage of humidity parameter seems to be an important factor from several reasons. It provides a direct information about the water deficit in the air which is especially important during hot summer days when garden and lawns watering is common. Additionally this parameter is commonly available, since it is being measured at a majority of meteorological stations. Based on that, it seems that humidity is a sound alternative to evaporation for which estimation various formulas and several other meteorological parameters are used. However, this study assumed that humidity, which is unknown prior to day j , can be forecasted with 100% accuracy. Taking into account errors in weather forecasts may remove the positive effect of using this variable in water demand forecasting.

The analysis showed that ANN is a suitable method for forecasting water demand in a small city. This confirmed the results of research conducted by other academics for greater cities in various climate zones [11, 13]. The main difference

is the structure of input variables sets. In the case of Torun, all meteorological variables except humidity turned out to be irrelevant. This is a result of the local climate. One good example is precipitation, which is characterized by significant variability on a quantitative and temporal scale. The major volume of rainfall occurs in the spring-to-summer months, but is delivered in relatively short and intensive events [26]. Therefore, the consumption of water for watering lawns or gardens is significantly distorted.

Forecasting water demand in a smaller city seems to be a more complicated task than in multi-million population cities. This results from the short-term dynamics of water consumption resulting from population mobility. Short trips out of town may significantly change water demand. In Torun these mainly occur during public holidays (family holidays) and during favourable weather in summer. On such days, water consumption may decrease by several dozen percent. In greater cities, the fluctuations in water-consuming population are greater in absolute terms but in percentage terms are relatively minor.

The ability to accurately forecast water demand is important from the perspective of all waterworks. The full equipment (both used to water treatments and delivery) readiness contributes significantly to the water price. Therefore, waterworks should have access to detailed and accurate water demand forecast. Short-term forecasts (hours, days) are used mainly to ensure an effective operation of the waterworks system and especially in order to minimize the time when the water is stored in pipelines and reservoirs. What directly translates into its quality [13].

In the case of Torun and other Polish cities, this is particularly important due to a declining tendency in water demand, as observed over recent years. In consequence, the hydraulic conditions in the water distribution system have changed and this is a serious threat to water quality which could change seasonally [27]. The application of accurate forecasting models allows the optimization of the water distribution process, assuring its appropriate quality whilst minimizing the occurrence of failures and risk [28] as well as increasing the reliability of the whole water supply system (from water intakes to households) [29, 30].

This paper presents an approach based on two forecasting methods. Their selection resulted not only from their performance as highlighted in previous papers, but also from their applicability in various cities [31]. The plethora of methods which appeared in the literature makes it very hard to point out the most accurate one. However, it seems that the precision obtained by the ANN approach is sufficient for practical application.

5 Conclusions

Two modeling strategies are employed in this paper to forecast water demand in Torun's waterworks system, where the models comprise Multi-Linear Regression and Artificial

Neural Networks. For both strategies, the same sets of input variables were used in order to single out the most effective approach to forecasting over a given period. The ANN model proved to be superior to the MLR approach; however, the difference in terms of MAPE criterion was very limited. This study also presents a novel approach to the division of sets into testing, training and validation subsets. Future studies should focus on analyzing the impact of humidity-forecast accuracy on water-demand predictions.

The results obtained in this research have been presented to the people responsible for managing Torun Waterworks Company. However, the potential economic benefits of improved water demand forecasts have not been estimated so far. Therefore, further works are needed in order to develop precise models for medium and long term water demand forecasting (months, years). Such forecast may deliver an important information which will impact the company operational decisions or even its strategy for future. What is more, a detailed knowledge about the future event facilitates the creation of company internal and external investment policy.

Because water demand shows a decreasing trend in Torun, it is important to accurately predict its real level. The conducted analysis confirmed the good accuracy of ANN and slightly lesser accuracy of MLR models. However, both those approaches were superior in comparison to a benchmark approach (previous day or the same day in the previous week as the naïve forecast).

The applied in this paper forecasting methods can be also used for different parts of the water and wastewater system in each city. It is especially important in case of the volume of wastewater which is delivered to the wastewater treatment plants or the changes of the water pressure in pipelines. In both cases accurate forecasts can provide important information for people involved in various decision making processes.

Acknowledgments

This research uses data from Torun Waterworks Company Ltd. (TWC). The authors thank the IT Specialists from TWC for assistance in obtaining these data.

References

- [1] 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". *Eksploracja i Niezawodność – Maintenance and Reliability*, 18(2), pp. 278–284. 2016. <https://doi.org/10.17531/ein.2016.2.16>
- [2] Misra, A. K. "Climate change and challenges of water and food security". *International Journal of Sustainable Built Environment*, 3(1), pp. 153–165. 2014. <https://doi.org/10.1016/j.ijbsbe.2014.04.006>
- [3] Donkor, E., Mazzuchi, T., Soyer, R., Alan Roberson, J. "Urban Water Demand Forecasting: Review of Methods and Models". *Journal of Water Resources Planning and Management*, 140(2), pp. 146–159. 2012. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000314](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000314)
- [4] Adamowski, K., Adamowski, J. F., Seidou, O., Ozga-Zieliński, B. "Weekly urban water demand forecasting using a hybrid wavelet–bootstrap–artificial neural network approach". *Annals of Warsaw University of Life Sciences, Land Reclamation*, 46(3), pp. 197–204. 2014. <https://doi.org/10.2478/ssgw-2014-0016>
- [5] Mohammed, H. J. R., Ibrahim, H. M. "Hybrid wavelet artificial neural network model for municipal water demand forecasting". *ARPN Journal of Engineering and Applied Sciences*, 7(8), pp. 1047–1065. 2012. <https://pdfs.semanticscholar.org/38c4/8cc5e58420488fce727e6c8d7c-401ca8c90d.pdf>
- [6] Tabesh, M., Dini, M. "Fuzzy and neuro-fuzzy models for short-term water demand forecasting in Tehran". *Iranian Journal of Science and Technology*, 33(B1), pp. 61–77. 2009. <https://doi.org/10.22099/IJSTC.2009.701>
- [7] Wang, X., Sun, Y., Song, L., Mei, C. "An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. Part II. Model application and results". *Journal of Environmental Management*, 90(8), pp. 2612–2619. 2009. <https://doi.org/10.1016/j.jenvman.2009.02.009>
- [8] Caiado, J. "Performance of combined double seasonal univariate time series models for forecasting water demand". *Journal of Hydrologic Engineering*, 15(3), pp. 215–222. 2010. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000182](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000182)
- [9] Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J., Rietveld, L. "Improving the performance of water demand forecasting models by using weather input". *Procedia Engineering*, 70(1), pp. 93–102. 2014. <https://doi.org/10.1016/j.proeng.2014.02.012>
- [10] Bárdossy, G., Halász, G., Winter, J. "Prognosis of urban water consumption using hybrid fuzzy algorithms". *Journal of Water Supply: Research and Technology - AQUA*, 58(3), pp. 203–211. 2009. <http://aqua.iwaponline.com/content/58/3/203>
- [11] Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., Sliusarieva, A. "Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada". *Water Resources Research*, 48(1), W01528. 2012. <https://doi.org/10.1029/2010WR009945>
- [12] Firat, M., Ali, Yurdusev, M., Erkan Turan, M. "Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling". *Water Resources Management*, 23(4), pp. 617–632. 2009. <https://doi.org/10.1007/s11269-008-9291-3>
- [13] Babel, M. S., Shinde, V. R. "Identifying Prominent Explanatory Variables for Water Demand Prediction Using Artificial Neural Networks: A Case Study of Bangkok". *Water Resources Management*, 25(1), pp. 1653–1676. 2011. <https://doi.org/10.1007/s11269-010-9766-x>
- [14] Piasecki, A., Marszelewski, W. "Short-term dynamics of a press-fit to the municipal water supply system on the example of Toruń – a preliminary analysis". In: *Water in the city*. Vol. 2, (pp. 205–213). 2011.
- [15] Shahin, M. A., Maier, H. R., Jaksa, M. B. "Evolutionary data division methods for developing artificial neural network models in geotechnical engineering". Department of Civil and Environmental Engineering. The University of Adelaide. Research Report No. R 171. 2000.
- [16] Wei, Q., Dunbrack Jr, R. L. "The role of balanced training and testing data sets for binary classifiers in bioinformatics". *PLoS ONE*, 8(7), e67863. 2013. <https://doi.org/10.1371/journal.pone.0067863>
- [17] Kutner, M. H. *Applied linear statistical models*. Chicago: Irwin. 1996.
- [18] Hagan, M. T., Demuth, H. B., Beale, M. H., De Jesús, O. *Neural network design*. Boston: PWS publishing company. 1996.

- [19] Huang, G. B., Babri, H. A. "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions". *IEEE Transactions on Neural Networks*, 9(1), pp. 224–229. 1998. <https://doi.org/10.1109/72.655045>
- [20] Xu, S., Chen, L., "A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining". In: *5th International Conference on Information Technology and Applications*, pp. 683–686. Cairns, Queensland, Australia. 2008. <https://pdfs.semanticscholar.org/254f/3f0fce2905675445d48ca8ab61e3761b1e9b.pdf>
- [21] Kavzoglu, T. "Determining optimum structure for artificial neural networks". In: *25th Annual Technical Conference and Exhibition of the Remote Sensing Society*, pp. 675–682. Nottingham, UK, 1999. <https://pdfs.semanticscholar.org/b9f8/173c375ad07f3e43c3ca9cdd3419c8cec923.pdf>
- [22] Shibata, K., Ikeda, Y. "Effect of number of hidden neurons on learning in large-scale layered neural networks". In: *ICROS-SICE International Joint Conference*, pp. 5008–5013, 2009.
- [23] Kutylowska, M. "Prediction of failure frequency of water-pipe network in the selected city". *Periodica Polytechnica Civil Engineering*, 61(3), pp. 548–553. 2017. <https://doi.org/10.3311/PPci.9997>
- [24] Piasecki, A., Jurasz, J., Skowron, R. "Forecasting surface water level fluctuations of lake Serwy (Northeastern Poland) by artificial neural networks and multiple linear regression". *Journal of Environmental Engineering and Landscape Management*, 25(4), pp. 379–388. 2017. <https://doi.org/10.3846/16486897.2017.1303498>
- [25] Nash, J. E., Sutcliffe, J. V. "River flow forecasting through conceptual models part I – A discussion of principles". *Journal of Hydrology*, 10(3), pp. 282–290. 1970. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- [26] Wdowikowski, M., Kaźmierczak, B., Ledvinka, O. "Maximum daily rainfall analysis at selected meteorological stations in the upper Lusatian Neisse River basin". *Meteorology, Hydrology and Water Management*, 4(1), pp. 53–63. 2016. <https://doi.org/10.26491/mhwm/63361>
- [27] Zimoch, I., Łobos, E. "Evaluation of health risk caused by chloroform in drinking water". *Desalination and Water Treatment*, 57(3), pp. 1027–1033. 2016. <https://doi.org/10.1080/19443994.2015.1033134>
- [28] Tchórzewska-Cieślak, B. "Matrix method for estimating the risk of failure in the collective water supply system using fuzzy logic". *Environment Protection Engineering*, 37(3), pp. 111–118. 2011. http://epe.pwr.wroc.pl/2011/3_2011/12tchorzewska.pdf
- [29] Tchórzewska-Cieślak, B. "Bayesian model of urban water safety management". *Global NEST Journal*, 16(4), pp. 667–675. 2014. https://journal.gnest.org/sites/default/files/Submissions/gnest_01344/gnest_01344_published.pdf
- [30] Zimoch, I., Szymik-Gralska, J. "Improving the operation of a surface water intake with exceeded redundant capacity using reliability analysis and life cycle costing". *Environment Protection Engineering*, 42(2), pp. 41–53. 2016. <https://doi.org/10.5277/epel60203>
- [31] Kutylowska, M. "Neural network approach for availability indicator prediction". *Periodica Polytechnica Civil Engineering*, 61(4), pp. 873–881. 2017. <https://doi.org/10.3311/PPci.10429>