

## Forecasting failure rate of water pipes

M. Kutylowska

### ABSTRACT

This paper presents the results of failure rate prediction by means of support vector machines (SVM) – a non-parametric regression method. A hyperplane is used to divide the whole area in such a way that objects of different affiliation are separated from one another. The number of support vectors determines the complexity of the relations between dependent and independent variables. The calculations were performed using Statistical 12.0. Operational data for one selected zone of the water supply system for the period 2008–2014 were used for forecasting. The whole data set (in which data on distribution pipes were distinguished from those on house connections) for the years 2008–2014 was randomly divided into two subsets: a training subset – 75% (5 years) and a testing subset – 25% (2 years). Dependent variables ( $\lambda_r$  for the distribution pipes and  $\lambda_p$  for the house connections) were forecast using independent variables (the total length –  $L_r$  and  $L_p$  and number of failures –  $N_r$  and  $N_p$  of the distribution pipes and the house connections, respectively). Four kinds of kernel functions (linear, polynomial, sigmoidal and radial basis functions) were applied. The SVM model based on the linear kernel function was found to be optimal for predicting the failure rate of each kind of water conduit. This model's maximum relative error of predicting failure rates  $\lambda_r$  and  $\lambda_p$  during the testing stage amounted to about 4% and 14%, respectively. The average experimental failure rates in the whole analysed period amounted to 0.18, 0.44, 0.17 and 0.24 fail./km-year for the distribution pipes, the house connections and the distribution pipes made of respectively PVC and cast iron.

**Key words** | failure frequency, prediction, support vector machines, water pipelines

### M. Kutylowska

Wrocław University of Science and Technology,  
Faculty of Environmental Engineering,  
Wybrzeże Wyspińskiego 27, Wrocław 50-370,  
Poland  
E-mail: [malgorzata.kutylowska@pwr.edu.pl](mailto:malgorzata.kutylowska@pwr.edu.pl)

### INTRODUCTION

The condition of water-pipe networks should be a major concern not only for their operators, but also for scientists who are able to propose more accurate ways of condition deterioration modelling. The necessity to ensure the proper protection and management (Hamchaoui *et al.* 2015) of water supply systems is increasingly often highlighted. These are undoubtedly vital issues which together with reliability analyses, water demand analyses and the properly planned modernization (Tscheikner-Gratl *et al.* 2016) of the pipelines and the whole water supply infrastructure should be and currently are the subject of numerous

studies and projects. The research findings indicate that such studies need to be continued in order to gain deeper knowledge in this field, especially with regard to mathematical modelling, which owing to the development of computing techniques is constantly improved and uses increasingly more accurate modelling methods (Scheidegger *et al.* 2015).

Prior to modelling it is necessary to investigate the number and kinds of failures occurring in the water-pipe networks as well as the causes and effects (Iwanek *et al.* 2017; Pietrucha-Urbanik & Studziński 2017) of the failures and

the level of risk (Boryczko & Tchórzewska-Cieślak 2014). The analysis of the failure frequency of water pipes has been the subject of many investigations. For example, Hu & Hubble (2007) studied conduits made of asbestos cement. They demonstrated that the climate and the soil surrounding the pipe had a great influence on the failure rate. The deterioration of old water pipes was examined by Shahata & Zayed (2012). The authors concluded that relatively old conduits (dating back to the 19th century) were less deteriorated than ones from the second half of the 20th century. In contrast to this, Arai *et al.* (2010) found that in Japan at the beginning of the 21st century the water-pipe network built about 60 years ago needed to be renovated. The type of conduits (water mains, distribution pipes or house connections) and fluctuations in pressure inside the pipe have a great influence on the level of failure frequency (Pelletier *et al.* 2005; Piratla *et al.* 2015; Martínez-Codina *et al.* 2016).

Failure analyses used to be based on typical mathematical modelling. For example, Shamir & Howard (1979) proposed a model in which the failure rate exponentially depended on time (Shamir & Howard 1979). A few years later the model was expanded by Walski & Pelliccia (1982). Many statistical and physically based models concerning water-pipe deterioration were discussed in Kleiner & Rajani (2001) and Rajani & Kleiner (2001). There have been numerous studies relating to failure rate prediction and new investigations are still undertaken.

Nowadays typical statistical modelling is substituted by other kinds of mathematical modelling, e.g. Bayesian models (Tchórzewska-Cieślak 2014), which are successfully used in environmental engineering. For instance, sediment transport in sewers and the failure frequency of water pipes are estimated by artificial neural networks (ANNs) (Tabesh *et al.* 2009; Jafar *et al.* 2010; Nishiyama & Fillion 2014; Ebtehaj *et al.* 2016a; Kutylowska 2017). Genetic algorithms are used to model and optimize the failure frequency or the time between failures (Xu *et al.* 2011; Sattar *et al.* 2016). The risk level of a water distribution system has been assessed (as part of a failure analysis) by means of artificial intelligence and Monte Carlo simulation (Yung *et al.* 2011). Also environmental engineering problems have been investigated using mathematical modelling. For example, the degradation of organic compounds in the

environment has been assessed using the K-nearest neighbours method (Manganaro *et al.* 2016) and raw water quality has been modelled for chemical dosing process control in a water purification plant, by means of support vector machines (SVM) (Wang 2016).

The SVM method is used in many fields, e.g. to predict bus arrivals in municipal transport systems (Bin *et al.* 2006) and to forecast the cash demand in cash machines (Ramírez & Acuña 2011). Shirzad *et al.* (2014) proposed applying SVM and ANNs to predict the rate of failure of water pipes. The authors suggested that neural networks generated results more convergent with experimental data than the results yielded by support vector models. Using the SVM method one can also locate water leakages from pipes (Mashford *et al.* 2012; Candelieri *et al.* 2014). The prediction of sludge transport, which is essential for the proper operation of sewers, can be based on SVM modelling (Ebtehaj *et al.* 2016b). Also the condition of sewerage systems can be assessed (Mashford *et al.* 2011) and the inspection schedule can be planned (Harvey & McBean 2014) by means of SVM. The SVM can be a valuable tool for solving hydrological and hydrogeological problems relating to, e.g., flood wave height (Liu & Pender 2015), surface water quality (Kisi & Parmar 2016) and hydraulic conductivity (Elbisy 2015). However, there are very few studies concerning the deterioration and failure analysis of water conduits by means of SVM. Therefore this subject was undertaken by the present author.

The main aim of this paper was to verify whether the regression method called SVM could be used to forecast the failure rate ( $\lambda$ , fail./km-year) of the water pipelines (distribution pipes and house connections) in a selected zone of the water supply system in a Polish city.

## MATERIALS AND METHODS

The SVM method is a regression and classification algorithm which takes nonlinear decision space into account. A hyperplane divides the whole area in such a way that objects of different affiliation are separated from one another. It is also necessary to keep a maximum margin of error, i.e. the distance from the separating plane. The number of support vectors determines the complexity of

the relations between dependent and independent variables (Statistica Electronic Manual). In the case of a qualitative analysis of such a dependent variable as the failure rate of water conduits, no classification but regression is performed. Four kinds of SVMs, characterized by four types of kernel functions, linear (SVM-L), polynomial (SVM-P), sigmoidal (SVM-S) and radial basis functions (SVM-RBF), are distinguished (Statistica Electronic Manual). The notion of kernel functions derives from investigations of linear vector spaces. In the case of the problem considered here, all SVM models (based on all kernel functions) were built and verified. In the course of a regression analysis a relation between the dependent variable and the independent variables (predictors) is sought. This relation should possibly most accurately generate a dependent variable value for new cases (testing sample data), i.e. the ones which the SVM model has not 'seen' before, having been trained on a training sample. The mapping function  $\varphi(x)$  is called a kernel function which meets the Mercer condition while the feature map for the Mercer kernel is as follows (Guo *et al.* 2014):

$$k(x, y) = \varphi(x)^T \varphi(y) \quad (1)$$

The kernel functions are described by equations (2)–(5), respectively for linear, polynomial, sigmoidal and RB functions (Guo *et al.* 2014):

$$k(x, y) = (x \cdot y) \quad (2)$$

$$k(x, y) = (s(x \cdot y) + \gamma)^d \quad (3)$$

$$k(x, y) = \tanh(s(x \cdot y) + \gamma) \quad (4)$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (5)$$

where:

- $\gamma$  – learning rate,
- $x$  – independent variable,
- $y$  – dependent variable,

- $d$  – degree of polynomial,
- $\sigma$  – dispersion parameter,
- $s$  – indicator of kernel function (parameter similar to dispersion in radial function).

The prediction function is calculated from the relation (Aydogdu & Firat 2015):

$$y(X) = \mathbf{w}^T \varphi(x) + b \quad (6)$$

where:

- $y$  – dependent variable,
- $b$  – bias,
- $\mathbf{w}$  – matrix of weights,
- $\varphi$  – mapping function.

There are many advantages of SVM modelling, e.g. the size of the learning vector can be relatively small, outliers do not have a significant influence on the modelling quality (Williams 2011), the modelling is possible even if the relationships between dependent and independent variables are complicated and the application of typical mathematical models is difficult and limited (Bin *et al.* 2006), and in fact the solution cannot reach the so-called local minimum (Cristianini & Shave-Taylor 2014). Several regression methods are used for prediction purposes, e.g. ANNs, which belong to 'black box' models. SVM models seem to be a little bit easier in application in comparison with ANNs. Neural networks require a relatively large data set for training, validating and testing the model. Moreover, artificial networks are not so resistant to outliers as the SVM method. ANNs are trained using training methods, learning epochs and neurons activated by functions. The number and kind of model parameters depend on the problem being solved and often can be determined by trial and error. A proper activation function and training method as well as a number of hidden layers and hidden neurons need to be selected. Generally, a hidden layer behaves like a 'black box' and it is impossible to identify the procedures and the relationships occurring in it, which is the main disadvantage of neural networks. Because of fewer limitations the SVM method seems to be easier in application and in modelling.

The calculations were performed using Statistica 12.0. Operating data for a selected zone of the water supply

system for the period 2008–2014 were used for forecasting. The whole set of data (for respectively distribution pipes and house connections) for the years 2008–2014 was divided randomly into two subsets: a training subset – 75% (5 years) and a testing subset – 25% (2 years). Dependent variables ( $\lambda_r$  for the distribution pipes and  $\lambda_p$  for the house connections) were forecast using independent variables (the total length –  $L_r$  and  $L_p$  and number of failures –  $N_r$  and  $N_p$  of the distribution pipes and the house connections, respectively). The independent variables, the total length and the number of failures (basic information about the water pipes), were adopted to check, using a relatively simple case, if the SVM algorithm was suitable for failure frequency forecasting. This paper continues the subject of the author's previous investigations in which SVM modelling was applied to another water distribution system (Kutylowska 2016). In that case (Kutylowska 2016), the diameter, the year of installation and the material were used as independent variables. In the present work more basic parameters (the length and the number of failures) are used to explore the possibilities of SVM modelling based on less complicated information about the water pipeline.

The whole city, with c. 230,000 inhabitants, was divided into 55 water supply zones. The failure frequency of the distribution pipes and the house connections was investigated in only one selected zone in which the pressure inside the pipe-network amounted to about 0.4 MPa. The total length of the distribution pipes and the house connections amounted to 17.5 km and 14.2 km, respectively. The distribution pipes were made mainly of grey cast iron (48.6%, 8.5 km) and PVC (38.9%, 6.8 km), and the remaining 12.5% of the total length was made of PE. The analysed zone has the area of c. 41 km<sup>2</sup> and about 10,000 citizens who are all connected to the water-pipe network. The water is supplied, in the amount of 1,920 m<sup>3</sup>/d, from a well. The water-pipe network architecture is shown in Figure 1.

## RESULTS AND DISCUSSION

The values of the dependent and independent experimental variables for the years 2008–2014 are shown in Table 1. The

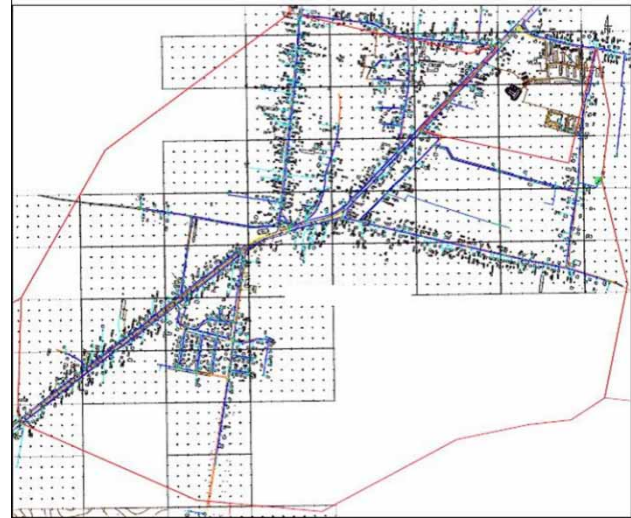


Figure 1 | Structure of the water-pipe network in the analysed zone.

data are for one zone selected from the whole water supply system. The detailed temporal and spatial clustering of pipe failures within this zone will be the subject of future investigations. The values of failure frequency  $\lambda_r$  and  $\lambda_p$  were calculated for the whole length of each kind of water pipeline in the analysed zone. Moreover, Table 1 shows the length ( $L_{r\text{ PVC}}$ ,  $L_{r\text{ CI}}$ ), the number of failures ( $N_{r\text{ PVC}}$ ,  $N_{r\text{ CI}}$ ) and the failure rate of the distribution pipes ( $\lambda_{r\text{ PVC}}$  and  $\lambda_{r\text{ CI}}$ ) for two kinds of material (PVC and cast iron – CI). The total number of failures in the selected area over the whole analysed time was equal to 22, 44, 8 and 14 for the distribution pipes, the house connections and the distribution pipes made of respectively PVC and cast iron. The average experimental failure rates in the whole analysed period amounted to 0.18, 0.44, 0.17 and 0.24 fail./ (km-year) for the distribution pipes, the house connections and the distribution pipes made of respectively PVC and cast iron. All types of kernel functions (L, P, S and RBF) were applied. As mentioned above, the whole data set (2008–2014) was randomly divided into a training sample (5 years: 2008–2010, 2012 and 2014) and a testing sample (2 years: 2011 and 2013).

The main model parameters are shown in Table 2. The polynomial degree was equal to 3 in all the SVM-P models. Since the SVM method is a kind of nonparametric regression, the correlations between the dependent variables (the predicted values) and the independent variable need

**Table 1** | Dependent and independent variables

$L_r$ , km	$L_{r\text{ PVC}}$ , km	$L_{r\text{ CI}}$ , km	$L_p$ , km	$N_r$	$N_{r\text{ PVC}}$	$N_{r\text{ CI}}$	$N_p$	$\lambda_r$ , fail./(km-year)	$\lambda_{r\text{ PVC}}$ , fail./(km-year)	$\lambda_{r\text{ CI}}$ , fail./(km-year)	$\lambda_p$ , fail./(km-year)
17.5	6.8	8.5	14.2	2–5	0–3	1–3	3–10	0.11–0.29	0.00–0.44	0.12–0.35	0.21–0.70

not be known.  $V$ -fold cross-validation was used to find the optimal model parameters. In this type of cross-validation, data are divided into  $V$  randomly selected disjoint parts. Using the  $V-1$  parts of the data as training examples the dependent variable is predicted and the prediction error is calculated on the basis of the residual sum of squares. The procedure is executed for all the  $V$  data segments. Then a model quality measure is determined on the basis of the averaged errors of the particular cycles. The optimal model parameters are selected during a quality analysis.

The parameters determined in the course of the  $V$ -fold cross-validation are: gamma, capacity, epsilon and the number of SVM (including localized vectors) ([Statistica Electronic Manual](#)). Tenfold ( $V=10$ ) cross-validation was applied to the considered problem, whereby it was possible to select proper values for such parameters (learning constants) as capacity ( $C$ ) and epsilon ( $\epsilon$ ), since they are not *a priori* known.

The data presented in [Table 2](#) should be analysed together with the prediction results shown in [Table 3](#) and

**Table 2** | Model parameters

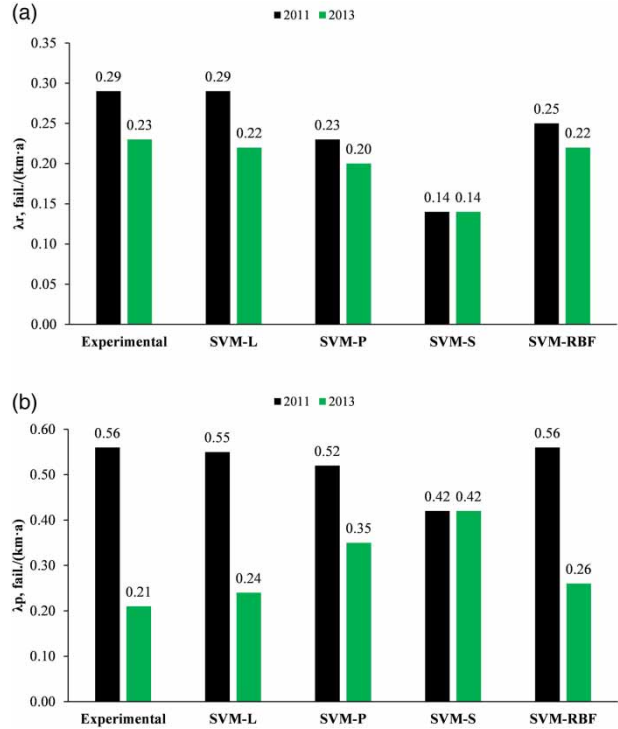
Type of conduit/parameter	Distribution pipes	Distribution pipes – PVC	Distribution pipes – CI	House connections
SVM-L				
Gamma	–	–	–	–
Capacity ( $C$ )	4	2	2	4
Epsilon ( $\epsilon$ )	0.1	0.1	0.1	0.1
Number of support vectors (localized)	2 (0)	2 (0)	2 (0)	2 (0)
Cross-validation error	0.024	0.010	0.008	0.023
SVM-P				
Gamma	0.5	0.5	0.5	0.5
Capacity ( $C$ )	1	1	1	1
Epsilon ( $\epsilon$ )	0.5	0.5	0.5	0.5
Number of support vectors (localized)	2 (0)	2 (0)	2 (0)	2 (0)
Cross-validation error	0.024	0.010	0.008	0.023
SVM-S				
Gamma	0.5	0.5	0.5	0.5
Capacity ( $C$ )	1	1	1	1
Epsilon ( $\epsilon$ )	0.5	0.5	0.1	0.5
Number of support vectors (localized)	2 (2)	4 (4)	4 (4)	2 (2)
Cross-validation error	0.650	1.500	1.800	0.689
SVM-RBF				
Gamma	0.5	0.5	0.5	0.5
Capacity ( $C$ )	4	3	3	4
Epsilon ( $\epsilon$ )	0.1	0.1	0.1	0.1
Number of support vectors (localized)	2 (0)	2 (0)	2 (0)	2 (0)
Cross-validation error	0.069	0.010	0.008	0.072

**Table 3** | Experimental and predicted failure rates

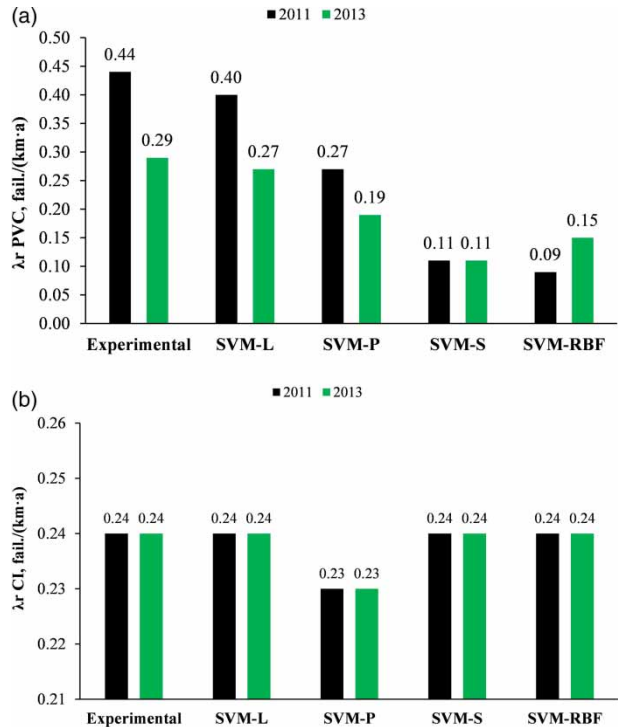
Experimental	SVM-L	SVM-P	SVM-S	SVM-RBF
$\lambda_r$ , fail./ (km-year)				
0.11	0.12	0.14	0.14	0.12
0.11	0.12	0.14	0.14	0.12
0.17	0.17	0.17	0.14	0.17
0.23	0.22	0.20	0.14	0.22
0.11	0.12	0.14	0.14	0.12
$\lambda_p$ , fail./ (km-year)				
0.28	0.30	0.39	0.42	0.30
0.70	0.68	0.60	0.42	0.68
0.49	0.49	0.49	0.42	0.49
0.35	0.36	0.42	0.42	0.36
0.49	0.49	0.49	0.42	0.49
$\lambda_{r\text{ PVC}}$ , fail./ (km-year)				
0.15	0.14	0.11	0.11	0.14
0.00	0.01	0.04	0.11	0.01
0.00	0.01	0.04	0.11	0.01
0.15	0.14	0.11	0.11	0.14
0.15	0.14	0.11	0.11	0.14
$\lambda_{r\text{ CI}}$ , fail./ (km-year)				
0.12	0.13	0.18	0.24	0.13
0.24	0.24	0.23	0.24	0.24
0.35	0.34	0.29	0.24	0.34
0.35	0.34	0.29	0.24	0.34
0.12	0.13	0.18	0.24	0.13

in Figures 2 and 3. The prediction results in Table 3 are for the training sample and they are compared with the experimental results.

An analysis of the failure rate ( $\lambda_r$ ) and ( $\lambda_p$ ) prediction results (Table 3) shows that the SVM models based on the linear kernel function and the radial basis kernel function are the optimal ones for each case (the distribution pipes, the house connections and the distribution pipes made of two different materials). The relative errors of the experimental values and the forecast values ranged from 0.00% to 9.09%. The SVM-S models (the sigmoidal function) yielded senseless results since the predicted failure rate was constant in all the cases. The models based on the polynomial kernel function forecast the failure rate with a higher error than models SVM-L and SVM-RBF. The results of



**Figure 2** | Experimental and predicted failure rates of (a) distribution pipes, (b) house connections.



**Figure 3** | Experimental and predicted failure rates of distribution pipes: (a) PVC, (b) CI.

prediction based on the testing sample (2011 – the black bar and 2013 – the green bar) are shown in Figures 2 and 3 for the distribution pipes, the house connections and the distribution pipes made of respectively PVC and CI.

The models based on the linear kernel function are the most suitable for forecasting failure rates  $\lambda_r$  and  $\lambda_p$  (Figure 2) during the testing stage.

The results of forecasting  $\lambda_r$  for the two different materials (Figure 3) are surprising. In comparison with the experimental values, the models based on all the kernel functions ideally predict the failure rate of the pipelines made of cast iron (Figure 3(b)). The failure frequency of the conduits made of PVC (Figure 3(a)) is predicted quite well by means of the linear kernel function. The other functions yielded senseless results. The quality and applicability of a model should be evaluated on the basis of the forecasting results obtained from testing since they are more representative (the model has no prior knowledge of the dependent variables and the predictors) than the solutions obtained from the learning phase. Considering the above, the SVM model based on the linear kernel function is the optimal one for predicting the failure rate of each kind of water conduit. In this model (the testing stage) the maximum relative error amounted to about 4% and 14% for predicting respectively  $\lambda_r$  and  $\lambda_p$ . For the RBF model the maximum errors were higher, amounting to about 13% and 23%, respectively. In the case of the SVM-P models (Table 2), the cross-validation error was the same as for the SVM-L models, but it did not influence the prediction quality. In the case of any kind of modelling, one should answer the question whether the aim is to obtain a perfect data fit at any cost, i.e. at the expense of model architecture complication. Even if the capacity ( $C = 1$ ) is lower in SVM-P and SVM-S models and the epsilon values are higher ( $\epsilon = 0.5$ ) than in linear models, one should choose the model which is characterized by the highest convergence between the real (experimental) and forecast failure rate values. One should bear in mind that water-pipe networks belong to the critical buried infrastructure and so the condition of the water pipelines should be estimated accurately. Model structure is important, but first of all one should consider the model which estimates the dependent variable in the most optimal way with the lowest error between the real and forecast values. The

number of localized vectors, whose weights are equal to  $\pm$  the capacity value, is also a crucial issue. The more localized vectors there are, the more difficult it is to divide the whole area by means of the hyperplane. This means that the problem becomes then more complicated. For example, when the sigmoidal kernel function (Table 2) was used, the model had more localized support vectors than the other models. In fact, all the support vectors were localized. The model architecture (the values of  $C$  and  $\epsilon$  and the number of support vectors) will change if more and other independent variables are included.

## CONCLUSIONS

The SVM method is useful for forecasting the failure rate of water conduits. The methodology is applicable to any water supply system, but the results presented in this paper are valid for only the particular water-pipe network and the particular pressure zone. Another model based on SVM needs to be built to predict the failure rate in another city. For the purposes of failure rate modelling, the length of the conduits and the number of registered failures (separately for the distribution pipes and the house connections) were treated as independent variables (predictors). The whole data set (time span: 2008–2014) was randomly divided into two subsets (for model training and testing). An analysis of the testing results indicated that the models based on the linear kernel function were the most optimal and suitable for failure rate prediction for all the types of pipelines and the two kinds of material. In the case of the optimal SVM-L model, the correlation between the experimental and predicted failure rates of the distribution pipes and the house connections was almost perfect for the testing sample. The same was observed for the distribution pipes made of respectively PVC and cast iron. The SVM-L model was characterized by the following parameters:

- capacity equal to 4 for the distribution pipes and the house connections and to 2 for the distribution pipes made of respectively PVC and CI;
- epsilon equal to 0.1;
- two support vectors and none of them localized;
- the cross-validation error ranging from 0.008 to 0.024;

- maximum relative errors (the testing sample) equal to 4.4% and 14.3% for respectively the distribution pipes and the house connections and to 9.1% and 0.0% for the distribution pipes made of respectively PVC and CI. From the engineering point of view such errors are acceptable.

As mentioned earlier, this paper continues the subject of the author's previous study concerned with the SVM modelling of a water distribution system (Kutylowska 2016). One should note that simple comparisons should be avoided because the previous water-pipe network was completely different and the operating conditions were not the same. In the earlier work more detailed information about the water pipes was taken into consideration as predictors, e.g. the year of construction and the diameter of the pipes. This approach strongly affected the model structure. In the earlier paper (Kutylowska 2016) the number of support vectors was larger, amounting to 56 and 14 for the distribution pipes and the house connections, respectively. This means that the model architecture was more complicated, but the prediction quality was not affected since the cross-validation error (for the optimal model based on the linear kernel function) was higher, amounting to 0.094 and 0.112 for the distribution pipes and the house connections, than in the case of the model proposed in the present paper. Generally, the optimal model should be relatively simple and forecast the dependent variable in good agreement with the experimental values. If detailed information about the pipelines is not available or some data are missing or are considered to be outliers, one can still use SVM modelling based on simple operational data as described above. For this reason the total length of the conduits and the registered number of their failures (basic information about the water conduits) were treated as independent variables. It is highly important to create a relatively simple model using available operational data. The methodology, the solutions and the prediction results meet the above requirement for model simplicity. Nevertheless, one should remember that each water supply system is different and it is necessary to check all the modelling possibilities with simple and more detailed predictors in order to select the optimal solution.

The proposed methodology can be useful for water utilities and their managers. The models can be used to forecast the failure frequency solely on the basis of two variables.

This approach does not require collecting a lot of operational data, which are sometimes very difficult to acquire, especially when the prognosis is to be made on the basis of sparse historic information not collected in the Geographic Information System. The advantage of SVM modelling is that it is possible to extend once-created models using other operational data (the pressure inside the pipe, the depth of laying, the diameter, the material, etc.) if this is required by the operators to understand the processes responsible for the failures of the water pipes. Moreover, building two separate models (for distribution pipes and house connections) is a good solution since the operation of the two types of conduits is completely different. Then the water utility can use the proposed methodology independently for larger and smaller pipes. One should note that damage to one distribution pipe has a more disruptive effect (e.g. a pressure drop or no supply of water for some hours) on the operation of the whole water supply system than even several failures of house connections. The next step can be an analysis of failures over time and their spatial clustering which will provide the water utility with detailed information about the failures and help it to draw up a modernization or replacement schedule.

---

## ACKNOWLEDGEMENTS

This work was carried out thanks to allocation No. 0401/0069/16 awarded to the Faculty of Environmental Engineering at Wroclaw University of Science and Technology by the Ministry of Science and Higher Education in 2017.

---

## REFERENCES

- Arai, Y., Koizumi, A., Inakazu, T., Watanabe, H. & Fujiwara, M. 2010 [Study on failure rate analysis for water distribution pipelines](#). *Journal of Water Supply: Research and Technology - AQUA* **59** (6-7), 429-435.
- Aydogdu, M. & Firat, M. 2015 [Estimation of failure rate in water distribution network using fuzzy clustering and LS-SVM method](#). *Water Resources Management* **29** (5), 1575-1590.
- Bin, Y., Zhongzhen, Y. & Baozhen, Y. 2006 [Bus arrival time prediction using support vector machines](#). *Journal of Intelligent Transportation Systems* **10** (4), 151-158.



- Boryczko, K. & Tchórzewska-Cieślak, B. 2014 Analysis of risk of failure in water main pipe network and of delivering poor quality water. *Environment Protection Engineering* **40** (4), 77–92.
- Candelieri, A., Soldi, D., Conti, D. & Archetti, F. 2014 Analytical leakages localization in water distribution networks through spectral clustering and support vector MACHINES. The ICeWater approach. *Procedia Engineering* **89**, 1080–1088.
- Cristianini, N. & Shave-Taylor, J. 2014 *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge, UK.
- Ebtehaj, I., Bonakdari, H. & Zaji, A. H. 2016a An expert system with radial basis function neural network based on decision trees for predicting sediment transport in sewers. *Water Science and Technology* **74** (1), 176–183.
- Ebtehaj, I., Bonakdari, H., Shamshirband, S. & Mohammadi, K. 2016b A combined support vector machine-wavelet transform model for prediction of sediment transport in sewer. *Flow Measurement and Instrumentation* **47**, 19–27.
- Elbisy, M. S. 2015 Support vector machine and regression analysis to predict the field hydraulic conductivity of sandy soil. *KSCE Journal of Civil Engineering* **19** (7), 2307–2316.
- Guo, Y. M., Wang, X. T., Liu, C., Zheng, Y. F. & Cai, X. B. 2014 Electronic system fault diagnosis with optimized multi-kernel SVM by improved CPSO. *Maintenance and Reliability – Eksploatacja I Niezawodność* **16** (1), 85–91.
- Hamchaoui, S., Boudoukha, A. & Benzerra, A. 2015 Drinking water supply service management and sustainable development challenges: case study of Bejaia, Algeria. *Journal of Water Supply: Research and Technology – AQUA* **64** (8), 937–946.
- Harvey, R. R. & McBean, E. A. 2014 Comparing the utility of decision trees and support vector machines when planning inspections of linear sewer infrastructure. *Journal of Hydroinformatics* **16** (6), 1265–1279.
- Hu, Y. & Hubble, D. W. 2007 Factors contributing to the failure of asbestos cement water mains. *Canadian Journal of Civil Engineering* **34** (5), 608–621.
- Iwanek, M., Suchorab, P. & Karpińska-Kielbasa, M. 2017 Suffosion holes as the result of a breakage of a buried water pipe. *Periodica Polytechnica Civil Engineering* **61** (4), 700–705.
- Jafar, R., Shahrour, I. & Juran, I. 2010 Application of artificial neural networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling* **51** (9–10), 1170–1180.
- Kisi, O. & Parmar, K. S. 2016 Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology* **534**, 104–112.
- Kleiner, Y. & Rajani, B. 2001 Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* **3** (3), 131–150.
- Kutylowska, M. 2016 Prediction of water conduits failure rate – comparison of support vector machine and neural network. *Ecological Chemistry and Engineering A* **23** (2), 147–160.
- Kutylowska, M. 2017 Comparison of two types of artificial neural networks for predicting failure frequency of water conduits. *Periodica Polytechnica Civil Engineering* **61** (1), 1–6.
- Liu, Y. & Pender, G. 2015 A flood inundation modelling using v-support vector machine regression model. *Engineering Applications of Artificial Intelligence* **46**, 223–231.
- Manganaro, A., Pizzo, F., Lombardo, A., Pogliaghi, A. & Benfenati, E. 2016 Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. *Chemosphere* **144**, 1624–1630.
- Martínez-Codina, A., Castillo, M., González-Zeas, D. & Garrote, L. 2016 Pressure as a predictor of occurrence of pipe breaks in water distribution networks. *Urban Water Journal* **13** (7), 676–686.
- Mashford, J., Marlow, D., Tran, D. & May, R. 2011 Prediction of sewer condition grade using support vector machines. *Journal of Water Resources Planning and Management* **25** (4), 283–290.
- Mashford, J., De Silva, D., Burn, S. & Marney, D. 2012 Leak detection in simulated water pipe networks using SVM. *Applied Artificial Intelligence* **26** (5), 429–444.
- Nishiyama, M. & Filion, Y. 2014 Forecasting breaks in cast iron water mains in the city of Kingston with an artificial neural network model. *Canadian Journal of Civil Engineering* **41** (10), 918–923.
- Pelletier, G., Mailhot, A. & Villeneuve, J.-P. 2003 Modeling water pipe breaks – three case studies. *Journal of Water Resources Planning and Management* **129** (2), 115–123.
- Pietrucha-Urbanik, K. & Studziński, A. 2017 Case study of failure simulation of pipelines conducted in chosen water supply system. *Eksploatacja I Niezawodność – Maintenance and Reliability* **19** (3), 317–323.
- Piratla, K. R., Yerri, S. R., Yazdekhasi, S., Cho, J., Koo, D. & Matthews, J. C. 2015 Empirical analysis of water-main failure consequences. *Procedia Engineering* **118**, 727–734.
- Rajani, B. & Kleiner, Y. 2001 Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water* **3** (3), 151–164.
- Ramírez, C. & Acuña, G. 2011 Forecasting cash demand in ATM using neural networks and least square support vector machine, In: *CIARP 2011* (C. San Martin & S.-W. Kim, eds), Springer-Verlag, Berlin and Heidelberg, Germany, pp. 515–522.
- Sattar, A. M. A., Gharabaghi, B. & McBean, E. A. 2016 Prediction of timing of water main failure using gene expression models. *Water Resources Management* **30** (5), 1635–1651.
- Scheidegger, A., Leitão, J. P. & Scholten, L. 2015 Statistical failure models for water distribution pipes – a review from a unified perspective. *Water Research* **83**, 237–247.
- Shahata, K. & Zayed, T. 2012 Data acquisition and analysis for water main rehabilitation techniques. *Structure and Infrastructure Engineering* **8** (11), 1054–1066.

- Shamir, U. & Howard, C. D. D. 1979 [An analytical approach to scheduling pipe replacement](#). *Journal of AWWA* **71** (5), 248–258.
- Shirzad, A., Tabesh, M. & Farmani, R. 2014 [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), 941–948.
- Statistica 12.0, Electronic Manual.
- Tabesh, M., Soltani, J., Farmani, R. & Savic, D. 2009 [Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling](#). *Journal of Hydroinformatics* **11** (1), 1–17.
- Tchórzewska-Cieślak, B. 2014 [Bayesian model of urban water safety management](#). *Global NEST Journal* **16** (4), 667–675.
- Tscheikner-Gratl, F., Sitzenfrei, R., Rauch, W. & Kleidorfer, M. 2016 [Enhancement of limited water supply network data for deterioration modelling and determination of rehabilitation rate](#). *Structure and Infrastructure Engineering* **12** (3), 366–380.
- Walski, T. M. & Pelliccia, A. 1982 [Economic analysis of water main breaks](#). *Journal of AWWA* **74** (3), 140–147.
- Wang, D. 2016 [Research on raw water quality assessment oriented to drinking water treatment based on the SVM model](#). *Water Science and Technology: Water Supply* **16** (3), 746–755.
- Williams, G. 2011 *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. Springer Science + Business Media, New York, USA.
- Xu, Q., Chen, Q. & Li, W. 2011 [Application of genetic programming to modeling pipe failures in water distribution systems](#). *Journal of Hydroinformatics* **13** (3), 419–428.
- Yung, B. B., Tolson, B. A. & Burn, D. H. 2011 [Risk assessment of a water supply system under climate variability: a stochastic approach](#). *Canadian Journal of Civil Engineering* **38** (3), 252–262.

First received 3 July 2017; accepted in revised form 4 April 2018. Available online 13 April 2018