Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-21 (2015) 1216–1223

# A Method of Leakage Location in Water Distribution Networks using Artificial Neuro-Fuzzy System

**D. Wachla, P. Przystałka, W. Moczulski**

*Institute of Fundamentals of Machinery Design, Silesian University of Technology, Gliwice, Konarskiego 18a Street, Poland, e-mails: {dominik.wachla, piotr.przystalka, wojciech.moczulski}@polsl.pl.*

**Abstract:** The paper deals with a method to locate uncontrolled leaks in water distribution networks. The idea of the method is based on discretization of the water supply system to the predefined areas and then identifying an approximate location where a leakage can occur. In the proposed method, the location of leakage is determined by means of the group of neuro-fuzzy classifiers. The number of classifiers corresponds to the number of areas in which the network is divided. The task of each classifier is to change the state of its output in the event of a leak in the network associated with this classifier. The input signals of classifiers are residues. They are computed using flow measurements and output signals of predictive models that describe the observed changes in the nominal flow conditions of the water supply network. In this paper the verification case study is conducted for the water distribution network covering industrial and individual consumers in the selected district of the town in the southern Poland.

© 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

*Keywords:* Water distribution networks, leakage detection and localization, neuro-fuzzy systems, support vector machine.

## 1. INTRODUCTION

The water supply system is one of the most important components of the technical infrastructure in towns and villages. This is a system of engineered hydrologic and hydraulic components which is subject to faults such as leaks. Leak detection and localization is not a trivial problem, because the outflow of water does not always manifest on the surface of the ground. Moreover, the problem increases when one takes into account such factors as the complexity of the network, repeatedly applied loads due to the change of behaviour of its users and the expansion and continuous modifications of the network. Undetected leaks generate certain economic losses and increase the operating costs of a water supply network. For this reason, there is the need to develop monitoring and diagnostic systems of water supply networks that can be used for detection and localization of small leakages as soon as possible.

The general survey on methods for detection and location of leaks in water supply networks was presented in the paper (Puust et al. (2010)). There were distinguished the following classes of leak location approaches: methods based on acoustic measurements, based on the data acquired with the ground penetrating radar, and based on measurements of pressures and/or flow rates of water. The idea of the leakage localization methods in the first of these groups is based on measurements carry out by means of vibration or hydrophone sensors that are temporarily or permanently mounted in selected junctions of the water supply network. The sensors are arranged in such a way that the distance between adjacent sensors is in the range from 200 m to 500 m. Then, the gates (installed on the network) are closed taking into account the schedule of the inspection in order to isolate the sub-areas of the test pipe network with one or two nodes for which acoustic signals can be gathered. The measurements are carried out at night between 02:00 and 4:00 am. Finally, the identification of the leak is usually achieved by the analysis of the collected data. The main advantage of this solution is the exact location of the leak. On the other hand, the fundamental limitations include time-consuming measurements and analysis of the results, the need to change the networks configuration, the occurrence of interferences from the environment, which might significantly affect the correct diagnosis.

The current research works (Poulakis et al. (2003); Gertler et al. (2010); Puust et al. (2010)) emphasizes leak location methods using measurements of pressures and/or flows of water registered by transducers installed permanently in specific sections of the water supply system. The measurement data is transmitted from the measuring devices (using a telemetry system) to SCADA system, where specialized modules make automatic analysis of the data. The analysis of the measurement data is the basis for generation the final diagnosis. These data analysis modules are developed using methods of leakage localization. The vast majority of these approaches directly (Poulakis et al. (2003)) or indirectly (Gertler et al. (2010)) applies the analytical model of the supervised water supply system. The authors of the paper (Poulakis et al. (2003)) proposed the location method based on optimization of the analytical model of a water supply system taking into account the uncertainty caused by measuring and modeling errors. In

2405-8963 © 2015, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2015.09.692

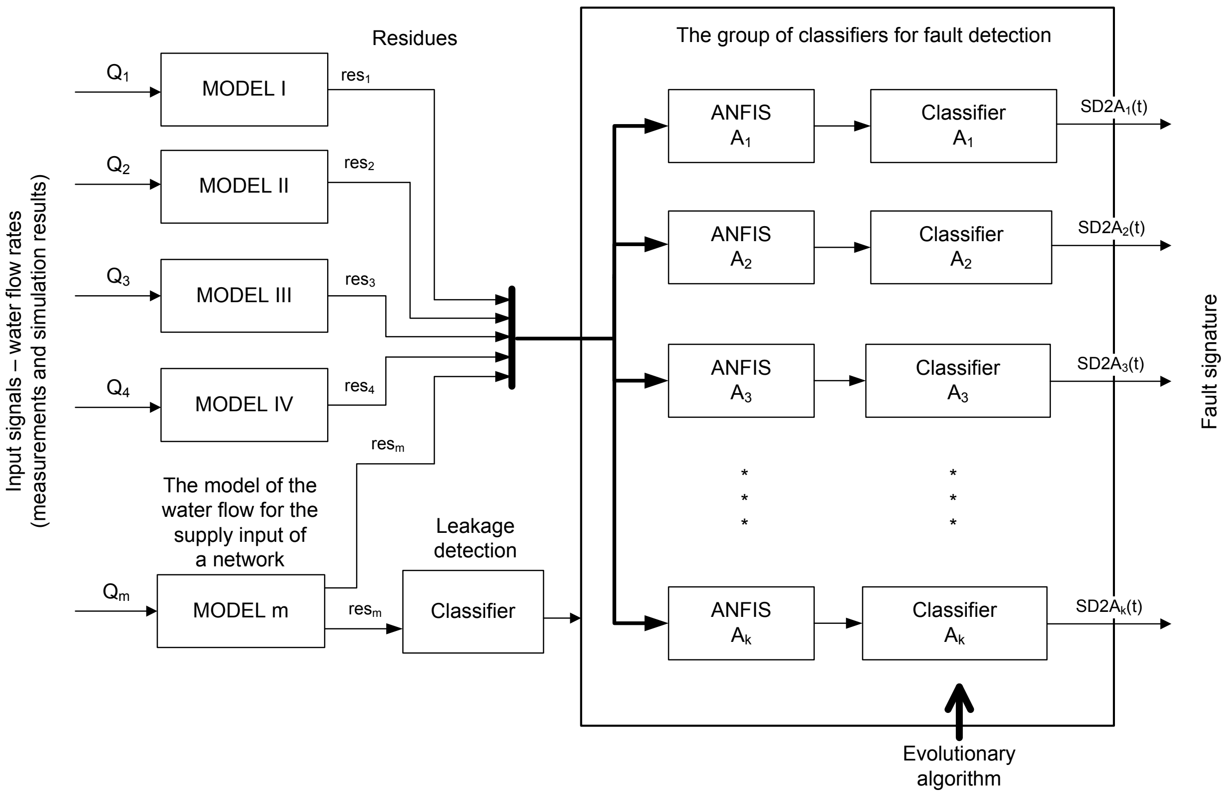


Fig. 1. A data flow diagram of the proposed method

this case, the location of the leak is determined by the designated set of model parameters. In the paper (Gertler et al. (2010)), the usage of PCA analysis to generate structural residues is suggested. PCA model was built based on the nominal simulation data taking into account the technical condition of water supply systems and the data representing various configurations of leaks. The main advantage of these methods is the short time of fault diagnosis process. Among the disadvantages are the need to update the analytical model and the models derived in case of changes in the structure of the monitored water supply network.

The paper describes a new study corresponds to the previous work conducted by a research team from the Institute of Fundamentals of Machine Design at Silesian University of Technology and domain experts from the Water and Sewage Company PWiK Rybnik. The details and results of this work can be found e.g. in (Wyczółkowski and Wysogląd (2007); Wyczółkowski (2008)). The related research dealt with methods of on-line and automated leak detection and location which were based on static neural network classifiers (a multilayer perceptron network was applied) created with the use of historical measurement data such as flow rates and pressures of water. The obtained result in the previous papers showed limitations of the method. In particular, the dynamic behaviour of the network and consumers was not mapped by the static neural classifiers and it was the reason of the inaccuracy of leak detection and localization classifiers. The other drawback of the previous method was that such kind of classifiers cannot be created using the knowledge of domain experts.

For this reason our motivation was to continue ongoing research. Hence, the new method is proposed to increase the accuracy and precision of leakage location in the water supply network. The novelty in this paper depends on that leak location is performed by means of support vector machine predictive models describing the dynamics of the observed variables and neuro-fuzzy state classifiers in which their inputs are supported by residual signals representing the difference between the predictive model output and the corresponding response signal measured directly on the network.

## 2. A METHOD FOR LEAKAGE LOCATION

The data flow diagram of the proposed method for locating small leaks in water supply networks is shown in Fig. 1. The presented concept is based on the drop-down approach in the field of process diagnostic, where the inference on the technical condition of the process is conducted analyzing residual signals (Kościelny (2001)). The residual signal is the difference between the observed process variable and the output of the model describing the dynamics of the object for the nominal operating conditions. This approach makes possible to obtain information about any changes occurring in the system, which differ from nominal conditions. The carrier of this information is the residual signal. Taking into consideration these reasons the authors developed the method based on:

- measurements of water flows in the specific nodes of the water supply system,
- predictive models describing the dynamics of the flow of water for nominal conditions of the network,

- a set of classifiers making the detection of leaks in areas of the water supply network.

The nodes where the water flows are measured were selected using the optimization procedure proposed in the paper (Przystałka and Moczulski (2012)). The optimization process applies the analytical model of the water supply network and takes into account the limitations such as the costs associated with the purchase, installation and operation of the measuring devices.

At the stage of formulating guidelines for the proposed method, it was assumed that each model of the water flow to be created for one process variable, that means the output and input signal of the model was the same process variable. This assumption was adopted due to the simplicity of the method. Despite these simplifications, in the general case, the method allows the use of any models, including models in which the input signals are (besides the modeled process variable) other process variables such as the water flow in the other nodes of a water supply system than the node associated with the process variable under consideration, as well as other meteorological factors. In addition, it is assumed that linear models will be considered. In case of using the linear models it can be expected that the change in residual signal value is proportional to the size of the leak. This approach allows to estimate the size of leakage by measuring the flow at the input of the water supply network.

The residues determined on the basis of measurements of water flows and adequate models are routed to the binary classifiers, each of which is tuned to recognize a leak in one of the areas to which it has previously been associated. Furthermore, a residual signal which is generated by the given flow model for the measuring point located at the supply to the network is the input of the classifier whose task is to detect the leak at the certain level of emission and, subsequently, run the set of basic classifiers for fault detection and location.

The basic structure of the classifiers for leak localization is composed of two parts: adaptive network-based fuzzy inference system (ANFIS) in the form proposed in (MATLAB (2009)) and a simple binary classifier, whose operation involves comparing the output of ANFIS with a fixed threshold. The task of neuro-fuzzy system is to evaluate the residuum in such a way that a binary signal is generated in order to represent two states: the occurrence of leakage in the area of an associated classifier (value 1) and faultless state in this area (value 0). Due to the properties of the neuro-fuzzy system, the output signal is not bivalent, but a signal whose values vary continuously. Therefore, in the structure of the proposed classifiers for leak location the additional binary classifier is applied in order to have the output with an adequate decision value.

The efficiency of the proposed classifier depends on the quality of a neuro-fuzzy system as well as on the threshold of the decision-making binary classifier. As it was mentioned above, in the presented method a neuro-fuzzy network is specified in the form of ANFIS model (MATLAB (2009)). The choice of a structure is implemented with fuzzy subtractive clustering (Chiu (1994)), while the choice of behavioural parameters (the number of rules and the type, form and number of membership function, coefficients of polynomials in conclusions of rules) with the error back propagation algorithm (Jang and Sun (1997)) and the method of gradient descent (Jang (1993)) . In turn, the decision threshold value of a binary classifier is selected as a result of the optimization process. The optimization problem in this case is formulated as a minimizing task of the classification error. In order to determine the optimum value of a decision threshold, the evolutionary algorithm is proposed (Michalewicz (1996)), and in the general case it is possible to apply any searching method which requires the definition of the criterion function.

## 3. VERIFICATION RESEARCH

The proposed method consists of two main steps. The first step is generation of residuum signals using models of water flow, while the second one is leakage location using both neuro-fuzzy classifiers and residuum signals. Each classifier is tuned to detect a leakage only in one subarea of the water distribution network which was partitioned *a'priori* by domain experts. This solution results from application of the Matlab (MATLAB (2009)) implementation of anfis algorithm, where a one-output neuro-fuzzy inference system is assumed. Taking this into account, the verification of the proposed method consists of the two steps:

(1) identification of models of water flow through selected nodes of the water distribution network,
(2) training and testing of classifiers for leakage location.

The additional assumption for verification of the proposed method was that the leakage in the supply network was detected by means of a classifier for residuum signal evaluation of the water flow at the input node in the network. Based on expert opinions, the detection level for this classifier was set to $5m^3/h$. Thus, classifiers for leakage location was trained and tested using data representing leakage with the emission level equal to $5m^3/h$ as well.

### *3.1 Identification of models of water flow*

In the domain of mathematical modeling of systems and processes, there is a variety of existing model types and methods of identification of model parameters. In particular, they include: linear parametric models (AR, ARX, MAX, ARMA, ARMAX, . . . ) (Box et al. (1994); Ljung (1986), nonlinear parametric models (Janiszowski (2002)), neural networks (Korbicz et al. (2004)), neuro-fuzzy systems (Kościelny (2001)), models of Support Vector Machine (Vapnik (1995); Schölkopf and Smola (2001)), and many others. The output signal, input signals and model structure are required to identify parameters values for all of the models mentioned above. Identification of the model parameters values is realized using appropriate objective function.

In the conducted research, the Support Vector Machine (SVM) (Vapnik (1995); Schölkopf and Smola (2001)) was used as a modeling method. This method was chosen arbitrary, however investigation into selecting the appropriate modeling method of flows in water distribution networks should be carried out.

The SVM has a numerous advantages, e.g. it allows for modeling functional relationships, the SVM model can be

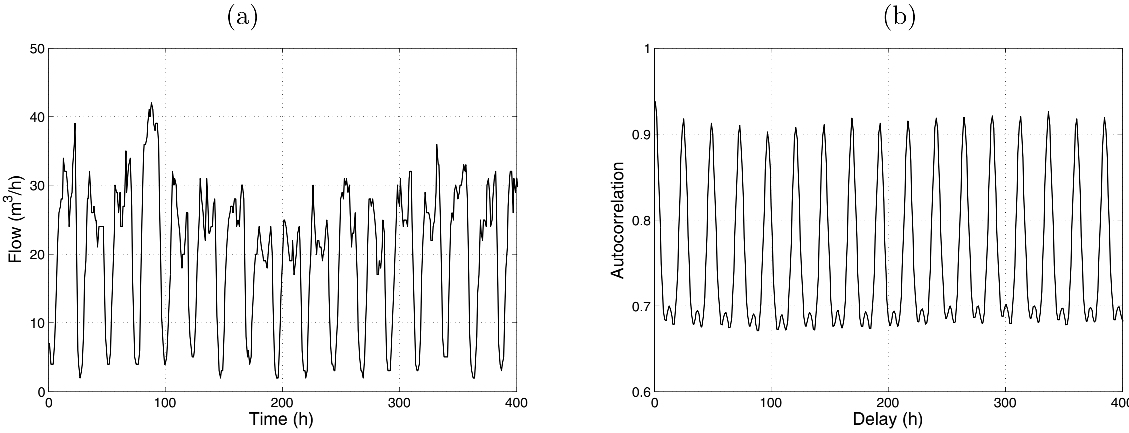


Fig. 2. Correlation analysis of the process variable $Q_6$ (Fig. 4): (a) water flow and (b) autocorrelation analysis

Table 1. Structures of models of water flow through selected nodes in the considered water distribution network (Fig. 4)

| Model ID | Model Output | Model Inputs | Inputs Order |
|---|---|---|---|
| $mQ_{21}$ | $Q_2(k)$ | $Q_2(k)$ | $[12]$ |
| $mQ_{22}$ | $Q_2(k)$ | $Q_2(k), Q_2(k-68\Delta t), Q_2(k-664\Delta t)$ | $[12, 20, 20]$ |
| $mQ_{23}$ | $Q_2(k)$ | $Q_2(k), Q_2(k-664\Delta t)$ | $[12, 20]$ |
| $mQ_{31}$ | $Q_3(k)$ | $Q_3(k)$ | $[12]$ |
| $mQ_{32}$ | $Q_3(k)$ | $Q_3(k), Q_3(k-68\Delta t), Q_3(k-664\Delta t)$ | $[12, 20, 20]$ |
| $mQ_{33}$ | $Q_3(k)$ | $Q_3(k), Q_3(k-664\Delta t)$ | $[12, 20]$ |

Table 2. Exemplary results of SVM modeling

| Model ID | $\varepsilon$ | $C$ | Kernel Function | Number of Support Vectors |
|---|---|---|---|---|
| $mQ_{21}$ | 0.11 | 1.0 | linear | 10 |
| $mQ_{22}$ | 0.13 | 1.0 | linear | 10 |
| $mQ_{23}$ | 0.10 | 1.0 | linear | 8 |
| $mQ_{32}$ | 0.11 | 1.0 | linear | 10 |

linear or nonlinear depending on a kernel function, etc. It can be used to model dynamics of systems and processes as well. At the first stage, the model output, inputs and the model order and delays of input signals must be defined (Table 1). Based on this information, training and testing data sets are created (Wachla (2006); Wachla and Moczulski (2007)). Next, the kernel function type, values of its parameters and values of metaparameters of the SVM are established. Finally, after the SVM learning has taken place, a set of support vectors, their weights and other parameters of the SVM model are obtained.

In this work, models of water flow were identified using data from the measurement system installed in the water distribution network (Fig. 4) in the city of Rybnik in the southern region of Poland. The total length of pipelines of the network exceeds 25 km. The network provides water to over a thousand individual and industrial customers. The locations of flow meters are shown in Fig. 4. A more detailed description of the network can be found in (Wyczółkowski and Wysogląd (2007); Wyczółkowski (2008)).

Measurement data obtained from 7th September 2009 to 8th November 2009 was used to identify the SVM models. A data sampling interval Δt was equal to 15 minutes. The data described above was selected due to their completeness, the autumn weather stability and leakage absence. The data was divided into two sets, i.e. a training data set and a testing data set. The training set was formed by measurement data from 7th September 2009 to 8th October 2009, while the testing set was formed by measurement data from 9th October 2009 to 8th November 2009. Both data sets have the same number of learning examples. The SVM model identification was realized using the *LibSVM* toolbox (Chang and Lin (2001)). Each data set was scaled to the interval $[0, 1]$ due to the stability of the calculation realized by the *LibSVM* software.

Six process variables $Q_1$,...,$Q_6$ representing water flow ($m^3/h$) through selected nodes of the water distribution network in Rybnik were considered (Fig. 4). The process variable $Q_6$ represents water flow through the node located at the input pipeline to the network.

Based on the correlation analysis (Fig. 2), the SVM model input signals and their delays and orders were established (Table 1). The analysis of the autocorrelation of each process variable has indicated that there exists a strong relation between the actual water flow and water flow from the period of 1, 2, 3, ... days ago (Fig. 2b). The analysis of the cross correlation between process variables has indicated the same results as the autocorrelation analysis. Based on these results, the structures (e.g. kernel function and values of its parameters) of the SVM models of water flow were defined (Table 2).

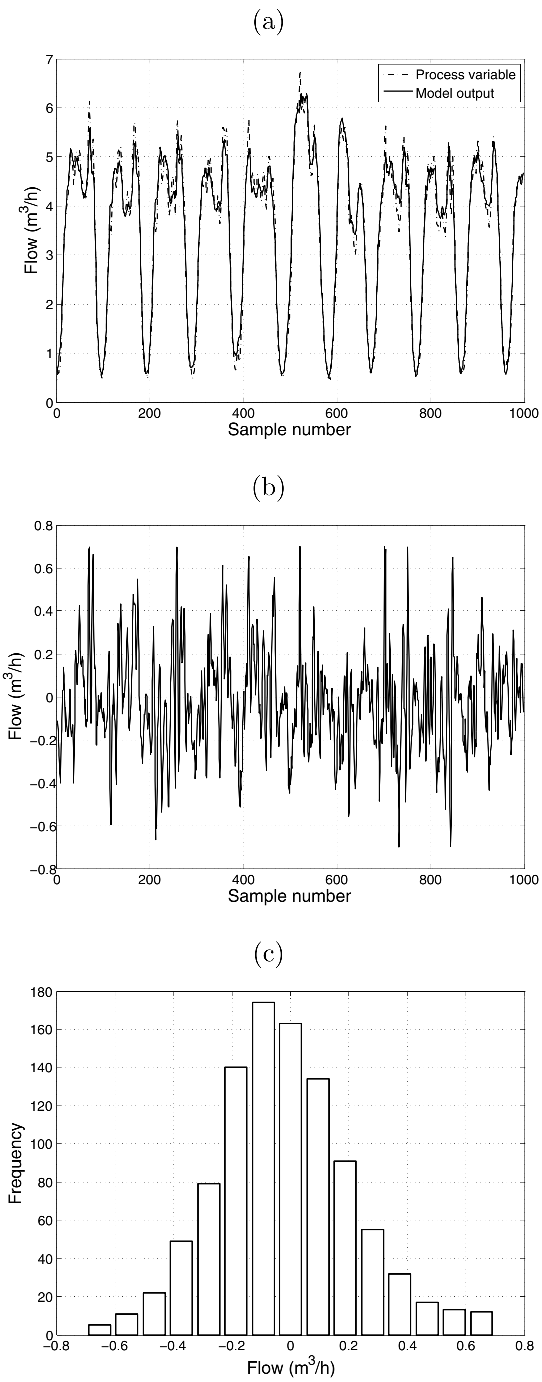


Fig. 3. Modeling results of the process variable $Q_2$ with application of the $mQ_{23}$ model: (a) the output signal, (b) the residuum signal, (c) the histogram analysis

A well-known problem of the application of the SVM is the selection of the right values of metaparameters. The fundamental regression approach of the SVM (the $\varepsilon$-SVR (Vapnik (1995))) has two metaparameters. These are:

- the $\varepsilon$ parameter which is used to control the width of the insensitive zone in loss function,
- the regularization parameter $C$ (Vapnik (1995); Schölkopf and Smola (2001)).

The search for the optimal value of the $\varepsilon$ parameter in relation to the number of support vectors and model accuracy was realized in an exhaustive manner. The optimal value of the $\varepsilon$ parameter was searched for within the limits $\langle 0.001; 0.5\rangle$ and the step value $\Delta\varepsilon = 0.001$. For every value of the $\varepsilon$, the adequate SVM model was identified using training data. The value of the metaparameter $C$ was 1.0. This process was realized in an automatic manner.

Table 3. Models evaluation: MSE – Mean Square Error, R – Correlation coefficient, MAPE – Mean Absolute Percentage Error (%)

| Model ID | MSE | R | MAPE |
|---|---|---|---|
| $mQ_{21}$ | 1.96E-3 | 0.978 | 4.01 |
| $mQ_{22}$ | 2.99E-3 | 0.968 | 4.79 |
| $mQ_{23}$ | 1.44E-3 | 0.977 | 3.92 |
| $mQ_{32}$ | 1.74E-3 | 0.967 | 3.87 |

Table 3 and Figure 3 presents results of models evaluation at the testing phase. In general, the obtained results confirm the hypothesis that models with additional input signals of water flow formed by measurement data at one, two, . . . days ago allow to identify more sensitive models due to the leakage level.

The modeling detailed analysis results shows that families of models $mQ_{i2}$ and $mQ_{i3}$ (Table 1) have better performance than others. It should be noted that the process variable $Q_i(k)$ and its historical values were used to form inputs of these models. Furthermore, for models $mQ_{23}$ and $mQ_{32}$ a high value of the correlation coefficient $R$ and the smallest values of the Mean Square Error (MSE) and the Mean Absolute Percentage Error (MAPE) (Table 3) indicators were obtained. Additionally, the evaluation results have been confirmed by means of a comparison analysis of the model output with the adequate process variable (Fig. 3a) and as well as by means of histogram (Fig. 3c) and residuum signal analysis (Fig. 3b).

### 3.2 *Training and testing classifiers for leakage location*

The data for training and testing classifiers were obtained using the analytical model of the water distribution network of the Kamień district in the city of Rybnik. This model was prepared in the EpaNET (Rossman (2000)) software. Based on the EpaNet model of the water distribution network, instances of leakage were simulated. The level of every simulated leakage instance was 5 $m^3/h$. The response of the network was observed at nodes related to locations of flow meters installed on the real water distribution network in the Kamień district. In the Figure 4, the exact location of flow meters was presented. The partition of the water distribution network into 23 subareas was made in connection to the early investigations conducted by research teams of Silesian University of Technology and the Water Supply and Sewage Company in Rybnik city. The model of the water demand by means of the network end users was included in simulations as well.

The number of leakage emissions and their location was determined for each subarea of the water distribution network. It was assumed that the number of leakage emissions at every single subarea will be equal to 4. The first 3 emissions were the source of the training data while the 4th emission was the source of the testing data. A uniform distribution of the location of all leakage emissions at every single subarea was also assumed. Based on these

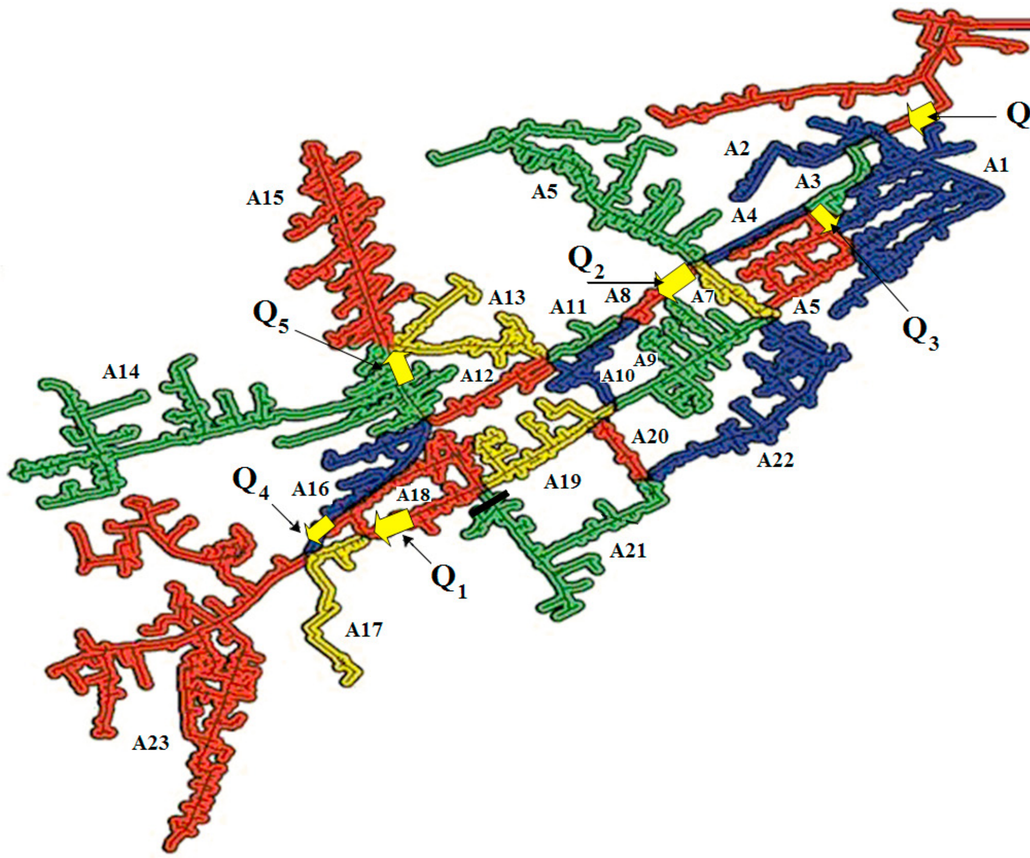


Fig. 4. The structure and subareas partition of the water distribution network of the Kamień district in the city of Rybnik, Poland

assumptions, a simulation of operation of the water distribution network was conducted. A normal condition of the network, i.e. a network state without leakage, was included as well. Due to the dynamic properties of the water distribution network and requirements of water flow models, one assumed that a simulation of the water distribution system operation would be 4 weeks with a time interval $\Delta t = 15$ minutes.

In the next phase, models of water flow through selected nodes of the water distribution network were used to generate residuum signals. Based on the residuum signals, data sets for training and testing the classifiers were prepared.

As mentioned above (Section 2), the neuro-fuzzy system ANFIS (MATLAB (2009)) was developed mainly for regression tasks. Nevertheless, ANFIS system can be used in data classification. If one assumes that ANFIS output value can be 0 or 1 then the ANFIS system may be used in data classification. Values 1 and 0 represent two different classes, e.g. a leakage at the 14th subarea of the water distribution network (value 1) and instances of leakage in other subareas (value 0) (Fig. 5(b)). This approach leads to a nonuniform data distribution over both classes. It means that the class with positive examples is smaller than the class with negative ones. In other words, the amount of simulation data for one subarea of the water distribution network is always lower than the sum of simulation data from other subareas.

Taking the above into consideration, the training data sets were prepared by means of connection of residuum signals containing information about leakage. In particular, from each residuum signal the 96 samples (one day) were extracted. Next, the subsignals were connected together to form the input signals for the neuro-fuzzy classifiers. A uniform distribution of the learning data was obtained by several repetitions of the data connected to the subarea containing leakage in relation to other subareas of the water distribution network without leakage (Fig. 5(a)).

The training of classifiers for leakage location was initiated by establishing the initial structure and parameters value of the neuro-fuzzy models. This was realized using the subtractive clustering method (Chiu (1994)). In particular, the squash factor equal 1.25, the accept ratio equal 0.5, the reject ratio equal 0.15 and the gaussian membership functions on each input were used. Finally, the structure of the neuro-fuzzy inference system was generated. It was composed with three rules, three membership functions on the every single input and three linear membership function on output. Subsequently, the *anfis* (Jang (1993)) algorithm was used to train neuro-fuzzy models created at the previous step.

Because the output of the neuro-fuzzy model is continuous in the amplitude domain, the assumption of the proposed method was such that the output signal of the neuro-fuzzy model would be the input to the simple binary classifier. The simple binary classifier compares the output value of the neuro-fuzzy model with threshold value (Fig. 5(c)). If

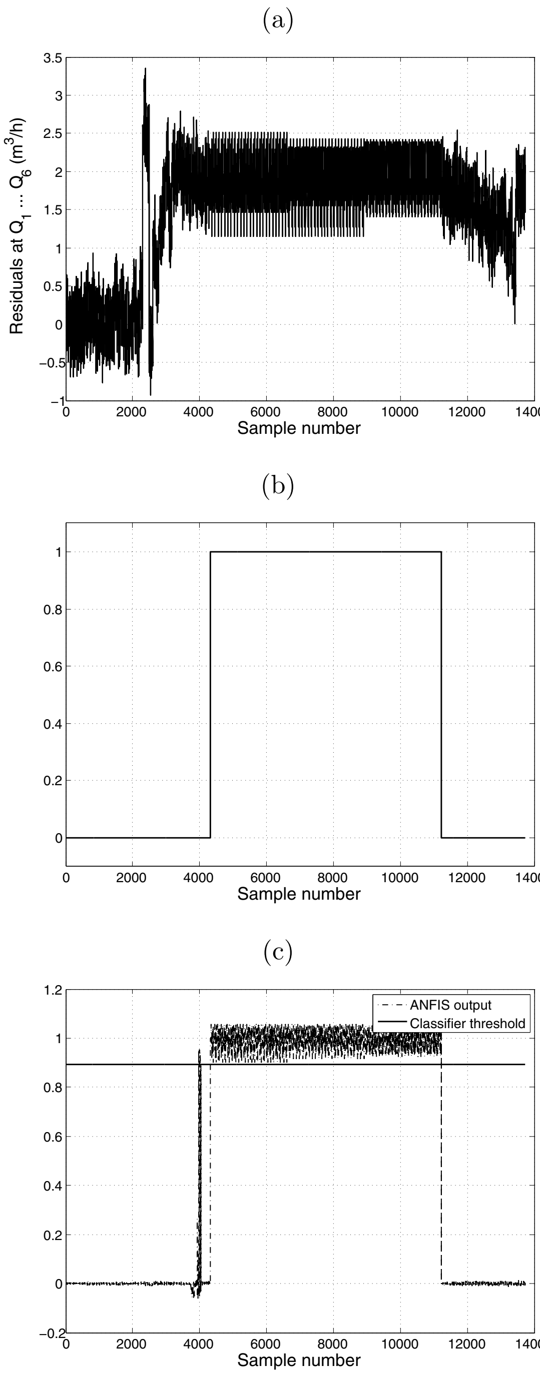


Fig. 5. Training data and training result of a classifier connected to the subarea A14: (a) input training data, (b) output training data, (c) anfis output and the threshold value

the output value of the neuro-fuzzy model is greater than the threshold value then the classifier output is 1. Otherwise, the classifier output is 0. The training of the simple binary classifier consists of identification of the right value of the threshold. The identification of the threshold value was realized using evolutionary algorithm (Michalewicz (1996)) and the objective function was based on a classification error. Output signals from neuro-fuzzy models were used to train binary classifiers.

Table 4. Classifiers performance: MRE – Mean Relative Error, MSE – Mean Square Error

| Classifier ID | Decision Threshold | ANFIS | | Classification Error | |
|---|---|---|---|---|---|
| | | MRE | MSE | Training | Testing |
| **C1** | 0.617 | 6.17E-2 | 5.95E-2 | 0.061 | 0.067 |
| **C2** | 0.721 | 2.04E-2 | 8.59E-2 | 0.007 | 0.022 |
| **C3** | 0.551 | 5.34E-2 | 5.03E-2 | 0.057 | 0.081 |
| **C4** | 0.640 | 6.69E-2 | 6.63E-2 | 0.075 | 0.060 |
| **C5** | 0.589 | 6.25E-2 | 4.76E-2 | 0.045 | 0.046 |
| **C6** | 0.485 | 5.69E-2 | 5.36E-2 | 0.057 | 0.049 |
| **C7** | 0.588 | 7.12E-2 | 5.72E-2 | 0.050 | 0.024 |
| **C8** | 0.587 | 5.71E-2 | 3.39E-2 | 0.023 | 0.014 |
| **C9** | 0.575 | 7.21E-2 | 5.64E-2 | 0.049 | 0.027 |
| **C10** | 0.620 | 5.78E-2 | 3.71E-2 | 0.031 | 0.013 |
| **C11** | 0.615 | 5.34E-2 | 2.92E-2 | 0.024 | 0.012 |
| **C12** | 0.732 | 4.51E-2 | 3.78E-2 | 0.035 | 0.040 |
| **C13** | 0.654 | 3.12E-2 | 8.79E-2 | 0.001 | 0.001 |
| **C14** | 0.718 | 2.50E-2 | 1.37E-2 | 0.016 | 0.013 |
| **C15** | 0.894 | 1.01E-2 | 3.45E-2 | 0.001 | 0.002 |
| **C16** | 0.770 | 4.41E-2 | 2.47E-2 | 0.012 | 0.081 |
| **C17** | 0.886 | 2.06E-2 | 1.99E-2 | 0.021 | 0.022 |
| **C18** | 0.593 | 3.86E-2 | 1.39E-2 | 0.007 | 0.000 |
| **C19** | 0.631 | 5.68E-2 | 5.55E-2 | 0.054 | 0.051 |
| **C20** | 0.695 | 6.03E-2 | 6.00E-2 | 0.048 | 0.065 |
| **C21** | 0.676 | 5.54E-2 | 5.48E-2 | 0.051 | 0.041 |
| **C22** | 0.553 | 7.42E-2 | 6.22E-2 | 0.064 | 0.060 |
| **C23** | 0.846 | 2.80E-2 | 2.56E-2 | 0.025 | 0.021 |

The results presented in Table 4 do not consider cases of the classifier behavior when leakage occurs in other subareas of the water distribution network except the nominal one. Therefore, the confusion matrix (Fig. 6) was presented as well. The values in the confusion matrix vary from 0 to 1 and they represent the leakage detection probability. The confusion matrix clearly shows that every classifier properly detects leakage occurred in related subarea (the main diagonal of the confusion matrix). On the other hand, some of them are sensitive to leak occurrence in other subareas than nominal one as well. In particular, this observation is applied to the two groups of the network subareas, i.e. $A1 \div A6$ and $A18 \div A22$, where the listed subareas are adjacent to each other (Fig. 4). In that case, it is not possible to identify the right subarea with leakage by using one classifier. Therefore,the analysis of the responses of the other classifiers from neighborhood is required.

## 4. SUMMARY

A method of leakage location in water distribution networks was proposed. The method is based on a set of neuro-fuzzy classifiers. Each classifier is tuned to detect leakage only in one subarea of a water distribution network partitioned *a'priori* by domain experts. Models of water flow through selected nodes of the network form the first stage of this method. Next, all outputs of these models become an input of each classifier. The main purpose of the method development was to increase the accuracy of leakage detection and location. The assumptions of the proposed method were confirmed by results from verification research.

The elaborated method can be a foundation of a diagnostic system for leakage detection and location in water distribution networks. Due to the complexity of the method, a user interface for diagnosis presentation should be simplified.

| Classifier ID | Subarea ID | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 | A12 | A13 | A14 | A15 | A16 | A17 | A18 | A19 | A20 | A21 | A22 | A23 |
| C1 | 0,96 | 0 | 0,98 | 0,19 | 0 | 0,98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C2 | 0 | 1 | 0 | 0 | 0,99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C3 | 0,77 | 0 | 0,93 | 1 | 0 | 0,27 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C4 | 0,95 | 0 | 1 | 1 | 0 | 0,8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C5 | 0,02 | 1 | 0,17 | 0,93 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C6 | 0,93 | 0 | 0,9 | 0,13 | 0 | 1 | 0,3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C7 | 0,36 | 0 | 0,27 | 0,03 | 0 | 0,39 | 1 | 0 | 0,04 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C8 | 0,02 | 0 | 0,01 | 0 | 0 | 0,14 | 0 | 1 | 0 | 0 | 0,48 | 0 | 0,01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C9 | 0 | 0 | 0 | 0 | 0 | 0 | 0,11 | 0 | 1 | 0,07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,25 | 0 | 0,8 | 0 |
| C10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,05 | 0,1 | 1 | 0,03 | 0 | 0 | 0 | 0 | 0 | 0 | 0,01 | 0,07 | 0,23 | 0,04 | 0,05 | 0 |
| C11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0,56 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0,92 | 0 | 0,94 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0,06 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,99 | 0 | 0,38 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,03 | 0 | 0,09 | 0 | 0,84 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| C18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0,01 | 0 | 0 |
| C19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,42 | 1 | 0,84 | 1 | 0,06 | 0 |
| C20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,02 | 0,73 | 0,99 | 0,81 | 0,97 | 0 |
| C21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,02 | 1 | 0,72 | 1 | 0,17 | 0 |
| C22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0,01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,02 | 0,07 | 0,81 | 0,36 | 0,98 | 0 |
| C23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0,98 | 0 | 0 | 0 | 0 | 0 | 1 |

Fig. 6. The confusion matrix of classifiers for leakage detection and location in subareas of the water distribution network

## REFERENCES


Box, G.E., Jenkins, G.M., and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall.

Chang, C.C. and Lin, C.J. (2001). *LIBSVM: a library for support vector machines*. Software available at `http://www.csie.ntu.edu.tw/ cjlin/libsvm`.

Chiu, S. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent & Fuzzy Systems*, 2(3).

Gertler, J., Romera, J., Puig, V., and Quevedo, J. (2010). Leak detection and isolation in water distribution networks using pca and structured residuals. In *Control and Fault-Tolerant Systems (SysTol)*, 191–196.

Jang, J. (1993). Anfis: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685.

Jang, J. and Sun, C. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ.

Janiszowski, K. (2002). *Identyfikacja modeli parametrycznych w przykładach*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Korbicz, J., Kościelny, J.M., Kowalczuk, Z., and Cholewa, W. (Eds.) (2004). *Fault diagnosis. Models, artificial intelligence, applications*. Springer Berlin/Heidelberg. doi:10.1017/S0263574704241133. URL `http://dx.doi.org/10.1017/S0263574704241133`.

Kościelny, J.M. (2001). *Diagnostyka zautomatyzowanych procesów przemysłowych*. Akademicka Oficyna Wydawnicza EXIT, Warszawa.

Ljung, L. (1986). *System Identification: Theory for the User*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

MATLAB (2009). *Fuzzy Logic Toolbox User's Guide*. The MathWorks, Inc.

Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin Heidelberg.

Poulakis, Z., Valougeorgis, D., and Papadimitriou, C. (2003). Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 18, 315–327.

Przystałka, P. and Moczulski, W. (2012). Optimal placement of sensors and actuators for leakage detection and localization. In *C 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Mexico City Mexico*, 666–671.

Puust, R., Kapelan, Z., Savic, D., and Koppel, T. (2010). A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7/1, 25–45.

Rossman, L.A. (2000). *EPANET 2 Users Manual*.

Schölkopf, B. and Smola, A.J. (2001). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.

Wachla, D. (2006). *Identyfikacja dynamicznych modeli diagnostycznych metodami odkryć wiedzy w bazach danych*, volume 130. Katedra Podstaw Konstrukcji Maszyn, Politechnika Śląska, Gliwice.

Wachla, D. and Moczulski, W.A. (2007). Identification of dynamic diagnostic models with the use of methodology of knowledge discovery in databases. *Engineering Applications of Artificial Intelligence*, 20(5), 699–707.

Wyczółkowski, R. (2008). Intelligent monitoring of local water supply system. *Eksploatacja i Niezawodnosc Maintenance and Reliability*, 1, 33–36.

Wyczółkowski, R. and Wysogląd, B. (2007). An optimization of heuristic model of water supply network. *Computer Assisted Methods in Engineering and Science*, 14, 767–776.