UDC 004.8:628.1
DOI: 10.20535/SRIT.2308-8893.2026.2.06

HYBRID COMPUTING INTELLIGENT SYSTEM FOR ASSESSING THE STABILITY OF THE WATER DISTRIBUTION SYSTEM AND DETERMINING THE OPTIMUM LOCATIONS OF PRESSURE SENSORS

YU.P. ZAYCHENKO, T.V. STAROVOIT

Abstract. This study introduces a method for determining the best locations for pressure sensors in water supply networks and for assessing network conditions using artificial intelligence techniques. The goal is to identify the network nodes that would provide the most important information for detecting water leaks and evaluating the overall network status. The selection of sensor locations was based on data sets of pressure changes caused by various leak scenarios generated by EPANET simulations. Genetic algorithms were used to rank candidate nodes and determine the optimal number of sensor locations. The next step involved assessing the network state using the ANFIS neuro-fuzzy network and the Mamdani neuro-fuzzy logical inference algorithm. These algorithms were implemented in the Google Colab environment and tested on a section of the water supply network in Kyiv, Ukraine.

Keywords: sensor placement, WDN, artificial intelligence, hydraulic modeling, ANFIS, Mamdani, genetic algorithms, EPANET 2.2.

INTRODUCTION

A water supply network (WDN) is a complex engineering network made up of pipelines and other elements that ensure the functioning of the water supply system from the source to the end users (end nodes). The primary function of the WDN is to deliver the expected volume of water at sufficient pressure. Real-time monitoring of the network’s condition is crucial for stable operation to assess its current performance and ability to fulfill its core functions.

Most studies focus on quantifying the performance of the entire WDN. Shin et al. [28] proposed quantitative methods for determining the stability of the water supply network. Analyzing the network at the global level using a single metric does not provide information about the level of performance of the nodes in the network. Therefore, it is important to monitor the performance of each network node. The pressure information at each node and the corresponding nodal outflow information are necessary to evaluate the state of the network at the node level. It also helps in adaptive network management for increased returns and maximum quality of service. It follows that data collection from each demand node is necessary. However, it is not possible to install measuring devices throughout the network. Since the more sensors, the higher the cost of initial installation and maintenance of the network. There is also a need to power the sensors for their operation.

Determining the ideal placement for pressure sensors is a key challenge for monitoring water distribution networks. It's not feasible to install sensors at every point within the network, which consists of thousands of nodes. Therefore, only a select few nodes can have sensors installed. The main goal is to figure out the best locations for these sensors so that they can provide relevant data for assessing hydraulic variables at unmonitored points and help with various monitoring algorithms, such as leak detection [23]. However, the data from these sensors alone might not be sufficient to accurately pinpoint leaks and evaluate the network's condition, so additional sensors may need to be installed in other areas [24]. One solution to this problem is to install more pressure sensors, as they are more cost-effective and easier to install and maintain compared to flow rate sensors. Furthermore, pressure readings are more sensitive to leaks than flow rates, so many localization algorithms rely on network pressure measurements.

Once a sufficient number of measurements are obtained, they can be used to estimate the unmeasured states of the WDN at any point in time. The problem of restoring missing information or assessing the state of the system from damaged data is not new. There are several methods based on spatial, spectral, temporal, and statistical approaches. Hu et al. [29] worked on the estimation of discrete-time networks. In this work, it is assumed that the measurements follow a certain distribution. In the study [30], state assessment was performed for networks with a variable topology. Liu et al. [31] developed a decentralized state estimator for spatially distributed systems. The implementation involves state space matrices A and C, which have the shape of a block diagonal. This method of matrix filling (MS) for state estimation is quite popular in various fields.

In order to enhance network reliability, Chandramouli and Malleswarvrao [11] utilized fuzzy logic based on excess pressure available at demand nodes. Prasad and Park employed genetic algorithms, taking into account both cost minimization and network reliability maximization [12]. Recent advancements include improving algorithm convergence by employing a constructed starting population instead of a random one [13], enhancing computing efficiency by reducing the search space [14], and combining GA and mathematical programming with the inclusion of new elements, such as reduction valves [15]. Other methods involve the use of artificial neural networks (ANNs) instead of hydraulic and simulation models for water quality, along with differential evolution (DE) for optimization [16], as well as the development of the Harris Hawks optimization algorithm (HHO) for WDN optimization [16].

While gathering information about the water supply network in Kyiv, we encountered difficulties due to limited and unstructured data, as well as a large amount of paper records. To address this, active efforts are underway to digitally transform and develop GIS systems. In this study, Mamdani's method [22], a reliable and flexible fuzzy logic inference method, was adapted to convert non-quantitative "expert" knowledge into a quantitative scale. This adaptation is valuable for mapping a relatively complex water supply network and allows for the integration of non-statistically independent components.

This paper proposes a method for determining the optimal locations of pressure sensors using genetic algorithms and assessing the state of the water supply network using the ANFIS neuro-fuzzy network and the Mamdani fuzzy logic inference algorithm. The results of this study make it possible to assess the state of the water supply network and find the best places for the placement of pressure sensors. Based on the previous research of Khorshidi [16], who also performed the relevant work, this method not only helps to find the leakage points but also evaluates the state of the water supply network.

The proposed intelligent model can be used for controlling pressure or flow rate sensors in water drainage networks. It is necessary to utilize software tools for hydraulic modeling of water drainage networks, as pipeline breaks are also present in sewage networks [18–20]. Additionally, this method can help identify the best locations for water quality monitoring sensors. With the assistance of neuro-fuzzy neural networks and fuzzy logical inference algorithms, it can also be used to assess network conditions and predict water quality.

DESCRIPTION OF MODEL DEVELOPMENT AND TRAINING METHODOLOGY

Modeling of water demand

The cor-PRP model, which was proposed in the study [2], was used to form the demand at the common node. In this model, the nodal frequency of water use in residential premises corresponds to a non-stationary Poisson process. In particular, the probability that Pr has exactly k pulses generated during a time period of duration $\Delta\tau$[c] is determined by the following equation:

$$Pr(k) = e^{-\lambda\Delta\tau}\,\frac{(\lambda\Delta\tau)^k}{k!}, \qquad (1)$$

where the parameter λ[c-1] is the average pulse arrival frequency, which is considered constant during $\Delta\tau$. After generating pulse arrivals, the duration D [s] and intensity I [l/s] of the total pulses are sampled from probability distributions such as the log-normal and beta distributions [2]. An important feature of the cor-PRP model is the correlation between pulse duration and intensity, which improves pulse stability [2].

Hydraulic model of the water supply network

A hydraulic model is commonly used to calculate hydraulic parameters such as water pressure and flow rate for water network design. Hydraulic formulas describe the conservation of mass and conservation of energy, taking into account the topological characteristics of the water supply network. The hydraulic model takes into account fluctuations in water demand and leaks that affect network performance. The main formulas used for hydraulic modeling are indicated in the equations below [4].

Formula (2) calculates the mass transfer in the pipe node; it indicates that in the absence of leakage, the inflow of water to the pipe node should be equal to the outflow of water [4]:

$$\sum_{p\in P_n} q_{p,n} - D_n^{act} = 0, \forall n \in N, \qquad (2)$$

where P_n is a set of pipes connected to node n; $q_{p,n}$ is the flow of water in node n from pipe p (m^3/s), D_n^{act}. Formula (3) describes the energy transfer, that is, the total water head, which includes components describing kinetic energy (kinetic water head), hydraulic potential energy (head) and gravitational potential energy (height head) [4]:

$$h_A = \frac{u_A^2}{2_g} + \frac{p_A}{\gamma_\omega} + z_A = h_B + H_L = \frac{u_B^2}{2_g} + \frac{p_B}{\gamma_\omega} + z_B + H_L\,, \tag{3}$$

where h is the total water head; u is the water velocity at each node; and z is the height of each node. H_L is the energy loss value between node A and node B [4]. Energy consumption in the pipe flow can be distributed or localized. The distributed energy consumption is determined by the flow rate V, the internal diameter of the pipe d, the length of the pipe L, and the roughness of the pipe wall, which is determined by the Hazen–Williams formula [5], formula (4):

$$H(m) = \left(\frac{6.78L}{d^{1.65}}\right)\left(\frac{V}{C}\right)^{1.85}, \tag{4}$$

where C is the roughness coefficient of the pipe wall. Localized energy losses occur due to turbulence associated with changes in flow conditions (such as flow velocity, direction, etc.) determined by the topology of water supply network connections [4].

For the water supply network, the main thing is the consumption or demand for water. Two models are used for water demand in nodes: a demand-driven model and a pressure-driven model [4]. This study uses a pressure-driven water demand model to consider the effects of pressure loss due to changes in water demand or leaks [4]:

$$D = \begin{cases} 0 & p \le P_0 \\ D_f\left(\frac{p - P_0}{P_f - P_0}\right)^{\frac{1}{2}} & P_0 \le p \le P_f \\ D_f & p > P_f \end{cases}, \tag{5}$$

where D is the demand in each node; D_f is the desired demand (m^3/s); p is the water pressure; P_f is the pressure above which the desired demand D_f must be satisfied; P_0 is the pressure below which water will not be supplied to the node [4].

Genetic algorithms of sensor placement

In the water supply network, the main focus is on water consumption or demand. There are two models used for water demand in nodes: a demand-driven model and a pressure-driven model [4]. This study utilizes a pressure-driven water demand model to account for the impacts of pressure loss caused by variations in water demand or leaks [4].

Each possible solution of the optimization problem, using GA, is called a chromosome. The mathematical formulation of the optimization problem is based on the fact that each chromosome consists of a series of genes (decision variables) that represent a possible solution to the optimization problem. In an N-dimensional optimization problem, a chromosome is an array of size $1 \times N$. This array is defined as follows [3]:

$$Chromosome = X = (x_1, x_2, \dots, x_i, \dots, x_N),$$

where X is a possible solution to the optimization problem; x_i is the i-th decision variable (or gene) of decision X; and N is the number of decision variables.

A genetic algorithm begins by randomly generating a population of chromosomes or possible solutions. The size of the population, or the number of possible solutions, is presented in the form of a matrix of chromosomes of size $M \times N$ [3]:

$$Population = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_i \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,i} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,i} & \cdots & x_{2,N} \\ & & & \vdots & & \\ x_{j,1} & x_{j,2} & \cdots & x_{j,i} & \cdots & x_{j,N} \\ & & & \vdots & & \\ x_{M,1} & x_{M,2} & \cdots & x_{M,i} & \cdots & x_{M,N} \end{bmatrix},$$

where X_j is the j-th solution (or chromosome); $x_{j,i}$ is the i-th solution variable (or gene) of the j-th solution; and M is the population size. Each decision variable $x_{j,i}$ can be represented as a floating-point number (real values), or as a predefined set of values for discrete problems. Some of the initially generated possible solutions are selected as parents to create a new generation [3].

Selection in GA is a procedure by which R ($R < M$) individuals are selected from a population for reproduction. The selected individuals are the parents of the next generation and make up the parental population. The ranking option ranks all chromosomes based on their match values. The best solution gets rank 1, and the worst gets the lowest rank. The decision is assigned a probability that is proportional to its rank according to the following linear function [3]:

$$P_k = U - (S_k - 1) \times Z\,, \tag{6}$$

$$S_k = Rank\ (X_k), \tag{7}$$

$$\sum_{j=1}^{M} P_j = 1, \tag{8}$$

$$U = \frac{Z\,(M-1)}{2} + \frac{1}{M}\,, \tag{9}$$

where S_k is the rank of the *k-th* solution in the population; $S_k = 1$ indicates that the k-th solution is the best solution; and Z is a user-defined value. Fig. 1 shows the sorting of solutions according to the fit function (F) in the maximization problem [3].

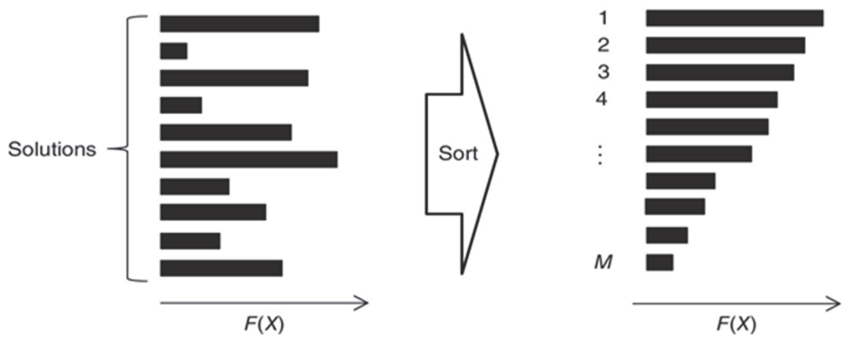

Fig. 1. Ranking of chromosomes according to the correspondence function (F) [3]

An alternative method ranks all solutions according to their fitness values. Then, $M-S$ copies of each solution are generated [3].

Pressure sensor placement method using genetic algorithms

The proposed sensor placement method is based on a nodal pressure data set that simulates typical variations due to leaks of different sizes at all network nodes. Data on pressure in the network were obtained using hydraulic modeling of the water supply network of the city of Kyiv. Each pressure data point is labeled with a leak class, for further classification of the data.

The method of placement of pressure sensors is the stage of selecting a function. In order to select the characteristics (a subset of nodes where the sensors will be placed), an algorithm is proposed that seeks to maximize the relevance of the selected characteristics (node pressure) for the response variable (leakage node). Each response variable avoids capturing information already contributed by others, that is, redundancy is minimized. The definitions of relevance and redundancy proposed in the study [32] are used as a basis for determining the methodology:

a) Relevance is an indicator of the relevance of a subset of nodal pressures. It is calculated according to the formula:

$$Rel = (\mathcal{S}) \stackrel{\text{def}}{=} \frac{1}{S} \sum_{x \in \mathcal{S}} I\,(x, y)\,, \tag{10}$$

where x is any function in $\mathcal{S}$; and $S = |\mathcal{S}|$ is the number of functions in $\mathcal{S}$ (power).

b) Redundancy – the metric of information redundancy in a subset of the function $\mathcal{S}$, determined by the formula:

$$Red = (\mathcal{S}) \stackrel{\text{def}}{=} \frac{1}{S^2} \sum_{x, x' \in \mathcal{S}} I\,(x, x')\,, \tag{11}$$

where x and x' are the quality of the feature in $\mathcal{S}$.

To apply the above formulas in order to calculate the location of the pressure sensor, a data set of pressures at the nodes is first created, which provides different scenarios that take into account leaks of different sizes at all nodes of the network. A series of nodal pressure samples, one sample for each leakage scenario, were obtained using hydraulic simulation. If M different leakage scenarios are simulated in a network containing N nodes, the result of the simulation is a set of MN-dimensional vectors, x and x' in (10) and (11), corresponding to N candidate nodes (it is assumed that all nodes are potential definition nodes). Additionally, an output vector y (Equation (10)) containing integers is generated to indicate the leak node corresponding to each simulated scenario.

Finding the optimal subset of sensors $\mathcal{S}$ requires testing 2^N different combinations, which would require impractical computation time in multi-node networks. Therefore, the proposed method [32] is used to rank pressures in nodes using an iterative forward scheme that requires only (NS) combinations. With the help of this scheme, you can rank all the pressures in the nodes in order of importance with the calculated costs (N^2).

The next step involves creating a genetic algorithm to rank node pressures based on their importance in identifying different classes of leaks (nodes with leaks). The output list will start with the nodes considered most important for identifying leak locations based on the information in the dataset. The process of selecting nodes will begin with an empty subset and, at each step, the node with the best ranking among those not yet selected will be added. Each iteration will evaluate the relevance of each available function (nodal pressure) in relation to the output (leak node) and its redundancy with respect to previously selected variables.

The number of sensors needed for pressure monitoring and leak detection depends on the available equipment. The goal is to minimize the number of sensors based on the method's use of available information, measurement noise, sensor calibration quality, and resolution. It's important to consider that increasing the number of sensors doesn't always result in better outcomes. The process involves starting with one sensor (the best-ranked one) and gradually increasing the number of sensors while evaluating the leak localization performance for each new set of sensors. This continues until adding a new sensor no longer provides a significant advantage.

Intelligent models for assessing and forecasting the state of the water network

Mamdani Fuzzy Logic Inference System (MFISM). The model of fuzzy logic was presented by Zadeh in 1965 [33] in order to create systems close to human thinking. Works [34, 35] proved that this method is effective in the development of complex systems with uncertain conditions [37].

Fuzzy Logic Inference (FIS) consists of three components: fuzzification, inference, and defuzzification. In this system, knowledge is represented as a set of fuzzy linguistic rules that allow making fuzzy decisions (not 0 or 1, but values from 0 to 1). The human expert can be replaced by a combination of a fuzzy rule-based system (FRBS) and a block called a defuzzifier [37].

Fuzzification is a process of decomposition of input and output data into one or more fuzzy sets. The fuzzy set is defined using a membership function that represents the region of interest, i.e., the main interests in the interval [0, 1]. The shape of the curves shows the membership function for each data set and can be expressed in various geometric shapes, such as: trapezoid, triangle, etc. The membership function represents the degree of belonging of the values in the data set [36].

The membership function of the set A defined on the domain X has the form: $\mu(A)$: X [0; 1], the set A is defined through its membership function μ by the equation [37]:

$$\mu\ (A)\epsilon \begin{cases} = 1, if\ x\ is\ full\ member\ of\ A \\ \in [0,1]\ if\ x\ is\ partial\ member\ of\ A \\ = 0,\ if\ x\ is\ not\ member\ of\ A \end{cases}. \quad (12)$$

The following set for the trapezoidal membership function f is calculated according to the formula [37]:

$$f\ (x,a,b,c,d) = \begin{cases} 0, if\ x < a\ or\ x > d \\ \dfrac{(a-x)}{(a-\ b')}\ if\ \le x\ \le b \\ 1, if\ b\ \le x\ \le c \\ \dfrac{(d-x)}{(d-c)}, if\ c\ \le x\ \le d \end{cases}. \quad (13)$$

Inference rules represent relationships between subsets of inputs and outputs. Inference rules should create a new output subset. Each rule consists of two parts: "If" and "Then" [37].

The last stage of processing is the defuzzification procedure. This process allows you to convert the results obtained as fuzzy sets into numerical values. The center of gravity, the average of the maxima, and the smallest of the maxima are widely used as defuzzification methods [38].

Adaptive neuro-fuzzy logical inference system (ANFIS). The adaptive neuro-fuzzy inference system consists of five different levels. The first level is responsible for identifying input data and output variables and defining their descriptors. The second level defines membership functions for each input and output variable. The third level creates a rule base. Level 4 performs rule evaluation. The last level (Level 5) performs defuzzification [39].

At the **first level**, each "i" node of this equation is an adaptive node and a node membership function (MF) [39]:

$$O_i^1 = \mu_{Ai}(x), i = 1, 2, 3 \ldots \tag{14}$$

$$O_i^1 = \mu_{Bi}(x), i = 1, 2, 3 \ldots \tag{15}$$

Fuzzy MFs have different shapes, such as Gaussian, triangular, and trapezoidal.

Level 2. Calculation of power of rules [39]:

$$O_i^2 = W_i = \mu_{Ai}(x) * \mu_{Bi}(x), i = 1, 2, 3 \ldots \tag{16}$$

Level 3. Normalization of calculated values [39]:

$$O_i^3 = \frac{W_i}{\sum W_i}, i = 1, 2, 3, 4 \ldots \tag{17}$$

Level 4. In this level, each node represents a sequential part of a fuzzy rule [39]:

$$O_i^4 = \underline{W_i} * f_1 = \underline{W_i} * (P_k * x + q_k * y + r_k) \; where \; i = 1, 2 \ldots n. \tag{18}$$

At **level 5**, the sequential part of the rules is defuzzified by summing the outputs of all rules [39]:

$$Final\ Output = O_i^5 = \sum_i \bar{W}_i * f_1 = \bar{W}_i * (P_k * x + q_k * y + r_k). \tag{19}$$

SIMULATION RESULTS

Determination of the number of pressure control points and optimum placement scheme

For this study, the hydraulic zone of the digital double of the water network of the city of Kyiv (Fig. 2) with a length of approximately 30 km, 1335 connections and an average input flow of about 65 m3/h was used. A hydraulic simulation model was developed in EPANET 2.2 [40], with an hourly demand pattern. The nodal flow was divided into water consumption by the user and leakage. To establish the parameters of the model, a relationship between leakage and pressure was constructed.

Fig. 2. Hydraulic zone of the network of the city of Kyiv (Ukraine)

To construct the pressure data set, leaks of different magnitudes were simulated at each connection node using the hydraulic modeling program EPANET 2.2 [40]. The process of creating a dataset using the EPANET program, training and predictive use of classifiers for leak detection are described in [41]. The data set generated by simulations for this study considered leaks at all connection nodes with flow rates starting at 50 L/s. To simulate nodal leaks, the demand assigned to a given node in the EPANET hydraulic model was modified by increasing the demand by an amount equal to the flow of the simulated leak. The maximum number of installed sensors, N_{max}, was determined based on the characteristics of the network, namely the length of the network (one sensor per km).

As a result, two pressure sensitivity matrices were calculated. Pipe roughness indicators were determined by the Hazen–Williams formula (20) (for the first matrix):

$$F(N) = \frac{(aN)}{(b+N)}, \tag{20}$$

where a, b, c and d are function parameters.

The second matrix was obtained by generating a burst of fixed size for each node of the hydraulic model with a single emitter factor of 0.25; hydraulic modeling was performed in EPANET 2.2 [40].

The optimal placement of pressure sensors for a given quantity was formulated as an unconstrained multi-objective optimization problem. The decision variables were the nodes where the pressure sensors could potentially be installed, and all nodes were considered as possible placement locations. The objective functions were aimed at maximizing the sensitivity of the nodal pressure both to the variations of the pipe roughness coefficient (f_1) and to the cases of pipe rupture (f_2).

The problem was solved once for each number of sensors in each of the five discrete sets. The algorithm was implemented in the Google Colab environment on Python, using the Pymoo package [42]. Each discrete location (i.e., node) was translated as an integer value. A population member is a set of pressure sensor locations, and each population member variable represents a possible pressure sensor location. Fig. 3 shows the general diagram of nodes and edges (joints and pipes).

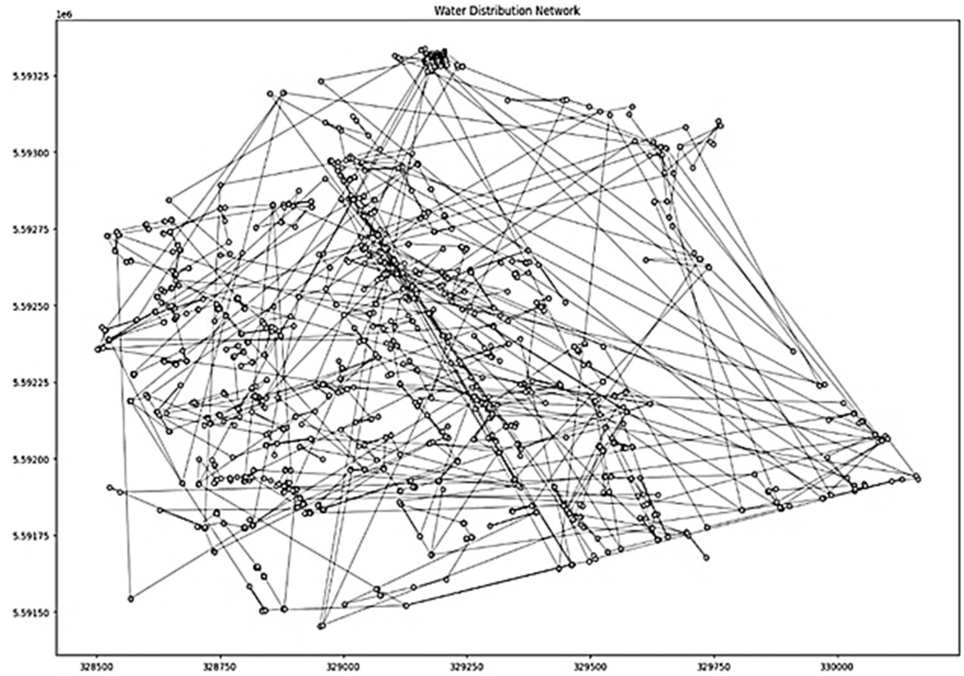

Fig. 3. General scheme of the network

The following parameters were used to develop the GA method: random integer sampling and selection operators, integer mutation with probability p_m = 0.05, and index parameters n_m = 20. The population size (100) was considered, and all operations were conducted for 500 generations. This solution resulted in 500 × 100 = 50,000 objective function evaluations. A PC with an Intel(R) Core(TM) i5-9400 CPU @ 2.90GHz 2.90 GHz and 16GB of memory was used for this work with a total running time of ≈ 1 hour. The results of the proposed options for placement of pressure sensors and the final placement scheme are shown in Figs. 4–6, respectively.

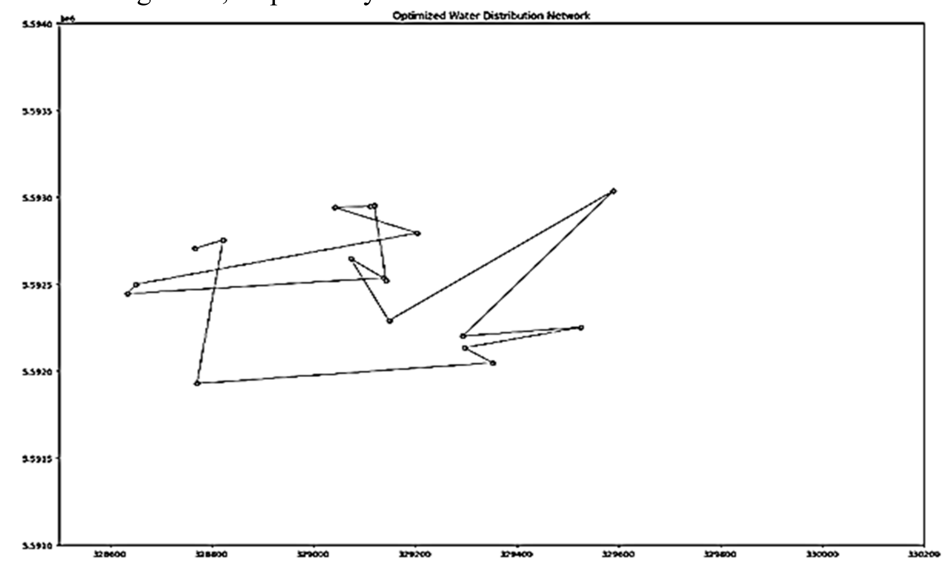

Fig. 4. Proposed variant 1 of the location of pressure sensors

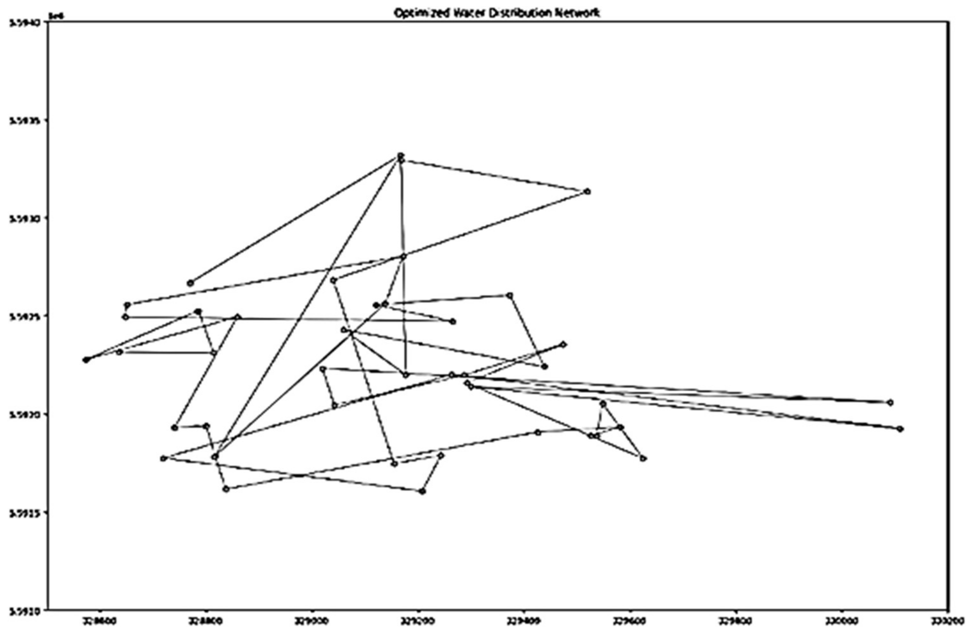

Fig. 5. Proposed variant 2 location of pressure sensors

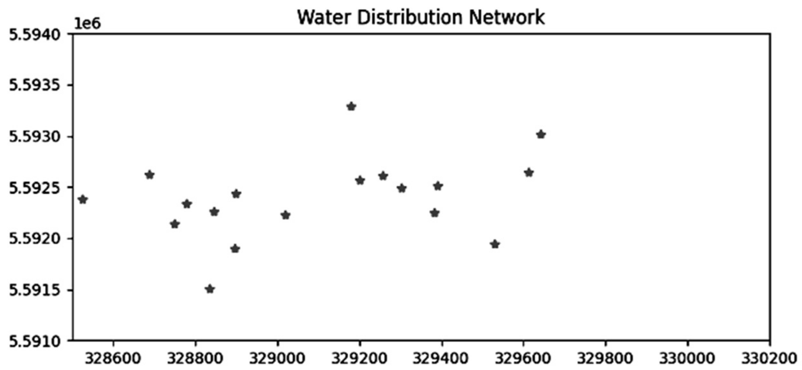

Fig. 6. The final proposed arrangement of pressure sensors

Figs. 4, 5 show that the calculated locations of the pressure monitoring sensors do not exhibit geometric regularity, as no geometric or spatial criteria were used to distribute the sensors in the network. However, despite the geometric irregularity, testing of the leak locations at these locations showed that pressure measurements at these nodes provided the most useful information for distinguishing between different leak scenarios. This can be explained informally: “Two teams of players cannot achieve similar results using different players”. There can be different sensor placements that provide high performance in leak detection. The next section describes the evaluation of the network condition using the results described above.

Assessment and forecasting of the state of the water network of the city of Kyiv

Adaptive neuro-fuzzy network (ANFIS). The main advantage of the adaptive neural network-based fuzzy logic inference system over other systems is that the parameters used for its membership function can be changed. The first 8 rules for each membership function are shown in Figs. 7, 8. Therefore, we can apply machine learning (ML) algorithms to train these parameters and build a model that fits the given data set.

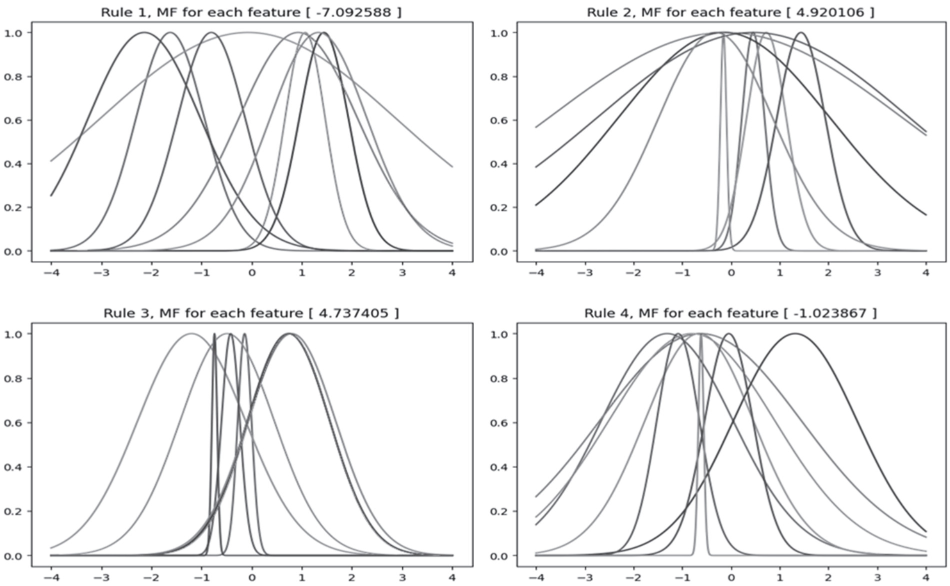

Fig. 7. Graphic representation of rules no. 1, no. 2, no. 3, and no. 4 for the membership function

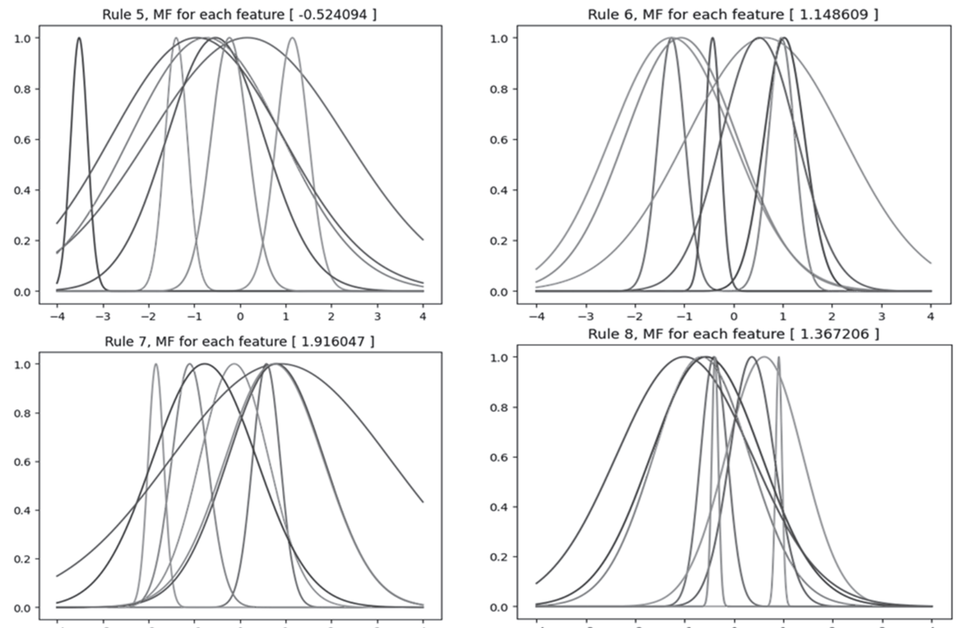

Fig. 8. Graphic representation of rules no. 5, no. 6, no. 7, and no. 8 for the membership function

The input parameters of ANFIS are: flow rate, pressure, material, diameter, material resistance, data on the height of the pipeline location, volume of water consumption, year of laying the pipe, type of soil. The data with the selected parameters were first transformed into the number of principal components and then loaded as input data into ANFIS using a Gaussian membership function of the input parameters. The performance results of ANFIS neural fuzzy network training on the test and training data sets are shown in Table 1.

Table 1. Assessment of the accuracy of the ANFIS model

No.		Test	Train
1.	**Epoch**	**200**	
	Loss	0.28776	0.170298
	Accuracy	0.89362	0.94715
	F1-score	0.87805	0.94222
	Precision	0.9	0.91379
	Recall	0.85714	0.97248
2.	**Epoch**	**400**	
	Loss	0.31423	0.13415
	Accuracy	0.914893	0.951219
	F1-score	0.90476	0.946903
	Precision	0.90476	0.91453
	Recall	0.90476	0.98165
3.	**Epoch**	**600**	
	Loss	0.34790	0.12462
	Accuracy	0.91489	0.95121
	F1-score	0.90476	0.94737
	Precision	0.90476	0.90756
	Recall	0.90476	0.99082
4.	**Epoch**	**800**	
	loss	0.37367	0.12151
	Accuracy	0.91489	0.95528
	F1-score	0.90476	0.95196
	Precision	0.90476	0.90833
	Recall	0.90476	1.0
5.	**Epoch**	**1000**	
	Loss	0.40200	0.11145
	Accuracy	0.91489	0.95528
	F1-score	0.90476	0.95154
	Precision	0.90476	0.91525
	Recall	0.90476	0.99082

For model training, 14 general fuzzy rules were set, and binary cross-entropy was used as the loss function. During training, we used the Adam optimizer with the parameter $\alpha = 0.01$, for 1000 epochs. Data points for plotting were taken from every tenth epoch. A Gaussian membership function was used for each fuzzy rule. Weights μ, σ were randomly initialized from a normal distribution. Graphical results of model performance indicators are shown in Fig. 9.

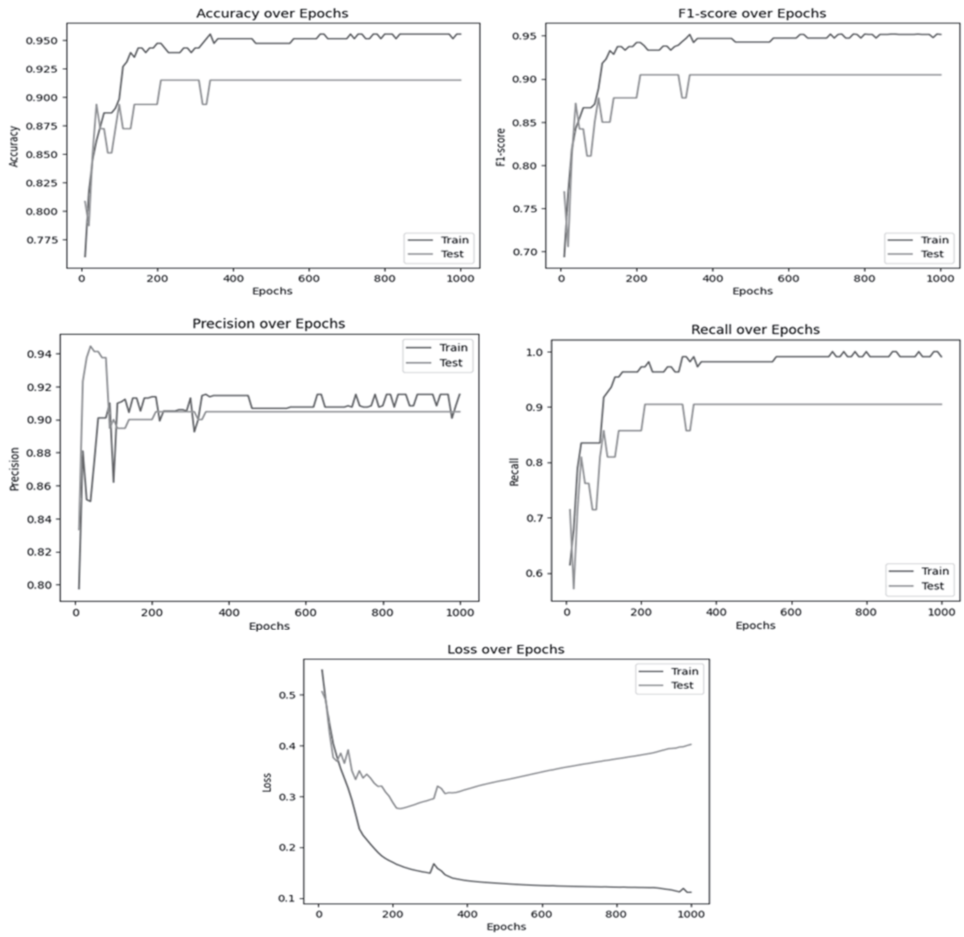

Fig. 9. Graphical visualization of ANFIS accuracy assessment

The obtained results indicate that the adaptive neuro-fuzzy model is successfully trained and works well.

The result of Mamdani's neuro-fuzzy logic inference method. A key feature of Mamdani's algorithm is its ability to adapt to transfer non-quantitative, "expert" knowledge into a quantitative scale that is useful for mapping a relatively complex water supply system. This algorithm makes it possible to integrate components that are not statistically independent into a single index value – this makes it possible to fully use all available data. An example of the transformation of data on the location of pipes in the area and data on network failures is shown in Figs. 10, 11.

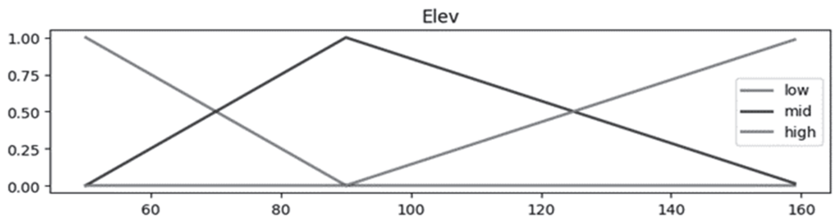

Fig. 10. Transformation of "expert" knowledge about the location of the network into a quantitative scale

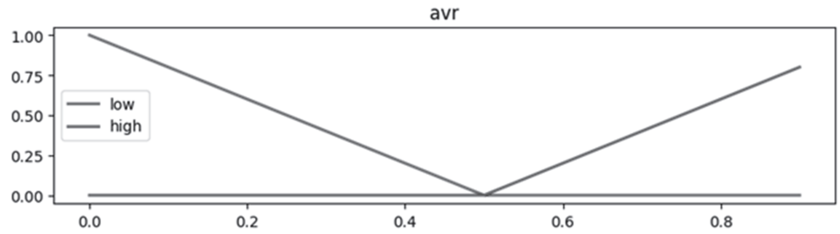

Fig. 11. Example of transmission of data about failures on the network in a quantitative scale

The Mamdani method consists of three steps:

Step1: fuzzification – input values are transformed into categorical variables (for example, "high", "low") using membership functions;

Step 2: fuzzy logical inference – a system of rules is defined that defines an aggregated function based on the categorical values attributed to an individual component;

Step 3: Defuzzification – the final value of the index is determined based on the geometry of the aggregated function. The accuracy results of the Mamdani algorithm are described in Table 2.

Table 2. Evaluation result of Mamdani's neuro-fuzzy logic inference algorithm

No.	Metrics	Result
1.	Accuracy	0.47773
2.	Roc-auc score	0.50365
3.	F1-score	0.27528

Hence, Mamdani's neuro-fuzzy logic inference algorithm is well-suited for translating non-quantitative, "expert" knowledge into a quantitative scale that is useful for mapping a relatively complex water supply system. Since large amounts of data on water distribution networks are non-quantitative or approximate.

The assessment of the state of the water supply network of the city of Kyiv was carried out under normal and abnormal conditions. Performance evaluation at nodes is important to determine the performance variation between different parts of the network, which requires the selection of critical regions (nodes) for further operation and maintenance. The main factors that lead to the deterioration of the performance of nodes are their location – if the nodes are located at a high altitude, they have low pressure and flow. If the nodes are far from the treatment station, they have a lower concentration of residual chlorine (other cleaning reagents). Thus, nodes with low productivity cannot provide sufficient quantity and quality of water for the user at a given time, which minimizes the provision of consumer needs. In addition, there is a need to classify nodes according to the level of performance.

CONCLUSIONS

In this study, a hybrid computing intelligent system is proposed for assessing the stability of the water distribution system and determining the optimal locations of pressure sensors. The results of the study showed that:

- the application of artificial intelligence methods to the field of water resources management indicates a great informational potential, which makes it possible to control water supply systems in real time, to automate and implement revolutionary new methods of analysis and forecasting of the state of engineering networks;

- fuzzy forecasting and network condition assessment models have a significant advantage as they require less information about water supply systems than conventional probabilistic models. In addition, this information may be vague and inaccurate;
- those nodes that received minimum pressure (less than 20 m) and maximum pressure (more than 50 m) during all simulation periods have low productivity and negatively affect the quantity and quality of water provided to the consumer. There are several options recommended for increasing productivity in critical areas (nodes). For example, replacing pipes, near nodes, laying parallel pipes, building new tanks for emergency sources of water and chlorine;
- in the failure state, the results of the reliability calculation show that the network can perform its function, that is, the ability to provide a sufficient amount of water at the desired pressure and water quality by 64.81 %, which has a sufficient level of performance. Similarly, the resilience analysis shows that the network has a 53.21 % probability of quickly meeting demand after an event of insufficient supply, low pressure or a quality that has a high-performance level. In addition, the research results show that the network has a 17.5 % susceptibility to failure, which is in the medium vulnerability range.

REFERENCES

1. X. Zhou et al., "Deep learning identifies accurate burst locations in water distribution networks," *Water Research*, vol. 166, 115058, 2019. doi: https://doi.org/10.1016/j.watres.2019.115058
2. E. Creaco, M. Blokker, S. Buchberger, "Models for generating household water demand pulses: Literature review and comparison," *Journal of Water Resources Planning and Management*, 143(6), 04017013, 2017. doi: https://doi.org/10.1061/(ASCE)WR.1943-5452.0000763
3. Omid Bozorg-Haddad, Mohammad Solgi, Hugo A. Loáiciga. *Meta-heuristic and Evolutionary Algorithms for Engineering Optimization.* John Wiley & Sons, Inc., 2017. doi: https://doi.org/10.1002/9781119387053
4. C.P. Liou, "Limitations and proper use of the Hazen–Williams equation," *Journal of Hydraulic Engineering*, 124(9), pp. 951–954, 1998. doi: https://doi.org/10.1061/(ASCE)0733-9429(1998)124:9(951)
5. Xudong Fan, Xijin Zhang, Xiong (Bill) Yu. "Machine learning model and strategy for fast and accurate detection of leaks in water supply network," *Journal of Infrastructure Preservation and Resilience*, vol. 2, article no. 10, 2021. doi: https://doi.org/10.1186/s43065-021-00021-6
6. A.E. Eiben, S.K. Smit, "Evolutionary algorithm parameters and methods to tune them," in *Y. Hamadi, E. Monfroy, F. Saubion (eds) Autonomous search*. Springer, Berlin, Heidelberg, 2011, pp. 15–36. doi: https://doi.org/10.1007/978-3-642-21434-9
7. N. Irfan, M.C.E. Yagoub, K. Hettak, "Genetic algorithm based on efficient tag detection in RFID reader networks," *2011 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings, Ottawa, ON, Canada, 2011*, pp. 1–4. doi: 10.1109/CIMSA.2011.6059921
8. D. Rani, S.K. Jain, D.K. Srivastava, M. Perumal, "Genetic algorithms and their applications to water resources systems," in *X.-S.Yang, A.H. Gandomi, S. Talatahari,*

A.H. Alavi (eds) Metaheuristics in water, geotechnical and transport engineering. Elsevier, 2013, pp. 43–78. doi: https://doi.org/10.1016/B978-0-12-398296-4.00003-9

9. A.R. Simpson, G.C. Dandy, L.J. Murphy, "Genetic algorithms compared to other techniques for pipe optimization," *Journal of Water Resources Planning and Management*, 120(4), pp. 423–443, 1994. doi: https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(423)
10. F. Di Pierro, S.-T. Khu, D. Savić, L. Berardi, "Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms," *Environmental Modelling & Software*, 24(2), pp. 202–213, 2009. doi: https://doi.org/10.1016/j.envsoft.2008.06.008
11. T.D. Prasad, N.S. Park, "Multiobjective genetic algorithms for design of water distribution networks," *Journal of Water Resources Planning and Management*, 130(1), pp. 73–82, 2004. doi: https://doi.org/10.1061/(ASCE)0733-9496(2004)130:1(73)
12. D. Kang, K. Lansey, "Revisiting optimal water-distribution system design: Issues and a heuristic hierarchical approach," *Journal of Water Resources Planning and Management*, 138(3), pp. 208–217, 2012. doi: https://doi.org/10.1061/(ASCE)WR.1943-5452.0000165
13. M. Cisty, Z. Bajtek, L. Celar, "A two-stage evolutionary optimization approach for an irrigation system design," *Journal of Hydroinformatics*, 19(1), pp. 115–122, 2017. doi: https://doi.org/10.2166/hydro.2016.032
14. B. Martínez-Bahena, M.A. Cruz-Chávez, E.Y. Ávila-Melgar, M.H. Cruz-Rosales, R. Rivera-Lopez, "Using a genetic algorithm with a mathematical programming solver to optimize a real water distribution system," *Water*, 10(10), 1318, 2018. doi: https://doi.org/10.3390/w10101318
15. W. Bi, G. Dandy, "Optimization of water distribution systems using online retrained metamodels," *Journal of Water Resources Planning and Management*, 140(11), 2014. doi: https://doi.org/10.1061/(ASCE)WR.1943-5452.0000419
16. S. Khalifeh, S. Akbarifard, V. Khalifeh, E. Zallaghi, "Optimization of water distribution of network systems using the Harris Hawks optimization algorithm (Case study: Homashahr city)," *MethodsX*, vol. 7, 100948, 2020. doi: https://doi.org/10.1016/j.mex.2020.100948
17. M.S. Khorshidi, M.R. Nikoo, N. Taravatrooy, M. Sadegh, M. Al-Wardy, G.A. Al-Rawas, "Pressure sensor placement in water distribution networks for leak detection using a hybrid information-entropy approach," *Information Science*, vol. 516, pp. 56–71, 2020. doi: https://doi.org/10.1016/j.ins.2019.12.043
18. N. Sho, A. Maki, O. Naoya, H. Jun, F. Naoki, "Monitoring of iodine species during water purification at a public water treatment plant in Japan," *Water Supply*, 19(2), pp. 580–587, 2019. doi: https://doi.org/10.2166/ws.2018.104
19. P. Ami, O. Avi, "Multiobjective Contaminant Sensor Network Design for Water Distribution Systems," *Journal of Water Resources Planning and Management*, 134(4), pp. 366–377, 2008. doi: https://doi.org/10.1061/(ASCE)0733-9496(2008)134:4(366)
20. Z. Hu, W. Chen, B. Chen, D. Tan, Y. Zhang, D. Shen, "Robust hierarchical sensor optimization placement method for leak detection in water distribution systems," *Water Resources Management*, vol. 35, pp. 3995–4008, 2021. doi: https://doi.org/10.1007/s11269-021-02922-3
21. Y. Huang, Y. Lan, S.J. Thomson, A. Fang, W.C. Hoffmann, R.E. Lacey, "Development of soft computing and applications in agricultural and biological

engineering," *Computers and Electronics in Agriculture*, 71(2), pp. 107–127, 2010. doi: https://doi.org/10.1016/j.compag.2010.01.001

22. E.H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Proceedings of the Institution of Electrical Engineers*, 121(12), 1974. doi: https://doi.org/10.1049/piee.1974.0328
23. L. Sela, S. Amin, "Robust sensor placement for pipeline monitoring: Mixed integer and greedy optimization," *Advanced Engineering Informatics*, vol. 36, pp. 55–63, 2018. doi: https://doi.org/10.1016/j.aei.2018.02.004
24. R. Pérez et al., "Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona [applications of control]," in *IEEE Control Systems Magazine,* vol. 34, no. 4, pp. 24–36, Aug. 2014. doi: 10.1109/MCS.2014.2320336
25. D. Kang, K. Lansey, "Optimal Meter Placement for Water Distribution System State Estimation," *Journal of Water Resources Planning and Management*, 136(3), pp. 337–347, 2010. doi: https://doi.org/10.1061/(ASCE)WR.1943-5452.0000037
26. I. Santos-Ruiz, F.R. López-Estrada, V. Puig, J. Blesa, "Estimation of Node Pressures in Water Distribution Networks by Gaussian Process Regression," *2019 4th Conference on Control and Fault Tolerant Systems (SysTol), Casablanca, Morocco, 2019*, pp. 50–55. doi: https://doi.org/10.1109/SYSTOL.2019.8864793
27. J. Mankad, B. Natarajan, B. Srinivasan, "Integrated approach for optimal sensor placement and state estimation: A case study on water distribution networks," *ISA Transactions*, vol. 123, pp. 272–285, 2021. doi: https://doi.org/10.1016/j.isatra.2021.06.004
28. S. Shin et al., "A systematic review of quantitative resilience measures for water infrastructure systems," *Water*, 10(2), 164, 2018. doi: https://doi.org/10.3390/w10020164
29. D. Xu, Z. Lv, H. Zeng, H. Bessaih, B. Sun, "Sensor placement optimization in the artificial lateral line using optimal weight analysis combining feature distance and variance evaluation," *ISA Transactions*, vol. 86, pp. 110–121, 2019. doi: https://doi.org/10.1016/j.isatra.2018.10.039
30. H. Dong, N. Hou, Z. Wang, W. Ren, "Variance-Constrained State Estimation for Complex Networks with Randomly Varying Topologies," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2757–2768, July 2018. doi: https://doi.org/10.1109/TNNLS.2017.2700331
31. H. Liu, H. Yu, "Decentralized state estimation for a large-scale spatially interconnected system," *ISA Transactions*, vol. 74, pp. 67–76, 2018. doi: https://doi.org/10.1016/j.isatra.2018.01.007
32. C. Ding, H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 03, no. 02, pp. 185–205, 2005. doi: https://doi.org/10.1142/S0219720005001004
33. L.A. Zadeh, "Fuzzy sets," *Information and Control*, 8(3), pp. 338–353, 1965. doi: https://doi.org/10.1016/S0019-9958(65)90241-X
34. R. Mirabbasi, S.M. Mazloumzadeh, M.B. Rahnama, "Evaluation of irrigation water quality using fuzzy logic," *Research Journal of Environmental Science*, 2(5), pp. 340–352, 2008. Available: https://lnk.ua/MHVvjKtE8
35. H. Gharibi, A.H. Mahvi, R. Nabizadeh, H. Arabalibeik, M. Yunesian, M.H. Sowlat, "A novel approach in water quality assessment based on fuzzy logic," *Journal of Environmental Management*, 112, pp. 87–95, 2012. doi: https://doi.org/10.1016/j.jenvman.2012.07.007

36. M. Vadiati, A. Asghari-Moghaddam, M. Nakhaei, J. Adamowski, A.H. Akbarzadeh, "A fuzzy-logic-based decision-making approach for identification of groundwater quality based on groundwater quality indices," *Journal of Environmental Management*, 184, pp. 255–270, 2016. doi: https://doi.org/10.1016/j.jenvman.2016.09.082
37. Soumaya Hajji et al., "Using a Mamdani Fuzzy Inference System Model (MFISM) for Ranking Groundwater Quality in an Agri-Environmental Context: Case of the Hammamet-Nabeul Shallow Aquifer (Tunisia)," *Water*, 13(18), 2507, 2021. doi: https://doi.org/10.3390/w13182507
38. A.S. Nasr, M. Rezaei, M.D. Barmaki, "Groundwater contamination analysis using Fuzzy Water Quality index (FWQI): Yazd province, Iran," *Geopersia*, 3, pp. 47–55, 2013. Available: https://www.researchgate.net/publication/243055680_Groundwater_contamination_analysis_using_Fuzzy_Water_Quality_index_FWQI_Yazd_province_Iran#full-text
39. Abu Rashid, Sangeeta Kumari, "Performance evaluation of ANN and ANFIS models for estimating velocity and pressure in water distribution networks," *Water Supply*, 23 (9), pp. 3925–3949, 2023. doi: https://doi.org/10.2166/ws.2023.224
40. L.A. Rossman, H. Woo, M. Tryby, F. Shang, R. Janke, T. Haxton, *EPANET 2.2 User Manual; Technical Report EPA/600/R-20/133*. U.S. Environmental Protection Agency: Washington, DC, USA, 2020.
41. L. Ferrandez-Gamot et al., "Leak localization in water distribution networks using pressure residuals and classifiers," *IFAC-PapersOnLine*, 48(21), pp. 220–225, 2015. doi: https://doi.org/10.1016/j.ifacol.2015.09.531
42. J. Blank, K. Deb, "Pymoo: Multi-Objective Optimization in Python," in *IEEE Access*, vol. 8, pp. 89497–89509, 2020. doi: https://doi.org/10.1109/ACCESS.2020.2990567

Received 30.11.2024

INFORMATION ON THE ARTICLE

Yuriy P. Zaychenko, ORCID: 0000-0001-9662-3269, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zaychenkoyuri@ukr.net

Tetiana V. Starovoit, ORCID: 0009-0008-6335-7679, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: starovoyt.tania@lll.kpi.ua

ГІБРИДНА ОБЧИСЛЮВАЛЬНА ІНТЕЛЕКТУАЛЬНА СИСТЕМА ДЛЯ ОЦІНЮВАННЯ СТАБІЛЬНОСТІ СИСТЕМИ РОЗПОДІЛУ ВОДИ ТА ВИЗНАЧЕННЯ ОПТИМАЛЬНОГО РОЗТАШУВАННЯ ДАТЧИКІВ ТИСКУ / Ю.П. Зайченко, Т.В. Старовойт

Анотація. Подано метод визначення найкращих місць розташування датчиків тиску в мережах водопостачання та оцінювання стану мережі за допомогою методів штучного інтелекту. Мета – визначення вузлів мережі, які нададуть найважливішу інформацію для виявлення витоків води та оцінювання загального стану мережі. Вибір місць розташування датчиків ґрунтувався на наборах даних про зміни тиску, спричинені різними сценаріями витоків, згенерованими моделюванням EPANET. Для ранжування вузлів-кандидатів та визначення оптимальної кількості місць розташування датчиків використано генетичні алгоритми. Наступний крок – оцінювання стану мережі за допомогою нейронечіткої мережі ANFIS та нейронечіткого алгоритму логічного висновку Mamdani. Алгоритми реалізовано в середовищі Google Colab та протестовано на ділянці мережі водопостачання в Києві, Україна.

Ключові слова: розміщення датчиків, WDN, штучний інтелект, гідравлічне моделювання, ANFIS, Mamdani, генетичні алгоритми, EPANET 2.2.