Civil Engineering Infrastructures Journal, 49(2): 311 – 326, December 2016
Print ISSN: 2322-2093; Online ISSN: 2423-6691
DOI: 10.7508/ceij.2016.02.008

Probabilistic Contaminant Source Identification in Water Distribution Infrastructure Systems

Barandouzi, M.[1] and Kerachian, R.[2*]

[1] Ph.D. Candidate, Department of Civil and Environmental Engineering, Virginia Tech, Falls Church, USA.
[2] Professor, School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran.

Received: 28 Mar. 2016; Revised: 07 Oct. 2016; Accepted: 09 Oct. 2016

ABSTRACT: Large water distribution systems can be highly vulnerable to penetration of contaminant factors caused by different means including deliberate contamination injections. As contaminants quickly spread into a water distribution network, rapid characterization of the pollution source has a high measure of importance for early warning assessment and disaster management. In this paper, a methodology based on Probabilistic Support Vector Machines (PSVMs) is proposed for identifying the contamination source location in drinking water distribution systems. To obtain the required data for training the PSVMs, several computer simulations have been performed over multiple combinations of possible contamination source locations and initial mass injections for a conservative solute. Then the trained probabilistic SVMs have been effectively utilized to identify the upstream zones that are more likely to have the positive detection results. In addition, the results of this method were compared and contrasted with Bayesian Networks (BNs) and Probabilistic Neural Networks (PNNs). The efficiency and versatility of the proposed methodology were examined using the available data and information from water distribution network of the City of Arak in the western part of Iran.

Keywords: Bayesian Networks (BNs), Probabilistic Neural Networks (PNNs), Support Vector Machines (SVMs), Water Contamination, Water Distribution Infrastructure Systems.

INTRODUCTION

Public awareness has increased profoundly respecting water supply systems security after September 11, 2001, attacks. Historically monitoring drinking water quality has been focused on the water treatment plant or reservoirs and supply systems, even though the distribution system presents many security challenges (Murphy and Kirmeyer, 2005). As water distribution networks are extensively scattered spatially, they are intrinsically susceptible to contamination or injection of toxicants. As suggested by Ostfeld (2006), to reduce the risk of this hazard, the physical security of the system should be improved and the existing water quality monitoring and early warning systems should be enhanced.

* Corresponding author E-mail: kerachian@ut.ac.ir

The distribution network is one of the most vulnerable parts of a water supply system, and there are many opportunities to breach into the water network (Gupta and Sayyed, 2013). Deliberate injection of a contaminant into the water network can be easily accomplished by a small pump. Over the past decade, significant amount of research work have been conducted in the field of optimal design of sensor placement for water quality monitoring and water quality data analysis in water distribution systems (e.g., Kessler et al., 1998; Ostfeld and Salomons, 2004; Berry et al., 2005, 2006; Isovitsch and VanBriesen, 2008; Aral et al., 2010, Shen and McBean, 2010; Davis et al., 2013 and Rathi and Gupta, 2015). Storey et al. (2011) reviewed advances and new emerging technologies in online drinking water quality monitoring. They concluded that despite recent improvements, there is no universal monitoring method for contaminant detection and water quality monitoring; moreover, there is a need to have platforms that can benefit from automated meter readings and wireless technologies. Hart and Murray (2010) explored the state of the art sensor placement in water distribution infrastructure systems, surveying a broad range of strategies including sensor characteristics, methodologies in sensor placement objectives, optimization approaches and scalability, simulations quality and solution features. They concluded that there is not a robust method that can be used in a large-scale water distribution system by an end-user. In addition, they suggested that the quality of input data including water demands, population estimates, seasonal operational rules as well as inherited uncertainties in data could increase the robustness of sensor placement designs. Moreover, they advised for improvement of decision support systems, as selecting high-quality sensor locations are significantly dependent on decision-making strategies in the water industry sector. Rathi et al. (2015) built on the review work of Hart and Murray (2010) suggesting several methodologies for sensor placement in water distribution networks. Rathi and Gupta (2016) compared results of two methods for finding the optimal location of sensors considering two objectives namely demand coverage and detection likelihood using hydraulic analyses. They used Genetic Algorithm (GA) to optimize the sensors' places for small water distribution networks and suggested to use Simple method for large and complicated water distribution systems due to its computational efficiency.

Islam at al. (1997) proposed an inverse model for calculating the chlorine concentrations needed at the network sources for meeting a specific chlorine concentration at a particular node in a water distribution system with unsteady flow conditions. Shang et al. (2002) presented an input-output model for providing information about the relationships between water quality at input and output locations by tracking water parcels along their paths. Laird et al. (2005) developed a non-linear origin-tracking algorithm for solving the inverse problem of contamination source identification. Preis and Ostfeld (2006) proposed a methodology for contaminant source identification in water distribution systems using a hybrid trees-linear programming algorithm. Guan et al. (2006) proposed a simulation-optimization method to solve nonlinear contaminant source identification in a complex water distribution system. Di Cristo and Leopardi (2008) formulated a methodology for identifying the source location of an accidental contamination in a water distribution network. In their methodology, among all candidate nodes, the site of origin was identified, minimizing the differences between simulated and measured concentrations. Guidorzi et al. (2009) proposed a procedure for detecting the presence of a contaminant in a water distribution system and implementing actions to isolate and/or expel it rapidly. The process consists of two consecutive optimization processes, both of them performed off-line assuming a specific 24-hour water demand sequence in each network node, whereas the accidental/intentional injection of contaminant could occur in any node and at any hour of the day. De Sanctis et al. (2009) used Particle Backtracking Algorithm (PBA) to identify the possible sources of contamination. They suggested that this algorithm is more suitable for real-time applications and can identify the area of the network that contamination had been originated. Yusta et al. (2011) comprehensively studied various applications and new methodologies in security and risk assessment of critical infrastructures including water supply systems. They showed that among different infrastructure sectors, water supply systems only had received 13% attention in referenced employed methodologies. Zechman (2011) used an agent-based modeling framework, which combined different exposure and managerial scenarios to simulate contamination events in a mid-size water distribution network. It was concluded in this study that consumers' behavior could alter the overall impacts of contamination events through changes in the hydraulics of the water distribution system. Perelman and Ostfeld (2013) used Bayesian Networks (BN) to estimate the likelihood of contaminant location. They used this methodology for two water distribution systems. They grouped the nodes of the network into clusters and tried to identify clusters that are the sources of contamination. Che and Liu (2014) applied eight detection parameters on a pilot-scale system using two different test contaminants. They showed that the results vary based on the type of the contaminant as well as contaminant's concentration. Wang and Harrison (2014) coupled Markov Chain Monte Carlo (MCMC) methods with Support Vector Regression (SVR) to identify the spatial and temporal properties of a contaminant. Eliades et al. (2014) investigated a model-based approach to detect contamination events using chlorine measurements. They used Monte-Carlo simulations in parallel with a real system that can produce expected range of chlorine concentration at selected sensor locations. Then, they compared the sensory measurements with estimated ranges along with event logic rules. If the measurements fall outside the defined ranges, an alarm flag will rise notifying that a contamination event has occurred.

In this paper, a new methodology is proposed for identifying pollution source in contamination events in water distribution systems by utilizing Probabilistic Support Vector Machines (PSVM), Bayesian Belief Networks (BNs) and Probabilistic Neural Network (PNN). The inputs of these probabilistic simulation models are concentrations and temporal variations of the concentrations of a water pollutant at designated monitoring points in a water distribution system. To evaluate the applicability and accuracy of the methodology, it was applied to a water distribution system in City of Arak, Iran.

METHODOLOGY

Figure 1 illustrates the flowchart of the proposed methodology. As it is shown in this figure, at first, required data and information regarding the physical characteristics of the water distributions systems and water demands should be gathered. In the next step, an indicator of pollutant was selected. Then, an EPANET simulation model was calibrated for simulating the flow and water quality in the water distribution system. EPANET has the capability of providing the spatial and temporal variations of the water quality indicators in water distribution systems.

To calculate the statistical characteristics of the concentrations of water quality indicators at chosen nodes, a Monte Carlo analysis was performed considering the probability distribution functions of uncertain primary inputs of the EPANET, namely the pollutant injection zone and the mass of the injected pollutant. To facilitate the Monte Carlo analysis, EPANET software and EPANET-Toolkit were merged in Visual Basic environment. EPANET-Toolkit is a dynamic link library (DLL) of functions that allows customizing EPANET's computational engine for specific needs including executing EPANET software for water quality simulations. Then, PSVM, BNs, and PNNs models were trained using the results of the Monte Carlo analysis.

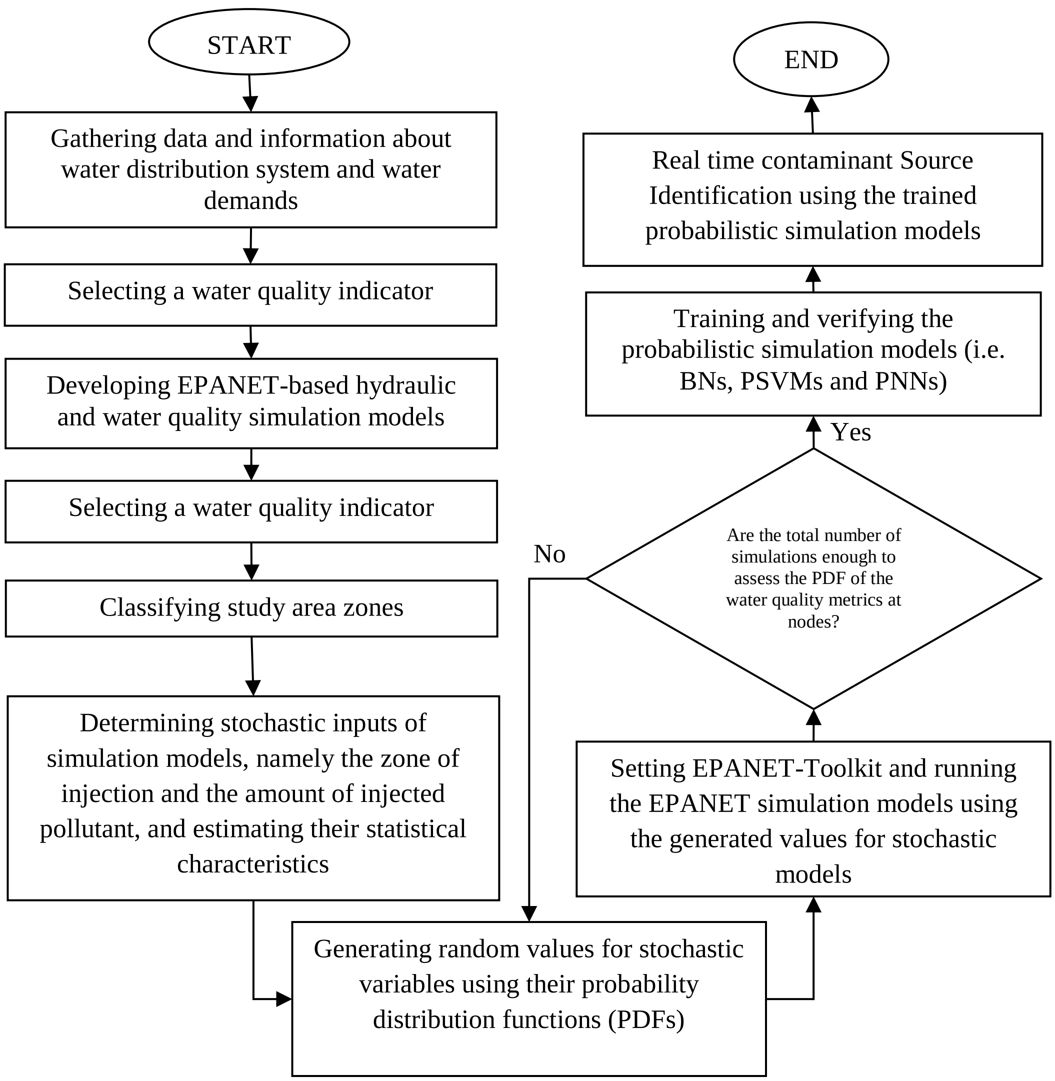

Fig. 1. Flowchart of the proposed methodology

The inputs of these probabilistic simulation models were the concentrations and the temporal variations of the concentrations of the water quality indicators at selected monitoring points in the water distribution system. In the proposed method, the goal is to identify the zone of contamination injection. In real-time water quality monitoring, the trained probabilistic simulation models can provide the probability of injected contamination in each zone based on the observed water quality data. The main parts of the proposed methodology are described in the following sections:

Support Vector Machines (SVMs)

Support vector machines (SVMs), which was introduced by Vapnik (1995), is a supervised pattern recognition method mostly used for classification and regression. SVMs have been used in different fields of water engineering such as in groundwater monitoring (Asefa et al., 2004; Bashi-Azghadi and Kerachian, 2010; Bashi-Azghadi et al., 2010), in wave height prediction (Malekmohamadi et al., 2011), in water quality zoning (Nikoo and Mahjouri, 2013), and in rainfall-runoff modeling (Hosseini and Mahjouri, 2016).

In data classification, SVMs separate dataset linearly into two distinct sets with constructing hyperplanes to maximize the margin between the two data sets. The best hyperplanes have the largest margin of support vectors (samples at the edges of the margins) and minimal empirical classification error (Figure 2).

Whenever the decision function is not a linear function of the data, kernels can be used for mapping out the data onto a higher-dimensional feature space where the data can be linearly separable. Several well-known kernels are linear, polynomial, radial basis function (RBF) and sigmoid.

Multiclass categorization problems are usually solved by reformulating the multiclass problem with M classes into a set of binary classification problems. A widely used method for multiclass categorization is one-against-one, which constructs $(k-1)/2$ classifiers each one trained on data from two classes. After all, $(k-1)/2$ classifiers were developed, and a voting approach was utilized. In max-wins voting strategy, each classifier assigns the instance to one of the two classes, and eventually the class with most votes determines the instance classification.

A probabilistic version of the SVM can be used to measure the prediction confidence. The probability of membership in class y, $y \in \{+1,-1\}$ is given by (Wu et al., 2004):

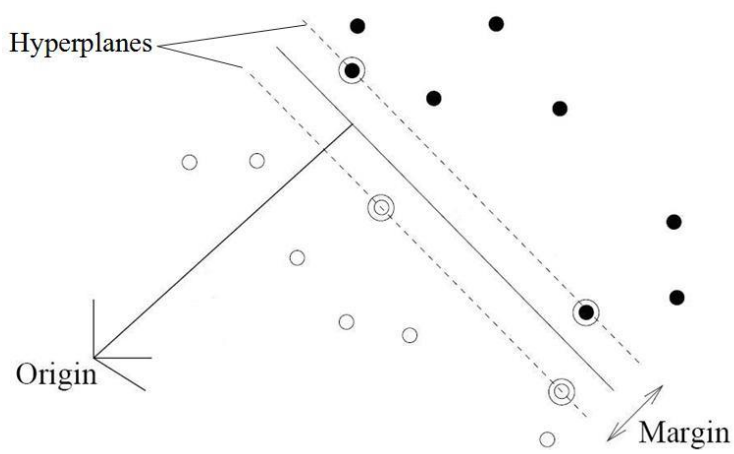

Fig. 2. Linear separating hyperplanes for the separable case. The support vectors are circled (Burges, 1998)

$$p(y=i|y=i \text{ or } j,x)=\frac{1}{1+e^{A\hat{f}+B}} \quad (1)$$

where A and B are estimated by minimizing the negative log-likelihood function using known training data and their decision values $\hat{f}$. As labels and decision values should be independent, the parameters A and B were fitted using cross-validation on the data sets as the classifier has been trained on. For more details, please see Chang and Lin (2001) and Wu et al. (2004).

Bayesian Networks (BNs)

Pearl (1988) proposed BNs to represent knowledge based on Bayes' theorem. A Bayesian network is a graphical representation of a probabilistic dependency model and consists of a set of interconnected nodes, where each node represents a variable, and the connecting arcs represent the causal relationships among variables.

By receiving a new evidence, the belief in the state of the evidence node changes, causing shifts in the belief of all nodes. In BNs, the belief in hypothesis h in response to evidence e is updated using the Bayes' theorem:

$$P(h|e)=\frac{P(e|h)\times P(h)}{P(e)} \quad (2)$$

The relationships among nodes and the conditional probability table are learned from a training data set. Learning BNs include two different processes: structure learning and parameter learning. Structure learning provides the best graph structures considering the relationships suggested by the training data. The conditional probability distributions are estimated in the parameter learning process. The values of parameters are usually defined by maximizing the likelihood of the training data (Buntine, 1996). A review of applications of BNs in water engineering can be found in Malekmohammadi et al. (2009), Mesbah et al. (2009), and Bashi-Azghadi -Azghadi et al. (2016).

Probabilistic Neural Networks (PNNs)

Artificial neural networks have received lots of attention over the past two decades. They have been used in the areas of prediction and classification, also in areas where regression models and other related statistical techniques have traditionally been used. PNNs are nonlinear, nonparametric pattern recognition modeling techniques that were originally introduced by Specht (1990). They train quickly and do not need a validation data set (i.e., wasted cases) to search for over-fitting. Therefore, all available data can be used for training of the model.

PNNs used in this paper have a 4-layer, feed-forward, one pass structure, which can classify data by estimating the probability density functions (PDFs) of the different classes. Unlike other ANNs, it is based on Bayes' decision strategy and non-parametric kernel based on estimators of probability density functions. Figure 3 shows the architecture of a typical PNN. The input layer unit simply distributes the input to the neurons in the pattern layer. On receiving a pattern x from the input layer, the neuron x_{ij} of the pattern layer computes its output (Hunter, 2000):

$$\phi_{ij}(x)=\frac{1}{(2\pi)^{d/2}\sigma^d}\exp\left[-\frac{(x-x_{ij})^T(x-x_{ij})}{2\sigma^2}\right] \quad (3)$$

where d: denotes the dimension of the pattern vector x, σ: is the smoothing parameter and x_{ij}: is the neuron vector.

Each neuron in the summation layer computes the maximum likelihood of pattern x being classified into class C_i by summarizing and averaging the output of all neurons that belong to the same class (Hunter, 2000):

$$P_i(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} \frac{1}{N_i} \sum_{j=1}^{N_i} \exp\left[-\frac{(x - x_{ij})^T (x - x_{ij})}{2\sigma^2} \right] \quad (4)$$

where N_i: denotes the total number of samples in class C_i. PNNs: are used for classification problems where the objective is to assign cases to one of a number of discrete classes

CASE STUDY

The efficiency of the proposed methodology is evaluated using available data from a real water distribution network. This network is a part of water distribution system of the City of Arak in the western part of Iran. The main characteristics of the network are shown in Table 1. The simulation model which is used in this paper has been borrowed from Water and Wastewater Company of the Markazi Province and has been calibrated by experts of that company for basic operating conditions considering pressure and flow rate variables.

In the first step, the network was divided into three zones, and four sensors were placed in critical nodes in a way that they can provide data regarding the quality of water in different parts of the network. Zone 1 is in an upstream of all four sensors, zone 2 is in an upstream of sensors 1 and 2, and zone 3 in is upstream of sensor 2. An effort was made to maximize the sensors' upstream coverage, by engineering judgment; however, no rigorous mathematical optimization was performed.

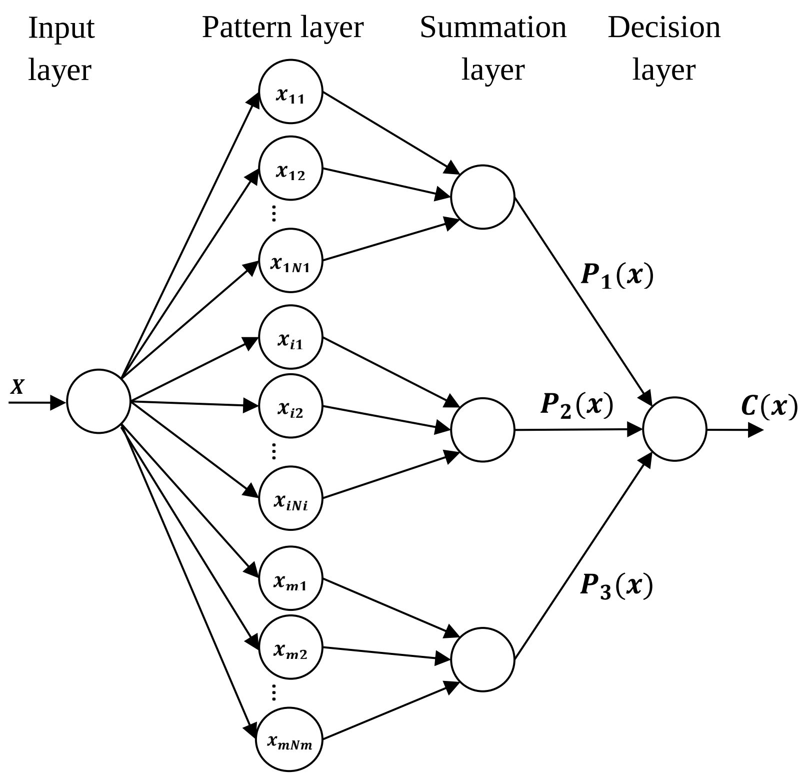

Fig. 3. A schematic illustration of a PNN (Adopted from Hunter, 2000)

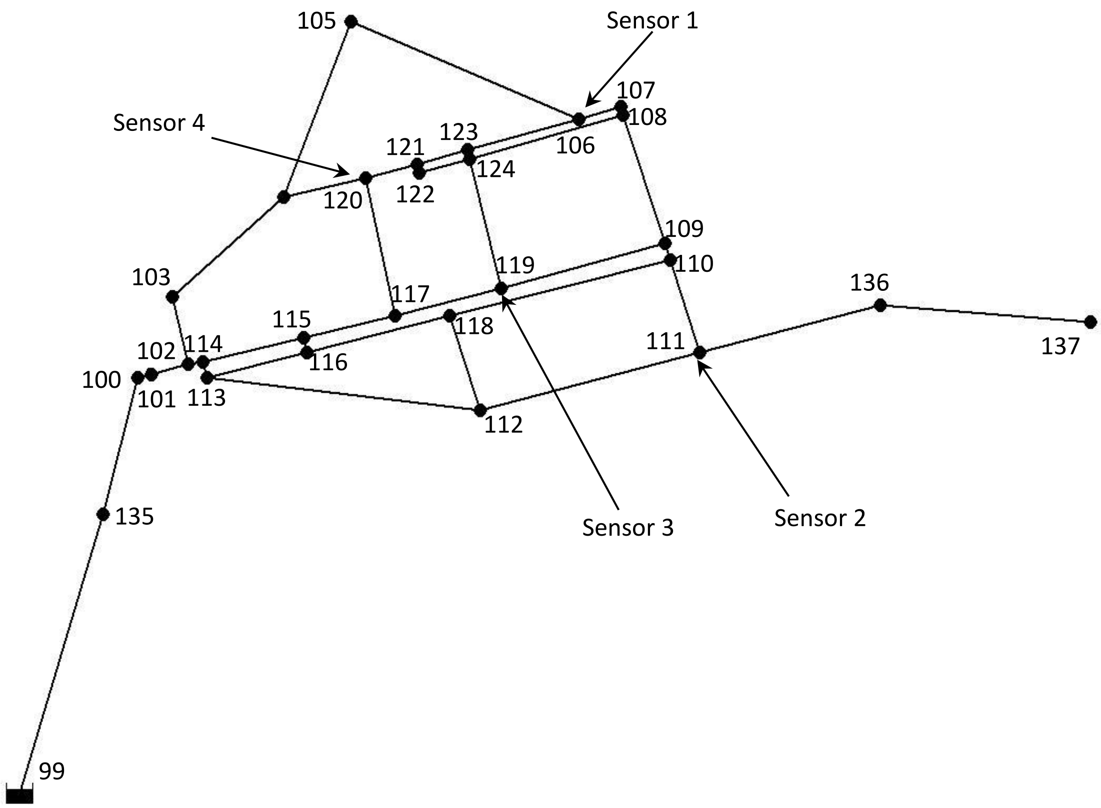

Fig. 4. Part of water distribution network of the City of Arak and the location of the water quality monitoring sensors

To simulate the flow and contaminant transport, the EPANET software (Rossman, 2000) was used. The Libsvm (Lin, 2008), Hugin® (Hugin Expert A/S, 2007) and PNN toolbox (MATLAB version R2007b) were used for developing the SVM, BN, and PNN-based probabilistic simulation models. EPANET in this setting serves for simulating contamination injection scenarios, and the trained probabilistic simulation models provide probabilistic estimations for the location of a pollutant injection using the water monitoring data.

RESULTS

The contaminant intrusion was modeled as a single injection of a pure conservative solute for 2 minutes using mass booster option in EPANET. To evaluate the efficiency of the proposed methodology, three distinct injection zones were considered and it was assumed that each contamination injection can occur in each node of any zone. The selected zones are shown in Figure 5. Each contamination injection in these three zones can be detected at least by one of the sensors. The pollutant transport in distribution network was simulated considering different contaminant injection locations in each zone. In these simulations, the injected mass was assumed to vary from 0 to 6000 grs and injection can take place at any time during a day. The upper limit of the injected mass reflects practical considerations for likely intentional intrusions. In this regard, the injection time was also assumed to be two minutes. The selected nodes for pollutant injection in each zone are presented in Table 2.

In each water quality simulation, the solute concentrations were measured by the sensors during a 60-minute simulation. This duration was chosen to make sure that the simulation would acquire the topmost concentration at each sensor. Each time step in simulating water quality was assumed to be one minute. Therefore, the sensors measure the quality of water in every minute during the simulation period.

Table 1. The main characteristics of pipes in the distribution network

Pipe Number	The First Node Number	The Second Node Number	Length (m)	Diameter (cm)	Roughness
100	101	102	40	150	105
200	101	102	40	350	115
101	102	103	71	100	105
201	102	103	71	200	115
102	102	114	13	150	105
502	102	114	13	250	115
103	113	114	19	300	115
104	113	115	105	300	115
105	115	116	19	100	115
106	114	116	105	150	105
107	115	118	155	400	105
108	112	118	101	150	115
109	110	118	233	400	105
110	110	111	103	100	105
510	110	111	103	150	115
111	109	110	19	100	105
511	109	110	19	150	115
112	109	119	176	150	105
113	108	109	130	100	105
213	108	109	130	150	115
114	107	108	11	100	105
214	107	108	11	150	115
115	106	107	46	100	105
215	106	107	46	150	115
116	105	106	462	100	115
117	104	105	200	150	115
118	103	104	256	200	115
119	104	120	86	100	115
120	117	120	141	100	105
121	119	124	137	100	105
122	122	124	53	300	110
123	106	123	118	100	105
124	120	121	57	100	105
125	121	123	53	100	105
126	108	124	162	300	110
127	111	112	235	100	115
128	112	113	368	100	115
140	100	135	166	350	120
241	99	135	330	350	120
142	123	124	11	100	115
143	117	119	111	150	105
144	116	117	81	150	105
145	136	137	263	100	115
146	111	136	187	100	115
406	121	122	11	100	105

Table 2. Location of injection nodes in each zone as well as the location of sensors

Sensor Location (Node)	Injection Nodes		
	Zone 1	Zone 2	Zone 3
106	102	108	113
111	104	109	118
119	117	119	-
120	120	122	-
-	-	124	-

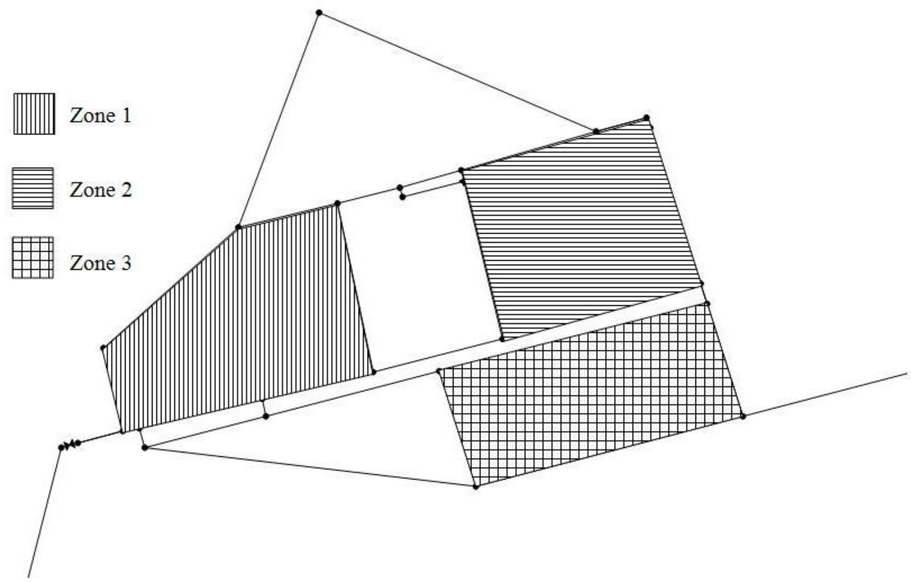

Fig. 5. The selected contaminant injection zones in the study area

To consider the real capability of the sensors, it was assumed that they can measure the concentrations more than 0.01 mg/L of the pollutant. For simulating the contamination injection scenarios, the EPANET and EPANET-Toolkit are linked in Visual Basic environment. The EPANET-Toolkit is a dynamic link library (DLL) that allows customizing EPANET's computational engine for specific needs including running EPANET for chemical and mass transport analyses. Considering the range of the injected pollutant mass and the injection nodes, 110 contamination scenarios are developed and simulated in this study.

The inputs of the probabilistic simulation models are the concentration and concentration gradient of the pollutant measured by the four sensors, and the outputs of the simulation models are the probability of membership of the pollution source in each zone. Therefore, each probabilistic simulation model has eight inputs and three outputs. The simulated scenarios provided 385 input-output data, which could be used for training and testing the probabilistic simulation models. 70% of the generated data were selected for training the models, and the trained models were tested using the rest of data. Most of the classifiers are known to be sensitive to the way features are scaled. As a result, it is essential to normalize either the data or the kernel itself. Normalization can be performed at the level of the input features or the level of the kernel (normalization in feature space) (Patki and Kelkarm, 2013). In this study, the input and output data was scaled to the range of [-1, 1].

In training phase of PSVM, Cross Validation (CV) via parallel grid search was performed for kernel and parameter selection. As shown in Table 3, Polynomial kernel function provides the best 10-fold cross-validation accuracy. Then, all grid points of (C; γ) were examined to identify the highest cross-validation accuracy. Table 4 presents sample results of n-fold cross validation for training PSVM with polynomial kernel function

As shown in Table 4, the best results were obtained using a 15-fold cross validation for a probabilistic SVM with $\gamma = 10^6$ and $C = 10^5$. Figure 6 shows the results of verified trained PSVM in estimating the zone of pollutant injection. As depicted in this figure, the trained PSVM can be effectively used for estimating pollutant injection zone.

The generated input-output data was also used for training a BN for identifying pollutant source location. Figure 7 shows the structure of developed BN. The probabilities presented in this figure, have been estimated using the training data set. As it can be seen in this figure, each input variables has been classified into 5 or 6 classes, and 3 classes have been considered for the output variable. Each class of the output variable is corresponding to a pollutant injection zone. In Figure 7, attributes number 1, 3, 5 and 7 shows the concentration of the pollutant measured by sensors 1 to 4, respectively. The gradients of the concentration of the pollutant in sensors 1 to 4 are also presented by attributes 2, 4, 6 and 8, respectively. Training of the BN was carried out using the Hugin® software based on the Estimation-Maximization (EM) method. Figure 8 shows the result of using the trained BN for a sample input data set. The verification accuracy of the train BN in estimating the location of an unknown contaminant injection is presented Table 5. The training data set also has been used for training the PNN. Table 6 presents the verification accuracy of the train PNN in estimating the location of an unknown contaminant injection. In Table 7 estimated location of an unknown contaminant injection using PSVM has been presented. Figure 9 shows a comparison of the accuracy of the probabilistic simulation models in the verification process. As shown in this figure, the PSVM can provide more reliable results in estimating the location of an unknown contaminant injection.

Table 3. Cross-validation accuracy obtained using PSVM with different kernel functions

Kernel Type	Linear	Polynomial	Tangent hyperbolic	Radial Basis Function (RBF)
Accuracy (%)	56	82	56	22

Table 4. Sample results of *n*-fold cross validation for training PSVM with polynomial kernel function

Parameters	Number of Folds in Cross-Validation	Deterministic Estimates		Probabilistic Estimates	
		Run Time (sec)	Accuracy (%)	Run Time (sec)	Accuracy (%)
$\gamma = 0.07$ $C = 1$	5	0.7	51.9	٣.١	55.9
	10	2.3	52.4	٣.5	55.2
	15	3.3	52.4	3.9	55.2
	20	3.9	52.4	5	55.2
$\gamma = 10^2$ $C = 10^5$	5	3.5	63.4	3.9	51.8
	10	3.6	65.9	4.9	61.5
	15	4.2	79.9	9.2	71.3
	20	6.2	82	13.5	79.9
$\gamma = 10^6$ $C = 10^5$	5	2.8	84.3	6.4	86.4
	10	3.2	87.1	7.2	87.8
	15	4.9	87.5	7.8	88.6
	20	5.2	87.5	12.3	87.1

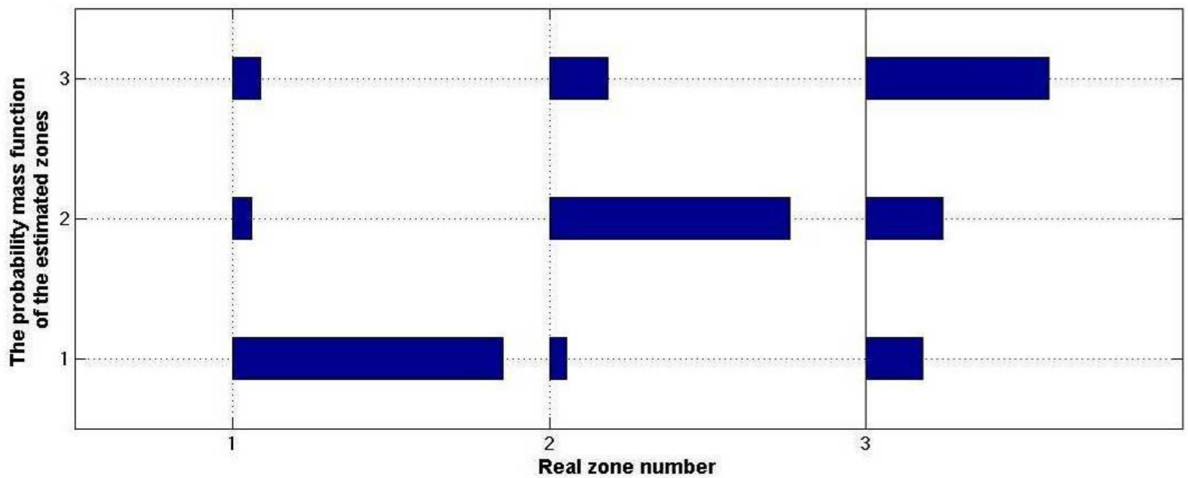

Fig. 6. The results of verified trained PSVM in estimating the zone of pollutant injection

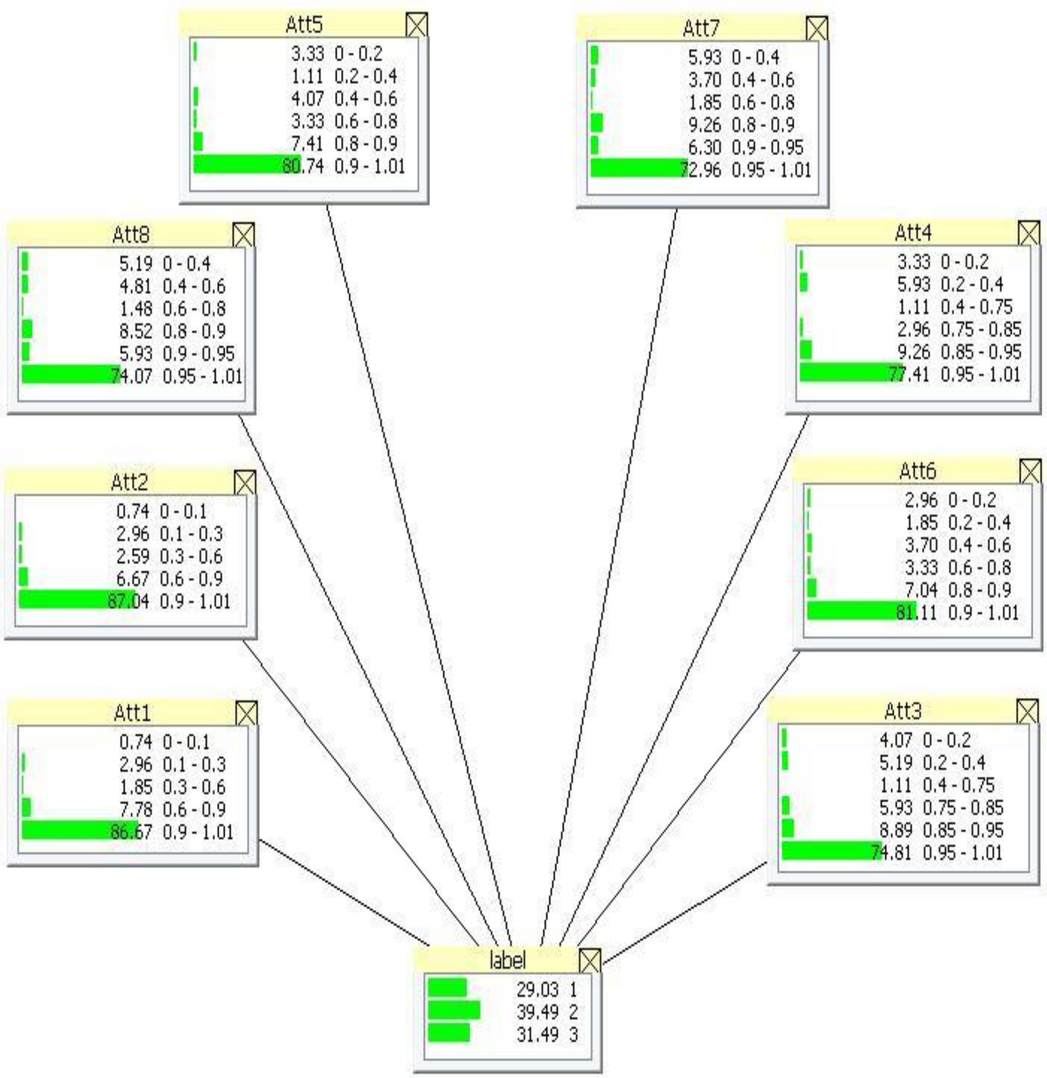

Fig. 7. The structure of the developed Bayesian Network for identifying pollutant source location in water distribution networks

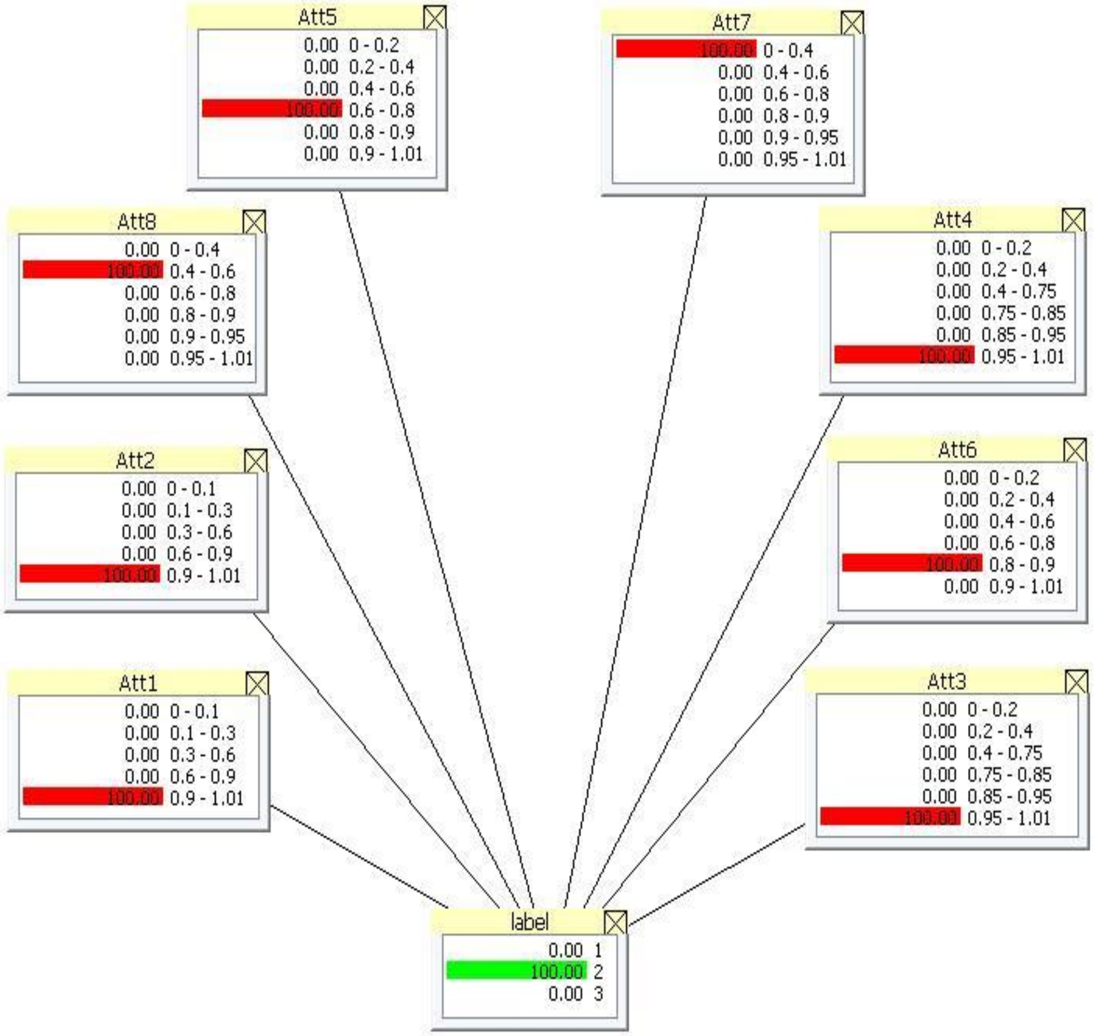

Fig. 8. The result of using the trained BN for a sample input data set. The BN has accurately identified the location of a contaminant injection (zone 2)

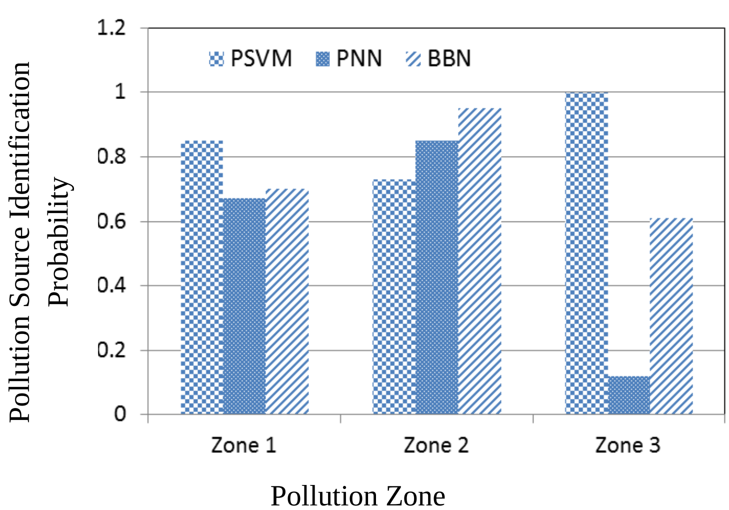

Fig. 9. Comparison the results of the probabilistic simulation models in the verification process

Table 5. Verification accuracy of trained BN in estimating the location of an unknown contaminant injection

Probabilistic Location of Contamination Injection in Percentage			Total
Zone 1	**Zone 2**	**Zone 3**	**Accuracy**
69.70	94.12	61.29	75.03

Table 6. Verification accuracy of trained PNN in estimating the location of an unknown contaminant injection

Probabilistic Location of Contamination Injection in Percentage			Total
Zone 1	**Zone 2**	**Zone 3**	**Accuracy**
66.67	86.27	12.90	55.28

Table 7. Verification accuracy of trained PSVM in estimating the location of an unknown contaminant injection

Probabilistic Location of Contamination Injection in Percentage			Total
Zone 1	**Zone 2**	**Zone 3**	**Accuracy**
84.85	72.55	100.00	83.48

CONCLUSIONS

After the events of September 11, 2001, intentional contamination intrusions to water distribution networks are considered as one of the most important menaces to public health. Online contamination monitoring can play an important role in improving water distribution infrastructures' security by providing early warnings for intentional contaminant injections. This paper presents a new methodology for contamination source identification in drinking water infrastructure distribution systems using three probabilistic simulation models, namely PSVM, BN and PNN. The efficiency and accuracy of the proposed methodologies were demonstrated using the existing data from the water distribution system of the City of Arak in Iran. The results showed that in this case study, the PSVM could provide more reliable results in estimating the location of an unknown contaminant injection.

ACKNOWLEDGMENTS

The technical contribution made by Dr. Babak N. Aaraabi, Associate Professor at the School of Electrical and Computer Engineering, University of Tehran, is hereby acknowledged.

REFERENCES

Aral, M. M., Guan, J. and Maslia, M. L. (2009). "Optimal design of sensor placement in water distribution networks", *Journal of Water Resources Planning and Management*, *136*(1), 5-18.

Asefa, T., Kemblowski, M. W., Urroz, G., McKee, M. and Khalil, A. (2004). "Support vectors–based groundwater head observation networks design", *Water Resources Research*, 40(11), W11509.

Bashi-Azghadi, S.N. and Kerachian, R. (2010). "Locating monitoring wells in groundwater systems using embedded optimization and simulation models", *Science of the Total Environment*, *408*(10), 2189-2198.

Bashi-Azghadi, S.N., Kerachian, R., Bazargan-Lari, M.R. and Solouki, K. (2010). "Characterizing an unknown pollution source in groundwater resources systems using PSVM and PNN", *Expert Systems with Applications*, 37(10), 7154-7161.

Bashi-Azghadi, S.N., Kerachian, R., Bazargan-Lari, M.R. and Nikoo, M.R. (2016). "Pollution source identification in groundwater systems: Application of Regret Theory and Bayesian Networks", *Iranian Journal of Science and Technology - Transaction of Civil Engineering*, 40(3), 241-249.

Berry, J.W., Fleischer, L., Hart, W.E., Phillips, C.A. and Watson, J.P. (2005). "Sensor placement in municipal water networks", *Journal of Water Resources Planning and Management, ASCE*, 131(3), 237-243.

Berry, J., Hart, W.E., Phillips, C.A., Uber, J.G. and Watson, J.P. (2006). "Sensor placement in municipal water networks with temporal integer programming models", *Journal of Water Resources Planning and Management*, *132*(4), 218-224.

Buntine, W. (1996). "A guide to the literature on learning graphical models", *IEEE Transactions on Knowledge and Data Engineering*, 8(2), 195-210.

Burges, C.J. (1998). "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, 2(2), 121-167.

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

Che, H. and Liu, S. (2014). "Contaminant detection using multiple conventional water quality sensors in an early warning system", *16th Conference on Water Distribution System Analysis, WDSA*, Bari, Italy.

Davis, M.J., Janke, R. and Phillips, C.A. (2013). "Robustness of designs for drinking water contamination warning systems under uncertain conditions", *Journal of Water Resources Planning and Management, 140*(10), 04014028.

Di Cristo, C. and Leopardi, A. (2008). "Pollution source identification of accidental contamination in water distribution networks", *Journal of Water Resources Planning and Management, ASCE*, 134(2), 197–202.

De Sanctis, A.E., Shang, F. and Uber, J.G. (2009). "Real-time identification of possible contamination sources using network backtracking methods", *Journal of Water Resources Planning and Management*, *136*(4), 444-453.

Eliades, D.G., Lambrou, T.P., Panayiotou, C.G. and Polycarpou, M.M. (2014). "Contamination event detection in water distribution systems using a model-based approach", *16th Conference on Water Distribution System Analysis, WDSA*, Bari, Italy.

Guan, J., Aral, M.M., Maslia, M.L., Maslia, and Grayman, W.M. (2006). "Identification of contaminant sources in water distribution systems using simulation–optimization method: Case study", *Journal of Water Resources Planning and Management*, 132(4), 252-262.

Guidorzi, M., Franchini, M., and Alvisi, S. (2009). "A multi-objective approach for detecting and responding to accidental and intentional contamination events in water distribution systems", *Urban Water Journal*, 6(2), 115-135.

Gupta, R. and Sayyed, M.A.H.A. (2013). "Predicting deficient condition performance of water distribution networks", *Civil Engineering Infrastructures Journal*, 46(2), 161-173.

Hart, W.E. and Murray, R. (2010). "Review of sensor placement strategies for contamination warning systems in drinking water distribution systems", *Journal of Water Resources Planning and Management*, 136(6), 611-619.

Hosseini, S.M. and Mahjouri, N. (2016). "Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling", *Applied Soft Computing*, 38, 329-345.

Hugin Expert A/S (2007). *Hugin® Software*, Aalborg, Denmark.

Islam, M.R., Chandhry, M.H., and Clark, R.M. (1997). "Inverse modelling of chlorine concentration in pipe networks under dynamic condition", *Journal of Environmental Engineering*, 123(10), 1033-1040.

Isovitsch, S. L., and VanBriesen, J. M. (2008). "Sensor placement and optimization criteria dependencies in a water distribution system", *Journal of Water Resources Planning and Management, ASCE*, 134(2), 186-196.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks", *Journal of Water Resources Planning and Management, ASCE*, 124(4), 192-198.

Laird, C.D., Biegler, L.T., van Bloemen Waanders, B.G. and Bartlett, R.A. (2005). "Contamination source determination for water networks", *Journal of Water Resources Planning and Management*, 131(2), 125-134.

Lin, C. (2008). "A library for Support Vector Machines-LIBSVM", (http://www.csie.ntu.edu.tw/~cjlin).

Malekmohamadi, I, Bazargan-Lari, M.R., Kerachian, R., Nikoo, M.R., (2011). "Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction", *Ocean Engineering*, 38(2), 487-497.

Malekmohammadi, B., Kerachian R., & Zahraie, B. (2009). "Developing monthly operating rules for a cascade system of reservoirs: Application of Bayesian Networks", *Environmental Modelling & Software*, 24(12), 1420-1432.

Mesbah, S.M., Kerachian, R. and Nikoo, M.R. (2009). "Developing real time operating rules for trading discharge permits in rivers: application of Bayesian Networks", *Environmental Modelling & Software*, 24(2), 238-246.

Murphy, B.M., and Kirmeyer, G.J. (2005). "Developing a phased distribution system, security enhancement program", *Journal of American Water Works Association.*, 97(7), 93-103.

Nikoo, M.R. and Mahjouri, N. (2013). "Water quality zoning using probabilistic support vector machines and self-organizing maps", *Water Resources Management*, 27 (7), 2577-2594.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security*", Journal of*

Water Resources Planning and Management, ASCE, 130(5), 377-385.

Patki, P.S., and Kelkar, V.V. (2013). “Classification using different normalization techniques in Support Vector Machine”, *International Conference on Communication Technology*, 17-19 November, Guilin, China.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*, Morgan Kaufmann Publishers, San Francisco, CA, USA.

Preis, A., and Ostfeld, A. (2006). “Contamination source identification in water systems: A hybrid model trees–linear programming scheme”, *Journal of Water Resources Planning and Management, ASCE*, 132(4), 263-273.

Rathi, S., and Gupta, R. (2015). “Optimal sensor locations for contamination detection in pressure-deficient water distribution networks using genetic algorithm”, *Urban Water Journal*, Oct., 1-13.

Rathi, S., Gupta, R., and Ormsbee, L. (2015). “A review of sensor placement objective metrics for contamination detection in water distribution networks”, *Water Science and Technology: Water Supply*, 15(5), 898-917.

Rathi, S., Gupta, R. (2016). “A simple sensor placement approach for regular monitoring and contamination detection in water distribution networks”, *KSCE Journal of Civil Engineering*, 20(2), 597-608.

Rossman, L.A. (2000). *EPANET2 user’s manual*, U.S. Environmental Protection Agency, Washington, D.C. (http://www.epa.gov/ORD/NRMRL/wswrd/epanet.html)

Shang, F., Uber, J.G. and Polycarpou, M.M. (2002). “Particle backtracking algorithm for water distribution system analysis”, *Journal of Environmental Engineering*, 128(5), 441-450.

Shen, H. and McBean, E. (2010). “Pareto optimality for sensor placements in a water distribution system”, *Journal of Water Resources Planning and Management*, *137*(3), 243-248.

Specht, D.F. (1990). “Probabilistic neural networks”, *Neural Networks*, 3(1), 110-118.

Storey, M.V., van der Gaag, B. and Burns, B.P. (2011). “Advances in on-line drinking water quality monitoring and early warning systems”, *Water Research*, 45(2), 741-747.

Vapnik, V. (1995). *The nature of statistical learning theory*, Springer-Verlag, New York, USA.

Wu, T.F., Lin, C.J. and Weng, R.C. (2004). “Probability estimates for multi-class classification by pairwise coupling”, *Journal of Machine Learning Research*, 5, 975-1005.

Yusta, J.M., Correa, G.J. and Lacal-Arantegui, R. (2011). “Methodologies and applications for critical infrastructure protection: State-of-the-art”, *Energy Policy*, 39(10), 6100-6119.

Zechman, E.M. (2011). “Agent-based modeling to simulate contamination events and evaluate threat management strategies in water distribution systems”, *Risk Analysis*, 31 (5), 758-772.