Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-24 (2018) 407–413

Comparison of Classifiers for Leak Location in Water Distribution Networks ⋆

Marcos Quiñones-Grueiro * **José M. Bernal-de Lázaro** *
Cristina Verde ** **Alberto Prieto-Moreno** *
Orestes Llanes-Santiago *

* *Department of Automation and Computing, CUJAE, Habana, Cuba (e-mail: marcosqg@automatica.cujae.edu.cu).*
** *Universidad Nacional Autónoma de México, Instituto de Ingeniería, México (e-mail: verde@unam.mx)*

Abstract: In this paper, the use of supervised classifiers for leak location in water distribution networks (WDN) is discussed. A comparative study is presented in the context of a benchmark network under the same leak and sensor placement scenarios. The comparison considers four classification tools widely used in the pattern recognition framework: Nearest Neighbor, Bayes Classifier, Artificial Neural Networks and Support Vector Machines. The classifiers' selection is made by considering their different working principles and application advantages. Training and testing sets are formed by the residuals generated by using the EPANET hydraulic simulator. The robustness of the methods is compared with respect to the leak location performance under model parameter uncertainty, demand uncertainty, leak size uncertainty and sensor noise. The SVM performs similar or better than the other classifiers when all uncertainties are present.

© 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Water distribution systems, Leak location, Supervised classifiers, Data-driven methods for FDI.

1. INTRODUCTION

Water scarcity has become a critical issue of the 21st century. Surface water resources are being depleted in some countries, while groundwater extraction is experiencing problems associated with quantity and quality (Mckenzie and Seago, 2005). According to the results of a World Bank survey (Kingdom et al., 2006), about 15 % of treated water is lost annually from the water distribution systems in developed countries, and an average of 35 % and even as high as 60 % of the treated water is lost in developing countries.

Water distribution networks (WDNs) are designed for delivering drinking water such that they play an important role in urban life. Networks' aging and other phenomena cause damage to the network's infrastructure and possibly cause water losses (Colombo and Karney, 2002). Leakages represent an economical, environmental and sustainability issue with consequences to the health and safety of the consumers (Pérez et al., 2014).

Water losses' detection and location can be tackled by using inference methods based on analytic or data-driven models of the network behavior. According to the IWA proposal, WDNs are subdivided into district metered areas (DMAs), such that a model is developed for characterizing the normal flow and pressure variations of each one (Loveday and Dixon, 2005).

Analytic approaches employ a model of the DMA based on the main physical laws describing it for simulating the normal operation scenarios. The core of this approach is the calibration of the model (Sanz et al., 2016). The difference between the synthetic and the metered data is used for diagnosing the system's state by using similarity measures. Although the identification of a hydraulic model is not an easy task, successful achievements have been presented for leak location in real DMAs (Pérez et al., 2014). The application of the analytic-based methods is however limited because of the bad influence of the model's uncertainties.

Data-driven approaches develop pattern recognition and statistic-based models by using historical data available from the network's behavior (Mounce et al., 2011; Arsene et al., 2012; Palau et al., 2012; Wu et al., 2016). This approach has proven to be useful for leak detection in real-life DMAs because it does not depend on detailed knowledge of the system and there is usually an abundance of normal data (Laucelli et al., 2016). The main limitation of data-driven methods is the development of leak location strategies because data from leakage scenarios are generally scarce (Wu and Liu, 2017).

A mixed analytic/data-based approach has been recently presented for improving the leakage location task (Soldevila et al., 2016, 2017). These works propose to combine a calibrated hydraulic model of the DMA with a classification method for leak location. A primary advantage of this approach is that data-based methods can handle uncertainty, while the hydraulic model can be

⋆ Paper supported by DGAPA-UNAM IT100716, , Proyecto 280170 conv 2016-3, Fondo Sectorial CONACyT-Secretaría de Energía-Hidrocarburos, II-UNAM and Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE).

2405-8963 © 2018, 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.2018.09.609

used for generating scenarios to train the classifier. A better performance can be achieved compared to classic methods such as the leak sensitivity matrix approach.

The aim of this paper is to assess the performance achieved by popular classification tools for the leak location task in DMAs under uncertain scenarios and steady-state conditions. The comparison's usefulness is that it allows determining the performance of different classifier(s) for the leak location problem. The compared classifiers are Nearest Neighbor (kNN), Bayes Classifier (BC), Artificial Neural Networks (ANN) and Support Vector Machines (SVM). These methods were selected by considering their different working principles and application advantages.

The content of this work is the following. In Section 2 the leak location methodology is described. The classification methods are presented in Section 3. Section 4 introduces the Hanoi network as the case study and the uncertain scenarios considered. The results and discussion are presented in Section 5. In Section 6 conclusions and directions for future work are drawn.

2. LEAK LOCATION SCHEME

The basic idea behind inference-based leak location is to make a decision about the DMA steady-state behavior by comparing the measurements with the model output signals. Thus, the leak location scheme follows the fault diagnosis principle presented in Fig. 1.

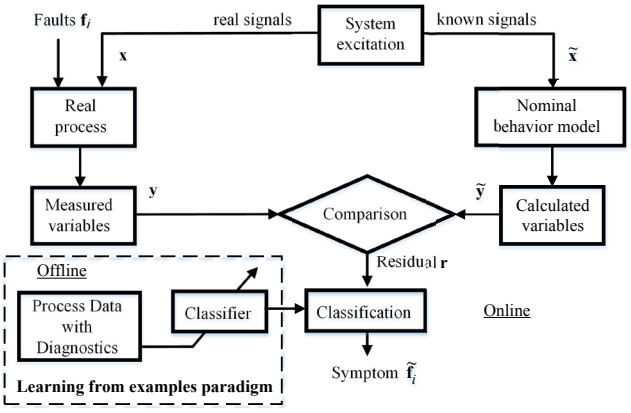

Fig. 1. General description of the fault diagnosis principle

In the context of WDNs' steady-state behavior, the system excitation is the consumers demand $\mathbf{x} = \mathbf{d} \in \Re^N$ at N nodes, and the measured variables are flows $\mathbf{q} \in \Re^{n_1}$ in n_1 pipes and pressure $\mathbf{h} \in \Re^{n_2}$ at n_2 nodes such that $\mathbf{y} = [\mathbf{q}, \mathbf{h}] \in \Re^{p=n_1+n_2}$. The estimated demands $\tilde{\mathbf{x}} = \tilde{\mathbf{d}} \in \Re^N$ are used for generating the calculated variables $\tilde{\mathbf{y}} \in \Re^p$ by using the nominal behavior hydraulic model. The faults considered $\Omega = \{\mathbf{f_1}, \mathbf{f_2}, ..., \mathbf{f_z}\}$ are single leaks occurring at z network nodes. The classification is the task of estimating the leak location $\tilde{\mathbf{f}}_\mathbf{i}$.

The classification process can be defined as the task of mapping the feature space ($\mathbf{r} \in \Re^p$) onto a set of z classes by using a decision function: $g(\mathbf{r})$: $\Re^p \rightarrow \Omega$ (Heijden et al., 2004). The decision function which guarantees the perfect separability among the classes is unknown. Thus, the *learning from examples* paradigm establishes that the parameters of $g(\mathbf{r})$ can be estimated by using some objects (samples) from the classes population (Heijden et al., 2004). If the classes to which the objects belong are known, the process of using data to determine the parameters of the decision function according to the classifier is known as supervised learning or training the classifier. The training process is performed offline, and the resulting classifier is used online for estimating the current fault.

The leak location process consists of estimating the node where the leak occurs by using the classification results (probable leaky node) within a Bayesian framework. In the context of classification tools, a temporal analysis is made by applying the Bayes rule to the probable leak locations throughout a time window of length L (Soldevila et al., 2017). Finally, the candidate location is the most probable after L samples.

3. CLASSIFICATION METHODS

This section briefly describes the selected classification tools for the comparative study.

3.1 k-Nearest Neighbor

The k-Nearest Neighbor is a pioneer nonparametric classifier which makes no assumptions regarding the statistical properties of the data set. The classification principle of kNN is simple: an observation $\mathbf{y}$ is assigned to the class with the maximum number of votes coming from the κ nearest samples. The decision function is suboptimal and can be represented as (Heijden et al., 2004)

$$g(\mathbf{r}) = f_{i(\kappa)} \; with \; \kappa = \max_{i=1,2,...,z} \{\kappa_i\} \tag{1}$$

where κ_i is the number of k-nearest neighbors that belong to class i and $i(\kappa)$ is the class with κ votes. The distance metric used for determining the neighbors is generally the Euclidean distance for continuous variables. There is essentially no training involved in the kNN classification process; it is considered a lazy learning algorithm. A serious drawback is the complexity of searching for the nearest neighbor(s) when the training set is large (Theodoridis and Koutroumbas, 2009). The probability associated with each class is estimated as $P_i = \kappa_i/\kappa_{max}$.

3.2 Bayes Classifier

The Bayes classifier is inspired in Bayes decision theory, and it is optimal with respect to minimizing the classification error probability (Theodoridis and Koutroumbas, 2009). The discriminant functions (minimum-error-rate classification) among the classes are the posterior probabilities for the Bayes classifier, i.e. $g_i(\mathbf{r}) = P(f_i|\mathbf{r})$. The multivariate Gaussian is the most frequently used (class-conditional) probability density function (Dougherty, 2013). In addition, when different covariance matrices are assumed for each class and *a priori* probabilities are considered equal, the decision function becomes

$$g(\mathbf{r}) = \max_{i=1,2,...,z} \{g_i(\mathbf{r})\} \tag{2}$$

$$g_i(\mathbf{r}) = -\frac{(\mathbf{r}-\mu_\mathbf{i})^{'}\Sigma_i^{-1}(\mathbf{r}-\mu_\mathbf{i})}{2} - \frac{ln|\Sigma_i|}{2} \quad (3)$$

where μ_i and Σ_i are the mean vector and covariance matrix for class i. The resulting discriminant functions $g_i(\mathbf{r})$ are inherently quadratic such that the Bayesian classifier is a quadratic classifier in these cases. The probability of each class is determined by using the probability density function for multivariate Gaussian processes (Dougherty, 2013). The training process consists of estimating the set of parameters $\{\mu_i, \Sigma_i\}$ for each class i. The Bayes classifier is optimal (minimum classification error) when the classes are Gaussian distributed and the distribution information is available (Venkatasubramanian et al., 2003).

3.3 Artificial Neural Networks

Artificial neural networks are a classifier inspired by the functioning of the brain and the human nervous system. ANNs provide a mechanism for pattern recognition by modeling the non-linear relationships among system's variables (Dougherty, 2013). The multilayer feedforward neural network is selected because it is commonly used in pattern recognition tasks (Patan, 2008). There are three types of layers according to their location: input, which receives information; output, where the results of the processing are given; and hidden, layers in between input and output. The central unit of each layer is the perceptron unit modeled by the McCulloch-Pitts equation (Patan, 2008):

$$\varphi^{(j)}(\mathbf{r}) = \phi^j(\sum_{i=1}^{p} \omega_{ij} r_i + b_j) \quad (4)$$

where j is the layer in which the perceptron is located, $\mathbf{r}$ is the input vector, b_j is a bias, ω_{ij} is a weight coefficient, and ϕ^j is a non-linear activation function. The resulting decision function can then be described by

$$g(\mathbf{r}) = \varphi^{j=(J)}(\varphi^{j=(\ldots)}(\varphi^{(j=2)}(\varphi^{(j=1)}(\mathbf{r})))) \quad (5)$$

where φ^j is the output function of layer j and J is the total number of layers neglecting the input. The training of a neural network is the determination of weight coefficient values between the neighboring processing units. The weights are adjusted according to the back-propagation error principle by using the scaled conjugate gradient method. A single hidden layer is selected given the Universal Approximation Theorem for ANNs (Dougherty, 2013). The soft-max transfer function is used for the output layer such that normalized output vectors sum to 1.0, and that can be interpreted as the class probabilities.

3.4 Support Vector Machines

Support vector machines are a classifier based on the structural risk minimization. The SVMs objective is to find the optimal separating hyperplane that maximizes the margin w between the closest sample points in the training data set, which are called support vectors (Scholkopf and Smola, 2002). Given a data set $X \in \{\mathbf{r_i}, \mathbf{y_i}\}^m$ with two classes where $\mathbf{r} \in \mathbb{R}^\mathrm{p}$ denotes the measured variables, m is the number of training examples, and the label vector $\mathbf{y} \in \{1, -1\}$, it is possible to determine the hyperplane $g(\mathbf{r})$ separating the two classes by

$$g(\mathbf{r}) = \mathbf{w}^T\mathbf{r} + b \quad (6)$$

where $\mathbf{w}$ and b (bias) are used to define the position of a separating hyperplane. To find the optimum values of w and b, solving the following dual optimization problem using the method of Lagrange multipliers $\mathbf{a} \in \Re^m$ is required.

$$\max\ W(\mathbf{a}) = \left(\sum_{i=1}^{m} a_i - \frac{1}{2}\sum_{i,j=0}^{m} a_i a_j g_i g_j\ K(\mathbf{r_i}, \mathbf{r_j})\right)$$
$$\text{subject to } \sum_i g_i a_i = 0\,; \quad 0 \le a_i \le C \quad (7)$$

where C represents the error penalty and $\mathrm{K}(\mathbf{r_i}, \mathbf{r_j})$ is a kernel function that allows access to spaces of higher dimensions without knowing the mapping function explicitly. In this work, the Radial Basis Function kernel shown in Eq. (8), was selected because of its generality and successful results in fault diagnosis applications (Bernal-de Lázaro et al., 2015; Zhang et al., 2016).

$$\mathrm{K}(\mathbf{r_i}, \mathbf{r_j}) = \exp\left(-\gamma\|\mathbf{r_i} - \mathbf{r_j}\|^2\right) \quad (8)$$

Here the term γ defines the geometric separation of the mapped data in the high dimensional space. The extension of the two-class to multi-class classification problems is performed by applying discriminant strategies. The one-against-one strategy is selected in this work based on previous comparisons (Hsu and Lin, 2002). Finally, the optimal posterior probabilities for the classes are estimated according to Platt's method (Platt, 1999).

4. CASE STUDY

The planned water distribution trunk network of Hanoi (Vietnam) presented by Fujiwara and Khang (1990) is used as a case study. The network topology is shown in Fig. 2. The gravity-fed network is formed by 32 nodes and 34 pipes. The design parameters (pipe diameters) were obtained from (Sedki and Ouazar, 2012), and the link lengths can be found in (Fujiwara and Khang, 1990).

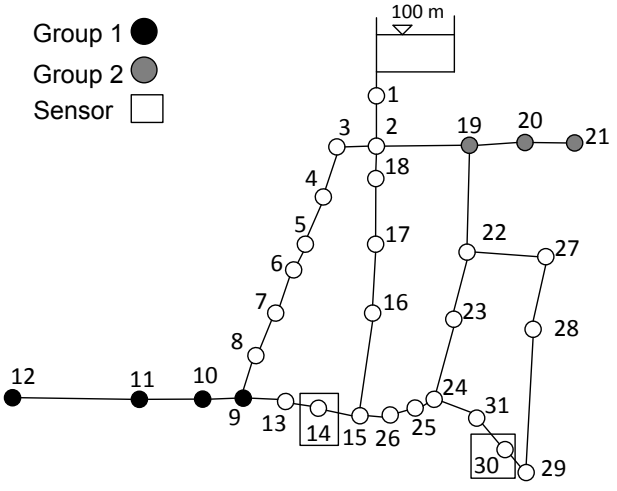

Fig. 2. Hanoi water distribution network

4.1 Model simulation and uncertain scenarios

The network behavior is simulated by using the package EPANET (Rossman, 2000) together with MATLAB. The

Hazen-Williams equation is utilized for the calculation of the friction factor (Houghtalen et al., 2010). A sampling period of 10 minutes is used for simulating the network, but hourly average values of the variables are used for reducing the uncertainty effect (Soldevila et al., 2016). The pressure head sensors are assumed to be located at nodes 14 and 30 according to previous works (Soldevila et al., 2016, 2017).

The residuals obtained by comparing the model outputs with the system measurements are sensitive to the uncertain conditions of the real WDN. Therefore, the performance of the classifiers depends on the amount and the types of uncertainty. The parameter affected by the uncertainty is obtained by using the following equation:

$$\theta_r(t) = \theta_d(t) + \theta_u(t) \tag{9}$$

where $\theta_u(t)$ represents the uncertainty; and $\theta_r(t)$ and $\theta_d(t)$ are the uncertain and the deterministic value, respectively. All uncertainty values $\theta_u(t)$ were generated from a uniformly sampled distribution. The robustness of the methods is assessed with respect to the following unknown disturbance conditions:

(1) Leak size uncertainty. Instead of generating a constant leak outflow, a realistic approach is to simulate a leak by setting an emitter coefficient C_e that will generate an outflow of magnitude f_i given by
$$f_i(t) = C_e h_i(t)^{\gamma} \tag{10}$$
where $h_i(t)$ is the pressure head at node i and $\gamma = 0.5$ (Rossman, 2000). The leak size uncertainty is associated with the emitter coefficient such that $C_e \in [10, 28]$. Thus, the leak outflow f_i ranges between 25 lps and 77 lps (0.45% and 1.39% of the maximum value of water demanded, which is 5539 lps (Fujiwara and Khang, 1990)).
(2) Measurements uncertainty. Measurements are corrupted such that the residuals are affected by noise with an amplitude within the range $[-5\%, 5\%]$.
(3) Pipe roughness uncertainty. The uncertainty of the model's parameter directly affects the simulation results of the WDN. The accurate identification of the pipe roughness in WDNs is complicated because it cannot be directly measured, and it decreases over time because of the pipe deterioration (Hutton et al., 2014). Thus, the Hazen-Williams coefficient (CHW) is simulated such that $CHW \in [125, 130]$.
(4) Estimated demand uncertainty. An uncertain demand is considered for each node $\tilde{\mathbf{d}} \in [-10\%, 10\%]$.
(5) All previous uncertainties are considered.

The residuals among the estimated variables and the measured variables (affected by the uncertainty of the real WDN) form each data set. Therefore, each residual vector contains information about the input flow and the two pressure measurements. Data from the network behavior is obtained by considering the given features and a single leak at a specific node. An emitter coefficient of 18 ($\simeq$ 50 lps) is considered with the exception of cases 1 and 5, as well as $CHW = 130$ excluding cases 3 and 5. A data set of 240 samples (hourly averages) is generated for each class and scenario by considering minimum night flow conditions such that node outflows represent approximately 30 % of the respective maximum demand.

Residuals at different nodes may cause the same leak signature such that they are practically indistinguishable. A nonparametric test which compares multidimensional data from two samples by using a measure based on statistical energy is employed for identifying overlapped residuals of different nodes (Soldevila et al., 2017). Thus, the nodes with a similar signature are clustered within the same class. As a result, two grouped classes are identified for the Hanoi network which are represented in Fig. 2.

5. RESULTS AND DISCUSSION

5.1 Performance and classifiers' parameter selection

The performance considered for the classifiers is estimated by analyzing the confusion among classes. Thus, a confusion matrix $A = [A(i, j)]$ can be formed with the aim of defining performance measures. The matrix is defined so that its element $A(i, j)$ is the number of samples with a class label i that are classified as class j (Theodoridis and Koutroumbas, 2009). From A, one can directly extract the overall accuracy and error. The overall accuracy (Ac) is the percentage of data that has been correctly classified. Given a problem of z classes, Ac can be computed from the confusion matrix according to $Ac = \frac{1}{m}\sum_{i=1}^{z} A(i, i)$ where m is the number of samples. Hence, the overall percentage error is $Err = 100 - Ac$.

The overall accuracy depends on the best fitting of the parameter(s) of each classification tool for the specific problem. Consequently, the parameters of each classifier should be adjusted for guaranteeing the maximum possible accuracy. The procedure for parameter estimation is based on a ten-fold cross validation (Dougherty, 2013). The classifier parameters are readjusted in an iterative fashion such that the best possible accuracy is obtained for the training set. Finally, the classifiers' overall accuracy is estimated by using the test data. The Friedman and Wilcoxon non-parametric hypothesis tests were used for estimating if the difference among the ten values of classifiers' accuracy is significant (Demsar, 2006).

The number of neighbors is an important parameter for the performance of the kNN classifier. Therefore, κ was selected for achieving the best test accuracy. The estimated value of κ is 5. In the neural networks' case, the training set was split. Part of the data (2/3) were used to optimize the layers' weights, and the other part (1/3) was used to decide the number of hidden neurons such that the maximum accuracy is achieved. The number of neurons for the hidden layer equals 26. For the SVM classifier, the parameters $\{C, \sigma\}$ were adjusted by using a grid-search for the interval $C \in 2^{\eta}$, $\eta \in [-2, 5]$ and $\gamma \in 2^{\eta}$, $\eta \in [-5, 3]$. The LIBSVM library was used for this purpose (Chang and Lin, 2011). The strategy of employing the test accuracy as a performance measure is used to prevent the classifiers from overfitting and the loss of generalization capacity (Dougherty, 2013).

5.2 Leak location results

The average overall validation and test accuracy of each classifier are shown in Table 1 for each scenario. The

classifiers with the best performance for each scenario are highlighted in bold letter, and this is determined by using the hypothesis tests with a 5 % significance level.

Table 1. Performance for the uncertain scenarios

	Ac_{train}/Ac_{test} [%]			
Uncertainty	kNN	BC	ANN	SVM
Leak size	98.8/97.3	98.8/97.3	99.7/**99.6**	99.26/99.1
Measurements	100/**100**	100/**100**	99.6/99.5	99.8/99.79
Pipe roughness	99.9/99.7	99.9/99.8	99.7/99.7	99.99/**99.98**
Estimated demand	81.2/71.7	77.1/**76.3**	77/**76**	77.19/**76.44**
Combined	70.8/57.22	64/**63**	63.3/**62.8**	64.6/**62.41**

The loss of generalization power is worse for the kNN classifier with a performance degradation between 10 % and 13 % for scenarios 4 and 5. The nearest neighbor is a lazy classifier as previously stated such that its generalization capability is limited because it suffers from the overfitting problem. The leak size uncertainty does not seem to have a significant negative impact on the classifiers' performance. Although the results seems to be similar from a practical point of view, the hypothesis tests reveal that the ANN performs better for the first scenario. The degradation of the overall test accuracy of the classifiers is not serious when the measurements are affected by uniformly distributed noise and pipe roughness uncertainty. The overall accuracy decreases when there is demand uncertainty for the fourth scenario. This result confirms the importance of the characterization of the demand for the leak detection and location problem as stated in previous works (Quiñones-Grueiro et al., 2017).

The overall test accuracy of the classifiers BC, ANN and SVM is similar for the fifth scenario according to the tests. The leak location results can be further improved by applying the Bayes rule over a time window. The leak location performance of the classifiers is assessed for $L \in [1\ 24]$ given that no significant improvement is accomplished for greater window sizes. The mean overall test accuracy results of applying the Bayesian approach for the fifth scenario are shown in Fig. 3. The fifth scenario was considered for the analysis because it resembles the real conditions of the network where all kind of uncertainties are present.

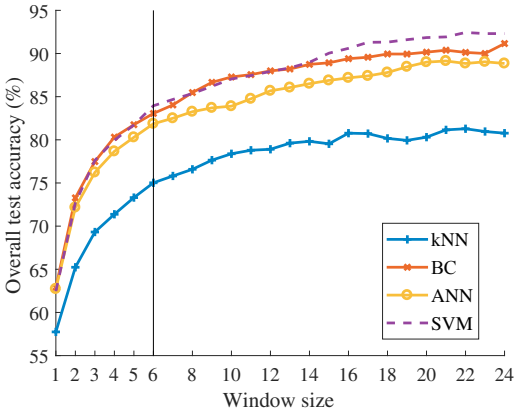

Fig. 3. Performance for different window sizes.

Hypothesis tests were applied for each window size for assessing significant differences in the median accuracy. Thus, the classifier with the best overall accuracy is the SVM since it performs best for cases 7, 9 and 10, and its performance was tied with the classifiers BC and ANN for the rest. In addition, the mean overall accuracy of SVM surpasses the BC for a window size greater than 15.

The performance achieved by using data of a single day is of interest for the early diagnosis of the network. The time window analyzed is 6 by considering that the low consumption period of the WDN can last up to six hours in the early morning. Among the classifiers' performance, the overall test accuracy of the SVM is slightly superior. The hypothesis tests, however, fail to reject the null hypothesis of significant differences in the median accuracy of the BC, ANN and SVM methods. Thus, the median performance of the three classifiers can be considered as similar. Consequently, it can be expected that by using BC, ANN or SVM for 83 % of the times that a leak appears, it can be correctly located at the node where it occurs within the same day.

Note that the BC achieves a similar performance for most scenarios compared to advanced classifiers such as the ANN and SVM. The performance of the BC depends mainly on the variables' distribution and the mixing of the data from different classes. The effect of pre-filtering the data approximates the data distribution to Gaussian given the central limit theorem: if a random variable is the outcome of a summation of a number of independent random variables, its probability distribution function approaches the Gaussian function as the number of summands tends to infinity (Theodoridis and Koutroumbas, 2009). Thus, the achieved performance for the BC could be considered as nearly optimal, and the results achieved by classifiers such as ANN and SVM are not quite superior.

The results obtained in this paper are conditioned to the case of study given the No Free Lunch Theorem (Dougherty, 2013). Therefore, the analysis developed should be further extended to networks with a different topology. The influence of the sensor placement on the leak location results is another element to be considered. Therefore, the relationship between the topology, number of sensors and sensor placement with respect to each classification method should be further analyzed. Overall, these three factors determine the number of classes and the data overlapping, as well as the variables statistical features.

The impact of the data-set size and the number of variables on the performance of the classifiers is should also be analyzed. In general, both factors are expected to grow for greater networks. It would also be interesting to assess the performance of the classifiers for a real WDN. Nonetheless, for greater networks:

(1) The performance of KNN will worsen with respect to the achieved results for the network topology analyzed in this paper. This fact is explained because of the increased uncertainty.
(2) Tuning the parameters for ANN and SVM will be harder because the number of classes (possible leaks) will increase.
(3) The performance of the BC will worsen if the data overlapping increases. Thus, the network topology is

an important element to take into account rather than the data-set size.

6. CONCLUSIONS

A comparative study of the use of classifiers for leak location in Water Distribution Networks was presented in this paper. The performance of four classification tools was assessed under different uncertain scenarios. It has been shown that the uncertainty associated with the estimated demand has the strongest impact on the degradation of the performance in the leak location task. From a practical point of view, there is not a significant performance difference among the classifiers for the uncertain scenarios associated with the leak size, measurements and pipe roughness. The methods BC, ANN and SVM achieve a similar performance from a statistical outlook. When the classification results are combined with the Bayes rule for different window sizes, there is a maximum improvement between 23 % to 30 %. Thus, the SVM performs similar or better than the BC and ANN for most window sizes analyzed in the scenario of all uncertainties.

ACKNOWLEDGEMENTS

The authors acknowledge the support provided by DGAPA-UNAM IT100716, Proyecto 280170 conv 2016-3, Fondo Sectorial CONACyT-Secretaría de Energía-Hidrocarburos, II-UNAM and the Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE).

REFERENCES

Arsene, C.T.C., Gabrys, B., and Al-dabass, D. (2012). Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection. *Expert Systems with Applications*, 39(18), 13214–13224.

Bernal-de Lázaro, J., Prieto-Moreno, A., Llanes-Santiago, O., and Silva-Neto, A. (2015). Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems. *Computers and Industrial Engineering*, 87, 140–149.

Chang, C.c. and Lin, C.j. (2011). LIBSVM : A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27:1—-27:27.

Colombo, A.F. and Karney, B.W. (2002). Energy and costs of leaky pipes: toward comprehensive picture. *Journal of Water Resource Planning and Management*, 128, 441–450.

Demsar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, 7, 1–30.

Dougherty, G. (2013). *Pattern Recognition and Classification*. Springer, New York, USA.

Fujiwara, O. and Khang, D.B. (1990). A Two-Phase Decomposition Method for Optimal Design of Looped Water Distribution Networks. *Water Resources Research*, 26(4), 539–549.

Heijden, F.V.D., Duin, R., Ridder, D.D., and Tax, D. (2004). *Classification , Parameter Estimation and State Estimation*. John Wiley & Sons, West Sussex, England.

Houghtalen, R.J., Akan, A.O., and Hwang, N.H.C. (2010). *Fundamentals of Hydraulic Engineering Systems*. Prentice Hall.

Hsu, C.w. and Lin, C.j. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.

Hutton, C.J., Kapelan, Z., Vamvakeridou-lyroudia, L., and Savic, D.A. (2014). Dealing with Uncertainty in Water Distribution System Models : A Framework for Real-Time Modeling and Data Assimilation. *Water Resources*, 140(February), 169–183.

Kingdom, W.D., Limberger, R., and Marin, P. (2006). The challenge of reducing NRW in developing countries.

Laucelli, D., Romano, M., Savic, D., and Giustolisi, O. (2016). Detecting anomalies in water distribution networks using EPR modelling paradigm. *Journal Of Hydroinformatics*, 18(3), 409–427.

Loveday, M. and Dixon, J. (2005). DMA sustainability in developing countries. In *Proceedings IWA Leakage Conference*. Halifax, Canada.

Mckenzie, R. and Seago, C. (2005). Assessment of real losses in potable water distribution systems: some recent developments. *Water Science and Technology*, 5(1), 33–40.

Mounce, S.R., Mounce, R.B., and Boxall, J.B. (2011). Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics*, 13(4), 672–686.

Palau, C., Arregui, F., and Carlos, M. (2012). Burst Detection in Water Networks Using Principal Component Analysis. *Journal of Water Resources Planning and Management*, 138(1), 47–54.

Patan, K. (2008). *Artificial Neural Networks for the Modelling and Fault Diagnosis of Technical Processes*. Springer-Verlag, Berlin, Germany.

Pérez, R., Sanz, G., Puig, V., Quevedo, J., Cugueró-Escofet, M.A., Nejjari, F., Meseguer, J., Cembrano, G., Mirats Tur, J.M., and Sarrate, R. (2014). Leak Localization in Water Networks. *IEEE Control Systems Magazine*, 34(4), 24–36.

Platt, J. (1999). *Probabilistic outputs for Support Vector Machines and comparisons to regularized likelihood methods*. MIT Press.

Quiñones-Grueiro, M., Verde, C., and Llanes-Santiago, O. (2017). Demand Model in Water Distribution Networks for Fault Detection. *IFAC-PapersOnLine*, 50(1), 3263–3268.

Rossman, L.A. (2000). *Water supply and water resources division. National Risk Management Research Laboratory. Epanet 2 User's Manual. Tech. Rep.* United States Envionmental Protection Agency. URL `http://www.epa.gov/nrmrl/wswrd/dw/epanet.html`.

Sanz, G., Pérez, R., Kapelan, Z., and Savic, D. (2016). Leak detection and localization through demand components calibration. *Journal of Water Resources Planning and Management*, 142(2), 1097–1098.

Scholkopf, B. and Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

Sedki, A. and Ouazar, D. (2012). Hybrid particle swarm optimization and differential evolution for optimal design of water distribution systems. *Advanced*

Engineering Informatics, 26(3), 582–591.

Soldevila, A., Blesa, J., Tornil-sin, S., Duviella, E., Fernandez-canti, R.M., and Puig, V. (2016). Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, 55, 162–173.

Soldevila, A., Fernandez-canti, R.M., Blesa, J., Tornil-sin, S., and Puig, V. (2017). Leak localization in water distribution networks using Bayesian classifiers. *Journal of Process Control*, 55, 1–9.

Theodoridis, S. and Koutroumbas, K. (2009). *Pattern Recognition*. Elsevier.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., and Yin, K. (2003). A review of process fault detection and diagnosis Part III: Process history based methods. *Computer and Chemical Engineering*, 27, 327–346.

Wu, Y. and Liu, S. (2017). A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal*, 14(9), 1–12.

Wu, Y., Liu, S., Wu, X., Liu, Y., and Guan, Y. (2016). Burst detection in district metering areas using a data driven clustering algorithm. *Water Research*, 100, 28–37.

Zhang, Q., Wu, Z.Y., Zhao, M., Qi, J., Huang, Y., and Zhao, H. (2016). Leakage Zone Identification in Large-Scale Water Distribution Systems Using Multiclass Support Vector Machines. *Journal of Water Resources Planning and Management*, 142(11), 1–15.