Online statistical hypothesis test for leak detection in water distribution networks

Radhia Fezai[a], Majdi Mansouri[b,*], Kamaleldin Abodayeh[c], Hazem Nounou[b], Mohamed Nounou[d], Vicenç Puig[e] and Kais Bouzrara[a]

[a]*Laboratory of Automatic Signal and Image Processing, National School of Engineers of Monastir, University of Monastir, Tunisia*

[b]*Electrical and Computer Engineering Program, Texas A&M University at Qatar, Qatar*

[c]*Department of Mathematical Sciences, Prince Sultan University, Riyadh, Saudi Arabia*

[d]*Chemical Engineering Program, Texas A&M University at Qatar, Qatar*

[e]*Department of Automatic Control, Polytechnic University of Catalonia, Barcelona, Spain*

Abstract. This paper aims at improving the operation of the water distribution networks (WDN) by developing a leak monitoring framework. To do that, an online statistical hypothesis test based on leak detection is proposed. The developed technique, the so-called exponentially weighted online reduced kernel generalized likelihood ratio test (EW-ORKGLRT), is addressed so that the modeling phase is performed using the reduced kernel principal component analysis (KPCA) model, which is capable of dealing with the higher computational cost. Then the computed model is fed to EW-ORKGLRT chart for leak detection purposes. The proposed approach extends the ORKGLRT method to the one that uses exponential weights for the residuals in the moving window. It might be able to further enhance leak detection performance by detecting small and moderate leaks. The developed method's main advantages are first dealing with the higher required computational time for detecting leaks and then updating the KPCA model according to the dynamic change of the process. The developed method's performance is evaluated and compared to the conventional techniques using simulated WDN data. The selected performance criteria are the excellent detection rate, false alarm rate, and CPU time.

Keywords: Leak detection, water distribution networks, kernel principal component analysis, online reduced kernel generalized likelihood ratio test, exponentially weighted moving average

1. Introduction

Health, safety, and environmental issues have gained significant importance world-wide. These issues are closely related to the availability and quality of water in many industrial and domestic applications. Water is considered as a unique commodity because nothing else can substitute for it, especially in areas exposed to drought weather conditions, such as in Qatar. Also, in most water distribution networks (WDN), it is estimated that between 10% to 30% of the water is lost in transportation from treatment plants to consumers. Water loss can be due to excessive physical activity, metering errors, leakage, pipe flushing, etc. The water losses can be categorised as "real losses" and "apparent losses." The apparent losses are constructed by errors in the measurements and measurements under-registration (e.g., consumption made by illegal connections). On the other hand, the real losses are the leakage in the WDN. Monitoring of leaks in WDN is, therefore, an important way to enhance

*Corresponding author. Majdi Mansouri, Electrical and Computer Engineering Program, Texas A&M University at Qatar, Qatar. E-mail: majdi.mansouri@qatar.tamu.edu.

the water management. Therefore, this paper evaluates the state of the network, i.e., to determine whether a leak exists or not. The idea behind leak detection in water distribution systems is to maintain disinfection levels, pressure and reduce water loss.

To address the fault detection purposes, several approaches have been developed, including principal component analysis (PCA) [1–3], nonlinear PCA (NPCA) [4], Multi-Regional PCA (MRPCA) [5], probabilistic PCA (PPCA) [6], attribute PCA (APCA) [7] and interval PCA (IPCA) [8]. Other detection indices, including the Hotelling T^2 statistic [9], the sum of squared residuals SPE and the generalized likelihood ratio test (GLRT) have been used. The GLRT has been proposed first in [10] for model identification and it has shown to be more favorable than the classical χ^2 test. Then, in [11], the GLRT chart has been used for fault detection (FD) based model. FD-based model [12] is performed into two steps: residual generation and residual evaluation. The GLRT chart is based on determining analytical relationships between measured variables to extract information about anomalies that may be caused by faults. However, it is difficult and even nearly impossible to determine an accurate and complete mathematical process model for complex and large-scale industrial processes. For instance, PCA based GLRT chart is used for fault detection when a process model is not available [13–15]. PCA-based GLRT aims to construct PCA model and then the GLRT chart is used for fault detection. The PCA-based GLRT technique depends on using linear PCA which is an input-space. However, most real processes are described by input-output models. To solve the issue, partial least squares (PLS) based GLRT technique is developed [15, 16]. This technique's main idea is to extract the covariation in both process and quality variables and model the relationship between them. Then, the process is monitored using the GLRT chart. However, PCA and PLS-based GLRT techniques assume that the process data are linear which makes them inappropriate for nonlinear industrial processes. To solve this problem, kernel scheme has been integrated to extend linear PCA and PLS techniques to their nonlinear versions (kernel PCA (KPCA) and kernel PLS (KPLS)) [17–19]. For the data to be linear, the kernel based technique consists of mapping the data into a higher-dimensional space. The choice of KPCA is supported by its distribution capabilities and its flexibility with a large variety of kernel functions. However, the classical KPCA technique is not appropriate for dealing with dynamic systems since it is a non-adaptive procedure. This technique requires to store the training data set and then use the KPCA model for detection purposes. For a large number of samples, the detection using KPCA method imposes a high computational cost. Then, KPCA is not adequate for real processes that are prone to drifting off due to various phenomena since the system may be subject to significant changes. This drift can change the relationship between process variables and can lead to poor detection results. To deal with changing process conditions and high computational cost, we propose to update the KPCA model and reduce the amount of training data. Thus, a reduced KPCA method is developed. This method aims to extract a new reduced data set, so-called dictionary, to build the reduced KPCA model. The resulting dictionary is formed by linearly independent kernel functions. The reduced method is applied in an online phase to construct the reduced model and it takes into account the dynamic behavior of the process by changing the model structure. Then, it uses the new KPCA model for detection purposes. Thus, the first objective of this paper is to develop an online reduced KPCA model that reduces the computation time and enhances the modeling accuracy of the classical KPCA model. The second objective is to enhance the detection accuracy of the classical GLRT. To do that, a new statistical leak detection chart based on merging the benefits of the exponentially weighted moving average (EWMA) filter with those of the GLRT is proposed. The proposed strategy, the so-called EW-GLRT, provides effective properties, such as smaller detection speed and smaller false alarm and missed detection rates. The main idea behind the proposed EW-GLRT is computing a new detection solution that considers the current and previous data information in a decreasing exponential fashion giving more importance to the more recent data.

Hence, in the paper, we develop a novel leak detection approach that uses online reduced KPCA to identify the model and then introduce the built model to EW-GLRT for detection purposes. The developed leak detection approach is the so-called exponentially weighted online reduced Kernel generalized likelihood ratio test (EW-ORKGLRT).

To summarize, this paper's main contribution is the improvement of the leak detection capability in water distribution systems by developing an effective online monitoring strategy called EW-ORKGLRT technique. The developed leak detection method will be able to detect different kinds and sizes of leaks.

This technique captures the normal, leak-free behavior and contrasts it with the new measurements to evaluate the network's state. The results demonstrate the effectiveness of the developed EW-ORKGLRT method over the ORKGLRT and the KGLRT methods for detecting leaks in different nodes. The leak monitoring efficiencies are assessed using false alarm rate (FAR), good detection rate (GDR), and CPU-time (CT).

The rest of the paper is organized as follows. In Section 2, a related work that surveys previous efforts in leak detection and monitoring is presented. Section 3 presents KPCA model-based detection technique. In Section 4, the improved leak detection approach using EW-ORKGLRT is developed. In Section 5, the leak detection performance is evaluated using simulated WDN data. In the end, the conclusions are presented in Section 6.

2. Related works

The main goal of this paper is to improve the operation of water distribution systems by developing a leak detection framework. Different solutions to analyze and model the state of the water distribution networks have been studied, where the analysis tool is usually trained (not necessary for all techniques) with data (real or artificial). Then, the state of the network is compared with the trained data and analyzed using these techniques.

Several techniques have been developed for leak detection and localization in WDN. In [20, 21], flow measurements or pressure are applied for classification in leak location problem. In [22], the evidence theory is used to locate the leak through Pipe Burst Model Prediction (PBMP). In [23], the authors proposed to detect and locate leaks in a benchmark network using Bayesian reasoning (BR). In [24], the authors developed a leak detection approach based on Support Vector Regressions (SVRs). The authors in [25], proposed a leak detection and localization methodology using model falsification. An ensemble of Change Detection Tests (CDTs) is implemented in [26] to detect the time the leak appears by considering that in healthy conditions, the flow measurements at the inlet of the WDN follow an unknown distribution. However, the distribution gets changes when the leak occurs. This approach uses an ensemble of CDTs to be robust against false alarms. In [27], a similar approach is applied but a hierarchy of CDTs is used instead of an ensemble. In [28], the periodicity of water consumption (the measurement of the flow at the inlet) is used to apply Self-Similarity (SS) between the recent measurements and a set of stored recorded data, which represents the current behavior of the WDN in a healthy state. In [29], an Artificial Immune System (AIS) network is used to detect and locate the leak. In [30], the leaks are considered to occur in the pipes, and the number of potential leaks is reduced using flow and pressure measurements (obtained from a hydraulic simulator) and spectral clustering. Then, SVM trained with the clustered data are used to classify the current measurements and perform the leak localization task.

In [31], all the measurements (flow and pressure) are used to develop a data model-based on the multicase Evolutionary Polynomial Regression (EPR) to predict the measurements and match them to the current measurements. The residuals are then compared with the thresholds to detect leaks. In [32], a database of normalized inlet flow patterns from different networks is used to retrieve the most similar pattern using k-Nearest Neighbors (k-NN) and to compare the degree of similarity with a threshold for detecting abnormal consumption. This approach is extended in [33] to exploit measurements from AMRs using Big Data and Cloud computing technologies. In [34], a leak detection and localization technique is considered in this article. The detection technique uses the inlet flow measurements at a high sampling rate and a Kalman Filter (KF) is applied to remove noise.

In [35], the repetitive flow pattern consumption is exploited to create a cyclic PCA technique (different PCA models are built depending on the different hourly water consumption pattern) using the pressure sensors inside the network, where the statistics T^2 and SPE are used and compared to detect leaks. In [36, 37], neuro-fuzzy classifiers are used to detect and locate leaks in a discretized network (i.e., the networks are reduced to discrete zones) using flow measurements. A similar approach is considered in [38] using pressure measurements instead. In [39], a fuzzy recognition system is applied to detect and locate the leak in the area that mismatched the expected behavior. In [40], the leak is detected using the information extracted from acoustic measurements and wavelets. Then PCA is applied to reduce the dimensionality. Finally, measurements are classified as a leak or not, using a Multi-Layer Perception (MLP) ANN (the ANN is trained beforehand with representative data measurements). In [41], an ANN with feed-forward MLP is used to build Gaussian Mixture Models (GMMs), which results in a network that provides the conditional probability density of the output data and is called a Mixture Density Network (MDN).

In [42], an MLP back propagation ANN is used hierarchically. First, the leak is detected, then the leak is located, and its size is estimated.

3. Kernel GLRT based on KPCA model

3.1. Description of KPCA model

The main idea of KPCA is to map the original data into a feature space using a nonlinear mapping function ϕ. Then a linear PCA is performed in feature space. For a given training data matrix $X = \left[x(1) \cdots x(N) \right]^T \in \mathbb{R}^{N\times m}$, where N is the number of samples and m is the number of measured variables. The mapped data matrix is defined as, $\mathcal{X} = \left[\phi(x(1)) \cdots \phi(x(N)) \right]^T \in \mathbb{R}^{N\times h}$, where h is a large value representing the dimension in feature space and the columns of$\mathcal{X}$ are assumed to be centered.

The covariance matrix is determined in the higher-dimensional space **F** by

$$\begin{aligned} Q &= \tfrac{1}{N-1}\mathcal{X}^T\mathcal{X} \\ &= \tfrac{1}{N-1}\textstyle\sum_{i=1}^{N} \phi(x(i))\phi^T(x(i)). \end{aligned} \tag{1}$$

The eigenvector problem for Q is expressed as,

$$Qv = \mu v. \tag{2}$$

where v and μ are the eigenvalue and eigenvector of Q, respectively.

To resolve Equation 2, it is required to compute the inner product of two kernel functions by introducing a kernel function of the form,

$$k(x(i), x(j)) = \phi^T(x(i))\phi(x(j)). \tag{3}$$

The commonly used kernel function is the radial basis kernel such that,

$$k(x, y) = exp(-\tfrac{(x-y)^T(x-y)}{c}), \tag{4}$$

where c is the kernel parameter.

Denoting $K = \mathcal{X}\mathcal{X}^T$ for the kernel matrix. According to Equation 1, Equation 2 can be expressed as,

$$(N-1)\mu v = \mathcal{X}^T\mathcal{X}v. \tag{5}$$

Hence, multiplying both sides of Equation 5 by $\mathcal{X}$ from the left, we get

$$\lambda\mathcal{X}v = \mathcal{X}\mathcal{X}^T\mathcal{X}v, \tag{6}$$

where $\lambda = (N-1)\mu$.

By using $K = \mathcal{X}\mathcal{X}^T$, the eigenvector problem is defined as,

$$\lambda\mathcal{X}v = K\mathcal{X}v. \tag{7}$$

Hence, if we define

$$\alpha = \mathcal{X}v, \tag{8}$$

then Equation 7 can be written in the form

$$\lambda\alpha = K\alpha. \tag{9}$$

According to Equation 9, α and λ are the eigenvector and eigenvalue of the kernel matrix K, respectively. Then, the eigenvector v is determined by multiplying with $\mathcal{X}^T$ from the left of both sides in Equation 8 and then use Equation 5

$$\mathcal{X}^T\alpha = \mathcal{X}^T\mathcal{X}v = \lambda v. \tag{10}$$

Thus, v is given by,

$$v = \lambda^{-1}\mathcal{X}^T\alpha. \tag{11}$$

Define $P_f = \left[v_1 \cdots v_\ell \right]$ to be the matrix of ℓ principal eigenvectors. From Equation 11, P_f can be expressed as,

$$\begin{aligned} P_f &= \left[\tfrac{1}{\lambda_1}\mathcal{X}^T\alpha_1 \; \tfrac{1}{\lambda_2}\mathcal{X}^T\alpha_2 \cdots \tfrac{1}{\lambda_\ell}\mathcal{X}^T\alpha_\ell \right] \\ &= \mathcal{X}^T P\Lambda^{-1}, \end{aligned} \tag{12}$$

where $\Lambda = diag(\lambda_1 \ldots \lambda_\ell)$ and $P = \left[\alpha_1 \cdots \alpha_\ell \right]$ are the ℓ largest eigenvalues and principal eigenvectors of the matrix K, respectively.

The number of kernel principal components (KPCs) ℓ that constitute the optimal KPCA model is determined using the cumulative percent variance (CPV) criterion. The CPV criterion is a measure of the percent variance captured by the first ℓ KPCs as per the following :

$$CPV(\ell) = 100\frac{\sum_{j=1}^{\ell}\lambda_j}{\sum_{j=1}^{N}\lambda_j}. \tag{13}$$

The CPV is considered to select the KPCs for which, say, over 95% of the cumulative variance is captured. The number ℓ satisfies the CPV criterion is defined as,

$$\ell = arg(CPV(\geq 95)). \tag{14}$$

For the test sample x and its mapped vector $\phi(x)$, its projections on the eigenvectors, also called the kernel components, are determined as,

$$t = P_f^T \phi(x). \tag{15}$$

According to Equation 12, the vector t can be written as

$$t = \Lambda^{-1} P \mathcal{X} \phi(x) = \Lambda^{-1} P k(x), \tag{16}$$

where $k(x) = \left[k(x_1, x)\ k(x_2, x)\ \ldots\ k(x_N, x) \right]^T$.

3.2. Description of kernel GLRT based detection chart

The exponentially weighted kernel GLRT (EW-KGLRT) detection chart combines the advantages of the exponentially weighted moving average (EWMA) and kernel GLRT charts. The EW-KGLRT chart aims to determine a new KGLRT statistic that integrates current and previous data information in a decreasing exponential mode by providing more weight to the more recent data. The KGLRT chart is one of the most used techniques for solving composite hypotheses testing problems by maximizing the likelihood ratio function overall faults [43]. The main idea begins with one N-dimensional vector $\phi(x)$, the hypothesis testing problem can be expressed as, [18]:

$$\begin{aligned} &H_0 : \phi(x) = \omega \text{ (null hypothesis)} \\ &H_1 : \phi(x) = P\theta + \omega \text{ (alternative hypothesis)}, \end{aligned} \tag{17}$$

where ω represents a white noise following a normal distribution $N(0, \sigma^2 I)$, P is an orthogonal matrix, $P^T P = I$, with I is an identity matrix, and θ is the mean vector (which is also the value of the fault).

For a new data x, the likelihood ratio test method chooses between H_0 and H_1 [18] as

$$\rho = \frac{f_1(\phi(x)|H_1)}{f_0(\phi(x)|H_0)} \lessgtr \gamma, \tag{18}$$

where γ is the threshold value of KGLRT statistic. $f_1(\phi(x) \mid H_0)$ and $f_1(\phi(x) \mid H_1)$ are conditional probability densities which obey Gaussian distributions [16]:

$$\begin{aligned} H_0 : f_0(\phi(x) \mid H_0) &\sim N(0, \sigma_0^2 I) \\ &= \frac{1}{(2\pi\sigma_0^2)^{\frac{N}{2}}} exp(-\frac{1}{2\sigma_0^2} \|\omega_0\|^2), \end{aligned} \tag{19}$$

$$\begin{aligned} H_1 : f_1(\phi(x) \mid H_1) &\sim N(P\theta, \sigma_1^2 I) \\ &= \frac{1}{(2\pi\sigma_1^2)^{\frac{N}{2}}} exp(-\frac{1}{2\sigma_1^2} \|\omega_1\|^2), \end{aligned} \tag{20}$$

where $w_0 = \phi(x)$ and $w_1 = \phi(x) - P\theta$. Here, the KGLRT chart changes the unknown parameter θ, σ_0 and σ_1 by its maximum likelihood estimate $\hat{\theta}$, $\hat{\sigma_0}$ and $\hat{\sigma_1}$. The maximum likelihood estimate of θ is identical to the least square estimate of ω_1 [44].

$$\begin{aligned} &\hat{\omega_0} = \phi(x) \\ &\hat{\omega_1} = \phi(x) - P_f \hat{\theta} = (I - C_f)\phi(x), \end{aligned} \tag{21}$$

where $C_f = P_f P_f^T$.

The estimate of $\hat{\sigma_0}$ and $\hat{\sigma_1}$ are given by,

$$\begin{aligned} &\hat{\sigma_0} = \frac{1}{N} \|\hat{\omega}_0\|^2 \\ &\hat{\sigma_1} = \frac{1}{N} \|\hat{\omega}_1\|^2. \end{aligned} \tag{22}$$

Substituting the maximum likelihood estimates of the parameters (Equation 22) into Equation 18 and taking $N/2$ root, KGLRT is defined as,

$$\begin{aligned} \mathbf{KG} &= \frac{\|\hat{\omega_0}\|^2}{\|\hat{\omega_1}\|^2} = \frac{\phi^T(x) I \phi(x)}{\phi^T(x)(I - C_f)\phi(x)} \\ &= \frac{\phi^T(x)\phi(x)}{\phi^T(x)\phi(x) - \phi^T(x) P_f P_f^T \phi(x)}. \end{aligned} \tag{23}$$

Then, from Equation 12, the KGLRT statistic can be computed as,

$$\begin{aligned} \mathbf{KG} &= \frac{1}{1 - \phi^T(x)\mathcal{X}^T P \Lambda^{-1} P^T \mathcal{X}^T \phi(x)} \\ &= \frac{1}{1 - k^T(x) P \Lambda^{-2} P^T k(x)}. \end{aligned} \tag{24}$$

To improve the performance of the classical KGLRT test, a new statistic that combines the advantages of the exponentially weighted filter with the KGLRT chart is proposed. The exponentially weighted-KGLRT statistic provides an effective and fast detection while conserving a low false alarm rate and higher good detection rate. The EW-KGLRT chart **EKG** can be computed as :

$$\mathbf{EKG} = \varpi \mathbf{KG} + (1 - \varpi)\mathbf{EKG}, \tag{25}$$

where ϖ define the smoothing parameter of the EWMA filter. To determine the threshold for the filtered KGLRT, its distribution should be determined. The new filtered KGLRT statistic follows a Chi-square distribution χ^2 since KGLRT distributed according to a Chi-square distribution [16]. Thus, the control limit for **EKG** statistic is obtained using the χ^2-distribution and it is given by [44]:

$$\mathbf{EKG}_\alpha = g_{\mathbf{EKG}} \chi^2_{h_{\mathbf{EKG}},\alpha}, \tag{26}$$

where $g_{\mathbf{EKG}} = \frac{b_{\mathbf{EKG}}}{2a_{\mathbf{EKG}}}$ and $h_{\mathbf{EKG}} = \frac{2a^2_{\mathbf{EKG}}}{b_{\mathbf{EKG}}}$, a and b are the mean and variance of the **KG**.

The **EKG** statistic suggests the existence of an abnormal situation in the data when

$$\mathbf{EKG}(x) > \mathbf{EKG}_\alpha. \tag{27}$$

4. Description of the developed leak detection technique

The developed leak detection approach is a method based on statistics and linear algebra techniques, used for data dimensionality reduction required in order to speed up and increase the performance of leak detection algorithms. The idea behind the developed technique is to evaluate the residuals obtained from the reduced KPCA model at each instant. Then, to make the decision if a leak is present or not, the EW-ORKGLRT statistic is compared to a threshold from the chi-square distribution. This technique aims to use a reduced KPCA to model the WDN system and at the same time to extract a reduced data set for online leak detection using EW-ORKGLRT statistic. Firstly, two dimension reduction (DR) metrics are used to extract only relevant samples in the feature space that are useful for analysis while eliminating redundant and unnecessary samples. An approximation criterion and the evaluation of the EW-ORKGLRT statistic are used as DR metrics to reduce the time and effort required to extract valuable information and enhance the processing speed. The two metrics transform the initial data set having high dimensionality and transform it into a new data set representing low dimensionality while preserving as much as possible the original meanings of the data. The low-dimensional representation of the initial data overcomes the dimensionality curse problem. Then, the low dimensional data can be easily, analyzed, processed, and visualized. Thus, by eliminating irrelevant and redundant features, EW-ORKGLRT statistic can become helpful for online leak detection. Secondly, an online leak detection scheme is adopted by using the EW-ORKGLRT statistic to enhance the performance of KPCA-based KGLRT. The evaluation of the EW-KGLRT statistic and the approximation criterion, determines whether the dictionary should be updated or remains unchanged for a new sample.

4.1. Dimension reduction (DR) metrics

The selection of a suitable dimension reduction metric according to the type of data is a big issue that needs to be considered in this paper. It is essential to extract the relevant and important samples from the initial data without affecting low-dimensional mapping performance. Another essential aspect to consider is to remove the leaky samples. The minor changes in the samples can affect the leak detection performance. In an online setting, determining the appropriate dictionary $\mathcal{D}_k = \left[\phi(x(w_1))\cdots\phi(x(w_r))\right]^T$, where $\phi(x(w_i))$ are the r selected kernel functions from the k kernel functions available so far, namely, $\{w_1, ..., w_r\} \subset \{1, ..., k\}$, at each instant is an important step before applying leak detection. Thus, in this paper, we evaluate first EW-ORKGLRT statistic at each instant to eliminate the leaky samples. Then, we use the linear approximation metric to remove irrelevant and redundant samples. The approximation criterion is based on building a dictionary with a high approximation measure. For online kernel principal component analysis, the approximation criterion operates as follows: the current kernel function is not added in the dictionary, if it can be sufficiently represented by a linear combination of kernel functions already belonging to the dictionary; otherwise, it is added in the dictionary. The norm of the residual approximation of the kernel function $\phi(x_k)$ by the r kernel functions is determined as,

$$\varepsilon_k = \|\phi(x_k) - P^r\phi(x_k)\|^2, \tag{28}$$

where P^r is the projection operator onto the subspace spanned by $\left[\phi(x(w_1))\cdots\phi(x(w_r))\right]^T$.

Then, ε_k can be computed by minimising the residual approximation as,

$$\varepsilon_k = min_\beta\|\phi(x_k) - \textstyle\sum_{j=1}^r \beta_j\phi(x(w_j)). \tag{29}$$

The optimal value of each coefficient β_j is determined by the minimization Equation 29, which leads to

$$\varepsilon_k = min_\beta \sum_{j,i=1}^{r} \beta_j \beta_i k(x(w_j), x(w_i)) \\ -2\sum_{j=1}^{r} \beta_j k(x(w_j), x(k)) + k(x(k), x(k)) \\ = min_\beta \beta^T K_k^r \beta - 2\beta^T k^r(x(k)) + k(x(k), x(k)), \tag{30}$$

where $k^r(x(k)) = [k(x(w_1), x(k)) \cdots (x(w_r), x(k))]^T$.

By solving Equation 30, the vector $\beta = \left[\beta_1 \cdots \beta_r\right]^T$ is given by,

$$\beta = (K_k^r)^{-1} k^r(x(k)), \tag{31}$$

where $K_k^r \in \mathbf{R}^{r\times r}$ is the reduced Gram matrix with elements $k(w_j), w_i))$. By inserting Equation 31 into Equation 30, we get the following expression of ε_k

$$\varepsilon_k = k(x(k), x(k)) - k^r(x(k))^T \beta, \tag{32}$$

Our approximation linear metric rule consists of including, at each time instant k, the kernel function $\phi(x(k))$ into dictionary $\mathcal{D}_k$ if

$$\varepsilon_k > \nu, \tag{33}$$

where ν is a positive threshold parameter that controls the level of sparseness. Thus, Equation 33 ensures the linear independence of the elements of the dictionary. The resulting dictionary, called ν-approximate, verifies the following relation

$$\min_{i=1\cdots r} \min_{\beta_1\cdots\beta_N} \|\phi(x(w_i)) - \sum_{\substack{j=1 \\ i\neq j}}^{r} \beta_j \phi(x(w_j))\| \geq \sqrt{\nu}. \tag{34}$$

Consequently, dimensionality reduction metrics offer an efficient way to reduce the number of samples before applying online leak detection.

4.2. Online leak detection based EW-ORKGLRT statistic

In this study, an online leak detection based EW-ORKGLRT statistic algorithm is derived. There are two possible cases in this algorithm: keep the dictionary unchanged or expand it with the new kernel function. A different dictionary will yield a different Gram matrix, eigenvector, and detection EW-ORKGLRT statistic. These parameters are updated only due to the dictionary change. Now, we consider the case when the sample $x(k+1)$ is leaky. It implies that EW-ORKGLRT statistic verifies this relation; $\mathbf{EKG}(k+1) \leq \mathbf{EKG}_{\alpha,k}$ and the sample is not included in the dictionary. Thus, the dictionary remains unchanged. For leaky data $(\mathbf{EKG}(x(k+1)) \leq \mathbf{EKG}_\alpha)$, the dictionary remains unchanged. However, for normal process change $(\mathbf{EKG}(x(k+1)) > \mathbf{EKG}_\alpha)$, two cases may arise :

In the first case, the dictionary is unchanged and the kernel function is discarded from the dictionary $\phi(x(k+1))$.

$$\mathcal{D}_{k+1} = \{\mathcal{D}_k\}. \tag{35}$$

The kernel function is not included to the dictionary if

$$\varepsilon_{k+1} = k(x(k+1), x(k+1)) - k^r(x(k+1))^T \beta < \nu. \tag{36}$$

This means that the kernel function $\phi(x(k+1))$ can be approximated by a linear combination of the model kernel functions. The vector $k^r(x(k+1))$ is updated as,

$$k^r(x(k+1)) = \left[k(x(w_1), x(k+1)) \cdots (x(w_r), x(k+1))\right]^T. \tag{37}$$

The vector $\beta = \left[\beta_1 \cdots \beta_r\right]^T$ with the dictionary unchanged can be updated as

$$\beta = (K_{k+1}^r)^{-1} k^r(x(k+1)), \tag{38}$$

where $K_{k+1}^r \in \mathbf{R}^{r\times r}$ is the reduced Gram matrix obtained from the dictionary D_{k+1} which is defined as

$$K_{k+1}^r = \begin{bmatrix} k(x(w_1), x(w_1)) & \cdots & k(x(w_r), x(w_1)) \\ \vdots & \ddots & \vdots \\ k(x(w_1), x(w_r)) & \cdots & k(x(w_r)), x(w_r)) \end{bmatrix}. \tag{39}$$

According to Equation 11, the eigenvector V_{k+1}^r corresponding to the dictionary $\mathcal{D}_{k+1}$, can be updated as,

$$v_{k+1}^r = \lambda_{k+1}^{-1} \sum_{i=1}^{r} \alpha_{k+1,i}^r \phi(x(w_i)). \tag{40}$$

For the second case when the new kernel function $\phi(x(k+1))$ is added into the dictionary and the size of the dictionary is increased by one to become $r+1$. The new dictionary $\mathcal{D}_{k+1}$ becomes,

$$\mathcal{D}_{k+1} = \{D_k, \phi(x(k+1))\}. \tag{41}$$

This case arises when the kernel function is significantly different from the previously selected elements of the dictionary. The dimensionality of the Gram matrix increases. Thus, the new Gram matrix $K_{k+1}^r \in$

$\mathbf{R^{r+1\times r+1}}$is updated as

$$K^r_{k+1} = \begin{bmatrix} K^r_k & k^r(x(k+1)) \\ & \\ k^r(x(k+1))^T & k(x(k+1), x(k+1)) \end{bmatrix}. \quad (42)$$

where
$k^r(x(k+1)) = [\, k(x(w_1), x(k+1)) \cdots k(x(w_{r+1}), x(k+1)) \,]^T$.

The inverse of the kernel matrix $(K^r_{k+1})^{-1}$ is calculated iteratively from the Woodbury matrix identity to avoid the problem of the higher computational complexity [45]:

$$\begin{aligned} {K^r_{k+1}}^{-1} &= \begin{bmatrix} (K^r_k)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \\ &+ \frac{1}{\varepsilon_{k+1}} \begin{bmatrix} -(K^r_k)^{-1} k^r(x(k+1)) \\ 1 \end{bmatrix} \begin{bmatrix} -k^r(x(k))^T (K^r_k)^{-1} & 1 \end{bmatrix} \\ &= \begin{bmatrix} (K^r_k)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\varepsilon_{k+1}} \begin{bmatrix} -\beta_k \\ 1 \end{bmatrix} \begin{bmatrix} -\beta^T_{k+1} & 1 \end{bmatrix}. \end{aligned} \quad (43)$$

The updating of the eigenvector V^r_{k+1} becomes

$$V^r_{k+1} = \lambda^{-1}_{k+1} \sum_{i=1}^{r+1} \alpha^r_{k+1,i} \phi(x(w_i)), \quad (44)$$

where $w_{r+1} = k+1$.

According the matrix updating rule in Equation 43, we have

$$\beta_{k+1} = (K^r_{k+1})^{-1} k^r(x(k+1)). \quad (45)$$

The parameters of the reduced KPCA model are updated and introduced to the ORKGLRT ($\mathbb{KG}$) statistic for online leak detection purpose,

$$\begin{aligned} &\mathbf{KG}(x(k+1)) \\ &= \frac{1}{1-\phi^T(x(k+1))(\mathcal{X}^r)^T P^r (\Lambda^r)^{-1} (P^r)^T (\mathcal{X}^r)^T \phi(x(k+1))} \\ &= \frac{1}{1-(k^r)^T(x(k+1)) P^r (\Lambda^r)^{-2} (P^r)^T k^r(x(k+1))}, \end{aligned} \quad (46)$$

where $\mathcal{X}^r = \mathcal{D}_{k+1}$, $\Lambda^r = diag(\lambda_{\ell_1} ... \lambda_{\ell_{r+1}})$, $P^r = \left[{\alpha^r}_{\ell_1} \; {\alpha^r}_{\ell_2} \cdots {\alpha^r}_{\ell_{r+1}} \right]$ and ℓ_{r+1} is the number of retained kernel principal components using the dictionary $\mathcal{D}_{k+1}$, and EW-ORKGLRT statistic (**EKG**) is defined as

$$\mathbf{EKG}(k+1) = \varpi \mathbf{KG}(k+1) + (1-\varpi)\mathbf{EKG}(k), \quad (47)$$

where, ϖ is the smoothing parameter between 0 and 1. ϖ the weight that defines the trade-off between the **KG** and $\mathbf{EKG}(k)$ indices. The control limit for EW-ORKGLRT statistic is updated as,

$$\mathbf{EKG}_{\alpha,k+1} = g_{\mathbf{EKG}_{k+1}} \chi^2_{h_{\mathbf{EKG}_{k+1}},\alpha}, \quad (48)$$

where $g_{\mathbf{EKG}_{k+1}} = \frac{b_{\mathbf{EKG}_{k+1}}}{2a_{\mathbf{EKG}_{k+1}}}$ and $h_{\mathbf{EKG}_{k+1}} = \frac{2a^2_{\mathbf{EKG}_{k+1}}}{b_{\mathbf{EKG}_{k+1}}}$, b and a are the variance and the mean of the EW-ORKGLRT index.

The computational complexity issue is extremely important for online leak detection. The proposed technique consists of three main parts : the dictionary selection, the update of the parameters of the model and the determination of leak detection statistic. As we can see from Equation 48 , the computational cost of evaluating EW-ORKGLRT depends on the update of the kernel vector. In this case, the update of the kernel vector depends only on the data in the dictionary whose size is less than the initial training data set ($r < N$), where N is the size of the initial training data set. As a result, using a reduced date set to update the EW-ORKGLRT statistic improves the computational speed. The proposed method not only reduces the computation time and memory usage as the datasize increases, but also enhances online leak detection by combining the benefits of ORKGLRT statistic and the EWMA chart. The use of the EW-ORKGLRT statistic can improve the online leak detection by reducing the false alarm rate and increasing the good detection by using EWMA filter. Indeed, the EW-ORKGLRT computes a new ORKGLRT detection statistic by taking into account the current and the previous data information by giving more weight to the more recent sample. This gives a more accurate estimation of the EW-ORKGLRT statistic and provides a stronger memory which will allow better decision making with respect to leak detection.

The online leak detection algorithm is illustrated schematically in Fig. 1.

5. Leak detection in water distribution networks

The EW-ORKGLRT is developed in order to improve the leak detection capabilities. The effectiveness of the proposed leak detection technique is assessed and compared to the ORKGLRT and KGLRT techniques in terms of three detection criteria:

1. False alarm rate (FAR) (%): percentage of wrong leak declared in leak free region,
2. Good detection rate (GDR) (%): percentage of leaky observations undetected,

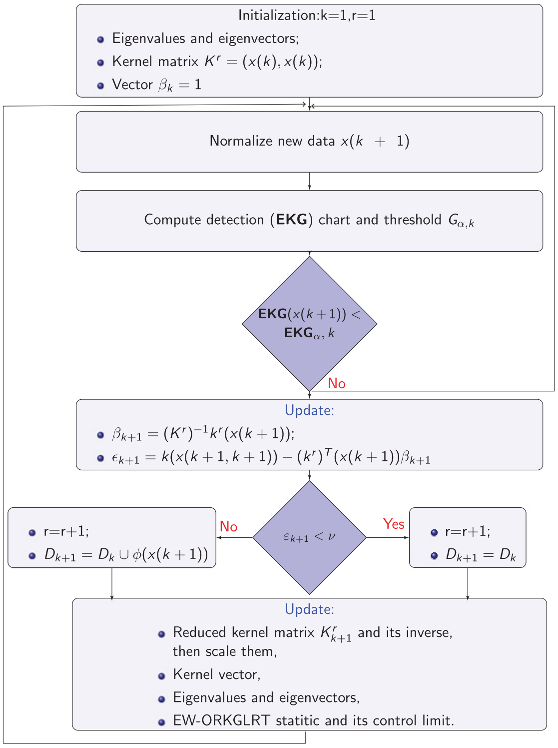

Fig. 1. Diagram of the online leak detection algorithm.

3. CPU-time (CT): The time required for leak detection.

5.1. Water distribution system description

WDN is a complex process consisting of hydraulic elements that are connected (including reservoirs or tanks and consumption nodes), linked by interconnecting links (including pipes, pumps, and valves). In this study, the leak detection is validated using Hanoi benchmark [46]. The diagram of the network is depicted on Fig. 2. The network is built using 34 pipes and 32 nodes arranged in two branches and three loops. The process is gravity fed by a single reservoir. Figures 3 and 4 show the time evolution of demands and pressures of nodes 3, 13 and 25, respectively.

At this stage, the goal is to detect the leaks in the network using the developed technique. Here, we consider a simulation of 24 hours with a sampling time of 15 minutes. The network has 31 demand nodes with index from 2 to 32 (see Fig. 2).

Assume that we have a set of sensors in the network, placed in the nodes 3, 10, 16, 23 and 25, (as depicted on Fig. 2). Then the KPCA model that we

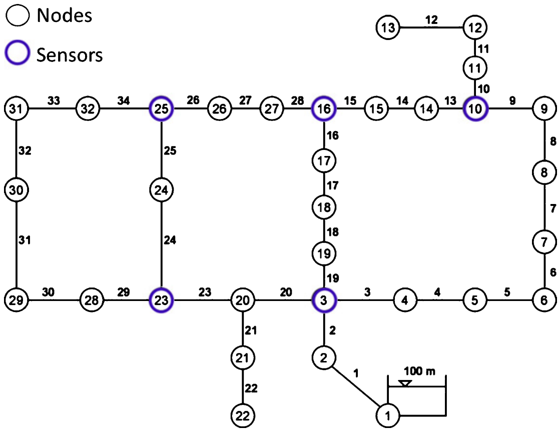

Fig. 2. Hanoi network with sensors placement.

can build concerns only those nodes. The X data is obtained using the five measured pressures in nodes 3, 10, 16, 23, and 25, and is split into training and testing data sets of 97 observations each to carry out leak detection. These data were scaled to zero mean and unit variance.

For KPCA modeling, a radial basis kernel function $k(x, y) = exp(-\frac{(x_i - x_j)^T (x_i - x_j)}{\sigma})$ is selected as the kernel function. The width of the Gaussian function σ is selected beforehand. The value of this parameter can affect the performance of the leak detection. A small value of σ would make the exponential really large argument, making the value of the kernel function very small or near 0. However, a huge value of σ would yield kernel function values very close to 1. To obtain an optimal parameter, a cross-validation methodology is used. Fig. 5 shows the time evolution of pressures of nodes 3, 10, 16, 23 and 25.

Once the KPCA model is identified, we can proceed with leak detection. Statistical confidence limits are set as 95%. The performance of the developed EW-ORKGLRT leak detection method is illustrated and compared to KGLRT and ORKGLRT. The comparison is assessed through different leaks.

Case 1: A leak in node 4

In this case, a leak is simulated in node 4 between the samples 30 and 97. The time evolution of process variables with a leak in node 4 is illustrated in Fig. 6.

The leak detection results using KGLRT, ORKGLRT and EW-ORKGLRT are presented in Figures 7, 8 and 9, respectively. If the statistic values are higher than confidence limit values then there are leaks in the process. We can show from Fig. 7 that there is no leak in the system as expected when

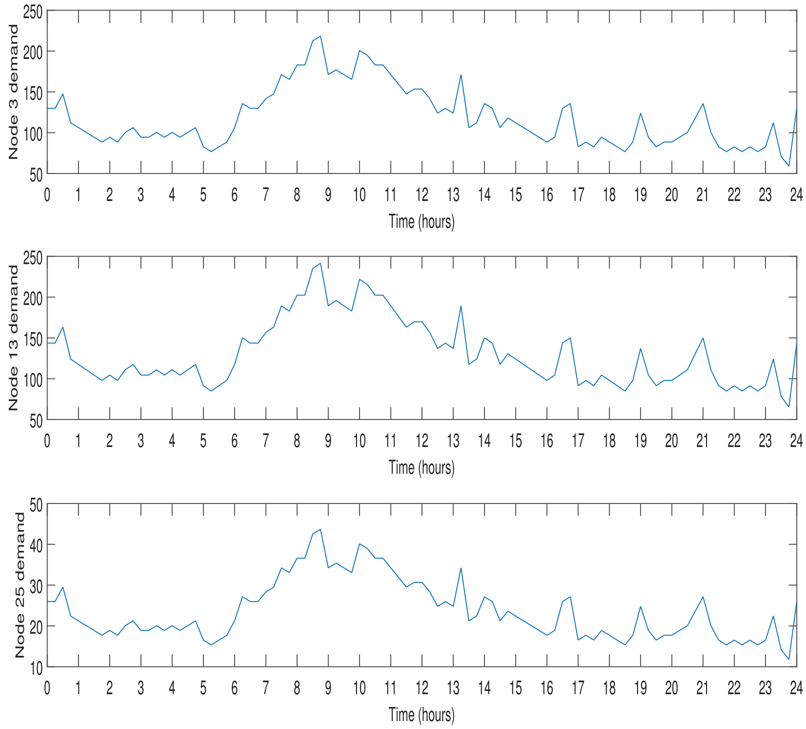

Fig. 3. Time evolution of demand in nodes 3, 13 and 25, respectively.

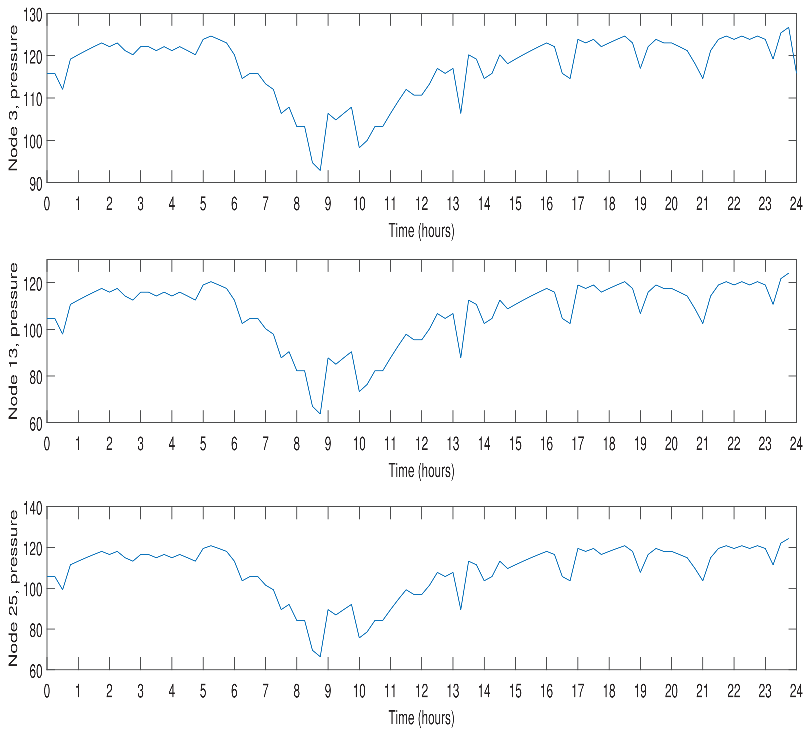

Fig. 4. Time evolution of pressure in nodes 3, 13 and 25, respectively.

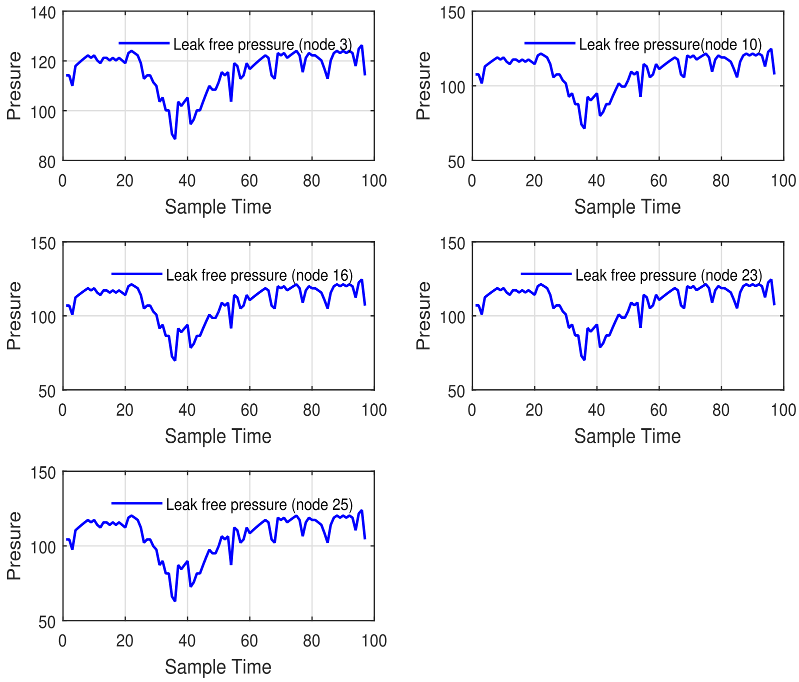

Fig. 5. Time evolution of process variables.

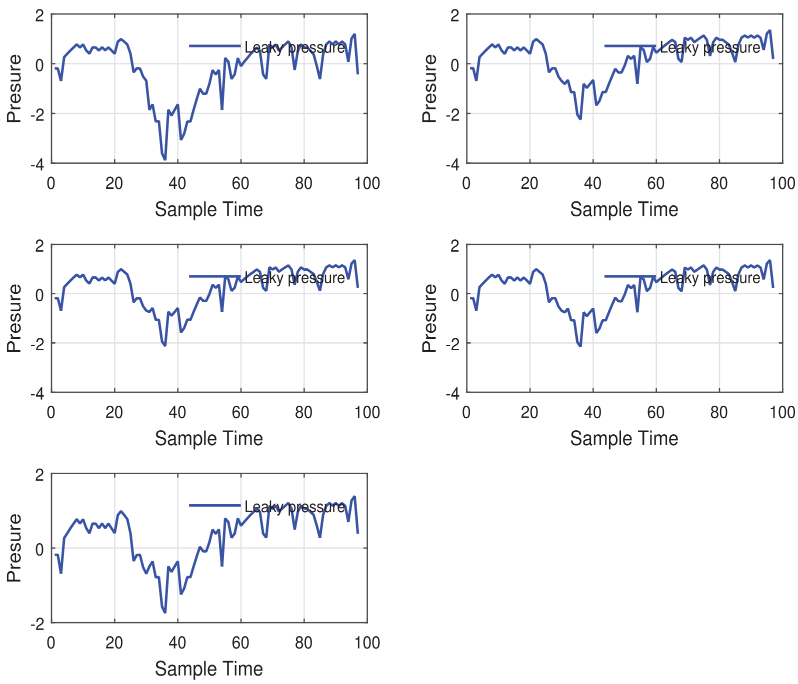

Fig. 6. Time evolution of process variables with a leak in node 4.

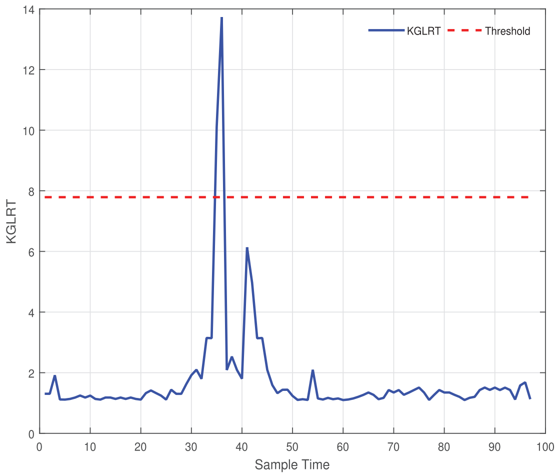

Fig. 7. Time evolution of KGLRT method with a leak in node 4.

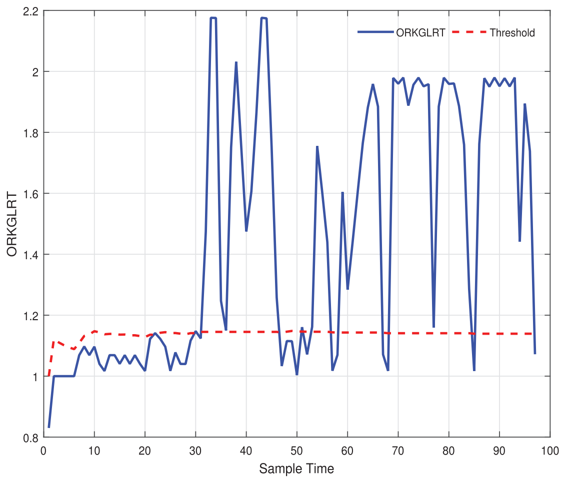

Fig. 8. Time evolution of ORKGLRT method with a leak in node 4.

using the KGLRT chart. However, the ORKGLRT and EW-ORKGLRT methods detect clearly the leak between 30 and 97, this might be due to the adaptation of the threshold for both of them.

The results of the leak detection are illustrated in Table 1 in terms of FAR, GDR, and CPU-time values with a leak in node 4. The results show that both ORKGLRT and EW-ORKGLRT provide better leak detection performance than the KGLRT and they are able to detect the leak in node 4. The control chart obtained using the KPCA model is not adapted according to process changes. However, as shown in Table 1, applying reduced KPCA based adaptive control chart to the same samples, provides better capabilities of adaptation to real behaviour

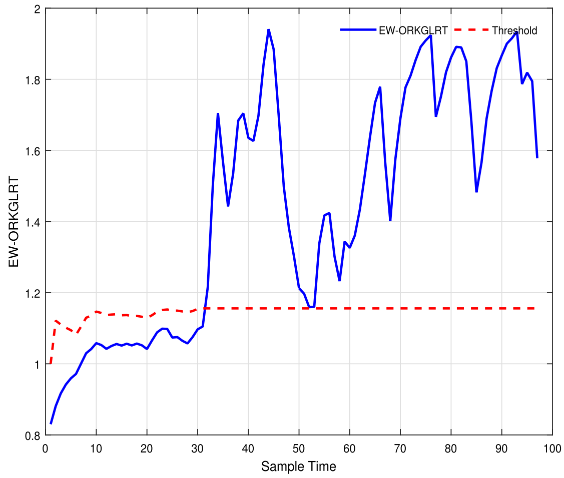

Fig. 9. Time evolution of EW-ORKGLRT method with a leak in node 4.

Table 1
Summary of FAR, GDR and CPU-Time with a leak in node 4

	FAR	GDR(%)	CPU-Time(s)
KGLRT	0	2.9412	0.8125
ORKGLRT	3.4483	82.3529	3.488610^{-4}
EW-ORKGLRT	0	97.0588	2.455210^{-4}

changes of the process. These results show also that the CPU-Time needed for leak detection using ORKGLRT and EW-ORKGLRT is approximately 3300 and 2320 times faster than using KGLRT technique, respectively. This fact is useful in many industrial applications where the updating procedure requires to be processed with other steps in a short time.

Case 2: A leak in node 5

A leak in node 5 is simulated from sample time 30 to 97. Fig. 10 shows time evolution of the process variables with a leak in node 5.

Figures 11, 12 and 13 show the leak detection results using the KGLRT, ORKGLRT and EW-ORKGLRT, respectively. We can show that the considered leak is not detected using the KGLRT technique (Fig. 11). Using the ORKGLRT (Fig. 12), the leak is detected but with higher missed detection rate and some false alarm rate. While, the developed technique presents a good detection abilities when compared to KGLRT, ORKGLRT. EW-ORKGLRT is able to detect the leak between the samples 30 and 96 as illustrated in Fig. 13 which is due the fact that the EW-ORKGLRT is able to take into consideration the information provided by the current and previous

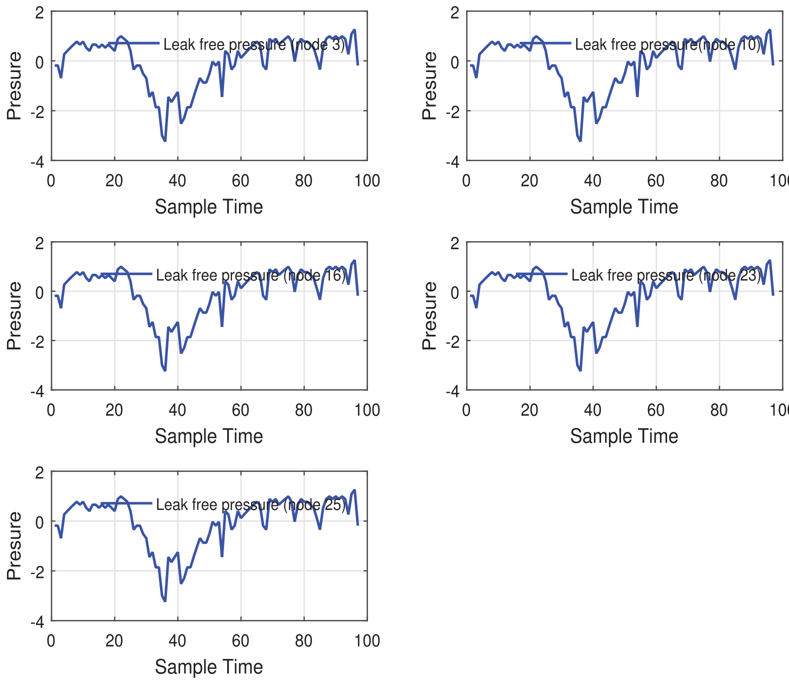

Fig. 10. Time evolution of process variables with a leak in node 5.

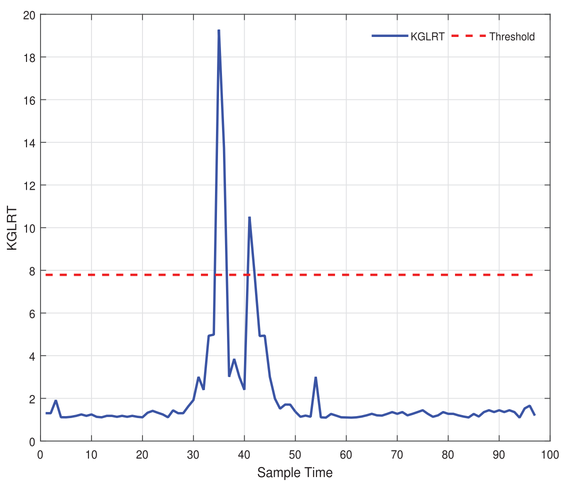

Fig. 11. KGLRT with a leak in node 5.

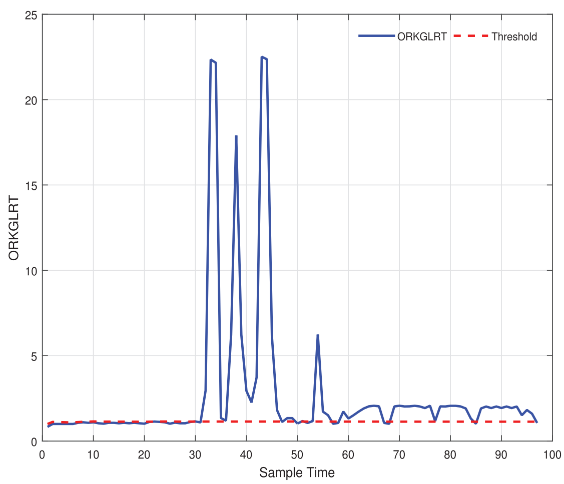

Fig. 12. ORKGLRT with a leak in node 5.

sample by giving significant importance to the newest data. Thus, the use of EW filter improves the leak detection performances by reducing the false alarm and missed detection rates. Also, we should note that the control limits of ORKGLRT and EW-ORKGLRT obtained using the reduced KPCA model are variable over time due to the fact that they are updated when a new sample is available. This fact also provides a good adaptation to the condition in which the system is operating.

Table 2 shows the performances using KGLRT, ORKGLRT and EW-ORKGLRT techniques. As demonstrated in Table 2, the developed EW-ORKGLRT technique is able to detect the leak with higher GDR rate, lower FAR rate, and lower CPU-time.

Case 3: A leak in node 25

In this case, a leak in node 25 is simulated from sample time 30 to 97. Figures 14, 15 and 16, and

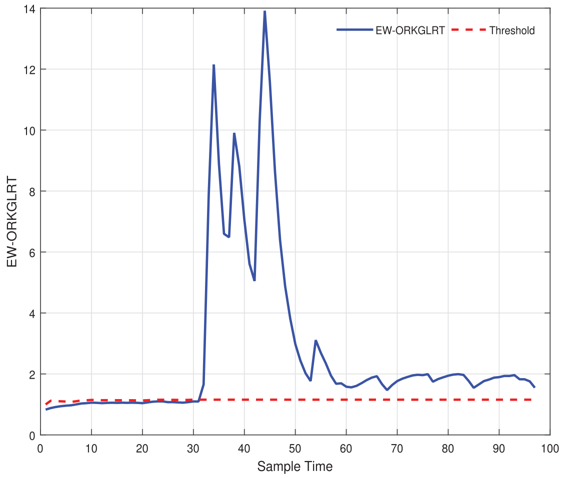

Fig. 13. EW-ORKGLRT with a leak in node 5.

Table 2
Summary of FAR, GDR and CPU-Time with a leak in node 5

	FAR	GDR(%)	CPU-Time(s)
KGLRT	0	4.4118	0.8750
ORKGLRT	3.4483	85.2941	3.532010^{-4}
EW-ORKGLRT	0	97.0588	2.629910^{-4}

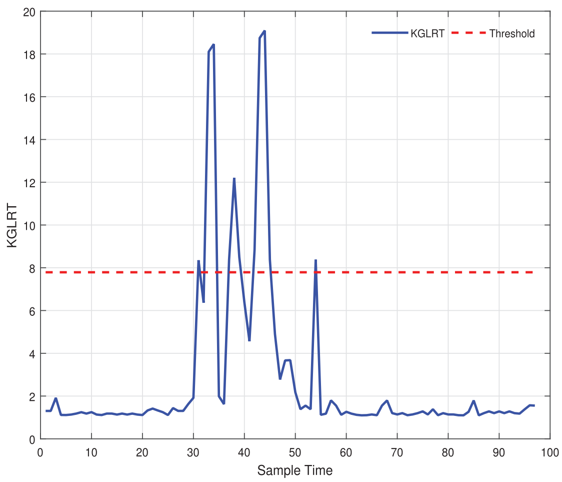

Fig. 14. Time evolution of KGLRT with a leak in node 25.

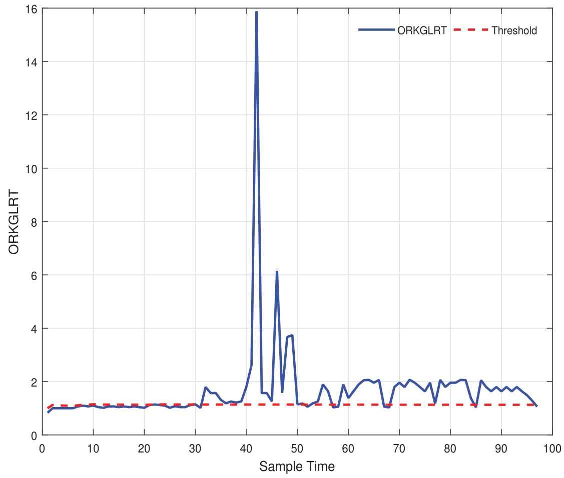

Fig. 15. Time evolution of ORKGLRT with a leak in node 25.

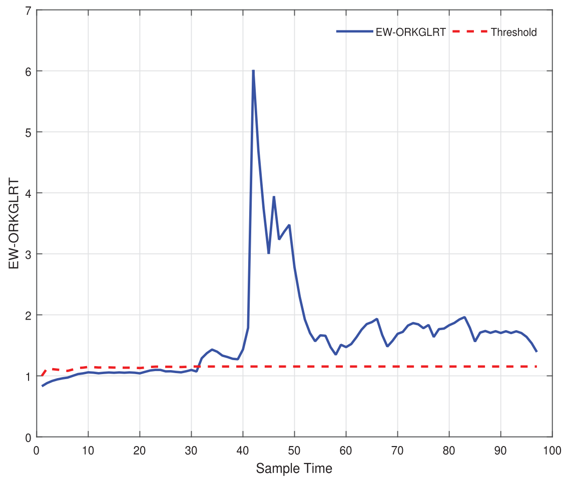

Fig. 16. Time evolution of EW-ORKGLRT with a leak in node 25.

Table 3
Summary of FAR, GDR and CPU-Time with a leak in node 25

	FAR	GDR(%)	CPU-Time(s)
KGLRT	0	4.4118	0.9688
ORKGLRT	3.4483	88.2353	4.588110^{-4}
EW-ORKGLRT	0	94.1176	2.89910^{-4}

Table 3 show the leak detection results of the KGLRT, ORKGLRT and EW-ORKGLRT techniques. We can show from Fig. 14, that KGLRT is not able to detect the leak between 30 and 97. While, the leak in node 25 is detected using the ORKGLRT method with false alarms and high missed detection rates. The leak detection results show also that the proposed method provides good improvement in terms of FAR and GDR, due to the advantages of the filter and the adaptability control limit according to the dynamic process change.

The leak detection performance in different nodes are illustrated in Table 4. These results show that the new method is able to detect the leaks with higher GDR, lower FAR and faster CPU-time. These enhanced results may be awarded to the fact that the

Table 4
Summary of false alarm rate, good detection rate and time computation

	KGLRT			ORKGLRT			EW-ORKGLRT		
leaks	FAR (%)	GDR(%)	CPU-time(s)	FAR (%)	GDR (%)	CPU-time(s)	FAR (%)	GDR (%)	CPU-time(s)
Leak in node 1	0	0	1.09	0	98.52	4.29^{-4}	0	97.05	3.24^{-4}
Leak in node 2	0	0	0.95	3.44	80.88	3.49^{-4}	0	89.70	2.80^{-4}
Leak in node 3	0	0	0.89	3.44	88.23	3.76^{-4}	0	88.23	2.99^{-4}
Leak in node 4	0	2.94	0.81	3.44	82.3	3.48^{-4}	0	97.05	2.45^{-4}
Leak in node 5	0	4.41	0.87	3.4483	85.29	3.53^{-4}	0	97.05	2.62^{-4}
Leak in node 6	0	5.84	1.01	3.44	86.76	3.45^{-4}	0	97.05	2.63^{-4}
Leak in node 15	0	16.17	1.03	3.44	88.23	4.50^{-4}	0	89.70	3.39^{-4}
Leak in node 19	0	2.94	0.95	3.44	85.29	4.17^{-4}	0	86.76	2.86^{-4}
Leak in node 25	0	4.48	0.96	3.44	88.23	4.58^{-4}	0	94.11	2.89^{-4}
Leak in node 26	0	14.70	0.76	0	88.23	3.46^{-4}	0	89.70	3.11^{-4}
Leak in node 28	0	14.70	1.07	3.44	88.23	3.92^{-4}	0	91.17	3.14^{-4}
Leak in node 29	0	14.70	0.95	3.44	88.23	3.51^{-4}	0	97.05	2.77^{-4}
Average	0	6.74	0.94	2.87	80.02	3.53^{-4}	0	92.89	2.91^{-4}

proposed method can capture the dynamic variation in the water distribution networks and also due to the fact that it is able to detect small leak with faster detection time.

6. Conclusion

In this paper, a novel technique for detecting leaks in the water distribution network is developed. In the proposed approach, the modeling phase was addressed using the reduced kernel PCA method, and the leaks were detected using a statistical hypothesis test. The results demonstrated the effectiveness of the proposed framework over the conventional techniques. The detection abilities were evaluated in terms of false alarm rate, good detection rate and CPU-time.

When it has been determined that there are leaks in the network, they should be identified. Therefore, as future work, the EW-ORKGLRT framework merged with the sensitivity analysis that characterizes the theoretical leak signatures developed and applied to leak identification.

References

[1] M. Swiercz and H. Mroczkowska, Multiway pca for early leak detection in a pipeline system of a steam boiler—selected case studies, *Sensors* **20**(6) (2020), 1561.

[2] J. Gertler, J. Romera, V. Puig and J. Quevedo, Leak detection and isolation in water distribution networks using principal component analysis and structured residuals, in: *Control and Fault-Tolerant Systems (SysTol), 2010 Conference on, IEEE*, (2010), 191–196.

[3] I.d.l. Santos-Ruiz, F.R. López-Estrada, V. Puig, E. Pérez-Pérez, J. Mina-Antonio and G. Valencia-Palomo, Diagnosis of fluid leaks in pipelines using dynamic pca, *IFAC-Papers On Line* **51**(24) (2018), 373–380.

[4] W. Ruilian and G. Shengjian, Comprehensive evaluation to distribution network planning schemes using principal component analysis method, *Indonesian Journal of Electrical Engineering and Computer Science* **12**(8) (2014), 5897–5904.

[5] K. Duzinkiewicz, A. Borowa, K. Mazur, M. Grochowski, M.A. Brdys and K. Jezior, Leakage detection and localisation in drinking water distribution networks by multiregional pca, *Studies in Informatics and Control* **17**(2) (2008), 135.

[6] D. Li, A. Jones and B. Engelhardt, Probabilistic contrastive principal component analysis, *arXiv preprint arXiv:2012.07977*.

[7] Z. Niu, C. Wang, Y. Zhang, X. Wei and X. Gao, Leakage rate model of urban water supply networks using principal component regression analysis, *Transactions of Tianjin University* (2017), 1–10.

[8] M.-F. Harkat, M. Mansouri, M. Nounou and H. Nounou, Fault detection of uncertain nonlinear process using interval-valued data-driven approach, *Chemical Engineering Science* **205** (2019), 36–45.

[9] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**(6) (1933), 417.

[10] Y. Hosoya, Hierarchical statistical models and a generalized likelihood ratio test, *Journal of the Royal Statistical Society: Series B (Methodological)* **51**(3) (1989), 435–447.

[11] M. Fouladirad and I. Nikiforov, Optimal statistical fault detection with nuisance parameters, *Automatica* **41**(7) (2005), 1157–1171.

[12] M. Basseville, Information criteria for residual generation and fault detection and isolation, *Automatica* **33**(5) (1997), 783–803.

[13] M.Z. Sheriff, C. Botre, M. Mansouri, H. Nounou, M. Nounou and M.N. Karim, Process monitoring using data-based fault detection techniques: Comparative studies, in: *Fault Diagnosis and Detection, In Tech* (2017), 237–261.

[14] M. Mansouri, M.-F. Destain, H. Nounou and M. Nounou, Enhanced monitoring of environmental processes, *International Journal of Environmental Science and Development* **7**(7) (2016), 525.

[15] C. Botre, M. Mansouri, M.N. Karim, H. Nounou and M. Nounou, Multiscale plsbased glrt for fault detection of chemical processes, *Journal of Loss Prevention in the Process Industries* **46** (2017), 143–153.

[16] M. Mansouri, M.N. Nounou and H.N. Nounou, Multiscale kernel pls-based exponentially weighted-glrt and its application to fault detection, *IEEE Transactions on Emerging Topics in Computational Intelligence* (99) (2017), 1–11.

[17] M. Mansouri, M. Nounou, H. Nounou and N. Karim, Kernel pca-based glrt for nonlinear fault detection of chemical processes, *Journal of Loss Prevention in the Process Industries* **40** (2016), 334–347.

[18] C. Botre, M. Mansouri, M. Nounou, H. Nounou and M.N. Karim, Kernel pls-based glrt method for fault detection of chemical processes, *Journal of Loss Prevention in the Process Industries* **43** (2016), 212–224.

[19] M. Mansouri, M.N. Nounou and H.N. Nounou, Improved statistical fault detection technique and application to biological phenomena modeled by s-systems, *IEEE Transactions on Nanobioscience* **16**(6) (2017), 504–512.

[20] D. De Silva, J. Mashford and S. Burn, Computer aided leak location and sizing in pipe networks, *Urban Water Security Research Alliance* (2009).

[21] J. Mashford, D. De Silva, S. Burn and D. Marney, Leak detection in simulated water pipe networks using svm, *Applied Artificial Intelligence* **26**(5) (2012), 429–444.

[22] J. Bicik, Z. Kapelan, C. Makropoulos and D.A. Savić, Pipe burst diagnostics using evidence theory, *Journal of Hydroinformatics* **13**(4) (2011), 596–608.

[23] Z. Poulakis, D. Valougeorgis and C. Papadimitriou, Leakage detection in water pipe networks using a bayesian probabilistic framework, *Probabilistic Engineering Mechanics* **18**(4) (2003), 315–327.

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

[25] J.-A. Goulet, S. Coutu and I.F. Smith, Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks, *Advanced Engineering Informatics* **27**(2) (2013), 261–269.

[26] C. Alippi, G. Boracchi, V. Puig and M. Roveri, An ensemble approach to estimate the fault-time instant, in *Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on. IEEE*, (2013), 836–841.

[27] G. Boracchi, V. Puig and M. Roveri, A hierarchy of change-point methods for estimating the time instant of leakages in water distribution networks, in *IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer*, (2013), 615–624.

[28] G. Boracchi and M. Roveri, Exploiting self-similarity for change detection, in *Neural Networks (IJCNN), 2014 International Joint Conference on. IEEE*, (2014), 3339–3346.

[29] T. Tao, H. Huang, F. Li and K. Xin, Burst detection using an artificial immune network in water-distribution systems, *Journal of Water Resources Planning and Management* **140**(10) (2014), 04014027.

[30] A. Candelieri, D. Conti, D. Soldi and F. Archetti, Spectral clustering and support vector classification for localizing leakages in water distribution networks–the icewater project approach, (2014).

[31] D. Laucelli, M. Romano, D. Savić and O. Giustolisi, Detecting anomalies in water distribution networks using epr modelling paradigm, *Journal of Hydroinformatics* **18**(3) (2016), 409–427.

[32] S. Mounce, R. Mounce, T. Jackson, J. Austin and J. Boxall, Pattern matching and associative artificial neural networks for water distribution system time series data analysis, *Journal of Hydroinformatics* **16**(3) (2014), 617–632.

[33] S. Mounce, C. Pedraza, T. Jackson, P. Linford and J. Boxall, Cloud based machine learning approaches for leakage assessment and management in smart water networks, *Procedia Engineering* **119** (2015), 43–52.

[34] Y. Kim, S.J. Lee, T. Park, G. Lee, J.C. Suh and J.M. Lee, Robust leakage detection and interval estimation of location in water distribution network, in *9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM*, (2015), 1264–1269.

[35] M. Quiñones-Grueiro, C. Verde and A. Prieto-Moreno, Leaks' detection in water distribution networks with demand patterns, in *Control and Fault-Tolerant Systems (SysTol)*, *2016 3rd Conference on. IEEE* (2016), 313–318.

[36] D. Wachla, P. Przystalka and W. Moczulski, A method of leakage location in water distribution networks using artificial neuro-fuzzy system, *IFAC-PapersOnLine* **48**(21) (2015), 1216–1223.

[37] W. Moczulski, R. Wyczolkowski, K. Ciupke, P. Przystałka, P. Tomasik and D. Wachla, A methodology of leakage detection and location in water distribution networksâĂŤ the case study, in *Control and Fault-Tolerant Systems (SysTol), 2016 3rd Conference on. IEEE*, (2016), 331–336.

[38] G. Sanz, R. Perez and A. Escobet, Leakage localization in water networks using fuzzy logic, in *Control & Automation (MED), 2012 20th Mediterranean Conference on. IEEE*, (2012), 646–651.

[39] X. Li and G.-j. Li, Leak detection of municipal water supply network based on the cluster-analysis and fuzzy pattern recognition, in *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on. IEEE*, (2010), 1–5.

[40] A. Daoudi, M. Benbrahim and K. Benjelloun, An intelligent system to classify leaks in water distribution pipes, *Small* **7**(13) (2005), 20.

[41] S.R. Mounce, A. Khan, A.S. Wood, A.J. Day, P.D. Widdop and J. Machell, Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system, *Information Fusion* **4**(3) (2003), 217–229.

[42] A.C. Caputo and P.M. Pelagagge, Using neural networks to monitor piping systems, *Process Safety Progress* **22**(2) (2003), 119–127.

[43] M. Mansouri, R. Baklouti, M.F. Harkat, H. Nounou and A. Hamida, Kernel generalized likelihood ratio test for fault detection of biological systems, *IEEE Transactions on NanoBioscience 1* **49**(17) (2018), 1–10.

[44] M. Mansouri, M. Hajji, M. Trabelsi, M.F. Harkat, A. Alkhazraji, A. Livera, H. Nounou and M. Nounou, An effective statistical fault detection technique for grid connected photovoltaic systems based on an improved generalized likelihood ratio test, *Energy* **159** (2018), 842–856.

[45] R. Fazai, O. Taouali, M.F. Harkat and N. Bouguila, A new fault detection method for nonlinear process monitoring, *The International Journal of Advanced Manufacturing Technology* **87** (9-12) (2016), 3425–3436.

[46] O. Fujiwara and D.B. Khang, A two-phase decomposition method for optimal design of looped water distribution networks, *Water Resources Research* **26**(4) (1990), 539–549.