The International Journal of Advanced Manufacturing Technology (2018) 99:2789–2802
https://doi.org/10.1007/s00170-018-2674-6

ORIGINAL ARTICLE

Online fault detection and isolation of an AIR quality monitoring network based on machine learning and metaheuristic methods

Radhia Fazai[1] · Khaoula Ben Abdellafou[2] · Maroua Said[1] · Okba Taouali[1]

Received: 17 May 2018 / Accepted: 6 September 2018 / Published online: 15 September 2018

Abstract

The dynamic process monitoring is discussed in this paper. Kernel principal component analysis (KPCA) is a nonlinear monitoring method that cannot be applied for dynamic systems. Reduced online KPCA (OR-KPCA)is used for fault detection of dynamic processes, which is developed to built a dictionary according to the process status and then, it update the KPCA model and uses it for process monitoring. Also the Tabu search metaheuristic algorithm is used in order to determine the optimal parameter of the kernel function. In this paper, new approaches for online fault isolation, which is a challenging problem in nonlinear PCA, are formulated. An extension of partial PCA and the elimination sensor identification (ESI) to the case of nonlinear systems are presented in a feature space. The partial OR-KPCA and the elimination sensor identification (ESI-KPCA) are generated based on the OR-KPCA method and they consist of developing a set of sub-models. The sub-models are selected according to a pre-designed fault-to-residual structure matrix and by eliminating sequentially one variable from the set of the variables. The proposed fault isolation methods are applied for monitoring an air quality monitoring network. The simulation results show that the proposed fault isolation methods are effective for KPCA.

Keywords Principal component analysis (PCA) · Kernel PCA · OR-KPCA · Dynamic process · Fault detection · Fault isolation · Partial OR-KPCA · ESI-KPCA

1 Introduction

Process monitoring and fault diagnosis of industrial processes have become ever-increasingly important because of the rising demands for ensuring process safety and improving product quality. As computer control systems are widely used in modern industry, abundant process data are collected and stored in historical database. Therefore, data-driven fault diagnosis methods have received significant interests from academics and engineers. Many data-driven methods have been developed, including principal component analysis (PCA) [1–4], independent component analysis (ICA) [5–7], and fisher discriminant analysis (FDA) [8–10]. PCA is the most widely used data-driven technique for process monitoring since it can effectively deal with high-dimensional, noisy, and highly correlated data by projecting the data onto a lower dimensional subspace which contains sufficient variance information of the original data.

Despite the proven performances of this technique, it is necessary to mention that it is applied on a system whose variables are mainly linearly related. In fact, PCA is a linear projection method, which cannot effectively capture the nonlinear variable correlations existing in real industrial processes. To overcome the nonlinear short coming of the classical PCA method, several nonlinear extensions of PCA have been developed [11–13]. One of these methods which is frequently used is Kernel principal component analysis (KPCA) [14–17] KPCA can efficiently compute the nonlinear principal components in high-dimensional feature spaces by means of integral operators and nonlinear kernel functions. The basic idea of KPCA is first to map the input space into a feature space via nonlinear mapping and then to perform PCA in that feature space. The main advantages of KPCA are that it only solves an eigenvalue problem and does not involve nonlinear

✉ Okba Taouali
taoualiokba@yahoo.fr

[1] Research Laboratory of Automation, Signal Processing and Image (LARATSI), National School of Engineering of Monastir, University of Monastir, Monastir, 4023, Tunisia

[2] MARS (Modeling of Automated Reasoning Systems) Research Lab LR17ES05, Higher Institute of Computer Sciences and Communication Technologies (ISITCom), University of Sousse, Sousse, Tunisia

optimization. KPCA has been applied successfully for process monitoring and it has proven superior process monitoring performance compared to linear PCA [18] and [19]. In process monitoring with KPCA, fault detection is performed with fault detection indices. Most used indices in the feature space are the Hotelling's T^2 index for monitoring the principal component subspace and the SPE statistic, which monitors the residual subspace [20] and [21].

However, a major limitation of KPCA-based monitoring is that the KPCA model is time-invariant. Most real industrial processes are time-varying. The time-varying characteristics of industrial processes include changes in the variance, changes in the mean, changes in the number of significant principal components (PCs). Monitoring processes with fixed KPCA model may lead to the occurrence of false alarms which significantly reduce the reliability of this method. To track the changing characteristics of the industrial process, attention should be paid more on the recent data and the adaptation of KPCA model is necessary. To update the KPCA model, several dynamic KPCA methods have been proposed [22–26]. One of these methods is online reduced KPCA (OR-KPCA) [27], which consists to extract a reduced number of kernel function and then it updates the old KPCA model and uses it for process monitoring. The OR-KPCA method is performed in an online phase (One phase) and takes into consideration the dynamic behaviors of the systems by changing the model structure. It reduces significantly the computation time required to detect faults in nonlinear systems while conserving the structure of the data in the feature space. It has been applied successfully for fault detection of dynamic process data.

Once a fault is detected by the statistical monitoring method, it is important to identify which variables are responsible for causing the abnormality in the process. This phase is named isolation. Based on the literature, only few papers propose new nonlinear method for isolation [18, 28, 29]. To locate the faulty variables, [18] propose a method for fault diagnosis employing the reconstruction method of [30], which looks at fault identification index obtained when a fault detection index has been reconstructed along the direction of a variable However, this technique applies to sensor faults or a process fault with a known direction, it cannot be applied to identify faulty variables with unknown directions. To cope with this problem, reconstruction-based contribution for KPCA (RBC-KPCA) [28] is proposed for fault identification. This method can guarantee the correct diagnosis of the variable with the largest contribution to the fault. However, these methods are applied to diagnosis fault for nonlinear statistical process. Thus, they do not consider the variation of parameters according to the process operation changes. To handle the problem posed by nonstationary data, extended versions of partial PCA [31–34] and the elimination sensor identification (ESI) [36] methods to nonlinear case have been proposed. These methods aim to identify the faulty variable by taking into account the changing characteristics of data. They consist to apply the OR-KPCA method to a set of sub-models. The sub-model is violated if a sensor or actuator associated with any of the variables in the subset is faulty, but is insensitive to faults associated with variables outside the subset. The effectiveness of the two proposed monitoring methods are evaluated using an air quality monitoring network (AIRLOR). The simulation results prove the performances of these methods.

The paper is outlined as follows. The concept of KPCA is introduced in Section 2. The online reduced KPCA method for fault detection is described in Section 3. The Tabu search metaheuristic algorithm is presented in Section 4. Section 5 presents the proposed online fault isolation methods. Section 6 describes the application of the proposed methods for monitoring the AIRLOR process. At the end, the conclusions are presented in Section 7.

2 Kernel principal component analysis

To derive KPCA, we first map the training set $X = \left[x(1)\ x(2)...x(N) \right]^T \in \mathbb{R}^{N\times m}$, where N is the number of samples and m is the number of process variables, into a high-dimensional feature space $\mathcal{F}$ via a nonlinear mapping $\phi_i = \phi(x(i))$ and then to apply a linear PCA in the feature space $\mathcal{F}$.

In the feature space, the inner product of two vectors ϕ_i and ϕ_j is determined via a kernel function of the corresponding vectors $x(i)$ and $x(j)$ as follows:

$$\phi_i^T \phi_j = k(x_i, x_j) \tag{1}$$

The function $k(.,.)$ is called kernel function. There exist several types of these functions [37]. The representative kernel functions are :

- Polynomial kernel : $k(x, y) =< x, y >^d$
- Sigmoid kernel : $k(x, y) = \beta_0 < x, y > +\beta_1$
- Radial basis kernel : $k(x, y) = \exp\left(-\frac{(x-y)^T(x-y)}{\sigma}\right)$

where d, β_0, β_1, c are specified a priori by the user.

Assuming that the vectors in the feature space are scaled to zeros mean, the training data are denoted as

$$\mathcal{X} = \left[\phi(x(1))\ \phi(x(2))\ ...\ \phi(x(N)) \right]^T \in \mathbb{R}^{N\times h} \tag{2}$$

where $h >> m$ is the dimension of the feature space.

The covariance Q can be expressed in the extended feature space as follows:

$$\begin{aligned} Q &= \frac{1}{N-1}\mathcal{X}^T\mathcal{X} \\ &= \frac{1}{N-1}\sum_{i=1}^{N}\phi(x(i))\phi^T(x(i)) \end{aligned} \tag{3}$$

The kernel principal components can be obtained by solving the eigenvector equation in the feature space:

$$\mu v = Qv \tag{4}$$

where μ and v are an eigenvector and eigenvalue of Q.

From Eqs. 3 and 4 can be expressed as follows:

$$\mu v = \frac{1}{N-1}\sum_{i=1}^{N}\phi(x(i))\phi^T(x(i))v \tag{5}$$

Unfortunately, the right-hand side of Eq. 5 is not tractable since the kernel mapping is defined by an inner product only. However, the following related matrix is tractable.

$$\begin{aligned} K &= \mathcal{X}\mathcal{X}^T \\ &= \begin{bmatrix} k(x(1),x(1)) & \cdots & k(x(1),x(N)) \\ \vdots & \ddots & \vdots \\ k(x(N),x(1)) & \cdots & k(x(N),x(N)) \end{bmatrix} \end{aligned} \tag{6}$$

Multiplying $\mathcal{X}$ with both side of Eq. 5, the following equation is obtained :

$$\lambda\mathcal{X}v = \mathcal{X}\mathcal{X}^T\mathcal{X}v \tag{7}$$

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

Using kernel trick $K = \mathcal{X}\mathcal{X}^T$, Eq. 6 can be expressed as a simplified form as follows:

$$\lambda\mathcal{X}v = K\mathcal{X}v \tag{8}$$

Denote $\alpha = \mathcal{X}v$ and rewrite (8) as follows:

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

Equation 8 shows that α and λ are an eigenvector and its corresponding eigenvalue of the kernel matrix K.

From the expression of $\alpha = \mathcal{X}v$, v can be represented by the following simple form:

$$v = \lambda^{-1}\mathcal{X}^T\alpha \tag{10}$$

In the feature space, to determine the PCA model (λ_i and v_i), we first perform eigen-decomposition of Eq. 9. Then, use Eq. 10 to determine v_i. To guarantee that $v_i^T v_i = 1$, Eq. 3 and the expression of $\alpha = \mathcal{X}v$ are used to derive the following form:

$$\begin{aligned} \alpha_i^T\alpha_i &= v_i^T\mathcal{X}^T\mathcal{X}V_i \\ &= v_i^T\lambda_i v_i \\ &= \lambda_i \end{aligned} \tag{11}$$

Therefore, α_i needs to have a norm of $\sqrt{\lambda_i}$. Let us define a unit norm eigenvector corresponding to λ_i by α_i^0, such that the following equation is satisfied:

$$\alpha_i = \sqrt{\lambda_i}\alpha_i^0 \tag{12}$$

In the feature space, the eigenvectors v_i form a matrix $P_f = \left[\, v_1\ v_2 ... v_\ell ... v_N \,\right]$, where ℓ is the number of retained kernel principal components. The first part $\hat{P}_f = \left[\, v_1\ v_2\ ...\ v_\ell \,\right]$ represents the principal subspace that spans the maximal variance between data. The second part $\tilde{P}_f = \left[\, v_\ell\ ...\ v_N \,\right]$ defines the residual subspace that contains the noises. The two subspaces are complementary. Using this configuration, we estimate the KPCA model. Therefore, the number of significant principal components (PCs) ℓ is determined based on the cumulative percent variance (CPV) criterion. The CPV is a measure of the percent variance determined by the first ℓ PCs:

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

In this paper, the number of principal components is retained when CPV attained a predetermined limit (95%).

From Eq. 10, $\hat{P}_f$ can be rewritten in the following form:

$$\begin{aligned} \hat{P}_f &= \left[\, \tfrac{1}{\lambda_1}\mathcal{X}^T\alpha_1\ \tfrac{1}{\lambda_2}\mathcal{X}^T\alpha_2\ ...\ \tfrac{1}{\lambda_\ell}\mathcal{X}^T\alpha_\ell \,\right] \\ &= \left[\, \mathcal{X}^T\alpha_1^0\lambda_1^{-\frac{1}{2}}\ \mathcal{X}^T\alpha_2^0\lambda_2^{-\frac{1}{2}}\ ...\ \mathcal{X}^T\alpha_\ell^0\lambda_\ell^{-\frac{1}{2}} \,\right] \\ &= \mathcal{X}^T P\Lambda^{-\frac{1}{2}} \end{aligned} \tag{14}$$

where $P = \left[\, \alpha_1^0\ \alpha_2^0\ ...\ \alpha_\ell^0 \,\right]$ and $\Lambda = \mathrm{diag}(\lambda_1 ... \lambda_\ell)$ are the ℓ principal eigenvectors and eigenvalues of the matrix K, respectively, corresponding to the largest eigenvalues in descending order.

Denote x as a new measured data of the sensor, $\phi(x)$ is its mapped vector in the feature space $\mathcal{F}$. The projection of $\phi(x)$ on the principal and residual spaces are defined by:

$$\hat{t} = \hat{P}_f\phi(x) \tag{15}$$

$$\tilde{t} = \tilde{P}_f\phi(x) \tag{16}$$

In the feature space, it is preferred to have centered data. Therefore, the centered vector $\overline{\phi}$ is given by:

$$\begin{aligned} \overline{\phi}(x) &= \phi(x) - \frac{1}{N}\sum_{i=1}^{N}\phi(x(i)) \\ &= \phi(x) - \left[\, \phi(x(1))\ \phi(x(2))\ ...\ \phi(x(N)) \,\right] I_N \end{aligned} \tag{17}$$

where $I_N = \frac{1}{N}\left[\, 1\ \ldots\ 1 \,\right]^T \in \mathbf{R}^N$.

From Eq. 17, the kernel function of two centered vectors $\overline{\phi}(x(i))$ and $\overline{\phi}(x(j))$ is expressed as follows:

$$\begin{aligned}\overline{k}(x(i),x(j)) &= \overline{\phi}(x(i))^T\overline{\phi}(x(j)) \\ &= k(x(i),x(j)) - k^T(x(i))I_N \\ &\quad -k^T(x(j))I_N + I_N^T K I_N \end{aligned} \tag{18}$$

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

Using Eq. 18, the centered kernel vector $\overline{k}(x)$ is determined as :

$$\begin{aligned}\overline{k}(x) &= \left[\,\overline{\phi}(x(1))\ \ldots\ \overline{\phi}(x(N))\,\right]^T\overline{\phi}(x) \\ &= F(k(x) - KI_N)\end{aligned} \tag{19}$$

where

$$F = I - E \tag{20}$$

with I is the identity matrix and $E = II_N^T$.

Finally, the centered kernel matrix $\overline{K}$ can be expressed from the non-centered kernel matrix K, as follows:

$$\begin{aligned}\overline{K} &= \left[\,\overline{\phi}(x_1)\ \ldots\ \overline{\phi}(x_N)\,\right]^T\left[\,\overline{\phi}(x_1)\ \ldots\ \overline{\phi}(x_N)\,\right] \\ &= FKF\end{aligned} \tag{21}$$

Once the KPCA model is estimated during the learning phase, faults can be detected by defining statistics in the feature space. In general, KPCA-based fault detection methods use the squared prediction error (SPE), which is defined as the norm of the residual vector in the feature space and is calculated as [27] and [38]:

$$\mathrm{SPE}(x) = \tilde{t}^T\tilde{t} \tag{22}$$

From Eqs. 16, 17, 22 becomes equal to

$$SPE(x) = \overline{\phi}(x)^T\tilde{P}_f\tilde{P}_f^T\overline{\phi}(x) \tag{23}$$

The product $\tilde{P}_f\tilde{P}_f^T$ can be determined as the projection orthogonal to the principal component space as follows:

$$\tilde{P}_f\tilde{P}_f^T = I - \hat{P}_f\hat{P}_f^T \tag{24}$$

Inserting (24) into (23), we obtain :

$$\begin{aligned}\mathrm{SPE}(x) &= \overline{\phi}(x)^T(1 - \hat{P}_f\hat{P}_f^T)\overline{\phi}(x) \\ &= \overline{\phi}(x)^T\overline{\phi}(x) - \overline{\phi}(x)^T\hat{P}_f\hat{P}_f^T\overline{\phi}(x)\end{aligned} \tag{25}$$

Inserting (14) and (19) into (25), the SPE index can be written as

$$\begin{aligned}\mathrm{SPE}(x) &= \overline{k}(x,x) - \overline{k}(x)^T P\Lambda^{-1}P^T\overline{k}(x) \\ &= \overline{k}(x,x) - \overline{k}(x)^T C\overline{k}(x)\end{aligned} \tag{26}$$

where $C = P\Lambda^{-1}P^T$. The processus is considered in normal operation if the following condition is satisfied:

$$\mathrm{SPE}(x) < \mathrm{SPE}_\alpha \tag{27}$$

The control limit SPE_α of the SPE statistic is calculated from the χ^2 distribution and it is given as

$$\mathrm{SPE}_\alpha = g\chi^2_{h,\alpha} \tag{28}$$

where $g = \frac{b}{2a}$ and $h = \dfrac{2a^2}{b}$, with a and b are the mean and variance of the SPE index.

3 Online Reduced KPCA for fault detection

Online Reduced KPCA (OR-KPCA) [27] is a method for feature extraction and dimension reduction. It aims to find a reduced data set, referred as dictionary, to represent the original training set. Then, it consists to update the KPCA model and use it for fault detection. The dictionary is formed by the evaluation of the SPE index and an approximation criterion [39–42]. Therefore, under normal operating condition, an approximation criterion determines whether it is necessary to update the dictionary or not.

3.1 Unchanged dictionary

In this case, the kernel function $\phi(x(k+1))$ corresponding to the observation $x(k+1)$ is not included in the dictionary, since it can be approximated by its elements. This can be done by comparing $\phi(x(k+1))$ to its projection onto the space spanned by the other r kernel functions , yielding the following optimization problem :

$$\varepsilon_{k+1} = min_\beta\|\phi(x(k+1)) - \sum_{j=1}^{r}\beta_j\phi(x(w_j))\|_H^2 < \nu \tag{29}$$

where ν is a positive threshold parameter that controls the level of sparseness. The optimal vector $\beta = \left[\,\beta_1\cdots\beta_r\,\right]^T$ is determined by the minimization of the left term of Eq. 29. By developing the left term of the expression (29), this leads to the following condition expression :

$$\begin{aligned}\varepsilon_{k+1} &= \min_\beta \sum_{j,i=1}^{r}\beta_j\beta_i k(x(w_j),x(w_i)) \\ &-2\sum_{j=1}^{r}\beta_j k(x(w_j),x(k+1)) + k(x(k+1),x(k+1)) \\ &= \min_\beta \beta^T K_r\beta - 2\beta^T k^r(x(k+1)) + k(x(k+1),x(k+1))\end{aligned} \tag{30}$$

where $K_k^r \in \mathbf{R}^{r\times r}$ is the Gram matrix of the dictionary D_k which is given by

$$K_k^r = \begin{bmatrix} k(x(w_1),x(w_1)) & \cdots & k(x(w_r),x(w_1)) \\ \vdots & \ddots & \vdots \\ k(x(w_1),x(w_r)) & \cdots & k(x(w_r)),x(w_r)) \end{bmatrix} \tag{31}$$

The vector $k^r(x(k+1))$ is defined as follows:

$$k^r(x(k+1)) = [k(x(w_1),x(k+1))\cdots(x(w_r),x(k+1))]^T \tag{32}$$

By solving (30), we obtain the vector β:

$$\beta = (K_k^r)^{-1} k^r(x(k+1)), \quad k = 1, ..., N \tag{33}$$

In this case, the size of the dictionary remains unchanged. Thus, $D_{k+1} = D_k$.

The average M_k of the dictionary D_k, at instant k, is defined by

$$M_k = \frac{1}{r} \sum_{i=1}^{r} \phi(x(w_i)) \tag{34}$$

The centering of the vector $\phi(x(k+1))$, at instant $k+1$, is given by

$$\overline{\phi}(x(k+1)) = \phi(x(k+1)) - M_k \tag{35}$$

The vector M_{k+1} of the dictionary D_{k+1} is not updated ($M_{k+1} = M_k$). Therefore, the centered Gram matrix remains unchanged, $\overline{K}_{k+1}^r = \overline{K}_k^r$.

Using Eq. 10, the eigenvector V_{k+1}^r, at the instant $k+1$, can be written in the following form:

$$v_{k+1}^r = \lambda_{k+1}^{-1} \sum_{i=1}^{r} \alpha_{k+1,i}^r \overline{\phi}(x(w_i)), \quad k = 1, ..., N \tag{36}$$

The centered kernel vector $\overline{k}^r(x(k+1))$ is given by

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

3.2 Changed dictionary

In this case, the kernel function $\phi(x(k+1))$ corresponding to the observation $x(k+1)$ cannot be efficiently approximated by the other elements of the dictionary and thus should be added in the dictionary. Thus, the size of the dictionary is incremented to $r+1$. Therefore, the dictionary will be given by

$$\mathcal{D}_{k+1} = \left[\phi(x(w_1)) \; \phi(x(w_2)) \; . \; . \; . \; \phi(x(w_r)) \; \phi(x(w_{r+1})) \right]^T \tag{38}$$

The update of the vector M_{k+1}, at instant $k+1$, is determined by

$$M_{k+1} = \frac{1}{r+1}(r M_k + \phi(x(w_{r+1}))) \tag{39}$$

The eigenvector is updated as follows:

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

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

The Gram matrix is updated by adding a column and a row to the previous one, as follows:

$$\overline{K}_{k+1}^r = \begin{bmatrix} \overline{K}_k^r & \overline{k}^r(x(k+1)) \\ \overline{k}^r(x(k+1))^T & \overline{k}(x(k+1), x(k+1)) \end{bmatrix} \in \mathbf{R^{r+1 \times r+1}} \tag{41}$$

The inverse of the Gram matrix $(K_{k+1}^r)^{-1}$ is obtained by applying the Woodbury matrix identity:

$$\begin{aligned} {K_{k+1}^r}^{-1} &= \begin{bmatrix} (K_k^r)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\varepsilon_{k+1}} \begin{bmatrix} -(K_k^r)^{-1} k^r(x(k+1)) \\ 1 \end{bmatrix} \\ &\quad \times \begin{bmatrix} -k^r(x(k))^T (K_k^r)^{-1} & 1 \end{bmatrix} \\ &= \begin{bmatrix} (K_k^r)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{\varepsilon_{k+1}} \begin{bmatrix} -\beta_k \\ 1 \end{bmatrix} \begin{bmatrix} -\beta_{k+1}^T & 1 \end{bmatrix} \end{aligned} \tag{42}$$

According to Eq. 42, the vector of the coefficients β_{k+1} can be expressed as

$$\beta_{k+1} = (K_{k+1}^r)^{-1} k^r(x(k+1)) \tag{43}$$

The SPE index of the observation $x(k+1)$ using the dictionary $\mathcal{D}_{k+1}$ is defined by

$$\begin{aligned} \mathrm{SPE}(x(k+1)) &= \overline{k}(x(k+1), x(k+1)) \\ &\quad - (\overline{k}^r(x(k+1)))^T C^r (\overline{k}^r(x(k+1))) \end{aligned} \tag{44}$$

where

$$\begin{cases} \overline{k}^r(x(k+1)) = \left[\overline{k}(x(w_1), x(k+1)) ... \overline{k}(x(w_{r+1}), x(k+1)) \right]^T \\ C^r = P^r (\Lambda^r)^{-1} (P^r)^T \\ \Lambda^r = \mathrm{diag}(\lambda_{\ell_1} ... \lambda_{\ell_{r+1}}) \\ P^r = \left[\bar{\alpha}^r{}_{\ell_1} \; \bar{\alpha}^r{}_{\ell_2} \; . \; . \; . \; \bar{\alpha}^r{}_{\ell_{r+1}} \right] \\ \ell_{r+1} \text{ is the number of retained principal components using the dictionary } \mathcal{D}_{k+1}. \end{cases}$$

During the abnormal conditions of the process ($\mathrm{SPE}(x(k+1)) > \mathrm{SPE}_{\alpha,k}$), the dictionary $\mathcal{D}_{k+1} = [\phi(x(w_1))\phi(x(w_2)) \ldots \phi(x(w_r))]^T$ remains unchanged.

4 Selection of kernel parameter using Tabu search algorithm

4.1 Principle

The kernel function is the core of the kernel method which helps it to get an optimal solution. In general, the RBF kernel, as a nonlinear kernel function, is a reasonable first choice. The parameter σ is key element of the RBF kernel and directly exerts considerable influence on the generalization ability of the online reduced KPCA. The selection of the kernel function and the corresponding

parameter is the key of KPCA. The parameter σ of kernel function has an effect on the partitioning outcome in the feature space. If the value of σ is too large, it will lead to over-fitting. If the value of σ is too small, it will lead to under-fitting. In this part, we present an approach to select what is optimal Gaussian kernel parameter to use when applying the proposed online reduced KPCA technique. The optimal kernel parameter is defined as the one that can improve the fault detection performance . For many industrial applications, minimizing the false alarm rate may be the greatest performance criterion. Therefore, the choice of the Gaussian kernel parameter needs to be selected based on the given application. The tabu search algorithm is applied to optimize the kernel parameter to use when applying the online reduced KPCA algorithm.

4.2 Initial solution

In this study, the determination of the initial solution in the tabu search algorithm is to optimize the σ for the current online reduced KPCA model. Firstly, an initialization solution is presented by random. To reduce the search space referring to previous literature using the online reduced KPCA model, it is recommended to introduce the constraints of the parameter σ which respectively attribute to the range $\sigma \in [2^{-6}, 2^{6}]$. The solution is computed by appending the nearest unused neighbor values of the parameter with respect improve the fault detection performance. The process repeats until all the neighbors are visited.

5 Online fault isolation

In this section, two online fault isolation methods are proposed. The partial OR-KPCA and the elimination sensor identification (ESI) algorithm based on the OR-KPCA method were developed in the nonlinear case. These methods consist of developing a set of sub-models. This sub-model is violated if a sensor or actuator associated with any of the variables in the subset is faulty, but is insensitive to faults associated with variables outside the subset.

5.1 Partial OR-KPCA

This method is an extension of partial PCA method which is based on the generation of structured residuals [31–35]. The concept of this method is to perform the OR-KPCA method on a reduced vector where some variables are discarded from the original vector. Therefore, the residual will only be sensitive to faults associated with the variables which are present in the reduced vector. These structured residuals are generated according to a strongly isolable incidence matrix. The rows of this matrix belong to residuals and its columns to faults. The value "0" at an intersection indicates that the residual does not respond to the fault while the value "1" indicates that the residual is sensitive to this fault. Columns of the incidence matrix are the Boolean fault codes obtained in response to the particular faults. For the construction of this matrix, it is necessary that

- The number of "0" contained in each column is equal to that contained in each line.
- The number of lines is equal to the number of faults.

The incidence matrix strongly isolating if no column can arise from another column by replacing a "1" with a "0".

The procedure of achieving a structured partial OR-KPCA set is as follows:

1. Apply the OR-KPCA method to the data matrix.
2. Construct an incidence matrix, preferably with strong isolation properties.
3. Perform a set of partial OR-KPCAs with each one implementing a row of the incidence matrix.

This procedure is illustrated in Fig. 1.

After the structured partial OR-KPCA set is obtained, it can be used in online fault isolation. The procedure of online isolation using partial OR-KPCA method is summarized by the following steps:

1. Determine the $\mathrm{SPE}_i(k), i \in 1, ..., q$ index and its control limit of each partial model.

   The control limit $\mathrm{SPE}_{\alpha,i}(k)$ of the $\mathrm{SPE}_i(k)$ statistic of each partial model is given as follows:

$$\mathrm{SPE}_{\alpha,i}(k) = g_i \chi^2_{h_i,\alpha} \tag{45}$$

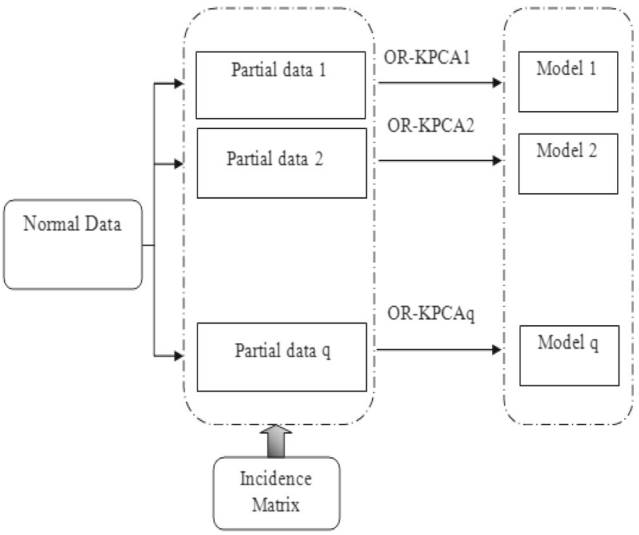

Fig. 1 Procedure of structured partial OR-KPCA

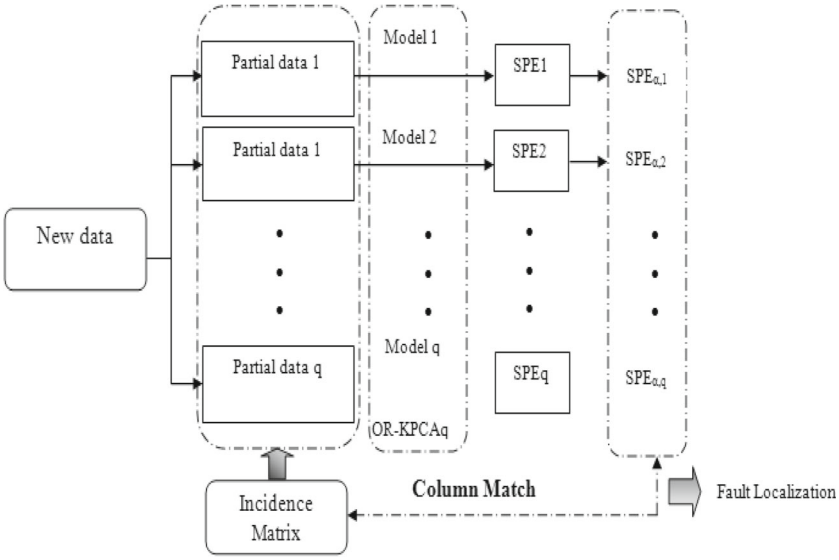

Fig. 2 Procedure of localization using partial OR-KPCA method

where $g_i = \frac{b_i}{2a_i}$ and $h_i = \frac{2a_i^2}{b_i}$, with a_i and b_i are the mean and variance of the statistic of each partial model.

2. Compare the indices to appropriate thresholds and form the fault code $\mathrm{SE}(k) = \left[\, \mathrm{SE}_1(k) \quad \ldots \quad \mathrm{SE}_q(k) \,\right]^T$ SE_i

$$\mathrm{SE}_i(k) = \begin{cases} 0 & \text{if } \mathrm{SPE}_i(k) \leq \mathrm{SPE}_{\alpha,i}(k) \\ 1 & \text{if } \mathrm{SPE}_i(k) > \mathrm{SPE}_{\alpha,i}(k) \end{cases} \tag{46}$$

3. Compare the fault code to the columns of the incidence matrix to arrive at a localization decision.

The fault localization procedure using partial OR-KPCA method is presented in Fig. 2.

5.2 The elimination sensor identification algorithm

The elimination sensor identification (ESI) algorithm has been proposed by [36] in the linear PCA method. In this paper, an extension of this method in the nonlinear case is proposed. The ESI-KPCA algorithm consists, after detecting the presence of a fault with the SPE index, to eliminate each of the m variables sequentially from the set of variables and recalculate the SPE index. Therefore, the new SPE index calculated after the elimination of the jth variable will be given by

$$\mathrm{SPE}_{-j}(x_{-j}(k)) = \bar{k}(x_{-j}(k), x_{-j}(k)) - \bar{k}(x_{-j}(k))^T C_{-j} \bar{k}(x_{-j}(k)) \tag{47}$$

where $x_{-j}(k)$ is the vector $x(k)$ after the elimination of the jth variables, $C_{-j} = P_{-j} \Lambda_{-j} P_{-j}^T$ with P_{-j} and Λ_{-j} are the ℓ eigenvectors and eigenvalues of the matrix K after elimination of the jth variables.

This quantity suggests the existence of an abnormal situation in the data when

$$\mathrm{SPE}_{-j} > \mathrm{SPE}_{(\alpha,-j)} \tag{48}$$

where $\mathrm{SPE}_{(\alpha,-j)}$ is the control limit of the SPE_{-j} index.

A quantity designated as the Q_r, or the ratio between the SPE_{-j} and its control limit $\mathrm{SPE}_{(\alpha,-j)}$, is accordingly calculated for each of the m OR-KPCA models as follows:

$$Q_r = \frac{\mathrm{SPE}_{-j}}{\mathrm{SPE}_{(\alpha,-j)}} \tag{49}$$

The eliminated variable, for which the ratio Q_r is the smallest, is considered as the faulty variable.

6 Application to air quality monitoring network

In this section, the proposed methods for isolation are applied to diagnosis of an air quality monitoring network [43–47].

6.1 Description of air quality monitoring network

The air quality monitoring network (AIRLOR) working in Lorraine, (France) consists of twenty stations located in rural, peri-urban, and urban sites. Each monitoring station consists of a set of sensors for measuring concentrations of pollutants: carbon monoxide CO, oxides of nitrogen (NO and NO_2) measured by the same analyzer, the dioxide sulfur $SO2$ and ozone O_3. Moreover, some stations are dedicated to the recording of additional meteorological parameters. The measured meteorological parameters are

– Temperature °C,

Fig. 3 Air quality monitoring station

- Global solar radiation (W/m^2),
- Relative humidity (%),
- Atmospheric pressure (hPa)
- wind direction (degree),
- wind speed (m/s)

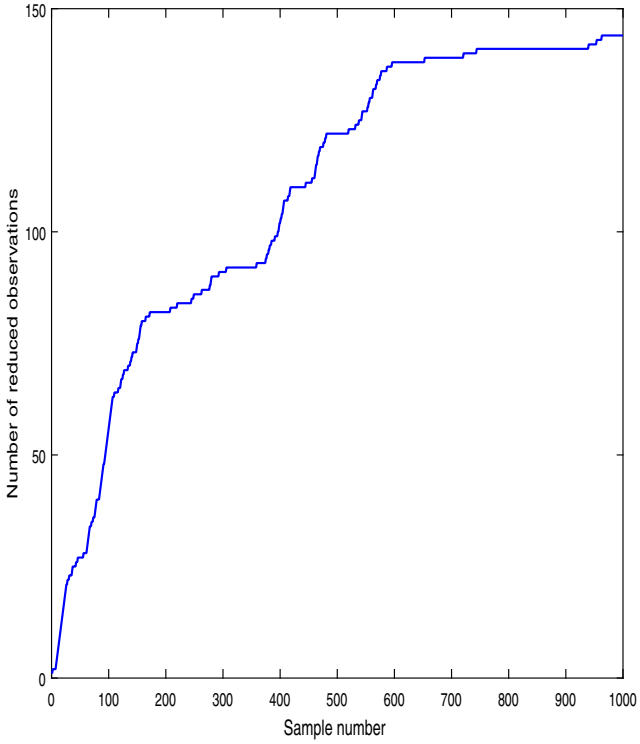

Fig. 5 Time evolution of the reduced number of observations

A measuring station is a room in which analyzers are located as it is shown in Fig. 3.

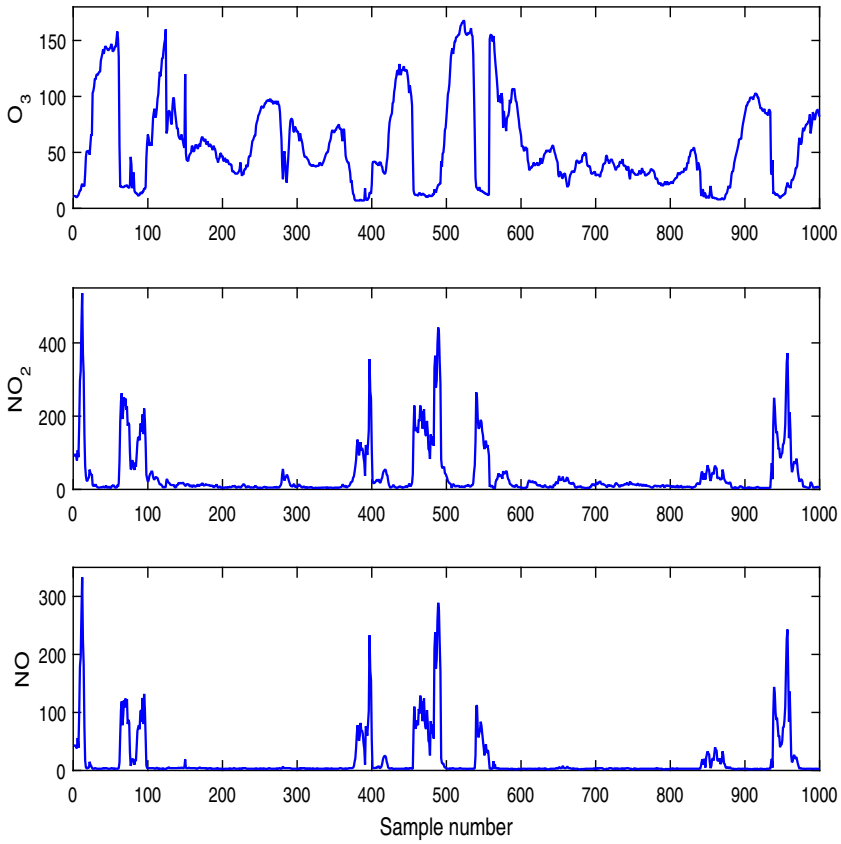

Fig. 4 Evolution of the measurements O_3, NO_2, and NO for the first station

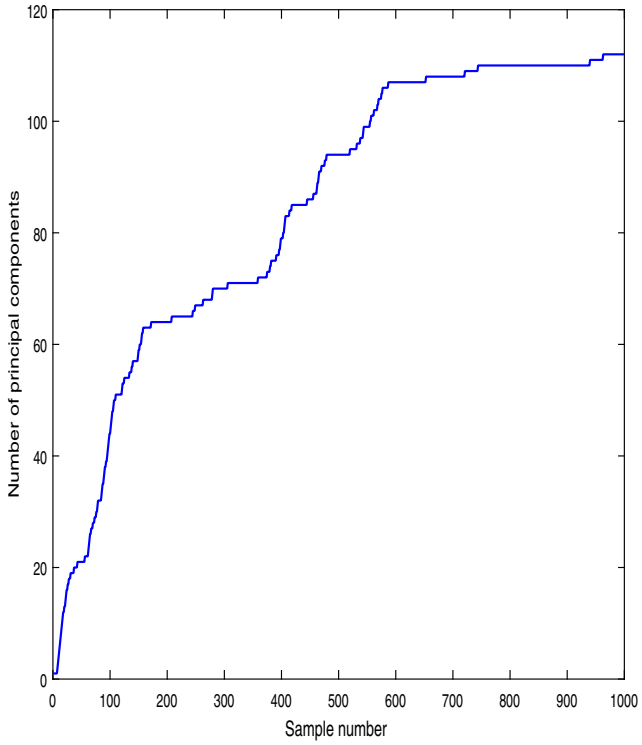

Fig. 6 Time evolution of the number of principal components

The outside air is pumped and brought to the analyzer which will measure its content of one or more specific pollutants. The air is drawn at about 3 m. The measurement is done continuously every quarter of an hour, 24 h a day and all year round. The measurement stations, spread over different geographic areas, are connected to a central computer via telephone line. The data is transmitted every day. They are validated manually, processed statistically, and analyzed before being transmitted to the media and a national database.

The purpose of data validation is to detect functioning abnormalities of the sensors principally those of the ozone concentration (O_3) and nitrogen oxides (NO and NO_2).

6.2 Sensor fault detection and localization

In the current work, only six neighbor measurement stations are considered. The dataset matrix X contains 18 variables, x_1 to x_{18} which correspond, respectively, to ozone O_3 and nitrogen oxides ($NO2$ and NO). In this study, a set of 1000 observations are generated by these six stations. In this paper, the tabu search algorithm is used to determine the optimal kernel parameter of the RBF function. The value of kernel parameter σ is equal to 41.5. Using the CPV criterion, the number of kernel principal components retained for the identification of the KPCA model is equal to 65. The detection indicator used is the SPE for a confidence limit equal to 95%.

The evolution of the measurements O_3, NO_2, and NO of the station 1 is presented in Fig. 4.

Using the OR-KPCA method, Fig. 5 presents the variation of the reduced number over time and Fig. 6 shows the number of principal components variation over time. As

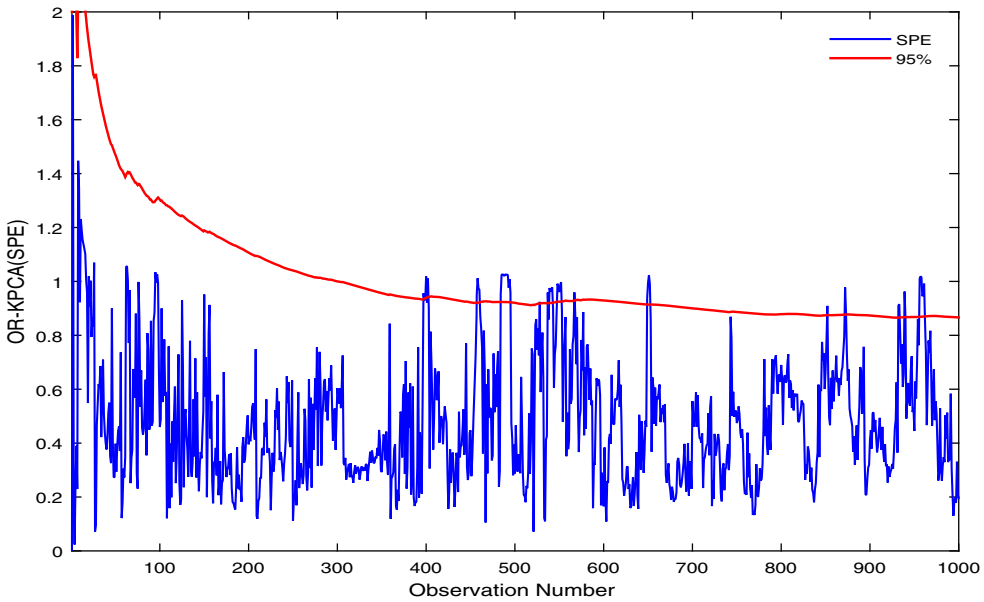

Fig. 7 Time evolution of the SPE index using the OR-KPCA method under normal operating condition

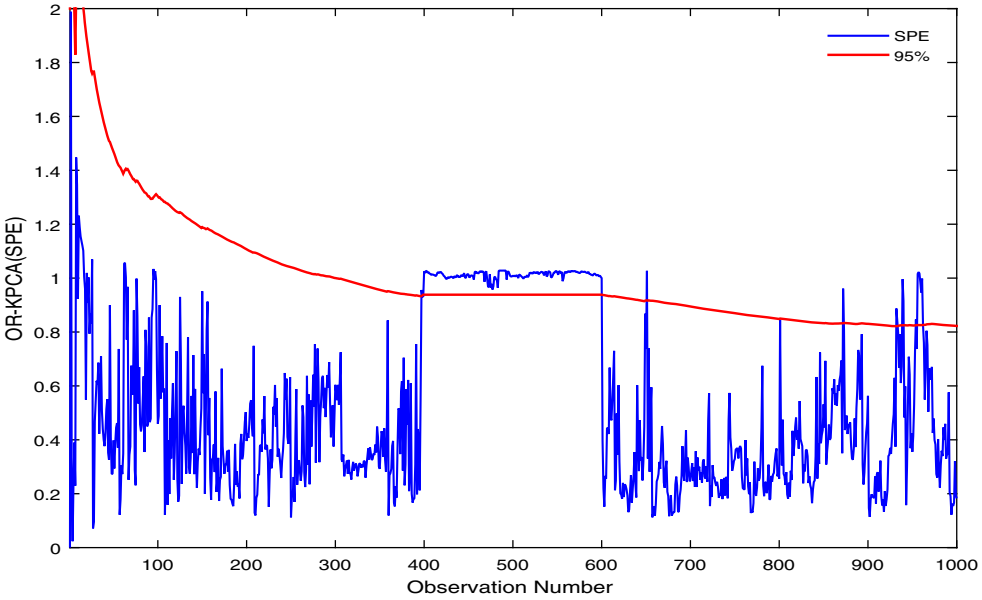

Fig. 8 Time evolution of the SPE index using the OR-KPCA method in the case of the fault on the variable x_2

shown in these figures, for a data set of 1000 observations, the number of reduced observations is equal to 144 which is equivalent to 14.40% of the number of initial data.

Figure 7 shows the fault detection results of the OR-KPCA using the SPE index under normal operating condition. For a control limit equal to 95%, it is shown from this figure that the SPE index present some false alarm rates.

To illustrate the OR-KPCA method, a bias fault is simulated on the variable x_2 between samples 400 and 600. The magnitude of the fault is equal to 35% of the range

Table 1 Incidence matrix for the AIRLOR

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}
r_1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
r_2	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
r_3	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
r_4	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
r_5	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
r_6	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
r_7	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1
r_8	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1
r_9	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
r_{10}	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1
r_{11}	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
r_{12}	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
r_{13}	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
r_{14}	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1
r_{15}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
r_{16}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1
r_{17}	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
r_{18}	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

of variation of this variable. Using the OR-KPCA method, the simulated fault is clearly detected with the SPE index between measurements 400 and 600 as shown in Fig. 8.

After the presence of faults has been detected, it is important to identify this fault. Therefore, we apply the proposed partial OR-KPCA method. The incidence matrix for the AIRLOR system is given in Table 1. According to this table, a set of 18 partial OR-KPCAs model can be constructed. Each model is insensitive to two variables.

The fault isolation results are shown in Fig. 9. The experimental signatures are obtained from Fig. 8. A SPE index greater than its control limit gives a "1" in the corresponding fault code, and a SPE index less than its control limit gives a "0." Thus, the experimental signature in the case of fault on variable x_2 is given by (0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1). To identify the fault, we compare between the experimental signature and the incidence matrix (Table 1). As can be seen from Fig. 9 that the simulation results showed complete agreement with the second column of the incidence matrix. Therefore, the variable x_2 is the faulty one.

The result of application of the proposed ESI-KPCA method to the AIRLOR process is shown in Figs. 10 and 11. These figures represent, respectively, the time evolution of the different indices $\text{SPE}_{-j}, ; j \in \{1, ..., m\}$ and the quantity Q_r. From Fig. 4, it is shown that the elimination of the variable x_2 makes it possible to obtain the index SPE_{-2} which is not affected by the fault. Therefore, the variable responsible for this out-of-control situation is x_2. Figure 11 shows that the variable x_2 has the smallest quantity Q_r compared to the other variables which indicates that the variable x_2 is the faulty variable.

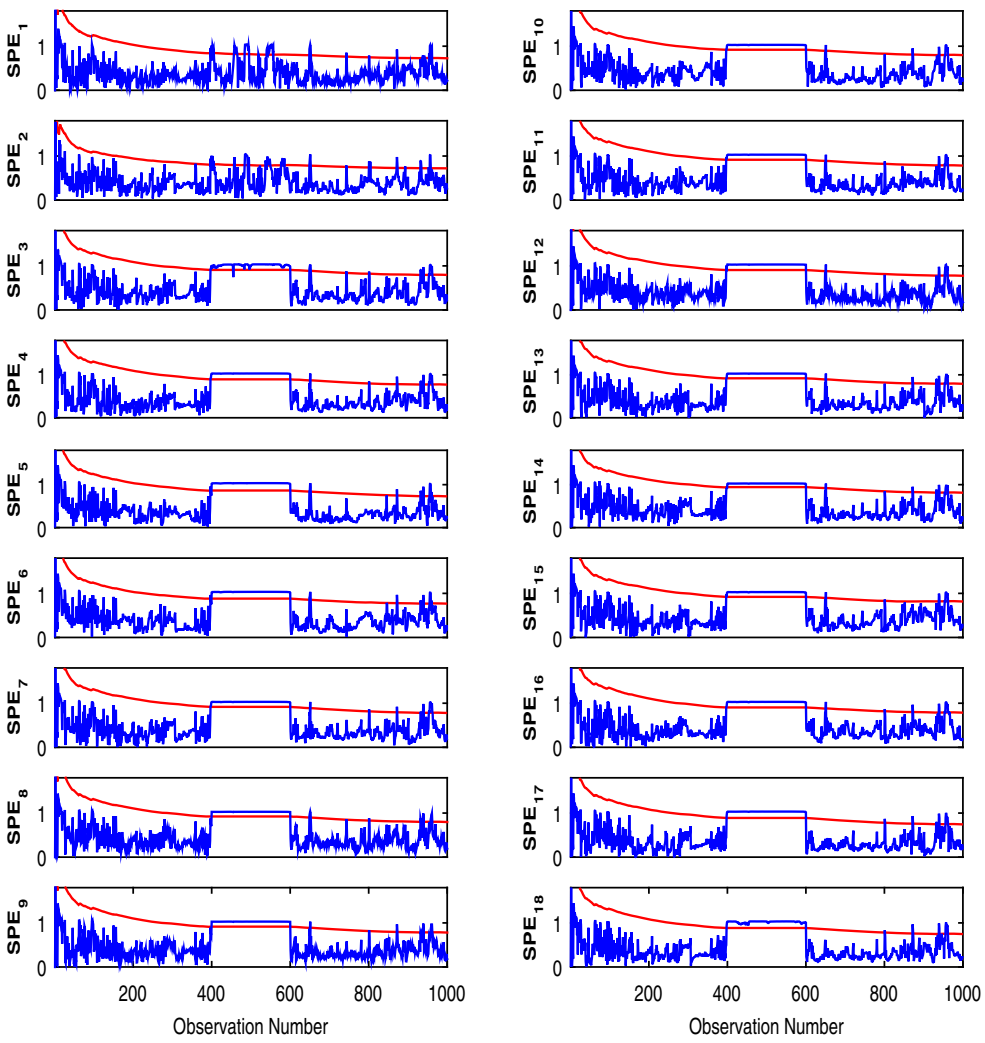

Fig. 9 Time evolution of different SPEs corresponding to the 18 partial OR-KPCA models in the case of the fault on the variable x_2

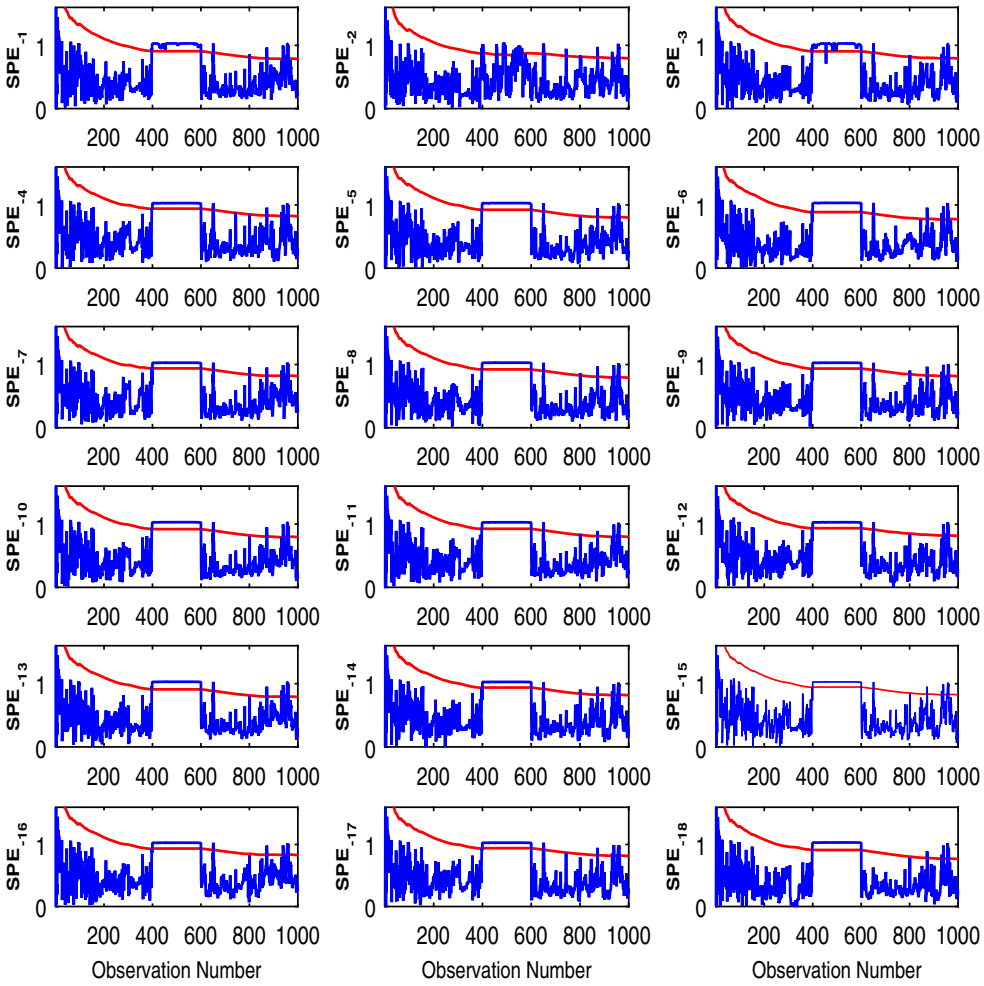

Fig. 10 The time evolution of the different SPEs obtained using the ESI-KPCA method with a fault on variable x_2

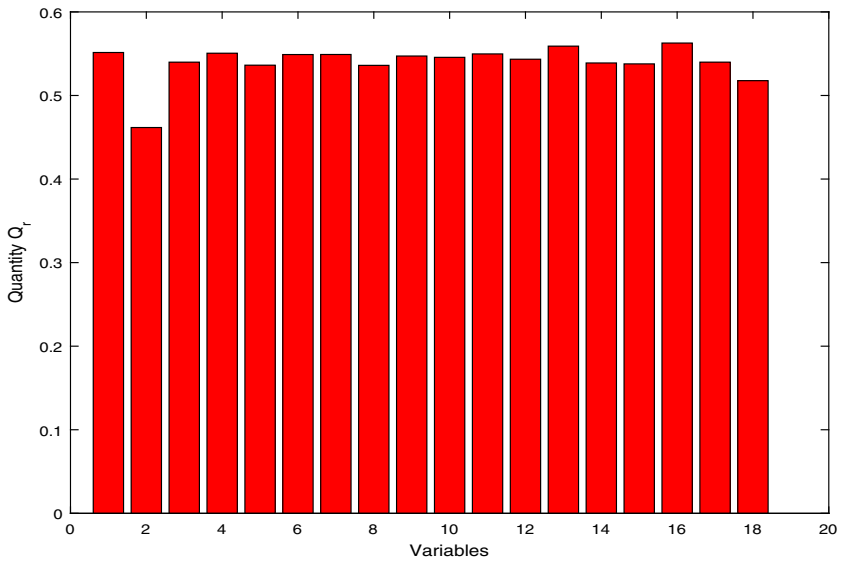

Fig. 11 Isolation of the faulty variable x_2 using the quantity Q_r

7 Conclusions

In this paper, we propose a new fault detection and isolation methods which are applicable to the process monitoring using OR-KPCA technique. Recently, the fault detection method using KPCA has been developed. In this paper, an online RKPCA (OR-KPCA) method with a tabu search algorithm are used in order to determine the optimal parameter of kernel function. However, the fault isolation scheme suitable for kernel PCA monitoring has rarely been found. The proposed partial OR-KPCA and the ESI-KPCA methods work like the standard partial PCA and the ESI in linear PCA, which does not require the fault direction to be known beforehand. The proposed partial OR-KPCA is a useful method for fault isolation. By performing OR-KPCA on subsets of variables, a set of structured residuals can be obtained. The structured residuals are used in composing a fault isolation scheme according to a properly designed incidence matrix. Simulation results showed complete agreement with the incidence matrix. An extension of the ESI method in the linear PCA to the nonlinear case was proposed in this paper, called ESI-KPCA. The ESI-KPCA consists, after detecting the presence of a fault with the SPE index, to eliminate each of the m variables sequentially from the set of variables and recalculate the SPE index. The fault isolation power of the proposed methods have been tested with real data of an air quality monitoring network process with sensor faults. It was showed that, the proposed methods were effective to correctly diagnose simple faults.

References

1. Qin SJ (2012) Survey on data driven industrial process monitoring and diagnosis. Ann Rev Control 36(2):220–234
2. Wang X, Kruger U, Irwin GW (2005) Process monitoring approach using fast moving window PCA. Ind Eng Chem Res 44(15):5691–5702
3. Sheriff MZ, Mansouri M, Karim MN, Nounou H, Nounou M (2017) Fault detection using multiscale PCA-based moving window GLRT. J Process Control 54:47–64
4. Wang H, Qian L, Dougherty E (2010) Inference of gene regulatory networks using S-system: a unified approach. IET Syst Biol 4(2):145–156
5. Ge Z, Song Z (2007) Process monitoring based on independent component analysis principal component analysis (ICA-PCA) and similarity factors. Ind Eng Chem Res 46(7):2054–2063
6. Lee JM, Qin SJ, Lee IB (2006) Fault detection and diagnosis based on modified independent component analysis. Aiche J 52(10):3501–3514
7. Kano M, Tanaka S, Hasebe S, Hashimoto I, Ohno H (2003) Monotoring independent components for fault detection. Aiche J 49(4):969–976
8. Chiang LH, Russell EL, Braatz RD (2000) Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemometr Intell Lab Syst 50(2):243–252
9. He QP, Qin SJ, Wang J (2005) A new faul diagnosis method using fault directions in Fisher discriminant analysis. Aiche J 51(2):555–571
10. Yu J (2011) Localized Fisher discriminant analysis based complex chemical process monitoring. Aiche J 57(7):1817–1828
11. Chen J, Liao CM (2002) Dynamic process fault monitoring based on neural network and PCA. J Process Control 12(2):277–289
12. Shijian Z, Yongmao X (2005) Multivariate staistical process monitoring using robust nonlinear principal component analysis. Tsinghua Sci Technol 10(5):582–586
13. Maulud A, Wang D, Romagnolti JA (2006) A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring. J Process Control 16(7):671–683
14. Sheriff MZ, Karim MN, Nounou MN, Nounou H, Mansouri M (2017) Monitoring of chemical processes using improved multiscale KPCA. In: 4th International Conference on Control, Decision and Information Technologies (CODIT). IEEE, pp 0049–0054)
15. Baklouti R, Mansouri M, Nounou H, Nounou M, Slima MB, Hamida AB (2017) Fault detection of chemical processes using KPCA-based GLRT technique. In: International conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, pp 1-6
16. Mansouri M, Nounou M, Nounou H, Karim N (2016) Kernel PCA-based GLRT for nonlinear fault detection of chemical processes. J Loss Prev Process Ind 40:334–347
17. Lee JM, Yoo C, Lee IB (2004) Fault detection of batch processes using multiway kernel principal component analysis. Comput Chem Eng 28(9):1837–1847
18. Choi SW, Lee C, Lee JM, Park JH, Lee IB (2005) Fault detection and identification of nonlinear processes based on kernel PCA. Chemometr Intell Lab Syst 75(1):55–67
19. Lee JM, Yoo C, Choi SW, Vanrolleghem PA, Lee IB (2004) Nonlinear process monitoring using kernel principal component analysis. Chem Eng Sci 59(1):223–234
20. Cho JH, Lee JM, Choi SW, Lee D, Lee IB (2005) Fault identification for process monitoring using kernel principal component analysis. Chem Eng Sci 60(1):279–288
21. Cui P, Li J, Wang G (2008) Improved kernel principal component analysis for fault detection. Expert Syst Appl 34(2):1210–1213
22. Galiaskarov MR, Kurkina VV, Rusinov LA (2017) Online diangnostics of time varying nonlinear chemical processes using moving window kernel principal component analysis and Fisher discriminant analysis. J Chemom 31(8):e2866
23. Khediri IB, Liman M, Weihs C (2011) Variable window adaptative kernel principal component analysis for nonlinear nonstationary process monitoring. Comput Ind Eng 61(3):437–446
24. Jaffel I, Taouali O, Harkat MF, Messaoud H (2017) Kernel principal component analysis with reduced complexity for nonlinear dynamic process monitoring. Int J Adv Manuf Technol 88(9-12):3265–3279
25. Fazai R, Taouali O, Harkat MF, Bouguila N (2016) A new fault detection method for nonlinear process monitoring. Int J Adv Manuf Technol 87(9-12):3425–3436
26. Jaffel I, Taouali O, Harkat MF, Messaoud H (2016) Moving window KPCA with reduced complexity for nonlinear dynamic process monitoring. ISA Trans 64:184–192
27. Fezai R, Mansouri M, Taouali O, Harkat MF, Bouguila N (2018) Online reduced kernel principal component analysis for process monitoring. J Process Control 61:1–11

28. Alcala CF, Qin SJ (2010) Reconstruction-based contribution for process monitoring with kernel principal component analysis. Ind Eng Chem Res 49(17):7849–7857
29. Kallas M, Mourot G, maquin D, Ragot J (2015) Fault estimation of nonlinear processes using kernel principal component analysis. In: control Conference (ECC), European, IEEE, pp 3197–3202)
30. Dunia R, Qin SJ, Edgar TF, McAovy TJ (1996) Identification of faulty sensors using principal component analysis. Aiche J 42(10):2797–2812
31. Gertler J, Cao J (2005) Desiggn of optimal strucured residuals from partial principal component models for fault diagnosis in linear systems. J Process Control 15(5):585–603
32. Huang Y, Gertler J, McAvoy TJ (2000) Sensor and actuator fault isolation by structured partial PCA with nonlinear extensions. J Process Control 10(5):459–469
33. Gertler J, Li W, Huang Y, McAvoy T (1999) Isolation enhanced principal component analysis. Aiche J 45(2):323–334
34. Huang Y, Gertler J, McAvoy TJ (1999) Fault isolation by partial PCA and partial NLPCA. IFAC Proc Vol 32(2):7647–7652
35. Said M, Fazai R, Adellafou KB, Taouali O (2018) Decentralized fault detection and isolation using bond graph and PCA methods. The International Journal of Advanced Manufaturing Technology, pp 1–13. https://doi.org/10.1007/s00170-018-2526-4
36. Stork CL, Veltkamp DJ, Kowalski BR (1997) Identification of multiple sensor disturbances during process monitoring. Anal Chem 69(24):5031–5036
37. Cristianini N, Shawe Taylor J (2000) An introduction to support vector machines and other kernel based learning methods. Cambridge university press, Cambridge
38. Taouali O, Jaffel I, Lahdhiri H, Harkat MF, Messouad H (2016) New fault detection method based on reduced kernel principal component analysis (RKPCA). Int J Adv Manuf Technol 85(5-8):1547–1552
39. Honeine P (2015) Approximation errors of online sparsification criteria, IEEE Trans. Signal Process 63(17):4700–4709
40. Honeine P (2012) Online kernel principal component analysis: reduced order model. IEEE Transactions on Patern Analysis and Machine Intelligence 34(9):1814–1826
41. Gao W, Chen J, Richard C, Huang J (2014) Online dictionary learning for kernel LMS. IEEE Trans Signal Process 62(11):2765–2777
42. Honeine P (2015) Analyzing sparse dictionaries for online learning with kernels. IEEE Trans Signal Process 63(23):6343–6353
43. Harkat MF, Mourot G, Ragot J (2006) An improved PCA scheme for sensor FDI: application to an air quality monitoring network. J Process Control 16(6):625–634
44. Harkat MF, Mourot G, Ragot J (2009) Multiple sensor fault detection and isolation of an air quality monitoring network using RBF-NLPCA model. IFAC Proc Vol 42(8):828–833
45. Downs JJ, Vogel EF (1993) A plant wide industrial process control problem. Comput Chem Eng 17(3):245–255
46. Rato T, Reis M, Schmitt E, Hubert M, De Ketelare B (2016) A systematic comparison of PCA based Statistical Process Monitoring methods for high-dimensional, time dependent Processes. AIChE J 62(5):1478–1493
47. Harkat MF, Mansouri M, Nounou M, Nounou H (2018) Enhanced data validation strategy of air quality monitoring network. Environ Res 160:183–194