# Data-driven approach for fault detection and isolation in nonlinear system

Maya Kallas, Gilles Mourot, Didier Maquin, José Ragot

# Data-driven approach for fault detection and isolation in nonlinear system

Maya Kallas | Gilles Mourot | Didier Maquin | José Ragot

Centre de Recherche en Automatique de Nancy, Université de Lorraine, CNRS, 2 Avenue de la Forêt de Haye, TSA 60604, 54518 Vandoeuvre-lès-Nancy, France

**Correspondence**
José Ragot, Centre de Recherche en Automatique de Nancy, Université de Lorraine, CNRS, 2 Avenue de la Forêt de Haye, TSA 60604, 54518 Vandoeuvre-lès-Nancy, France.
Email: Jose.ragot@univ-lorraine.fr

**Summary**

System diagnosis has been of a great interest for all aspects of industrial processes more precisely to gain in quality. It is based essentially on the analysis of the links between the variables of a system and more precisely on the changes of the relations between these variables, which testify to the presence of faults or anomalies. For that purpose, data modeling is the process of finding a mathematical expression that provides a good fit between given finite sample values of the independent variables and the associated values of the dependent variables of the process.

The aim of this paper is to detect and, above all, localize faults affecting a system with nonlinear behavior, when its model is not known a priori. An important part of the presentation is dedicated to the construction of fault indicators capable of locating faults, ie, recognizing the input or output of a system affected by a fault. The first part of this paper is devoted to how to predict the output of a nonlinear behavior system. The second part proposes a way for the detection and isolation of measurement faults based on the proposed prediction model. The relevance of the proposed technique, for modeling and system diagnosis, is illustrated on a simulated example in the context of SIMO and MIMO systems.

**KEYWORDS**

data-driven approach, fault detection, fault isolation, kernel, nonlinear systems

## 1 | INTRODUCTION

With the emergence of high performance industrial processes, the systems become more complicated. The need of important gain in products requires the supervision of these systems. Systems supervision covers many aspects, including the diagnosis of these systems. The latter is achieved by many techniques to detect, isolate, and identify the presence of faults on sensors, actuators, and components affecting the operating system, as shown in the work of Qin.[1] The sooner the fault is detected, the sooner a recovery can be applied. Many techniques have been proposed in literature, some are based on model predictions, whereas others on data-driven approaches.

In this paper, we consider the second case, where the behavior model of the system is not a priori known and must then be built from measurements made on the system. The case of linear systems has been studied intensively for fault detection and isolation (FDI) and, in particular, by PCA-type techniques, as in the works of Harkat et al[2] and Benaicha et al.[3] In the works of Samuel and Cao[4] and Fazai et al,[5] the authors addressed the case of nonlinear systems using KPCA but only for fault detection.

In all FDI approaches, the main idea is the use of deviations between the real and the estimated values for fault detection, whereas the isolation is frequently done using the technique of structured residuals. With the nonlinear aspect of the systems nowadays, new techniques have been proposed in literature, such as the use of a nonlinear transformation to describe the nonlinearities between data. Considering most of the literature, it is clear that the generally treated aspect, for the nonlinear case, is the fault detection. A much more limited number of communications have been dedicated to the location or isolation of detected faults affecting a variable. In the works of Alcala and Qin[6] and Kallas et al,[7] the pre-image technique offers an isolation step and estimating faults. The problem is due to the nonlinear transformation that induces some difficulties in the identification procedure, which is only partially resolved by using a regularization function.

For the linear case, the relationship between the input and the output can be obtained using a least squares estimator. Extension to the nonlinear case using nonlinear regressions has also been proposed. In the work of Bates and Watts,[8] the authors presented the nonlinear regression with the possible applications in chemical engineering. Smyth[9] defined in his book the nonlinear regression by giving the most common models and data transformation for converting nonlinear models into linear ones. Kung[10] provided also a detailed study on the nonlinear case with a focus on kernel-based learning theory. A study has been made in the nonlinear case for time prediction using kernel methods by Kini and Sekhar,[11] whereas, in the work of Caswell,[12] the author proposed a prediction of geomagnetic fluctuations. Another method consists to estimate the output based on previous output estimation and with optimal test signal; this technique has been applied on a discrete nonlinear three-input MISO model for the brake torque output of an engine.[13]

Up to the authors' knowledge, the use of a kernel approach for fault diagnosis adapted to real processes is barely exposed in the literature.[14] On the contrary, the literature often has applications only with simulations or laboratory processes. In related works,[5,15-17] fault detection on a chemical processes is proposed; in the work of Schwantes and Pande,[18] for estimating some characteristics, molecular dynamics is proposed; in the work of Park et al,[19] fault diagnosis for induction machines is proposed; in other works,[20-22] an application to water treatment is presented; and in the work of Navi et al,[23] an industrial gas turbine is considered.

In the work of Exterkate et al,[24,25] the authors extended the kernel ridge regression methodology to enable its use for economic time series forecasting, by including lags of the dependent variable; moreover, they evaluated the potential of kernel ridge regression when many predictor variables are present. In the work of Chaouch,[26] FDI of ECG process is performed. Various fields in environmental sciences use learning methods based on kernel (see, for example, the works of Pulkkinen[27] and Gilbert et al[28]).

Generally, in most of the published papers, squared prediction error (*SPE*) and Hotelling's $T^2$ as two kinds of statistics are used to monitor whether or not a fault occurs in the process. The fault detection logic uses a threshold for *SPE* and $T^2$ and provides correct results in detecting the presence of faults but is often not very effective in terms of fault isolation, this is essentially due to the very global nature of these tests. These tests globally analyze a function of the corrective terms taken together, which weakens the contribution of the variable really carrying the fault. However, if these tests are limited to a subset of variables, evaluation may be improved and more efficient for fault isolation.

In this paper, we present a model for the prediction of the output of a nonlinear system using a Kernel formulation. Once this prediction model is identified using reference and training data, a novel technique for diagnosis is presented. The main idea is to create a set of nonlinear redundancy equations to isolate a fault and define its location (the input or the output or precisely one in case of multiple inputs or outputs). The remaining of this paper is organized as follows. Section 2 is devoted to the output prediction of a system; firstly, considering the linear case and, secondly, the nonlinear case by using a kernel formulation. Section 3 is dedicated to the detection and isolation of faults affecting input or output. Finally, Section 4 presents the implementation of the proposed technique on a simulated data set, for output prediction and diagnosis to isolate some added faults.

The main contribution of this paper returns to the use of a kernel approach for fault diagnosis and mainly fault isolation. If there are a number of publications on the detection of faults,[5,29-33] to the authors' knowledge, there are a very few number on the location of these faults[4,34] in which kernel density estimation technique was used to estimate an upper control limit more robust than those based on the Gaussian assumption. It is precisely this point of fault location that is addressed here by proposing a structuring of the fault-indicating residues thanks to a judicious choice of the redescription variables for the construction of kernel-based regression models.

## 2 | PREDICTION USING NON LINEAR APPROACH

Time series prediction guide decisions in a variety of fields and mainly for control and diagnosis. For the last point of view, obtaining useful predictions is the key of fault detection because it will be possible to compare the prediction of a variable with its realization when available at a next time. Prediction generally starts from a sequence of past observations of the

input and the output of the considered process. To this end, it is used as a system's identification technique by finding the relation between the input and the output of the system. In a general form, let us consider a data set of variables $(\boldsymbol{x}_k, y_k)$, with $k = 1, \ldots, N$, where $\boldsymbol{x}_k \in \mathcal{R}^s$ and $y_k \in \mathcal{R}$. Each variable $y_k$ can be explained by variables $\boldsymbol{x}_k$ using a nonlinear regression technique.

## 2.1 | Prediction using linear dependency

In its simplest form, prediction model can be defined as a linear regression described as $y_k = < \boldsymbol{\theta}, \boldsymbol{x}_k >$, where $\boldsymbol{\theta}$ and $\boldsymbol{x}_k \in \mathcal{R}^s$, and $< ., . >$ is the dot product in $\mathcal{R}^s$. The estimation of $\boldsymbol{\theta}$ can be obtained by different techniques, the simplest approach is to minimize with regard to $\boldsymbol{\theta}$ the following criterion:

$$\Phi(\boldsymbol{\theta}) = \frac{1}{2}\sum_{k=1}^{N}\left(y_k - \boldsymbol{\theta}^T\boldsymbol{x}_k\right)^2 + \frac{\gamma}{2}\|\boldsymbol{\theta}\|^2, \tag{1}$$

where $\gamma$ is a positive term used to control the compromise between the two terms of the criterion. The first term quantifies the similarity between the estimated and measured outputs, whereas the second term provides more regular solutions, based on the regularization applied. The choice of the regularization parameter $\gamma$ controls the balance between training error and degree of regularity. As well known, the ridge regression reduces the variability between the coefficients by shrinking them; as result, a more prediction accuracy is obtained at the cost of only a small increase of bias in the estimated parameters. In this paper, we will not seek for an optimal value of $\gamma$; our goal is essentially dedicated to the detection of fault.

Let us define the following:

$$\begin{aligned} \boldsymbol{y} &= [y_1, \ldots, y_N]^T \quad \in \mathcal{R}^N \\ \boldsymbol{X} &= [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^T \quad \in \mathcal{R}^{N\times s}. \end{aligned} \tag{2}$$

The estimation that minimizes (1) is given by

$$\hat{\boldsymbol{\theta}} = (\gamma\boldsymbol{I} + \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{3}$$

Taking into account that $(a\boldsymbol{I} + \boldsymbol{A}^T\boldsymbol{B})^{-1}\boldsymbol{A}^T = \boldsymbol{A}^T(a\boldsymbol{I} + \boldsymbol{B}\boldsymbol{A}^T)^{-1}$ for a scalar $a$ and matrices $\boldsymbol{A}, \boldsymbol{B}$ with appropriate dimensions, (3) may also be expressed as follows:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{X}^T(\gamma\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y}. \tag{4}$$

Consequently, the estimation of $\boldsymbol{y}$ can be expressed as follows:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{X}^T(\gamma\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y}. \tag{5}$$

In this expression, it is important to note that the estimates are evaluated as dot products between data. This remark will be widely exploited in the following for the establishment of the prediction models for nonlinear systems. It should be noted that (5) needs the calculation of the inverse of a matrix of dimension $N$, the latter however being evaluated by a recursive technique.

This relationship can also be applied to any new observation $\boldsymbol{x}^{new}$, for which the estimated output value can be established using the previously defined model (we therefore implicitly assume that the system has invariant parameters, ie, that the new observation is made when the system behaves no differently from the one used to estimate its parameters)

$$\hat{y}^{new} = (\boldsymbol{x}^{new})^T\boldsymbol{X}^T(\gamma\boldsymbol{I} + \boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{y}. \tag{6}$$

Based on the so-called ridge regression, expression (6) allows to predict the output of a system using its inputs. However, this technique fails to estimate the relationship between the inputs and outputs of nonlinear systems.

*Remark* 1. Interest of the ridge regression.

As it is well known, ridge regression reduces significantly the effects of multicollinearity of the variables. However, the obtained estimate is sensitive to data when corrupted by outliers and leads to unreliable results for prediction. This drawback can be minimized by combining M-estimation and ridge regression, as proposed in the works of Wibowo[35] and Ertaş et al.[36]

## 2.2 | Output prediction of MISO system using nonlinear dependency

Due to the fact that expressions (5) and (6) are written using a dot product between data, it is worth noting that the extension to the nonlinear case can be done by replacing this product by a nonlinear function according to the use of kernel. Therefore, the case of nonlinear MISO system with inputs $u_i, i = 1, \dots, m$ and output $y$ is considered using this technique. To this end, a data set of variables $(\boldsymbol{x}_k, y_k)$, for $k = r+1, \dots, N$ is used, where $r = \max(p, q)$ and

$$\boldsymbol{x}_k = \left[y_{k-1:k-p}, u_{1,k-1:k-q}, \dots, u_{m,k-1:k-q}\right]^T \quad \in \mathcal{R}^{m+m.q} \tag{7}$$

regroups the last $p$ output measures and the last $q$ measures of each input. In all what follows, the notation $z_{k_1:k_2}$ means the collection of data $z$ between indexes $k_1$ and $k_2$. To deal with a nonlinear representation, it is a common practice to map the original input data $\boldsymbol{x}_k$ into another feature space. For that purpose, using a nonlinear function $\varphi$, we consider the projection of the original data

$$\boldsymbol{x}_k \in \mathcal{R}^{p+m.q} \quad \rightarrow \quad \varphi(\boldsymbol{x}_k) \in \mathcal{R}^r. \tag{8}$$

A new data matrix is therefore considered

$$\boldsymbol{X} = [\varphi(\boldsymbol{x}_1), \dots, \varphi(\boldsymbol{x}_N)]^T \quad \in \mathcal{R}^{N \times r}, \tag{9}$$

and expression (4) then recalculates the model parameters based on the coordinate change (8). It is remarkable to note that this calculation can be done without ever knowing the coordinates in this new space, but rather by simply calculating the scalar products of the images of the data in this space. This operation is known as the trick of the kernel. In this case, we consider a kernel $\boldsymbol{K}$ that describes these dot products using their corresponding feature vectors. This matrix $\boldsymbol{K}$ is thus defined by the elements $[\boldsymbol{K}]_{i,j} = < \varphi(\boldsymbol{x}_i), \varphi(\boldsymbol{x}_j) >$. To this end, (4) can be written as follows:

$$\hat{\boldsymbol{y}} = \boldsymbol{K}(\gamma I + \boldsymbol{K})^{-1}\boldsymbol{y}. \tag{10}$$

*Remark* 2. Dimension of the Gram matrix $\boldsymbol{K}$.

In practice, a large data set leads to a large $\boldsymbol{K}$, and storing $\boldsymbol{K}$ may become a problem. One way to deal with this is to perform clustering on the data set and populate the kernel with the means of those clusters. Moreover, to control the model order $p$ and $q$, see, for example, the works of Honeine[37] or Drineas and Mahoney[38] and Hussein and Abd-Almageed[39] to construct a sparse matrix approximation to Gram matrix $\boldsymbol{K}$.

Therefore, for a new observation $\boldsymbol{x}^{new}$, the predicted output using the formulation (6) can be written as follows:

$$\hat{\boldsymbol{y}}^{new} = \kappa^T(\boldsymbol{x}^{new})(\gamma \boldsymbol{I} + \boldsymbol{K})^{-1}\boldsymbol{y}, \tag{11}$$

with the following definitions:

$$\boldsymbol{\kappa}(\boldsymbol{x}^{new}) = \begin{bmatrix} < \varphi(\boldsymbol{x}_{r+1}), \varphi(\boldsymbol{x}^{new}) > \\ \vdots \\ < \varphi(\boldsymbol{x}_{r+N}), \varphi(\boldsymbol{x}^{new}) > \end{bmatrix}$$

$$\boldsymbol{K} = \begin{bmatrix} < \varphi(\boldsymbol{x}_{r+1}), \varphi(\boldsymbol{x}_{r+1}) > & \cdots & < \varphi(\boldsymbol{x}_{r+1}), \varphi(\boldsymbol{x}_{r+N}) > \\ < \varphi(\boldsymbol{x}_{r+2}), \varphi(\boldsymbol{x}_{r+1}) > & \cdots & < \varphi(\boldsymbol{x}_{r+2}), \varphi(\boldsymbol{x}_{r+N}) > \\ \vdots & & \\ < \varphi(\boldsymbol{x}_{r+N}), \varphi(\boldsymbol{x}_{r+1}) > & \cdots & < \varphi(\boldsymbol{x}_{r+N}), \varphi(\boldsymbol{x}_{r+N}) > \end{bmatrix}.$$

It is important to again mention that $\varphi(.)$ is not explicitly given; however, the kernel function $\boldsymbol{K}(.,.)$ may perfectly describe the nonlinear relations between data, according to a judicious choice of the scalar product in $\boldsymbol{K}(.,.)$ because some kernel have the capability of universal approximation.[40,41] Many nonlinear functions can be used to define this dot product between data.[42] There is no theoretical works about the comparison of the performances of the difference kernels; however, there is a lot of works for this comparison in specific domains and specific applications.[43,44] In this paper, the radial basis function is used (but there exist other choices such as quadratic, polynomial, hyperbolic tangent kernels, sigmoïd), taking into consideration the distance (or similarity) between data. It is given by

$$\kappa(\boldsymbol{x}_k, \boldsymbol{x}_\ell) = \exp\left(-\frac{1}{c}\|\boldsymbol{x}_k - \boldsymbol{x}_\ell\|^2\right) \tag{12}$$

where parameter $c$ controls the sensitivity of the kernel function in respect the observation $\boldsymbol{x}_{(.)}$ (smaller values of $c$ will cause the function to overfit the data points, whereas larger values will cause it to underfit). Although choice of the range of used in the Gaussian kernel is still an open problem,[45] we may define some reasonable values for $c$. In Gaussian kernel, if the value of $c$ is very small, then $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is close to zero. If the value of $c$ is very large, then $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is close to 1. Consequently, we can identify from some numerical tests a lower bound and an upper bound for $c$ and then select $c$ between these two values.

*Remark* 3. Weighted quadratic norm.

Aside from the two-norm used in definition (9), the most useful norm is the weighted two-norm where each of the coordinates of a vector space is given its own weight. Thus, the quadratic term in (9) is replaced by $\|\boldsymbol{x}_k - \boldsymbol{x}_\ell\|^2_{\boldsymbol{W}}$, where $\boldsymbol{W}$ can be chosen as a diagonal matrix, which makes it possible to modulate the importance given to certain components of $\boldsymbol{x}_k - \boldsymbol{x}_\ell$. This is particularly important if the domains of variation of the variables are very different from each other.

*Remark* 4. Redescription's variables.

In definition (12), using (7), we can notice that

$$\|\boldsymbol{x}_k - \boldsymbol{x}_\ell\|^2 = \sum_{i=1}^{p} (y_{k-i} - y_{\ell-i})^2 + \sum_{i=1}^{q} (u_{k-i} - u_{\ell-i})^2, \tag{13}$$

and that the expression obtained introduces nonlinearities on the variables $y$ and $u$ independent of each other. For systems with pronounced nonlinear behavior, it may be desirable to take into account this characteristic by means of redescription variables. For example,

$$\boldsymbol{x}_k = [y_{k-1},\ y_{k-1}y_{k-2},\ y_{k-2}u_{k-1}]^T \tag{14}$$

introduces more inter and intracoupling between the input $u$ and the output $y$ of a SISO system when computing $\|\boldsymbol{x}_k - \boldsymbol{x}_\ell\|^2$.

*Remark* 5. Methods for identifying model parameters.

As can be seen from the expression of $\boldsymbol{x}_k$ in (7), the parametric identification technique has been defined using the output and input measurements $y$ of a system (1) and what is commonly called the equation error. It is well known that this identification technique can give a biased estimate if the system output is affected by white noise. However, there exists another method of system's identification, namely, the output error model, which is based on the output prediction and the input.[46,47]

In summary, the prediction of the output of a nonlinear system from a new observation $\boldsymbol{x}^{new}$ is explained as follows:

$$\hat{y}^{new} = \kappa^T(\boldsymbol{x}^{new})\boldsymbol{\beta}, \tag{15}$$

where $\boldsymbol{\beta}$ is a vector containing the influence of the training data set and is defined by

$$\boldsymbol{\beta} = (\gamma \boldsymbol{I} + \boldsymbol{K})^{-1} \boldsymbol{y}. \tag{16}$$

If the variable $y^{new}$ is known, it is then possible to compare the estimate $\hat{y}^{new}$ to the measurement $y^{new}$ to identify the presence of a fault.

Given the wording (15), the implicit assumption that is made is that the new observation $\boldsymbol{x}^{new}$ belongs to the same operating domain as that used to determine the nonlinear prediction model parameters $\boldsymbol{\beta}$. In other words, the parameters of this model are assumed to e adequate to represent the system operation for this new observation. Finally, the prediction that is made is more of an interpolation than an extrapolation for which few guarantees could be given.

## 2.3 | Output prediction of MIMO system's output

The first phase of diagnosis is to identify the system using healthy data and provide an output prediction model, whereas the second phase is the detection and isolation of faults using the identified model. Here, the proposed technique for output prediction is applied in the case of a multiple input output system. As the physical model of the system is not available, the structure of the prediction models are a priori unknown, but we can hypothetically create, for each

output $y_{j,k}(j = 1, \dots, n, \quad k = r, \dots, N)$, dependency relations given the prediction at time instant $k + 1$ according to measurements available at time instant $k$, so, for each of the $n$ system outputs,

$$\begin{aligned} \hat{y}_{1,k+1} &= f_1(\boldsymbol{x}_{1,k}) \\ \hat{y}_{2,k+1} &= f_2(\boldsymbol{x}_{2,k}) \\ &\vdots \\ \hat{y}_{n,k+1} &= f_n(\boldsymbol{x}_{n,k}), \end{aligned} \tag{17}$$

where $f_1, \dots, f_n$ will be obtained with the approach of Section 2.2 and where the vectors of explanatory variables are defined as follows:

$$\begin{aligned} \boldsymbol{x}_{1,k} &= [\, y_{1,k:k-p}, \quad u_{1,k:k-q}, \quad \dots, \quad u_{m,k:k-q} \,]^T \\ \boldsymbol{x}_{2,k} &= [\, y_{2,k:k-p}, \quad u_{1,k:k-q}, \quad \dots, \quad u_{m,k:k-q} \,]^T \\ &\vdots \\ \boldsymbol{x}_{n,k} &= [\, y_{n,k:k-p}, \quad u_{1,k:k-q}, \quad \dots, \quad u_{m,k:k-q} \,]^T, \end{aligned} \tag{18}$$

in which $p$ and $q$ are the orders of the nonlinear ARMA process. In view of (17), (18), we assumed that each output depended on all inputs, which is the most general situation. In the future, by comparing them to measurements when they become available, predictions (17) will generate fault occurrence indicators. The presence of a fault must be completed by a fault location phase, which requires dedicated indicators for this task. Those indicators are designed in such a way that each sensitive to some faults while insensitive to other faults. For this, it is necessary to generate system output predictions using particular subsets of variables.

For that purpose, due to the fact that both outputs (17) depend on the same inputs $u_{1,i:i-p}, \dots, u_{n,i:i-p}$(18), we can explore a new relation between the outputs independent from one particular input or a subset of inputs; this relation being known as an interredundancy. The interest of such a relationship appears clearly in the perspective of the diagnosis where the objective is to locate the presence of a fault affecting a particular variable. As model (17) is definitely not known, therefore, it is impossible to analytically eliminate a particular input $u_\ell$ (for example) between two given outputs with a classical technique. However, we can hypothetically create a dependency function between these two outputs independent of $u_\ell$ but depending the other input $u_1, \dots, u_{\ell-1}, u_{\ell+1}, \dots, u_n$. In fact, to facilitate fault isolation, it is advisable to show, if possible, only one input per model. In the case of a MIMO system with as many inputs as outputs ($n = m$), we can therefore construct $n$ models, each predicting an output according to the other outputs and a single input

$$\begin{aligned} \hat{y}_{n+1,k+1} &= f_{n+1}(\boldsymbol{x}_{1,k}) \\ \hat{y}_{n+2,k+1} &= f_{n+2}(\boldsymbol{x}_{2,k}) \\ &\vdots \\ \hat{y}_{2n,k+1} &= f_{2n}(\boldsymbol{x}_{n,k}), \end{aligned} \tag{19}$$

where $f_{n+1}, \dots, f_{2n}$ will be obtained with the approach of Section 2.2 and where the new vector of explanatory variables is defined in the following way:

$$\begin{aligned} \boldsymbol{x}_{n+1,k} &= [\, y_{1,k:k-p}, \quad \dots, \quad y_{n,k:k-p}, \quad u_{1,k:k-q} \,]^T \\ \boldsymbol{x}_{n+2,k} &= [\, y_{1,k:k-p}, \quad \dots, \quad y_{n,k:k-p}, \quad u_{2,k:k-q} \,]^T \\ &\vdots \\ \boldsymbol{x}_{2n,k} &= [\, y_{1,k:k-p}, \quad \dots, \quad y_{n,k:k-p}, \quad u_{n,k:k-q} \,]^T. \end{aligned} \tag{20}$$

*Remark* 6. About elimination of entries.

The particular situation $n \neq m$ is treated in a similar way, knowing that the number of inputs that can potentially be eliminated depends on the number of inputs compared to the number of outputs. In case $n \geq m + 1$, it is possible to formally eliminate all $u_i$ entries in Equations (17) and thus obtain output predictions independent of inputs.

Therefore, the idea is to identify the parameters of the $2n$ structures $f_i$ aforementioned (17) and (19) to generate some indicators that can be further use to detect the variables affected by faults. Identification of the model's parameters $f_i, i = 1, \dots, 2n$ is performed using classical least square method that we had completed with a cross-validation technique. Cross-validation is one of the most commonly used methods of evaluating predictive performances of a model. Basically,

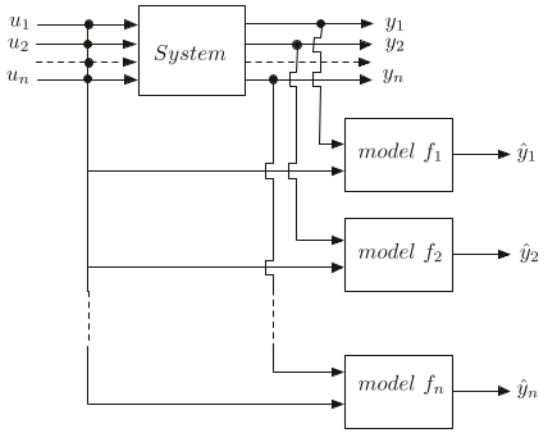


**FIGURE 1** Primary models

part of the available data is used for fitting each competing model and the rest of the data is used to measure the predictive performances of the models by the validation errors, and the model with the best overall performance is the one that is kept. Thus, we use two data sets, ie, the first one for identification and the second for training. The role of these two data sets is then swapped.

For cross validation purposes $2n$ variables $\boldsymbol{x}^t_{j,k}$, $k = r+1, \cdots\cdots, r+M$, of the same structure as those defined in (18) and (20), are used for the training data set

$$\begin{aligned}
\boldsymbol{x}^t_{j,k} &= \left[\, y^t_{j,k:k-p}, \quad u^t_{1,k:k-q}, \quad \ldots, \quad u^t_{m,k:k-q} \,\right]^T, \quad j = 1, \ldots, n \\
\boldsymbol{x}^t_{j,k} &= \left[\, y^t_{1,k:k-p}, \quad \ldots, \quad y^t_{n,k:k-p}, \quad u^t_{j,k:k-q} \,\right]^T, \quad j = n+1, \ldots, 2n.
\end{aligned} \tag{21}$$

Using these two sets of variables $\boldsymbol{x}_{j,k}$ and $\boldsymbol{x}^t_{j,k}$, we can estimate the $\boldsymbol{\beta}'s$ parameters (16) and then generate prediction models for the $n$ outputs of the two model structures as follows:

$$\begin{aligned}
\hat{y}_{i,k+1} &= \sum_{j=r+1}^{r+N} \beta_{i,j-r} \kappa_i \left( \boldsymbol{x}_{i,j}, \boldsymbol{x}^t_{i,k} \right), \quad i = 1, \ldots, n \\
\hat{y}_{i,k+1} &= \sum_{j=r+1}^{r+N} \beta_{i,j-r} \kappa_i \left( \boldsymbol{x}_{i,j}, \boldsymbol{x}^t_{i,k} \right), \quad i = n+1, \ldots, 2n.
\end{aligned} \tag{22}$$

Thus, $2n$ sets of coefficients $\boldsymbol{\beta}_{i,.}$ where $i = 1, \ldots, 2n$, are defined, one for each prediction model (17) and (19) and identified with a classical least square technique. Then, the quality of the resulting model is tested from the gaps between prediction and measurements.

To complete the cross-validation procedure, the role of the two databases $\boldsymbol{x}_{j,k}$ and $\boldsymbol{x}^t_{j,k}$ is swapped, the model parameters are re-identified and the predictions are evaluated in a similar way to (17) and (19). Finally, the model chosen is the one that provides the closest prediction of the measurements. It is important to remember that the predictions $\hat{y}_{i,k+1}, i = 1, \ldots, n$ on the one hand and $\hat{y}_{i,k+1}, i = n+1, \ldots, 2n$ on the other hand use sets of partially different variables. Finally, Figures 1 and 2 summarize how to generate primary models that predict each output based on inputs and structured secondary models that take advantage of interredundancies by explaining the prediction of each output based on other outputs and an input.

## 3 | DIAGNOSIS STRATEGY

The benefit of having a prediction model is to generate an indicator that compares each output to its predicted value. This indicator is sensitive to both the input and output faults of the system. The detection part can be done easily, however, the isolation of a fault on the input or on the output is somehow difficult. The isolation can be applied on a system with different outputs to have some redundancies. In order to satisfy fault isolation, FDI systems are based on structured residues, where some residues are sensitive only to a subset of faults and where the whole set of residues can isolate the

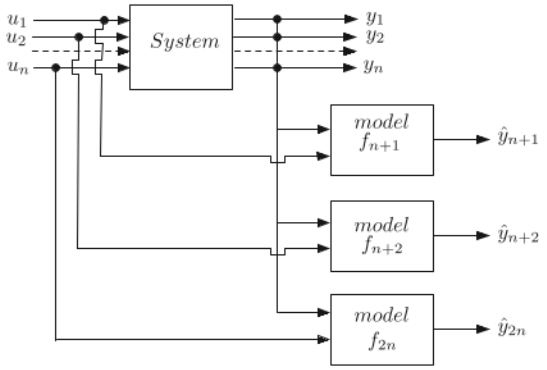


**FIGURE 2** Structured secondary models

**TABLE 1** Fault signatures

| | $\delta u_1$ | $\delta u_2$ | … | $\delta u_n$ | $\delta y_1$ | $\delta y_2$ | … | $\delta y_n$ |
|---|---|---|---|---|---|---|---|---|
| $f_1$ | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| $f_2$ | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| … | | | | | | | | |
| $f_n$ | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| $f_{n+1}$ | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| $f_{n+2}$ | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| … | | | | | | | | |
| $f_{2n}$ | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |

set of faults. Therefore, using predictions (17) and (19), different residuals are generated and they can help to detect and isolate faults. At each time instant $k$, these residuals are evaluated numerically according to

$$\tilde{y}_{i,k} = y_{i,k} - \hat{y}_{i,k}, \quad i = 1, \ldots, 2n \tag{23}$$

and are compared, at each time instant, to a threshold to trigger a fault alarm, ie, what constitutes the experimental signature of the faults.

On another side, based on Equations (17), (19) with definitions (18) and (20), the theoretical signature's table can be established. It takes into consideration the influence of a fault than can affect the input or one of the outputs. It is given by Table 1 where a "one" would typically correspond to an observed symptom of a fault, whereas a "zero" would indicate normal behavior of the system. A fault signature (not indicated in this table) with all zeros would represent normal functioning of the system. Another way to read this table is to consider that each column corresponds to a specific fault and that each row corresponds to a particular symptom. For example, the $f_2$ model is sensitive to faults affecting any input and the second output. A fault affecting the third input sensitizes all prediction models $f_i, i = 1, \ldots, n$ and $f_{n+3}$. We can see from this table that the $2n$ possible faults $\delta u_i, \delta y_i, i = 1, \ldots, n$ are detectable and isolable because their signatures are independent.

This theoretical signature table serves as a reference for fault detection and localization. The procedure is this time numerical, the prediction residues are calculated at each moment from (23). A threshold on these residues then leads to a binary signature. A comparison is established between the observed signature and each signature of Table 1, associated with the different faults. A perfect match allows determining which component is faulty. Where there is no exact match, the theoretical fault signature vector (Table 1) with a minimum distance with respect to the experimental fault signature vector is postulated as the possible fault.

To summarize the steps of the diagnostic strategy is as follows.

- Offline steps:
  - E1: establishment of a nonlinear prediction model for each output:
  - E2: establishment of a structured nonlinear prediction model for each output;

- E3: validation of prediction models;
- E4: establishment of the theoretical table of fault signatures.

- Online steps:
  - E5: at any time, collect measurements and evaluation of residue indicators of faults;
  - E6: generation of experimental fault indicators by residue thresholding;
  - E7: comparison of theoretical and experimental signatures.

From a practical point of view, to complete these steps, it may be important to distinguish between a fugitive fault and a persistent fault, the duration of persistence is of course specific to each system and a function of operational safety requirements. The diagnostic approach summarized earlier can be easily complemented by a sequential test combining detection and location findings over a given period of time.

*Remark* 7. Models structuration for diagnosis.

As previously said, the redescription variables are a key point for the detection and localization of faults affecting variables in a nonlinear system. Obviously, vectors (18) and (20) favor the occurrence of certain variables to satisfy the requirement of fault isolation. In addition to the notion of occurrence intervenes that of strengthening the influence of a variable on the prediction of the system's outputs. This remains at the discretion of the user who may prefer to accentuate the role of a variable by subjecting it to an ad hoc nonlinear transformation (see Remark 3).

## 4 | NUMERICAL RESULTS

The following examples illustrate the proposed approach. They voluntarily come from the same basic structure, but their complexity is increasing. The first addresses a nonlinear dynamic system with one input and two outputs without direct interaction between them. Since both outputs depend on the same input, generating redundancy between outputs is exploited by constructing a prediction model of one output based on the other. For the second, an interaction between the two outputs is taken into account, which can induce a certain difficulty of isolation because the prediction of an output uses the measurements of the two outputs where one of them can be affected by a fault. The third uses a MIMO system (also with a coupling between the outputs), which makes it possible to use more redundancy for the generation of fault-indicating residuals and thus to increase the capacity to isolate these faults.

### 4.1 | First example

Let us illustrate the relevance of the proposed technique on a two-output nonlinear system defined by

$$\begin{aligned} y_{1,k+1} &= \frac{0.8\, y_{1,k}}{1 - 0.09\, y_{1,k}} + 0.20\, u_k \\ y_{2,k+1} &= -\frac{0.2\, y_{2,k}}{1 + 0.09\, y_{2,k}} + 0.35\, u_k, \end{aligned} \tag{24}$$

where $k$ states for the discrete time. The main idea is to detect and isolate faults that could affect input $u$, output $y_1$, or output $y_2$. It is important to have models capable of such isolation. To this end, we propose four models for predicting system outputs. We start by a priori defining the following redescription variables:

$$\begin{aligned} \boldsymbol{x}_{1,k+1} &= [\, y_{1,k:k-p}, \quad u_{k:k-q} \,]^T \\ \boldsymbol{x}_{2,k+1} &= [\, y_{2,k:k-p}, \quad u_{k:k-q} \,]^T, \end{aligned} \tag{25}$$

to which, for interredundancy purpose, we add the following variables:

$$\begin{aligned} \boldsymbol{x}_{3,k+1} &= [\, y_{1,k:k-p} \quad y_{2,k:k-p} \quad y_{1,k:k-p} \otimes y_{2,k:k-p} \,]^T \\ \boldsymbol{x}_{4,k+1} &= [\, y_{1,k:k-p} \quad y_{2,k:k-p} \quad y_{1,k:k-p} \otimes y_{2,k:k-p} \,]^T, \end{aligned} \tag{26}$$

where $\otimes$ is the Hadamard operator applied here to the product of two vectors. Expressions (25) and (26) are used to define the models (17) and (19) of Section 2.3 with $n = 2$, which generates the predictions $\hat{y}_{i,k+1}, i = 1, \ldots, 4$. The orders $p$ and $q$ have been a priori chosen and set at 2. We could take more previous value, but we found that this configuration is sufficient

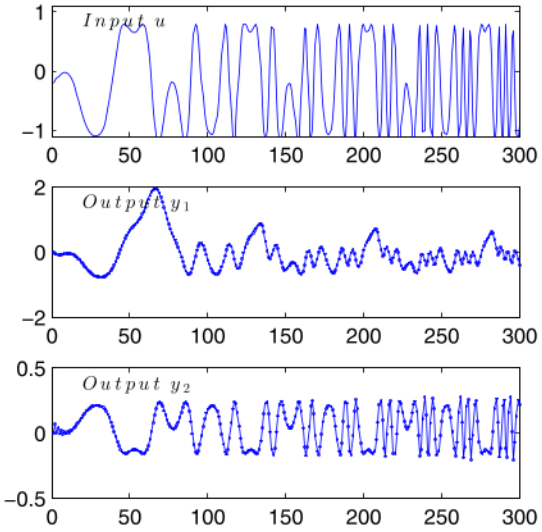


**FIGURE 3** System's input and output for the reference data [Colour figure can be viewed at wileyonlinelibrary.com]

for the aforementioned example. Hadamard's product used in (26) allows to introduce nonlinearities in redescription variables.

We use the Gaussian kernel (12) defining the distance between data. As aforementioned, we have to fix a value to the kernel bandwidth $c$. Many experimental results have shown that the detection and localization are effective for different values of $c$ in a wide range. The same statement can be applied on the regularization parameter $\gamma$ (1). However, in order to obtain a generalized model for adequate FDI, some precautions on the construction of databases must be respected. In particular, the reference data set must describe the system operating conditions and the identification database should be comparable to those of the reference data. Consequently, this can increase the dimensions of matrix $\boldsymbol{K}$, its size being directly linked to the number of observations. However, there exists algorithm to compute an easily interpretable low-rank approximation to Gram matrix $\boldsymbol{K}$.[38]

The identification procedure of Section 2 is illustrated with Figures 3 and 4. The first graph in Figure 3 represents the input of the system. The following two graphs compare the outputs to their estimations using only the variables (25). Figure 4 relates to validation using a second database. The first graph shows the input used. The following four graphs

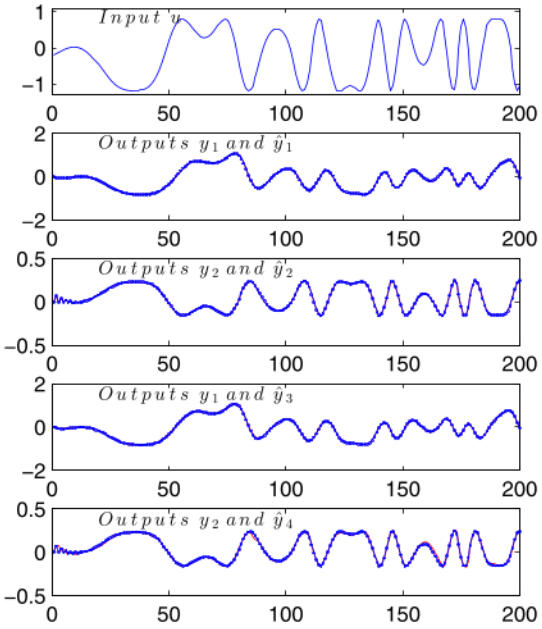


**FIGURE 4** Measured and estimated outputs (training data set) [Colour figure can be viewed at wileyonlinelibrary.com]

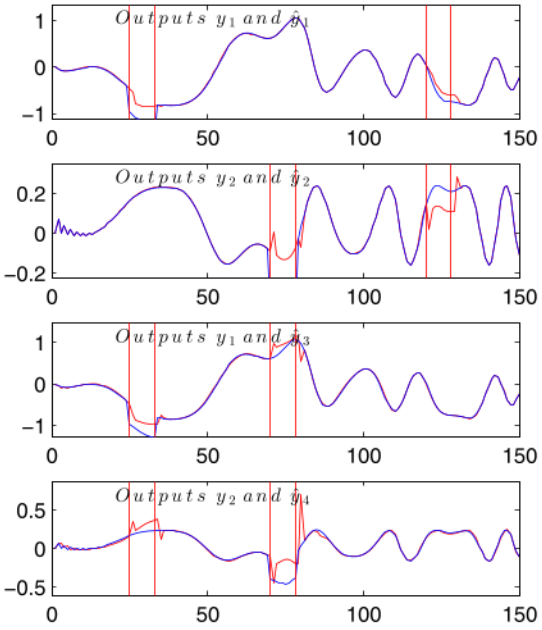


**FIGURE 5** Measured and estimated outputs (testing data set), $c = 5, p = q = 2, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 2** Characteristic of the faults

| | $t_d$ | $t_f$ | *amp.* |
|---|---|---|---|
| $\delta y_1$ | 25 | 31 | -0.4 |
| $\delta y_2$ | 45 | 61 | -0.4 |
| $\delta u$ | 70 | 76 | -0.8 |

compare the outputs with their respective estimates, on the one hand, with input-output models using variables (25) and on the other, hand with interredundancy models not using system input (26). We find that the difference between the measured and the estimated output is very small.

Figure 5 concerns abrupt intermittent fault diagnosis using a testing data set. On the basis of the previous system, some faults were added respectively on the two outputs and the input during the intervals [25, 31], [45, 51], and [70, 76], as mentioned in Table 2. The graphs compare the two measured outputs to their predictions using the four models. Figure 6 is deduced directly from Figure 5, showing the differences between the measured and the estimated outputs values (19). These residuals reveal clearly the time intervals where faults occur and the fault isolation is consistent with Table 1. In order to free themselves from insignificant fluctuations in the residues, the latter were filtered using a trimming technique.[48,49] Figure 7 shows the filtering effect especially for the last residue, and Figure 8 indicates the detection result binary with a threshold of 0.07. In this figure, two indicators are presented, ie, the red line corresponds to the time intervals where the faults were created (Table 2) and the blue line corresponds to the intervals detected on the residues. The correspondence between theoretical intervals and intervals obtained is good, the latter being often wider due to the memory related to the *AR* structure of the models generating the residuals.

The previous thresholding can be replaced by a CUSUM-type test, and the one proposed in the work of Hinkley[50] can be used to detect changes on the residuals and thus to specify the time instants of these changes. It can be seen that the sensitivity of residuals to faults is different and can vary in different operating regions of system behavior. Indeed, it appears that a residual has better detectability performance for a fault in a particular time region, but another residual has better performance in another operating region. Ideally, the threshold for detection has to be adapted and that has been proposed in model-based diagnosis. However, for the data-driven approach, at our knowledge, the problem is still open. In what follows, the detection threshold was set by learning from some tests. Obviously, the choice of this threshold is linked, on the one hand, to the structure of the model, which conditions its ability to make a correct prediction of the

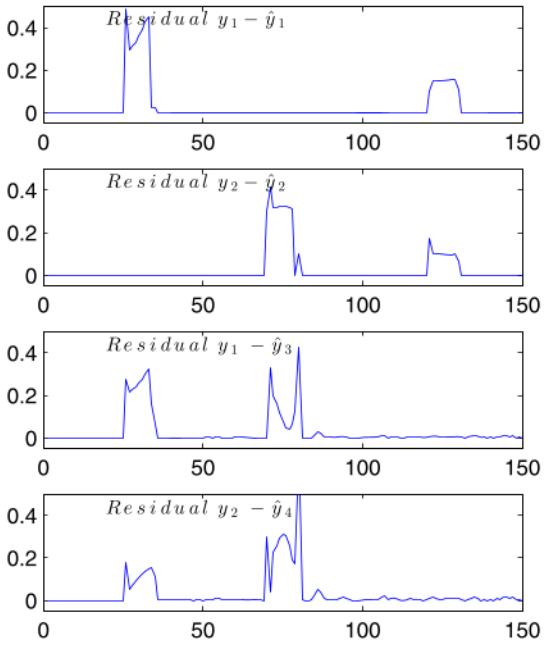


**FIGURE 6** Residual indicators. $c = 5, p = q = 2, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

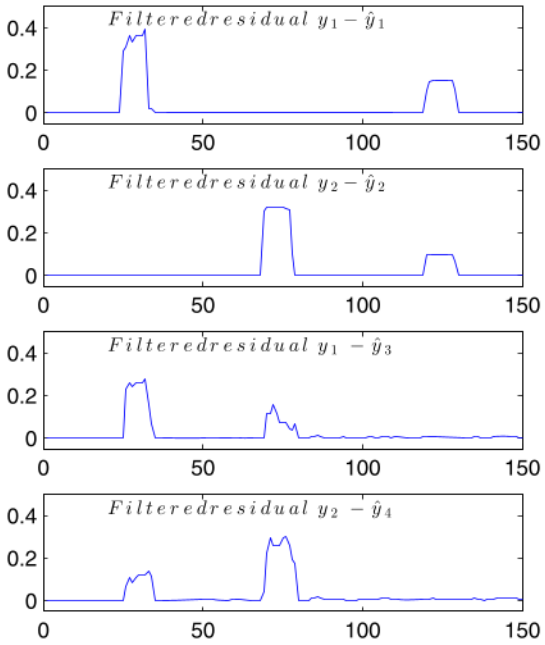


**FIGURE 7** Filtered residuals. $c = 5, p = q = 2, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

outputs of the system and, on the other hand, to the measurement noise, which can make difficult the interpretation of the residues. Since the system under consideration has a nonlinear behavior characterized by a nonlinear regression model, a theoretical study of the influence of these factors is problematic. On the other hand, an experimental analysis of the influence of these two factors is quite possible.

Figures 9 and 10 show qualitatively the influence of a change in $p$ and $q$ orders of the AR part of the prediction model, orders now set at 6. By comparing Figures 7 and 9, on the one hand, and Figures 8 and 10, on the other hand, we note the role of the increase of the order, which overestimates the time intervals corresponding to the presence of a fault; this phenomenon is due to the memory of the system, which increases with the AR order. The influence of the measurement

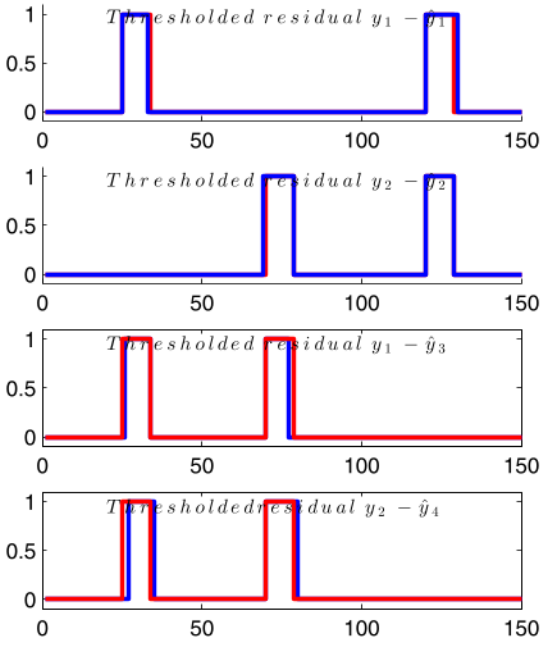


**FIGURE 8** Faults indicators. $c = 5, p = q = 2, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

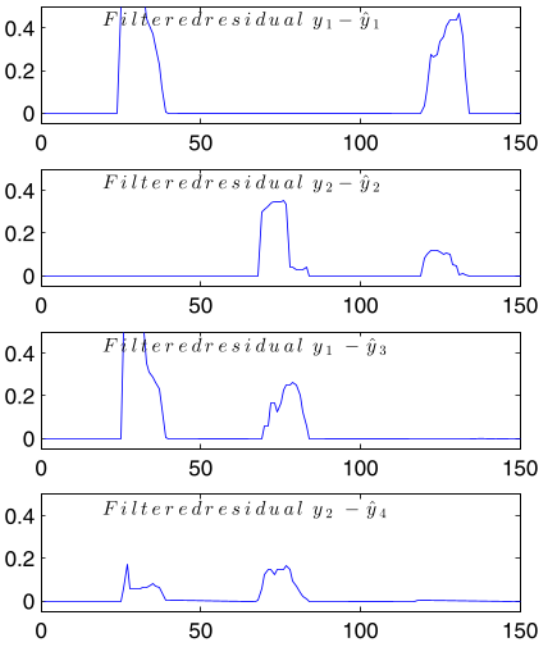


**FIGURE 9** Filtered residuals. $c = 5, p = q = 6, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

noise is visualized in Figures 11 and 12, relative to the residues and their thresholding. This influence is highlighted by comparing Figures 7 and 11, on the one hand, and Figures 8 and 12 on the other hand. Faults are still detected and localized but with some under and overdetection errors. The last test proposed is that of the combined influence of noise and the order of the AR part of the model. Figures 13 and 14 were obtained for a uniform random noise (5% of the maximum value of the output) and an order $p = q = 6$, where the influence of the noise is countered by an increase in the size of the model. Indeed, the increase in the order of the model plays a significant role in the detection and location of faults to the detriment of an increase in false alarms but without nondetection.

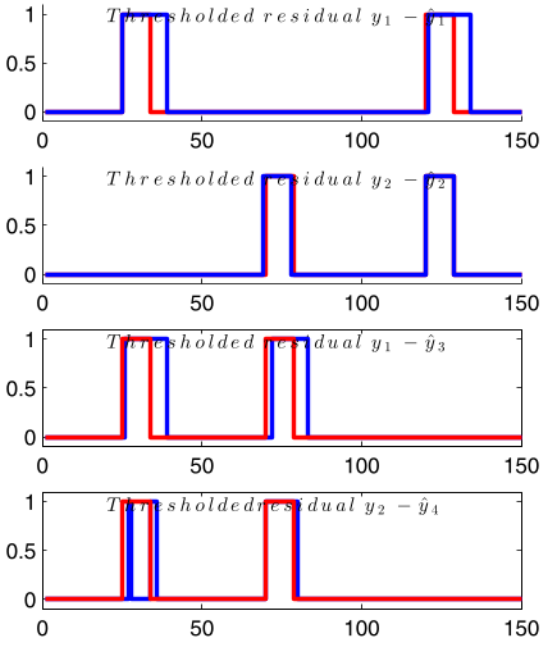


**FIGURE 10** Faults indicators. $c = 5, p = q = 6, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

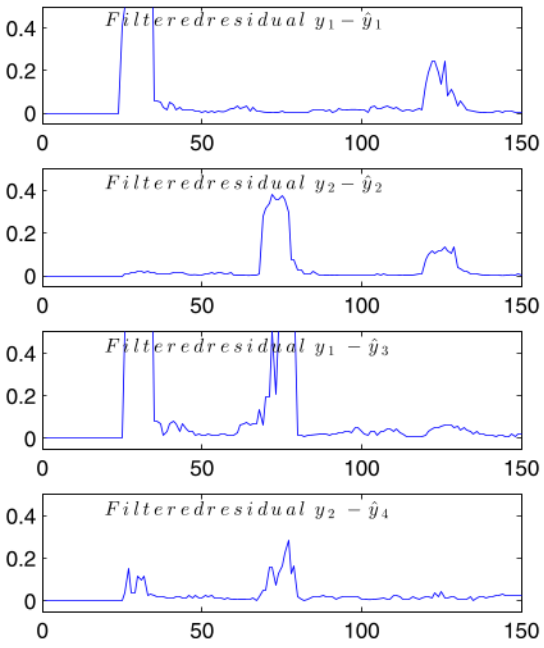


**FIGURE 11** Filtered residuals. $c = 5, p = q = 2, \gamma = 10^{-5}, noise : 5\%$ [Colour figure can be viewed at wileyonlinelibrary.com]

*Remark* 8. Fault estimation magnitude.

In the foregoing, emphasis has been placed on detection and localization. The estimation of the magnitude of the faults is part of the inverse problems, intimately related to the notion of pre-image, which consists of finding an approximate solution by identifying data in the input space based on their corresponding features in the high-dimensional feature space.[7]

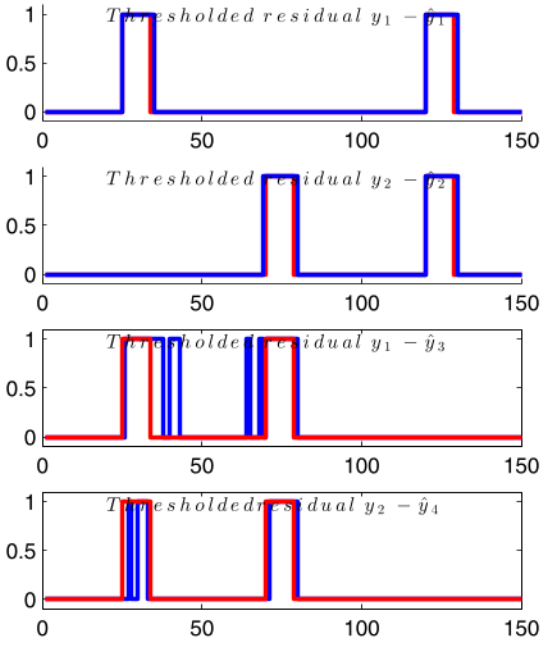


**FIGURE 12** Residual indicators. $c = 5, p = q = 2, \gamma = 10^{-5}$, *noise* : 5% [Colour figure can be viewed at wileyonlinelibrary.com]

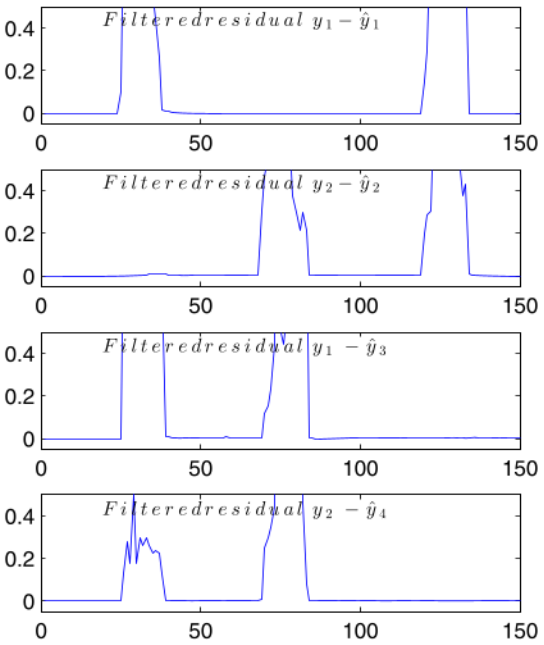


**FIGURE 13** Filtered residuals. Noise: 5%, $c = 5, p = q = 6, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

## 4.2 | Second example

Model (24) is now modified for taking into account a coupling between the two outputs $y_1$ and $y_2$, which will increase the difficulty of fault isolation

$$\begin{aligned} y_{1,k+1} &= \frac{0.60y_{1,k} + 0.20y_{2,k}}{1 - 0.09\,y_{1,k}} + 0.20\,u_k \\ y_{2,k+1} &= -\frac{0.10\,y_{2,k} + 0.30y_{1,k}}{1 + 0.09\,y_{2,k}} + 0.35\,u_k. \end{aligned} \tag{27}$$

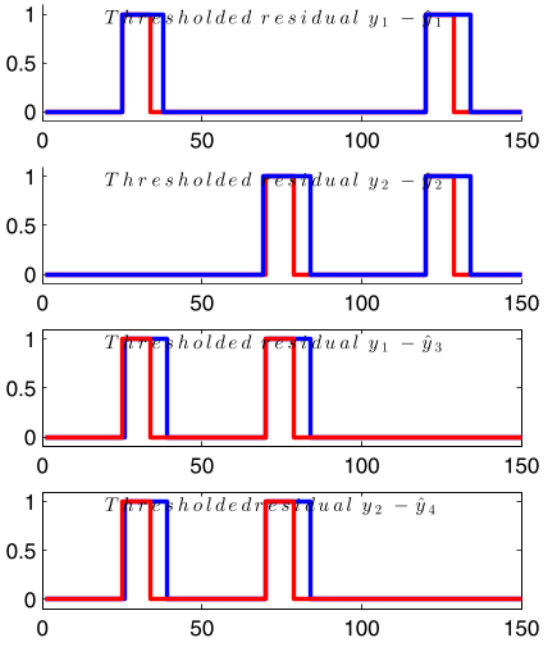


**FIGURE 14** Residual indicators. $c = 5$ and $\gamma = 10^{-5}$, noise: 5%, $c = 5, p = q = 6, \gamma = 10^{-5}$ [Colour figure can be viewed at wileyonlinelibrary.com]

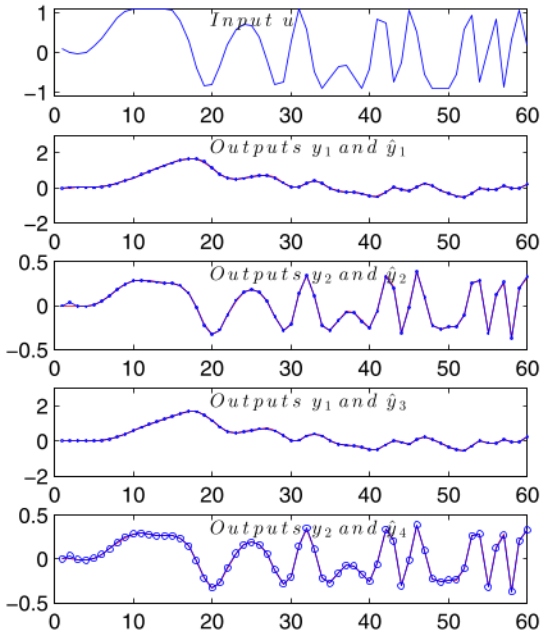


**FIGURE 15** Measured and estimated outputs (training data set) [Colour figure can be viewed at wileyonlinelibrary.com]

Figures 15 and 16 compare the outputs to their estimations for the training and the testing data sets. The residuals depicted on Figure 17 clearly reveal the time intervals where faults occur and the results are consistent with the theoretical signature of Table 1.

## 4.3 | Third example

Model (21) is now extended to MIMO case, which will give more possibilities for residuals generation and which is favorable to the faults isolation

$$y_{1,k+1} = \frac{0.8\,y_{1,k}}{1 - 0.09\,y_{1,k}} + 0.10\,u_{1,k} + 0.05\,u_{2,k}$$
$$y_{2,k+1} = -\frac{0.2\,y_{2,k}}{1 + 0.09\,y_{2,k}} + 0.35\,u_{1,k} - 0.15\,u_{2,k}. \tag{28}$$

Since the goal is to detect and locate measurement faults that may affect the input $u_1$ or $u_2$ and the output $y_1$ or $y_2$, it is necessary to design suitable models to this isolation. For this, it is proposed to establish different model structures.

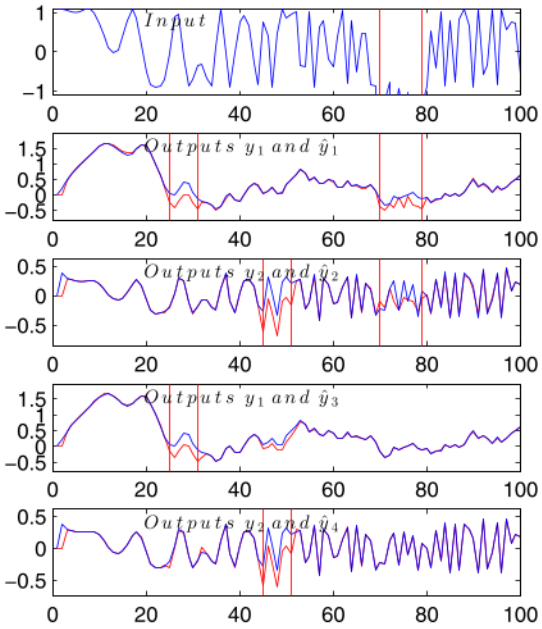


**FIGURE 16** Measured and estimated outputs (testing data set) [Colour figure can be viewed at wileyonlinelibrary.com]

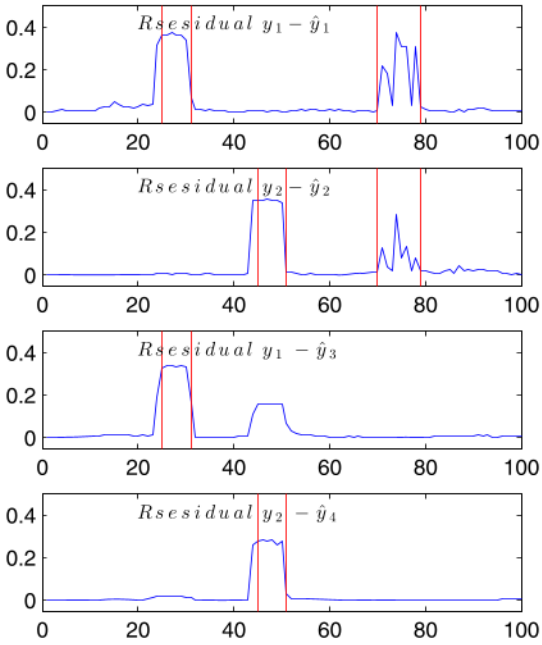


**FIGURE 17** Residual indicators. $c = 7$ and $\gamma = 10^{-9}$ [Colour figure can be viewed at wileyonlinelibrary.com]

**TABLE 3** Fault signatures

| | | $y_1$ | $y_2$ | $u_1$ | $u_2$ |
|---|---|---|---|---|---|
| $M_1$ | $f_1$ | × | . | × | × |
| | $f_2$ | . | × | × | × |
| $M_2$ | $f_3$ | × | × | . | . |
| | $f_4$ | × | × | . | . |
| $M_3$ | $f_5$ | × | × | × | . |
| | $f_6$ | × | × | . | × |
| $M_4$ | $f_7$ | × | . | . | . |
| | $f_8$ | . | × | . | . |

**TABLE 4** Characteristic of the faults

| | $t_d$ | $t_f$ | *amp.* |
|---|---|---|---|
| $\delta u_1$ | 150 | 158 | -0.8 |
| $\delta u_2$ | 200 | 208 | -0.8 |
| $\delta y_1$ | 50 | 58 | -0.4 |
| $\delta y_2$ | 100 | 108 | -0.4 |

The former explains the dependence of each output $y_1$ or $y_2$ on the two inputs $u_1$ and $u_2$. The following translate the dependencies between the two outputs $y_1$ and $y_2$ and one of the two inputs $u_1$ or $u_2$.

Beforehand, the explanatory variables of each model must be defined. The $M_1$ model explains each output based on the two inputs, whereas the $M_2$ model explains one output based on the other. The interredundancy model $M_3$ explains one output according to the other but also taking into account one of the two inputs. Finally, $M_4$ is an $AR$ model

$$
\begin{aligned}
M_1 &: \begin{cases} \hat{y}_{1,k+1} &= f_1(y_{1,k:k-p}, u_{1,k:k-p}, u_{2,k:k-p}) \\ \hat{y}_{2,k+1} &= f_2(y_{2,k:k-p}, u_{1,k:k-p}, u_{2,k:k-p}) \end{cases} \\
M_2 &: \begin{cases} \hat{y}_{1,k+1} &= f_3(y_{1,k:k-p}, y_{2,k:k-p}) \\ \hat{y}_{2,k+1} &= f_4(y_{1,k:k-p}, y_{2,k:k-p}) \end{cases} \\
M_3 &: \begin{cases} \hat{y}_{1,k+1} &= f_5(y_{1,k:k-p}, y_{2,k:k-p}, u_{1,k:k-p}) \\ \hat{y}_{2,k+1} &= f_6(y_{1,k:k-p}, y_{2,k:k-p}, u_{2,k:k-p}) \end{cases} \\
M_4 &: \begin{cases} \hat{y}_{1,k+1} &= f_7(y_{1,k:k-p}) \\ \hat{y}_{2,k+1} &= f_8(y_{2,k:k-p}), \end{cases}
\end{aligned}
\tag{29}
$$

where we recall that the notation $z_{.,k:k-p}$ allows to take into account the variables $z$ over a time horizon $[k : k - p]$.

Table 3 displays the dependencies between the variables translated by these eight models and testifies to the independence of fault signatures. The careful examination of signatures shows that we can reduce the number of models while maintaining the independence character of these signatures.

The redescription variables for these eight models (29) can use linear or nonlinear combinations of the variables of the initial system, the choice of nonlinear combinations remaining to be studied. The preceding models explain the predictions of the output from its measurements. The design of prediction models at time $k + 1$ using the previous predictions and without using the measurements of this output at times instant before $k$ is also possible. As here, the objective is the diagnosis and the search for a model as accurate as possible did not prove necessary.

The $\delta u_i$ and $\delta y_i$ faults respectively affect the inputs and outputs between the instants $t_d$ and $t_f$, as shown in Table 4, where the column "amp." indicates the amplitudes of the faults. The corrections, highlighted in Figure 18, correspond to the differences between measurements and estimates. The time instants where they are significantly nonzero (materialized by vertical lines of green color) are to be put in relation with the data of Tables 2 and 3.

As mentioned previously, the choice of parameter $c$ strongly conditions the values taken by Gram matrix $\boldsymbol{K}$ (11). In particular, values of $c$ that are too large or too small leading to values close to 1 or 0 of the elements of $\boldsymbol{K}$,

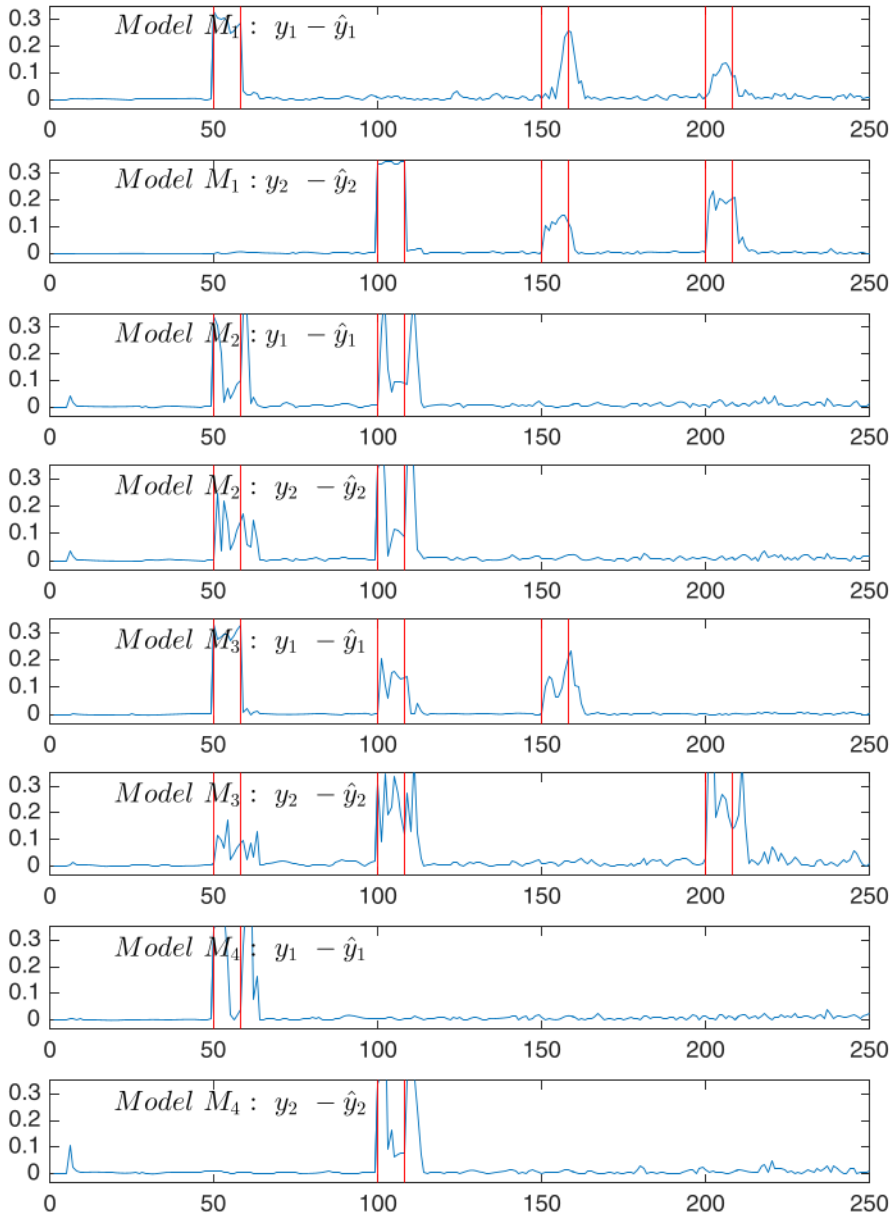


**FIGURE 18** Fault indicators. $c = 5, \gamma = 10^{-9}, p = 5, q = 4$ [Colour figure can be viewed at wileyonlinelibrary.com]

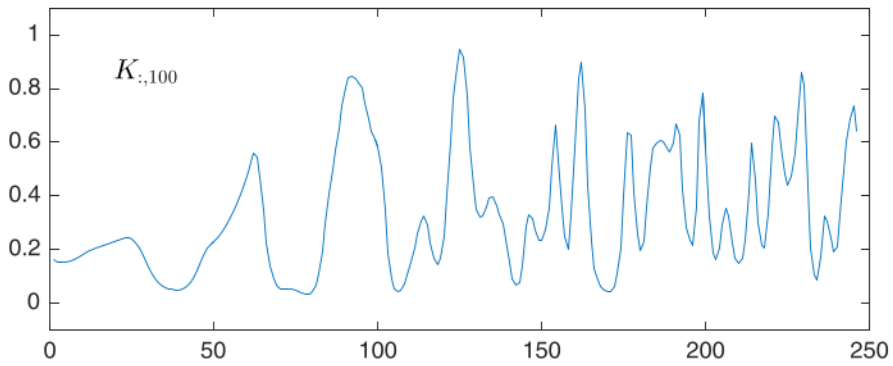


**FIGURE 19** Gram matrix. $c = 5, \gamma = 10^{-9}, p = 5, q = 4$ [Colour figure can be viewed at wileyonlinelibrary.com]

which in a certain way, kills the information transmitted by variables $x_{(.)}$. Figure 19, which visualizes one of the rows of matrix $\boldsymbol{K}$, corresponds to an acceptable choice of parameter $c$, the values taken by $\boldsymbol{K}$ being well distributed in the interval [0 1].

## 5 | CONCLUSION

The paper is devoted to the problem of fault detection (sensors and / or actuators) of a technical system characterized by nonlinear relationship between its variables. However, as the model of the system is a priori unknown, the so-called "model-free" or "data-driven" method is used to solve the problem of diagnosis. Therefore, the main difficulty and the contribution of this approach is due to the nonavailable model of the system.

First of all, in order to make explicit the redundancy of information contained in the measurements, we used nonlinear prediction models. At a given time, the equation error model is defined by a nonlinear regression, depending on the previous measurements of the input and the output of the system. Second, comparing the output with its prediction to its measurement allows to construct a residual, which is a priori sensitive to faults. Thus, the magnitude of these residuals allows to detect when a fault occurs. Then, in order to clarify the location of the fault, a method for isolation is proposed. Isolation need to design reduced prediction output model, involving only a part of the process variables. Variable selection means a subset of relevant variables usually with the aim to have a less complex model to predict an output variable. Furthermore, lower computational power is needed when the model learning is performed on the reduced variable set.

The relevance of the proposed approaches is illustrated on simulated examples with SIMO and MIMO structures, allowing to have some redundancies for the residues generation. Redundancy has been intensively used to generate structured residuals, each of them involving a particular set of variables allowing FDI.

As far as future work is concerned, three directions are possible. The first concerns the taking into account of multiple faults, ie, affecting several inputs or outputs of the system simultaneously. While detection appears to be a fairly direct extension, fault isolation is a difficult problem that will likely require special structuring of prediction models. The second development concerns the detection of faults or variations affecting the system itself, ie, its parameters. Here, too, the difficulty is significant because a single parametric fault can affect all the system outputs. The last development concerns the sensitivity analysis of a system with respect to faults, a subject that remains, in the published works, very little approached from an analytical point of view.

## REFERENCES


1. Qin SJ. Statistical process monitoring: basis and beyond. *J Chemom*. 1999;17(8-9):480-502.
2. Harkat MF, Mourot G, Ragot J. An improved PCA scheme for sensor FDI: Application to an air quality monitoring network. *J Process Control*. 2006;16(6):625-634.
3. Benaicha A, Mourot G, Benothman K, Ragot J. Determination of principal component analysis models for sensor fault detection and isolation. *Int J Control Autom Syst*. 2013;11(2):296-305.
4. Samuel RT, Cao Y. Nonlinear process fault detection and identification using kernel PCA and kernel density estimation. *Syst Sci Control Eng*. 2016;4(1):165-174.
5. Fazai R, Taouali O, Harkat MF, Bouguila N. A new fault detection method for nonlinear process monitoring. *Int J Adv Manuf Technol*. 2016;87(9-12):3425-3436.
6. Alcala CF, Qin SJ. Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Ind Eng Chem Res*. 2010;49(17):7849-7857.
7. Kallas M, Mourot G, Maquin D, Ragot J. Diagnosis of nonlinear systems using kernel principal component analysis. Paper presented at: 11th European Workshop on Advanced Control and Diagnosis (ACD); 2014; Berlin, Germany.
8. Bates DM, Watts DG. *Nonlinear Regression Analysis and Its Applications*. Wiley Series in Probability and Mathematical Statistics - Applied Probability and Statistics Section Series. New York, NY: John Wiley & Sons; 1988.
9. Smyth GK. *Nonlinear Regression*. Hoboken, NJ: John Wiley and Sons; 2006.
10. Kung SY. *Kernel Methods and Machine Learning*. Cambridge, UK: Cambridge University Press; 2014.
11. Sekhar CC, Kini VB. Kernel auto-regressive model with exogenous inputs for nonlinear time series prediction. Paper presented at: International Conference on Computing: Theory and Applications; 2007; Kolkata, India.

12. Caswell JM. A nonlinear autoregressive approach to statistical prediction of disturbance storm time geomagnetic fluctuations using solar data. *J Signal Inf Process*. 2014;5:42-53.
13. Fang K, Shenton AT. Constrained optimal test signal design for improved prediction error. *IEEE Trans Autom Sci Eng*. 2014;11(4):1191-1202.
14. Ren Z, Hou J, Zhou H. Fault Detection and process monitoring of industrial process based on spherical kernel T-PLS. Paper presented at: 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON); 2016; Florence, Italy.
15. Mansouri M, Nounou M, Nounou H, Karim N. Kernel PCA-based GLRT for nonlinear fault detection of chemical processes. *J Loss Prev Process Ind*. 2016;40:334-347.
16. Gao Q, Liu W, Zhao X, Li J, Yu X. Research and application of the distillation column process fault prediction based on the improved KPCA. Paper presented at: IEEE International Conference on Mechatronics and Automation (ICMA); 2017; Takamatsu, Japan.
17. Jia Q, Zhang Y. Quality-related fault detection approach based on dynamic kernel partial least squares. *Chem Eng Res Des*. 2016;106:242-252.
18. Schwantes CR, Pande VS. Modeling molecular kinetics with tICA and the kernel trick. *J Chem Theory Comput*. 2015;11(2):600-608.
19. Park J-H, Lee D-J, Chun M-G. Fault diagnosis for induction machines using kernel principal component analysis. In: Wang J, Yi Z, Zurada JM, Lu B-L, Yin H, eds. *Advances in Neural Networks - ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part III*. Berlin, Germany: Springer-Verlag Berlin Heidelberg; 2006:406-413. *Lecture Notes in Computer Science*; vol. 3973.
20. Huang D, Zhang D, Liu Y, Zhang S, Zhu W. A KPCA based fault detection approach for feed water treatment process of coal-fired power plant. Paper presented at: 11th World Congress on Intelligent Control and Automation; 2014; Shenyang, China.
21. Jun B-H, Park J-H, Lee S-I, Chun M-G. Kernel PCA based faults diagnosis for wastewater treatment system. In: Wang J, Yi Z, Zurada JM, Lu B-L, Yin H, eds. *Advances in Neural Networks - ISNN 2006: Third International Symposium on Neural Networks, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part III*. Berlin, Germany: Springer-Verlag Berlin Heidelberg; 2006:426-431. *Lecture Notes in Computer Science*; vol. 3973.
22. Ren Z, Liu J, She Z, Yang C, Yu H. Data-driven approach of FS-SKPLS monitoring with application to wastewater treatment process. Paper presented at: IEEE International Conference on Industrial Technology; 2016; Taipei, Taiwan.
23. Navi M, Meskin N, Davoodi M. Sensor fault detection and isolation of an industrial gas turbine using partial adaptive KPCA. *J Process Control*. 2018;64:37-48.
24. Exterkate P, Groenen PJF, Heij C, Van Dijk D. Nonlinear forecasting with many predictors using kernel ridge regression. *Int J Forecast*. 2016;32(3):736-753.
25. Exterkate P. Model selection in kernel ridge regression. *Comput Stat Data Anal*. 2013;68:1-16.
26. Chaouch H. Fault detection and isolation of ECG process using kernel principal components. Paper presented at: International Conference on Engineering & MIS (ICEMIS); 2017; Monastir, Tunisia.
27. Pulkkinen S. Nonlinear kernel density principal component analysis with application to climate data. *Stat Comput*. 2016;26(1-2):471-492.
28. Gilbert RC, Trafalis TB, Richman MB, Leslie LM. A data-driven kernel method assimilation technique for geophysical modelling. *Optim Methods Softw*. 2016;32(2):237-249.
29. Deng X, Tian X, Chen S. Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis. *Chemom Intell Lab Syst*. 2013;127:195-209.
30. Lahdhiri H, Taouali O, Elaissi I, Jaffel I, Harakat MF, Messaoud H. A new fault detection index based on Mahalanobis distance and kernel method. *Int J Adv Manuf Technol*. 2017;91(5-8):2799-2809.
31. Zhang Y, Fu Y, Wang Z, Feng L. Fault detection based on modified kernel semi-supervised locally linear embedding. *IEEE Access*. 2018;6:479-487.
32. Zhang L, Lin J, Karim R. Adaptive kernel density-based anomaly detection for nonlinear systems. *Knowl Based Syst*. 2018;139(1):50-63.
33. Jamil F, Abid M, Adil M, Haq I, Khan AQ, Khan SF. Kernel approaches for fault detection and classification in PARR-2. *J Process Control*. 2018;64:1-6.
34. Navi M, Davoodi M, Meskin N. Sensor fault detection and isolation of an autonomous underwater vehicle using partial kernel PCA. Paper presented at: IEEE Conference on Prognostics and Health Management (PHM); 2015; Austin, TX.
35. Wibowo A. Robust kernel ridge regression based on M-estimation. *Comput Math Model*. 2009;20(4):438-446.
36. Ertaş H, Kaçiranlar S, Güler H. Robust Liu-type estimator for regression based on M-estimator. *Commun Stat Simul Comput*. 2017;46(5):3907-3932.
37. Honeine P. Online kernel principal component analysis: a reduced-order model. *IEEE Trans Pattern Anal Mach Intell*. 2012;34(9):1814-1826.
38. Drineas P, Mahoney MW. On the Nyström method for approximating a Gram matrix for improved kernel-based learning. *J Mach Learn Res*. 2005;6:2153-2175.
39. Hussein M, Abd-Almageed W. Efficient band approximation of Gram matrices for large scale kernel methods on GPUs. In: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis; 2009; Portland, OR.
40. Park J, Sandberg IW. Universal approximation using radial-basis-function networks. *Neural Comput*. 1991;3(2):246-257.
41. Wu X, Wilamowski BM. A greedy incremental algorithm for universal approximation with RBF newtorks. In: Fodor J, Fullér R, eds. *Advances in Soft Computing, Intelligent Robotics and Control*. Topics in Intelligent Engineering and Informatics; vol. 8. Cham, Switzerland: Springer International Publishing Switzerland; 2014.

42. Alam MA, Fukumizu K. Hyperparameter selection in kernel principal component analysis. *J Comput Sci*. 2014;10(7):1139-1150.
43. Haridas N, Sowmya V, Soman KP. GURLS vs LIBSVM: Performance comparison of kernel methods for hyperspectral image classification. *Indian J Sci Tech*. 2015;8(24).
44. Soh Y, Hae Y, Mehmood A, Ashraf RH, Kim I. Performance evaluation of various functions for kernel density estimation. *Open J Appl Sci*. 2013;3:58-64.
45. Chen S. Optimal bandwidth selection for kernel density functionals estimation. *J Probab Stat*. 2015. ID242683.
46. Ljung L. *System Identification: Theory for the User*. 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR; 1999.
47. Duan H, Jia J, Ding R. Two-stage recursive least squares parameter estimation algorithm for output error models. *Math Comput Model*. 2012;55(3-4):1151-1159.
48. Huber PJ, Ronchetti EM. *Robust Statistics*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2009.
49. Beliakov G. Fast computation of trimmed means. *J Stat Softw*. 2011; 39(2):1-6.
50. Hinkley DV. Inference about the change-point from cumulative sum tests. *Biometrika*. 1971;58(3):509-523.