Int. J. Appl. Math. Comput. Sci., 2022, Vol. 32, No. 4, 635–655
DOI: 10.34768/amcs-2022-0044

A DATA DRIVEN FAULT ISOLATION METHOD BASED ON REFERENCE FAULTY SITUATIONS WITH APPLICATION TO A NONLINEAR CHEMICAL PROCESS

JOSÉ RAGOT [a], GILLES MOUROT [a,*], MAYA KALLAS [a]

[a]CRAN
University of Lorraine/CNRS
54000, Nancy, France
e-mail: gilles.mourot@univ-lorraine.fr

The diagnosis of systems is one of the major steps in their control and its purpose is to determine the possible presence of dysfunctions, which affect the sensors and actuators associated with a system but also the internal components of the system itself. On the one hand, the diagnosis must therefore focus on the detection of a dysfunction and, on the other hand, on the physical localization of the dysfunction by specifying the component in a faulty situation, and then on its temporal localization. In this contribution, the emphasis is on the use of software redundancy applied to the detection of anomalies within the measurements collected in the system. The systems considered here are characterized by non-linear behaviours whose model is not known *a priori*. The proposed strategy therefore focuses on processing the data acquired on the system for which it is assumed that a healthy operating regime is known. Diagnostic procedures usually use this data corresponding to good operating regimes by comparing them with new situations that may contain faults. Our approach is fundamentally different in that the good functioning data allow us, by means of a non-linear prediction technique, to generate a lot of data that reflect all the faults under different excitation situations of the system. The database thus created characterizes the dysfunctions and then serves as a reference to be compared with real situations. This comparison, which then makes it possible to recognize the faulty situation, is based on a technique for evaluating the main angle between subspaces of system dysfunction situations. An important point of the discussion concerns the robustness and sensitivity of fault indicators. In particular, it is shown how, by non-linear combinations, it is possible to increase the size of these indicators in such a way as to facilitate the location of faults.

Keywords: fault detection, fault isolation, prediction, nonlinear systems, data modelling, kernel.

1. Introduction

1.1. Process diagnoses. Many supervision techniques proposed in the literature are based on behavioural predictions from models characterizing their good functioning, or using approaches based on data collected on the system.

To approach this problem, some historical concepts and definitions should be recalled. Although frequently the terms "fault isolation" and "fault detection" are used synonymously, fault detection means determining that a problem has occurred, whereas fault isolation pinpoints the exact location of this problem. We can add here fault identification which consist in estimating the magnitude and type of the fault. In common usage, fault diagnosis often includes fault detection, fault isolation and fault identification. As explained by Frank (1990) and Blanke *et al.* (2006), one of the basic concepts of model-based diagnosis requires a model of the nominal process as a reference. The first step in diagnosis is then to generate estimates of the actual process outputs. These estimates are then compared with the corresponding process outputs to generate residuals. The final step is to analyze the variations in the residuals. In one way or another, all strategies involve the analysis of redundant information from models and measurements.

Model-based and data-based fault diagnoses have been investigated for decades. Both the approaches used to fault diagnosis have significant advantages and disadvantages. The choice of one or the other approach is often linked to the possibility or not of having a

*Corresponding author

model of the system that accurately reflects its behaviour. As soon as the physical laws, supplemented by some simplifying assumptions, can be applied, the resulting model can then be used for diagnosis (Wang and Wan, 2017). The case of complex systems (large systems or systems with poorly known reaction mechanisms) is generally handled by the second type of approach (Simanio and Farso, 2018), with the data being directly used to provide a functional diagnosis. However, the dichotomy stated may be questionable because, often, database approaches also use models: for example, detecting an operational inconsistency by analyzing a correlation between different quantities is based on the assumption of a model between variables. Perhaps one could say that database methods do not use an *a priori* known model. This is the case, for example, for techniques using principal component analysis where the system model is not known, but the diagnostic technique is based on the search for empirical linear or non-linear redundancies between system variables.

Very often, system operation diagnosis techniques are based on the analysis of the adequacy between a model reflecting good functioning (which therefore plays the role of reference) and the data collected on the system. With this type of technique, anomaly detection is generally simple to implement and is very effective. The same cannot be said for the localization of these anomalies, which requires much finer comparison techniques that are able to isolate the types of malfunctions from each other. To this end, techniques have been developed for structuring fault indicators and, more precisely, by defining subsets of indicators specific to the recognition of particular fault subsets. The search for the ad-hoc dimension of these indicators and their structuring remains a difficult key point, to which are added the need to compromise between the robustness and sensitivity of the indicators. These various observations prompted us to focus our contribution on the structuring of indicators and, consequently, on the processing of data adapted to their development.

Thus, we consider here the case where the behavioural model of the system is not known and must be constructed from measurements made on the system. For linear systems, in order to extract redundancies useful for diagnosis, the relationships between the inputs and outputs of a system are usually obtained using the least-squares estimator. For systems with non-linear behaviour, modelling is much more delicate and requires the use of variable transformations to describe non-linearities between data. Bates and Watts (1988) presented non-linear regression with possible applications in chemical engineering. Smyth (2006) describes the most common models and transformations to convert non-linear models to linear models.

Given most of the work on diagnosis, it is clear that the aspect generally treated, for the non-linear case, is limited to fault detection. A much lower number of papers were devoted to faults isolation. In the works of Alcala and Qin (2010) or Kallas *et al.* (2014), the pre-image technique offers a step of isolation and fault estimation. The complexity of the problem is due to the non-linear transformation that induces some difficulties in the system identification procedure which is only partially solved by using a regularization function.

In order to cope with nonlinear systems, several works applied kernel principal component analysis KPCA. In the works of Botre *et al.* (2016) or Ren *et al.* (2016a), for example, kernel partial least squares regression (KPLS) is used as a modeling framework to generate residuals which are evaluated using a generalized likelihood ratio (GLR) statistic. To the best of our knowledge, the use of a kernel approach to diagnosis of failures appropriate for real processes is only partially addressed and documented in the literature (Ren *et al.*, 2016a). Concerning more applicable aspects of simulated systems, in the works of Manikandan *et al.* (2013), Fazai *et al.* (2016), Gao *et al.* (2017), Wang (2018), Simanio and Farso (2018) or Jia and Huang (2016) the detection of faults in a chemical process is proposed, in those by Huang *et al.* (2014), Jun *et al.* (2006) or Ren *et al.* (2016b) an application to water treatment is presented and in that of Navi *et al.* (2018) an industrial gas turbine is considered. Among multivariate statistical process monitoring methods applied to chemical processes, principal component analysis (PCA) (Benaicha *et al.*, 2013) has been widely used for fault detection in chemical processes, although this approach is devoted to static linear systems. However, some extensions have been proposed for time varying systems (see Gu *et al.*, 2015) and nonlinear systems (kernel PCA) (see Kallas *et al.*, 2014). In the work of Nithya and Vijayachitra (2018) PCA is proposed to detect faults of an actuator, but some results are surprising since PCA is applied in the context of dynamic nonlinear functioning.

1.2. Main idea and contribution. In our paper, we present a novel technique for diagnosis without using an *a priori* known model. Only using data collected from the system, the main idea is to create a set of non-linear structured redundancy equations in order to isolate a fault and define its time localization, in the case where this fault may affect one of the inputs or outputs of the system. In this paper several contributions are to be noted.

Based on the idea that physical models of systems can be difficult to develop, the proposed approach is based on the construction of grey box nonlinear prediction models if some knowledge on causality between variables is available. In order to describe the non-linear behaviour of systems, a formalization using Gaussian type kernels

is used; these kernels have a parameter to adjust the importance of non-linearity.

Usually, diagnostic procedures use the comparison of the current situation to be analyzed with a well functioning model. Here the approach adopted is quite different. The non-linear prediction model is used to construct a set of reference situations for the system in a dysfunctioning state, i.e., for different measurement fault configurations. A classification technique, based on subspace angle estimation, then makes it possible to compare the current situation to be analyzed with the different dysfunctional situations.

If there are a number of publications on faults detection, to the best of our knowledge there are few published works on the isolation of these faults without using an *a priori* known model. It is precisely this point of *fault localization* that is addressed in our paper by proposing a structuring of the fault-indicating residuals, on the one hand thanks to a judicious choice of the re-description variables for the construction of kernel-based regression models and, on the other hand, by using a bank of predictors, each of them using a specific subset of variables in such a way as to improve the isolation of faults.

The use of non-linear prediction models remains quite classic, but we have extended its use to *structured prediction*, i.e., involving particular subsets of explanatory variables, in order to decouple certain models from specific variables. As a direct consequence, this leads to a significant improvement in the separation of faults.

A particularly difficult point for the diagnosis of systems is the lack of *significant sensitivity* of some residuals to faults to be detected. In the case where the residuals depend linearly on the variables and, consequently, on the faults affecting them, the sensitivity analysis is simple. The non-linear case is much more complex, which probably explains the lack of published work in this field. The difficulty stems from the fact that the sensitivity depends closely on the nature of the excitations affecting the system. To solve this problem, we propose to create a reference dictionary containing fault signatures representative of a set of excitation situations of the system. Since insensitivity may also be due to the structure of the models and the values of their parameters, increasing the number of models, at the expense of a slight increase in complexity, favours the isolability of faults. This increase in the number of models is simply due to a combination of the system's primary models, combinations that allow for a decoupling of fault indicators from certain variables.

Finally, the common thread of our text is that of a continuous stirred-tank reactor (CSTR), the structure chosen being sufficiently complex in terms of non-linearities and the number of variables to highlight the interest of a data-based diagnosis technique. This example, although small in size (5 inputs and 5 outputs), serves as a guideline throughout the text to implement the contributions mentioned above. We thought it was wise to use this example systematically in all stages of the proposed approach. However, many developments in the text clearly show that this approach remains generic, which will allow the future reader to use it for his or her own application.

In view of the main ideas underlying our approach, it is clear that the very principle of FDI, as set out in the seminal work, remains in use, namely the generation of fault-indicating residuals. However, the way in which they are constructed and used covers a number of personal developments in our presentation: prediction from a non-linear grey box model, structuring residuals by elimination of variables, dictionary of reference residuals covering different operating regimes, situation recognition by estimating angles between subspaces.

1.3. Outline. The remainder of the paper is organized as follows. Section 2 is devoted on the one hand to the model structures that can be proposed to describe the behaviour of a CSTR and, on the other hand, to the output prediction of a system considering the non-linear case by using a kernel formulation. Section 3 is dedicated to the detection and isolation of faults affecting input or output. This section alerts the reader to the problem of the sensitivity of fault detection indicators by providing some elements for discussion and the beginning of a solution. Section 4 is devoted to our contribution: designing a structured model, designing fault indicators allowing isolation, estimating the magnitude of faults affecting the measurements of the input and output variables. A conclusion is presented in Section 5 and suggests some possible research directions.

2. System under consideration and its different model structures

In the following the systems considered are represented by the set of differential equations :

$$\begin{aligned} 0 &= g_1(\mathrm{d}/\mathrm{d}t, x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m), \\ &\vdots \\ 0 &= g_n(\mathrm{d}/\mathrm{d}t, x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m), \end{aligned} \tag{1}$$

where the variables x_i and u_i are generally referred to as state variables and command or input variables, respectively, and where the operator $\mathrm{d}/\mathrm{d}t$ reminds that the variables x_i and u_i may appear as their derivatives with respect to time.

In the following, the structures of the models that can reflect how the CSTR operates are presented. We

recall the equations resulting from the knowledge of reaction mechanisms, then those resulting from a black box approach, since the FDI approach proposed in the following will be established in the absence of a phenomenological model.

In order to address the objective of detecting and isolating faults that may affect the measurements of these variables, we are led to construct indicators that can highlight the presence of these faults and, above all, locate them. To do this, these indicators must have a particular structure so that each of them is sensitive to a fault or possibly to a group of faults. This structuring is made possible if the components themselves of the system model (1) have ad-hoc sensitivities to the different variables. Obviously, this is not the case of the system (1) whose components $g_i, i = 1, \ldots, n$ are sensitive to all the variables and thus to all the faults that can affect them. The situation would be different if some components g_i of the model were independent of particular variables. For example, this independence can be obtained, under certain conditions, by trying to eliminate the variable x_1 between the two equations g_1 and g_2.

When the functions g_i are perfectly known, tools adapted to this elimination are available (Ritt, 1950; Nur *et al.*, 2018; Diop, 1989) and, in particular, in the case of polynomial differential systems (Buchberger and Kauers, 2010). In our proposal, the situation is quite different since only the occurrence of the variables in these models is known. Therefore, the elimination principle previously mentioned cannot be directly applied. However, from a strictly structural point of view, it is quite possible to assume the existence of a model

$$0 = g_{n+1}(\mathrm{d}/\mathrm{d}t, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m)$$

resulting from the elimination of the variable x_1 between two models g_i and g_j of (1) where this variable intervenes.

In the following, this idea will be widely exploited using the specific example of a chemical reaction, but as indicated above, the proposed approach remains generic and does not use any particularity linked to this example. Moreover, although known, the equations describing the chemical reaction mechanism are not used in order to remain in the context of data-based FDI. The following sections are devoted to the summary description of the reaction mechanism, the grey box model structure, the kernel-based prediction model structure and their parametric identification.

2.1. Physical model example: The continuous stirred-tank reactor. The following continuous model $\mathcal{M}_c$ describes the chemical kinetics of a reactor in which five species A, B, X, Y, Z are considered. Without affecting the generality of the proposed diagnostic approach, the feed rate q as well as the volume V of the reactor

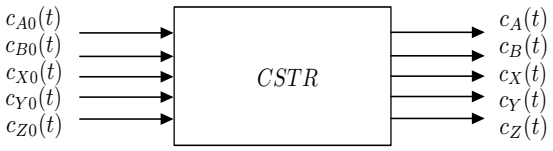


Fig. 1. System structure.

are assumed to be constant. The input concentrations are denoted by $c_{A0}, c_{B0}, c_{X0}, c_{Y0}, c_{Z0}$ and the output concentrations by c_A, c_B, c_X, c_Y, c_Z. Figure 1 shows that the system is of the MIMO type with five inputs and five outputs. The equations of the model of this system are derived from the knowledge of the reaction mechanisms between the five species considered but also from simplifying hypotheses on the operation of the reactor (in particular, homogeneity). From a qualitative point of view, the unbalanced reaction mechanism is described by

$$\begin{aligned} A + B &\xrightarrow{k_1} X, \\ B + X &\xrightarrow{k_2} Y, \\ B + Y &\xrightarrow{k_3} Z, \end{aligned} \tag{2}$$

the term unbalance meaning that it is only listed the reactants and the products, respectively, on the left- and right-hand sides of the arrow. The unknown reaction rate coefficients $k_i, i = 1, \ldots, 3$ are assumed to be constant to simplify the presentation and implementation of the proposed diagnostic approach. In the following, for notational simplicity, the input and output variables are respectively gathered in the vectors:

$$\begin{aligned} u &= \begin{bmatrix} c_{A0} & c_{B0} & c_{X0} & c_{Y0} & c_{Z0} \end{bmatrix}^T, \\ x &= \begin{bmatrix} c_A & c_B & c_X & c_Y & c_Z \end{bmatrix}^T. \end{aligned} \tag{3}$$

It is important to note that in the following, the system model is assumed to be unknown, and only its structured is known. Obviously, in order to have measurements of the inputs and outputs of the system, a non-linear dynamic model was used, but in the proposed diagnostic approach this model was then completely ignored, our objective being to detect and locate faults only from the input and output data collected on the system in operation.

2.2. Grey-box model of the continuous stirred-tank reactor. Based on the data collected from a system, different modelling strategies can be considered to characterize its functioning. In the absence of knowledge on reaction mechanisms, i.e., by ignoring the model equations, an elementary way to represent the MIMO system is provided by a set of equations which expresses each predicted output as a function of the five inputs and this without taking into account any coupling

between the outputs. This type of model is known as a black-box model in the sense that the choice of its structure does not result from any phenomenological knowledge of the system.

Still ignoring the reaction mechanisms, we can also consider building simpler models where each output of the MIMO system only uses certain inputs. Of course, in the absence of knowledge about reaction mechanisms, we are rather helpless about how to design this explanatory model of outputs as a function of certain inputs, especially when they are dynamic systems with non-linear behaviour. With regard to the diagnosis purpose, the only interest in this reduced black-box model lies in the reduction in the number of explanatory variables but above all, in the context of fault detection, in the fact that the non-inclusion of a variable in an equation is free from the influence of the possible fault of that variable.

A more reasonable situation uses the notion of a grey-box model because it uses partial knowledge of the behaviour of the physical system. In the following, the proposed approach uses this grey-box model because it is only based on the known structure (2) of the reaction mechanism governing the combined evolution of the different chemical species. It should be noted, however, that this knowledge remains rather superficial in the sense that kinetics remain unknown and that only the occurrences of the variables in the concentration evolution equations are used.

Given the known structure of the reaction mechanism (2) and, in particular, the occurrence of variables in this model, a smaller number of explanatory variables can be used in the non-linear regression model than that used in the black-box models. In general, this reduction principle remains applicable for any system whose physical or chemical mechanisms are known without knowing the numerical values of the parameters of the model of this system. According to the description (2), because of the nature of the interactions between the five species involved, the following structures, also illustrated with Fig. 2, are proposed:

$$\mathcal{M}\begin{cases} \hat{x}_{1,i} = f_1(x_1^{h_p}, x_2^{h_p}, u_1^{h_q}), \\ \hat{x}_{2,i} = f_2(x_1^{h_p}, x_2^{h_p}, x_3^{h_p}, x_4^{h_p}, u_2^{h_q}), \\ \hat{x}_{3,i} = f_3(x_1^{h_p}, x_2^{h_p}, x_3^{h_p}, u_3^{h_q}), \\ \hat{x}_{4,i} = f_4(x_2^{h_p}, x_3^{h_p}, x_4^{h_p}, u_4^{h_q}), \\ \hat{x}_{5,i} = f_5(x_2^{h_p}, x_4^{h_p}, x_5^{h_p}, u_5^{h_q}), \end{cases} \tag{4}$$

where $h_p = i-1 : i-p, h_q = i-1 : i-q$ f_i will be further defined and the notation $i-1 : i-g$ allows to take into account the information of the time interval $[i-g, i-1]$. The advantage of this structure is, on the one hand, the reduction in the number of explanatory variables in each model and, on the other hand, subsets of explanatory variables specific to each model, which will promote the isolation of faults affecting

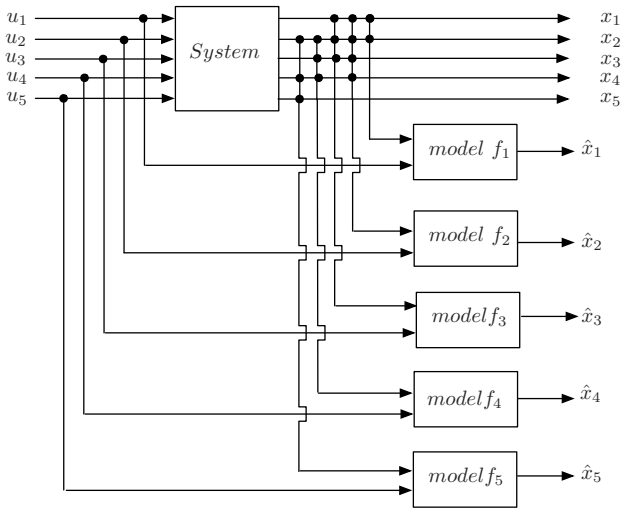


Fig. 2. Functional diagram of system model $\mathcal{M}$.

Table 1. Variable occurrence in models $\mathcal{M}$.

	x_1	x_2	x_3	x_4	x_5	u_1	u_2	u_3	u_4	u_5
f_1	×	×	.	.	.	×	.	.	.	.
f_2	×	×	×	×	.	.	×	.	.	.
f_3	×	×	×	.	.	.	.	×	.	.
f_4	.	×	×	×	.	.	.	.	×	.
f_5	.	×	.	×	×	.	.	.	.	×

the measurements. Table 1 provides the appearance of variables in models (4). A column in this table indicates the occurrence of a variable in the five equations, while a row lists the variables involved in a particular equation. With regard to our diagnosis objective, the examination of this table clearly indicates the possibility of detecting all faults that may affect an input or output of the system; we observe that the signatures of the variables in the $\mathcal{M}$ models are distinct from each other except for u_5 and x_5 for which the faults are not isolable.

To conclude, since the phenomenological model is not known *a priori*, only the structure (4) of $\mathcal{M}$ will be used for the diagnosis. The proposed model is of the grey box type because it uses partial knowledge of the real system, i.e., the causality between inputs and outputs.

2.3. Output prediction using dynamic non-linear regression. The prediction of variables generated by time series is a major tool in all areas of system monitoring and diagnosis. Indeed, obtaining predictions directly serves to predict the future evolution of a system, which is essential, on the one hand, to anticipate its behaviour, whether normal or abnormal, and, on the other hand, to make a decision on the control law to be applied to the

system. This prediction is generally based on a sequence of past observations of the input and output of the process under consideration and uses a pre-established model of the normal functioning of the system.

To do this, a preliminary step of identifying the system was carried out, by constructing a relationship between the input and output of the system based on data deemed to be healthy. In general, independently of our application to the CSTR, if we consider the input and output variables (u, y) of a SISO system, y can be explained by means of u using a non-linear regression technique. Of course, the underlying difficulty is the choice of the regression structure. First of all, we recall the structure of a linear static regression that will serve as a starting point for the synthesis of a non-linear dynamic regression.

In its simplest form, a prediction model can be defined as a linear regression described, at time i, as $y_i = f(z_i, \theta)$, where θ and $z_i \in \mathbb{R}^{p+q}$. The estimates of the model parameters θ can be obtained by different techniques; the simplest approach is to minimize the following penalized criterion:

$$\Phi(\theta) = \frac{1}{2} \sum_{i=r+1}^{r+N} (y_i - f(z_i, \theta))^2 + \frac{\gamma}{2} \parallel \theta \parallel^2, \quad (5)$$

where $r+1$ and $r+N$ are the bounds of the time horizon used for the identification and where the explanatory z_i variables depend on the past outputs and inputs of the system:

$$z_i = \left[\; y_{i-1}, \ldots, y_{i-p}, u_{i-1}, \ldots, u_{i-q} \; \right]^T \in \mathbb{R}^{p+q} \quad (6)$$

As in any estimation procedure using a black box model, the choice of the structure of the function f is up to the user and is linked to the system to be characterized.

Remark 1. This identification principle, presented here for a single output variable, obviously applies to all the output variables of the system. In the case of the CSTR, it is the variables $x_i, i = 1, \ldots, 5$ that will be concerned.

In (5), N is the number of data pairs available and γ 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, while the second term provides more regular solutions. The choice of the regularization parameter γ controls the balance between the training error and the degree of regularity. As is well known, this so-called ridge regression reduces the variability between the coefficients of θ by shrinking them; as a result, more prediction accuracy is obtained at the cost of only a small increase in the bias in the estimated parameters. In this paper, we will not seek for an optimal value of γ, our goal being essentially dedicated to the detection of a fault.

To deal with a non-linear prediction model, it is a common practice to map the original input data z_i into another feature space. For that purpose, using a non-linear function φ, we consider the projection of the data:

$$z_i \in \mathbb{R}^{p+q} \;\; \rightarrow \;\; \varphi(z_i) \in \mathbb{R}^r. \quad (7)$$

In this case, we consider a kernel K defined by the elements $[K]_{i,j} = \langle \varphi(z_i), \varphi(z_j) \rangle$. This makes the prediction of the system output explicit (Kallas *et al.*, 2018):

$$\begin{aligned} \hat{y} &= K\,\beta, \\ \beta &= (\gamma\, I_N + K)^{-1}\, y \end{aligned} \quad (8)$$

with

$$\begin{aligned} cy &= \left[y_1, \ldots, y_N \right]^T \quad \in \mathbb{R}^N \\ c\beta &= \left[\beta_1, \ldots, \beta_N \right]^T \quad \in \mathbb{R}^N. \end{aligned}$$

Moreover, for a new observation at time $p+1$, we have

$$z_p^{\text{new}} = \left[y_p^{\text{new}}, \ldots, y_1^{\text{new}}, u_q^{\text{new}}, \ldots, u_1^{\text{new}} \right]. \quad (9)$$

The predicted output can be written as

$$\hat{y}_{p+1}^{\text{new}} = \kappa^T(z_p^{\text{new}})\, \beta \quad (10)$$

with the following definitions:

$$\kappa(z_p^{\text{new}}) = \begin{bmatrix} \langle \varphi(z_{r+1}), \varphi(z_p^{\text{new}}) \rangle \\ \vdots \\ \langle \varphi(z_{r+N}), \varphi(z_p^{\text{new}}) \rangle \end{bmatrix} \in \mathbb{R}^N,$$

$$K = \begin{bmatrix} \langle \varphi(z_{r+1}), \varphi(z_{r+1}) \rangle \ldots \langle \varphi(z_{r+1}), \varphi(z_{r+N}) \rangle \\ \langle \varphi(z_{r+2}), \varphi(z_{r+1}) \rangle \ldots \langle \varphi(z_{r+2}), \varphi(z_{r+N}) \rangle \\ \vdots \\ \langle \varphi(z_{r+N}), \varphi(z_{r+1}) \rangle \ldots \langle \varphi(z_{r+N}), \varphi(z_{r+N}) \rangle \end{bmatrix},$$

$$z_i = \left[\; y_{i-1}, \quad \ldots, \quad y_{i-p}, \quad u_{i-1}, \quad \ldots, \quad u_{i-q}, \; \right]^T.$$

It is important to mention that the non-linear function $\varphi(\cdot)$ is not necessarily given in an explicit form, however the kernel function $K(\cdot, \cdot)$ can perfectly describe the non-linear relations between data only by defining the dot product. Many non-linear functions can be used to define this dot product $\langle \cdot, \cdot \rangle$ between data. In this paper, we consider the Gaussian function which is a radial basis function (but there exist other choices such as quadratic, polynomial, hyperbolic tangent kernels), taking into consideration the distance (or similarity) between data. The coefficients $\kappa(z_i, z_j)$ of $K(\cdot, \cdot)$ are given by

$$\kappa(z_i, z_j) = \exp\Big(-\frac{1}{c} \parallel z_i - z_j \parallel^2 \Big), \quad (11)$$

where the parameter c controls the sensitivity of the kernel function with respect to the observation $z_{(\cdot)}$ (smaller values of c will make the function overfit the data points, while larger values will make it underfit). Moreover, if the measurement output variable y^{new} is known, it is possible to compare the estimate $\hat{y}^{\text{new}}$ with this measurement in order to generate a residual indicator whose magnitude may reveal a measurement anomaly on u^{new} or y^{new}:

$$\tilde{y}^{\text{new}} = \hat{y}^{\text{new}} - y^{\text{new}}. \tag{12}$$

2.4. Identification of model parameters. As the principle of kernel transformation has been developed in Section 2.3, we limit ourselves here to giving the structure of the models and the identification results showing the quality of the forecast of the five outputs. The procedure of identifying the system model of correct functioning is a completely classical approach integrating validation and testing of the model. For the identification of this model, it was implicitly assumed that all input and output variables were measured. Obviously, in a more general framework, the number of available measurement sensors can lead to constraints on the number of models to be built.

The following figures reflect the results of identifying the parameters of the functions $f(z_i)$ involved in the model $\mathcal{M}$. Figure 3(a) shows the five entries in the training database $\mathcal{B}^e$. Using these data, the model was developed and validated with the values $p = 8, q = 8$. Figure 3(b) shows the five noisy outputs of this same database. A database $\mathcal{B}^t$ is used for testing the identified model $\mathcal{M}$. Figure 4(a) shows the inputs used for this test. The choice of these inputs complies with the range of variation of the inputs used in the database $\mathcal{B}^e$. Figure 4(b) contains five graphs relating to the five outputs model $\mathcal{M}$. This figure also compares the five measured system outputs with the corresponding predicted outputs of model $\mathcal{M}$. We note that the structure chosen for models $\mathcal{M}$ is quite adequate to reproduce the variations in the outputs despite the noise affecting the measurements.

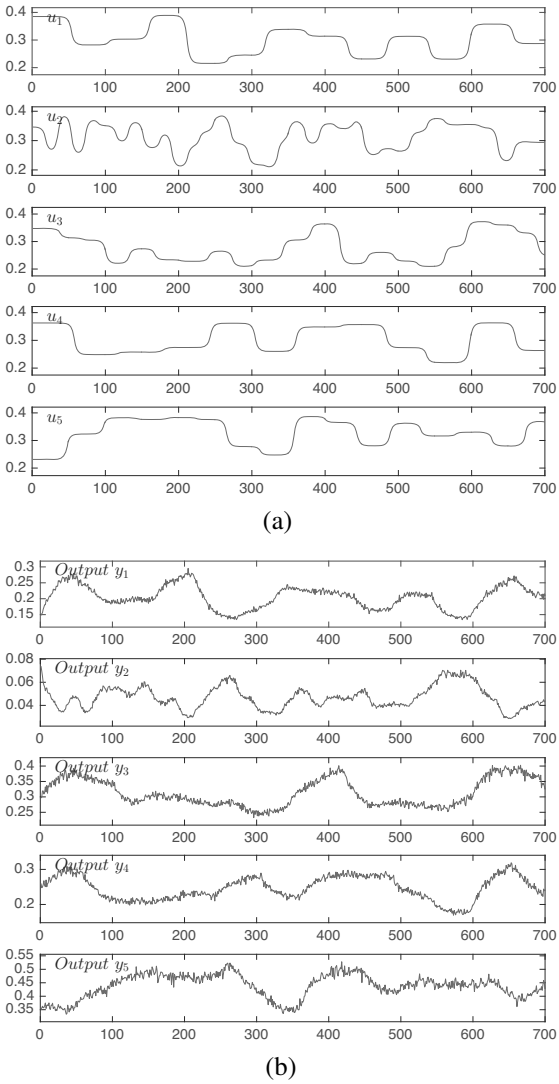
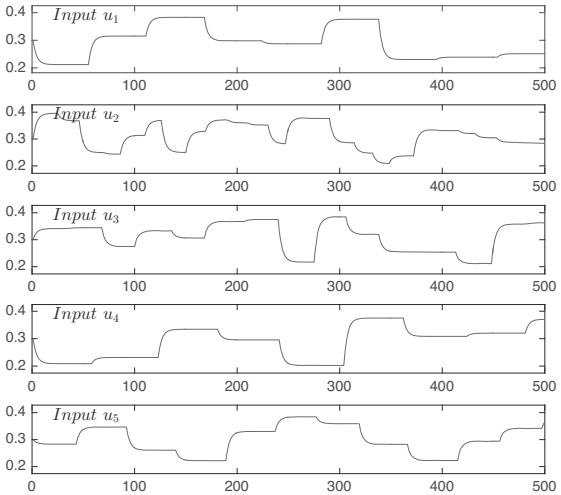


Fig. 3. Training database $\mathcal{B}^e$: input (a) and output (b).

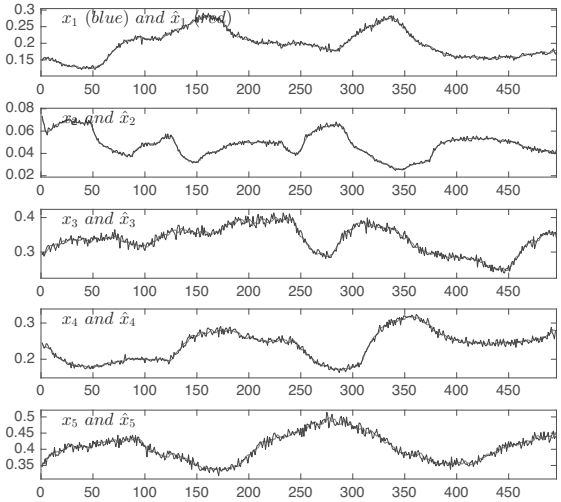
Remark 2. (*Influence of hyper-parameters*) The identification technique has a number of hyper-parameters offering degrees of freedom to improve the quality of the models resulting from this identification. These parameters include: c, the range of functions used to calculate the distances between observations (11), p and q the orders of the regression models (6), γ, the regularization parameter (5) used to estimate the parameters for regressions. However, this important aspect is not discussed here for reasons of space and the reader can refer to Kallas *et al.* (2018) for an analysis of the choice of parameters. Let us simply mention that for this application, the values of c can be between 4 and 8, those of γ between 10^{-3} and 10^{-9} and the orders p, q between 4 and 12. In practice, these values must be adapted according to the severity of the non-linearity of the system and its dynamics.

3. Principle of fault localization

Let us recall some essentials of the procedure usually used in the field of diagnosis before presenting the strategy adopted and the results obtained. There are generally three important successive phases, often of increasing complexity: detection, localization or isolation and fault estimation. In all that follows we consider only faults affecting the measurements of the inputs or outputs of the system, which in practice leads to the presence of abnormal measurements. As indicated in the introduction, the fundamental point of the diagnosis concerns the synthesis of indicators, often called residuals, capable of highlighting and recognizing dysfunctions.



(a)



(b)

Fig. 4. Test database $\mathcal{B}^t$: input (a) and output (b).

Definition 1. (*Residual*) In the diagnostic sense, a residue is a function of the data collected from the system to be monitored that takes the zero value in the absence of data errors. On the other hand, in the presence of faults, under certain structural conditions, the residual is expected to manifest this presence, i.e., to be sensitive.

Definition 2. (*Detectability*) Given a set R of residuals and a D set of faults, a fault $d \in D$ is detectable if there is a subset of residuals $\bar{R} \in R$ depending on d.

Definition 3. (*Isolability*) Given a set R of residues sensitive to two faults $d_i \in D$ and $d_j \in D$, d_i is isolatable from d_j if there exists a subset of residuals $\bar{R}_i \in R$ depending on d_i but not on d_j, and a subset of residuals $\bar{R}_j \in R$ depending on d_j but not on d_i.

Definition 4. (*Fault signature*) A fault signature vector typically consists of values 0 or 1. A value of 1 corresponds to a symptom indicating a fault, while a zero indicates normal behavior. A signature consisting of zeros only represents normal operation. A set of signatures can then constitute a fault signature matrix, each column corresponding to a fault.

Detection is limited to looking for the presence of a fault affecting a variable by estimating its occurrence time, while isolation specifies which variable is subject to the detected fault. The estimation aims to quantify the amplitude of the fault. The detection/localization technique is based on the generation of fault indicators, each indicator being characterized by a particular signature. Fault isolation consists in comparing the experimental signature obtained from the measurements with all the theoretical signatures. This comparison remains a delicate and difficult point of this approach, as it often requires a normalization of indicators and the use of thresholds.

In fact, behind the apparent simplicity of the detection/localization procedure lie a number of difficulties. First of all, since the system under consideration is non-linear, the impact of measurement errors can vary in accordance with the operating point of the system, i.e., according to the inputs applied to it. This variation in sensitivity also affects the isolability of the faults, as faults affecting two different variables can have very similar influences.

Secondly, linked to the previous problem of sensitivity, the choice of detection thresholds then becomes problematic and the use of adaptive thresholds in the case of non-linear systems remains an almost unexplored approach to the best of our knowledge. If the choice of thresholds to be applied to fault indicators remains problematic, the original reason undoubtedly comes from the difficulty of designing these indicators offering *a priori* the fundamental property of separating fault signatures.

A third, and not the least, difficulty is that of fault isolation, which is eminently linked to the structure of the system and that of the indicators performing the diagnosis. If the number of variables subject to failure is high (10 in our case), these indicators must have independence properties, which is partly related to their dimensions. Again, while isolability studies exist in the linear case, to the best of our knowledge no methodological work has been published on this subject.

In Sections 3 and 4, we do not attempt to provide theoretical solutions to these difficulties, but they are explained and the proposed approaches provide practical answers. Section 3.1 presents an example of residual generation and their temporal evolution, indicating how they are sensitive to faults. In Section 3.2 FDI will be dealt with according to the classical procedure based on

thresholds and its limits will be highlighted, which will justify the approach proposed in Section 4. As mentioned in Section 3.3, the presence of non-linearities in the system under study reinforces this difficulty because the sensitivity of the fault indicators is highly dependent, on the one hand, on the operating point of the system and, on the other hand, on the nature of the exciting inputs.

3.1. Residual generation. In order to cover all situations, the faults were created on the five outputs x_i in the respective time intervals $I_j, j = 1, \ldots, 5$ and on the five inputs u_i in the respective time intervals $I_j, j = 6, \ldots, 10$:

$$I_1 = [50 : 110], \qquad I_2 = [140 : 200],$$
$$I_3 = [230 : 290], \qquad I_4 = [320 : 380],$$
$$I_5 = [410 : 470], \qquad I_6 = [560 : 620],$$
$$I_7 = [650 : 710], \qquad I_8 = [740 : 800],$$
$$I_9 = [830 : 890], \qquad I_{10} = [920 : 980].$$

The model for the correct functioning of the system was identified based on the healthy database (Section 2.4). Diagnostic tests were systematically performed on model $\mathcal{M}$ with several sets of data collected in situations of sensor failures. For each observation, the prediction model then allows the calculation of residuals (12). Two examples of state residuals are shown Figs. 5 and 6 obtained with parameters $\gamma = 10^{-6}, c = 5, p = 8, q = 8$ and two data sets. We observe that the residuals are sensitive to the fault that affects the measurement of the state x_i and to all the inputs $u_i, i = 1, \ldots, 5$.

According to Table 2 of theoretical fault signatures, for several time instants there is a good agreement between experimental and theoretical signatures. In Figs. 5 and 6 the grey rectangles indicate for each residual of model $\mathcal{M}$ the time locations of the faults to be detected. At each sampling moment, the reader will be able to verify that the experimental signature generated by the five model residues, after comparison with the theoretical signatures of Table 2, allows us to locate most of the faults.

For example, in Figs. 5 and 6, up to time instant 49, the five residuals are close to zero. For times 50 to 110 (approximatively), only the first and the third residuals show a sensitivity to fault on x_1, which is in accordance with the values in Table 1. However, the second residual is not sensitive to this fault. The other time intervals are commented on in a similar way. Moreover, we can also notice that the sensitivity of the residuals to the faults depends on the data set that have been used. For example, the examination of Figs. 5 and 6 highlights that the fifth residual presents different sensitivities towards the fault on x_4 in the time interval $[320 : 380]$. Thereafter, it will be necessary that the diagnostic procedure takes into account these variations of sensitivity according to the excitations,

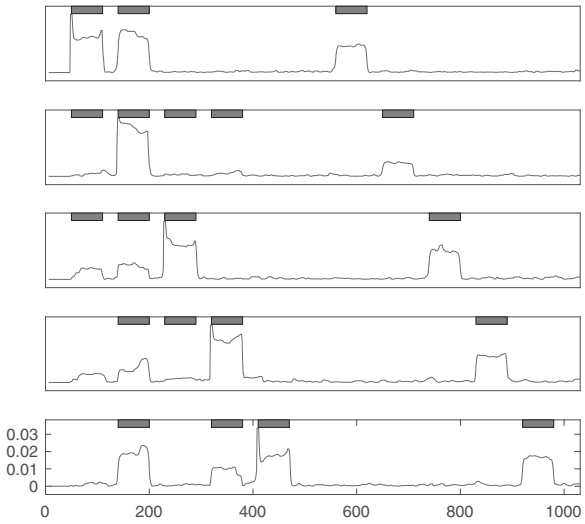


Fig. 5. Residuals for model $\mathcal{M}$ (1st data set).

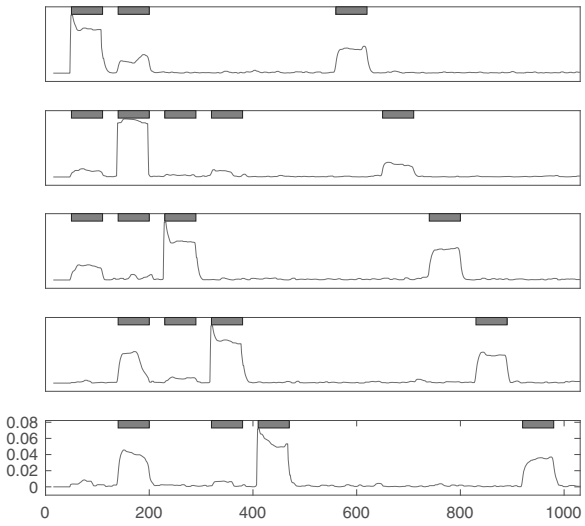


Fig. 6. Residuals for model $\mathcal{M}$ (2nd data set).

this being largely due to the non-linear behaviour of the system.

To conclude, the example presented but also all the others we have treated show that not all residuals are equally sensitive to faults (see the example of Figs. 5 and 6) which is also true for all the examples treated), this being due, on the one hand, to the non-linear nature of the system and, on the other hand, to the nature of the excitations; a combination of these two causes can also reinforce the problem of sensitivity or insensitivity. It is important to realize that the fault signature tables that reflect the influence of a variable on a model are only simple occurrence tables and, therefore, they do not report the weight of the contribution of a given variable to the output of the model. This binary influence

should therefore be complemented by a method using a more appropriate set of residuals and a fault classification technique.

3.2. Detectability and isolability. In order to highlight some potential difficulties in diagnosis for non-linear systems, it is necessary to return to the analysis of theoretical fault signatures and, in particular, to examine whether the faults are *a priori* detectable and localisable. Remember that detection is limited to recognizing the presence of a fault, whereas isolation is intended to locate it, i.e., assign it to a particular variable. The sensitivity problem has a particularly crucial impact on the possibility of isolation. It is therefore necessary to study *a priori* whether the various faults have signatures that allow them to be discriminated against. Beforehand, we remind the importance of the Hamming distance for the comparison of the theoretical binary fault signatures.

Definition 5. (*Localizable fault*) Let us recall that the Hamming distance $d(\boldsymbol{v}, \boldsymbol{w})$ between two vectors $\boldsymbol{v}, \boldsymbol{w} \in \{0,1\}^n$ is the number of coefficients in which they differ. In terms of diagnosis, a fault is said to be localizable with an order k (number of different bits) if its Hamming distance from the signature of the nearest fault is k digits.

This definition applies, of course, to evaluating the distances from a vector to a group of vectors representing particular faults, the minimum value of orders k then allowing a particular fault to be recognized. To illustrate its use, the $\mathcal{M}$ model (4) is considered. Table 2 directly obtained from Table 1 gathers the binary signatures s_i^{th} of faults that can affect the 10 variables $x_i, u_i, i = 1, \ldots, 5$ of the system. The structure of this table shows that all faults are theoretically detectable since each signature has at least one non-zero element. Since the signatures s_5^{th} and s_{10}^{th} are identical, faults affecting x_5 and u_5 are not isolable, but faults affecting the other variables are isolable.

Table 3 shows the theoretical degree of fault localizability. This table, built directly from Definition 4 applied to each column of Table 2, indicates the theoretical signatures s_i^{th} of the faults that can affect the variables $x_i, u_i, i = 1, \ldots, 5$. The columns and rows of Table 3 indicate the variables for which the Hamming distances are evaluated. For example, between the signatures s_4^{th} and s_8^{th}, which relate themselves to variables x_4 and u_3, the distance is 4 as indicated in Table 3 by the light grey cell. The signatures of the faults on x_4 and u_3, both coded in 5 digits (the number of models), therefore different from 4 digits which indicates a good separation of these signatures. This is different for many other situations where the separation of signatures is only 2 digits, with an even more particular situation for signatures of u_5 and x_5 with zero distance. The observations that have just been made concern only theoretical signatures. In the measurement exploitation phase, the situation is different and more subtle since it concerns the comparison of the theoretical signatures with an experimental signature, the latter being sensitive to the fault but also, unfortunately, to random measurement errors.

Table 2. Fault signatures for model $\mathcal{M}$.

	x_1 s_1^{th}	x_2 s_2^{th}	x_3 s_3^{th}	x_4 s_4^{th}	x_5 s_5^{th}	u_1 s_6^{th}	u_2 s_7^{th}	u_3 s_8^{th}	u_4 s_9^{th}	u_5 s_{10}^{th}
f_1	1	1	.	.	.	1	.	.	.	.
f_2	1	1	1	1	.	.	1	.	.	.
f_3	1	1	1	.	.	.	.	1	.	.
f_4	.	1	1	1	.	.	.	.	1	.
f_5	.	1	.	1	1	.	.	.	.	1

Table 3. Hamming's distances between theoretical signatures for model $\mathcal{M}$.

	x_1	x_2	x_3	x_4	x_5	u_1	u_2	u_3	u_4	u_5
x_1	0	2	2	4	4	2	2	2	4	4
x_2	2	0	2	2	4	4	4	4	4	4
x_3	2	2	0	2	4	4	2	2	2	4
x_4	4	2	2	0	2	4	2	4	2	2
x_5	4	4	4	2	0	2	2	2	2	0
u_1	2	4	4	4	2	0	2	2	2	2
u_2	2	4	2	2	2	2	0	2	2	2
u_3	2	4	2	4	2	2	2	0	2	2
u_4	4	4	2	2	2	2	2	2	0	2
u_5	4	4	4	2	0	2	2	2	2	0

3.3. Principle of fault localization. First of all, it should be recalled that fault detection and localization are carried out by comparing the actual signature of the model residual with the theoretical ones of Table 2. As the latter are in binary form, it is therefore appropriate that the actual signature is also in binary form. This is done by numerically evaluating the model residuals and then comparing them to a given threshold. Of course, this thresholding can lead to an error in evaluating the influence of a fault (due to the choice of the threshold but also according to the sensitivity level of the system to faults) and therefore can compromise the result of comparison of experimental and theoretical signatures. Therefore, it is well understood that it is necessary to examine the robustness of detection/localization against an error due to inadequate thresholding and to adopt a detection/localization technique that is tolerant against such an error.

From a practical point of view, at a particular time instant let us consider, e.g., the experimental signature

(from evaluation of 5 model residuals followed by a thresholding):

$$s^{\text{ex}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \end{bmatrix}^T.$$

This signature does not correspond to any theoretical signature of Table 2. Nevertheless, we can calculate the k index which quantifies the number of different digits between this experimental signature and the theoretical signatures of Table 2. Using Definition 4 with s^{ex} and s_i^{th} of Table 2, we obtain

$$\Delta = \begin{bmatrix} 1 & 1 & 1 & 3 & 5 & 3 & 3 & 3 & 3 & 5 \end{bmatrix}^T.$$

The experimental signature is therefore close to the signatures relating to the faults in x_1, x_2 and x_3, this undecided situation being the result of the thresholding of the residuals, but also the size of the signatures. Although simplistic, this example clearly shows the limitations of the localization technique using Boolean residuals. Let us recall that the major difficulty lies in the choice of the threshold allowing the change from a residual to its Boolean form, a choice that is all the more delicate as our purpose concerns a non-linear system whose residuals have very variable sensitivities depending on its operating point.

4. Robust alternative to fault localization

Noting the difficulty of choosing a threshold for the binarization of residuals, the insufficient separating power of the residuals from the five models (Tables 2 and 3), the significant variations in residual sensitivities over time, we propose a diagnostic procedure using a classification approach, the implementation of which is based on angle evaluation between subspaces. In addition to the above, signatures derived from model $\mathcal{M}$ are limited to five components which can make their comparisons difficult in the presence of low sensitivity to certain faults. For this reason, the proposed approach aims to make these comparisons more robust by increasing the number of output prediction models and, consequently, the size of signatures as explained in Section 4.1. Based on these suplementary models, fault isolation is presented in Section 4.2 and illustrated in Section 4.3.

4.1. Extended signature table. It should be remembered that Table 2 of fault signatures was constructed only by observing the occurrence of the ten system variables in the five equations of the model. This table can be largely completed by considering combinations of these five equations, these combinations being selected in order to free themselves from certain variables in order to generate structured fault indicators. For example, the f_1 and f_2 models of Table 1 have x_1 and x_2 as common variables. We can therefore form two additional models from f_1 and f_2 by a simple aggregation followed by a suppression of x_1 or x_2. It is important to note that this operation applies to variable occurrences and does not require knowledge of the f_i functions.

This aggregation/suppression principle leads to Table 4 which totals $N_m = 24$ models but they only present a part of these complementary models. The first column of the table locates the model numbers, the second column specifies which models were merged and the name of the variable whose influence is removed. Columns 3 and 12 of this table indicate the occurrences of the variables in the models but also generate the theoretical signatures s_i^{th} of the faults affecting the variables $x_i, u_i, i = 1, \ldots, 5$. The rows f_1 to f_5 of this table are nothing but those of Table 1 describing model $\mathcal{M}$.

Its construction was done in a systematic way by examining the occurrences of each variable in all the equations. The rows f_6 to f_{11} of model $\mathcal{M}_e$ in Table 4 come from the virtual carry-over of variables from f_1 (i.e., x_1 and x_2) in models f_2 to f_5. For example, f_6 comes from the virtual carryover x_1 from model f_1 in model f_2 and f_7 comes from the carryover x_2 from model f_1 in model f_2. Similarly, rows f_{12} to f_{19} come from the elimination of f_2 (so x_1 or x_2 or x_3 or x_4) with the models f_3 to f_5. Rows f_{20} to f_{22} come from the elimination of f_3 from models f_4 and f_5, rows f_{23} and f_{24} come from the elimination of f_4 from f_5. Other eliminations are possible, but it is to be noted that it is structurally impossible to apply them to u_5 (or x_5) because of the occurrence of this variable which only appears in model f_5.

The following equations give the structure of the 24 non-linear models $\mathcal{M}_e$ whose parameters have been identified according to the procedure indicated in Section 2.4:

$$\begin{aligned}
\hat{x}^1_{1,i} &= f_1(x_1^{h_p}, x_2^{h_p}, u_1^{h_q}),\\
\hat{x}^2_{2,i} &= f_2(x_1^{h_p}, x_2^{h_p}, \quad x_3^{h_p}, x_4^{h_p}, u_3^{h_q}),\\
\hat{x}^3_{3,i} &= f_3(x_1^{h_p}, x_2^{h_p}, x_3^{h_p}, u_3^{h_q}),\\
\hat{x}^4_{4,i} &= f_4(x_2^{h_p}, x_3^{h_p}, x_4^{h_p}, u_4^{h_q}),\\
\hat{x}^4_{5,i} &= f_5(x_2^{h_p}, x_4^{h_p}, x_5^{h_p}, u_5^{h_q}),\\
\hat{x}^2_{6,i} &= f_6(u_1^{h_q}, u_2^{h_q}, x_2^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^3_{7,i} &= f_7(u_1^{h_q}, u_2^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^2_{8,i} &= f_8(u_1^{h_q}, u_3^{h_q}, x_2^{h_p}, x_3^{h_p}),\\
\hat{x}^1_{9,i} &= f_9(u_1^{h_q}, u_3^{h_q}, x_1^{h_p}, x_3^{h_p}),
\end{aligned}$$

Table 4. Table of 24 models $\mathcal{M}_e$.

		x_1 s_1^{th}	x_2 s_2^{th}	x_3 s_3^{th}	x_4 s_4^{th}	x_5 s_5^{th}	u_1 s_6^{th}	u_2 s_7^{th}	u_3 s_8^{th}	u_4 s_9^{th}	u_5 s_{10}^{th}
f_1		×	×	.	.	.	×	.	.	.	.
f_2		×	×	×	×	.	.	×	.	.	.
f_3		×	×	×	.	.	.	.	×	.	.
f_4		.	×	×	×	.	.	.	.	×	.
f_5		.	×	.	×	×	.	.	.	.	×
f_6	$x_1 : f_1, f_2$	.	×	×	×	.	×	×	.	.	.
f_7	$x_2 : f_1, f_2$	×	.	×	×	.	×	×	.	.	.
f_8	$x_1 : f_1, f_3$	.	×	×	.	.	×	.	×	.	.
f_9	$x_2 : f_1, f_3$	×	.	×	.	.	×	.	×	.	.
f_{10}	$x_2 : f_1, f_4$	×	.	×	×	.	×	.	.	×	.
f_{11}	$x_2 : f_1, f_5$	×	.	.	×	×	×	.	.	.	×
f_{12}	$x_1 : f_2, f_3$	.	×	×	×	.	.	×	×	.	.
f_{13}	$x_2 : f_2, f_3$	×	.	×	×	.	.	×	×	.	.
f_{14}	$x_3 : f_2, f_3$	×	×	.	×	.	.	×	×	.	.
f_{15}	$x_2 : f_2, f_4$	×	.	×	×	.	.	×	.	×	.
f_{16}	$x_3 : f_2, f_4$	×	×	.	×	.	.	×	.	×	.
f_{17}	$x_4 : f_2, f_4$	×	×	×	.	.	.	×	.	×	.
f_{18}	$x_2 : f_2, f_5$	×	.	×	×	×	.	×	.	.	×
f_{19}	$x_4 : f_2, f_5$	×	×	×	.	×	.	×	.	.	×
f_{20}	$x_2 : f_3, f_4$	×	.	×	×	.	.	.	×	×	.
f_{21}	$x_3 : f_3, f_4$	×	×	.	×	.	.	.	×	×	.
f_{22}	$x_2 : f_3, f_5$	×	.	×	×	×	.	.	×	.	×
f_{23}	$x_2 : f_4, f_5$	.	.	×	×	×	.	.	.	×	×
f_{24}	$x_4 : f_4, f_5$	.	×	×	.	×	.	.	.	×	×

$$
\begin{aligned}
\hat{x}^4_{10,i} &= f_{10}(u_1^{h_q}, u_4^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^5_{11,i} &= f_{11}(u_1^{h_q}, u_5^{h_q}, x_1^{h_p}, x_4^{h_p}, x_5^{h_p}),\\
\hat{x}^2_{12,i} &= f_{12}(u_2^{h_q}, u_3^{h_p}, x_2^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^3_{13,i} &= f_{13}(u_2^{h_q}, u_3^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^4_{14,i} &= f_{14}(u_2^{h_q}, u_3^{h_q}, x_1^{h_p}, x_2^{h_p}, x_4^{h_p}),\\
\hat{x}^4_{15,i} &= f_{15}(u_2^{h_q}, u_4^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^4_{16,i} &= f_{16}(u_2^{h_q}, u_4^{h_q}, x_1^{h_p}, x_2^{h_p}, x_4^{h_p}),\\
\hat{x}^3_{17,i} &= f_{17}(u_2^{h_q}, u_4^{h_q}, x_1^{h_p}, x_2^{h_p}, x_3^{h_p}),\\
\hat{x}^5_{18,i} &= f_{18}(u_2^{h_q}, u_5^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}, x_5^{h_p}),\\
\hat{x}^5_{19,i} &= f_{19}(u_2^{h_q}, u_5^{h_q}, x_1^{h_p}, x_2^{h_p}, x_3^{h_p}, x_5^{h_p}),\\
\hat{x}^3_{20,i} &= f_{20}(u_3^{h_q}, u_4^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}),\\
\hat{x}^4_{21,i} &= f_{21}(u_3^{h_q}, u_4^{h_q}, x_1^{h_p}, x_2^{h_p}, x_4^{h_p}),\\
\hat{x}^5_{22,i} &= f_{22}(u_3^{h_q}, u_5^{h_q}, x_1^{h_p}, x_3^{h_p}, x_4^{h_p}, x_5^{h_p}),\\
\hat{x}^4_{23,i} &= f_{23}(u_4^{h_q}, u_5^{h_q}, x_3^{h_p}, x_4^{h_p}, x_5^{h_p}),\\
\hat{x}^5_{24,i} &= f_{24}(u_4^{h_q}, u_5^{h_q}, x_2^{h_p}, x_3^{h_p}, x_5^{h_p}),
\end{aligned}
\tag{13}
$$

$$h_p = i-1 : i-p, \quad h_q = i-1 : i-q.$$

The notation $\hat{x}^j_{m,i}$ reminds us that model f_m explains variable x_j at time i. The choice of the explanatory variable is at the user's discretion. For example, model f_6 has x_2 as a variable to explain but x_3 or x_4 could have been candidates.

In order to clarify the notation of the predicted variables, define the vector L containing the indices of the variables reconstructed by the different models $f_m, m = 1, \ldots, 24$:

$$
\begin{aligned}
L = \{&1\ 2\ 3\ 4\ 5\ 2\ 3\ 2\ 1\ 4\ 5\ 2\\
&3\ 4\ 4\ 4\ 4\ 5\ 5\ 3\ 4\ 5\ 4\ 5\},
\end{aligned}
\tag{14}
$$

which allows predictions to be written in the form of $\hat{x}^{L(m)}_{m,i}, m = 1, \ldots, 24$ and the prediction errors

$$\tilde{x}^{L(m)}_{m,i} = \hat{x}^{L(m)}_{m,i} - x_{L(m),i}, \quad m = 1, \ldots, 24. \tag{15}$$

For example, the prediction error for variable x_3 at time i for the 7-th model ($m = 7$) is $\tilde{x}^3_{7,i} = \hat{x}^3_{7,i} - x_{3,i}$

Considering Definition 4, Table 5 clearly shows that switching from theoretical signatures established with model $\mathcal{M}$ (f_1 to f_5, Table 2) to theoretical signatures established with model $\mathcal{M}_e$ (f_1 to f_{24}, Table 4) can improve the degree of separability of theoretical

Table 5. Hamming's distance between theoretical signatures.

	x_1	x_2	x_3	x_4	x_5	u_1	u_2	u_3	u_4	u_5
x_1	0	16	10	11	16	14	10	11	15	16
x_2	16	0	14	15	14	14	12	13	13	14
x_3	10	14	0	9	16	16	10	11	15	16
x_4	11	15	9	0	13	17	11	16	10	13
x_5	16	14	16	13	0	12	14	15	11	0
u_1	14	14	16	17	12	0	14	13	13	12
u_2	10	12	10	11	14	14	0	13	15	14
u_3	11	13	11	16	15	13	13	0	14	15
u_4	15	13	15	10	11	13	15	14	0	11
u_5	16	14	16	13	0	12	14	15	11	0

Table 6. Location of a fault in variable x_1.

Variable	x_1	x_2	x_3	x_4	x_5	u_1	u_2	u_3	u_4	u_5
Index k	2	15	14	9	15	12	10	12	12	16

signatures. Examination of these two tables clearly shows that the separation of signatures improves as the number of models increases. For example, to distinguish signatures related to variables x_1 and x_2, the use of models f_1 to f_5 gives an index k of 2 but increases to 16 if one takes models f_1 to f_{24}. Again the reader will nevertheless notice that the use of 24-bit signatures does not allow the isolability of faults on x_5 and u_5, which is confirmed by a null distance between x_5 and u_5.

Remark 3. Bear in mind the fundamental interest in Table 5 which significantly increases the size of the vectors of fault signature with the sole purpose of improving their classification. For example, consider the case of a fault affecting the variable x_1 for which the 24 model residuals (after application of a threshold) revealed the following experimental signature:

$$\tilde{\boldsymbol{x}}^b = [1\,1\,0\,0\,0\,0\,1\,0\,1\,1\,1\,0\,1\,1\,1\,1\,0\,1\,1\,1\,1\,1\,0\,0].$$

This signature does not exactly correspond to any theoretical signature of Table 4. Nevertheless, we can calculate the k index (see Definition 4) which quantifies the number of different digits between this experimental signature and the theoretical signatures. The examination of the values taken by this index (Table 6) shows, on the one hand, that the experimental signature $\tilde{\boldsymbol{x}}^b$ differs only in 2 digits (out of 24) from the first theoretical signature and that, on the other hand, the second theoretical signature closest to it differs in 9 digits. As a result, there is no ambiguity about the proximity of the experimental signature to the nearest theoretical signature. Despite the error due to the thresholding, we can therefore conclude that the experimental signature makes it possible to find the faulty variable, in this case x_1.

In conclusion, it should be noted that increasing the number of models can improve the separability of theoretical signatures. As a direct consequence, the recognition of an experimental signature to localize a fault is improved. Obviously, this improvement is at the cost of an increase in the volume of calculations, on the one hand, on the identification of model parameters, and on the other hand, on the procedure for generating and analyzing residuals. It should also be borne in mind that the system under consideration is non-linear, which may lead to significant variations in residuals in relation to faults. With this in mind, the following section proposes a new robust approach to the problem of fault detection and localization.

4.2. Proposed diagnosis strategy. As the models $\mathcal{M}$ (4) and $\mathcal{M}_e$ (13) give a very representative view of the functioning of the real system, they can be used for output prediction. They have very similar performances in terms of output prediction quality, but for reasons of separability of signatures, model $\mathcal{M}_e$ is the one used for the diagnostic step. The proposed strategy includes an off-line step dedicated to the generation of a so-called reference fault signature database and an online step that analyzes model residuals using the reference signature database.

Given the sensitivities of the residuals, which vary according to the operating points of the system, a fault signature database is constructed from the $\mathcal{M}_e$ prediction model, this database being the image of a set of various situations that can be reached by the real system. This set of situations was generated for different inputs and for different fault magnitudes, the fault affecting a particular variable being generated in the same time interval for all configurations and amplitude in the interval $[0.05, 0.15]$. For each variable, these time interval are given in Table 7.

4.2.1. Off-line design of an input references database. As previously mentioned, the construction of reference signatures requires prior generation of excitation signals for the system (its inputs) in order to generate data representative of the system's operating domain. This necessity is well known in the field of parameter identification of a system and in particular in the linear case where techniques for the construction of experimental designs are available (Dagnelie, 2012). Here, we are dealing with a problem of diagnosis of non-linearly operating systems for which, rather than the usual techniques, an experimental construction of this experimental design was preferred. The system inputs were generated by a series of random magnitude and duration pulses chosen in a bounded interval, $[0.2,\ 0.4]$ for the magnitudes and $[25,\ 65]$ for the durations. Such a signal is represented in the upper part of Fig. 7, the lower part resulting from low-pass filtering of this signal in order

Table 7. Time intervals for the presence of faults.

	x_1	x_2	x_3	x_4	x_5
I	[50:110]	[140:200]	[230:290]	[320:380]	[410:470]
	u_1	u_2	u_3	u_4	u_5
I	[560:620]	[650:710]	[740:800]	[830:890]	[920:980]

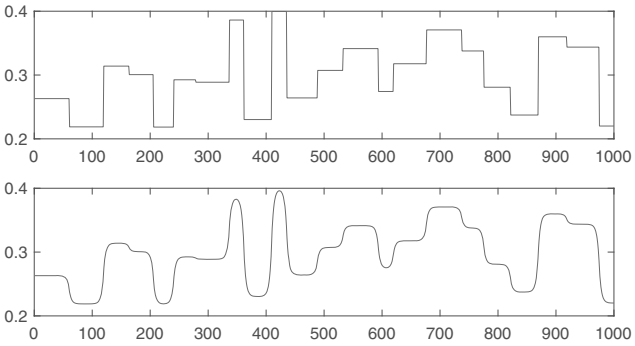


Fig. 7. Input u_1 reference design.

to avoid abrupt variations in the system excitation. The example presented is obviously valid for the five inputs of the system, which then makes it possible, thanks to the previously identified model, to generate the five outputs corresponding to these five inputs.

The construction just described is thus a set of input/output data, which is then called a configuration. Obviously, several configurations will be generated afterwards in order to have a set of reference situations. In the following, 12 configurations will be used.

4.2.2. Generation of an input/output reference database. From the previously constructed reference data, reference data for input/output faulty situations were then generated, with faults having an additive influence on the data. Table 7 specifies the time domains where these faults are active. In order to have a set of situations with faults, it is advisable, as for the generation of excitation inputs, to consider several fault magnitudes in a bounded domain, here $[0.02,\ 0.08]$. This domain has been chosen in relation to the range of variation of the inputs/outputs of the system. Table 8 displays the relative importance of the faults affecting the inputs and outputs of the system by indicating the ratios of the fault magnitudes and the mean values of the signals for the chosen configurations $C_j, j = 1, \ldots, 12$. We can see that the average importance is around 25%.

This database was developed with 12 input configurations denoted by $C_j, j = 1, \ldots, 12$ (this number being adapted by the user according to the systems to be analyzed) and for each configuration C_j the excitations $(u_i, i = 1, \ldots, 5)$ were chosen in the same operating domains.

4.2.3. Generation of a residual database with 12 configurations. Then, we have built a database containing residuals in the presence of faults and for different operating modes. For each configuration, model $\mathcal{M}_e$ generates a residual vector at 24 components. As an example, Fig. 8 shows the 24 residuals related to the first configuration C_1, where the faults were created on the five outputs and inputs (time intervals I in Table 7). The grey markers display the time intervals at which faults were created on the system's output and input measurements.

Obviously, for this configuration, a number of residuals show a lack of sensitivity to faults, which may change for other inputs: for other configurations, due to different excitations, the sensitivities of the residuals may be different and reflect differently the presence of faults. It should also be noted that the non-linear nature of the system leads to residual values that are sensitive to the operating point and the fault magnitude. The simple ratio of the signal magnitude to the fault magnitude, which is widely used in the linear case to adjust the detection threshold, may be ineffective here. Thus with all the twelve configurations (this number to be adapted) we can have an image of a large number of situations. Of course, the underlying difficulty is to ensure that the majority of situations are covered.

Remark 4. One may wonder about the interest in keeping models whose sensitivity to certain faults seems low, or even very low. To analyze this possibility and to give an idea of the sensitivity, Table 9 indicates for configuration C_1 the mean magnitudes of the residuals of 24 models when faults are applied to 10 variables in the time intervals defined in Table 7. Although this is only one configuration, some observations can be made, especially for the magnitudes marked by a light grey background. The model M_{13} has a sensitivity of 0.04 for a fault on x_4. However this model has a proven sensitivity to other variables such as x_3 and u_3 and for this reason its role must be kept. The same remark can be made about model M_{14}. Model M_{14} indicates a sensitivity of 0.06 with respect to variable x_2. However, as we note that the other variables also have weak influence, the role of this model may seem questionable and it could be removed from the list of models useful for diagnosis.

4.2.4. Generation of a fault signature database. The 12 configurations allow to extract 10 reference tables, each corresponds to a fault on the 5 states and the 5 inputs of the system. For example, Table 10, relating to a fault in x_1, contains the average values of the residues calculated on the time interval $I_1 = [50\ :\ 110]$, for 24 models M_e and for configurations $C_j, j = 1, \ldots, 12$. The reader will notice that the first column of table T_1 has been used to form the first column of Table 9.

The dispersion of the numerical values between

Table 8. Relative average magnitudes of input and output faults.

	u_1	u_2	u_3	u_4	u_5
C_1	0.10	0.10	0.10	0.10	0.10
C_2	0.14	0.14	0.14	0.14	0.14
C_3	0.17	0.17	0.17	0.17	0.17
C_4	0.21	0.21	0.21	0.21	0.21
C_5	0.22	0.21	0.22	0.22	0.22
C_6	0.26	0.25	0.26	0.26	0.25
C_7	0.26	0.26	0.26	0.26	0.26
C_8	0.26	0.26	0.26	0.26	0.26
C_9	0.26	0.27	0.26	0.26	0.27
C_{10}	0.27	0.27	0.27	0.27	0.27
C_{11}	0.31	0.31	0.31	0.31	0.31
C_{12}	0.35	0.35	0.35	0.35	0.36

	x_1	x_2	x_3	x_4	x_5
C_1	0.12	0.05	0.08	0.07	0.06
C_2	0.18	0.09	0.11	0.12	0.09
C_3	0.22	0.14	0.14	0.15	0.13
C_4	0.26	0.18	0.18	0.20	0.14
C_5	0.27	0.19	0.19	0.20	0.15
C_6	0.33	0.23	0.22	0.25	0.17
C_7	0.35	0.22	0.22	0.27	0.19
C_8	0.34	0.24	0.22	0.25	0.19
C_9	0.33	0.24	0.22	0.25	0.19
C_{10}	0.33	0.24	0.22	0.27	0.19
C_{11}	0.37	0.28	0.26	0.31	0.23
C_{12}	0.45	0.33	0.31	0.35	0.26

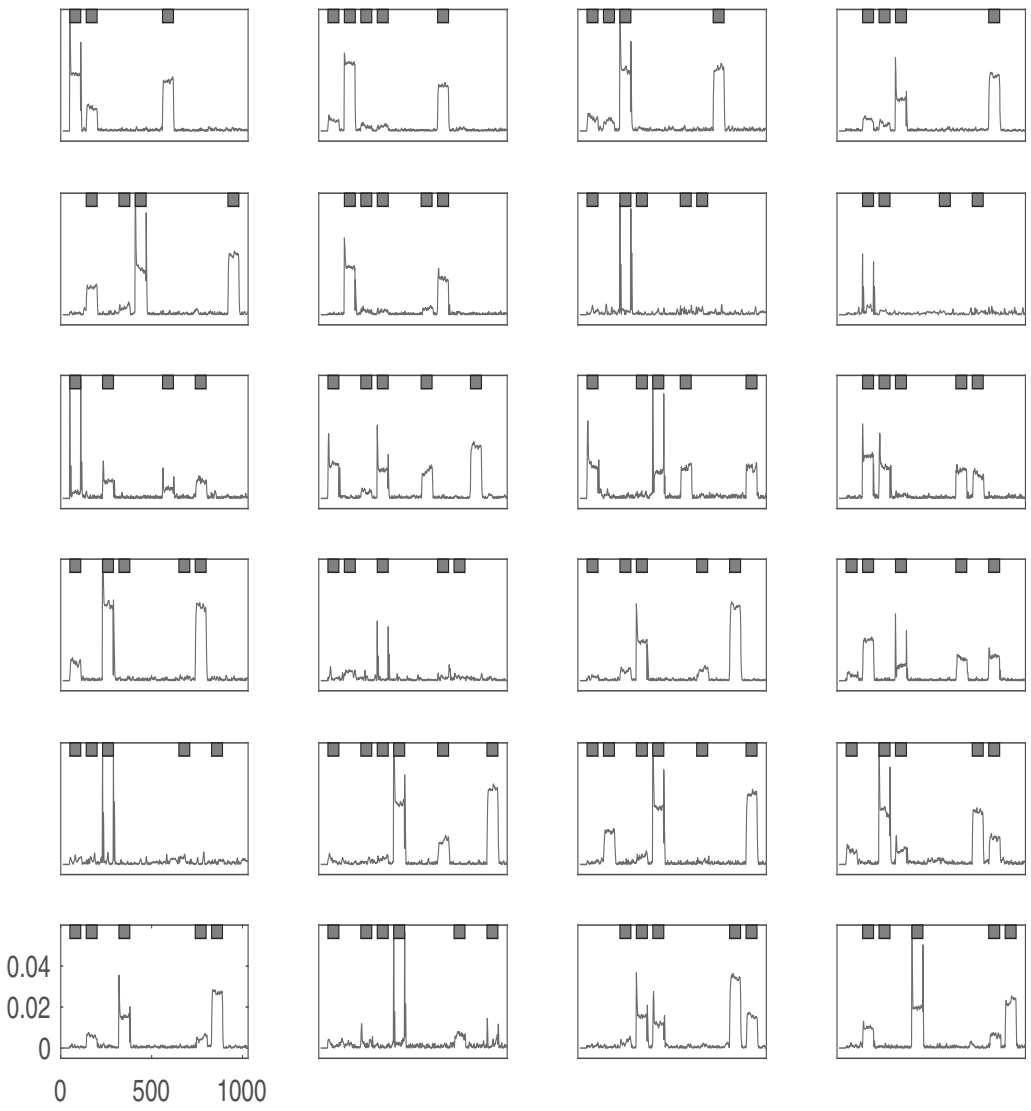


Fig. 8. Residuals from model $\mathcal{M}_e$ (configuration C_1).

Table 9. Residuals of the 24 models, obtained with configuration C_1.

	δx_1	δx_2	δx_3	δx_4	δx_5	δu_1	δu_2	δu_3	δu_4	δu_5
M_1	3.06	1.44	0.02	0.04	0.02	2.66	0.03	0.02	0.02	0.02
M_2	0.55	3.44	0.25	0.30	0.00	0.01	1.89	0.02	0.01	0.00
M_3	0.41	0.56	3.30	0.03	0.05	0.05	0.06	3.06	0.03	0.05
M_4	0.02	0.73	0.17	1.71	0.02	0.01	0.03	0.07	2.78	0.01
M_5	0.02	1.54	0.17	0.47	2.85	0.06	0.04	0.02	0.04	2.95
M_6	0.02	2.53	0.18	0.23	0.03	0.38	1.54	0.02	0.02	0.00
M_7	0.27	0.13	0.63	0.13	0.20	0.29	0.20	0.22	0.14	0.07
M_8	0.10	0.59	0.33	0.09	0.07	0.24	0.09	0.49	0.14	0.07
M_9	0.80	0.07	1.19	0.07	0.03	0.76	0.06	1.14	0.07	0.04
M_{10}	0.95	0.04	0.16	1.28	0.01	0.95	0.01	0.05	2.15	0.02
M_{11}	1.51	0.04	0.16	0.22	1.34	1.23	0.08	0.03	0.07	1.45
M_{12}	0.05	1.97	1.47	0.12	0.03	0.05	1.30	1.27	0.03	0.03
M_{13}	0.69	0.01	3.79	0.04	0.01	0.03	0.57	3.43	0.02	0.03
M_{14}	0.09	0.06	0.08	0.14	0.10	0.09	0.07	0.14	0.08	0.07
M_{15}	0.12	0.02	0.43	2.02	0.01	0.02	0.56	0.03	3.60	0.00
M_{16}	0.19	2.13	0.05	0.76	0.02	0.04	0.97	0.04	1.21	0.01
M_{17}	0.11	0.16	0.69	0.15	0.11	0.11	0.23	0.20	0.20	0.09
M_{18}	0.17	0.02	0.12	0.38	3.23	0.03	1.04	0.04	0.02	3.54
M_{19}	0.10	1.30	0.05	0.45	3.14	0.02	0.25	0.02	0.02	3.32
M_{20}	0.27	0.04	1.80	0.83	0.03	0.06	0.06	1.55	1.11	0.04
M_{21}	0.04	0.68	0.03	1.55	0.03	0.02	0.03	0.19	2.59	0.02
M_{22}	0.11	0.04	0.67	0.13	0.82	0.10	0.11	0.93	0.05	0.50
M_{23}	0.03	0.05	0.35	1.58	1.25	0.05	0.03	0.12	3.10	1.27
M_{24}	0.02	0.99	0.16	0.03	2.15	0.03	0.02	0.05	0.54	2.35

12 situations is obviously due to the non-linear nature of the system. These twelve residues are supposed to be representative of the different operating regimes of the system. For the other nine variables ($x_i, i = 2, \ldots, 5, u_i, i = 1, \ldots, 5$) similar tables $T_i, i = 2, \ldots, 10$ were developed. The ten tables T_i can then be used as references for situations with faults. One can, of course, compare the results of Tables 4 and 10 indicating respectively the occurrences of the faults of the variable x_1 and their sensitivities in the model equations. For example, the first three rows of Table 10 indicate a significant sensitivity of a fault in x_1 on the first three models, as opposed to the two following rows showing a very low sensitivity of this fault in the two following models, which is well in agreement with the occurrences of the first five rows of Table 4.

4.2.5. On-line diagnosis: Fault localization. In this step, where we try to identify the variable affected by a fault in each new observation, we will need to use a rather particular indicator which is the main angle. This main angle can be considered as a generalization of the concept of the angle between two straight lines or two planes to hyperplanes in spaces of any size. Subsequently, this angle will be used to measure the proximity between an experimental signature and the set of reference signatures established for the ten potential faults and different operating points of the system.

Definition 6. (*Angles between subspaces (Jordan, 1875; Gunawan, 2018)*) Let $(X, \langle \cdot, \cdot \rangle)$ be a real inner product space, which will be our ambient space throughout this definition. Let $U = \text{span}\{u_1, \ldots, u_p\}$ and $V = \text{span}\{v_1, \ldots, v_q\}$ be two subspaces of X with $1 \leq p \leq q$. The angle θ between U and V is given by

$$\cos^2 \theta = \frac{\det(M^T M)}{\det(N_u)\det(N_v)} \tag{16}$$

with $M := [\langle u_i, v_k \rangle]^T$ being a $q \times p$ matrix, $N_u := [\langle u_i, u_j \rangle]^T$ a $p \times p$ matrix and $N_v := [\langle v_k, v_\ell \rangle]^T$ a $q \times q$ matrix.

If the principle angle between the two subspaces is small, they are nearly dependent. The above definition allows for the effective calculation of the angle between subspaces. Many techniques have been developed to allow this calculation with the desired accuracy and the reader is referred to specialized work on this subject such as that by Nur *et al.* (2018) or Gunawan (2018) and the publications cited therein.

Therefore, we reduce the problem of fault localization to a problem of signal classification. Indeed,

Table 10. Residual table T_1 with a fault on x_1, obtained with 12 simulations.

	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$	$C_{1,5}$	$C_{1,6}$	$C_{1,7}$	$C_{1,8}$	$C_{1,9}$	$C_{1,10}$	$C_{1,11}$	$C_{1,12}$
M_1	3.06	3.57	3.23	3.38	3.39	3.2	3.28	3.31	3.31	3.27	3.30	3.30
M_2	0.55	0.80	0.59	0.73	0.57	0.64	0.70	0.65	0.71	0.64	0.64	0.59
M_3	0.41	1.24	0.84	0.99	0.87	0.69	0.89	0.94	0.92	0.83	0.92	0.84
M_4	0.02	0.10	0.03	0.05	0.02	0.02	0.05	0.01	0.02	0.03	0.00	0.06
M_5	0.02	0.06	0.08	0.06	0.05	0.02	0.07	0.02	0.03	0.05	0.03	0.05
M_6	0.02	0.07	0.01	0.03	0.02	0.02	0.03	0.01	0.00	0.00	0.00	0.01
M_7	0.27	0.14	0.22	0.46	0.25	0.31	0.36	0.26	0.40	0.50	0.24	0.45
M_8	0.10	0.20	0.09	0.08	0.07	0.08	0.07	0.04	0.06	0.07	0.04	0.10
M_9	0.80	0.37	0.58	0.52	0.55	0.61	0.56	0.51	0.54	0.62	0.52	0.58
M_{10}	0.95	1.14	1.13	1.15	1.31	1.08	1.08	1.18	1.07	1.07	1.11	1.10
M_{11}	1.51	1.84	1.69	1.74	1.79	1.62	1.73	1.74	1.70	1.65	1.69	1.69
M_{12}	0.05	0.07	0.02	0.05	0.03	0.03	0.03	0.01	0.04	0.03	0.01	0.03
M_{13}	0.69	1.19	0.86	0.97	0.83	0.88	0.91	0.96	0.94	0.83	0.94	0.80
M_{14}	0.09	0.12	0.13	0.08	0.10	0.07	0.05	0.08	0.04	0.03	0.05	0.06
M_{15}	0.12	0.12	0.15	0.14	0.16	0.13	0.12	0.12	0.12	0.13	0.11	0.13
M_{16}	0.18	0.43	0.32	0.39	0.29	0.30	0.39	0.34	0.37	0.35	0.34	0.32
M_{17}	0.11	0.21	0.06	0.14	0.13	0.17	0.13	0.09	0.12	0.20	0.05	0.18
M_{18}	0.17	0.31	0.23	0.26	0.28	0.25	0.24	0.30	0.26	0.16	0.27	0.19
M_{19}	0.10	0.05	0.09	0.05	0.09	0.08	0.02	0.07	0.03	0.04	0.05	0.06
M_{20}	0.27	0.53	0.54	0.52	0.49	0.39	0.50	0.53	0.52	0.53	0.52	0.50
M_{21}	0.04	0.02	0.10	0.14	0.09	0.06	0.14	0.07	0.11	0.09	0.08	0.07
M_{22}	0.11	0.38	0.20	0.23	0.23	0.22	0.28	0.27	0.23	0.10	0.23	0.15
M_{23}	0.03	0.09	0.05	0.03	0.05	0.03	0.04	0.02	0.02	0.03	0.02	0.05
M_{24}	0.02	0.09	0.07	0.05	0.03	0.01	0.03	0.02	0.00	0.05	0.03	0.02

the preceding tables ($T_i, i = 1, \ldots, 10$) finally represent a dictionary of reference situations corresponding to the ten faults, each reference situation being itself the result of 12 configurations obtained for different system inputs (see Table 10 for the fault example regarding x_1).

At each time instant i, the fault localization strategy is based on the comparison of the experimental residual vector $\tilde{x}_i$ with components $\tilde{x}_{m,i}^{L(m)}$ obtained from 24 models (13) with the T_i signature tables. This comparison is performed here by analyzing the angles between the experimental residual vector and the subspaces generated by the column vectors of the T_i tables. To this end, at time instant i, (16) is applied with

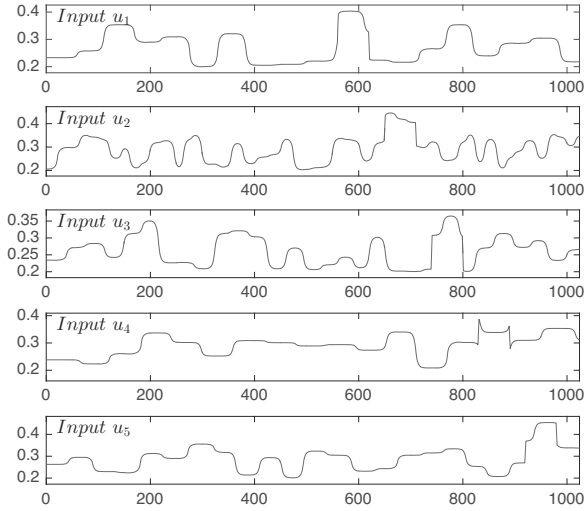
$$\begin{aligned}
&\tilde{x}_i = \begin{bmatrix} \tilde{x}_{1,i}^{L(1)} & \ldots & \tilde{x}_{24,i}^{L(24)} \end{bmatrix}^T, \quad i = 1, \ldots, N,\\
&U_i = \operatorname{span}\{\tilde{x}_i\},\\
&V_j = \operatorname{span}\{C_{j,1}, C_{j,2}, \ldots, C_{j,12},\}, \quad j = 1, \ldots, 10,\\
&\qquad \tilde{x}_i \in \mathbb{R}^{24}, \quad U_i \in \mathbb{R}^{24}, \quad V_j \in \mathbb{R}^{24\times 12},
\end{aligned} \tag{17}$$

where $C_{.,.}$ are defined in tables T_i, Table 10 showing the example of T_1 for a fault affecting x_1. As a result, at each time instant i, 10 angles $\theta_{i,j}, j = 1, \ldots, 10$ are obtained, reflecting the proximity of the experimental signature to the ten theoretical signatures, each of these signatures corresponding to a possible fault; the minimum angle then allows the fault to be isolated:

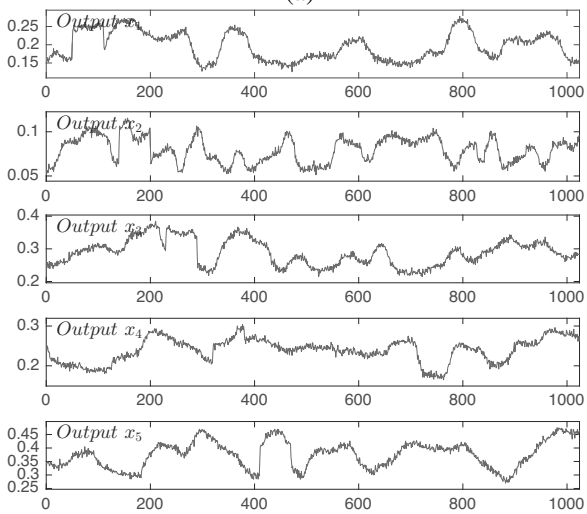
$$j_{\min,i} = \arg \min_{j\in(1,10)} \theta_{i,j}. \tag{18}$$

4.3. Some numerical results about fault localisation. Figures 9(a) and (b) respectively show a set of measurements of the system's inputs and outputs, the latter being corrupted by noise from a uniform distribution. The maximum noise magnitude has been set at 4% of the amplitude of the variation ranges of each signal. The faults were generated in the time intervals I defined in Table 7. The procedure in Section 4.2 is applied to these data; it includes fault detection and fault localization. At each time instant i, the residual vector of the 24 models is evaluated numerically. The main angle $\theta_{1,i}$ between this vector and the subspaces generated by the columns of T_1 (relative to a fault in variable x_1) is evaluated. This operation is repeated for all tables T_j and finally we have angles $\theta_{j,i}, j = 1, \ldots, 10$.

For example, Fig. 10 shows the evolution of the principal angles $\theta_{2,i}$ and $\theta_{6,i}$ over time, angles estimated using tables T_2 and T_6. The first 40 points of these graphs are not to be taken into account for initialization reasons, so we can clearly see principal angle values of $\theta_{2,i}$ and



(a)



(b)

Fig. 9. System inputs (a) and outputs (b).

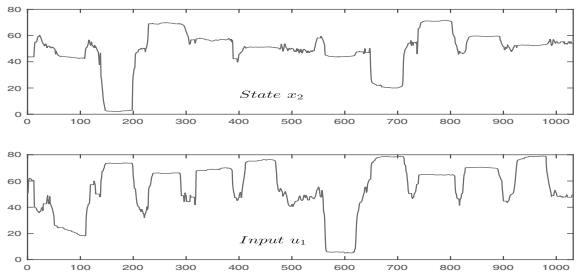


Fig. 10. Angles ($\theta_{2,i}$, $\theta_{6,i}$) associated to faults on (x_2, u_1).

$\theta_{6,i}$ respectively close to 0 in the intervals $[49 : 117]$ and $[561 : 624]$. Outside these intervals the principal angles are significantly different from zero. Taking these two graphs into account, we can therefore conclude that a fault affects the second variable, i.e., x_2 and the sixth variable, i.e., u_1 in the respective intervals $[49 : 117]$ and $[561 : 624]$. The previous step is of course performed at any time for the ten principal angles. Table 11 presents a comparison then the "true" and "estimated" time intervals for the presence of faults.

A more global and synthetic view of the results of fault location is proposed. For each of the ten time intervals (11) identified as fault carrying, the main angles have been replaced by their respective averages over these intervals. Table 12 illustrates the *localization* results using the classification principle based on the principal angle computation between subspaces. Thus, for the first time interval $\hat{I}_1$, where a fault exists, the value 3 is the angle between the subspace generated by the residual vector and the subspace generated by the 12 column vectors of the table T_1 which relates to a fault affecting variable x_1. The 9 other values of the column $\hat{I}_1$ a being calculated with the column vectors of the other tables $T_i, i = 2, \ldots, 10$. The value 3 is the smallest in column $\hat{I}_1$; the fault detected on the time interval $\hat{I}_1$ is therefore the one that affects variable x_1. The other time intervals $\hat{I}_i, i = 2, \ldots, 10$ are analyzed in a similar way.

The values on the diagonal of the table correspond to the smallest angles in each column: in the 10 intervals, the faults that affect the variables have been located. However, we again note the ambiguity of localization between the faults affecting the variables x_5 and u_5 (medium grey boxes), an ambiguity that is entirely justified by the structure of model M_e. We can also note for the time interval $\hat{I}_3$ the proximity of angles 4 and 9 which can possibly create a doubt on the failed variable, namely x_3 or u_3. A similar remark applies to interval $\hat{I}_8$ concerned by the proximity of values 3 and 6.

Remark 5. As mentioned above, since this is a system with non-linear behaviour, the prediction model obtained can only be valid in the operating domain in which it was developed. In other words, it is valid for the input variation range with which it was identified and tested. However, it is questionable to what extent it can be used in the margins of this area. An FDI test was carried out in this respect. In previous tests (see Fig. 9(a) for example) the variation range of the inputs was $[0.2,\ 0.4]$, it is now increased to $[0.25,\ 0.50]$ so a little disjointed from the domain that was used to build the output prediction model. The same fault configuration was applied and Table 13 gathers the detection/localization results. All the conclusions of Table 12 apply to this new situation, which shows that the proposed approach is somewhat robust. Of course, this conclusion applies to this example and its generalization requires further analysis. In the case of inconclusive localization results, the implementation of the procedure must be reviewed by questioning the database used to build the prediction model and the structure of the model, particularly the delays p and q of its dynamic part.

Table 11. Time intervals in the presence of faults.

	x_1	x_2	x_3	x_4	x_5
I	[50:110]	[140:200]	[230:290]	[320:380]	[410:470]
$\hat{I}$	[49:117]	[139:205]	[229:297]	[314:385]	[409:470]
	u_1	u_2	u_3	u_4	u_5
I	[560:620]	[650:710]	[740:800]	[830:890]	[920:980]
$\hat{I}$	[561:624]	[651:713]	[741:804]	[830:895]	[921:984]

Table 12. Fault isolation using the subspace angle approach.

	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$	$\hat{I}_4$	$\hat{I}_5$	$\hat{I}_6$	$\hat{I}_7$	$\hat{I}_8$	$\hat{I}_9$	$\hat{I}_{10}$
T_1	3	56	67	54	65	16	66	71	57	58
T_2	55	3	68	67	73	76	18	54	84	74
T_3	58	59	4	78	85	80	62	6	82	85
T_4	67	66	79	5	75	64	43	68	15	77
T_5	68	75	77	29	3	70	75	76	83	3
T_6	11	49	71	57	69	4	69	70	51	59
T_7	58	18	71	66	86	75	3	49	81	87
T_8	59	61	9	82	86	81	69	3	86	85
T_9	69	72	82	17	81	61	40	72	2	86
T_{10}	68	75	77	29	3	70	75	76	84	3

Table 13. Fault isolation for the validation test.

	$\hat{I}_1$	$\hat{I}_2$	$\hat{I}_3$	$\hat{I}_4$	$\hat{I}_5$	$\hat{I}_6$	$\hat{I}_7$	$\hat{I}_8$	$\hat{I}_9$	$\hat{I}_{10}$
T_1	14	41	49	49	55	18	48	49	51	43
T_2	65	7	56	42	63	66	22	58	55	62
T_3	52	49	5	49	53	36	36	10	57	51
T_4	55	34	38	6	49	56	46	49	13	43
T_5	47	61	61	60	5	51	60	56	57	7
T_6	21	45	53	58	57	10	57	54	57	53
T_7	59	20	51	43	67	57	7	46	57	59
T_8	59	60	13	58	51	45	43	5	64	45
T_9	56	38	45	13	54	60	61	57	5	53
T_{10}	49	61	63	59	7	53	58	61	59	7

In summary, the proposed approach is based on three essential points: the use of structured prediction models to promote the separation of fault signatures, the use of a reference signature database to cover different operating situations and reduce problems of variable residue sensitivity, and finally, the use of a signature recognition technique based on angles between subspaces.

5. Conclusion

The problem of detecting faults has been addressed here using a data processing technique. In the absence of an *a priori* physical model, the proposed technique is of a numerical nature and proceeds by establishing an empirical model providing redundancies between system outputs. This is a well-known technique in the case of systems with linear behaviour, but which is developed here in a non-linear framework. The fault detection aspect has been discussed earlier in a non-linear framework, in particular with non-linear principal component analysis, but it should be recalled that KPCA remains of problematic use as far as fault localization is concerned.

Our approach provides some answers to the problem of fault detection and localization, as well as a discussion of comments on the sensitivity of residuals to faults. First of all, we proposed a non-linear approach for the system's output prediction. At a given time the equation error model is defined by non-linear regression depending on the previous measurements of the input and the output of the system. Secondly, comparing the output with its prediction allows us to construct a residual which is *a priori* sensitive to faults. Thus the magnitude of these residuals allows us to detect when a fault occurs. Then, in order to clarify the localization of the fault, a method for isolation is proposed. Isolation needs to design reduced prediction output model involving only a part of the process variables. Our approach has focused on the design of structured fault indicators, i.e., those that are sensitive to specific faults. These residuals were developed from black box prediction models requiring no knowledge of the relationships between system variables. In order to reduce the complexity of these residuals, knowledge of the interactions between variables was used without knowing *a priori* the parameters describing these interactions. This second approach has made it possible to structure fault indicators in a more relevant way.

Modelling efforts are not sufficient to completely solve the fault isolation problem, largely due to the non-linearities of the system. Indeed, these non-linearities induce, depending on the operating regime of the system, significant changes in the sensitivity of the models to faults. This is the reason for a major difficulty in fault detection and isolation. To address this problem, we have proposed a classification-type approach by developing a database of reference residuals in the presence of faults, residuals covering a large number of operating situations. To identify a current situation, the residuals generated from the prediction of the system outputs are then compared with the reference residuals by calculating angles between subspaces. The proposed approach has proven effective for all cases treated, but it should be emphasized that particular excitations on the system can make the diagnosis ineffective.

The relevance of the proposed approaches is illustrated on a simulated CSTR characterized by a MIMO structure allowing to have some redundancies for the residual generation and, in particular, the inter-redundancy relationships between the outputs independent partly or entirely of the inputs. Redundancy has been intensively used to generate structured residuals each of them involving a particular set of variables

allowing fault detection and isolation.

Among the major problems that can be the subject of future work, the one of fault detection conditions remains in its entirety. It covers various aspects, in particular the choice of excitation inputs that effectively allow to detect and localise faults and the sensibility analysis of the residuals to the faults. In addition, it would be interesting to recognize for which excitation class fault detection/localization cannot be undertaken due to the insensitivity problem mentioned previously.

References

Alcala, C. and Qin, S. (2010). Reconstruction-based contribution for process monitoring with kernel principal component analysis, *Industrial and Engineering Chemistry Research* **49**(17): 7849–7857.

Bates, D.M. and Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.

Benaicha, A., Mourot, G., Benothman, K. and Ragot, J. (2013). Determination of principal component analysis models for sensor fault detection and isolation, *International Journal of Control, Automation and Systems* **11**(2): 296–305.

Blanke, M., Kinnaert, M., Lunze, J. and Staroswiecki, M.D. (2006). *Diagnosis and Fault-Tolerant Control*, Springer, Berlin.

Botre, C., Mansouri, M., Karim, N., Nounou, H. and Nounou, M. (2016). Kernel PLS-based GLRT method for fault detection of chemical processes, *Journal of Loss Prevention in the Process Industries* **43**: 212–223.

Buchberger, B. and Kauers, M. (2010). Bases de Gröbner, *Scholarpedia* **4249**(5): 7763, DOI: 10. 4249/scholarpedia.

Dagnelie, P. (2012). *Principes d'Expérimentation: Planification des Expériences et Analyse de leurs Résultats*, Presses Agronomiques, Gembloux.

Diop, S. (1989). A state elimination procedure for nonlinear systems, *in* J. Descusse *et al.* (Eds), *New Trends in Nonlinear Control Theory*, Springer, Berlin, pp. 190–198.

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

Frank, P.M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—A survey and some new results, *Automatica* **26**(3): 459–474.

Gao, Q., Liu, W., Zhao, X., Li, J. and Yu, X. (2017). Research and application of the distillation column process fault prediction based on the improved KPCA, *IEEE International Conference on Mechatronics and Automation, Takamatsu, Japan*, pp. 247–251.

Gu, S., Liu, Y., Zhang, N. and Du, D. (2015). Fault detection approach based on weighted principal component analysis applied to continuous stirred tank reactor, *Open Mechanical Engineering Journal* **9**(1): 966–972.

Gunawan, H. (2018). n-Inner products, n-norms and angles between two subspaces, *in* M. Rozhansky *et al.* (Eds), *Mathematical Analysis and Applications: Selected Topics*, Wiley, Hoboken, pp. 493–515.

Huang, D., Zhang, D., Liu, Y., Zhang, S. and Zhu, W. (2014). A KPCA based fault detection approach for feed water treatment process of coal-fired power plant, *11th World Congress on Intelligent Control and Automation, Shenyang, China*, pp. 3222–3227.

Jia, Q. and Huang, Y. (2016). Quality-related fault detection approach based on dynamic kernel partial least squares, *Chemical Engineering Research and Design* **106**: 242–252.

Jordan, C. (1875). Essai sur la géométrie n dimensions, *Bulletin de la Société Mathématique de France* **3**: 103–174.

Jun, B.H., Park, J.H., Lee, S.I. and Chun, M.G. (2006). Kernel PCA based faults diagnosis for wastewater treatment system, *in* Wang J. *et al.* (Eds), *Advances in Neural Networks*, Lecture Notes in Computer Science, Vol. 3973, Springer, Berlin, pp. 426–431.

Kallas, M., Mourot, G., Maquin, D. and Ragot, J. (2018). Data driven approach for fault detection and isolation in nonlinear system, *International Journal of Adaptive Control and Signal Processing* **32**(11): 1569–1590.

Kallas, M., Mourot, G., Maquin, D. and Ragot, J. (2014). Diagnosis of nonlinear systems using kernel principal component analysis, *11th European Workshop on Advanced Control and Diagnosis, Berlin, Germany*.

Manikandan, P., Geetha, M., Jubi, K. and Jovitha, J. (2013). Fault tolerant fuzzy gain scheduling proportional-integral-derivative controller for continuous stirred tank reactor, *Australian Journal of Basic and Applied Sciences* **7**(13): 84–93.

Navi, M., Meskin, N. and Davoodi, M. (2018). Sensor fault detection and isolation of an industrial gas turbine using partial adaptive KPCA, *Journal of Process Control* **64**: 37–48.

Nithya, N. and Vijayachitra, S. (2018). Fault detection and analysis using statistical data for continuous stirred tank reactor, *International Journal of ChemTech Research* **11**(2): 266–274.

Nur, M., Gunawan, H. and Neswan, O. (2018). A formula for the g-angle between two subspaces of a normed space, *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* **59**(1): 133–143.

Ren, Z., Hou, J. and Zhou, H. (2016a). Fault detection and process monitoring of industrial process based on spherical kernel T-PLS, *42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy*, pp. 7161–7166.

Ren, Z., Liu, J., She, Z., Yang, C. and Han, Y. (2016b). Data-driven approach of FS-SKPLS monitoring with application to wastewater treatment process, *IEEE International Conference on Industrial Technology, Taipei, Taiwan*, pp. 550–555.

Ritt, J.F. (1950). *Differential Algebra*, American Mathematical Society, New York.

Simani, S., Farsoni, S. and Castaldi, S. (2018). Data-driven techniques for the fault diagnosis of a wind turbine benchmark, *International Journal of Applied Mathematics and Computer Science* **28**(2): 247–268, DOI: 10.2478/amcs-2018-0018.

Smyth, G. (2006). *Nonlinear Regression*, Wiley, Hoboken.

Wang, L. (2018). Enhanced fault detection for nonlinear processes using modified kernel partial least squares and the statistical local approach, *Canadian Journal Chemical Engineering* **96**(5): 1116–1126.

Wang, Z.Y. and Wan, G.X. (2017). Temperature fault-tolerant control system of CSTR with coil and jacket heat exchanger based on dual control and fault diagnosis, *Journal of Central South University* **24**(3): 655–664.

José Ragot is a full professor of automatic control at Ecole Nationale Supérieure de Géologie (a French "Grande Ecole" in Institut National Polytechnique de Lorraine) and a researcher at Centre de Recherche en Automatique de Nancy (CRAN, CNRS UMR 7039), where he has been the head of the diagnosis group for 12 years. His major research fields include data validation and reconciliation, process diagnosis, fault detection and isolation, and fault tolerant control. Another part of his activities is devoted to modelling adapted to process diagnosis, mainly in the field of multi-models. He has successfully advised 70 PhD theses and published about 500 refereed technical communications, including 150 papers in international journals, 300 communications in international conferences, 100 communications in national conferences, four books and 12 chapters in collective volumes.

Gilles Mourot received his PhD in electrical engineering in 1993 from Institut National Polytechnique de Lorraine (INPL), Nancy, France. His research activities within the Automatic Control Research Center of Nancy (CRAN) include data-driven fault detection as well as isolation and multiple model identification of nonlinear system.

Maya Kallas was born in Zahle, Lebanon, in 1985. She received her engineering degree in computing and telecommunications in 2008 from the Holy Spirit University of Kaslik (USEK), Lebanon, her MS degree in industrial control in 2009 from Lebanese University (UL), Lebanon, and her MR in science, technology and health in 2009 from the University of Technology of Compiégne, France. In 2012, she received a PhD degree in optimization and safety of systems from the University of Technology of Troyes (UTT), France, and a PhD degree in engineering sciences from UL. She was a lecturer and a research member (ATER) at the UTT from September 2012 to August 2013. Since 2013, she has been an associate professor at the Research Centre for Automatic Control of Nancy (CRAN), University of Lorraine. Her research focuses on diagnosis of non-linear systems using data analysis methods, analysis of non-stationary signals, kernel methods, machine learning, pattern recognition, feature extraction, classification and prediction.

Received: 19 December 2021
Revised: 11 June 2022
Accepted: 4 August 2022