New contributions to nonlinear process monitoring through Kernel Partial Least Squares

José L. Godoy[a,1,*], David A. Zumoffen[b,2], Jorge R. Vega[a,3], Jacinto L. Marchetti[a]

[a] *Institute of Technological Development for the Chemical Industry (INTEC-CONICET-UNL), Güemes 3450, (3000) Santa Fe, Argentina. TE: +54-342-4559174, Fax: +54-342-4550944*
[b] *French-Argentine International Center for Information and Systems Sciences (CIFASIS-CONICET-UNR-AMU), Rosario, Santa Fe, Argentina.*

Abstract

The kernel partial least squares (KPLS) method was originally focused on soft-sensor calibration for predicting online quality attributes. In this work, an analysis of the KPLS-based modeling technique and its application to nonlinear process monitoring are presented. To this effect, the measurement decomposition, the development of new specific statistics acting on non-overlapped domains, and the contribution analysis are addressed for purposes of fault detection, diagnosis, and prediction risk assessment. Some practical insights for synthesizing the models are also given, which are related to an appropriate order selection and the adoption of the kernel function parameter. A proper combination of scaled statistics allows the definition of an efficient detection index for monitoring a nonlinear process. The effectiveness of the proposed methods is confirmed by using simulation examples.

Keywords: KPLS Modeling, Fault Detection, Fault Diagnosis, Prediction Risk Assessment, Nonlinear Processes.

1. Introduction

The design of monitoring systems for supervising the operation of industrial processes has acquired great relevance in the last decade. This fact is essentially due to the need of more demanding operating conditions related to security for equipments and personnel, operating costs, and environmental restrictions. Furthermore, the increasing complexity observed in the interactions between energy- and mass- transfer processes, and their corresponding control policies, require more sophisticated monitoring systems in aspects such as detection rate, robustness, user friendliness, easiness of understanding, modeling and data storage requirements, and adaptability, among others [1, 2].

*Corresponding author
Email address: jlgodoy@santafe-conicet.gov.ar (José L. Godoy)
[1] Also with Universidad Tecnológica Nacional - FRP, Paraná, Argentina
[2] Also with Universidad Tecnológica Nacional - FRRo, Rosario, Argentina
[3] Also with Universidad Tecnológica Nacional - FRSF, Santa Fe, Argentina

The multivariate statistical process monitoring is a well-known research topic where several strategies based on projection to latent structures have successfully been developed. Moreover, they are of great interest in industrial applications because of their excellent properties for handling noisy and highly correlated measurements, and large datasets [2–4]. Some of these approaches are summarized in [4–10] where the principal component analysis (PCA), independent component analysis (ICA), and partial least squares regression (PLSR) methodologies were addressed. There are also several modifications to these tools for including issues such as dynamics, adaptation, and non-linearity [2, 8, 11–16].

In this work, a nonlinear version of the partial least squares (PLS) approach -called kernel PLS (KPLS)- is addressed. KPLS is a powerful statistical tool for obtaining multivariate nonlinear relationships from historical data. In fact, it is a nonlinear regression method that computes the regression coefficients in a high-dimensional space; the input data are mapped via non-linear functions in this space and then they are linearly related to the measured outputs. Hence, the KPLS approach represents a suitable methodology for predicting online unmeasured quality variables in complex nonlinear processes. The overall procedure relies on classic linear algebra, similar to the linear projection methods, and the non-linearity degree is mainly given by the selected kernel function associated to the mapping functions [17]. Ever since the KPLS approach appeared, some modifications as well as applications have been published in the process monitoring area. For example, a kernel-based PLS system linked to orthogonal signal correction has been proposed for data preprocessing and prediction purposes [12]; and a modified PLS method of independent component regression has been used for complex processes monitoring [8]. An application of nonlinear multivariate quality prediction based on KPLS has also been presented [14]. In this context, new publications addressing the fault detection tasks based on KPLS have also appeared [18, 19]. In the last decade, KPLS or variants thereof have been applied for composition analysis of agricultural materials [20] and foods [21], process analysis [22], determination of structure activity relationships [23], studies on drug metabolism [24], and quality-related monitoring [25], among others.

KPLS method, as well as other kernel based modelling methods [26], are often used as a black box approach. However, in contrast to kernel PCA (KPCA) [11, 16], KPLS is able to properly determine the predictive importance of each input variable onto the final regression model. This result can then be used for reducing the number of inputs and therefore the complexity of the model. For instance, Posmat et al. [27] propose a method based on the principle of pseudo-sampled trajectories (representing the original variables) that help to visualizing and determining the most important variables for regression purposes. This method is able to detect poor predictor variables, providing the chance for improving the KPLS structure by eliminating interfering variables from the pre-selected inputs. The advantage of the KPLS modeling lays in the fact that only the outputs of interest are chosen, while the inputs are determined by their predictive importance, thus limiting the group of variables to be monitored.

The main objective of this article is to provide a deep analysis of the KPLS-based modeling technique

and its application to nonlinear process monitoring. Initially, the classic KPLS modeling is here extended by adding the projections of the outputs onto the latent space. The underlying structure of the KPLS modeling is highlighted in order to describe the functional relationships between the spaces induced by the KPLS procedure. Moreover, some practical insights are given for the proper selection of the number of latent variables and for setting the kernel function parameter. In fact, the latent space dimension is here defined by using a new balanced index designed to efficiently quantify the squared prediction error in both the input and output spaces. This approach is compared with the standard output prediction error via the Wold's R criterion [7, 14]. To deal with nonlinear processes, the kernel method is first embedded into the PLS algorithm. Then, new specific statistics (that act on non-overlapped domains) are combined into a single index able to detect process anomalies. Finally, the statistics pattern is used for diagnosing faults or process anomalies. In this regard, the present monitoring technique of nonlinear processes is an extension of our PLS-based strategy originally developed for monitoring linear processes [28]. Besides, contribution plots are frequently used to isolate the detected faulty variables without using historical fault patterns [26, 29]. However, it is difficult to build a contribution plot for a kernel based model [29]. In this paper, a new contribution plot based on the KPLS model is proposed for identifying faulty variables. The proposed supervision approach puts together the abnormal event detection, the diagnosis, and the isolation in a single method. Besides, a risk assessment index is also developed for online quantifying the predictive capabilities of the KPLS inferential model. The effectiveness of the proposed method is tested through simulated examples taken from the literature.

The article is organized as follows: Section 2 presents the basic background of the KPLS regression. Some details about the KPLS-based modeling approach are given in Section 3. The main contributions of this work are presented in Sections 4 and 5, where we analyze the KPLS model calibration (Section 4), the process monitoring and the statistics for fault detection (Section 5.1), the diagnosis method through the pattern of statistics (Section 5.2), the fault isolation via a contribution analysis (Section 5.3), and the prediction risk assessment (Section 5.4). Section 6 summarizes the simulation results and the overall conclusions are given in Section 7.

2. Basic concepts on KPLS

Consider a process with m measured input variables plus p measured output variables which are arranged in the vectors $\mathbf{x} = [x_1 \dots x_m]'$ and $\mathbf{y} = [y_1 \dots y_p]'$, respectively. Assume that N measurements of each variable are collected while the process is operating under normal conditions. In order to build a KPLS regression model, let us consider the calibration data set consisting of N centered and scaled samples for the input vector (predictor), i.e., $\{\mathbf{x}_j \in \mathbb{R}^m\}_{j=1}^N$, and the corresponding centered and scaled samples for the response vector, $\{\mathbf{y}_j \in \mathbb{R}^p\}_{j=1}^N$.

The KPLS theory utilizes a nonlinear function $\boldsymbol{\varphi}(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^c$ to map the vectors $\mathbf{x}_j \in \mathbb{R}^m$ onto a high-dimensional space $\mathbb{R}^c$, with $c \gg m$ (or even $c = \infty$). The classic PLS regression approach [6] is first applied to decompose the high-dimensional space into a low-dimension latent space plus a residual space; and then a linear regression model is developed to relate the data in the latent space with the original output data in $\mathbb{R}^p$ [17]. The nonlinear mapping is not implemented through an explicit function, $\boldsymbol{\varphi}(\cdot)$, instead a kernel function $k(\cdot, \cdot)$ is proposed for computing the following inner products,

$$k(\mathbf{x}_j, \mathbf{x}_r) = \boldsymbol{\varphi}(\mathbf{x}_j)'\boldsymbol{\varphi}(\mathbf{x}_r), \quad \text{with} \quad j = 1, \ldots, N \quad r = 1, \ldots, N. \tag{1}$$

Thus, by replacing each inner product $\boldsymbol{\varphi}(\mathbf{x}_j)'\boldsymbol{\varphi}(\mathbf{x}_r)$ with $k(\mathbf{x}_j, \mathbf{x}_r)$, both the explicit nonlinear mapping and the inner product computation can be avoided [17]. The KPLS approach only uses the inner product values for performing the nonlinear regression.

From Eq. (1) the so-called Gram Kernel matrix, $\mathbf{K} \in \mathbb{R}^{N \times N}$, can be obtained:

$$\mathbf{K} = \boldsymbol{\Phi}\,\boldsymbol{\Phi}', \quad \text{with} \quad \boldsymbol{\Phi} = [\boldsymbol{\varphi}(\mathbf{x}_1), \ldots, \boldsymbol{\varphi}(\mathbf{x}_N)]' \in \mathbb{R}^{N \times c}. \tag{2}$$

Similarly to PLSR, the nonlinear KPLS model includes zero-mean variables. The mapped input vectors $\boldsymbol{\varphi}(\mathbf{x}_j)$ are centered as follows:

$$\bar{\boldsymbol{\varphi}}(\mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_j) - \boldsymbol{\Phi}'\mathbf{e} \tag{3}$$

where $\mathbf{e}$ is a column vector with all its entries equal to $1/N$ [17]. In this way, $\bar{\boldsymbol{\Phi}} = [\bar{\boldsymbol{\varphi}}(\mathbf{x}_1), \ldots, \bar{\boldsymbol{\varphi}}(\mathbf{x}_N)]'$ is the centered version of $\boldsymbol{\Phi}$. Now the centered Gram Kernel matrix is given by

$$\bar{\mathbf{K}} = \bar{\boldsymbol{\Phi}}\,\bar{\boldsymbol{\Phi}}' = (\mathbf{I} - \mathbf{E})\mathbf{K}(\mathbf{I} - \mathbf{E}) \tag{4}$$

where $\mathbf{E}$ is a $(N \times N)$ matrix with all its entries equal to $1/N$ [17] and $\bar{k}(\mathbf{x}_j, \mathbf{x}_r) = \bar{\boldsymbol{\varphi}}(\mathbf{x}_j)'\bar{\boldsymbol{\varphi}}(\mathbf{x}_r)$ is the element (j, r) of $\bar{\mathbf{K}}$.

From the centered data matrices $\bar{\mathbf{K}}$ and $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]'$, a KPLS calibration algorithm can be developed by modifying the steps of the NIPALS algorithm [17] as shown in Algorithm 1. Specific details about the parameter setting for the kernel function and the optimal selection of the number of latent variables, A, are given in Section 4.

The prediction of the response variables by using the calibration data are given by [17]:

$$\hat{\mathbf{Y}} = \bar{\boldsymbol{\Phi}}\mathbf{B}_{\text{PLS}} = \bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\Phi}}'\mathbf{U}(\mathbf{T}'\bar{\mathbf{K}}\mathbf{U})^{-1}\mathbf{T}'\mathbf{Y} = \bar{\mathbf{K}}\mathbf{U}(\mathbf{T}'\bar{\mathbf{K}}\mathbf{U})^{-1}\mathbf{T}'\mathbf{Y} = \mathbf{T}\mathbf{T}'\mathbf{Y} = \mathbf{T}\mathbf{C}' \tag{5}$$

where the matrices $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_A]$ and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_A]$ are orthonormal by columns. Note that, although the regression coefficients matrix $\mathbf{B}_{\text{PLS}}$ might exist (for $\bar{\boldsymbol{\varphi}}(.) \in \mathbb{R}^c$ when $c \neq \infty$), the KPLS algorithm does not calculate this values explicitly, i.e. the kernel substitution avoids this evaluation.

Equation (5) shows that the response variables (outputs) can be obtained from the inner products of the mapped vectors. Hence, for a new observation $\mathbf{x}$ of the predictor vector, the outputs are estimated by

$$\hat{\mathbf{y}} = \mathbf{B}'_{\mathrm{PLS}}\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{Y}'\mathbf{T}\left[\mathbf{U}(\mathbf{T}'\bar{\mathbf{K}}\mathbf{U})^{-1}\right]'\bar{\mathbf{k}}(\mathbf{x}) = \mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}) \tag{6}$$

where $\bar{\mathbf{k}}(\mathbf{x}) = [\bar{k}(\mathbf{x}_1, \mathbf{x}), \ldots, \bar{k}(\mathbf{x}_N, \mathbf{x})]'$ is the vector of centered kernel functions evaluated in the pairs $(\mathbf{x}_j, \mathbf{x})$ for $j = 1, \ldots, N$. Note that matrices $\mathbf{U}$, $\mathbf{T}$, $\mathbf{C}$, and $\mathbf{V}$ are outputs of the KPLS algorithm.

From Eqs. (4) and (3) the following relationships can be defined,

$$\bar{\mathbf{k}}(\mathbf{x}) = \bar{\boldsymbol{\Phi}}\bar{\boldsymbol{\varphi}}(\mathbf{x}) = (\mathbf{I} - \mathbf{E})(\mathbf{k}(\mathbf{x}) - \mathbf{K}\mathbf{e}) = \mathbf{k}(\mathbf{x}) - \mathbf{K}\mathbf{e} - \mathbf{E}\mathbf{k}(\mathbf{x}) + \mathbf{E}\mathbf{K}\mathbf{e} \tag{7}$$

where $\mathbf{k}(\mathbf{x})$ is the vector of non-centered kernel functions and is defined analogously to $\bar{\mathbf{k}}(\mathbf{x})$.

Let us consider the latent vector structure. From Eqs. (5) and (6), it is obtained,

$$\mathbf{t}' = [t_1, \ldots, t_A]' = \bar{\boldsymbol{\varphi}}(\mathbf{x})'\mathbf{R} = \bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}, \quad \text{with} \quad \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_A] \tag{8}$$

where $\mathbf{R} = \bar{\boldsymbol{\Phi}}'\mathbf{U}(\mathbf{T}'\bar{\mathbf{K}}\mathbf{U})^{-1} = [\mathbf{r}_1, \ldots, \mathbf{r}_A]$ is the PLS weight matrix (see Eq. 5) with its components given by $\mathbf{r}_a = \sum_{j=1}^{N} \alpha_j \bar{\boldsymbol{\varphi}}(\mathbf{x}_j)$, with $\alpha_j \in \mathbb{R}$. Each latent variable can be estimated independently (i.e. $t_a = \bar{\mathbf{k}}'(\mathbf{x})\mathbf{v}_a$ in Eq. (8)), thus the prediction is computed as,

$$\hat{\mathbf{y}} = \mathbf{C}\mathbf{t}, \quad \text{with} \quad \mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_A]. \tag{9}$$

Note that, given a new observation $\mathbf{x}$ with absolute units the prediction also can be written as,

$$\hat{\mathbf{y}} = \mathbf{D}_y\mathbf{C}\mathbf{V}'\bar{\mathbf{k}}\left(\mathbf{D}_x^{-1}(\mathbf{x} - \bar{\mathbf{x}})\right) + \bar{\mathbf{y}} \tag{10}$$

where $\mathbf{D}_x = \mathrm{diag}(\hat{\sigma}_{x_1}, \ldots, \hat{\sigma}_{x_m})$ and $\mathbf{D}_y = \mathrm{diag}(\hat{\sigma}_{y_1}, \ldots, \hat{\sigma}_{y_p})$ are the estimated standard deviations for predictor and response variables, respectively. On the other hand, $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are the corresponding mean values.

3. Modeling based on KPLS

This section presents a KPLS-based extension of the PLSR modeling described in Godoy et al. [28], where three different residues are defined. The first residue $\tilde{\mathbf{Y}}_1$ represents the internal model error and the other two ($\tilde{\bar{\boldsymbol{\Phi}}}$ and $\tilde{\mathbf{Y}}_2$) are associated to the external model error as detailed below. The KPLS algorithm induces both an internal and an external model. By analogy of KPLS with PLSR [28], it is assumed here that there is an internal linear relationship between $\mathbf{t}_a$ and $\mathbf{u}_a$. Furthermore, since the KPLS algorithm scales these score-vectors to unit norm, the following internal model can be obtained,

$$\mathbf{U} = \mathbf{T} + \tilde{\mathbf{U}} \tag{11}$$

Algorithm 1: KPLS training via NIPALS steps

Data: Centered matrices $\bar{\mathbf{K}}$, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]'$

Result: $\mathbf{T}$, $\mathbf{U}$, $\mathbf{C}$, $\mathbf{V} = \mathbf{U}(\mathbf{T}'\bar{\mathbf{K}}\mathbf{U})^{-1}$

1 Set $a = 1$, $\bar{\mathbf{K}}_1 = \bar{\mathbf{K}}$, $\mathbf{Y}_1 = \mathbf{Y}$

2 Initialize the score-vector $\mathbf{u}_a$ $(N \times 1)$ of the latent variable u_a of $\mathbf{Y}_a$, as the maximum-variance column of $\mathbf{Y}_a$;

3 Compute the score-vector $\mathbf{t}_a$ $(N \times 1)$ of the latent variable t_a of $\bar{\mathbf{\Phi}}_a$, as: $\mathbf{t}_a = \bar{\mathbf{K}}_a\mathbf{u}_a / \left|\left|\bar{\mathbf{K}}_a\mathbf{u}_a\right|\right|$, $||\mathbf{t}_a|| = 1$;

4 Regress the columns of $\mathbf{Y}_a$ on $\mathbf{t}_a$: $\mathbf{c}_a = \mathbf{Y}_a'\mathbf{t}_a$, where $\mathbf{c}_a$ is a weighting vector;

5 Calculate the new score-vector: $\mathbf{u}_a = \mathbf{Y}_a\mathbf{c}_a / ||\mathbf{Y}_a\mathbf{c}_a||$, $||\mathbf{u}_a|| = 1$;

6 Repeat the steps 2 to 4 until the convergence of $\mathbf{t}_a$;

7 Deflate the matrices: $\bar{\mathbf{K}}_{a+1} = (\mathbf{I} - \mathbf{t}_a\mathbf{t}_a')\bar{\mathbf{K}}_a(\mathbf{I} - \mathbf{t}_a\mathbf{t}_a')$ and $\mathbf{Y}_{a+1} = \mathbf{Y}_a - \mathbf{t}_a\mathbf{t}_a'\mathbf{Y}_a$;

8 Save data in matrices: $\mathbf{T} \leftarrow \mathbf{t}_a$, $\mathbf{U} \leftarrow \mathbf{u}_a$, $\mathbf{C} \leftarrow \mathbf{c}_a$;

9 Set $a = a + 1$ and return to step 2. Stop when $a > A$, being A the selected number of latent variables.

where the a-th column of $\tilde{\mathbf{U}}$ represents the $\mathbf{t}_a - \mathbf{u}_a$ regression residuals.

The induced external model decomposes $\bar{\mathbf{\Phi}}$ and $\mathbf{Y}$ into latent variables and residuals matrices ($\tilde{\bar{\mathbf{\Phi}}}$ and $\tilde{\mathbf{Y}}_2$), via the following expressions:

$$\bar{\mathbf{\Phi}} = \mathbf{T}\mathbf{P}' + \tilde{\bar{\mathbf{\Phi}}}, \tag{12}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}' + \tilde{\mathbf{Y}}_2, \tag{13}$$

where $\mathbf{P} = \bar{\mathbf{\Phi}}'\mathbf{T} = \bar{\mathbf{\Phi}}'\bar{\mathbf{K}}\mathbf{V}$ and $\mathbf{C} = \mathbf{Y}'\bar{\mathbf{K}}\mathbf{V}$. For new $\bar{\mathbf{\Phi}}$ and $\mathbf{Y}$ matrices, by means of $\mathbf{R} = (\mathbf{P}')^-$ and $\mathbf{D} = (\mathbf{C}')^-$ where $(\cdot)^-$ denote the pseudo inverse operator (i.e. $\mathbf{P}'\mathbf{R} = \mathbf{I}$ and $\mathbf{C}'\mathbf{D} = \mathbf{I}$), the predictions of $\mathbf{T}$ and $\mathbf{U}$ can be represented as follows,

$$\mathbf{T} = \bar{\mathbf{\Phi}}\mathbf{R}, \tag{14}$$

$$\mathbf{U} = \mathbf{Y}\mathbf{D}, \tag{15}$$

since the row space of $\tilde{\bar{\mathbf{\Phi}}}$ ($\tilde{\mathbf{Y}}_2$) belongs to the null space of $\mathbf{R}$ ($\mathbf{D}$), then $\tilde{\bar{\mathbf{\Phi}}}\mathbf{R} = 0$ ($\tilde{\mathbf{Y}}_2\mathbf{D} = 0$). By means of kernel substitution in Eq. (14), the prediction of $\mathbf{T}$ also is given by $\mathbf{T} = \bar{\mathbf{K}}\mathbf{V}$. Note that, the external model in Eqs. (12) and (13) relates latent variables with responses and mapped inputs. On the other hand, the internal model in Eq. (11) links latent variables only. By combining both models, a prediction model based on kernel is obtained as shown in Eq. (16).

$$\mathbf{Y} = \bar{\mathbf{K}}\mathbf{V}\mathbf{C}' + \tilde{\mathbf{U}}\mathbf{C}' + \tilde{\mathbf{Y}}_2 = \hat{\mathbf{Y}} + \tilde{\mathbf{Y}}_1 + \tilde{\mathbf{Y}}_2, \tag{16}$$

where $\tilde{\mathbf{Y}}_2 = \mathbf{Y} - \mathbf{Y}\mathbf{D}\mathbf{C}'$ and $\tilde{\mathbf{Y}}_1 = \mathbf{Y}\mathbf{D}\mathbf{C}' - \hat{\mathbf{Y}}$ are the projection and transformation error matrices, respectively. It is particularly noteworthy that the KPLS algorithm does not compute the matrices $\mathbf{R}$ and $\mathbf{P}$, which would be computationally troublesome for typically high-values of c. Summarizing, the internal

model is represented by Eq. (11), the external relationships are displayed in Eqs. (12) and (13), and finally the nonlinear regression model is shown in Eq. (16).

3.1. Underlying decompositions

After synthesizing an in-control KPLS model, the measured vectors $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ and $\mathbf{y}$ can implicitly be decomposed in their projections (following the PLSR decomposition presented in [28]), as described bellow. Note that given a new mapped vector $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ the following theoretical decompositions would be valid:

$$\begin{aligned} \bar{\boldsymbol{\varphi}}(\mathbf{x}) &= \check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) + \tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}), && \in \mathbb{R}^c \\ \check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) &= \mathbf{P}\mathbf{R}'\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{P}\mathbf{t}, && \in W_M \equiv \text{Span}\{\mathbf{P}\} \subseteq \mathbb{R}^c \\ \tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) &= (\mathbf{I} - \mathbf{P}\mathbf{R}')\bar{\boldsymbol{\varphi}}(\mathbf{x}), && \in W_R \equiv \text{Span}\{\mathbf{R}\}^{\perp} \subseteq \mathbb{R}^c \end{aligned} \tag{17}$$

where $\mathbf{P}\mathbf{R}'$ $(\mathbf{I} - \mathbf{P}\mathbf{R}')$ is the projector on the model subspace W_M (W_R) along the residual subspace W_R (W_M). Here, $(\cdot)^{\perp}$ represents the orthogonal complement. The oblique projections in Eq. (17) decompose the high-dimensional space $\mathbb{R}^c$ in two complementary subspaces W_M and W_R [28].

On the other hand, generalizing the results presented in [28], the response space can be decomposed (via KPLS) in two complementary oblique subspaces as shown in Eq. (18),

$$\begin{aligned} \mathbf{y} &= \check{\mathbf{y}} + \tilde{\mathbf{y}}_2, && \in \mathbb{R}^p \\ \check{\mathbf{y}} &= \mathbf{C}\mathbf{D}'\mathbf{y}, && \in S_{MY} \equiv \text{Span}\{\mathbf{C}\} \subseteq \mathbb{R}^p \\ \tilde{\mathbf{y}}_2 &= (\mathbf{I} - \mathbf{C}\mathbf{D}')\mathbf{y}, && \in S_{RY} \equiv \text{Span}\{\mathbf{D}\}^{\perp} \subseteq \mathbb{R}^p \end{aligned} \tag{18}$$

where S_{MY} and S_{RY} denote the model subspace and residual subspace of $\mathbb{R}^p$, respectively.

The model subspaces, W_M and S_{MY}, are related via Eq. (19) as follows,

$$\begin{aligned} \check{\mathbf{y}} &= \hat{\mathbf{y}} + \tilde{\mathbf{y}}_1, && \in S_{MY} \\ \hat{\mathbf{y}} &= \mathbf{C}\mathbf{R}'\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}), && \in S_{MY} \\ \tilde{\mathbf{y}}_1 &= \mathbf{C}\mathbf{D}'\mathbf{y} - \mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}), && \in S_{MY} \end{aligned} \tag{19}$$

where $\tilde{\mathbf{y}}_1$ is the linear transformation error $\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) \rightarrow \check{\mathbf{y}}$ and $\hat{\mathbf{y}}$ represents the predictable part of $\check{\mathbf{y}}$ from $\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x})$.

Figure 1 shows the underlying decomposition in the KPLS procedure. The gray areas represent the so-called "in-control" or "normal operating" zones. The functional relationships among several spaces are shown in this figure, i.e., the links among input, high-dimensional, model, residual, and output spaces.

4. Calibration of the KPLS model

There are two main issues in any kernel-based latent structure approach: 1) the selection of the kernel function and its parameters, and 2) the determination of the latent space dimension (number of latent variables). Both decisions play a significant influence in the KPLS model performance for prediction as well as monitoring purposes.

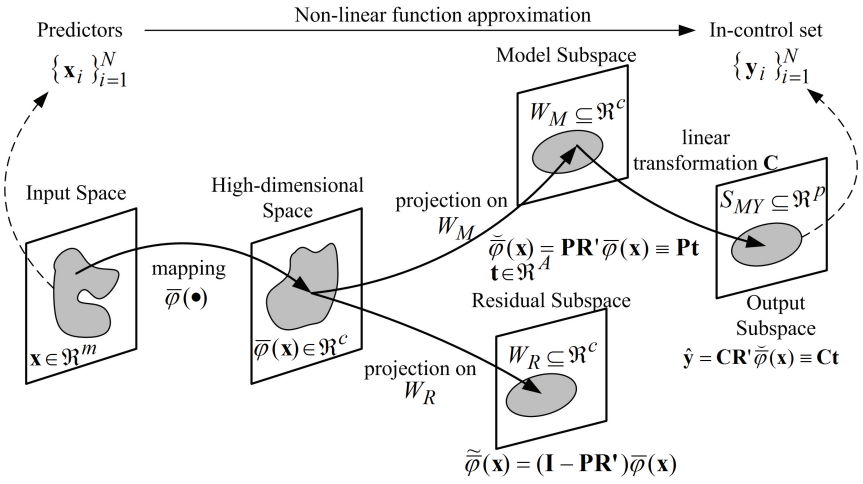


Figure 1: Underlying KPLS-decomposition of the input space and its relation to output subspace together with control zones.

A specific kernel function implicitly defines the mapping φ and features the high-dimensional space. Moreover, the monitoring characteristics are strongly influenced by the parameter settings of the selected kernel function, as shown in [15] for KPCA. Similarly, in KPLS these parameters determine the ability for detecting and identifying abnormal situations from the measurements [30]. How to select the proper kernel function for a specific application is still an open problem, however the most used kernel function is the Gaussian $k(\mathbf{x}_j, \mathbf{x}) = \exp(-||\mathbf{x}_j - \mathbf{x}||^2/h)$ [26]. This suitable choice is driven by the observation that most functions can be approximated fairly well by a sum of Gaussians. Indeed, mixture models and Radial Basis Function (RBF) neural nets [31] are based on this observation, and they are testimony of the power of Gaussians in fitting data too. However, a poorly chosen h will lead to a poor KPLS model [26]. A critical drawback of the KPLS model is the difficulty for selecting the parameter h.

The occurrence of an abnormal event in the process will alter the statistical behavior of the measurement vector: $\mathbf{x} = [x_1 \dots x_i \dots x_m]'$. Such event will significantly be propagated to the high-dimensional space (see Fig. 1), provided that the expected value of $\mathbf{k}(\mathbf{x})$ (see Eq. 7) exhibits a meaningful change. Additionally, the expected value of the j-th element of $\mathbf{k}(\mathbf{x})$, $k(\mathbf{x}_j, \mathbf{x})$, will be sensitive to the presence of an abnormal event when a relevant change is observed in the expected value of its argument, $||\mathbf{x}_j - \mathbf{x}||^2/h$. Therefore, an appropriate selection of h becomes important because: i) a high value of h could turn the argument too low, with the risk that an actual event is not detected when measuring the projections norms in Eqs. (17) and (19); and ii) a low value of h could excessively increase the argument, with the risk of producing a false alarm during a normal process operation. To circumvent this problem, we here propose to adopt $h = 2\sum_{i=1}^{m} Var(x_i)$. Since x_i is a standardized variable, $Var(x_i) = 1$ and consequently $h = 2m$. In what follows, we will prove that such selection of h provides us with a robust decision criterion for detecting the

presence of a process fault. In fact, note that:

$$\mathrm{E}\left\{\frac{||\mathbf{x}_j-\mathbf{x}||^2}{h}\right\}=\frac{\sum_{i=1}^{m}\mathrm{E}\left\{(x_{j,i}-x_i)^2\right\}}{2m}=\frac{\sum_{i=1}^{m}Var(x_i)+\sum_{i=1}^{m}Bias_j^2(x_i)}{2m}$$
$$=\begin{cases}=1/2, \text{ if } \mathbf{x}_j \text{ and } \mathbf{x} \text{ belong to the same in-control process}\\ >1/2, \text{ if at least one } x_i \text{ is a (fixed) disturbed variable}\end{cases} \tag{20}$$

where $\mathbf{x}_j$ is the j-th sample of the calibration data set, $x_{j,i}$ is the i-th element of $\mathbf{x}_j$, and x_i is considered an estimate of $x_{j,i}$. To interpret the use of Eq. (20) in some practical cases, assume first a normal process operation (i.e., with null bias and unit variance in all variables x_i). Then, Eq. (20) yields $\mathrm{E}\{.\}=1/2$, thus indicating no evidence of process fault. On the other hand, consider the presence of an offset in the i-th sensor. Such fault will induce on the variable x_i a meaningful bias Δx_i with respect to its normal value. Then, Eq. (20) predicts: $\mathrm{E}\{.\}=1/2(1+\Delta x_i^2/m)>1/2$, thus alerting on the presence of a fault in the process.

On the other hand, the number of latent variables retained (A) is an important parameter in any approach based on latent structures. In the KPLS methodology this parameter can be determined by considering monitoring as well as prediction purposes. In fact, the KPLS approach models both, the "$\mathbf{x}-\mathbf{y}$" relationship and the correlations within $\mathbf{x}$ and $\mathbf{y}$. Hence, the ultimate number of latent variables to be retained should be determined by the simultaneous adjustment of the prediction- and correlation models. Therefore, the modeling error is evaluated through the following total mean square error:

$$MSE_T \triangleq \mathrm{E}\left\{\left|\left|\tilde{\boldsymbol{\varphi}}(\mathbf{x})\right|\right|^2\right\}+\mathrm{E}\left\{||\mathbf{y}-\hat{\mathbf{y}}||^2\right\}=MSE_x+MSE_y. \tag{21}$$

In the present work, we use a generalization of the adjusted Wold's R criterion [7], which is given by $R(a+1)=MSE_T(a+1)/MSE_T(a)$, where $MSE_T(a)$ is the criterion in Eq. (21) parameterized by the first "a" latent variables. The inclusion of new latent variables into the model finishes when the ratio $R(a+1)$ exceeds a predefined threshold (e.g. 0.9) and hence $A=a$. In other words, an additional latent variable will not be included in the KPLS model unless this variable significantly improves the predictions (or explicated variability). Generally, the inclusion of an excessive number of latent variables produces an over-fitted (or overtrained) nonlinear model with poor predictive ability. Hence, to reliably determine the number A, the historical data set is divided in two subsets called calibration and validation data, respectively. Thus, the $MSE_T(a)$ criterion is tested on both subsets.

5. Process monitoring based on KPLS

Once an "in-control" KPLS model is developed using process data under normal operating conditions, the process state can be supervised by using the proper statistics on the current measurements. Similarly to the KPCA approach [16], the main idea behind the KPLS methodology is mapping and projecting the

input and the response measurements respectively, into the $\mathbb{R}^A$ latent space to get a linear distribution of the modeled data. These transformations are useful to perform the detection of abnormal events and the diagnosis by inspecting appropriate statistics. In this work, a combined index is proposed for monitoring both, the model and the residual subspaces. The diagnostic is performed by inspecting the pattern defined by the statistics once the anomaly is detected. Furthermore, a variable contribution analysis completes the diagnostic tasks allowing the isolation of disturbed variables.

5.1. KPLS-based fault detection

The "in-control" KPLS model is used for analyzing the current state of the process. By mapping and projecting the current measurements, $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ and $\mathbf{y}$, on the subspaces W_M, W_R, S_{MY}, and S_{RY}, the corresponding deviations are quantified and compared with their appropriate control limits. However, there are no explicit expressions for the projections $\check{\boldsymbol{\varphi}}$ and $\tilde{\check{\boldsymbol{\varphi}}}$ in Eq. (17). This trouble promotes the development of new statistics based on kernel substitution for estimating the measures needed for the monitoring task.

For detecting a significative change in the W_M subspace, the following Hotelling's statistic can be used:

$$T^2_{\mathbf{t}} = \mathbf{t}'\Lambda^{-1}\mathbf{t} = (N-1)\bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}) \tag{22}$$

where $\Lambda = (N-1)^{-1}\mathbf{T}'\mathbf{T} = (N-1)^{-1}\mathbf{I}$. Recall that this measure (Eq. 22) accounts for the process correlations.

When new events occur (not considered for the in-control model), the new mapped observation $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ will move out from W_M towards W_R. In this case, the square prediction error (SPE) is used for quantifying the distance from the model in W_M,

$$\begin{aligned} SPE_{\tilde{\check{\boldsymbol{\varphi}}}} &= \left|\left|\tilde{\check{\boldsymbol{\varphi}}}(\mathbf{x})\right|\right|^2 = ||\bar{\boldsymbol{\varphi}}(\mathbf{x}) - \check{\boldsymbol{\varphi}}(\mathbf{x})||^2 \\ &= \bar{\boldsymbol{\varphi}}(\mathbf{x})'\bar{\boldsymbol{\varphi}}(\mathbf{x}) - 2\bar{\boldsymbol{\varphi}}(\mathbf{x})'\check{\boldsymbol{\varphi}}(\mathbf{x}) + \check{\boldsymbol{\varphi}}(\mathbf{x})'\check{\boldsymbol{\varphi}}(\mathbf{x}) \\ &= \bar{k}(\mathbf{x},\mathbf{x}) - 2\bar{\mathbf{k}}'(\mathbf{x})\bar{\mathbf{K}}\mathbf{V}\mathbf{t} + \mathbf{t}'\mathbf{T}'\bar{\mathbf{K}}\mathbf{T}\mathbf{t} \end{aligned} \tag{23}$$

Thus, the $SPE_{\tilde{\check{\boldsymbol{\varphi}}}}$ statistic can be used for detecting changes in W_R. When the process is under normal operation, the $SPE_{\tilde{\check{\boldsymbol{\varphi}}}}$ index represents the fluctuations that can not be explained by the KPLS model. On the other hand, the distance from the regression model in S_{MY} is defined as

$$\begin{aligned} SPE_{\mathbf{y}_1} =& ||\tilde{\mathbf{y}}_1||^2 = \left|\left|\mathbf{C}\mathbf{D}'\mathbf{y} - \mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x})\right|\right|^2 \\ =& \mathbf{y}'\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{D}'\mathbf{y} - 2\mathbf{y}'\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}) + \bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}\mathbf{C}'\mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x}) \end{aligned} \tag{24}$$

Similarly, the distance from the model in S_{MY} for detecting changes in S_{RY} is

$$SPE_{\mathbf{y}_2} = ||\tilde{\mathbf{y}}_2||^2 = ||(\mathbf{I} - \mathbf{C}\mathbf{D}')\,\mathbf{y}||^2 . \tag{25}$$

Furthermore, the correlation matrices $\mathbf{R}_{\check{\boldsymbol{\varphi}}}$ and $\mathbf{R}_{\hat{\mathbf{y}}}$ are singular because $\check{\boldsymbol{\varphi}}$ and $\hat{\mathbf{y}}$ typically have collinear variables, as can be inferred from Eqs. (17) and (9). In this context, the generalized Mahalanobis distance

is considered for measuring these projections as follows:

$$D_{\breve{\boldsymbol{\varphi}}} = \breve{\boldsymbol{\varphi}}'(\mathbf{x})\mathbf{R}^{-}_{\breve{\boldsymbol{\varphi}}}\breve{\boldsymbol{\varphi}}(\mathbf{x}), \tag{26}$$

$$D_{\hat{\mathbf{y}}} = \hat{\mathbf{y}}'\mathbf{R}^{-}_{\hat{\mathbf{y}}}\hat{\mathbf{y}}, \tag{27}$$

where the correlation matrices are given by,

$$\mathbf{R}_{\hat{\mathbf{y}}} = (N-1)^{-1}\hat{\mathbf{Y}}'\hat{\mathbf{Y}} = (N-1)^{-1}\mathbf{C}\mathbf{T}'\mathbf{T}\mathbf{C}' = (N-1)^{-1}\mathbf{C}\mathbf{C}', \tag{28}$$

$$\mathbf{R}_{\breve{\boldsymbol{\varphi}}} = (N-1)^{-1}\breve{\boldsymbol{\Phi}}'\breve{\boldsymbol{\Phi}} = (N-1)^{-1}\mathbf{P}\mathbf{T}'\mathbf{T}\mathbf{P}' = (N-1)^{-1}\mathbf{P}\mathbf{P}'. \tag{29}$$

The following statement shows that not all these statistics are independent. More specifically, the statistics on $\breve{\boldsymbol{\varphi}}(\mathbf{x})$, $\hat{\mathbf{y}}$, and $\mathbf{t}$ are equivalents (see Proof in Appendix A), i.e.,

$$D_{\hat{\mathbf{y}}} = T^2_{\mathbf{t}} = D_{\breve{\boldsymbol{\varphi}}}. \tag{30}$$

The above identity suggests that the behavior of the response variables $\mathbf{y}$ can be monitored by a KPLS-based statistic applied to the input variables $\mathbf{x}$. Therefore, the monitoring of the complete measurement space can be implemented by four independent KPLS statistics: $T^2_{\mathbf{t}}$, $SPE_{\breve{\boldsymbol{\varphi}}}$, $SPE_{\mathbf{y}_1}$, and $SPE_{\mathbf{y}_2}$, each of them acting on different subspaces W_M, W_R, S_{MY}, and S_{RY}, respectively. Consequently, these statistics normalized by their control limits are combined together into a unified index called I_{KPLS} as shown in Eq. (31). The scalars τ^2_α, $\delta^2_{\mathbf{x}}$, $\delta^2_{\mathbf{y}_1}$, and $\delta^2_{\mathbf{y}_2}$ are the corresponding control (confidence) limits.

$$I_{KPLS}(\mathbf{x},\mathbf{y}) = \frac{T^2_{\mathbf{t}}}{\tau^2_\alpha} + \frac{SPE_{\breve{\boldsymbol{\varphi}}}}{\delta^2_{\mathbf{x}}} + \frac{SPE_{\mathbf{y}_1}}{\delta^2_{\mathbf{y}_1}} + \frac{SPE_{\mathbf{y}_2}}{\delta^2_{\mathbf{y}_2}} = I_{WM}(\mathbf{x}) + I_{WR}(\mathbf{x}) + I_{RY1}(\mathbf{x},\mathbf{y}) + I_{RY2}(\mathbf{y}). \tag{31}$$

The four statistics in Eq. (31) probably have non-Gaussian distributions due to the process nonlinearities. Hence, their control limits are estimated via a kernel density estimation (KDE) approach [32–34]. This methodology is based on the approximation of the probability density function (PDF) of an index, e.g. $T^2_{\mathbf{t}}$, by the sum of Gaussian basis functions. In this way, the $100(1-\alpha)\%$ confidence limit for $T^2_{\mathbf{t}}$ is given by $\tau^2_\alpha = G^{-1}(1-\alpha)$, where G^{-1} is the inverse of the cumulative density function $G(T^2_{\mathbf{t}})$. The KDE strategy is a well-known procedure for estimating the PDF when applied to univariate random processes [34].

In case that I_{KPLS} belong to a multidimensional elliptic region, it would be compatible with the assumption of multi-normal data. Therefore, the number of false alarms and undetected faults for this combined index is significantly reduced with respect to the performance typically given by the separated statistics [2, 4, 9].

5.2. Anomaly class diagnosis through its statistics pattern

The I_{KPLS} index (Eq. 31) is useful for simultaneous monitoring of product quality, process changes, and sensor problems. An anomaly is detected when a meaningful change in the measurements triggers the alarm

condition given by: $I_{KPLS} \geq I_\alpha$, where I_α denotes the $100(1-\alpha)\%$ confidence limit obtained through the KDE approach. Once an anomaly is detected, the diagnosis is made by comparing the patterns produced by the normalized statistics composing the combined index. Note that these statistics actuate on different domains and are affected by different scaling factors. In this way the significance level of each normalized statistic is 1. In summary, it is assumed here that each class of anomaly is characterized by a specific pattern of significant statistics.

Let us consider 6 different classes of anomalies: 1) faults in sensors associated to $\mathbf{x}$; 2) faults in sensors associated to $\mathbf{y}$; 3) changes in the nonlinear correlation structure of $\mathbf{x}$; 4) changes in the internal latent relationships; 5) changes in the correlation structure of $\mathbf{y}$; and 6) changes in the process population parameters. Ideally, each anomaly will produce an independent pattern of the statistics that compose I_{KPLS}; and therefore, a proper analysis of the measured statistics would allow the unambiguous identification of the anomaly. These anomalies can qualitatively be grouped into 3 categories: sensor fault (classes 1 and 2), change in the process correlations (classes 3, 4, and 5), and change in the process population parameters (class 6).

An artificial process was created to determine the characteristic patterns (see Appendix B). This process models the generation of ideal data obeying to a predetermined correlation/functional structure, in absence of anomalies. Then, each of the 6 anomalies were independently analyzed by assuming localized pure disturbances and then observing the mismatch with the available model. Figure 2 indicates how a disturbed measurement (in $\mathbf{x}$ or $\mathbf{y}$) goes through the KPLS model and generates a warning signal (in $\tilde{\boldsymbol{\varphi}}(\mathbf{x})$, $\tilde{\mathbf{y}}_1$, $\tilde{\mathbf{y}}_2$ or $\mathbf{t}$).

Table 1 summarizes the main results obtained in Appendix B, and can be seen as a generalization of the results in [28]. The symbols "$-$" and "$+$" respectively represent a "negligible" or a "high" value of the corresponding statistic. More specifically, the symbols "$+$" indicate the statistics that are activated as soon as the measurements $(\mathbf{x}, \mathbf{y})$ bring information about a localized model mismatch. The patterns in Table 1 can facilitate the fault diagnosis tasks. Furthermore, when some statistic is above its control limit a contribution analysis can be performed to identify the disturbed variables in $\mathbf{x}$ or $\mathbf{y}$ [29]. Unfortunately, the statistic patterns for classes 1 and 3 are coincident (see Table 1). However, a further contribution analysis can be used to determine the proper class, bearing in mind that, unlike Class 3, Class 1 leads a significant contribution in the faulty variable.

5.3. Isolation of disturbed variables

In order to localize a faulty sensor, the identification of the involved variables becomes helpful. A preliminary classification of the anomalous event according to Table 1 enables us to restrict the searching problem to one or two statistics. To isolate the abnormal events, the corresponding statistics can then be analyzed in their variable contributions, as typically proposed by several authors [29].

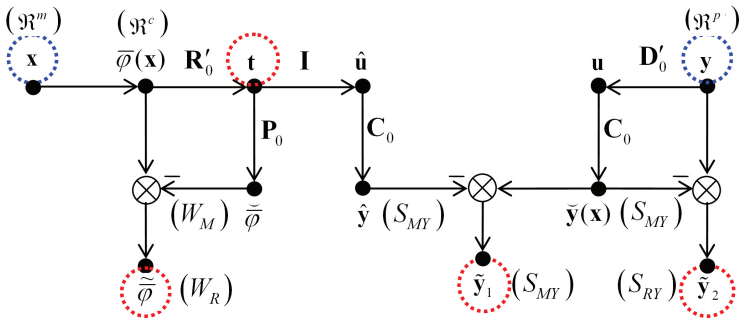


Figure 2: Measurements decomposition based on mapping and projections onto the subspaces created by a KPLS model

Table 1: Patterns of significant statistics to be used for diagnosis purposes

	Normalized statistic			
Event type	I_{WM}	I_{WR}	I_{RY1}	I_{RY2}
Class 1	–	+	+	–
Class 2	–	–	–	+
Class 3	–	+	+	–
Class 4	–	–	+	–
Class 5	–	–	+	+
Class 6	+	–	–	–
Associated subspace	W_M	W_R	S_{MY}	S_{RY}

A generic strategy for decomposing a quadratic index as a sum of variable contributions is given by [28]

$$I_S(\mathbf{x}) = \sum_{i=1}^{m} \left(\frac{x_i}{2} \frac{\partial I_S(\mathbf{x})}{\partial x_i} \right) = \sum_{i=1}^{m} cI_S(x_i), \tag{32}$$

where x_i represents the i-th variable of the current vector $\mathbf{x}$ and I_S can be I_{WM}, I_{WR}, I_{RY1} or I_{RY2}. Since each normalized statistic has a significance level of 1, then the significance level of their contributions is also adopted equal to 1 [28]. This decomposition is exact for I_{RY2} because it is a quadratic function of $\mathbf{y}$. On the other hand, I_{WM}, I_{WR} and I_{RY1} are quadratic expressions of the non-linear functions vector $\bar{\mathbf{k}}(\mathbf{x})$, but not of $\mathbf{x}$. However, around a given point $\mathbf{x}$, its second order Taylor approximation has a quadratic form on $\mathbf{x}$. Consequently, even in case that the decomposition in Eq. (32) is an approximation, we can define the contribution of the variable x_i to a normalized statistic as

$$cI_S(x_i) = \frac{x_i}{2} \frac{\partial I_S(\mathbf{x})}{\partial x_i}, \tag{33}$$

where $I_S = I_{WM}, I_{WR}, I_{RY1}, I_{RY2}$. The contributions of the variable y_i to the component statistics I_{RY1} and I_{RY2} are also defined by Eq. (33), but replacing y_i with x_i. The partial derivatives of I_{WM}, I_{WR}, I_{RY1}, and I_{RY2}, are detailed in Appendix C.

In summary, when an alarm in I_{KPLS} is detected, the significant statistics that compose I_{KPLS} (I_{WM}, I_{WR}, I_{RY1}, or I_{RY2}) are used for classifying the abnormality through Table 1. Then, depending on the classification, the variable contributions of any particular statistic with significant signal are analyzed for determining the fault source.

5.4. Risk assessment about the prediction accuracy

When the process outputs can not be measured online, we can still use Eq. (10) to predict quality variables $\mathbf{y}$ from the measurements $\mathbf{x}$. The prediction reliability depends on the accuracy of both, the inferential model (Eq. 10) and the measurements $\mathbf{x}$. Hence, it is convenient to validate online the prediction accuracy. An index based on an in-control KPLS model that depends only on $\mathbf{x}$, can be used for supervising the prediction reliability. In this case, the normalized statistics I_{WM} and I_{WR} depend on the $\mathbf{x}$ information only. Hence, recalling Eq. (31) the following combined inferential index is suggested,

$$I_c(\mathbf{x}) = I_{WM}(\mathbf{x}) + I_{WR}(\mathbf{x}), \tag{34}$$

and the risk assessment metric for the predictions can be stated as

$$I_{risk} = \frac{I_c(\mathbf{x})}{I_\alpha}, \tag{35}$$

where I_α is the confidence limit for I_c in Eq. (34). Thus, when $I_{risk} \geq 1$ the KPLS predictions are no longer reliable. Faults of class-1, class-3 and class-6 (see Table 1) trigger an alarm when $I_{risk} \geq 1$. The risk index depends on the KPLS model and therefore it is not reliable when the model is over-fitted or improperly calibrated.

In this work the KPLS-based online prediction is complemented with an additional control chart of I_{risk} to guarantee the reliability.

6. Simulation results

6.1. Case study no. 1. Soft-sensor with prediction risk assessment

A non-linear numerical simulation example is presented here for evaluating the proposed calibration and supervised prediction approach under several abnormal events. In fact, the multivariate simulation case used in Zhao et al. [13] and Zhang et al. [14] is reproduced here for the sake of comparison. The system is defined as follows,

$$\begin{cases} \mathbf{x} = \begin{cases} x_1 = t^2 - t + 1 + \varepsilon_1 \equiv f_1(t) + \varepsilon_1 \\ x_2 = \sin(t) + \varepsilon_2 \equiv f_2(t) + \varepsilon_2 \\ x_3 = t^3 + t + \varepsilon_3 \equiv f_3(t) + \varepsilon_3 \end{cases} \\ \\ y = x_1^2 + x_1 x_2 + 3\cos(x_3) + \varepsilon_4 \equiv f_4(t) + \varepsilon_4 \end{cases} \tag{36}$$

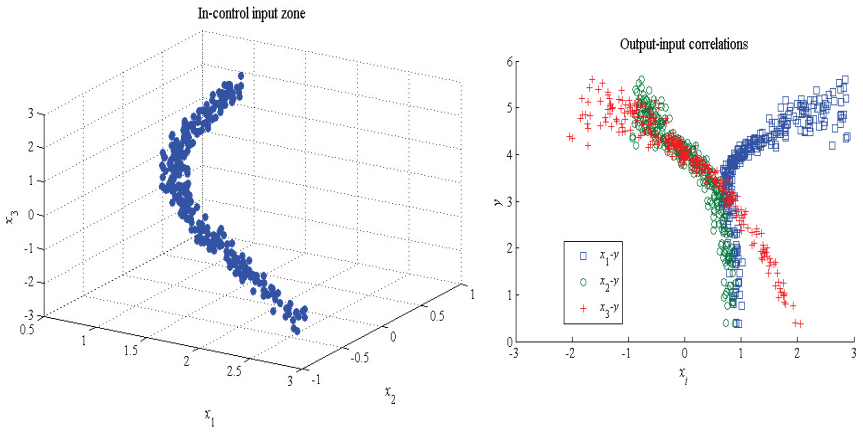
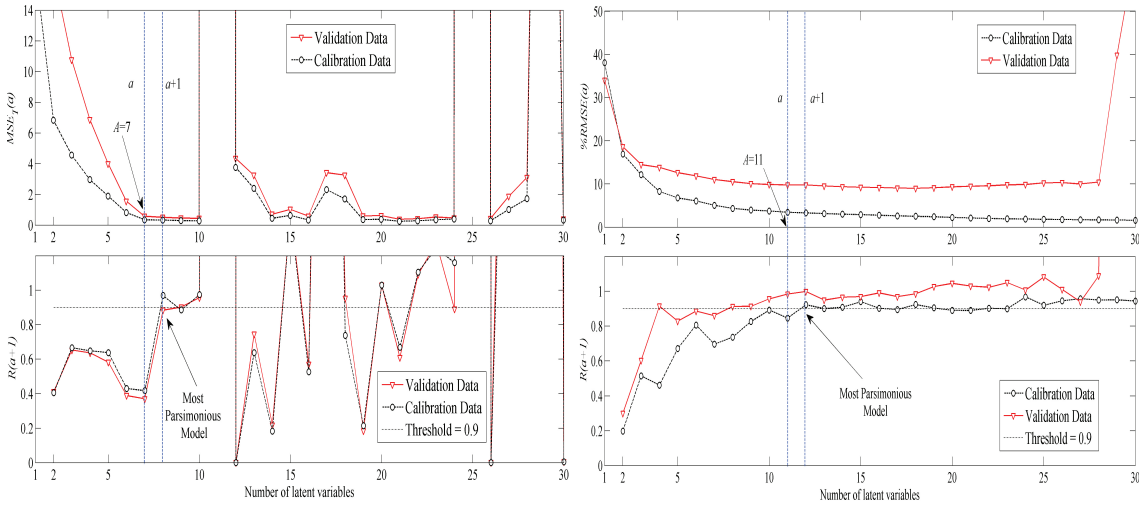


Figure 3: In-control input zone and output-input correlations.



(a) (b)

Figure 4: Selection of the KPLS model order - Case study no. 1. (a) MSE_T-based approach with $h = 2m = 6$. (b) MSE_y-based approach with $h = 0.06$.

where t and ε_i $(i = 1, 2, 3, 4)$ are uniformly distributed variables between $[-1, 1]$ and $[-0.1, 0.1]$, respectively. The variables ε_i are the noise components. The data set generated from 300 samples is divided in two subsets for calibration (the first 200 samples) and validation (the latest 100 samples) purposes. Equation (36) shows that inputs and output are functions of the internal variable t ($f_1(t)$ to $f_4(t)$). Figure 3 shows the in-control input zone, which depends on data and is related to the output range $0 - 6$. Figure 3 also shows that the output is non-linearly correlated with the inputs.

Figure 4 displays the procedure for calibrating the KPLS model based on the Wold's R criterion by using both, validation and calibration data. In this case, two scenarios for obtaining the optimal number of the

Table 2: Simulated abnormal events - Case study no. 1

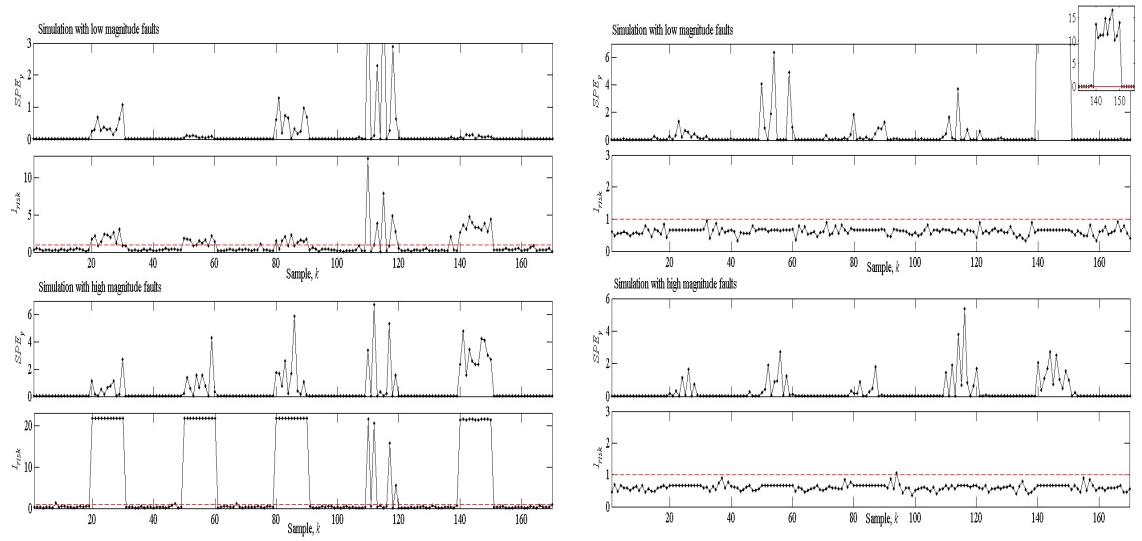
Fault type	Location	Low magnitude	High magnitude
Offset in x_1-sensor	$k = 20...30$	$\Delta x_1 = -0.34^{\ddagger}$	$\Delta x_1 = -10$
Offset in x_2-sensor	$k = 50...60$	$\Delta x_2 = 0.27^{\ddagger}$	$\Delta x_2 = 10$
Offset in x_3-sensor	$k = 80...90$	$\Delta x_3 = 0.49^{\ddagger}$	$\Delta x_3 = 10$
Correlation change	$k = 110...120$	$x_1 = 1.5t^3 - t + 1 + \varepsilon_1$	$x_1 = 3t^2 - t + 1 + \varepsilon_1$
Process upset	$k = 140...150$	$t = 1.1$ (fixed)	$t = 1.5$ (fixed)

‡ Equivalent to 0.5 in standardized units, hence $\mathrm{E}\{\cdot\} = 1/2(1 + 0.5^2/3) > 1/2$.

latent variables (model order) are compared. In fact, Figs. 4(a) and 4(b) summarize the mentioned criterion computed based on the MSE_T and MSE_y indexes, respectively. The first one, is the combined input and output prediction error suggested here in Eq. (21), and the second metric is the classical output prediction error MSE_y proposed by Zhang et al. [14], where the standard error is given by $\%RMSE_y = 100\sqrt{MSE_y}$. For the sake of comparison, the same settings as suggested in [14] are used here for the MSE_y-based approach: a Gaussian kernel function with $h = 0.06$ and a threshold of 0.9 for the Wold's criterion. The MSE_T-based methodology with $h = 2m = 6$ is shown in Fig. 4(a) which clearly gives the most parsimonious model with $a + 1 = 8$, hence the number of latent variables retained in the KPLS model is $A = 7$. On the other hand, the MSE_y-based approach in Fig. 4(b) suggests $A = 11$ as the best model order [14]. The over-fitting during the order selection is evaluated by using the validation data set (lines with "$\triangledown$" in Fig. 4).

The Fig. 4(b) (top) displays some evidence that the predictive ability of the model may be quite poor. In fact, the calibration and validation $\%RMSE_y$ deviate from each other, thus displaying the so-called over-fitting of the KPLS model. This fact is basically given by two simultaneous effects in the Zhang's procedure [14]: 1- they adopted $h = 0.06$, implicitly assuming a very low average variability in the data set, and 2- the MSE_y criterion does not consider the input modeling, nor the validation data for supervising the fitting reliability. Notice that the existence of over-fitting is observed by contrasting the Wold's criterion using the calibration data with the one based on validation data (see Fig. 4(b)). This result shows that the methodology proposed in section 4 improves the KPLS model calibration producing more reliable models as observed in Fig. 4(a).

Different faults affecting the normal process are simulated to test the ability of the risk index for warning a loss of prediction reliability. In this case, five types of abnormal events are considered as shown in Table 2 with low and high magnitudes in each case. In fact, this table summarizes three types of offset faults for the variables x_1, x_2, and x_3. The fourth abnormal event considers a nonlinear correlation change for the variable x_1 in the time period [110-120]. Finally, the uniformly distributed variable t is considered to be fixed at a given value between samples 140 and 150, representing a process upset. The in-control input zone



(a) (b)

Figure 5: Squared prediction error and prediction risk assessment during low and high magnitude faults for KPLS model calibration based on: (a) MSE_T, and (b) MSE_y

in Fig. 3 is related to an ellipsoidal zone in the latent space $\mathbb{R}^7$ which determines the control volume of $T^2_{\mathbf{t}}$.

The output squared prediction error $SPE_{\mathbf{y}} = ||\mathbf{y} - \hat{\mathbf{y}}||^2$ and the risk assessment index I_{risk} are displayed in Fig. 5 during the simulation of the anomalies listed in Table 2. Figures 5(a) and 5(b) summarize the risk assessment performance of the KPLS models calibrated via the MSE_T and the MSE_y methodologies, respectively. Note that, an alarm condition is triggered when the index I_{risk} exceeds the unitary threshold, i.e. $I_{risk} \geq 1$. Figure 5(a) show that the I_{risk} index timely alerts on intervals where the process is affected by abnormal events. In such intervals the predictions given by the model are unreliable and should not be considered. Hence the MSE_T criterion produces reliable KPLS models. On the other hand, the MSE_y criterion provides models that are unable to detect anomalies in the process (Fig. 5(b)). In fact, the I_{risk} index remains below the threshold along the simulation time, erroneously indicating a normal operation and reliable predictions. Figure 5(b) shows the insensitivity of I_{risk} to the abnormal events. In such cases $\mathbf{k}(\mathbf{x}) \cong 0$, causing a constant low value in I_{risk} due to the wrong selection of h and the model over-fitting. However, in our approach the index saturation occurs for the high magnitude faults, but above the threshold efficiently indicating the presence of the faults (see Fig. 5(a)).

6.2. Case study no. 2. Fault detection and diagnosis

An additional example is simulated for better understanding of the proposed methodology as a monitoring tool. The normal operation of the chosen non-linear process includes a uniform distribution in the intrinsic variables. The "measurements" of the external variables, $\mathbf{x}$ and $\mathbf{y}$, are generated by adding zero-mean

Gaussian random noises to the KPLS correlation structure characterized by the arbitrarily-selected process parameters and functions as follows:

$$\begin{cases} \mathbf{x} = \mathbf{f}(\mathbf{t}) + \boldsymbol{\varepsilon} \equiv \begin{cases} x_1 = 2t_1^2 + t_2^2 + \varepsilon_1 \\ x_2 = t_1^2 + 2t_2^2 + \varepsilon_2 \\ x_3 = 3t_1 + t_2 + \varepsilon_3 \\ x_4 = -t_1^3 + 3t_1^2 - t_2 + \varepsilon_4 \\ x_5 = -t_2^3 + 3t_2^2 - t_1 + \varepsilon_5 \end{cases} \\ \\ \mathbf{y} = \mathbf{C}\mathbf{t} + \boldsymbol{\eta} \equiv t_1\mathbf{c}_1 + t_2\mathbf{c}_1 + \boldsymbol{\eta} \quad \text{with} \quad \mathbf{c}_i = \mathbf{c}_i^*/||\mathbf{c}_i^*|| \\ \mathbf{c}_1^* = [1.5, 0.01, -0.1, 0.01, 0.05, 0.01, 2, 0.01, 0.5]' \\ \mathbf{c}_2^* = [0.1, -1.5, 0.01, 0.01, -0.05, 2.5, 0.01, -0.5, 0.01]' \end{cases} \tag{37}$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_5]'$ with $\varepsilon_i \sim N(0, 0.005^2)$ and $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_9]'$ with $\eta_i \sim N(0, 0.005^2)$ are independent noises, and $\mathbf{t} = [t_1, t_2]'$ are the internal variables every one distributed uniformly in the range $[0.01, 4]$.

This model was used to simulate multivariate observations under normal conditions, and the generated (100 samples) data set was used to fit the KPLS model and calculate the control limits. Figure 6 shows the linear or nonlinear correlations between the predictor variables (x_i) and the responses (y_i). For example, the correlations between x_3 and y_1,...,y_9 are linear, while most of the remaining correlations $x_i - y_i$ are clearly nonlinear. Figure 6 illustrates the high level of non-linearity present in the data. Figure 7 shows the calibration technique proposed in Section 4, where a model order $A = 4$ is selected.

The abnormal data set (250 samples) is also generated by using Eq. (37) and considering the six fault scenarios displayed sequentially in Table 3. This Table shows the six anomalies (one for each class in Table 1) simulated as follows: a) the bias faults of classes 1 and 2 are simulated by disturbing the measurements $\mathbf{x}$ and $\mathbf{y}$; b) the anomalies of classes 3, 4, and 5 are implemented by modifying the process parameters or functions; and c) the anomaly 6 consists in adding up a mean change $\Delta t_1 = 1$ to t_1. Each fault is simulated during 20 consecutive samples and immediately cancelled thereafter.

Figure 8 shows the time evolution of the detection index I_{KPLS} normalized by its control limit I_α determined by the KDE approach, and the statistics composing the index I_{KPLS} in order to interpret the anomaly class. In Fig. 8, the alarm condition is triggered at a given sample k, when the normalized global index overpasses the limit, i.e. when $I_{KPLS}(k)/I_\alpha \geq 1$. The inspection of such a figure leads to conclude that monitoring based on $I_{KPLS}(k)$ proved to be effective for detecting all simulated anomalies. The patterns presented by the contributing statistics in Fig. 8 along with the information given in Table 1 allow an unambiguous diagnosis of each kind of anomaly, except for the first two cases (see Table 3).

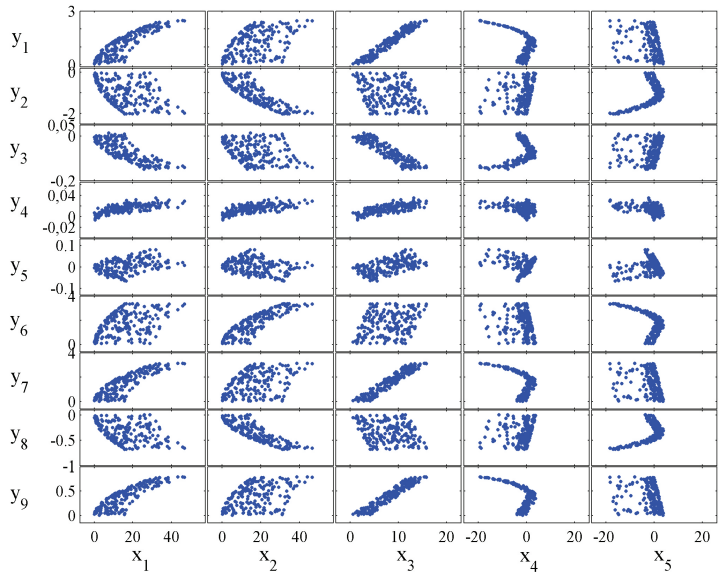


Figure 6: Scatter plots of the responses vs predictor variables using calibration data set - Case study no. 2

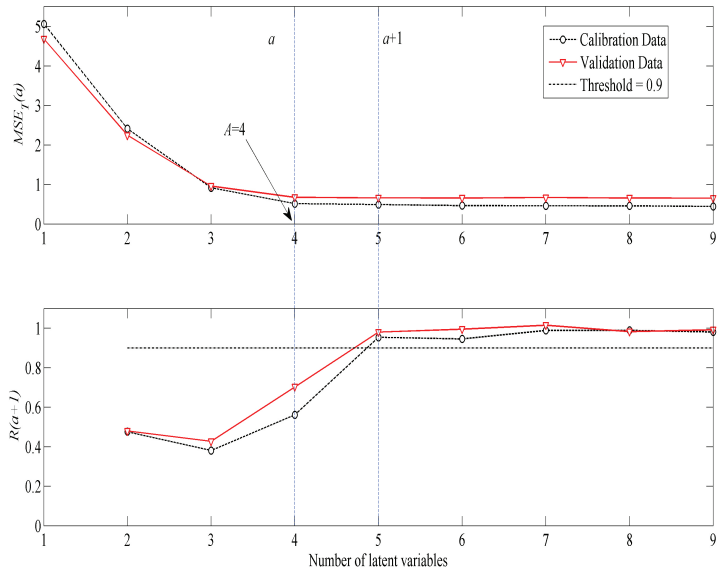


Figure 7: Calibration based on proposed methodology - Case study no. 2

As it was anticipated, faults of class-1 and class-3 produce the same pattern, hindering the discrimination between them. However, the x_3-sensor fault causes a unique significant contribution to I_{WR} due to the variable x_3 (Fig. 9). By contrast, a correlation change in $\mathbf{x}$ causes several significant contributions to I_{WR}, allowing discrimination from the previous case. Table 3 summarizes the classification of each detected abnormal event according to pattern of alarms in the statistics showed in Fig. 8. Figure 9 shows the variable

Table 3: Simulated fault scenarios - Case study no. 2

Fault type (Class)	Samples	Magnitude	Alarmed Statistics	Diagnosed Class
Offset in x_3-sensor (1)	20 to 40	$\Delta x_3 = 5$	I_{WR} and I_{RY1}	1/3
Correlation change in $\mathbf{x}$ (3)	60 to 80	$x_3^{new} = (x_3)^{1.2} + \varepsilon_3$	I_{WR} and I_{RY1}	1/3
Process upset (6)	100 to 120	$t_1^{new} = t_1 + 1$	I_{WM} and I_{WR}	6
Offset in y_3-sensor (2)	140 to 160	$\Delta y_3 = 0.03$	I_{RY2}	2
Correlation change in $\mathbf{y}$ (5)	180 to 200	$\Delta \mathbf{c}_1 = 0.01[-1, 1, 0, -1, 1, 0, -1, 0, 1]'$	I_{RY1} and I_{RY2}	5
Intrinsic gain change (4)	220 to 240	$\Delta \mathbf{I}_{1,1} = 0.7$, $u_1^{new} = 1.7t_1$	I_{RY1}	4

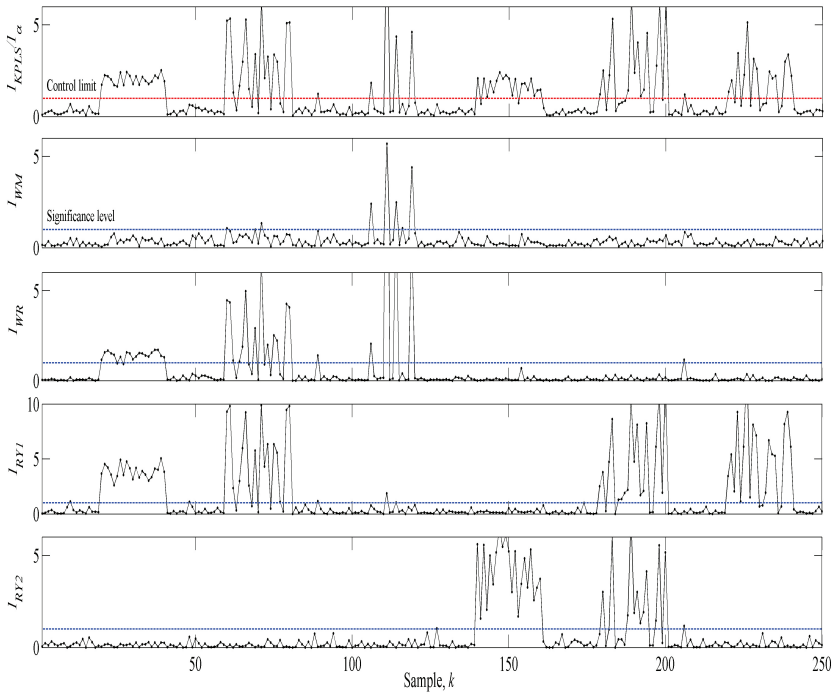


Figure 8: Time evolution of the index I_{KPLS} normalized and of its component statistics

contributions to the statistics I_{WR} and I_{RY2}, at the alarm locations $k = 31$ and $k = 151$. Figure 9 identifies the perturbed variables (x_3 and y_3) from the major (positive) contributions that more significantly affect the statistic showing alarm (cI_{WR} and cI_{RY2}, respectively), thus correctly reporting the faulty sensors. The analysis of the contributions to each statistic (at alarmed locations) allows the identification of broken relationships between the process variables. Consequently, the main contributions in cI_{WR} and cI_{RY2} characterize the correlation changes in $\mathbf{x}$ and $\mathbf{y}$, respectively, thus indicating major changes in the original external correlations.

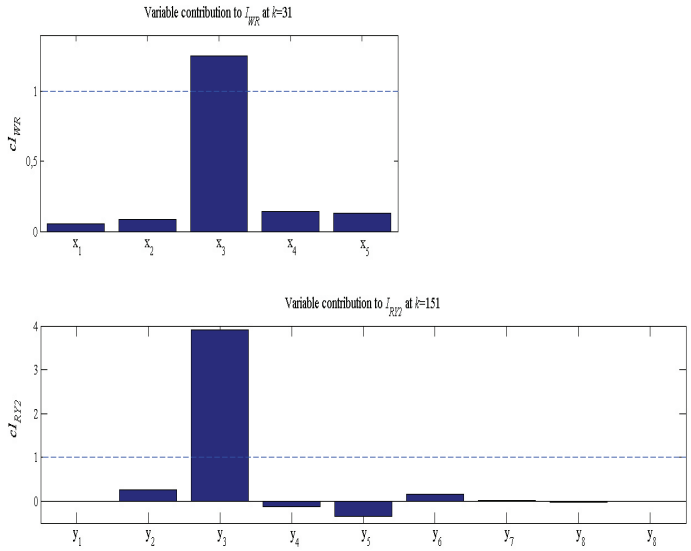


Figure 9: Contribution analysis at two alarmed locations (significance level = 1)

In these simulations, small deviations with respect to the normal behavior were adopted only, in order to evaluate the ability of the data-driven control volume associated to I_{KPLS} for detecting the simulated faults.

In summary, the proposed KPLS-technique for fault detection and diagnosis in strongly nonlinear processes has proven capable of: i) detecting an anomaly through a single combined index, ii) diagnosing the anomaly class from pattern presented by the four contributing statistics as compared to their respective confidence levels, and iii) identifying the disturbed variables based on the analysis of the main variable contributions to each significant statistic.

It is worth noting that if the normal operating range of $\mathbf{t}$ in Eq. (37) decreases, the nonlinearity of the system is less meaningful. Indeed, for the case study no. 2, when $[t_1, t_2]'$-operating zone is reduced to $[0.01, 2] \times [0.01, 2]$, the previous suggested methodology for selecting the KPLS model order gives $A = 2$. Moreover, if the same data set is used for developing a PLSR model [10, 28] the order is $A = 2$ once again, i.e., both methods converge to the same order (or number of latent variables). In summary, when the nonlinearity of the process is rather weak, we conclude the following: i) the linear PLSR approach is preferable when the model would mainly be used for monitoring purposes; and ii) the nonlinear KPLS approach is preferable when the model would mainly be used for prediction purposes. The reason for i) is that we would only need to know when the process moves out from the model; and in such case, the PLSR technique unambiguously differentiate classes 1 and 3. In contrast, the reason for ii) is the greater generalization ability of the KPLS method.

7. Conclusions

Monitoring techniques based on KPLS models designed under in-control conditions are especially useful for supervising strong nonlinear processes. In general, the results obtained in this work suggest that the proposed calibration strategy provides a comprehensive methodology for the systematic development of nonlinear KPLS models.

Meaningful deviations of the measurements from their expected behaviors are useful for detecting and diagnosing process anomalies. The proposed detection index I_{KPLS} combines several statistics of properly scaled metrics. This index represents a statistical distance that considers the linear/nonlinear correlation structure of the process as well as three Euclidean distances to the model. Unlike other existing data-driven techniques, the here proposed I_{KPLS} index allows a simultaneous monitoring of the process and the quality variables.

When an anomaly occurs in a process, the combination of alarm signals in the statistics composing I_{KPLS} is efficiently used for classifying the perturbation source. Such preliminary diagnosis is then completed through an analysis of contributions that allows the identification of the disturbed variables. Besides, the risk assessment index I_{risk} proved to be effective for validating the reliability of the KPLS predictions. The numerical simulations included in this study verify the effectiveness of the presented methodology and suggests the potential application of the presented monitoring techniques to real production systems.

Acknowledgements

The authors are grateful for the financial support received from CONICET, MinCyT, Universidad Nacional del Litoral, and Universidad Tecnológica Nacional (Argentina).

AppendixA. Proof of Eq. (30)

By substitution of Eqs. (9) and (28) into Eq. (27), we have $D_{\hat{\mathbf{y}}} = (N-1)\mathbf{t}'\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C}\mathbf{t}$. Then, using the singular value decomposition of the matrix $\mathbf{C}(p\times A)$ (with full-column rank) given by: $\mathbf{C}' = \mathbf{V}\left[\mathbf{\Sigma}\ \mathbf{0}\right]\mathbf{W}'$, where $\mathbf{\Sigma}(A\times A)$ is a non-singular diagonal matrix, and the matrices $\mathbf{W}(p\times p)$ and $\mathbf{V}(p\times A)$ are orthonormal, we have

$$\mathbf{C}'\left(\mathbf{C}\mathbf{C}'\right)^{-}\mathbf{C} = \mathbf{V}\left[\mathbf{\Sigma}\ \mathbf{0}\right]\mathbf{W}'\left(\mathbf{W}\begin{bmatrix}\mathbf{\Sigma}\\ \mathbf{0}\end{bmatrix}\mathbf{V}'\mathbf{V}\left[\mathbf{\Sigma}\ \mathbf{0}\right]\mathbf{W}'\right)^{-}\mathbf{W}\begin{bmatrix}\mathbf{\Sigma}\\ \mathbf{0}\end{bmatrix}\mathbf{V}' = \mathbf{V}\mathbf{\Sigma}\mathbf{\Sigma}^{-2}\mathbf{\Sigma}\mathbf{V}' = \mathbf{I}_{(A\times A)}. \quad \text{(A.1)}$$

Therefore, $D_{\hat{\mathbf{y}}} = (N-1)\mathbf{t}'\mathbf{t} = T^2_{\mathbf{t}}$. Furthermore, by replacing $\check{\boldsymbol{\varphi}} = \mathbf{P}\mathbf{t}$ (Eq. 17) and Eq. (29) into Eq. (26), it results: $D_{\check{\boldsymbol{\varphi}}} = (N-1)\mathbf{t}'\mathbf{P}'\left(\mathbf{P}\mathbf{P}'\right)^{-}\mathbf{P}\mathbf{t}$. Similarly, given that $\mathbf{P}$ is full-column rank can also be proved that $\mathbf{P}'\left(\mathbf{P}\mathbf{P}'\right)^{-}\mathbf{P} = \mathbf{I}(A\times A)$, hence $D_{\check{\boldsymbol{\varphi}}} = (N-1)\mathbf{t}'\mathbf{t} = T^2_{\mathbf{t}}$. The above relationships indicate that

$$D_{\hat{\mathbf{y}}} = T^2_{\mathbf{t}} = D_{\check{\boldsymbol{\varphi}}}, \quad \text{(A.2)}$$

i.e., the statistics on $\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x})$, $\hat{\mathbf{y}}$, and $\mathbf{t}$ are equivalents.

AppendixB. Determination of statistics pattern

In order to characterize the anomaly from the pattern of statistics composing I_{KPLS}, an artificial process system (identified with the subscript 0) is created for generating ideal data obeying to a predetermined correlation/functional structure. This artificial process is defined via the KPLS parameters such as Λ_0, $\mathbf{C}_0$, and $\mathbf{f}_0(\cdot)$, which represent a nonlinear model of the system under operating normal conditions. In this procedure, a random score vector $\mathbf{t}_0 \in \mathbb{R}^A \sim N(\mathbf{0}, \Lambda_0)$ is taken as an independent variable that models the associated input and output vectors through,

$$\begin{aligned} \mathbf{x}_0 &= \mathbf{f}_0(\mathbf{t}_0) \in \mathbb{R}^m, \\ \mathbf{y}_0 &= \mathbf{C}_0\mathbf{t}_0 \in \mathbb{R}^p \sim N(\mathbf{0}, (\mathbf{C}_0\Lambda_0\mathbf{C}_0')) \end{aligned} \tag{B.1}$$

where $\mathbf{y}_0 \in S_{MY} \equiv \text{span}\{\mathbf{C}_0\} \subseteq \mathbb{R}^p$ and $\mathbf{f}_0(\cdot)$ is associated to $\bar{\boldsymbol{\varphi}}$ (which is the inverse function of $\mathbf{f}_0$ in the in-control domain), such that $\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) = \bar{\boldsymbol{\varphi}}(\mathbf{f}_0(\mathbf{t}_0)) = \mathbf{P}_0\mathbf{t}_0 \in W_M \equiv \text{span}\{\mathbf{P}_0\} \subseteq \mathbb{R}^c$. Hence, Eq. (B.1) considers the common-cause variations only. Since these data stand for an ideal perfect model, the residuals $\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}_0)$, $\tilde{\mathbf{y}}_1$, and $\tilde{\mathbf{y}}_2$ are null, and there are no differences between model predictions and the data, i.e. $\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}_0) = \mathbf{P}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) = \bar{\boldsymbol{\varphi}}(\mathbf{x}_0)$, $\check{\mathbf{y}} = \mathbf{C}_0\mathbf{D}_0'\mathbf{y}_0 = \mathbf{y}_0$ and $\hat{\mathbf{y}} = \mathbf{C}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) = \mathbf{y}_0$.

Then, several alternatives to the normal condition are analyzed by assuming localized pure disturbances and observing the mismatch with the available model (Λ_0, $\mathbf{C}_0$, and $\mathbf{f}_0(\cdot)$). The sketch in Fig. 2 helps to visualize how a warning signal (at $\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})$, $\tilde{\mathbf{y}}_1$, $\tilde{\mathbf{y}}_2$ or $\mathbf{t}$) is generated as the disturbed measurements ($\mathbf{x}$ or $\mathbf{y}$) go through the KPLS model. In this context, the following classes of anomalies can be discriminated:

- *Class 1 (Sensor faults associated to* $\mathbf{x}$*)*: These faults are represented by a shift signal $\Delta\mathbf{x}$ producing $\mathbf{x}$-vector readings out of the pattern identified by the KPLS model. In this case, the input vector can be written as,

$$\mathbf{x} = \mathbf{x}_0 + \Delta\mathbf{x} \qquad \text{(disturbed measurements)} \tag{B.2}$$

  where $\mathbf{x}_0$ is the fault-free part of the input measurements. More specifically, let us assume that $\Delta\mathbf{x}$ is such that $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ is moved out of W_M and towards W_R. Then, using the first order Taylor approximation of $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ we have:

$$\begin{aligned} \check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) =& \mathbf{P}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}) \cong \mathbf{P}_0\mathbf{R}_0'(\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) = \bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \\ \tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) =& \bar{\boldsymbol{\varphi}}(\mathbf{x}) - \check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) \cong \Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \neq 0 \in W_R \text{ (disturbance detection)} \\ \tilde{\mathbf{y}}_1 =& \check{\mathbf{y}} - \hat{\mathbf{y}} = \mathbf{C}_0\mathbf{D}_0'\mathbf{y}_0 - \mathbf{C}_0\mathbf{R}_0'\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) = 0 \\ \tilde{\mathbf{y}}_2 =& \mathbf{y} - \check{\mathbf{y}} = \mathbf{y}_0 - \mathbf{C}_0\mathbf{D}_0'\mathbf{y}_0 = 0 \end{aligned} \tag{B.3}$$

On the other hand, if assuming that $\Delta\mathbf{x}$ is such that $\Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \in W_M$ then $\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) \cong \mathbf{P}_0\mathbf{R}_0'(\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) \neq \bar{\boldsymbol{\varphi}}(\mathbf{x}_0)$ affecting a different residue as follows:

$$\begin{aligned}\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) &= (\mathbf{I} - \mathbf{P}_0\mathbf{R}_0')\,\bar{\boldsymbol{\varphi}}(\mathbf{x}) \cong (\mathbf{I} - \mathbf{P}_0\mathbf{R}_0')\,(\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) = 0 \\ \tilde{\mathbf{y}}_1 &= \mathbf{C}_0\mathbf{D}_0'\mathbf{y}_0 - \mathbf{C}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}) \cong -\mathbf{C}_0\mathbf{R}_0'\Delta\mathbf{x}'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \neq 0 \in S_{MY} \quad \text{(disturbance detection)}\end{aligned} \tag{B.4}$$

Furthermore, since $\tau_\alpha^2 \gg \delta_{\mathbf{x}}^2$ then $D_{\check{\bar{\boldsymbol{\varphi}}}}/\tau_\alpha^2 \ll ||\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2/\delta_{\mathbf{x}}^2$ when $||\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2 = ||\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2$, hence the statistic I_{WM} is not affected. Therefore, the residues $\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})$ and $\tilde{\mathbf{y}}_1$ are used to detect this disturbance.

- *Class 2 (Sensor faults associated to* $\mathbf{y}$*)*

$$\mathbf{y} = \mathbf{y}_0 + \Delta\mathbf{y} \qquad \text{(disturbed measurements)} \tag{B.5}$$

where $\mathbf{y}_0$ is the fault-free part that follows the normal correlation structure. The disturbance is analyzed by assuming that $\Delta\mathbf{y} \in S_{RY}$ [28]. Hence, the disturbance track from generation to detection is as follows:

$$\begin{aligned}\check{\mathbf{y}} &= \mathbf{C}_0\mathbf{D}_0'\mathbf{y} = \mathbf{C}_0\mathbf{D}_0'\mathbf{y}_0 + \mathbf{C}_0\mathbf{D}_0'\Delta\mathbf{y} = \mathbf{y}_0 \\ \tilde{\mathbf{y}}_2 &= \mathbf{y} - \check{\mathbf{y}} = \mathbf{y}_0 + \Delta\mathbf{y} - \mathbf{y}_0 = \Delta\mathbf{y} \neq 0 \in S_{RY} \quad \text{(disturbance detection)}\end{aligned} \tag{B.6}$$

thus, a measurement disturbance is sent to the residual space S_{RY}. Note that, $\mathbf{t} = \mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{t}_0$ then $\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) = 0$ and $\tilde{\mathbf{y}}_1 = \check{\mathbf{y}} - \mathbf{C}_0\mathbf{t} = 0$.

- *Class 3 (Changes in the nonlinear correlation structure of* $\mathbf{x}$*)* A change in the correlations of $\mathbf{x}$ can be thought as an unknown functional change $\Delta\mathbf{f}(\cdot) = \mathbf{f}(\cdot) - \mathbf{f}_0(\cdot)$, i.e.

$$\mathbf{x} = \mathbf{f}(\mathbf{t}_0) = \mathbf{f}_0(\mathbf{t}_0) + \Delta\mathbf{f}(\mathbf{t}_0) = \mathbf{x}_0 + \Delta\mathbf{f}(\mathbf{t}_0) \qquad \text{(disturbed measurements)} \tag{B.7}$$

Now the Taylor approximation of the mapped measurements is given by

$$\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \bar{\boldsymbol{\varphi}}(\mathbf{x}_0 + \Delta\mathbf{f}(\mathbf{t}_0)) \cong \bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{f}(\mathbf{t}_0)'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \tag{B.8}$$

The changes in $\bar{\boldsymbol{\varphi}}(\mathbf{x})$ shown in Eq. (B.8) can belong to both subspaces W_M and W_R, thus generating the following no-null residual values

$$\begin{aligned}\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x}) &\cong (\mathbf{I} - \mathbf{P}_0\mathbf{R}_0')\,(\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{f}(\mathbf{t}_0)'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) = (\mathbf{I} - \mathbf{P}_0\mathbf{R}_0')\,\Delta\mathbf{f}(\mathbf{t}_0)'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) \neq 0 \in W_R \\ \tilde{\mathbf{y}}_1 &\cong \check{\mathbf{y}} - \mathbf{C}_0\mathbf{R}_0'\,(\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) + \Delta\mathbf{f}(\mathbf{t}_0)'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0)) = -\mathbf{C}_0\mathbf{R}_0'\Delta\mathbf{f}(\mathbf{t}_0)'\nabla\bar{\boldsymbol{\varphi}}(\mathbf{x}_0) \neq 0 \in S_{MY}\end{aligned} \tag{B.9}$$

Since $\tau_\alpha^2 \gg \delta_{\mathbf{x}}^2$ then $D_{\check{\bar{\boldsymbol{\varphi}}}}/\tau_\alpha^2 \ll ||\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2/\delta_{\mathbf{x}}^2$ when $||\check{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2 = ||\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})||^2$, hence the statistic I_{WM} is not affected. Then, the residues $\tilde{\bar{\boldsymbol{\varphi}}}(\mathbf{x})$ and $\tilde{\mathbf{y}}_1$ are used to detect the disturbance.

- *Class 4 (Changes in the intrinsic relations)* The identity matrix $\mathbf{I}$ at Fig. 2 is the core place where the KPLS model ties up input with output internal variables. Let us assume that an unknown change occurs in this relationship, i.e.

$$\check{\mathbf{y}} = \mathbf{C}_0(\mathbf{I} + \Delta\mathbf{I})\mathbf{t}_0 = \mathbf{y}_0 + \mathbf{C}_0\Delta\mathbf{I}\mathbf{t}_0 \qquad \text{(disturbed measurements)}. \tag{B.10}$$

whilst $\mathbf{x} = \mathbf{f}_0(\mathbf{t}_0) = \mathbf{x}_0$. Analyzing the effects on the statistics

$$\tilde{\mathbf{y}}_1 = \check{\mathbf{y}} - \hat{\mathbf{y}} = (\mathbf{y}_0 + \mathbf{C}_0\Delta\mathbf{I}\mathbf{t}_0) - \mathbf{C}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{C}_0\Delta\mathbf{I}\mathbf{t}_0 \neq 0 \in S_{MY} \quad \text{(disturbance detection)}, \tag{B.11}$$

while the remaining statistics are not affected.

- *Class 5 (Changes in the correlation structure of* $\mathbf{y}$*)* Let us assume an unknown change in the matrix $\mathbf{C}_0$, i.e.

$$\mathbf{y} = (\mathbf{C}_0 + \Delta\mathbf{C})\mathbf{t}_0 = \mathbf{y}_0 + \Delta\mathbf{C}\mathbf{t}_0 \qquad \text{(disturbed measurements)}. \tag{B.12}$$

Hence, the disturbance detection is characterized by

$$\begin{aligned}
\check{\mathbf{y}} &= \mathbf{C}_0\mathbf{D}_0'(\mathbf{y}_0 + \Delta\mathbf{C}\mathbf{t}_0) = \mathbf{y}_0 + \mathbf{C}_0\mathbf{D}_0'\Delta\mathbf{C}\mathbf{t}_0 \\
\tilde{\mathbf{y}}_2 &= (\mathbf{y}_0 + \Delta\mathbf{C}\mathbf{t}_0) - (\mathbf{y}_0 + \mathbf{C}_0\mathbf{D}_0'\Delta\mathbf{C}\mathbf{t}_0) = (\mathbf{I} - \mathbf{C}_0\mathbf{D}_0')\Delta\mathbf{C}\mathbf{t}_0 \neq 0 \in S_{RY} \\
\tilde{\mathbf{y}}_1 &= (\mathbf{y}_0 + \mathbf{C}_0\mathbf{D}_0'\Delta\mathbf{C}\mathbf{t}_0) - \mathbf{C}_0\mathbf{R}_0'\bar{\boldsymbol{\varphi}}(\mathbf{x}) = \mathbf{C}_0\mathbf{D}_0'\Delta\mathbf{C}\mathbf{t}_0 \neq 0 \in S_{MY}.
\end{aligned} \tag{B.13}$$

- *Class 6 (Significant change in the process population parameters)* This anomaly produces measurements that follow the correlation structure captured by the KPLS model, and can be represented by changes in the original process parameters $\mathbf{t}_0 \sim N(\mathbf{0}, \Lambda_0)$. Let us assume a displacement of $E\{\mathbf{t}_0\}$ from $\mathbf{0}$ to $\mu_{\mathbf{t}} \neq 0$, or a significant change in the variability from Λ_0 to $\Lambda_{\mathbf{t}}$, thus producing an arbitrary distribution $N(\mu_{\mathbf{t}}, \Lambda_{\mathbf{t}})$. Hence, the abnormal event produces,

$$\mathbf{t} = \mathbf{t}_0 + \Delta\mathbf{t} \sim N(\mu_{\mathbf{t}}, \Lambda_{\mathbf{t}}) \qquad \text{(disturbed measurements)}, \tag{B.14}$$

with a magnitude such that the $T_{\mathbf{t}}^2 = ||\Lambda_0^{-1/2}\mathbf{t}||^2 = ||\Lambda_0^{-1/2}(\mathbf{t}_0 + \Delta\mathbf{t})||^2 \geq \tau_\alpha^2$ (disturbance detection).

The previous analysis is summarized in Table 1, where the highlighted discrimination patterns indicate the statistics that are activated as soon as the measurements $(\mathbf{x}, \mathbf{y})$ bring information about a localized model mismatch.

AppendixC. Partial derivatives of the contributions

The partial derivatives of I_{WM}, I_{WR}, and I_{RY1}, are the following:

$$\frac{\partial I_{WM}}{\partial x_i} = \frac{(N-1)2}{\tau_\alpha^2}\bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}\mathbf{V}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i}, \tag{C.1}$$

$$\frac{\partial I_{WR}}{\partial x_i} = \frac{1}{\delta_{\mathbf{x}}^2}(-2\mathbf{e}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i} - 4\bar{\mathbf{k}}'(\mathbf{x})\bar{\mathbf{K}}\mathbf{V}\mathbf{V}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i} + 2\bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}\mathbf{T}'\bar{\mathbf{K}}\mathbf{T}\mathbf{V}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i}), \tag{C.2}$$

$$\frac{\partial I_{RY1}}{\partial x_i} = \frac{1}{\delta_{\mathbf{y}_1}^2}(-2\mathbf{y}'\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{V}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i} + \bar{\mathbf{k}}'(\mathbf{x})\mathbf{V}\mathbf{C}'\mathbf{C}\mathbf{V}'\frac{\partial\bar{\mathbf{k}}(\mathbf{x})}{\partial x_i}), \tag{C.3}$$

where $\partial\bar{\mathbf{k}}(\mathbf{x})/\partial x_i = [\partial\bar{k}(\mathbf{x}_1,\mathbf{x})/\partial x_i, \ldots, \partial\bar{k}(\mathbf{x}_N,\mathbf{x})/\partial x_i]'$. Given that $\bar{k}(\mathbf{x}_j,\mathbf{x}) = k(\mathbf{x}_j,\mathbf{x}) - \mathbf{k}'(\mathbf{x}_j)\mathbf{e} - \mathbf{k}'(\mathbf{x})\mathbf{e} + \mathbf{e}'\mathbf{K}\mathbf{e}$ (see Eq. (7)), the elements of this vector are given by,

$$\frac{\partial\bar{k}(\mathbf{x}_j,\mathbf{x})}{\partial x_i} = \frac{\partial k(\mathbf{x}_j,\mathbf{x})}{\partial x_i} - \frac{1}{N}\sum_{r=1}^{N}\frac{\partial k(\mathbf{x}_r,\mathbf{x})}{\partial x_i}. \tag{C.4}$$

where $\partial k(\mathbf{x}_j,\mathbf{x})/\partial x_i = (-2/h)(x_{j,i} - x_i)\exp\left(-||\mathbf{x}_j - \mathbf{x}||^2/h\right)$ for a Gaussian kernel function. The partial derivatives of I_{RY1} and I_{RY2} respect to y_i are given by,

$$\frac{\partial I_{RY1}}{\partial y_i} = \frac{1}{\delta^2_{\mathbf{y}_1}}(2\xi_i'\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{D}'\mathbf{y} - 2\xi_i'\mathbf{D}\mathbf{C}'\mathbf{C}\mathbf{V}'\bar{\mathbf{k}}(\mathbf{x})), \tag{C.5}$$

$$\frac{\partial I_{RY2}}{\partial y_i} = \frac{2}{\delta^2_{\mathbf{y}_2}}\xi_i'\left(\mathbf{I} - \mathbf{C}\mathbf{D}'\right)'\left(\mathbf{I} - \mathbf{C}\mathbf{D}'\right)\mathbf{y}, \tag{C.6}$$

where $\xi_i' = [0,\ldots,1,\ldots,0]$ is a vector with zeros entries except for the location i, which takes an unitary value. Then, the contributions of any variable x_i or y_i to the component statistics are computed via Eq. (33) and are denoted by cI_{WM}, cI_{WR}, cI_{RY1}, and cI_{RY2}; respectively.

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S. Kavuri, A rewiew of process fault detection and diagnosis: Process history based methods, Computers & Chemical Engineering 27 (2003) 327–346.

[2] D. Zumoffen, M. Basualdo, Fault Detection Systems Integrated to Fault-Tolerant Control. Application to Large-Scale Chemical Processes, LAP Lambert Academic Publishing, Germany, 2012.

[3] T. Kourti., Application of latent variable methods to process control and multivariate statistical process control in industry, International Journal of Adaptive Control and Signal Processing 19 (2005) 213246.

[4] H. Yue, S. Qin, Reconstruction-based fault identification using a combined index, Industrial & Engineering Chemistry Research 40 (2001) 4403–4414.

[5] T. Kourti, J. MacGregor, Process analysis, monitoring and diagnosis using multivariate projection methods, Chemometrics and Intelligent Laboratory Systems 28 (1995) 3–21.

[6] S. Wold, M. Sjöström, L. Eriksson., PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2001) 109–130.

[7] B. Li, A. Morris, E. Martin, Model selection for partial least squares regression, Chemometrics and Inhtelligent Laboratory Systems 64 (2002) 79–89.

[8] Y. Zhang, Y. Zhang, Complex process monitoring using modified partial least squares method of independent component regression, Chemometrics and Intelligent Laboratory Systems 98 (2009) 143–148.

[9] D. Zumoffen, M. Basualdo, A systematic approach for the design of optimal monitoring systems for large scale processes, Industrial & Engineering Chemistry Research 49 (2010) 1749–1761.

[10] J. L. Godoy, R. J. Minari, J. R. Vega, J. L. Marchetti, Multivariate statistical monitoring of an industrial SBR process. soft-sensor for production and rubber quality, Chemometrics and Intelligent Laboratory Systems 107 (2011) 258–268.

[11] S. Mika, B. Schölkopf, A. Smola, K. Müller, M. Scholz, G. Rätsch, Kernel PCA and de-noising in feature spaces, Proceedings of the 1998 conference on Advances in neural information processing systems II. MIT Press Cambridge (1999) 536–542.

[12] K. Kim, J. Lee, I. Lee, A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction, Chemometrics and Intelligent Laboratory Systems 79 (2005) 22–30.

[13] S. Zhao, J. Zhang, Y. Xu, Z. Xiong, Nonlinear projection to latent structures method and its applications, Industrial & Engineering Chemistry Research 45 (2006) 3843–3852.

[14] X. Zhang, W. Yan, H. Shao, Nonlinear multivariate quality estimation and prediction based on kernel partial least squares, Industrial & Engineering Chemistry Research 47 (2008) 1120–1131.

[15] M. Jia, F. Chu, F. Wang, W. Wang, On-line batch process monitoring using batch dynamic kernel principal component analysis, Chemometrics and Intelligent Laboratory Systems 101 (2010) 110–122.

[16] C. Alcala, S. Qin, Reconstruction-based contribution for process monitoring with kernel principal component analysis, Industrial & Engineering Chemistry Research 49 (2010) 7849–7857.

[17] R. Rosipal, L. Trejo, Kernel partial least squares regression in reproducing kernel hilbert space, Journal of Machine Learning Research 2 (2001) 97–123.

[18] Y. Hu, H. Ma, H. Shi, Enhanced batch process monitoring using just-in-time-learning based kernel partial least squares, Chemometrics and Intelligent Laboratory Systems 123 (2013) 15–27.

[19] Y. Zhang, Z. Hu., On-line batch process monitoring using hierarchical kernel partial least squares, Chemical engineering research and design 89 (2011) 2078–2084.

[20] P. Williams, Influence of water bands on prediction of composition and quality factors: the aquaphotomics of low moisture agricultural materials, Journal of Near Infrared Spectroscopy 17 (2009) 315–328.

[21] C. Tan, M. Li, Mutual information-induced interval selection combined with kernel partial least squares for near-infrared spectral calibration, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 71 (2008) 1266–1273.

[22] Y. Zhang, Y. Teng, Process data modeling using modified kernel partial least squares, Chemical Engineering Science 65 (2010) 6353–6361.

[23] Y. Wang, Y. Li, B. Wang, An in silico method for screening nicotine derivatives as cytochrome P450 2A6 selective inhibitors based on kernel partial least squares, International Journal of Molecular Sciences 8 (2007) 166–179.

[24] M. Embrechts, S. Ekins, Classification of metabolites with kernel-partial least squares (K-PLS), Drug Metabolism and Disposition 35 (2007) 325–327.

[25] K. Peng, K. Zhang, G. Li, D. Zhou, Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process, Control Engineering Practice 21 (2013) 360–369.

[26] S. Qin, Survey on data-driven industrial process monitoring and diagnosis, Annual Reviews in Control 36 (2012) 220–234.

[27] G. Postma, P. Krooshof, L. Buydens, Opening the kernel of kernel partial least squares and support vector machines, Analytica Chimica Acta 705 (2011) 123–134.

[28] J. Godoy, J. Vega, J. Marchetti, A fault detection and diagnosis technique for multivariate processes using a PLS-decomposition of the measurement space, Chemometrics and Intelligent Laboratory Systems 128 (2013) 25–36.

[29] C. Alcala, S. Qin, Analysis and generalization of fault diagnosis methods for process monitoring, Journal of Process Control 21 (2011) 322–330.

[30] A. Willis, Condition monitoring of centrifuge vibrations using kernel PLS, Computers and Chemical Engineering 34 (2010) 349–353.

[31] G. McLachlan, Finite mixture models, Wiley, New York, 2012.

[32] C. Beardah, M. Baxter, Matlab routines for kernel density estimation and the graphical representation of archaeological data, Tech. rep., Nottingham Trent University Department of Mathematics, Statistics, and Operational Research, United Kingdom (1995).

[33] C. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, UK, 1995.

[34] E. Martin, A. Morris, Non-parametric confidence bounds for process performance monitoring charts, Journal of Process Control 6 (6) (1996) 349–358.