Criteria to optimizing kernel methods in fault monitoring process: A survey

JM. Bernal-de Lázaro, O. Llanes-Santiago, A. Silva-Neto, and C. Cruz-Corona

Abstract—Over two last decade, diverse fault diagnosis strategies based on kernel methods have been studied intensively to ensure the safety and reliability of large-industrial processes. However, how choice a proper kernel function and its parameters are still two open research issues. This paper provides an overview of kernel-based preprocessing methods in fault monitoring tasks, with emphasis on helpful procedures to select the optimal kernel parameters in fault diagnosis applications. The discussion is also towards evaluating some kernel functions available from the machine learning literature but less used in fault diagnosis applications. With this aim, a KPCA-based monitoring scheme is employing to validate the effectiveness of sixteen kernel functions in two case studies.

Index Terms—Kernel methods, kernel function, fault monitoring, preprocessing methods, tuned parameters, kernel PCA

1 Introduction

Kernel learning has become an important research topic for the data-based monitoring field because it provides potential solutions to nonlinear problems, including feature extraction, classification, and clustering tasks. Accordingly, numerous kernel approaches have been addressed to deal with the increasing complexity of modern industrial plants [1–5]. For example, the Support Vector Machines (SVM) paradigms have widely used to build effective discriminant models that are applied in fault isolation tasks [5–8]. Kernelized regression procedures have been also addressed to enhance the fault identification and the failure time prediction in complex production processes [9–12]. On the other hand, adaptive fault diagnosis schemes using kernel clustering strategies have been proposed for handling imbalanced classes and measurements affected by noise and outliers [13–16]. Many other kernelized algorithms have also been successfully employed as data pre-treatments methods in fault diagnosis applications, including algorithms as the Principal Component Analysis (KPCA) [17–19], Partial Least Squares (KPLS) [20–23], Fisher Discriminate Analysis (KFDA) [24–26], Entropy Component Analysis (KECA) [27–32], Independent Component Analysis (KICA) [33–35], as well as, Canonical Component Analysis (KCCA) [36–38]. Among all these data-pretreatments methods, kernel PCA is the more popular and extensively studied for fault diagnosis applications.

The main advantage of the above kernel-based solutions stems from the inherent mapping of input data into a higher-dimensional space through the kernel trick, which allows us to handle more effectively their nonlinear relationships. However, two bottlenecks are limiting the application of these approaches: (a) the computation complexity usually increases with sample size and dimension size [4]; (b) the proper choice of the kernel function and its parameters must be ensured to achieve high performances. Most authors agree that the selection and tuning of kernel functions are two open research issues [22, 39, 40]. However, the number of published papers that explicitly address the adjusting of the kernel-based diagnosis schemes is still very limited. Likewise, a systematic review of kernel parameter optimization criteria for the data-based process monitoring has not been recently reported. Since the lack of rigorous theoretical framework and appropriate guidelines, many times the kernel parameters are arbitrary selected without there being any guarantee of achieving acceptable performances [39, 41, 42]. These type of intuitive solutions was referred to here as heuristic criteria. Other strategies for the kernel parameter selection will be classified into two categories: (a) direct evaluation criteria that are more closely related to the kernel matrix; (b) indirect evaluation criteria where the fault diagnosis performance is evaluated to tuning the kernel parameters. These categories differ from those proposed by Xiao et al. [43], and they are not necessarily focused on the type of searching algorithm that is utilized, but rather the objective function employed to find the optimal parameters of the candidate kernels.

The first aim of this paper is to provide readers an overview of the kernel-based preprocessing methods in fault monitoring tasks, with emphasis on helpful procedures to achieve the proper tuning of the kernel functions selected. From the above view of point, the present work provides additional remarks that complement the comprehensive literature review on fault diagnosis applications of data-driven methods that have been supplied by several papers recently published [3–6, 35, 44–49], wherein these topics are not particularly detailed. The second motivation of this paper is to identify kernel functions potentially useful, that are available from

- *A. Silva-Neto is with Department of Mechanical Engineering, Universidade do Estado do Rio de Janeiro, IPRJ-UERJ, RJ, Brazil.*
- *C. Cruz-Corona is with Departament of Computer Science and Artificial Intelligence, University of Granada, Spain.*
- *JM. Bernal-de Lázaro and O. Llanes-Santiago are with Department of Computer Systems and Automation Engineering, Universidad Tecnológica de La Habana "José Antonio Echeverría", CUJAE, Cuba. E-mail: oresteslls@tesla.cujae.edu.cu*

Manuscript received August xx, 2020; revised August xx, 2020.

the machine learning literature [50–54] but doesn't be usually evaluated in fault diagnosis applications. The main contribution of this paper stems from the proposal, analysis, and justification of several kernel optimization criteria that facilitates finding the optimal kernel parameters in fault diagnosis applications. In order to further compare their performance from a practical perspective, sixteen kernel functions are evaluated using two case studies. Even though the effectiveness of the proposal is validated through a KPCA-based monitoring scheme, all optimization criteria herein discussed can be easily extensible to other kernelized algorithms.

This paper is organized as follows. Section 2 is addressed to review the theoretical concepts and related works. Section 3 presents the kernel optimization criteria and the framework proposed for their application. Section 4 validates the procedures for the adjusting of the kernel functions through two benchmark process. Finally, the conclusions are given.

2 Related works

This section outlines and provides further references for kernel methods applied in fault diagnosis tasks.

2.1 Methodology

The literature reviewed in the paper preparation includes publications since 2004 that have been obtained from ScienceDirect, IOPscience, IEEE Xplore, Thomson Reuters, and Taylor & Francis citation databases. Aiming provided an overview of the kernel-based monitoring trends, some papers published in world-renowned congresses were also analyzed. Through the information associated with titles, abstracts, and keywords, were identify hundreds of works related to kernel-based monitoring applications. The reviewed articles were also categorized into two relevant areas of research: electro-mechanical systems and complex chemical processes. Finally, the works more closely to the following scopes were considered.

- Research papers developing comparatives studies about kernel methods in fault monitoring issues.
- Research papers using several kernelized algorithms to improve the fault diagnosis procedures.
- Research papers evaluating several kernel functions.
- Research papers considering the choice of kernel parameters part of the fault diagnosis procedure.

It is important to highlight that a detailed mathematical description for all reviewed kernel methods is beyond that the scope of this paper. For interested readers, it is suggested consider to original works.

2.2 Data-pretreatment through kernel methods

In the data-driven process monitoring the measurements routinely collecting from the industrial processes are represented in a finite-dimensional space through vectors (i.e., so-called observations). A kernel function represents the inner product of these observations, which are implicitly mapped in a high dimensional space to handle more effectively the non-linear relations between the variables of the process. Let $\{\mathbf{X}_{\mathrm{i}}\}_{\mathrm{i}=1}^{\mathrm{m}} \in \mathbf{R}^{\mathrm{p}}$ denote the historical records with m observations, a kernel function can be defined by:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle \tag{1}$$

where Φ indicates the implicit mapping of input data into the Reproducing Kernel Hilbert Space (RKHS). If a symmetric positive definite function that satisfies Mercer's conditions is employed, then the kernel represents the inner product in the feature space $\mathcal{F}$, which amounts to the angle between the observations. The kernel methods assume that similarities between input data can be completely represented by the kernel matrix $\mathbf{K}_{ij}$; $\forall i, j = \{1, \ldots, m\}$, resulting to evaluate the kernel functions.

$$\mathbf{K}_{ij} = \begin{bmatrix} \Phi_1^T\Phi_1 & \cdots & \Phi_1^T\Phi_m \\ \vdots & \ddots & \vdots \\ \Phi_m^T\Phi_1 & \cdots & \Phi_m^T\Phi_m \end{bmatrix} = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_m) \\ \vdots & \ddots & \vdots \\ k(x_m,x_1) & \cdots & k(x_m,x_m) \end{bmatrix} \tag{2}$$

Therefore, the spatial structures of the input data and their similarity values are both determined by the tuning parameters of the kernel function. In order to validate this idea from the fault diagnosis perspective, the Table 1 illustrates different kernel functions that were considered for the present study. According to Genton (2001) [50], these kernels may be classified as anisotropic stationary, isotropic stationary, compactly supported, locally stationary, nonstationary, and separable nonstationary functions. For Interested practitioners, it is also suggested review the works provides by Canu and Smola (2006) [55], Al-Daoud and Turabieh (2013) [52], Zhu et al. [53], and Tian and Wang (2017) [54], where more kernel functions are analyzed for applications that go beyond the fault diagnosis field.

From the literature review, it is remarkable the number of published papers wherein the kernel-based preprocessing methods are employed to provide enhanced diagnosis schemes. The data-pretreatment through kernel methods allows to reduce the dimension of the original data and the overload of computational operations but also increases the robustness against uncertainties by preserving the relevant information of observations [56]. Hence, the kernelized algorithms require that the embedding data sets to be transformed into a reduced subspace ($d \ll p$) utilizing a transformation matrix $\mathbf{P} \in \mathbf{R}^{m \times d}$, such that the dot products of images can be substituted by a kernel [57]. Then, considering that the projection vector is a linear combination of images in $\mathcal{F}$, the matrix $\mathbf{P}$ can be expressed as

$$\mathbf{P}_j = \sum_{i=1}^{m} \beta_{j,i}\Phi(x_i) = \Phi\beta_j \tag{3}$$

where $\mathbf{P}_j$ is the *jth*-transformation vector of $\mathbf{P} = [\mathbf{P}_1, \ldots, \mathbf{P}_d]$, and $\beta_j = [\beta_{j,1}, \ldots, \beta_{j,m}]^T$ is the *jth*-vector from the pseudo transformation matrix $\mathbf{B} = [\beta_1, \ldots, \beta_d] \in \mathbf{R}^{m \times d}$. Through the kernel trick, the Eq. (3) can be modify as

$$\mathbf{B}^T k(\cdot, x) = \mathbf{B}^T \mathbf{K}\alpha \tag{4}$$

where $\mathbf{k}(\cdot, x) = [k(x_1, x), \ldots, k(x_m, x)]^T = \Phi^T\Phi(x)$, and $\mathbf{K} = \Phi^T\Phi \in \mathbf{R}^{m \times m}$ is the kernel Gram matrix which is symmetric and positive semidefinite. According to Zhang et al. [57], the problem of the dimensionality reduction in the

TABLE 1
Kernel functions.

No.	Kernel	Comments
1	$k_1(x_i,x_j)=exp\left(-\|x_i-x_j\|^2/2\theta^2\right)$	Gaussian kernel
2	$k_2(x_i,x_j)=1/\left(1+\|x_i-x_j\|^2/\theta\right)$	Cauchy kernel
3	$k_3(x_i,x_j)=1-\|x_i-x_j\|^2/\left(\|x_i-x_j\|^2+\theta\right)$	Rational Quadratic kernel
4	$k_4(x_i,x_j)=\left(1+\|x_i-x_j\|^\theta\right)^{-1}$	Generalized T-Student Kernel
5	$k_5(x_i,x_j)=exp\left(-\|x_i-x_j\|/\theta\right)$	Laplacian kernel
6	$k_6(x_i,x_j)=exp\left(-\|x_i-x_j\|/2\theta^2\right)$	Exponential kernel
7	$k_7(x_i,x_j)=\left(\langle x_i^Tx_j\rangle+c\right)^\theta$	Polynomial kernel
8	$k_8(x_i,x_j)=\sqrt{\|x_i-x_j\|^2+\theta^2}$	Multiquadratic kernel
9	$k_9(x_i,x_j)=1/\sqrt{\|x_i-x_j\|^2+\theta^2}$	Inverse Multiquadratic kernel
10	$k_{10}(x_i,x_j)=-\|x_i-x_j\|^\theta$	Power kernel
11	$k_{11}(x_i,x_j)=-log\left(1+\|x_i-x_j\|^\theta\right)$	Log kernel
12	$k_{12}(x_i,x_j)=1-\sin\left(\pi\|x_i-x_j\|/2\theta\right)$	Nonparametric kernel (NNk13)
13	$k_{13}(x_i,x_j)=\omega\left(\alpha\langle x_i^Tx_j\rangle+c\right)^\theta+(1-\omega)\,exp\left(-\|x_i-x_j\|^2/2\theta^2\right)$	RBFpoly kernel
14	$k_{14}(x_i,x_j)=1-\sum_{i=1}^{3}\left(-\|x_i-x_j\|/\theta\right)^i/2^{i-1}$	Nonparametric kernel (NNk10)
15	$k_{15}(x_i,x_j)=1-\left[1+\sum_{i=1}^{3}(-1)^i/\left(1+\left(\|x_i-x_j\|/\theta\right)^i\right)\right]$	Nonparametric kernel (NNk12)
16	$k_{16}(x_i,x_j)=1-\sum_{i=1}^{3}\left(-\|x_i-x_j\|/\theta\right)^i/i^{-1}$	Nonparametric kernel (NNk11)

More information of nonparametric kernels in Al-Daoud and Turabieh (2013) [52]

kernel feature space can be therefore formulating through the pseudo-transformation matrix in RKHS. In general, by applying eigenvalue decomposition to the kernel matrix, the pseudo-transformation vectors in the higher-dimensional space may be obtained. In this way, new descriptors through different data projections can found. For example, KPCA is a dimensionality reduction technique wherein the projections preserve the meaningful variability information in the original data set. The KICA algorithm finds a non-linear representation of the original set of variables for which the statistical dependence of the components result minimized [58]. In this context, the KECA reveals the angular structure of the transformed data set using projections onto those KPCA axes that contribute to the entropy estimate but not the variances of the input space data set [28]. In KPLS, the dimensions of two matrix blocks (i.e., input variables X and response variables Y) result simultaneously reducing to find the latent variables through projections that maximizing the covariance between input and output scores. A similar concept is addressed by the KCCA algorithm that maximizing the multi-dimensional correlation between X and Y, which performs better in terms of prediction power. Likewise, KCVA finds relations between two sets of variables considering serial correlations between the time instant to p past and f future measurements. On the other hand, KFDA is a supervised data-pretreatment algorithm that finds the transformation vectors maximizing the separability between the different sets of variables.

2.3 Application Examples

Next, several applications of kernel methods for feature extraction in electro-mechanical systems and complex chemical processes are analyzed. To facilitate the analyses on the subsequent sections, those studies where the kernel parameter selection forms part of the fault diagnosis procedure were summarized in Table 2.

2.3.1 Applications in electro-mechanical systems.

The rotating machinery is essential to transmit power and motion in the modern manufacturing industries [89]. There are several kernel approaches reported in the literature for the health condition monitoring of such devices. For example, Zhang et al. [90] propose a KPCA-SVM scheme to predict the bearing degradation process. Dong et al. [91] applied the RBF-KPCA to reduce the feature mutual correlation, while a Morlet-SVM classifier [92] is used to identify the bearing running state. On the other hand, Shao et al. [93] utilize the RBF-KPCA for the fault feature extraction on gear systems. Otherwise, Liu et al. [94] integrate KPCA and TWSVM to analyze the health conditions of bearing and bevel gear faults. Likewise, Cheng et al. [69] investigate the joint working of KPCA and Learning Vector Quantization (LVQ) Neural Network for the planetary gear diagnosis. At around the same time, Wan et al. [95] analyzes the performance of the PCA, KPCA, and Isomap algorithms considering five different gear crack. Sakthivel et al. [96] discusses the benefits of

TABLE 2
Summary of papers: Kernel method, case studies, kernel functions, and fitness criterion used.

No.	Reference.	Year	Kernelized Methods	Benchmark evaluated	Kernel function	Fitness criteria	Scientific Journal	number of citations
1	Lee et al.[59]	2004	PCA	NE, WWTP	RBF	Heuristic	Chemical Engineering Science	915
2	Choi et al.[60]	2005	PCA	NE, CSTR	RBF	Heuristic	Chemom. Intell. Lab. Syst.	419
3	Cho et al.[61]	2005	PCA	NE, CSTR	RBF	Heuristic	Chemical Engineering Science	294
4	Yoo et al.[62]	2005	PCA	NE, WWTP	RBF	Heuristic	Process Biochemistry	39
5	Jia et al.[63]	2010	PCA	NE, PenSim	RBF	Heuristic	Chemom. Intell. Lab. Syst.	101
6	Khediri et al.[64]	2011	PCA	NE, TEP	RBF	Heuristic	Computers and Industrial Engineering	63
7	Ni et al.[65]	2011	PCA	PS	RBF	Direct	IEEE Transactions on Power Delivery	95
8	Deng et al.[66]	2013	PCA	NE, TEP	RBF	Heuristic	Neurocomputing	57
9	Wang et al.[67]	2015	PCA	SEP	RBF	Heuristic	Chemom. Intell. Lab. Syst.	24
10	Yao et al.[68]	2015	PCA	PenSim	RBF	Direct	Journal of Process Control	33
11	Cheng et al.[69]	2016	PCA	RM	RBF	Direct	Measurement	57
12	Bernal et al.[34]	2016	PCA	TEP	RBF	Indirect	Chemical Engineering Science	30
13	Zhang et al.[70]	2017	PCA	NE, PenSim	RBF	Indirect	IEEE Access	23
14	Fu et al.[23]	2017	PCA	DP	RBF	Indirect	Chemom. Intell. Lab. Syst.	8
15	Deng et al.[71]	2017	PCA	NE, TEP	RBF	Heuristic	ISA Transactions	17
16	Mansouri et al.[17]	2018	PCA	BP	RBF	Heuristic	IEEE Transactions on Nanobioscience	5
17	He et al.[72]	2018	PCA	RM	RBF	Direct	Journal of Vibroengineering	1
18	Harkat et al.[18]	2019	PCA	NE, TEP	RBF	Heuristic	Chemical Engineering Science	9
19	Qian et al.[73]	2020	PCA	NE, TEP	RBF	Heuristic	Journal of Automatica Sinica	–
20	Zhou et al.[74]	2020	PCA	TEP	RBF	Heuristic	Neurocomputing	1
21	Hamrouni et al.[19]	2020	PCA	TEP	RBF	Direct	Int. J. Adv. Manuf. Technol.	–
22	Zhang et al.[75]	2007	ICA	NPP	RBF	Heuristic	Ind. Eng. Chem. Res	111
23	Fan et al. [33]	2014	ICA	TEP	RBF	Heuristic	Information Sciences	103
24	Bernal et al.[34]	2016	ICA	TEP	RBF	Indirect	Chemical Engineering Science	30
25	Zhang et al.[70]	2017	ICA	NE, PenSim	RBF	Indirect	IEEE Access	23
26	Samuel et al.[76]	2015	CVA	TEP	RBF	Heuristic	IFAC-PapersOnLine	18
27	Bai et al.[31]	2020	ECA	RM	RBF	Indirect	Applied Soft Computing	–
28	He et al. [77]	2008	FDA	TEP	RBF	Direct	Chemom. Intell. Lab. Syst.	34
29	Zhu et al. [78]	2010	FDA	TEP	RBF	Heuristic	Chem Eng. Res. Des.	68
30	Zhu et al.[79]	2011	FDA	TEP	RBF	Heuristic	Expert Systems with Applications	82
31	Rong et al. [80]	2013	FDA	WWTP, TEP	RBF	Indirect	Computers and Chemical Eng.	12
32	Liu et al. [81]	2013	FDA	RM	RBF	Direct	Int. J. Adv. Manuf. Technol	52
33	Bernal et al. [82]	2015	FDA	TEP	RBF	Direct	Computers and Industrial Eng.	41
34	Feng et al. [24]	2016	FDA	TEP	RBF	Heuristic	IEEE Trans. Autom. Sci. Eng.	31
35	Deng et al. [26]	2016	FDA	NE, CSTR	RBF	Heuristic	Chemom. Intell. Lab. Syst.	30
36	Peng et al. [83]	2013	PLS	HSMP	RBF	Indirect	Control Engineering Practice	74
37	Zhang et al. [84]	2013	PLS	NE, TEP, HSMP	RBF	Heuristic	Math. Probl. Eng.	50
38	Godoy et al. [40]	2014	PLS	NE	RBF	Heuristic	Chemom. Intell. Lab. Syst.	25
39	Sheng et al.[20]	2015	PLS	TEP	RBF	Heuristic	IEEE Trans. Autom. Sci. Eng.	22
40	Jiao et al. [22]	2017	PLS	NE, TEP	RBF	Indirect	ISA Transactions	31
41	Fu et al. [23]	2017	PLS	DP	RBF	Indirect	Chemom. Intell. Lab. Syst.	8
42	Shi et al.[85]	2009	PCA	RM	RBFpoly	Direct	IFAC Proceedings Volumes	4
43	Jia et al.[86]	2012	PCA	NE, PenSim	RBF,Poly, Sigmoid	Indirect	Computers & Chemical Engineering	48
44	Van et al. [87]	2015	FDA	RM	RBF, Morlet	Indirect	IEEE Trans Instrum Meas.	37
45	Shi et al. [88]	2016	FDA	TEP	RBFpoly	Indirect	Int. J. Syst. Sci.	11
46	Pilario et al.[36]	2019	CVA	NE, CSTR	RBFpoly	Indirect	Computers & Chemical Engineering	11

NPP, Nosiheptide production process; **CSTR**, Continuous stirred-tank reactor; **WWTP**, Wastewater treatment plant; **PenSim**, Penicillin fermentation process; **NE**, Numerical example; **TEP**, Tennessee Eastman process; **HSMP**, Hot strip mill process; **RM**, Rotating machinery; **DP**, Distillation process; **SEP**, Industrial semiconductor etch process; **PS**, Power systems

the KPCA to reveal the health conditions of several pump components, including the bearing and the impeller of the centrifugal pump. Besides, Wong et al. [97] evaluate the KPCA-SVM (i.e., linear, RBF, and polynomial kernel) to detect gearbox faults for gas turbine generator systems. From a different perspective, Jiang et al. [98] studies the bearing degradation process using KICA and several kernel functions (i.e., Hermite, RBF, and polynomial kernel). The advantages of KICA and LS-SVM are further combining by Ma et al. [99] to recognize several faults in bearing elements. To reveal the operation state of the gearbox, Li et al. [100] utilized a Fuzzy k-Nearest Neighbor (FKNN) having as inputs the fault features extracted by a KICA processing stage. After this, Li et al. [101] suggest the faults detection of gears though KICA and BP Neural Network. Further, Liu et al. [81] utilize a hybrid kernel dimension reduction through RBF-KFDA for the fault level diagnosis of planetary gearboxes. Meanwhile, Jiang et al. [102] propose the Semi-supervised Kernel Marginal Fisher Analysis (SSKMFA) to extracts the low-dimensional characteristics from the raw high dimensional vibration signals in the gearbox. At the same work, the classifiers k-NN and RBF-SVM are also utilizing to evaluate the performances of five feature extraction methods (i.e., SAEMD, KPCA, KFDA, KMFA, and SSKMFA). Somewhat later, He et al. [72] investigates the feature extraction and fault gearbox recognition when the KPCA is regularizing through the WPSO-FDA strategy. More recently, Bai et al. [31] employed a KECA-SVM scheme to recognize the state information of rolling bearing and gear, while a whale optimization algorithm is using to find the kernel parameters.

2.3.2 Applications in complex chemical processes

The kernel methods have been widely investigated for fault diagnosis in complex chemical plants, which including continuous and batch processes. For example, dynamic extensions of kernel PCA [103–105], kernel PLS [106–108], and kernel ICA [33, 109] have been proposed to minimize the computation time and enhance fault detection performances. Kernelized algorithms have also exhibited a tremendous potential to handle the high nonlinearity [20, 22, 23, 40, 59, 60, 62, 82], the non-gaussianity [34, 75, 110–112], the multiple operation modes [113–115], and the inherently time-varying dynamics [70, 116, 117] that are typical on the chemical processes. Likewise, the kernel learning approaches have proven to be useful to detect incipient faults [34, 36, 76, 118–121], even at dynamically varying process operating conditions.

Nevertheless, the related literature suggests integrating the kernel solutions to ensure more efficient operations on the critical systems. In this context, Vitale et al. [122] deal with the non-linear data structures of the chemical plants by using a kernel PLS with a pseudo-sample projection strategy. Meanwhile to aid the monitoring process, Zhang et al. [123] develop the kernel concurrent projection to latent structure (KCPLS) method aiming to reveal the more fault-relevant directions. On the other hand, He et al [77] and Li et al [124] propose two modified versions of KFDA to extracting discriminative features in the fault monitoring tasks. For more reliable quality monitoring, Jia et al. [63] integrate the KPCA and ARMAX models to characterize the batch processes. Somewhat later, Jiao et al. [22] discuss the nonlinear quality-related faults detection when KPLS is used to build the linear relationship between kernel and output matrices. Moreover, Zhu et al. [79] evaluate two discriminant approaches using KFDA-GMM and KFDA-kNN to perform the fault classification process. Beyond that, Khediri et al. [64] introduce an adaptive KPCA scheme to facilitate the recursive calculation of chart control limits between different batches. Soon later, Deng et al. [66] propose a novel similarity factor using KPCA to reveal the statistics information hidden in original measured variables. Alternatively, Yao et al. [68] explores the concept of generalized additive KPCA for the batch monitoring processes.

On the other hand, Wang et al. [67] utilize the functional KPCA to handle nonlinear correlations between monitoring variables and/or sampling times. From a data-driven perspective, Feng et al. [24] take advantage of KFDA with local and global manifold to removes outliers caused by disturbances. With the above idea in mind, Hu et al. [125] suggests the spherical KPLS in order to handle data sets contaminated with outliers, while Deng et al. [71] uses KPCA with double-weighted local outlier factor (LOF). Furthermore, Zhu et al. [78] addressed the problem of imbalanced data through a KFDA-based fault diagnosis scheme. Besides, Mansouri et al. [17] propose a multiscale kernel generalized likelihood ratio test (MS-KGLRT) to enhance the fault monitoring abilities of nonlinear biological plants. Given the above analysis, Harkat et al. [18] and Hamrouni et al. [19] investigate the KPCA approach using a nonlinear interval of data values to address the problem of uncertainties in these systems. To enhance the fault detection performance, Deng et al. [26] considers a Bayesian perspective to built the fault monitoring statistics from the kernel feature space. More recently, Zhou et al. [74] make full use of low-order and high-order statistics of the process to preserves both local structure and global structure in the kernel feature space improving the fault monitoring of chemical processes.

3 CRITERIA TO CHOICE KERNEL PARAMETERS

As already mentioned, the present work grouped the procedures for tuning the kernel parameters in three categories. In the first group are the heuristical selection procedures. A criterion widely used in this context is $\theta = cm\sigma^2$, where m and σ^2 are related to the training data set, and c is an empirical constant. Table 2 shows several published papers wherein different variants of the above rule utilized. In second place were considered the direct evaluation criteria, which take advantage of the data spatial structure to find the optimal kernel parameters. As illustrated in Figure 1, the direct evaluation criteria are independent of the fault diagnosis scheme used. Therefore, they have low computational complexity. Besides, from a theoretical view of point, the kernel functions tuned by the direct evaluation criteria could be successfully reusable for other fault monitoring schemes without repeat the parameter optimization procedure.

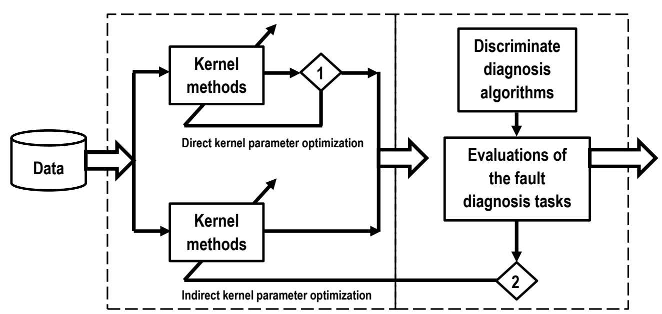

Fig. 1. Kernel evaluation approaches using for the selection of the optimal parameters. (1) A direct conception to built the objective function, (2) An indirect conception to built the objective function.

In contrast to this, the indirect evaluation criteria choose the kernel parameters considering the performance of fault diagnosis tasks. In general, higher fault diagnosis performances are achieving through this approach, but with more effort and computational costly during the training phase of algorithms. Furthermore, both direct and indirect evaluation approaches can be indistinctly employed with either optimization algorithms available in the literature.

3.1 Direct evaluation criteria

The direct kernel optimization approaches implement the diagnosis tasks as supervised learning problems, which depend on the learned spatial structure by the kernel gram matrix computed over the training data. Since the observations of each class can consider as one cluster, a kernel induced transformation will be appropriate if the separability of the data-classes in the inducing feature space is maximum. From this view of point, different inter-cluster distances have been using to choose the kernel parameters, but mostly restricted to gaussian kernels. For example, Teixeira et al. [126] choose the tuned parameter for the RBF kernel by the maximum distance of the samples from the mean vector of the class. Meanwhile, Ni et al. [65] consider the sum of the difference between the maximum and minimum for each variable in the data matrix. In the work reported by Zhang et al. [127], the median value of the reciprocal distances between each sample and the sample mean is utilizing as parameter selection criteria. More recently, Cheng et al. [69] employe the total distance among the sample centers of different fault gears in the feature space. On the other hand, Hamrouni et al. [19] suggest the average minimum distance between two points in the training data. According to Wu et al. [128], the following inter-cluster distance measures can be also used to find the kernel parameters.

1- The minimum inter-cluster distance:

$$\min J_f(\tilde{\theta}_f) = \|\phi(x_+) - \phi(x_-)\| \tag{5}$$

2- The maximum inter-cluster distance:

$$\max J_f(\tilde{\theta}_f) = \|\phi(x_+) - \phi(x_-)\| \tag{6}$$

3- The average inter-cluster distance:

$$\max\ J_f(\tilde{\theta}_f) = \frac{1}{N_+ N_-} \sum_{x_+ \in X_+} \sum_{x_- \in X_-} \|\phi(x_+) - \phi(x_-)\| \tag{7}$$

4- The distance between means of the two clusters:

$$\max J_f(\tilde{\theta}_f) = \|\mu_+ - \mu_-\| \tag{8}$$

5- The distance between the points of one cluster from the mean of the other cluster:

$$\max J_f(\tilde{\theta}_f) = \frac{\sum_{x_+ \in X_+} \|\phi(x_+) - \mu_-\| + \sum_{x_- \in X_-} \|\phi(x_-) - \mu_+\|}{N_+ + N_-} \tag{9}$$

where N_+ and N_- represents the number of data points in the positive class X_+ and the negative class X_+ respectively. The values $x_+ \in X_+$ and $x_- \in X_-$ are the observations with means μ_+ and μ_- in the kernel induced feature space.

Perhaps the inter-cluster distance more widely recognized in fault diagnosis applications [72, 82, 85, 129] is the Fisher measure based on the Rayleigh quotient [130]. Through the implicit information on the kernel matrix, the Fisher measure finds the projection ω that maximizes the separability of different classes in the empirical feature space, while it minimizes their spatial dispersion.

$$\min J_f(\tilde{\theta}_f) = \frac{\omega^T S_b \omega}{\omega^T S_w \omega} \tag{10}$$

$$S_b = (\mu_+ - \mu_-)(\mu_+ - \mu_-)^T \tag{11}$$

$$S_w = \sum_{x_+ \in X_+} (x_+ - \mu_+)(x_+ - \mu_+)^T + \sum_{x_- \in X_-} (x_- - \mu_-)(x_- - \mu_-)^T \tag{12}$$

where S_b and S_w represent the between-class and total within-class scatter matrices, respectively. This measure is optimal only for homoscedastic data distributions because all classes are assumed from a normal distribution and identical covariance matrices [131].

Among the direct evaluation criteria, there are also available kernel matrix-based approaches, where the class label information is employed to indicate how an ideal kernel matrix should be for discriminant tasks [77, 82]. For example, Kernel target alignment (KTA) is a measure commonly used to compare two positive semidefinite kernel matrices [132, 133]. The KTA is defining as the normalized Frobenius inner product between a candidate kernel matrix and their ideal kernel. It is interpreted as the cosine distance between two vectors [134], taking values in $[-1; 1]$. Though high KTA values amount to the kernel being well aligned to the class distribution of the ideal kernel, the opposite is not always true. For two kernel matrix, the objective function can be formulated as follow.

$$\max J_f(\tilde{\theta}_f) = \frac{\langle K_1, K_{ideal} \rangle_F}{\sqrt{\langle K_1, K_1 \rangle_F \langle K_{ideal}, K_{ideal} \rangle_F}} \tag{13}$$

Based on KTA, other kernel evaluation measures have also been derived recently, such as kernel polarization [135, 136], and feature space-based kernel matrix evaluation measure (FSM) [134]. Likewise, there are other measures less known as the α, β, and γ metrics proposed by [137]. In specific, $\alpha-$measure employ the Pearson correlation coefficient obtained for the triangular kernel matrix candidate and the ideal kernel matrix, both previously normalized. As a result, it is less computationally intensive than KTA. The objective function based on the $\alpha-$measure can be computed as follow.

$$\max J_f(\tilde{\theta}_f) = \frac{\sum_{i=1}^{m} \sum_{j=i+1}^{m} (K_{ij} - \overline{K})(S_{ij} - \overline{S})}{\sqrt{\sum_{i=1}^{m} \sum_{j=i+1}^{m} (K_{ij} - \overline{K})^2} \sqrt{\sum_{i=1}^{m} \sum_{j=i+1}^{m} (S_{ij} - \overline{S})^2}} \tag{14}$$

$$S_{ij} = \begin{cases} 1, & \text{if } y_i = y_j \\ 0, & \text{if } y_i \neq y_j \end{cases} \tag{15}$$

where the mean values of K and S are calculated over a triangular matrix without main diagonals.

$$\overline{K} = \frac{2}{m\,(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} K_{ij} \tag{16}$$

$$\overline{S} = \frac{2}{m\,(m-1)} \sum_{i=1}^{m} \sum_{j=i+1}^{m} S_{ij} \tag{17}$$

On the other hand, $\beta-$measure based on the $t-$student statistical test provides an evaluation for the equality of two mean vectors to determine the difference between the elements of the partitioned kernel matrix [82, 137]. For different operating conditions the $\beta-$measure may be used to determine the kernel parameter maximizing the separability between classes.

$$\max J_f(\tilde{\theta}_f) = \left| \frac{\bar{k}_w - \bar{k}_b}{\sqrt{\frac{\sigma_w^2}{m_w} + \frac{\sigma_b^2}{m_b}}} \right| \tag{18}$$

$$k_b = \{K_{ij} | i < j \wedge y_i \neq y_j\} \tag{19}$$

$$k_w = \{K_{ij} | i < j \wedge y_i = y_j\} \tag{20}$$

where $\bar{k}_x$ and σ_x^2 are the mean and variance of respective vectors k_w and k_b, with m_x observations.

Beyond that, $\gamma-$measure use the Brown$-$Forsythe test to evaluate the equality between the variance in the kernel block matrix, but without the normality assumption. Results shown in [137] suggest that the $\gamma-$measure could have good results independently of whether classes are balanced or not. The objective function based on this measure can be computed as follows.

$$\min J_f(\tilde{\theta}_f) = (m-2)\frac{m_w(\bar{Z}_w - \bar{Z}_{wb})^2 + m_b(\bar{Z}_b - \bar{Z}_{wb})^2}{\sum_{i=1}^{m_w}(\bar{Z}_i - \bar{Z}_w)^2 + \sum_{i=1}^{m_b}(\bar{Z}_i - \bar{Z}_b)^2} \tag{21}$$

then considering the Eqs (19-20), and $k_{wb} = \{k_w \cup k_b\}$ with $m_{wb} = m_w + m_b$ as the number values in k_{wb}, it can be obtained that for each of them $Z_x = kx - \bar{k}x$. In this case, the mean values $\bar{Z}_w$, $\bar{Z}_b$ and $\bar{Z}_{wb}$ are calculated using the following transformed variables.

$$\bar{Z}_x = \frac{1}{m_x}\sum_{i=1}^{m_x} Z_{ix} \tag{22}$$

3.2 Indirect evaluation criteria

These strategies choose the kernel parameters that ensure a high monitoring performance according to some objective function. For example, Jia et al. [86] relates several monitoring indicators using three kernel functions (i.e., RBF, Poly, and Sigmoid) and the following fitness function.

$$\min J_f(\tilde{\theta}) = \eta(\eta_A + \text{SPE}_{lim} \times d_{\eta_A}) \times d_\eta \tag{23}$$

where η is the correct monitoring rate with resolution d_η, and SPE_{lim} is statistical control limit of SPE. Meanwhile, η_A is referred to as the inverse of the retained component numbers of KPCA with a resolution d_{η_A}. In this case, the correct monitoring rate is calculated as:

$$\eta = \frac{\eta_{r,n}}{\eta_{t,f}} + \frac{\eta_{r,f}}{\eta_{t,f}} \tag{24}$$

where $\eta_{t,n}$ is the normal total number of samples by testing; $\eta_{r,n}$ is the normal total number of samples by correct testing; $\eta_{t,f}$ is the fault total number of samples by testing; $\eta_{n,f}$ is the fault total number of samples by correct testing. A similar approach is addressed in Shi et al. [88], which considers the kernel parameter selection by minimizing the ratio between the number of erroneous diagnoses, and the number of total training data. On the other hand, Zhang et al. [138] consider a combination of the false alarm rate dr and detection rate fr using the following fitness function.

$$\max\ J_f(\tilde{\theta}) = \sum_{i=1}^{2} \nu_i(dr_i - fr_i) \tag{25}$$

where ν_i is the weighting factor for the two monitoring indices T^2 and SPE, respectively. Another fitness function is reported by Bai et al. [31] for the feature selection in rotating machinery using the KECA approach.

$$\max\ J_f(\tilde{\theta}) = E(1)/\sum_{i=1}^{N} E(i) \tag{26}$$

where, E denotes the kernel entropy score of each kernel entropy component. Alternatively, Rong et al. [80] attempt to choose the kernel parameters achieving the lowest misclassification rate by three-fold cross-validation, such as:

$$\min J_f(\tilde{\theta}_f) = 2^{2+\tau/2} d\sigma^2 \tag{27}$$

where $\tau = \{0, 1, \ldots, 30\}$, d represent the dimension of the input space, and σ^2 is the variance of the training data set taking a unitary value after normalization process. On the other hand, Fu et al. [23] propose a framework to estimate the parameter corresponding with the RBF-KPCA and RBF-KPLS models using the following optimization function.

$$\min J_f(\tilde{\theta}_f) = \frac{1}{\text{m}_z\text{p}}\sum_{i=1}^{m} \varepsilon^{(i)^T}(\tilde{\theta}_f)\varepsilon^{(i)}(\tilde{\theta}_f) \tag{28}$$

where the prediction z corresponding to the data set $\{\mathbf{X}_\text{i}\}_{\text{i}=1}^{\text{m}} \in \mathbf{R}^\text{p}$ is defining by the inverse mapping of the kernel model, and the prediction error is $\varepsilon = (z - \hat{z})$. More recently, Bernal et al. [34] assumed the false alarm rate and missing alarm rate as two measures that can be unifying through the AUC index to select the kernel parameter that best distinguishes normal from faulty process conditions. Then, the following fitness function is formulated.

$$\min J_f(\tilde{\theta}_f) = 1 - \frac{1}{c}\sum_{i=1}^{c} \mathbf{AUC}_i(\tilde{\theta}_f) \tag{29}$$

where there are c mutually exclusive classes representing the fault operation conditions of the process. Beyond that, Pilario et al. [36] raised that a priori fault information could not be available for complex processes. Thereby, the normal operation data sets could be divided into two data sets (i.e., SET1: training data and SET2: validation data) to evaluate the monitoring schemes. In this case, Pilario et al. [36] employed the grid search method to decide the number of states and RBFpoly kernel parameters used for an MK-CVDA model, such that the false alarms in the validation data set are minimized.

The following optimization criterion is also a proposal of the present paper and will be used to show the experimental results later. This updated index combines the ideas suggest by [34, 36] both applied for the incipient fault monitoring.

$$\min_{\tilde{\theta}_f} J_f(\tilde{\theta}_f) = \text{FAR}(\tilde{\theta}_f|\text{fault-free}) + \left(1 - \frac{1}{c}\sum_{i=1}^{c} \mathbf{AUC}_i(\tilde{\theta}_f|\text{fault})\right) \tag{30}$$

where FAR($\tilde{\theta}_f$|fault-free) represent the performance of the monitoring scheme during the normal operating conditions and c are the monitoring fault conditions.

3.3 Kernel evaluation framework

The selection and adjusting of the kernel functions for fault diagnosis applications can be both highly dependent on the problem at hand. However, there are some general guidelines and procedures that might contribute to achieving better use of kernel-based methods in this field. Li et al. [139] provide a framework for kernel-based learning that is extending to fault diagnosis applications in the present work. Figure 2 depicts the modified framework as four consecutive phases: (i) Kernel selection, (ii) Parameters optimization, (iii) Training of methods, and (iv) Validation to fault diagnosis. As already emphasized, the kernel function candidates could be select by the researchers' experience/experimental studies. Depending on what kind of information it is expecting to extract in the fault diagnosis tasks should decide the best approach and objective function to choose the kernel parameters.

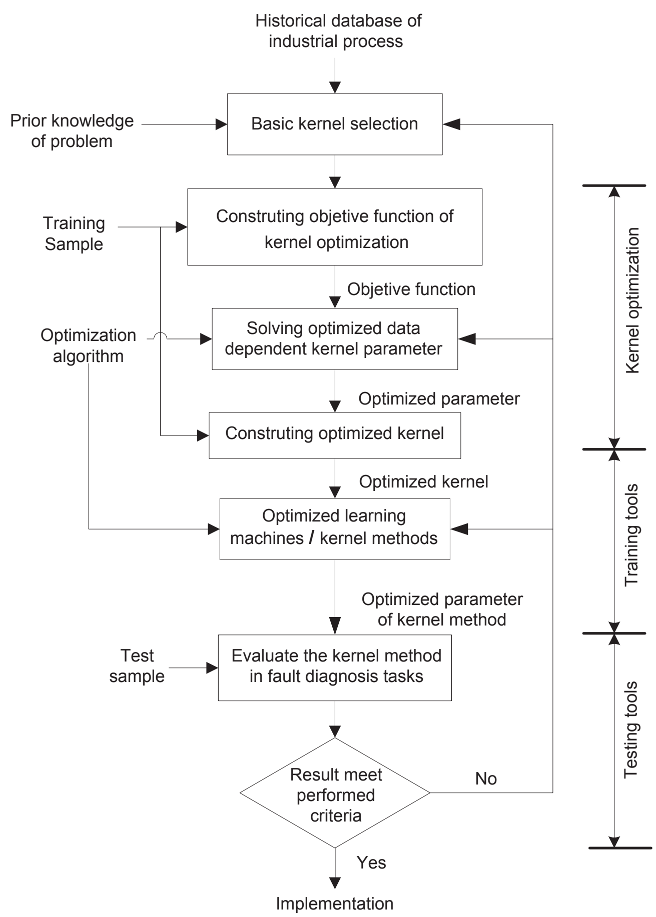

Fig. 2. Framework for kernel-based learning in faults diagnosis tasks.

4 Experimental studies and results

As has already mentioned, the main goal of this work is to provide a general overview of kernel-based preprocessing methods, giving particular emphasis to the adjusting of kernel solutions. However, several kernel functions were selected to show experimental results using two different optimization criteria.

4.1 Qualitative analysis of kernel functions

Although many studies have recognized the key role that has the kernel functions, there remains a systematic use of the kernel RBF for fault diagnosis applications without exploring other alternative functions. Next, different kernel functions are graphically exploring with the intention to find common rules that can help with their selection. In this context, the synthetic data set shown in Figure 3 is used to obtain each kernel matrices.

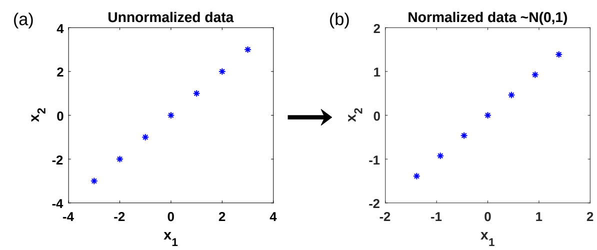

Fig. 3. Free-fault data set used for the qualitative study of kernel functions. (a) Synthetic data, (b) Standardized synthetic data

The following assumptions are considered throughout the exploratory study to conduct this from a fault diagnosis perspective. As can see from Figure 3, the first requirement is that the fault monitoring procedures begin with the data standardization by using the mean and covariance describing the normal operating condition. In this context, the centering and scaling of the kernel matrices are also assuming as necessary operations for the kernel-based fault monitoring. Let the kernel matrix $\mathbf{K}_{ij}$; $\forall i,j = \{1,\ldots,m\}$, resulting to evaluate the kernel functions, the centered kernel matrix is obtained as follows:

$$\mathbf{K}_c = \mathbf{K} - 1_m\mathbf{K} - \mathbf{K}1_m + 1_m\mathbf{K}1_m \tag{31}$$

where $\mathbf{1_m} \in \mathbf{R}^{m\times m}$ is a matrix in which each element is equal to $1/m$. Under the conditions stated above, Figure 4(a-b) compares the kernel matrices $\mathbf{K}(\mathbf{x}_j,\mathbf{x}_j)$ before and after of carry out the operations of centering and scaling. Note that all kernel function in Figure 4(a-b) are tuned with the same kernel parameter $\theta = 2$. For the additive kernel k13, which combining the polynomial and RBF kernel functions, were choose the parameters $\omega = 0.5$, $d = 1$, $\alpha = 1$, and $\theta = 2$. It is apparent that the gaussian kernel family k1, k3, k5, and k6 provides values closer to one to indicate a high degree of similarity between the data, and zero for point out the dissimilarities. The above kernels can be easily adjusting through direct evaluation criteria by using the class label information. However, this is less evident to the kernel functions k8, k10, k11, 13, and k14. However, given the representation of kernels k5 and k6 into a more reduced interval, for two different data set could be difficult to distinguish their features. Another unanticipated finding on this exploratory study was that kernel k2 and k3 shown identical results. The

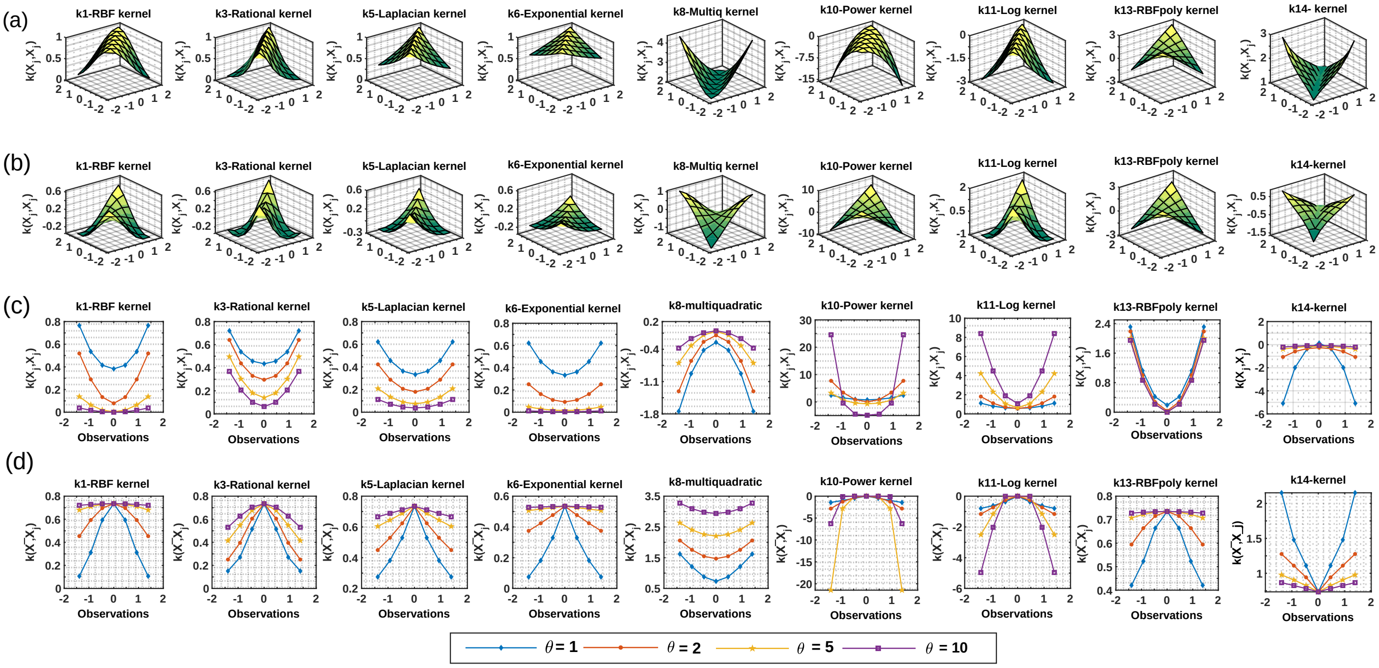

Fig. 4. A qualitative comparison. (a) Kernel matrices $\mathbf{K}(x_j, x_j)$ obtained with parameter $\theta = 2$; (b) Kernel matrices $\mathbf{K}(x_j, x_j)$ adjusting with parameter $\theta = 2$, after the operations of centering and scaling; (c) Values of the main diagonal for kernel matrices $\mathbf{K}(x_j, x_j)$ after the operations of centering and scaling, adjusting with different parameters $\theta = 1, 2, 5, 10$; (d) Kernel vectors from $\mathbf{K}(\bar{x}_j, x_j)$ after the operations of centering and scaling, adjusting with different parameters $\theta = 1, 2, 5, 10$.

same reasoning applies to the case of kernel k14 and k16. Figure 4(c) compares the variation on the main diagonal values from the above kernel functions, but with different tuning parameters. The parabola shown in the figure gives an idea about the kernel parameter sensibility, and how much the kernel emphasizes the differences between the input data. Note that the minimal variation of the kernel k13 due to their parameters. Figure 4(d) illustrates the kernel vector $\mathbf{K}(\bar{\mathbf{x}}, \mathbf{x}_j)$ obtained to compare the mean of the data set with each observation. As in the previous case, a parabola with less dilatation represents data more closely to the center of the cluster. The analysis of similar graphical could help decide if using the approaches based on the inter-cluster distances to choose the kernel parameters.

4.2 Fault detection in study cases

The following describes the kernel-based fault monitoring scheme utilized. Figure 5 illustrates the flowchart adopted to evaluate the kernel functions summarized in Table 1.

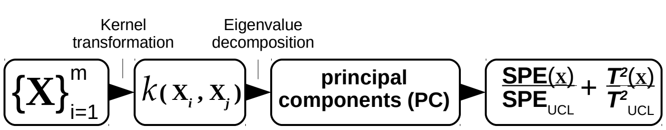

Fig. 5. Flowchart for KPCA-based fault monitoring scheme.

In this work, SPE and Hotelling's T^2 statistics were combining to derive the detectability conditions in the systematic and residual subspaces from the KPCA model. More details about the combination of these indices are given in [60, 140, 141]. Moreover, the criterion $\lambda_i/sum(\lambda_i) > 10^{-4}$ was employed to select the number of eigenvalues [142]. Upper control limits (UCL) with 99% significance were obtaining for these statistics through kernel density estimation (KDE) [36, 143]. The fault detection procedures were evaluated by using several performance indicators: detection delay (DD), false alarm rate (FAR), and missed detection rate (MDR). The detection delay also referred to as latency time, reflects the elapsed time before that faults are continuously detected. In this study, the latency is the time for which five consecutive alarms are detecting from the start of the simulated faults.

As earlier stated, FAR and MDR indicators were also computed by using the Eqs (32) and (33). However, the Area under the curve (AUC) was also used to give a joint interpretation of the two above indicators [34].

$$\mathrm{FAR} = \frac{\text{No. of samples } (J > J_{\mathrm{UCL}} | \text{fault} - \text{free})}{\text{total samples } (\text{fault} - \text{free})} \tag{32}$$

$$\mathrm{MDR} = \frac{\text{No. of samples } (J > J_{\mathrm{UCL}} | \text{fault})}{\text{total samples } (\text{fault})} \tag{33}$$

On the other hand, the Bayesian optimization algorithm available in the mathematical assistant Matlab®R2018 was utilized to chosen all kernel parameters. This approach is a particular case of the nonlinear optimization where the algorithm decides which point to explore next based on the analysis of distribution over functions [144, 145]. More specifically, the Bayesian optimization with an expected improvement search (i.e., expected-improvement-plus) was employed to escape from local minimums, considering the estimated variables in the range $[10^{-5}, 10^{5}]$ and sixty evaluations of the objective function. In this study, the fitness functions used to choose the kernel parameters were building from the indirect kernel evaluation approach with Eq. (30), and the direct kernel evaluation approach with Eq. (10). The tuning kernel parameters for the above criteria are provided in Table 6 at the Appendix of the present work. Finally, it is worth noting that other optimization algorithms could also apply to find the optimal kernel parameters.

4.2.1 Simulated nonlinear system

The first case of study is a modified nonlinear system that was suggested by Dong and McAvoy (1996) [146], and it is also employing in references [36, 59–61]. In this example, the goal is to distinguish the fault condition from the normal data as early as possible [36]. The system consists of three variables nonlinearly related, and two outputs y_1 and y_2. All measurements have also incorporated a white Gaussian noise, such that $e_i \sim N(0, 0.001)$ and $t \in [0.01, 2]$.

$$\begin{aligned} x_1 &= t + e_1, \\ x_2 &= t^2 - 3t + e_2, \\ x_3 &= -t^3 + 3t^2 + e_3, \\ y_1 &= -x_2(1 + x_1), \\ y_2 &= x_3 - sin(1.5x_2\pi) \end{aligned} \tag{34}$$

The training and test data sets, with 30 samples each one, were obtained according to these equations. The test data set was generated by a slow linear drift fault in y_2 at the 10th sample time. Furthermore, different data sets were considered for the validation stage. A graphical representation of the training (blue) and test (red) data samples is shows in Figure 6 to distinguish between the fault and the normal operating operations of the process.

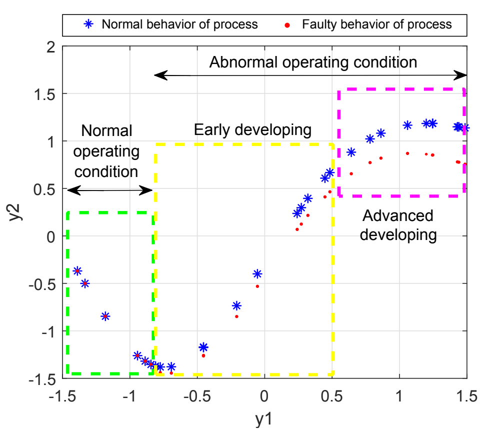

Fig. 6. Historical data set from the numerical example.

The fault detection performance for the validation data sets is summarized through box plots in Figure 7. In this case, the AUC index also provides integrating information of the FAR and MDR measures [34]. As shown in Figures 6 and 7, three possible zones marked in colors were considered to evaluate the fault detection results: normal operation (blue), early fault developing (red), and advanced fault developing (orange). A fourth marked zone (white) in Figure 7 shows the fault rates that were undetected from the evaluation of 10 validation data sets. An early fault detection ability was related to the area marked in red color. The dashed line indicates those kernel functions where the abnormal operating conditions are not detected. As can be seen from Figure 7(b), with the direct evaluation approach, the kernel functions k4, k12, and k15 fail to detect the abnormal operating conditions in the validation data sets. The same applies to the kernel functions k5 and k6, regardless of the evaluation approach used. Therefore, the above kernel functions were not to be considered for subsequent studies.

As shown in Figure 7, a smaller amount of false alarms is always achieving for all kernels. Except for the kernel function k10, Figure 7 shows an early fault detection for the kernel functions adjusting by the indirect evaluation approach. For the kernels k7, k8, k11, k13, k14, and k16, Figure 7(b) shown similar results in terms of detection times. In this case, the kernel functions were adjusting by the direct evaluation approach. Note there are three additive kernel functions with excellent results, whatever the evaluation approach adopted: k13, k14, and k16. Other interesting findings were showed in Figure 7(a), where the kernel functions k14 and k16 have identical fault monitoring performances. However, the kernel function k14 has a better fault monitoring performance than the kernel function k16, when the direct evaluation approach with Eq. (10) is using to adjust them. From a comparative view of point, kernel functions with the worst fault monitoring indicators have lower AUC values. Otherwise, kernel functions with higher fault monitoring indicators (i.e., smaller FAR, MDR, and latency) have AUC values over 0.9. In both evaluation approaches, the better fault detection indicators are achieving with the kernel functions k8, k11, k14, and k16. For a further robustness study of these four kernel functions, the variability of the validation data sets was increased by 5% and 10%, about the previous conditions. As illustrated in Figure 8, there is a progressive deterioration of the fault monitoring performance due to the increasing variability in the validation data sets. However, in terms of fault detection delay and AUC values, the kernel function k8 achieved a more robust performance than the other kernel functions by using the direct evaluation approach. Moreover, the kernel k14 and k16 tuned by the indirect evaluation approach, still have identical fault monitoring performances.

4.2.2 TE benchmark process

The Tennessee Eastman (TEP) consists of five units interconnected: a reactor, a vapor-liquid separator, a product condenser, a recycle compressor, and a product stripper. The process contains 21 preprogrammed faults and one normal operating condition data set. For each fault data set, it includes 960 samples where the faults are introducing at the 161st sample. The historical data set contains 33 variables available online, with a sampling time of 3 min and Gaussian noise incorporated in all measurements. In this study, the TEP benchmark is using with two purposes: (1) Evaluate the kernel functions using a realistic chemical system, (2) Evaluate the detection of novel faults through the kernel-based monitoring schemes. To accomplish these goals were considered six faults of the TEP, which are

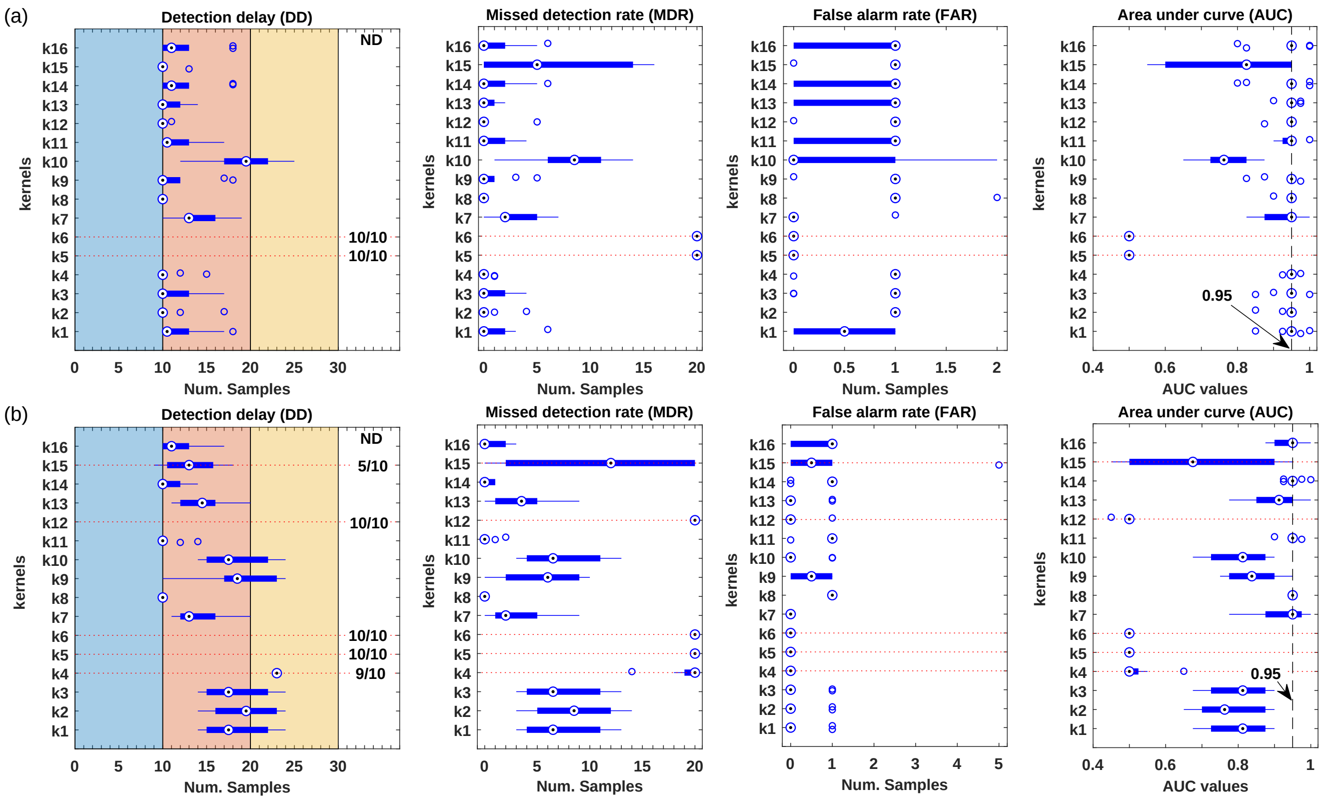

Fig. 7. Fault detection results obtained for the numerical example and kernel functions in Table 1. (a) Fault monitoring indicators for the indirect kernel evaluation approach with Eq. (30), (b) Fault monitoring indicators for the direct kernel evaluation approach with Eq. (10)

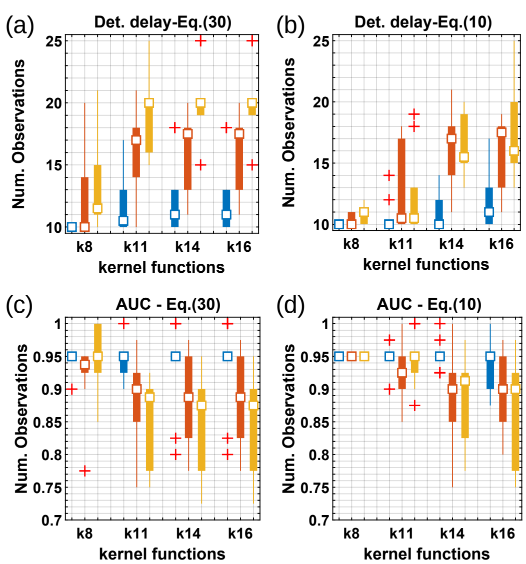

Fig. 8. Analysis of robustness. (a-c) Detection time and AUC index for the indirect evaluation approach with Eq. (30), (b-d) Detection time and AUC index for the direct evaluation approach with Eq. (10).

TABLE 3
Description of faults in TE process.

Fault	Process variable	Type	Status
F1(1)	A/C feed ratio, B composition constant	step	Learned
F2(8)	A, B, and C feed composition	Random variation	Learned
F3(14)	Reactor cooling water valve	Sticking	Learned
F4(6)	A feed loss	step	Unlearned
F5(12)	Condenser cooling water inlet temp.	Random variation	Unlearned
F6(15)	Condenser cooling water valve	Sticking	Unlearned

describing in Table 3. Besides, the monitoring procedures were training to recognize the data pattern of the first three faults (i.e., F1, F2, and F3). Thereupon, the data pattern corresponding to the other three abnormal operating conditions (i.e., F4, F5, and F6 that were not learned by the monitoring algorithms) are employed to validate the detection of novel faults.

Due to their previous results, the kernel functions k1, k3, k7, k9, k10, k13, k14, and k16 are also employing for the study with the TEP process. The kernel functions were further tuning through the indirect Eq. (30) and direct evaluation Eq. (10) criteria, respectively. In this context, Tables 4 and 5 provides a detailed description of the fault detection delay and values of the AUC(%) obtained. Note that the kernel functions with an acceptable performance are also indicating through a checkmark. As shown in Table 4, faults F1 and F3 are always detected effectively with the kernel-based monitoring scheme. However, the high detection delay for the fault F2 indicates a higher difficulty to recognize a random variation in the composition of

TABLE 4
Fault detection performance for known abnormal operations.

		Fobj1				Fobj2			
		F1	**F2**	**F3**	**S**	**F1**	**F2**	**F3**	**S**
k1	DD	**163.0**	**178.0**	**161.0**	✓	168.0	187.0	164.0	✗
	AUC	**99.25**	**98.94**	**100.0**		97.00	93.75	91.31	
k3	DD	163.0	180.0	161.0	✓	**163.0**	**180.0**	**161.0**	✓
	AUC	99.88	98.94	100.0		**99.88**	**98.94**	**100.0**	
k7	DD	168.0	189.0	162.0	✗	163.0	180.0	161.0	✓
	AUC	99.56	97.50	99.56		99.56	98.25	100.0	
k9	DD	166.0	183.0	162.0	✗	163.0	180.0	161.0	✓
	AUC	99.38	98.44	99.94		99.56	98.25	100.0	
k10	DD	166.0	181.0	161.0	✗	163.0	180.0	161.0	✓
	AUC	99.38	97.44	99.38		99.56	98.25	100.0	
k13	DD	167.0	186.0	162.0	✗	163.0	180.0	161.0	✓
	AUC	99.31	97.5	99.94		99.56	98.25	100.0	
k14	DD	163.0	180.0	161.0	✓	163.0	180.0	161.0	✓
	AUC	99.88	98.25	99.69		99.88	98.60	99.77	
k16	DD	163.0	180.0	161.0	✓	163.0	180.0	161.0	✓
	AUC	99.88	98.25	99.69		99.88	98.25	99.69	

Bold values indicate the best performance.

TABLE 5
Fault detection performance for novel abnormal operations.

		Fobj1				Fobj2			
		F4	**F5**	**F6**	**S**	**F4**	**F5**	**F6**	**S**
k1	DD	161.0	163.0	ND	✗	166.0	184.0	837.0	✓
	AUC	99.69	99.44	51.00		99.69	97.19	52.00	
k3	DD	161.0	163.0	836.0	✗	**161.0**	**163.0**	**790.0**	✓
	AUC	100.0	98.75	53.00		**100.0**	**98.44**	**54.00**	
k7	DD	167.0	182.0	836.0	✓	161.0	163.0	ND	✗
	AUC	99.63	98.13	51.00		99.69	99.88	51.00	
k9	DD	161.0	182.0	ND	✗	161.0	163.0	ND	✗
	AUC	99.69	98.81	51.00		99.69	99.88	51.00	
k10	DD	161.0	167.0	737.0	✓	161.0	163.0	ND	✗
	AUC	97.69	92.50	58.13		99.69	99.88	51.00	
k13	DD	165.0	176.0	787.0	✓	161.0	163.0	ND	✗
	AUC	99.75	95.81	56.00		99.69	99.69	51.00	
k14	DD	**161.0**	**167.0**	**791.0**	✓	**161.0**	**167.0**	**791.0**	✓
	AUC	**99.69**	**97.63**	**54.00**		**99.69**	**97.63**	**54.00**	
k16	DD	**161.0**	**167.0**	**791.0**	✓	**161.0**	**167.0**	**791.0**	✓
	AUC	**99.69**	**97.63**	**54.00**		**99.69**	**97.63**	**54.00**	

ND indicate undetected fault and bold values indicate the best performance.

the reactants. Based on this fact, kernel functions k1, k3, k14, and k16 have acceptable performances when using the indirect evaluation approach to choosing their parameters. Meanwhile, similar results are obtaining with the most kernel functions through the direct evaluation approach. The better monitoring performances are for the kernel functions k1 and k3, respectively. Beyond that, the kernel functions k3 and k16 showed high-performance indicators for both evaluation approaches. In this context, Table 5 summarizes the fault detection performance for the unlearned abnormal operations. In this case, fault F4 is always detecting by the kernel-based monitoring scheme. Note that the fault F5 is detected with low latency by a few kernel functions. Beyond that, F6 is an incipient fault which can hardly detect by using traditional statistics as have been reported in the specialized literature. In general terms, the kernel-based monitoring scheme achieves to recognize the fault F6 but with high latency and numerously missing alarms. From a comparative view of point, the better performance is achieved with the kernel functions k3, k14, k16, and k10. Note that the monitoring performance of kernel functions k14 and k16 showed the same behavior as in the previous experiments.

5 Conclusions

In this paper, recent developments of kernel-based process monitoring methods with applications for process industries have been reviewed. A detailed discussion of several evaluation criteria for adjusting the kernel-based diagnosis schemes was also provided. The results of this research support that the selection and adjusting of the kernel functions can be both highly dependent on the problem at hand. However, there are some general guidelines and procedures that might contribute to achieving better results with the kernel-based process monitoring methods. This paper grouped these procedures in four consecutive phases: (i) Kernel selection, (ii) Parameters optimization, (iii) Training of methods, and (iv) Validation to fault diagnosis. In this context, the proper choice of the kernel function and its parameters must be ensured to achieve high diagnosis performances. From the reviewed literature were found several criteria to choose the kernel parameters that can be used under the proposed framework. Although the kernel function candidates could be select by the researchers experience/experimental analyzes, the results of this study indicate there are kernel functions potentially useful in the machine learning literature that should be further addressed for fault diagnosis applications. Finally, the comparative analyses developed shown that higher performances can be achieving with other functions beyond that the conventional kernels used in the fault diagnosis applications.

Acknowledgments

The authors acknowledge the financial support provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico, CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, from Brazil. Furthermore, the authors appreciate the support provided by the Asociación Universitaria Iberoamericana de Postgrado (AUIP) in Spain, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) in Brazil, as well as the collaboration given by the University of Technologies of Habana ”José Antonio. Echeverría” (CUJAE) in Cuba.

APPENDIX A
KERNEL PARAMETERS

TABLE 6
Tuned kernel parameters for the numerical example.

Id.	Kernel	Obj. Func. Eq.(30)	Obj. Func. Eq.(10)
k1	Gaussian (RBF)	0.931137242	221.5572451
k2	Cauchy	1,205968833	20304.06342
k3	Rational	1.438203356	74936.50954
k4	Tstudent	3.4	931.9497116
k5	Laplacian	45762.52617	83975.35945
k6	Exponential	387.3407197	84353.27971
k7	Polynomial	44; 73.8	25; 86.5
k8	Multiquadratic	0.39	264.9163524
k9	Inv Multiquad.	0.67	61074.43672
k10	Power	17	2
k11	Log	2	46
k12	NNk13	5.96e-05	1.55733545
k13	RBFpoly	3; 0.94; 0.0039; 0.025	1; 1; 27.885; 1
k14	NNk10	0.001079886	0.219701507
k15	NNk12	0.608573494	8.620795622
k16	NNk11	0.001112046	0.166825671

TABLE 7
Tuned kernel parameters for the Tennessee Eastman process.

Id.	Kernel	Obj. Func. Eq.(30)	Obj. Func. Eq.(10)
k1	Gaussian (RBF)	187.7807	4938.1
k3	Rational	30940	39298e
k7	Polynomial	15; 1.8497e-05	1; 9990.1
k9	Inv Multiquad.	37.1904	6827.5
k10	Power	84	2
k14	NNk10	0.011582	1.4795
k16	NNk11	0.0078724	1.9839e-04

REFERENCES

[1] Y. Motai, "Kernel association for classification and prediction: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 208–223, 2014.

[2] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20 590–20 616, 2017.

[3] K. E. Pilario, M. Shafiee, Y. Cao, L. Lao, and S.-H. Yang, "A review of kernel methods for feature extraction in nonlinear process monitoring," *Processes*, vol. 8, no. 1, p. 24, 2020.

[4] Y. Liu and M. Xie, "Rebooting data-driven soft-sensors in process industries: A review of kernel methods," *Journal of Process Control*, vol. 89, pp. 58–73, 2020.

[5] A. Apsemidis, S. Psarakis, and J. M. Moguerza, "A review of machine learning kernel methods in statistical process monitoring," *Computers & Industrial Engineering*, vol. 142, p. 106376, 2020.

[6] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, 2016.

[7] B.-S. Yang and A. Widodo, "Support vector machine for machine fault diagnosis and prognosis," *Journal of System Design and Dynamics*, vol. 2, no. 1, pp. 12–23, 2008.

[8] S. Cuentas, R. Peñabaena-Niebles, and E. Garcia, "Support Vector Machine in statistical process monitoring: A methodological and analytical review," *The International Journal of Advanced Manufacturing Technology*, vol. 91, no. 1-4, pp. 485–500, 2017.

[9] J. Yu, "A Bayesian inference based Two-stage Support Vector Regression framework for soft sensor development in batch bioprocesses," *Computers & Chemical Engineering*, vol. 41, pp. 134–144, 2012.

[10] L. Zhou, J. Chen, L. Yao, Z. Song, and B. Hou, "Similarity based robust Probability Latent Variable Regression model and its kernel extension for process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, pp. 88–95, 2017.

[11] X. Li, F. Duan, I. Bennett, and D. Mba, "Canonical Variate Analysis, probability approach and Support Vector Regression for fault identification and failure time prediction," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3771–3783, 2018.

[12] M. Zhang, Z. Cai, and W. Cheng, "Multi-model quality prediction approach using fuzzy C-means clustering and support vector regression," *Advances in Mechanical Engineering*, vol. 9, no. 8, pp. 1–17, 2017.

[13] K. Zhu, F. Mei, and J. Zheng, "Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM," *Neurocomputing*, vol. 240, pp. 127–136, 2017.

[14] A. Ramos Rodríguez, J. M. Bernal-de Lázaro, A. Prieto-Moreno, A. J. Silva Neto, and O. Llanes-Santiago, "An approach to robust fault diagnosis in mechanical systems using computational intelligence," *Journal of Intelligent Manufacturing*, vol. 30, no. 4, pp. 1601–1615, 2019.

[15] A. Rodríguez-Ramos, J. M. Bernal-de Lázaro, A. J. Silva-Neto, and O. Llanes-Santiago, "Fault detection using kernel computational intelligence algorithms," in *Computational Intelligence, Optimization and Inverse Problems with Applications in Engineering*. Springer, 2019, pp. 263–281.

[16] X. Bi, J. Lin, D. Tang, F. Bi, X. Li, X. Yang, T. Ma, and P. Shen, "VMD-KFCM algorithm for the fault diagnosis of diesel engine vibration signals," *Energies*, vol. 13, no. 1, p. 228, 2020.

[17] M. Mansouri, R. Baklouti, M. F. Harkat, M. Nounou, H. Nounou, and A. B. Hamida, "Kernel Generalized Likelihood Ratio test for fault detection of biological systems," *IEEE Transactions on Nanobioscience*, vol. 17, no. 4, pp. 498–506, 2018.

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

[19] I. Hamrouni, H. Lahdhiri, K. ben Abdellafou, and O. Taouali, "Fault detection of uncertain nonlinear process using reduced interval kernel principal component analysis (RIKPCA)," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 9, pp. 4567–4576, 2020.

[20] N. Sheng, Q. Liu, S. J. Qin, and T. Chai, "Comprehensive monitoring of nonlinear processes based on concurrent kernel projection to latent structures," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 1129–1137, 2015.

[21] K. Peng, K. Zhang, B. You, J. Dong, and Z. Wang, "A quality-based nonlinear fault diagnosis framework focusing on industrial multimode batch processes," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2615–2624, 2016.

[22] J. Jiao, N. Zhao, G. Wang, and S. Yin, "A nonlinear quality-related fault detection approach based on modified Kernel Partial Least Squares," *ISA Transactions*, vol. 66, pp. 275–283, 2017.

[23] Y. Fu, U. Kruger, Z. Li, L. Xie, J. Thompson, D. Rooney, J. Hahn, and H. Yang, "Cross-validatory framework for optimal parameter estimation of KPCA and KPLS models," *Chemometrics and Intelligent Laboratory Systems*, vol. 167, pp. 196–207, 2017.

[24] J. Feng, J. Wang, H. Zhang, and Z. Han, "Fault diagnosis method of joint Fisher Discriminant Analysis based on the local and global manifold learning and its kernel version," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 122–133, 2015.

[25] Z. Ge, S. Zhong, and Y. Zhang, "Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1403–1411, 2016.

[26] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Fault discriminant enhanced Kernel Principal Component Analysis incorporating prior fault information for monitoring nonlinear processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 162, pp. 21–34, 2017.

[27] Y. Yang, X. Li, X. Liu, and X. Chen, "Wavelet Kernel Entropy Component Analysis with application to industrial process monitoring," *Neurocomputing*, vol. 147, pp. 395–402, 2015.

[28] H. Zhang, Y. Qi, L. Wang, X. Gao, and X. Wang, "Fault detection and diagnosis of chemical process using enhanced KECA," *Chemometrics and Intelligent Laboratory Systems*, vol. 161, pp. 61–69, 2017.

[29] H. Zhou, T. Shi, G. Liao, J. Xuan, L. Su, Z. He, and W. Lai, "Using supervised Kernel Entropy Component Analysis for fault diagnosis of rolling bearings," *Journal of Vibration and Control*, vol. 23, no. 13, pp. 2167–2178, 2017.

[30] Y. Xia, Q. Ding, Z. Li, and A. Jiang, "Fault detection for centrifugal chillers using a Kernel Entropy Component Analysis (KECA) method," in *Building Simulation*. Springer, 2020, pp. 1–9.

[31] L. Bai, Z. Han, J. Ren, and X. Qin, "Research on feature selection for rotating machinery based on supervision Kernel Entropy Component Analysis with Whale Optimization algorithm," *Applied Soft Computing*, p. 106245, 2020.

[32] C. Peng, D. Chunhao, and Z. Qiankun, "Fault diagnosis of microbial pharmaceutical fermentation process with non-gaussian and nonlinear coexistence," *Chemometrics and Intelligent Laboratory Systems*, vol. 199, p. 103931, 2020.

[33] J. Fan and Y. Wang, "Fault detection and diagnosis of non-linear non-gaussian dynamic processes using Kernel Dynamic Independent Component Analysis," *Information Sciences*, vol. 259, pp. 369–379, 2014.

[34] J. M. Bernal-de Lázaro, O. Llanes-Santiago, A. Prieto-Moreno, D. Knupp, and A. Silva-Neto, "Enhanced dynamic approach to improve the detection of small-magnitude faults," *Chemical Engineering Science*, vol. 146, pp. 166–179, 2016.

[35] M. Quiñones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multimode continuous processes: A review," *Chemometrics and Intelligent Laboratory Systems*, 2019.

[36] K. E. S. Pilario, Y. Cao, and M. Shafiee, "Mixed kernel canonical variate dissimilarity analysis for incipient fault monitoring in nonlinear dynamic processes," *Computers & Chemical Engineering*, vol. 123, pp. 143–154, 2019.

[37] L. Shang, J. Liu, and Y. Zhang, "Efficient recursive Kernel Canonical Variate Analysis for monitoring nonlinear time-varying processes," *The Canadian Journal of Chemical Engineering*, vol. 96, no. 1, pp. 205–214, 2018.

[38] J. Yu, K. Wang, L. Ye, and Z. Song, "Accelerated kernel Canonical Correlation Analysis with fault relevance for nonlinear process fault isolation," *Industrial & Engineering Chemistry Research*, vol. 58, no. 39, pp. 18 280–18 291, 2019.

[39] M. Žvokelj, S. Zupan, and I. Prebil, "Non-linear multivariate and multiscale monitoring and signal denoising strategy using Kernel Principal Component Analysis combined with ensemble empirical mode decomposition method," *Mechanical Systems and Signal Processing*, vol. 25, no. 7, pp. 2631–2653, 2011.

[40] J. L. Godoy, D. A. Zumoffen, J. R. Vega, and J. L. Marchetti, "New contributions to non-linear process monitoring through Kernel Partial Least Squares," *Chemometrics and Intelligent Laboratory Systems*, vol. 135, pp. 76–89, 2014.

[41] D. S. Lee, M. W. Lee, S. H. Woo, Y.-J. Kim, and J. M. Park, "Multivariate online monitoring of a full-scale biological anaerobic filter process using kernel-based algorithms," *Industrial & Engineering Chemistry Research*, vol. 45, no. 12, pp. 4335–4344, 2006.

[42] X. Wang, H. Mao, H. Hu, and Z. Zhang, "Crack localization in hydraulic turbine blades based on Kernel Independent Component Analysis and Wavelet Neural Network," *International Journal of Computational Intelligence Systems*, vol. 6, no. 6, pp. 1116–1124, 2013.

[43] Y. Xiao, H. Wang, L. Zhang, and W. Xu, "Two methods of selecting gaussian kernel parameters for one-class SVM and their application to fault detection," *Knowledge-Based Systems*, vol. 59, pp.

75–84, 2014.
[44] P. Kadlec, R. Grbić, and B. Gabrys, "Review of adaptation mechanisms for data-driven soft sensors," *Computers & Chemical Engineering*, vol. 35, no. 1, pp. 1–24, 2011.
[45] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, 2012.
[46] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Industrial & Engineering Chemistry Research*, vol. 52, no. 10, pp. 3543–3562, 2013.
[47] K. Tidriri, N. Chatti, S. Verron, and T. Tiplica, "Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges," *Annual Reviews in Control*, vol. 42, pp. 63–81, 2016.
[48] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 16–25, 2017.
[49] N. M. Nor, C. R. C. Hassan, and M. A. Hussain, "A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems," *Reviews in Chemical Engineering*, 2019.
[50] M. G. Genton, "Classes of kernels for machine learning: A statistics perspective," *Journal of Machine Learning Research*, vol. 2, no. Dec, pp. 299–312, 2001.
[51] S. R. Gunn, "Support Vector Machines for classification and regression," *ISIS Technical Report*, vol. 14, no. 1, pp. 5–16, 1998.
[52] E. Al Daoud and H. Turabieh, "New empirical nonparametric kernels for Support Vector Machine classification," *Applied Soft Computing*, vol. 13, no. 4, pp. 1759–1765, 2013.
[53] C. Zhu, Z. Wang, and D. Gao, "Globalized and localized canonical correlation analysis with multiple empirical kernel mapping," *Neurocomputing*, vol. 154, pp. 257–275, 2015.
[54] M. Tian and W. Wang, "Some sets of orthogonal polynomial kernel functions," *Applied Soft Computing*, vol. 61, pp. 742–756, 2017.
[55] S. Canu and A. Smola, "Kernel methods and the exponential family," *Neurocomputing*, vol. 69, no. 7-9, pp. 714–720, 2006.
[56] J. Wang, Y. Zhang, H. Cao, and W. Zhu, "Dimension reduction method of Independent Component Analysis for process monitoring based on minimum mean square error," *Journal of Process Control*, vol. 22, no. 2, pp. 477–487, 2012.
[57] L. Zhang, W.-D. Zhou, P.-C. Chang, J. Liu, Z. Yan, T. Wang, and F.-Z. Li, "Kernel sparse representation-based classifier," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1684–1695, 2011.
[58] Y. Zhang and S. J. Qin, "Improved nonlinear fault detection technique and statistical analysis," *AIChE Journal*, vol. 54, no. 12, pp. 3207–3220, 2008.
[59] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using Kernel Principal Component Analysis," *Chemical Engineering Science*, vol. 59, no. 1, pp. 223–234, 2004.
[60] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 1, pp. 55–67, 2005.
[61] J.-H. Cho, J.-M. Lee, S. W. Choi, D. Lee, and I.-B. Lee, "Fault identification for process monitoring using Kernel Principal Component Analysis," *Chemical Engineering Science*, vol. 60, no. 1, pp. 279–288, 2005.
[62] C. K. Yoo and I.-B. Lee, "Nonlinear multivariate filtering and bioprocess monitoring for supervising nonlinear biological processes," *Process Biochemistry*, vol. 41, no. 8, pp. 1854–1863, 2006.
[63] M. Jia, F. Chu, F. Wang, and W. Wang, "On-line batch process monitoring using batch dynamic Kernel Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 101, no. 2, pp. 110–122, 2010.
[64] I. B. Khediri, M. Limam, and C. Weihs, "Variable window adaptive Kernel Principal Component Analysis for nonlinear nonstationary process monitoring," *Computers & Industrial Engineering*, vol. 61, no. 3, pp. 437–446, 2011.
[65] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1960–1971, 2011.
[66] X. Deng and X. Tian, "Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor," *Neurocomputing*, vol. 121, pp. 298–308, 2013.
[67] H. Wang and M. Yao, "Fault detection of batch processes based on multivariate functional Kernel Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 78–89, 2015.
[68] M. Yao and H. Wang, "On-line monitoring of batch processes using generalized additive Kernel Principal Component Analysis," *Journal of Process Control*, vol. 28, pp. 56–72, 2015.
[69] G. Cheng, X. Chen, H. Li, P. Li, and H. Liu, "Study on planetary gear fault diagnosis based on entropy feature fusion of ensemble empirical mode decomposition," *Measurement*, vol. 91, pp. 140–154, 2016.
[70] H. Zhang, X. Tian, and X. Deng, "Batch process monitoring based on multiway global preserving Kernel Slow Feature Analysis," *IEEE Access*, vol. 5, pp. 2696–2710, 2017.
[71] X. Deng and L. Wang, "Modified Kernel Principal Component Analysis using double-weighted local outlier factor and its application to nonlinear process monitoring," *ISA Transactions*, vol. 72, pp. 218–228, 2017.
[72] Y. He and Z. Wang, "Regularized kernel function parameter of KPCA using WPSO-FDA for feature extraction and fault recognition of gearbox," *Journal of Vibroengineering*, vol. 20, no. 1, pp. 225–239, 2018.
[73] J. Qian, L. Jiang, and Z. Song, "Locally linear back-propagation based contribution for nonlinear process fault diagnosis," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 3, pp. 764–775, 2020.

[74] B. Zhou and X. Gu, "Multi-block statistics local Kernel Principal Component Analysis algorithm and its application in nonlinear process fault detection," *Neurocomputing*, vol. 376, pp. 222–231, 2020.

[75] Y. Zhang and S. J. Qin, "Fault detection of nonlinear processes using multiway Kernel Independent Component Analysis," *Industrial & Engineering Chemistry Research*, vol. 46, no. 23, pp. 7780–7787, 2007.

[76] R. T. Samuel and Y. Cao, "Kernel Canonical Variate Analysis for nonlinear dynamic process monitoring," *IFAC-PapersOnLine*, vol. 48, no. 8, pp. 605–610, 2015.

[77] X. B. He, Y. P. Yang, and Y. H. Yang, "Fault diagnosis based on variable-weighted Kernel Fisher Discriminant Analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 93, no. 1, pp. 27–33, 2008.

[78] Z.-B. Zhu and Z.-H. Song, "Fault diagnosis based on imbalance modified Kernel Fisher Discriminant Analysis," *Chemical Engineering Research and Design*, vol. 88, no. 8, pp. 936–951, 2010.

[79] ——, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6895–6905, 2011.

[80] G. Rong, S.-Y. Liu, and J.-D. Shao, "Fault diagnosis by locality preserving discriminant analysis and its kernel variation," *Computers & Chemical Engineering*, vol. 49, pp. 105–113, 2013.

[81] Z. Liu, J. Qu, M. J. Zuo, and H.-b. Xu, "Fault level diagnosis for planetary gearboxes using hybrid kernel feature selection and Kernel Fisher Discriminant Analysis," *The International Journal of Advanced Manufacturing Technology*, vol. 67, no. 5-8, pp. 1217–1230, 2013.

[82] J. M. Bernal-de Lázaro, A. Moreno-Prieto, O. Santiago-Llanes, and A. J. Silva-Neto, "Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems," *Computers & Industrial Engineering*, vol. 87, pp. 140–149, 2015.

[83] K. Peng, K. Zhang, G. Li, and D. Zhou, "Contribution rate plot for nonlinear quality-related fault diagnosis with application to the hot strip mill process," *Control Engineering Practice*, vol. 21, no. 4, pp. 360–369, 2013.

[84] H. Zhang, K. Peng, K. Zhang, and G. Li, "Quality-related process monitoring based on total kernel PLS model and its industrial application," *Mathematical Problems in Engineering*, vol. 2013, p. 707953, Dec. 2013.

[85] H. Shi, J. Liu, and Y. Zhang, "An optimized Kernel Principal Component Analysis algorithm for fault detection," *IFAC Proceedings Volumes*, vol. 42, no. 8, pp. 846–851, 2009.

[86] M. Jia, H. Xu, X. Liu, and N. Wang, "The optimization of the kind and parameters of kernel function in KPCA for process monitoring," *Computers & Chemical Engineering*, vol. 46, pp. 94–104, 2012.

[87] M. Van and H.-J. Kang, "Wavelet kernel local Fisher Discriminant Analysis with Particle Swarm Optimization algorithm for bearing defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 12, pp. 3588–3600, 2015.

[88] H. Shi, J. Liu, Y. Wu, K. Zhang, L. Zhang, and P. Xue, "Fault diagnosis of nonlinear and large-scale processes using novel modified Kernel Fisher Discriminant Analysis approach," *International Journal of Systems Science*, vol. 47, no. 5, pp. 1095–1109, 2016.

[89] M. S. Kan, A. C. Tan, and J. Mathew, "A review on prognostic techniques for non-stationary and non-linear rotating systems," *Mechanical Systems and Signal Processing*, vol. 62, pp. 1–20, 2015.

[90] Y. Zhang, H. Zuo, and F. Bai, "Classification of fault location and performance degradation of a roller bearing," *Measurement*, vol. 46, no. 3, pp. 1178–1189, 2013.

[91] S. Dong, D. Sun, B. Tang, Z. Gao, Y. Wang, W. Yu, and M. Xia, "Bearing degradation state recognition based on kernel PCA and wavelet kernel SVM," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 229, no. 15, pp. 2827–2834, 2015.

[92] J. Sheng, S. Dong, and Z. Liu, "Bearing fault diagnosis based on intrinsic time-scale decomposition and improved Support Vector Machine model," *Journal of Vibroengineering*, vol. 18, no. 2, pp. 849–859, 2016.

[93] R. Shao, W. Hu, Y. Wang, and X. Qi, "The fault feature extraction and classification of gear using Principal Component Analysis and Kernel Principal Component Analysis based on the Wavelet packet transform," *Measurement*, vol. 54, pp. 118–132, 2014.

[94] Z. Liu, W. Guo, J. Hu, and W. Ma, "A hybrid intelligent multi-fault detection method for rotating machinery based on RSGWPT, KPCA and Twin SVM," *ISA Transactions*, vol. 66, pp. 249–261, 2017.

[95] X. Wan, D. Wang, W. T. Peter, G. Xu, and Q. Zhang, "A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions," *Measurement*, vol. 78, pp. 138–150, 2016.

[96] N. Sakthivel, S. Saravanamurugan, B. B. Nair, M. Elangovan, and V. Sugumaran, "Effect of kernel function in Support Vector Machine for the fault diagnosis of pump," *Journal of Engineering Science and Technology*, vol. 11, no. 6, pp. 826–838, 2016.

[97] P. K. Wong, Z. Yang, C. M. Vong, and J. Zhong, "Real-time fault diagnosis for gas turbine generator systems using extreme learning machine," *Neurocomputing*, vol. 128, pp. 249–257, 2014.

[98] L. Jiang, B. Zeng, F. R. Jordan, and A. Chen, "Kernel function and parameters optimization in kica for rolling bearing fault diagnosis," *Journal of Networks*, vol. 8, no. 8, p. 1913, 2013.

[99] B. Ma, J. D. Wu, J. Ma, X. D. Wang, and Y. G. Fan, "Fault monitoring and classification method of rolling bearing based on KICA and LSSVM," in *Advanced Materials Research*, vol. 971. Trans Tech Publ, 2014, pp. 476–480.

[100] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.

[101] Z. C. Li, "New detection method for gear faults based on kernel independent component analysis and bp neural network," in *Advanced Materials Research*, vol. 909. Trans Tech Publ, 2014, pp. 371–374.
[102] L. Jiang, J. Xuan, and T. Shi, "Feature extraction based on semi-supervised Kernel Marginal Fisher Analysis and its application in bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 113–126, 2013.
[103] S. W. Choi and I.-B. Lee, "Nonlinear dynamic process monitoring based on dynamic kernel PCA," *Chemical Engineering Science*, vol. 59, no. 24, pp. 5897–5908, 2004.
[104] I. Jaffel, O. Taouali, M. F. Harkat, and H. Messaoud, "Moving window KPCA with reduced complexity for nonlinear dynamic process monitoring," *ISA transactions*, vol. 64, pp. 184–192, 2016.
[105] Q. Zhang, P. Li, X. Lang, and A. Miao, "Improved dynamic Kernel Principal Component Analysis for fault detection," *Measurement*, p. 107738, 2020.
[106] Y. Zhang and Z. Hu, "On-line batch process monitoring using hierarchical Kernel Partial Least Squares," *Chemical Engineering Research and Design*, vol. 89, no. 10, pp. 2078–2084, 2011.
[107] Y. Zhang, S. Li, Z. Hu, and C. Song, "Dynamical process monitoring using dynamical hierarchical Kernel Partial Least Squares," *Chemometrics and Intelligent Laboratory Systems*, vol. 118, pp. 150–158, 2012.
[108] Q. Jia and Y. Zhang, "Quality-related fault detection approach based on dynamic Kernel Partial Least Squares," *Chemical Engineering Research and Design*, vol. 106, pp. 242–252, 2016.
[109] G. Stefatos and A. B. Hamza, "Dynamic Independent Component Analysis approach for fault detection and diagnosis," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8606–8617, 2010.
[110] J. Mori and J. Yu, "Quality relevant nonlinear batch process performance monitoring using a kernel based multiway non-Gaussian latent subspace projection approach," *Journal of Process Control*, vol. 24, no. 1, pp. 57–71, 2014.
[111] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-gaussian processes using gaussian mixture model-based weighted Kernel Independent Component Analysis," *IEEE Transactions on neural networks and learning systems*, vol. 28, no. 1, pp. 122–135, 2015.
[112] Y. Liu, F. Wang, Y. Chang, F. Gao, and D. He, "Performance-relevant Kernel Independent Component Analysis based operating performance assessment for nonlinear and non-Gaussian industrial processes," *Chemical Engineering Science*, vol. 209, p. 115167, 2019.
[113] X. Deng, N. Zhong, and L. Wang, "Nonlinear multimode industrial process fault detection using modified Kernel Principal Component Analysis," *IEEE Access*, vol. 5, pp. 23 121–23 132, 2017.
[114] Y. Chen, X. Deng, and Y. Cao, "Nonlinear soft sensor modeling method based on multimode Kernel Partial Least Squares assisted by improved KFCM clustering," in *Chinese Automation Congress (CAC)*. IEEE, 2019, pp. 4245–4250.
[115] R. Tan, T. Cong, J. R. Ottewill, J. Baranowski, and N. F. Thornhill, "An on-line framework for monitoring nonlinear processes with multiple operating modes," *Journal of Process Control*, vol. 89, pp. 119–130, 2020.
[116] Y. Hu, H. Ma, and H. Shi, "Enhanced batch process monitoring using just-in-time-learning based Kernel Partial Least Squares," *Chemometrics and Intelligent Laboratory Systems*, vol. 123, pp. 15–27, 2013.
[117] H. Zhang, X. Tian, X. Deng, and Y. Cao, "Batch process fault detection and identification based on discriminant global preserving Kernel Slow Feature Analysis," *ISA Transactions*, vol. 79, pp. 108–126, 2018.
[118] W. Li and S. Hongbo, "Improved kernel PLS-based fault detection approach for nonlinear chemical processes," *Chinese Journal of Chemical Engineering*, vol. 22, no. 6, pp. 657–663, 2014.
[119] X. Deng and J. Deng, "Incipient fault detection for chemical processes using two-dimensional weighted SLKPCA," *Industrial & Engineering Chemistry Research*, vol. 58, no. 6, pp. 2280–2295, 2019.
[120] X. Deng, P. Cai, Y. Cao, and P. Wang, "Two-step localized Kernel Principal Component Analysis based incipient fault diagnosis for nonlinear industrial processes," *Industrial & Engineering Chemistry Research*, vol. 59, no. 13, pp. 5956–5968, 2020.
[121] P. Cai and X. Deng, "Incipient fault detection for nonlinear processes based on dynamic multi-block probability related Kernel Principal Component Analysis," *ISA Transactions*, 2020.
[122] R. Vitale, O. E. de Noord, and A. Ferrer, "A kernel-based approach for fault diagnosis in batch processes," *Journal of Chemometrics*, vol. 28, no. 8, pp. S697–S707, 2014.
[123] Y. Zhang, R. Sun, and Y. Fan, "Fault diagnosis of nonlinear process based on KCPLS reconstruction," *Chemometrics and Intelligent Laboratory Systems*, vol. 140, pp. 49–60, 2015.
[124] Z. Li, G. Tan, and Y. Li, "Fault diagnosis based on improved Kernel Fisher Discriminant Analysis." *Journal of Software*, vol. 7, no. 12, pp. 2657–2662, 2012.
[125] Y. Hu, H. Ma, and H. Shi, "Robust online monitoring based on spherical Kernel Partial Least Squares for nonlinear processes with contaminated modeling data," *Industrial & Engineering Chemistry Research*, vol. 52, no. 26, pp. 9155–9164, 2013.
[126] A. R. Teixeira, A. M. Tomé, K. Stadlthanner, and E. W. Lang, "KPCA denoising and the pre-image problem revisited," *Digital Signal Processing*, vol. 18, no. 4, pp. 568–580, 2008.
[127] L. Zhang, W. Zhou, P. Chang, J. Liu, Z. Yan, T. Wang, and F. Li, "Kernel sparse representation-based classifier," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1684–1695, 2012.
[128] K.-P. Wu and S.-D. Wang, "Choosing the kernel parameters for Support Vector Machines by the inter-cluster distance in the feature space," *Pattern Recognition*, vol. 42, no. 5, pp. 710–717, 2009.
[129] R. Ziani, A. Felkaoui, and R. Zegadi, "Bearing fault diagnosis using multiclass Support Vector

Machines with binary Particle Swarm Optimization and regularized Fishers criterion," *Journal of Intelligent Manufacturing*, vol. 28, no. 2, pp. 405–417, 2014.

[130] H. Xiong, M. Swamy, and M. O. Ahmad, "Optimizing the kernel in the empirical feature space," *IEEE Transactions on Neural Networks*, vol. 16, no. 2, pp. 460–474, 2005.

[131] B. Chen, H. Liu, and Z. Bao, "Optimizing the data-dependent kernel under a unified kernel optimization framework," *Pattern Recognition*, vol. 41, no. 6, pp. 2107–2119, 2008.

[132] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. S. Kandola, "On kernel-Target Alignment," in *Advances in Neural Information Processing Systems*, 2002, pp. 367–373.

[133] T. Wang, D. Zhao, and S. Tian, "An overview of kernel alignment and its applications," *Artificial Intelligence Review*, vol. 43, no. 2, pp. 179–192, 2015.

[134] C. H. Nguyen and T. B. Ho, "An efficient kernel matrix evaluation measure," *Pattern Recognition*, vol. 41, no. 11, pp. 3366–3372, 2008.

[135] Y. Baram, "Learning by kernel polarization," *Neural Computation*, vol. 17, no. 6, pp. 1264–1275, 2005.

[136] J.-g. WANG, X.-j. CHEN, and W.-x. ZHANG, "A research on the bearing fault diagnosis algorithm of Least Squares Support Vector Machine with multiple kernels," *Modular Machine Tool & Automatic Manufacturing Technique*, no. 6, p. 19, 2017.

[137] P. Chudzian, "Evaluation measures for kernel optimization," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1108–1116, 2012.

[138] N. Zhang, X. Gao, Y. Li, and P. Wang, "Fault detection of chiller based on improved KPCA," in *Chinese Control and Decision Conference (CCDC)*. IEEE, 2016, pp. 2951–2955.

[139] J.-B. Li, Y.-H. Wang, S.-C. Chu, and J. F. Roddick, "Kernel self-optimization learning for kernel-based feature extraction and recognition," *Information Sciences*, vol. 257, pp. 70–80, 2014.

[140] H. H. Yue and S. J. Qin, "Reconstruction-based fault identification using a combined index," *Industrial & Engineering Chemistry Research*, vol. 40, no. 20, pp. 4403–4414, 2001.

[141] C. F. Alcala and S. J. Qin, "Analysis and generalization of fault diagnosis methods for process monitoring," *Journal of Process Control*, vol. 21, no. 3, pp. 322–330, 2011.

[142] J.-M. Lee, S. J. Qin, and I.-B. Lee, "Fault detection of non-linear processes using Kernel Independent Component Analysis," *The Canadian Journal of Chemical Engineering*, vol. 85, no. 4, pp. 526–536, 2007.

[143] P.-E. P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using Canonical Variate Analysis and kernel density estimations," *IEEE Transactions on Industrial Informatics*, vol. 6, no. 1, pp. 36–45, 2009.

[144] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.

[145] R. Martinez-Cantin, "Bayesopt: A bayesian optimization library for nonlinear optimization, experimental design and bandits," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3735–3739, 2014.

[146] D. Dong and T. J. McAvoy, "Nonlinear Principal Component Analysis based on principal curves and neural networks," *Computers & Chemical Engineering*, vol. 20, no. 1, pp. 65–78, 1996.

PLACE PHOTO HERE

José M. Bernal de Lázaro is graduated in Automatic Control Engineering from the Universidad Tecnológica de La Habana José A. Echeverría (CUJAE, 2010). He received the M.Sc. in Mathematical Modeling applied to Engineering in 2014, and the Ph.D. degree in Applied Sciences from this university in 2016. He is currently working as a Researcher and Full Professor at the Computer Systems and Automation Department of the Universidad Tecnológica de La Habana. His scientific interests are in the fields of Control and Process Systems Engineering, with emphasis on problems of pattern recognition and fault diagnosis in industrial environments.

PLACE PHOTO HERE

Orestes Llanes Santiago obtained the Degree of Electrical Engineer from Universidad Tecnológica de la Habana José A. Echeverría (CUJAE, 1981). He pursued graduate studies from 1989 to 1994 at the Universidad de Los Andes, Mérida-Venezuela, where he obtained the degree of Master of Science in Control Engineering in 1990 and the degree of Doctor in Applied Sciences in 1994. He is currently a Researcher and Full Professor at the Computer Systems and Automation Department of the Universidad Tecnológica de La Habana in Cuba. His areas of interest are the Fault diagnosis in industrial systems, Nonlinear control and Computational intelligence with applications to control.

PLACE PHOTO HERE

Antônio J. da Silva Neto is a Mechanical/Nuclear Engineer (Universidade Federal do Rio de Janeiro−UFRJ, Brazil, 1983), M.Sc. in Nuclear Engineering (UFRJ, 1989), and Ph.D. in Mechanical Engineering, with a minor in Computational Mathematics (North Carolina State University, USA, 1993). In 1997 he joined the UERJ, being currently a Full Professor at the Polytechnic Institute. His areas of interest are Mechanical Engineering and the Applied and Computational Mathematics, with emphasis in numerical methods and inverse problems.

PLACE PHOTO HERE

Carlos Cruz Corona is graduated in Automatic Control Engineering from the Universidad Tecnológica de La Habana José A. Echeverría (CUJAE, 20??). He received the M.Sc. in Mathematical Modeling applied to Engineering in 20??, and the Ph.D. degree in Applied Sciences from this university in 20??. He is currently working as a Researcher and Full Professor at the Computer Systems and Automation Department of the Universidad Tecnológica de La Habana.