Journal of Process Control 44 (2016) 207–223

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Analysis and comparison of an improved unreconstructed variance criterion to other criteria for estimating the dimension of PCA model

Baligh Mnassri*, El Mostafa El Adel, Mustapha Ouladsine

Aix Marseille Université, CNRS, ENSAM, Université de Toulon, LSIS UMR 7296, 13397, Marseille, France

ARTICLE INFO

Article history:
Received 30 June 2015
Received in revised form 24 May 2016
Accepted 5 June 2016
Available online 18 June 2016

Keywords:
Principal component analysis
Data driven process modelling
Selection criteria

ABSTRACT

This paper provides a new criterion to select the significant components of an empirical process model using the principal component analysis approach. The proposed criterion is an improved unreconstructed variance (IUV) applied to a changing of process data representation. Four other criteria are studied to perform fundamental analyses and comparisons to each other. They are well known in the literature as the minimum description length (MDL), the imbedded error (IE), the equality of the eigenvalue (EOE) and the variance of reconstruction error (VRE). The selection of the significant components is usually constrained by three main difficulties such as the noise included in data, the presence of independent and quasi-independent process variables and the size of training samples. This paper presents two fundamental proofs that clarify the limitations of both criteria which are IE and VRE. The consistency of the MDL and EOE criteria improves by increasing the number of training observations. The purpose of the IUV criterion is to enhance the VRE in order to remedy the encountered limitations. The proposed criterion shows a promising consistency as well as a highly robustness versus the mentioned difficulties. Its potential and the limitations of the other criteria are illustrated using two numerical examples and the CSTR process.

1. Introduction

The use of multivariate statistical process analysis techniques has made tremendous successes and benefits for the modelling, control and diagnosis of complex and over instrumented physical processes in various industrial and even non-industrial areas including chemical and petrochemical engineering, microelectronics manufacturing, pharmaceutical, steel, power and desalination industries, air quality monitoring network, etc. [1–4].

Industrial processes provide regularly large volumes of data that can be stored in databases. These data represent an enormous amount of historical or training measurements collected from sensors usually connected to many process variables. Multivariate statistical process monitoring handles the exploitation of these data to ensure the normal operating states for their industrial processes. Such exploitation includes the timely detection of an abnormal event, diagnosing their causal origins and then taking the appropriate decisions and actions to bring the process back to its normal and safe operating state. Until a few decades ago, this process data exploitation is neither easy nor efficient due to many problems usually related to the nature of data which are often enormous, highly correlated and non-causal in nature. As well and due to low signal-to-noise ratios, the information contained in any process variable is often very small. Also, the collected measurements are often missing on many process variables. To adequately handle these databases, empirical modelling methods have the capacity to deal effectively with all these difficulties. The literature has provided several academia research efforts interesting on models developed by using latent variable methods such as principal component analysis (PCA). In multivariate statistical monitoring area, this approach has been successfully applied to detecting the abnormal situations in industrial processes. It handles and addresses all the mentioned problems in a straightforward manner and provides quite presentable and interpretable analysis tools [1,5].

PCA is a modelling approach which reduces the dimensionality of the measurement data space by projecting them onto a low-dimensional latent space. PCA models reveal variations in input data that represent the process variables by respecting a descending order of variances. Notice that these models are usually built using historical or training process data collected during normal operating conditions. For a successful PCA-based multivariate data analysis procedure, the first step involves the estimation of the relevant or the more significant components in the training data matrix. This step requires the determination of the PCA model dimension which consists of such significant principal components

* Corresponding author: Tel.: +33 491 056 041; fax.: +33 491 056 033.
E-mail addresses: baligh.mnassri@lsis.org (B. Mnassri), mostafa.eladel@lsis.org (E.M. El Adel), mustapha.ouladsine@lsis.org (M. Ouladsine).

http://dx.doi.org/10.1016/j.jprocont.2016.06.004
0959-1524/

(PCs). Ideally, the optimal retained number of PCs should be equal to the correct rank of the covariance or the correlation matrices of the training process data. However, in practice, its determination is not unique because real data often contain artifacts such as baseline problems, sensor outputs that are usually disturbed by different noise types, errors generated by data pre-processing and so on [6,7].

Many tests, stopping rules or criteria remedying to the model selection problem using PCA have been provided in the literature, and several comparisons have been made to check the applicability and limitations of these ones on different datasets [8–13]. Most of these criteria provide excellent results when the noise distribution is Gaussian, normal and homoscedastic. However, they may fail to yield easily interpretable results when these conditions are not fulfilled. In the case of training data including a white noise, one can find the optimal number of PCs from the multiplicity of the smallest eigenvalue of the covariance or the correlation matrices. Nevertheless, these matrices are unknown in practice requiring then their estimates. When estimating them from a finite number of training samples, the resulting eigenvalues are all different with probability one, thus making it difficult to determine the optimal number of PCs merely by observing the eigenvalues [14].

In the literature and in a non-exhaustive way, the selection criteria were classified in several categories and solidly discussed in [10–13]. In general, it is possible to distinguish four mainly groups. The first one consists of empirical and heuristic criteria whose the justification of their uses, despite some attempts to put them on a more formal basis, is often subjective such as the scree test criterion [15], the Guttman or Kaiser criterion [16], the cumulative percent variance criterion, etc. These criteria may be grouped together to form a subcategory called eigenvalue-based criteria [12]. On the contrary, the imbedded error (IE, [17]), the autocorrelation criterion [18] and some other criteria have more objective and simple decision rules that are often based on the minimization of the criterion with respect to the number of PCs. Nevertheless, the subjectivity lies in the theory of the criterion itself. The second category of criteria is based on a series of hypotheses that are usually very sophisticated and often overestimate the model size. For instance, Bartlett [19] and Lawley [20] have developed methods from which the confronted problem is related to the choice of the threshold levels for the different tests. A unique statistical framework for consistently estimating the model plane and extracting Gaussian as well as non-Gaussian source signals has been introduced by developing a maximum likelihood formulation. The proposed stopping rule, called equality of the eigenvalue (EOE) test, relies on a hypothesis for testing the equality of the discarded eigenvalues [12,21].

The third criteria category consists of intensive computing methods or cross-validation-based criteria. The cross-validation method has provided the well-known PRESS criterion [22,23]. To avoid its costly computation, Taylor expansions have helped to approximately recalculate the PRESS criterion. By assuming the existence of large number of training samples, the PRESS expression is equivalent to the scree test criterion belonging to the first criteria category [9,24]. The fourth category includes criteria that are mainly outcome from the signal processing field. Such a category is called information theoretic criteria and usually consists of the Akaike information criterion (AIC, [25]) and the minimum description length (MDL, [26,27]). The use of these two criteria was reformulated and adapted by Wax and Kailath [14] to deeply address the model selection problem.

The literature has also provided other significant criteria that are difficult to attribute them to one of the mentioned categories such as the Velicer's partial correlation procedure. Other criteria have been provided to retain the PCA model that ensures the best fault detection and isolation rather than to provide a better estimation of the process data [7,28]. The criterion which represents a particular aim in this work is based on the minimization of the variance of reconstruction error (VRE) also known as the unreconstructed variance [29,30]. This criterion has been categorized by Kruger and Xie [12] within the third category, yet it is not really an intensive method. The variable reconstruction is an estimate of its measurements from those of the other variables using different PCA models. The reconstruction accuracy is related to the capacity of the retained model to reveal the redundancy relations among the variables. Indeed, if too many PCs are chosen, each variable tends to rely too much on itself, which means its relationship to other variables is weakened. If too few PCs are used, the model is inaccurate to represent the normal variation of the data and thus results in poor reconstruction. This criterion features a minimum corresponding to the best reconstruction. However, we will show that it helps to identify the number of the main directions linking only the highly correlated process variables.

The objective in this paper is to provide analysis and comparison of four selection criteria such as MDL, IE, VRE and EOE. The minimization with respect to the number of PCs is the common feature on the use of the first three criteria. However, the EOE should be compared to a threshold to find out the optimal PCA model. The choice to investigate this criterion is based mainly in its likeness in part to the MDL criterion. Through fundamental analyses, we will prove that these criteria are impeded by some limitations and weaknesses that are usually related to the presence of independent and quasi-independent variables. Based on the principle of the unreconstructed variance and a changing of the data representation, a new criterion will be proposed to improve the model selection. After presenting an overview on the PCA principle and some fundamental properties related to the process modelling procedure in Section 2, the followed section presents the MDL criterion theory. It also provides the principle of the IE criterion as well as our proof showing its limitation. The same section details the EOE criterion. The fourth and last reviewed criterion in this section is the VRE. This one is solidly discussed and concluded by our fundamental proof that proves its limitation given in Appendix A. Section 4 presents the motivation, theory and consistency of our proposed approach. All the studied criteria are compared to each other through three illustrative examples in Section 5. Conclusions are given at the end of the paper.

2. PCA-based process modelling and properties

PCA-based process modelling relies on the use of normal training data to build the process model. PCA seeks new directions usually called PCs that explain the more significant variability of data.

2.1. PCA principle

Let $\mathbf{x}(k) \in \mathbb{R}^m$ denote a sample vector of m recorded variable measurements which are collected at the kth instant and under normal operating conditions. Assuming that there are N training samples, a historical process data matrix $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N)]^T \in \mathbb{R}^{N\times m}$ can be built in which each row represents a sample $\mathbf{x}^T(k)$. The training process data matrix $\mathbf{X}$ should be rich in normal variations to be representative to the common-cause variability of the process [31]. As pre-processing, it is conventionally considered that $\mathbf{X}$ should be scaled to zero-mean and usually unit-variance for covariance and correlation matrices-based PCA, respectively.

PCA approach consists of seeking an optimal linear transformation of the historical data matrix $\mathbf{X}$ into a new one noted $\mathbf{T} \in \mathbb{R}^{N\times m}$ and defined as follows:

$$\mathbf{T} = \mathrm{XP} \tag{1}$$

where $\mathbf{T} = [\mathbf{t}(1), \ldots, \mathbf{t}(N)]^T$. The kth row of the matrix $\mathbf{T}$ can be defined with those of the matrix $\mathbf{X}$ as $\mathbf{t}(k) = \mathbf{P}^T\mathbf{x}(k)$ with $k \in \{1, \ldots, N\}$. The

new variables constituting the columns of $\mathbf{T}$ are known as scores or PCs. $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_m] \in \mathbb{R}^{m\times m}$ is an orthonormal matrix whose the columns represent the loadings or eigenvectors. These ones are associated with the eigenvalues of the covariance or correlation matrix $\boldsymbol{\Sigma}$ such as

$$\boldsymbol{\Sigma} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T \tag{2}$$

$\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ is a diagonal matrix whose the entries that represent the eigenvalues of $\boldsymbol{\Sigma}$, are usually placed in decreasing order.

PCA-based process modelling is founded on the selection of an optimal number of PCs noted ℓ that permits the partition of the eigenvector matrix into two submatrices as follows:

$$\mathrm{P} = [\,\hat{\mathrm{P}} \quad \tilde{\mathrm{P}}\,] \tag{3}$$

where $\hat{\mathbf{P}} \in \mathbb{R}^{m\times\ell}$ and $\tilde{\mathbf{P}} \in \mathbb{R}^{m\times(m-\ell)}$.

The first ℓ eigenvectors constitute the representation or the principal component subspace (PCS). The orthogonal complement to this one is known as the residual subspace (RS). It is spanned by the remaining eigenvectors. The projection of $\mathbf{x}(k)$ onto these two subspaces provides its modelled and residual variations given by, respectively:

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{P}}\hat{\mathbf{P}}^T\mathbf{x}(k) = \hat{\mathbf{C}}\mathbf{x}(k) \tag{4}$$

and

$$\tilde{\mathbf{x}}(k) = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T\mathbf{x}(k) = \tilde{\mathbf{C}}\mathbf{x}(k) \tag{5}$$

$\hat{\mathbf{C}}$ and $\tilde{\mathbf{C}} = (\mathbf{I}_m - \hat{\mathbf{C}})$ are the projection matrices onto the PCS and the RS, respectively.

The analyse of the PCA robustness versus the measurement noise can lead to the definition of some properties related to the determination of an adapted structure for the PCA model. Due to the presence of noise in data and by assuming the absence of process disturbances as well as faults, it is possible to structure and model $\mathbf{x}(k)$ as follows:

$$\mathbf{x}(k) = \overset{\circ}{\mathbf{x}}(k) + \mathbf{e}(k) = \mathbf{E}_s\mathbf{s}(k) + \mathbf{e}(k) \tag{6}$$

where $\overset{\circ}{\mathbf{x}}$ is the sample vector of noise-free data. $\mathbf{E}_s \in \mathbb{R}^{m\times q}$ is a parameter matrix of rank $q<m$. $\mathbf{s} \in \mathbb{R}^q$ is a vector of $q<m$ zero-mean stochastic source variables or signals representing the common cause variation of the process. The source variables are assumed to be statistically independent from each other. The noise is represented by the error vector $\mathbf{e} \in \mathbb{R}^m$. It is assumed (i) to follow a zero-mean multivariate Gaussian distribution and (ii) to be statistically independent from each other as well as from the common cause variables. The formal expression of the error covariance matrix is the following:

$$\mathbf{S}_e = \mathcal{E}\{\mathbf{e}\mathbf{e}^T\} \tag{7}$$

where $\mathcal{E}$ designates the expectation operator.

2.2. Noise-free data based process modelling

Let $\overset{\circ}{\mathbf{x}} = \left[\overset{\circ}{\mathbf{x}}(1), \ldots, \overset{\circ}{\mathbf{x}}(N)\right]^T \in \mathbb{R}^{N\times m}$ and $\overset{\circ}{\boldsymbol{\Sigma}}$ denote the noise-free data matrix and its covariance matrix, respectively. The singular value decomposition of $\overset{\circ}{\boldsymbol{\Sigma}}$ is a factorization of the form:

$$\overset{\circ}{\boldsymbol{\Sigma}} = \frac{1}{N-1}\overset{\circ}{\mathbf{x}}^T\overset{\circ}{\mathbf{x}} = \overset{\circ}{\mathbf{P}}\overset{\circ}{\boldsymbol{\Lambda}}\overset{\circ}{\mathbf{P}}^T \tag{8}$$

If the process variables are linearly correlated, the optimal number of PCs is equal to the rank q of the matrix $\overset{\circ}{\boldsymbol{\Sigma}}$ because its last $(m-q)$ eigenvalues are usually null. Consequently, it is possible to deduce the following:

$$\overset{\circ}{\tilde{\mathbf{P}}}^T\overset{\circ}{\mathbf{x}}(k) = \mathbf{0} \in \mathbb{R}^{(m-q)} \tag{9}$$

This relationship is often referred to as a system of $(m-q)$ linear equations. The $(m-q)$ zero eigenvalues in $\overset{\circ}{\boldsymbol{\Sigma}}$ imply the existence of $(m-q)$ linear relations linking the process variables. In the case of noise-free data, the q non-zero eigenvalues involve the existence of q source signals or common cause variables that are linearly independent to each other.

2.3. Noisy data based process modelling

According to Li and Qin [32], PCA provides an unbiased process model only if the used data include white noise, i.e. *independent and identically distributed*. By considering this type of noise with identical variance equals to σ^2, the expression of the covariance matrix of the noisy data can be expressed as

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T = \overset{\circ}{\boldsymbol{\Sigma}} + \mathbf{S}_e = \overset{\circ}{\boldsymbol{\Sigma}} + \sigma^2\mathbf{I}_m \tag{10}$$

Anderson [33] has shown that the eigenvectors of both matrices $\boldsymbol{\Sigma}$ and $\overset{\circ}{\boldsymbol{\Sigma}}$ are identical. Consequently, and for $\ell = q$ PCs, $\tilde{\mathbf{C}} = \overset{\circ}{\tilde{\mathbf{P}}}\overset{\circ}{\tilde{\mathbf{P}}}^T$. Therefore, the residual and estimated of the noisy sample vector $\mathbf{x}(k)$ can be rewritten as follows, respectively:

$$\tilde{\mathbf{x}}(k) = \tilde{\mathbf{C}}\left(\overset{\circ}{\mathbf{x}}(k) + \mathbf{e}(k)\right) = \tilde{\mathbf{C}}\mathbf{e}(k) \tag{11}$$

and

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{C}}\left(\overset{\circ}{\mathbf{x}}(k) + \mathbf{e}(k)\right) \tag{12}$$

In the case of data including a white noise, Eq. (11) shows that the noise-free data cannot be projected into the RS when the PCA model is built using q PCs. This subspace should contain only the noise. However, Eq. (12) implies that the PCS may include a mixture of noise-free data and noise. The zero eigenvalues of the process covariance and/or correlation matrices usually indicate the existence of linear relations linking the process variables. Nevertheless, the smallest eigenvalues reveal the existence of quasi-linear relations. The selection of the relevant PCs plays a major role in the elaboration of the redundancy relations and, therefore, the more adequate process model.

3. Overview and analysis of some selection criteria

Due to the presence of noise in data, the process modelling task is not evident, thus revealing a long lasting research issue. Consequently, many criteria have already been reported in the literature in order to offer solutions that help in estimating the optimal dimension of PCA model. We have chosen four criteria among the most widely used ones to perform an analysis and comparative study of their performances versus a new proposed criterion in this paper. These criteria are separately discussed below.

3.1. Minimum description length criterion

The information theoretic criteria are more popular in signal processing literature and there are common attributes about them. They are applied only under the following hypothesis:

Hypothesis 1. $\boldsymbol{\Sigma}$ needs to be a covariance and not a correlation matrix. The process data include a white noise, i.e. independent and identically distributed with identical variance of σ^2.

Respecting these required assumptions, the information theoretic criteria usually demonstrate a minimum over the number of PCs. Introduced by Akaike [25], Schwarz [26] and Rissanen [27], these criteria address the following general problem for model selection: Given a set of N samples $\mathbf{X} = \{\mathbf{x}(1), \ldots, \mathbf{x}(N)\}$ and a family of models as a parameterized family of probability densities $f(\mathbf{X}|\boldsymbol{\Theta})$, select the model that best fits the process data.

Inspired by Akaike's work, Schwarz and Rissanen have approached the problem from quite different points of view. Schwarz's approach is based on Bayesian arguments. He assumed that each competing model can be assigned a prior probability, and proposed to select the model that yields the maximum posterior probability. Rissanen's approach is based on information theoretic arguments. Since each model can be used to encode the observed process data, Rissanen proposed to select the model that yields the minimum code length, whence the name of minimum description length (MDL) criterion. It turns out that in the large sample limit, both Schwarz's and Rissanen's approaches yield the same criterion given by:

$$\mathrm{MDL} = -\ln f(\mathbf{X}|\hat{\boldsymbol{\Theta}}) + \frac{1}{2}\eta \ln N \tag{13}$$

where ln is the natural logarithm, $\hat{\boldsymbol{\Theta}}$ is the maximum likelihood estimate of the parameter vector $\boldsymbol{\Theta}$ and η is the number of free adjusted parameters in $\boldsymbol{\Theta}$. The first term of the right side of Eq. (13) is the well-known log-likelihood estimator of the model parameters. The second is a bias correction term, inserted so as to make the criterion an unbiased estimate of the mean Kullback-Leibler distance between the modelled and the estimated densities $f(\mathbf{X}|\boldsymbol{\Theta})$ and $f(\mathbf{X}|\hat{\boldsymbol{\Theta}})$, respectively.

The information theoretic criteria, which are originally proposed for the detection of signal sources in signal processing, can be directly applied to the selection of the significant components in data based process modelling field. In this context, Wax and Kailath [14] have taken a different approach by converting the detection problem to a model selection problem. They have reformulated and adapted the application of the classical information theoretic criteria for the model selection. Under Hypothesis 1, the formal expression of the covariance matrix can be expressed with the following form:

$$\mathbf{R} = \mathcal{E}\{\mathrm{x}\mathrm{x}^T\} = \bar{\mathbf{R}} + \sigma^2\mathbf{I}_m \tag{14}$$

where $\bar{\mathbf{R}}$ has a rank $q \leq m$. Denoting by $r_1 \geq r_2, \ldots, \geq r_m$ the eigenvalues of the matrix $\mathbf{R}$, the smallest $(m-q)$ ones are equal to σ^2.

Since the model's covariance matrix is given by Eq. (14), it seems natural to consider the following family of covariance matrices:

$$\mathbf{R}^{(\ell)} = \sum_{a=1}^{\ell}(r_a - \sigma^2)\mathbf{v}_a\mathbf{v}_a^T + \sigma^2\mathbf{I}_m \tag{15}$$

where $r_1, \ldots, r_\ell$ and $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$ are the eigenvalues and eigenvectors of $\mathbf{R}^{(\ell)}$, respectively. Denoting by $\boldsymbol{\Theta}^{(\ell)}$ the parameter vector of the PCA model, it follows that:

$$\boldsymbol{\Theta}^{(\ell)} = \left[r_1, \ldots, r_\ell, \sigma^2, \mathbf{v}_1^T, \ldots, \mathbf{v}_\ell^T\right]^T \tag{16}$$

Since the N samples are regarded as statistically independent Gaussian vectors with zero-means, Wax and Kailath [14] have shown by referring to Anderson [33] that the maximum likelihood of $\boldsymbol{\Theta}^{(\ell)}$ is given by:

$$\hat{\boldsymbol{\Theta}}^{(\ell)} = \left[\lambda_1, \ldots, \lambda_\ell, \frac{1}{m-\ell}\sum_{a=\ell+1}^{m}\lambda_a, \mathbf{p}_1^T, \ldots, \mathbf{p}_\ell^T\right]^T \tag{17}$$

where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues and $\mathbf{p}_1, \ldots, \mathbf{p}_m$ the eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$ given by Eq. (2). By applying some straightforward manipulations, these authors proved that the MDL criterion has the following expression:

$$\mathrm{MDL}(\ell) = -\ln\left(\frac{\prod_{a=\ell+1}^{m}\lambda_a^{\frac{1}{m-\ell}}}{\frac{1}{m-\ell}\sum_{a=\ell+1}^{m}\lambda_a}\right)^{(m-\ell)N} + \frac{1}{2}\ell(2m-\ell)\ln N \tag{18}$$

The optimal number of PCs is considered as the value of ℓ that minimizes the MDL criterion:

$$\ell_{op} = \underset{\ell}{\mathrm{argmin}}\left\{\mathrm{MDL}(\ell)\right\} \tag{19}$$

Wax and Kailath [14] were proved that this criterion is a consistent estimator since any criterion having the following form

$$-\ln f(\mathbf{X}|\hat{\boldsymbol{\Theta}}) + \varpi(N)\ell \tag{20}$$

such that $\varpi(N) \to \infty$ and $\varpi(N)/N \to 0$, yields a consistent estimate with probability one if N tends to infinity. As a result, it is followed that the difference $\left(\mathrm{MDL}(q) - \mathrm{MDL}(\ell)\right)$ yields negative and positive values with probability one for $\ell < q$ and $\ell > q$, respectively (see, e.g. [14] for more details). This implies that the MDL criterion reaches its minimum at $\ell = q$ PCs. Note that this consistency was proved considering the mathematical meaning of the infinity-sample size limit. In practice, this brings up a doubt about the value of N from which such a criterion will be able to estimate the appropriate optimal number of PCs.

3.2. Imbedded error criterion

Factor analysis is rapidly gaining importance in many industrial areas. It is a computer technique for solving multidimensional problems. This technique expresses a data point as a linear sum of product functions. On the other hand, abstract factor analysis is concerned only with data reproduction using the abstract factors. Its first step uses PCA to determine the number of factors in the sum. A reduction is then performed in order to reproduce the data from a space consisting of only the significant factors that were determined by factor analysis. Since the data reproduction provides certainly errors, Malinowski [17] has distinguished three main error types such as real error (RE), imbedded error (IE) and extracted error (XE). By investigating the IE behaviour, this one has been considered by the author as a criterion that may be able to determine the size of factor space or PCA model.

3.2.1. Principle and consistency of the IE criterion

Notice that the IE criterion can be only applied with Hypothesis 1. It is expressed using the residual eigenvalues of the covariance matrix as follows:

$$\mathrm{IE}(\ell) = \left(\frac{\ell}{Nm(m-\ell)}\sum_{a=\ell+1}^{m}\lambda_a\right)^{1/2} \tag{21}$$

Since the process data usually represent a mixture of signals and measurement noise or errors, each PC will contain a portion of the signals as well as a portion of imbedded errors. The aim is to find the point at which all signals are extracted, and only the imbedded errors that are kept. The optimal number of PCs is the value of ℓ that minimises the IE criterion:

$$\ell_{op} = \underset{\ell}{\mathrm{argmin}}\left\{\mathrm{IE}(\ell)\right\} \tag{22}$$

One can express the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$ as being the diagonal entries of the following matrix:

$$\boldsymbol{\Lambda} = \overset{\circ}{\boldsymbol{\Lambda}} + \sigma^2\mathbf{I}_m = \begin{bmatrix} \overset{\circ}{\hat{\boldsymbol{\Lambda}}} + \sigma^2\mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I}_{m-q} \end{bmatrix} \tag{23}$$

In order to provide a theoretical consistency for the IE criterion, Malinowski [17] was limited in his study by showing that such a criterion increases for $\ell \geq q$ PCs since from Eqs. (21) and (23), the IE criterion can be written as follows

$$\mathrm{IE}(\ell \geq q) = \left(\frac{\ell\sigma^2}{Nm}\right)^{1/2} \tag{24}$$

3.2.2. Limitation of the IE criterion

Indeed, the suitable interpretation of Eq. (24) is that the IE criterion does not overestimate the model size. To ensure that its minimum is reached at the qth PC, the behaviour of this one should be analysed for all the values of ℓ. In other words, it is required to prove or establish a necessary condition guaranteeing the decrease of the IE criterion with respect to ℓ for $\ell < q$. To achieve this aim, we showed the following:

$$\mathrm{IE}(\ell < q) = \left(\frac{\ell}{Nm(m-\ell)}\sum_{a=\ell+1}^{q}\overset{\circ}{\lambda}_a + \frac{\ell\sigma^2}{Nm}\right)^{1/2} \tag{25}$$

Since $\mathrm{IE} \geq 0$, we can then checking the sign of the difference $(\mathrm{IE}(\ell-1)^2 - \mathrm{IE}(\ell)^2)$ to analyse the IE behaviour. We have showed that the IE criterion may be monotonically decreasing with respect to $\ell < q$ if

$$\overset{\circ}{\lambda}_\ell - \frac{m}{(m-\ell)(\ell-1)}\sum_{a=\ell+1}^{q}\overset{\circ}{\lambda}_a \geq \left(\frac{m-\ell+1}{\ell-1}\right)\sigma^2 \tag{26}$$

Eq. (26) illustrates a necessary but not a sufficient condition that guaranteeing the decrease of the IE criterion for $\ell < q$. This condition was not provided in the literature. It represents a fundamental consistency analysis that reveals the limitation of the IE criterion. Since such a condition is not sufficient to guarantee the minimum of the studied criterion at the qth PC, it should be also required to ensure that $\mathrm{IE}(q-1) \geq \mathrm{IE}(q)$. As a result, the following condition should be satisfied:

$$\overset{\circ}{\lambda}_q \geq \left(\frac{m-q+1}{q-1}\right)\sigma^2 \tag{27}$$

Although it was assumed that the qth PC has a non-zero variance when considering the noise-free data, the IE criterion cannot guarantee the selection of a such PC if its variance $\overset{\circ}{\lambda}_q$ does not satisfy the condition given by Eq. (27). As can be seen, this condition is parameterized according to the number of process variables m as well the optimal number q of PCs. This parameterization can illustrate the major drawback on the use of the IE criterion. As example, one can assume that the qth PC really represents an independent or quasi-independent process variable. Therefore, this one keeps its variance $\overset{\circ}{\lambda}_q$ even if changing the number of process variables. By adding redundant process variables, the number m will increase without thereby changing the value of q. Consequently, the right term of Eq. (27) will be more severe, increasing then the chance to remove the qth PC from the model.

3.3. Equality of the eigenvalue criterion

The identification of a linear non-causal model based on the data structure of Eq. (6) commences with consistent estimation of the column space of the parameter matrix $\mathbf{E}_s$. According to the authors of [12,21], this relies on the use of maximum likelihood PCA (MPCA) and includes the simultaneous estimation of the model plane and the m diagonal elements of the error covariance matrix $\mathbf{S}_e$ given by Eq. (7). This is accompanied by invoking an approach for determining the number q of source signals even if $\mathbf{S}_e \neq \sigma^2\mathbf{I}_m$. The method is based on a statistical test for testing the equality of the residual eigenvalues of the transformed, also called weighted or scaled, covariance matrix.

3.3.1. Principle of the EOE criterion

The statistic that tests the equality of the $(m-\ell)$ discarded eigenvalues is initially introduced by Jolliffe [11] and Anderson [34] as follows:

$$\mathrm{EOE}(\ell) = \frac{(N-1)}{\tau^2}\left((m-\ell)\ln\left(\frac{1}{m-\ell}\sum_{i=\ell+1}^{m}\hat{\tilde{\lambda}}_i\right) - \sum_{i=\ell+1}^{m}\ln\hat{\tilde{\lambda}}_i\right) \tag{28}$$

τ^2 is a critical value or a statistical confidence limit. In the literature, the expression of this one has had two forms. The authors of [11,12,34] have introduced the following:

$$\tau^2 = \chi^2_{((1/2)(m-\ell+2)(m-\ell+1),\alpha)} \tag{29}$$

which represents a χ^2 distribution with $(1/2)(m-\ell+2)(m-\ell+1)$ degrees of freedom and a significance level α. However, in [21], the expression of τ^2 has been proposed as follows:

$$\tau^2 = \frac{1}{2}\left(c_{\alpha/2} + \sqrt{(m-\ell-1)(m-\ell+2)-1}\right)^2 \tag{30}$$

with $c_{\alpha/2}$ representing the $\alpha/2$ percentile of a standardized Gaussian distribution.

The hypothesis that ℓ denoting the true number of source signals is accepted if $\mathrm{EOE}(\ell) \leq 1$ and rejected if $\mathrm{EOE}(\ell) > 1$. To estimate q, the test commences iteratively for $\ell = 1$ and testing whether $\mathrm{EOE}(\ell) \leq 1$. If this is not the case, incrementing ℓ iteratively until $\mathrm{EOE}(\ell) \leq 1$.

$\hat{\tilde{\lambda}}_i$ is the ith eigenvalue of the weighted covariance matrix: $\hat{\tilde{\boldsymbol{\Sigma}}} = \mathbf{L}^{-1}\boldsymbol{\Sigma}\mathbf{L}^{-T}$, where $\mathbf{L}$ being the lower triangular matrix that can be obtained by applying a Cholesky decomposition to the error covariance matrix, i.e. $\mathbf{S}_e = \mathbf{L}\mathbf{L}^T$. This transformation or scaled version of the covariance matrix $\boldsymbol{\Sigma}$ aims to estimate the true number of source signals even in case of data including a coloured noise, i.e. $\mathbf{S}_e \neq \sigma^2\mathbf{I}_m$, because

$$\hat{\tilde{\boldsymbol{\Sigma}}} = \mathbf{L}^{-1}\mathbf{E}_s\mathcal{E}\{\mathbf{s}\mathbf{s}^T\}\mathbf{e}_s^T\mathbf{L}^{-T} + \mathbf{I}_m = \mathbf{L}^{-1}\overset{\circ}{\boldsymbol{\Sigma}}\mathbf{L}^{-T} + \mathbf{I}_m \tag{31}$$

yields a transformed error covariance matrix $\tilde{\mathbf{S}}_e = \mathbf{I}_m$ that is of the type $\tilde{\sigma}^2\mathbf{I}_m$ where $\tilde{\sigma}^2 = 1$. The dominant eigenvalues of $\mathbf{L}^{-1}\overset{\circ}{\boldsymbol{\Sigma}}\mathbf{L}^{-T}$ are equal to the dominant eigenvalues of $\hat{\tilde{\boldsymbol{\Sigma}}}$ minus one. Note that this concept is based on a priori knowledge about the error covariance matrix. Otherwise, it would be necessary to estimate such a matrix. Narasimhan and Shah [35] were the first to introduce an iterative algorithm for estimating simultaneously the model plane and the diagonal elements of $\mathbf{S}_e$ by assuming that

$$q \leq m + \frac{1}{2} - \sqrt{2m + \frac{1}{4}} \tag{32}$$

Soon after, a slightly different version of this algorithm has been proposed in order to obtain a unified statistical framework for monitoring systems with unknown source and error signals [12,21].

3.3.2. Consistency of the EOE criterion

The EOE test refers to a statistical confidence limit and uses the eigenvalues of the scaled version of the covariance matrix given

in Eq. (31) to select the relevant PCs. In order to analysis its consistency, we suggest firstly to express the EOE criterion for $\ell = q$:

$$\mathrm{EOE}(\ell = q) = \frac{(N-1)}{\tau^2}\left((m-q)\ln\left(\frac{1}{m-q}\sum_{i=q+1}^{m}\hat{\tilde{\lambda}}_i\right) - \sum_{i=q+1}^{m}\ln\hat{\tilde{\lambda}}_i\right) \tag{33}$$

By considering that $\hat{\tilde{\lambda}}_i \approx 1$ for $i \geq q+1$, then $\mathrm{EOE}(q) \approx 0$ and consequently

$$\mathrm{EOE}(\ell \geq q) \approx 0 \tag{34}$$

Since the statistical confidence limit of the EOE test is equal to the unity, the above result proves that the residual PCs have EOE values that cannot usually exceed such a threshold:

$$\mathrm{EOE}(\ell \geq q) \leq 1 \tag{35}$$

Indeed, the only property that can be illustrated is that the EOE criterion does not usually overestimate le model size.

Remark 1. *The MDL, IE and EOE criteria are applied with only covariance but not correlation matrices, which restricts their usefulness still further.*

3.4. Variance of reconstruction error criterion

The variance of reconstruction error (VRE) criterion determines the number of PCs providing the best reconstruction of variables [29,30]. The purpose of the reconstruction approach is to remove the fault effects from these variables. To introduce its principle, it is useful to define the expression of a sample vector with the presence of unidimensional and single faults. Denoting $\mathcal{F}_j$, $\boldsymbol{\xi}_j \in \mathbb{R}^m$ and $\mathbf{x}^* \in \mathbb{R}^m$ as an actual fault, the jth column of the identity matrix and the sample vector of fault-free data, respectively. In the presence of $\mathcal{F}_j$, the vector $\mathbf{x}$ can be modelled as follows:

$$\mathbf{x}(k) = \mathbf{x}^*(k) + \boldsymbol{\xi}_j f(k) \tag{36}$$

where the scalar $f(k)$ represents the fault magnitude at the kth instant.

3.4.1. Principle of the VRE criterion

To remove the actual fault effect, the reconstruction of the jth faulty process variable yields the following reconstructed sample vector:

$$\mathbf{x}_j(k) = \mathbf{x}(k) - \boldsymbol{\xi}_j f(k) = \mathbf{x}(k) - \boldsymbol{\xi}_j \hat{f}_j(k) \tag{37}$$

where $\hat{f}_j$ is the estimated of f on the direction of the jth process variable. This estimate is optimal by minimizing the residual squared error $\|\tilde{\mathbf{C}}\mathbf{x}_j(k)\|^2$ as follows:

$$\hat{f}_j(k) = \underset{f(k)}{\mathrm{argmin}}\left\{\|\tilde{\mathbf{C}}\mathbf{x}_j(k)\|^2\right\} = \left(\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\xi}_j\right)^{-1}\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\mathbf{x}(k) \tag{38}$$

Notice that the reconstruction of the jth process variable can be performed if and only if $\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\xi}_j \neq 0$. This implies that $\tilde{\mathbf{C}}\boldsymbol{\xi}_j$ should not be a zero vector. The formal expression of the individual unreconstructed variance related to the reconstruction of the jth process variable is defined by:

$$\sigma_j^2(\ell) = var\left\{\boldsymbol{\xi}_j^T(\mathbf{x}-\mathbf{x}_j)\right\} = var\{\hat{f}_j\} = \frac{\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\Sigma}\tilde{\mathbf{C}}\boldsymbol{\xi}_j}{\left(\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\xi}_j\right)^2} \tag{39}$$

where *var* designates the operator variance.

As can be seen, the unreconstructed variance is the variance of the estimated fault magnitude. This feature justifies the use of such a variance to select the relevant PCA model because the training data are assumed to be free of faults. The reconstruction of a given variable using a judiciously selected PCA model should provide a minimum. Since m process variables need to be simultaneously considered, it should then identify the PCA model that minimizes the overall variance of reconstruction error represented by the following sum:

$$\mathrm{VRE}(\ell) = \sum_{j=1}^{m}\frac{\sigma_j^2(\ell)}{\boldsymbol{\xi}_j^T\boldsymbol{\Sigma}\boldsymbol{\xi}_j} = \sum_{j=1}^{m}\frac{\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\Sigma}\tilde{\mathbf{C}}\boldsymbol{\xi}_j}{(\boldsymbol{\xi}_j^T\boldsymbol{\Sigma}\boldsymbol{\xi}_j)\left(\boldsymbol{\xi}_j^T\tilde{\mathbf{C}}\boldsymbol{\xi}_j\right)^2} \tag{40}$$

In order to avoid the scale problem between the individual unreconstructed variances, it is important to weighting each of them by its process variable variance that is equal to $\boldsymbol{\xi}_j^T\boldsymbol{\Sigma}\boldsymbol{\xi}_j$. If the process data are scaled to unit-variances, then $\boldsymbol{\xi}_j^T\boldsymbol{\Sigma}\boldsymbol{\xi}_j = 1$. The choice of an optimal PCA model using the VRE criterion is achieved by selecting the optimal number ℓ_{op} of PCs that offer the best reconstruction of data, i.e.

$$\ell_{op} = \underset{\ell}{\mathrm{argmin}}\left\{\mathrm{VRE}(\ell)\right\} \tag{41}$$

3.4.2. Consistency of the VRE criterion

By referring to Dunia and Qin [29,30], the VRE criterion should necessarily have a minimum corresponding to the best reconstruction. To show the existence of such a minimum, these authors [29] have decomposed the individual unreconstructed variance of the jth process variable as follows:

$$\sigma_j^2(\ell) = \tilde{\sigma}_j^2(\ell) + \hat{\sigma}_j^2(\ell) \tag{42}$$

They proved that $\tilde{\sigma}_j^2$ is monotonically decreasing with ℓ. However, $\hat{\sigma}_j^2$ goes to infinity when ℓ is close to m. As a result, σ_j^2 will have a minimum for $\ell \in [1, m]$. This feature can be then extended to the VRE criterion which represents the sum of all the individual unreconstructed variances. Nevertheless, one can wonder whether the number of PCs that corresponds to the minimum of VRE, coincides really with the theoretical number q stated before. Therefore, Valle, Li and Qin [10] have extended their studies to establish some theorems on the VRE consistency. They showed that this criterion can correctly estimate the optimal number of PCs under some conditions that usually depend on the distribution of the noise in data.

Theorem 1. *No matter whether $\boldsymbol{\Sigma}$ is a covariance or a correlation matrix, if the process data include a white noise, i.e. independent and identically distributed with identical variances equal to σ^2, then*

1 $\forall(\ell \geq q)$, $VRE(\ell) \geq VRE(q)$

2 $\forall(\ell < q)$, $\ell_{op} = q$ if $\lambda_q^{\circ} \geq \dfrac{\boldsymbol{\xi}_j^T\tilde{\mathbf{P}}\tilde{\mathbf{P}}^T\boldsymbol{\xi}_j}{\boldsymbol{\xi}_j^T\tilde{\mathbf{P}}_q\tilde{\mathbf{P}}_q^T\boldsymbol{\xi}_j}\sigma^2$

Theorem 1 implies that the VRE criterion does not overestimate the PCA model size. The second constraint requires that the variance of the last PC, which should be theoretically retained, should be larger than the noise variance.

Theorem 2. *No matter whether $\boldsymbol{\Sigma}$ is a covariance or a correlation matrix, if the process data include a coloured noise, i.e. independent and non-identically distributed, such that the eigenvalue matrix has the following form:*

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_q, \sigma_{q+1}^2, \ldots, \sigma_m^2\right)$$

Then $\ell_{op} = q$ if

$$1\ \forall(\ell \geq q),\ \frac{\sigma^2_{q+1}}{\sigma^2_m} \leq \frac{\xi_j^T \tilde{P}_q \tilde{P}_q^T \xi_j}{\xi_j^T \tilde{P}\tilde{P}^T \xi_j}$$

$$2\ \forall(\ell < q),\ \lambda_q \geq \left(1 + \frac{\xi_j^T \tilde{P}\tilde{P}^T \xi_j}{\xi_j^T \tilde{P}_q \tilde{P}_q^T \xi_j}\right) \sigma^2_{q+1}$$

Theorem 2 involves that the variance range of the noise should be small. This implies that the values of these variances should be very close to each other. The second condition indicates that the qth PC should have a variance at least twice as high as the variance of the $(q+1)$th PC.

3.4.3. Limitation of the VRE criterion

In earlier works [36–38], we have noticed from some simulation examples that the VRE criterion does not take into account the presence of independent and quasi-independent process variables although each of these ones represents a full PC. These variables are PCs that should be retained into the model. Indeed, the criterion limitation has been observed only on examples without being shown theoretically. In Appendix A, we provide the theoretical proof of such a limitation. As stated before, this criterion can be applied using either a covariance or a correlation matrix. Nevertheless, our proof will be performed using only the correlation matrix. The demonstration, which is not so easy, is actually much more feasible considering this matrix category. We find that the limitation of this criterion may still be worse when considering a covariance matrix. The aim to deal with a correlation matrix is related to the priori knowledge about the location of the eigenvalues of the PCs that correspond to the independent and quasi-independent variables. Generally, the variances of these PCs are close to the unity. Thereby, they are the last PCs that should be retained into the model.

The proof given in Appendix A shows that the VRE criterion is insensitive to the presence of the independent and/or quasi-independent process variables. The number of PCs that corresponds to the minimum of this criterion when considering scaled process data is simply the number of the main directions expressing the variability between only the highly correlated process variables. These main directions represent really the source signals or the common cause variables that are involved in the generation of at least two recorded variables. A source signal that remains as a recorded variable is simply ignored by the VRE criterion.

4. Proposed criterion: improved unreconstructed variance

So far, the criteria presented in the previous sections suffer from some limitations and weaknesses for adequately selecting the optimal PCA model. As remedy, we propose a new criterion which was initially investigated by Mnassri, El Adel and Ouladsine [38]. This paper provides its theoretical detail.

4.1. Motivations and theory

By considering scaled process data, the VRE criterion can correctly determine the optimal number of PCs only in the absence of independent and/or quasi-independent variables. Indeed, its selection performance may be more limited and weakened using non-scaled process data. As example, an independent process variable that has large variance can occupy the highest ranked into the variance hierarchy of the overall PCs when using a covariance matrix. In this case, the minimum of the VRE criterion will match to a number of PCs that cannot exceed the rank of such a variable. Thus, some PCs that are really significant will be removed from the retained model. These removed PCs can be selected using this criterion only if their variances are higher than the variances of the independent and quasi-independent process variables. To adequately select the PCA model using the VRE criterion, we recommend scaling the process data because the PCs that represent the independent variables are hierarchically placed behind the more significant PCs, which express strong linear correlations in data. In this context, the mainly confronted problem focuses on the identification of such variables.

Indeed, the independent process variables are usually characterized by invariability in terms of independence and variance values even in the case of inverting the correlation matrix. Building on this feature, the main idea of the new proposed criterion focuses on the principle of inverse problems by changing the process data representation. The new criterion uses a reversed variances-based PCA. Since the scaling of data often affects the real linear relationships linking the process variables, we can assume that the scaled data $\mathbf{X}$ are initially issued from a linear transformation of other process data $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)]^T \in \mathbb{R}^{N\times m}$ that are not necessarily scaled to unit-variances. In a vector notation context, this assumption could result the following:

$$\mathbf{x}(k) = \mathrm{A}\mathrm{y}(k) \tag{43}$$

$\mathbf{A} \in \mathbb{R}^{m\times m}$ is a mixing or transformation matrix. $\mathbf{y} \in \mathbb{R}^m$ denotes a sample vector of data from which are extracted those of vector $\mathbf{x}$. Our aim, from the changing of data representation, is to ensure that the obtained data matrix $\mathbf{X}$ is already scaled to unit-variances. In this framework, we could formulate this purpose through the following expression:

$$\mathcal{E}\{\mathrm{x}\mathrm{x}^T\} = \mathbf{A}\mathcal{E}\{\mathrm{y}\mathrm{y}^T\}\mathbf{A}^T = \mathrm{A}\mathrm{S}\mathrm{A}^T = \boldsymbol{\Sigma} \tag{44}$$

where $\boldsymbol{\Sigma}$ and $\mathbf{S}$ are the correlation and covariance matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Through such a linear transformation, we hope that the eigenvalues of $\boldsymbol{\Sigma}$ and those of $\mathbf{S}$ should be linked by maintaining a particular order. Therefore, the determination of the matrices $\mathbf{A}$ and $\mathbf{S}$ can be performed according to three possible propositions. However, only one proposition can satisfy the desired aims.

Proposition 1. *From Eqs. (2) and (44), one can assume that* $\mathbf{A} = \mathbf{P}$ *and* $\mathbf{S} = \boldsymbol{\Lambda}$. *Under these assumptions, the problem to be resolved is brought back to that of the classical case;*

Proposition 2. *Since* $\boldsymbol{\Sigma}$ *should necessarily to be a correlation matrix,* $\mathbf{S}$ *can be considered as the covariance matrix of* $\mathbf{X}$ *and* $\mathbf{A}$ *as a diagonal matrix whose the diagonal elements are the inverses of the variable standard deviations. In this case, we would be faced with a study on non-scaled data matrix* $\mathbf{X}$. *Therefore, the unreconstructed variance applied on these data cannot ensure an optimal selection of the relevant PCs, particularly in the presence of independent process variables that have significant variances;*

Proposition 3. *The only possible assumption is to assume that* $\mathbf{A} = \boldsymbol{\Sigma}$ *and* $\mathbf{S} = \boldsymbol{\Sigma}^{-1}$.

Even though the correlation matrix $\boldsymbol{\Sigma}$ is often invertible when the process data include a noise that have non-zero variances, the use of Proposition 3 requires some precautions. If $\boldsymbol{\Sigma}$ is singular with $\mathrm{rank}(\boldsymbol{\Sigma}) = r < m$, the use of the pseudo-inverse of $\boldsymbol{\Sigma}$ leading to a study for reduced rank correlation matrix, $\mathbf{S} = \boldsymbol{\Sigma}^+$, which is the Moore–Penrose pseudo-inverse.

As stated above, $\boldsymbol{\Sigma}^{-1}$ will represent the covariance matrix of the new data matrix $\mathbf{Y}$. Therefore, the sample vector $\mathbf{y}(k)$ can be expressed as follows:

$$\mathbf{y}(k) = \boldsymbol{\Sigma}^{-1}\mathbf{x}(k) \tag{45}$$

Since $\mathbf{P}$ is an orthonormal matrix, i.e. $\mathbf{P}^{-1} = \mathbf{P}^T$, the singular value decomposition of the covariance matrix $\mathbf{S}$ of data $\mathbf{Y}$ yields the following result:

$$\mathbf{S} = \boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T \tag{46}$$

Eq. (46) shows that the new data of $\mathbf{Y}$ and the original data of $\mathbf{X}$ have PCs that are carried by the same directions. The difference can be illustrated through only their variance values. Let $\mathbf{h} \in \mathbb{R}^m$ denote the sample vector of PCs of the new data matrix $\mathbf{Y}$. Its expression is given below:

$$\mathbf{h}(k) = \mathbf{P}^T\mathbf{y}(k) = \mathbf{P}^T\boldsymbol{\Sigma}^{-1}\mathbf{x}(k) = \boldsymbol{\Lambda}^{-1}\mathbf{t}(k) \tag{47}$$

where $\mathbf{t}$ is the sample vector of PCs of the original data matrix $\mathbf{X}$. Since the variances of these ones are equal to the eigenvalues of $\boldsymbol{\Sigma}$, i.e. $\mathcal{E}\{\mathbf{t}\mathbf{t}^T\} = \boldsymbol{\Lambda}$, we deduce the following:

$$\mathcal{E}\{\mathbf{h}\mathbf{h}^T\} = \boldsymbol{\Lambda}^{-1}\mathcal{E}\{\mathbf{t}\mathbf{t}^T\}\boldsymbol{\Lambda}^{-1} = \boldsymbol{\Lambda}^{-1} \tag{48}$$

Hence, the PCs of data $\mathbf{Y}$ keep the same directions as those of data $\mathbf{X}$ but with reversed variances. On the other hand, and from Eq. (46), we note that this decomposition produces eigenvalues that are placed in increasing order. However, the commonly known principle about the PCA approach is to arrange the obtained eigenvalues in a decreasing order. To respect this principle, Eq. (46) can be rewritten as follows:

$$\mathbf{S} = \mathrm{GDG}^T \tag{49}$$

where

$$\mathbf{G} = \left[\mathbf{g}_1, \ldots, \mathbf{g}_{m-q}, \mathbf{g}_{m-q+1}, \ldots, \mathbf{g}_m\right] = \left[\mathbf{p}_m, \ldots, \mathbf{p}_{q+1}, \mathbf{p}_q, \ldots, \mathbf{p}_1\right] \tag{50}$$

and

$$\begin{aligned}\mathbf{D} &= \mathrm{diag}\left(d_1, \ldots, d_{m-q}, d_{m-q+1}, \ldots, d_m\right)\\ &= \mathrm{diag}\left(\lambda_m^{-1}, \ldots, \lambda_{q+1}^{-1}, \lambda_q^{-1}, \ldots, \lambda_1^{-1}\right)\end{aligned} \tag{51}$$

The matrices $\mathbf{G}$ and $\mathbf{D}$ are similar to, respectively $\mathbf{P}$ and $\boldsymbol{\Lambda}$ whose the columns are flipped in the left-right direction.

Since $\mathbf{S}$ is a covariance matrix, it seems inconsistent with our recommendation about the nature of data on which it is preferable to apply the unreconstructed variance for selecting the optimal PCA model. Nevertheless, the organization of the eigenvalues of the new covariance matrix will provide a major benefit for our new proposed approach. The new covariance matrix category that represents the inverse of the correlation matrix of the original process data will be the more suitable matrix for applying the unreconstructed variance on non-scaled data.

In the presence of independent and/or quasi-independent process variables, the qth PC would certainly be one of these variables. In addition, the inversion of the correlation matrix will not affect their natures. From Eqs. (50) and (51), the q significant PCs in data $\mathbf{X}$ transform to be redundant relations in $\mathbf{Y}$ and vice versa, the $(m-q)$ redundant relations in $\mathbf{X}$ transform to be significant PCs in $\mathbf{Y}$. These changes are due to the inversion of the residual eigenvalues of $\boldsymbol{\Sigma}$ that provide large eigenvalues in $\boldsymbol{\Sigma}^{-1}$. If there is a set of independent process variables in $\mathbf{X}$, these ones mutate into $\mathbf{Y}$ without significant change in their independence. The last PC among them will be associated with the eigenvector $\mathbf{g}_{m-q+1}$ in Eq. (50). In order to select the first q PCs of the data matrix $\mathbf{X}$, we envisage that it is easier to identify the first $(m-q)$ PCs of the matrix $\mathbf{Y}$ based on the unreconstructed variance principle.

4.2. Principle of the proposed IUV criterion

Let $\mathbf{y}^*(k)$ denote the sample vector of fault-free measurements collected at the kth instant. In the presence of a simple fault $\mathcal{W}_i$ carried by the ith direction $\boldsymbol{\zeta}_i$ and having a magnitude $w(k)$, the sample vector $\mathbf{y}(k)$ can be expressed as follows:

$$\mathbf{y}(k) = \mathbf{y}^*(k) + \boldsymbol{\zeta}_i w(k) \tag{52}$$

where the vector $\boldsymbol{\zeta}_i \in \mathbb{R}^m$ represents the ith column of the identity matrix. The reconstruction of the supposed faulty variable generates the following reconstructed sample vector of data:

$$\mathbf{y}_i(k) = \mathbf{y}(k) - \boldsymbol{\zeta}_i\hat{w}_i(k) \tag{53}$$

where $\hat{w}_i$ is the estimate of w. This one is considered optimal by minimizing the residual squared error $\|\tilde{\mathbf{Z}}\mathbf{y}_i(k)\|^2$ as follows:

$$\hat{w}_i(k) = \underset{w(k)}{\mathrm{arg\,min}}\left\{\|\tilde{\mathbf{Z}}\mathbf{y}_i(k)\|^2\right\} = \left(\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i\right)^{-1}\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\mathbf{y}(k) \tag{54}$$

$\tilde{\mathbf{Z}} = \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T$ is the projection matrix of data $\mathbf{Y}$ into the new residual subspace. $\tilde{\mathbf{G}} \in \mathbb{R}^{m\times(m-\kappa)}$ consists of the last $(m-\kappa)$ eigenvectors of the matrix $\mathbf{G}$. κ is the new parameter that represents the number of PCs in the new data of $\mathbf{Y}$.

Notice that the ith variable can be reconstructed if and only if $\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i \neq 0$, i.e. $\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i$ should be a non-zero vector. The new data of $\mathbf{Y}$ are with zero-means since the data matrix $\mathbf{X}$ is centred. Therefore, the average of $\hat{w}_i$ is zero.

The individual unreconstructed variance corresponding to the reconstruction of the ith variable of $\mathbf{Y}$ expressed in the residual subspace is defined by:

$$u_i^2(\kappa) = var\left\{\boldsymbol{\zeta}_i^T(\mathbf{y} - \mathbf{y}_i)\right\} = var\left\{\hat{w}_i\right\} = \frac{\boldsymbol{\zeta}_i^T\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T\boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i\right)^2} \tag{55}$$

The expression of the new improved unreconstructed variance criterion, noted IUV and depending on κ, is the following:

$$\mathrm{IUV}(\kappa) = \sum_{i=1}^{m}\frac{u_i^2(\kappa)}{\boldsymbol{\zeta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i} = \sum_{i=1}^{m}\frac{\boldsymbol{\zeta}_i^T\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T\boldsymbol{\zeta}_i}{(\boldsymbol{\zeta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i)\left(\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i\right)^2} \tag{56}$$

Let κ_{op} denote the optimal number of PCs in the data of $\mathbf{Y}$. We suggest that the minimum of the proposed IUV criterion corresponds to κ_{op} and equals to $(m-q)$ PCs, i.e.

$$\kappa_{op} = \underset{\kappa}{\mathrm{arg\,min}}\left\{\mathrm{IUV}(\kappa)\right\} = m - q \tag{57}$$

As a result, the optimal number ℓ_{op} of PCs in the original data of $\mathbf{X}$ can be deduced as follows:

$$\ell_{op} = q = m - \kappa_{op} = m - \underset{\kappa}{\mathrm{arg\,min}}\left\{\mathrm{IUV}(\kappa)\right\} \tag{58}$$

4.3. Consistency of the proposed IUV criterion

Let $\tilde{u}_i^2$ and $\hat{u}_i^2$ denote two unreconstructed variances such as

$$\tilde{u}_i^2(\kappa) = \frac{\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i\right)} \tag{59}$$

and

$$\hat{u}_i^2(\kappa) = \tilde{u}_i^2(\kappa)\left(\frac{1 - \boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T\tilde{\mathbf{Z}}\boldsymbol{\zeta}_i}\right) \tag{60}$$

From Eq. (55), the weighted individual unreconstructed variance of the ith variable can be expressed as follows:

$$\frac{u_i^2(\kappa)}{\boldsymbol{\zeta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i} = \frac{\tilde{u}_i^2(\kappa) + \hat{u}_i^2(\kappa)}{\boldsymbol{\zeta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i} \tag{61}$$

Since $\boldsymbol{\zeta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\zeta}_i$ is a constant which does not depend on κ, the behaviour study of the proposed IUV criterion may be limited to those of the two functions $\tilde{u}_i^2$ and $\hat{u}_i^2$.

As the new residual subspace is spanned by the last $(m-\kappa)$ eigenvectors, i.e. $\tilde{\mathbf{G}} = \left[\mathbf{g}_{\kappa+1}, \ldots, \mathbf{g}_m\right]$, the removing of the $(\kappa+1)$th

PC from this subspace provides a new matrix $\tilde{\mathbf{G}}_r$ having the following form:

$$\tilde{\mathbf{G}}_r = \left[\mathbf{g}_{\kappa+2}, \ldots, \mathbf{g}_m\right] \tag{62}$$

which involves that:

$$\tilde{\mathbf{G}} = \left[\mathbf{g}_{\kappa+1}, \quad \tilde{\mathbf{G}}_r\right] \tag{63}$$

The same considerations still hold on the residual eigenvalues:

$$\tilde{\mathbf{D}} = \begin{bmatrix} d_{\kappa+1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{D}}_r \end{bmatrix} \tag{64}$$

From Eqs. (63) and (64), Eq. (59) can be rewritten as follows:

$$\begin{aligned} \tilde{u}_i^2(\kappa) &= \left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{D}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^{-1} \\ &= \left(\tilde{u}_i^2(\kappa+1) + d_{\kappa+1}\frac{\boldsymbol{\zeta}_i^T \mathbf{g}_{\kappa+1}\mathbf{g}_{\kappa+1}^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i}\right) \\ &\quad \times \left(\frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \mathbf{g}_{\kappa+1}\mathbf{g}_{\kappa+1}^T \boldsymbol{\zeta}_i + \boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i}\right) \end{aligned} \tag{65}$$

where

$$\tilde{u}_i^2(\kappa+1) = (\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{D}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i)(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i)^{-1} \tag{66}$$

Therefore, it is straightforward to calculate the following:

$$\tilde{u}_i^2(\kappa) - \tilde{u}_i^2(\kappa+1) = (\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r(d_{\kappa+1}\mathbf{I}_r - \tilde{\mathbf{D}}_r)\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i)\Delta(\kappa) \tag{67}$$

such as

$$\Delta(\kappa) = \frac{\boldsymbol{\zeta}_i^T \mathbf{g}_{\kappa+1}\mathbf{g}_{\kappa+1}^T \boldsymbol{\zeta}_i}{(\boldsymbol{\zeta}_i^T \mathbf{g}_{\kappa+1}\mathbf{g}_{\kappa+1}^T \boldsymbol{\zeta}_i + \boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i)(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_r\tilde{\mathbf{G}}_r^T \boldsymbol{\zeta}_i)} \geq 0 \tag{68}$$

where $\mathbf{I}_r$ is an identity matrix having the same size as $\tilde{\mathbf{D}}_r$. $\left(d_{\kappa+1}\mathbf{I}_r - \tilde{\mathbf{D}}_r\right)$ is a diagonal matrix that disposes of the following form and property:

$$d_{\kappa+1}\mathbf{I}_r - \tilde{\mathbf{D}}_r = \begin{bmatrix} d_{\kappa+1} - d_{\kappa+2} & 0 & \cdots & 0 \\ 0 & d_{\kappa+1} - d_{\kappa+3} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & d_{\kappa+1} - d_m \end{bmatrix} \geq 0 \tag{69}$$

From Eqs. (67)–(69), we can deduce that

$$\tilde{u}_i^2(\kappa) \geq \tilde{u}_i^2(\kappa+1) \tag{70}$$

The previous inequality proves that $\tilde{u}_i^2$ is monotonically decreasing with respect to κ. On the other hand, the term $\boldsymbol{\zeta}_i^T \tilde{\mathbf{Z}} \boldsymbol{\zeta}_i$ is close to zero when κ goes to m. This involves that $\hat{u}_i^2$ grows rapidly when κ tends to m. Consequently, the expression given by Eq. (61) should necessarily have a single minimum corresponding to a number of PCs within the interval [1, m]. As well, the proposed IUV criterion will have a single minimum corresponding to a number of PCs within the same interval. However, it is crucial to establish theoretically the conditions guaranteeing to reach this minimum at $(m-q)$ PCs.

Let $\tilde{\mathbf{G}}_q \in \mathbb{R}^{m\times q}$ and $\tilde{\mathbf{D}}_q \in \mathbb{R}^{q\times q}$ denote two matrices that consist of the last q eigenvectors and eigenvalues of $\mathbf{G}$ and $\mathbf{D}$, respectively. For $\kappa = m-q$, u_i^2 can be expressed as follows:

$$u_i^2(\kappa = m-q) = \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{D}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i\right)^2} \leq \frac{d_{m-q+1}}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i} \tag{71}$$

Two cases can be usually illustrated to prove the conditions guaranteeing the minimum of the IUV criterion at $(m-q)$ PCs:

4.3.1. Case of $\kappa \geq$ m − q

The consideration of this case leads to the following result:

$$u_i^2(\kappa) = \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{D}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^2} \geq \frac{d_m}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i} \tag{72}$$

In order for u_i^2 to be monotonically increasing, i.e. $u_i^2(\kappa \geq m-q) \geq u_i^2(\kappa = m-q)$, it is necessary to ensure that

$$\frac{d_{m-q+1}}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i} \leq \frac{d_m}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i} \tag{73}$$

which requires

$$\frac{d_{m-q+1}}{d_m} \leq \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i} \tag{74}$$

4.3.2. Case of $\kappa <$ m − q

The matrices $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{D}}$ can be partitioned respectively as follows:

$$\tilde{\mathbf{G}} = \left[\tilde{\mathbf{G}}_1 \quad \tilde{\mathbf{G}}_q\right] \tag{75}$$

and

$$\tilde{\mathbf{D}} = \begin{bmatrix} \tilde{\mathbf{D}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{D}}_q \end{bmatrix} \tag{76}$$

where $\tilde{\mathbf{G}}_1 \in \mathbb{R}^{m\times(m-\kappa-q)}$ and $\tilde{\mathbf{D}}_1 \in \mathbb{R}^{(m-\kappa-q)\times(m-\kappa-q)}$. The matrices $\tilde{\mathbf{G}}_q$ and $\tilde{\mathbf{D}}_q$ are those used in the previous case.

Since $\tilde{\mathbf{G}}_1$ and $\tilde{\mathbf{G}}_q$ are orthogonal, we can prove the results given below:

$$u_i^2(\kappa < m-q) = \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_1\tilde{\mathbf{D}}_1\tilde{\mathbf{G}}_1^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^2} + \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{D}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^2} \tag{77}$$

and

$$\begin{aligned} u_i^2(\kappa) - u_i^2(m-q) \geq &-\left(d_{m-q+1}\frac{\boldsymbol{\zeta}_i^T(\tilde{\mathbf{G}}\tilde{\mathbf{G}}^T + \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T)\boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i\right)^2}\right) \\ &\times \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_1\tilde{\mathbf{G}}_1^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^2} + d_{m-q}\frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_1\tilde{\mathbf{G}}_1^T \boldsymbol{\zeta}_i}{\left(\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i\right)^2} \end{aligned} \tag{78}$$

To guarantee that $u_i^2(\kappa < m-q) \geq u_i^2(\kappa = m-q)$, the right side term of the previous inequality should be positive or zero. This constraint requires the following condition:

$$d_{m-q} \geq \left(1 + \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}\tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q\tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}\right) d_{m-q+1} \tag{79}$$

Indeed, we can extend the two conditions given by Eqs. (74) and (79) in order to be expressed with the original eigenvalues of

the correlation matrix $\boldsymbol{\Sigma}$. From Eq. (51), they can be transformed respectively as follows:

$$\frac{\lambda_1}{\lambda_q} \leq \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i} \quad \text{for} \quad \kappa \geq m-q \tag{80}$$

and

$$\lambda_q \geq \left(1 + \frac{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}} \tilde{\mathbf{G}}^T \boldsymbol{\zeta}_i}{\boldsymbol{\zeta}_i^T \tilde{\mathbf{G}}_q \tilde{\mathbf{G}}_q^T \boldsymbol{\zeta}_i}\right) \lambda_{q+1} \quad \text{for} \quad \kappa < m-q \tag{81}$$

Eq. (81) reveals an ordinary requirement since its interpretation involves that the variance of any signal should be larger than the noise variance. However, the constraint given by Eq. (80) means that the more significant eigenvalues $(\lambda_1, \ldots, \lambda_q)$ should be very close to each other. The presence of independent and/or quasi-independent variables can lighten such a constraint guaranteeing then the minimum of the new proposed criterion at $(m-q)$ PCs when $\kappa \geq m-q$. This feature can be supported through the proof established in Appendix A, which is still valid even for the IUV criterion. This proof shows that PCs corresponding to weak linear correlations are considered as residual PCs. Indeed, the aim of our proposed approach by applying a changing of the process data representation is to obtain, contrary to the classical approach, a reversed residual subspace. Into this latter, the first PCs represent such a category of variables followed by the more significant PCs of $\mathbf{X}$. The feature of the IUV criterion can ensure an optimal selection of PCs while considering the independent and quasi-independent process variables.

5. Illustrative examples

This section presents three examples in order to illustrate the theoretical analysis of the studied criteria. The first and second ones are simple numerical examples. The third represents the CSTR process with feedback control. For better readability of the figures, all of these ones are displayed in a base 10 logarithmic scale for the y-axes. The legends VRE_{cov} and VRE_{cor} mean the VRE criterion computed using the covariance and correlation matrices, respectively.

5.1. Numerical example 1

This first study focuses on the simulation of a process numerical example. Their noise-free data are collected from some recorded variables that are bonded to each other by respecting the following equations:

$$\overset{\circ}{\mathbf{x}}(k) = \begin{bmatrix} \overset{\circ}{x}_1(k) \\ \overset{\circ}{x}_2(k) \\ \overset{\circ}{x}_3(k) \\ \overset{\circ}{x}_4(k) \\ \overset{\circ}{x}_5(k) \\ \overset{\circ}{x}_6(k) \\ \overset{\circ}{x}_7(k) \\ \overset{\circ}{x}_8(k) \end{bmatrix} = \mathbf{E}_s \mathbf{s}(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.03 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1(k) \\ s_2(k) \\ s_3(k) \\ s_4(k) \\ s_5(k) \end{bmatrix} \tag{82}$$

where

$$\begin{aligned} s_1(k) &= 1 + \vartheta(k)^2 + \sin\left(\frac{k}{3}\right) \\ s_2(k) &= 2\sin\left(\frac{k}{6}\right)\cos\left(\frac{k}{4}\right)\exp\left(-\frac{k}{N}\right) \\ s_3(k) &= \ln\left(4\sin\left(\frac{k}{6}\right)^2 \cos\left(\frac{k}{4}\right)^2 \exp\left(-\frac{2k}{N}\right)\right) \\ s_4(k) &= \varsigma(k)^2 \\ s_5 &\sim \mathcal{N}(0,1) \end{aligned} \tag{83}$$

ϑ, ς and s_5 are uncorrelated random signals that are normally distributed with zero-means and 0.35, 0.35 and 1 as standard deviations, respectively. The presence of noise in data is manifested through the superimposition of independent random signals to the collected noise-free data. These signals are normally distributed with zero-means. The centred noisy data of the kth sample vector $\mathbf{x}(k)$ are generated using Eq. (6) by assuming that the recorded variables are independent to the error variables, these last are Gaussian and both error and source covariance matrices are diagonal, i.e.

$$\mathcal{E}\left\{\overset{\circ}{\mathbf{x}}\mathbf{e}^T\right\} = \mathbf{0}_8 \tag{84}$$

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}_8, \mathbf{S}_e) \tag{85}$$

$$\mathcal{E}\left\{\text{ss}^T\right\} = \text{diag}(\mathbf{var}\{\mathbf{s}\}) \tag{86}$$

where $\mathbf{0}_8$ is a zero matrix of order eight. The diagonal elements of $\mathbf{S}_e \in \mathbb{R}^{8\times 8}$ and the components of the vector $\mathbf{var}\{\mathbf{s}\} \in \mathbb{R}^5$ represent the variances of the error and source variables, respectively.

Centred training data matrices $\mathbf{X}$ consisting of different sample number N are built. These data are firstly used without preprocessing to compute their covariance matrices in order to analyse the behaviours of the MDL, IE and EOE criteria. Secondly, they are scaled to unit-variances in order to investigate the new proposed IUV criterion, which is assumed to be applied using only correlation matrices. Concerning the VRE criterion, both scaled and non-scaled process data are used. The statistical confidence limit given by Eq. (29) is considered for the investigation of the EOE criterion.

5.1.1. Case of data including a white noise

A white noise that has an identical variance $\sigma^2 = 0.25$ and different values of N have been considered to illustrate the robustness of the studied criteria to the presence of noise in data as well as to the training sample number. Fig. 1 displays their results using $N = 2000$ samples. This figure shows that the MDL, IE and EOE criteria have selected 4 PCs to built the optimal PCA model. The VRE_{cov} and VRE_{cor} have retained 3 and 2 PCs, respectively. The proposed IUV criterion has indicated 5 PCs. Indeed, the constraint given by Eq. (27) cannot be fulfilled involving that the IE criterion cannot guarantee an optimal selection of the model size. By considering $m = 8$ and $q = 5$, such a constraint requires that $\overset{\circ}{\lambda}_q \geq \sigma^2 = 0.25$. However, and in the case of non-scaled noise-free data, $\overset{\circ}{\lambda}_q$ is approximately equal to 0.03. Consequently, the mentioned inequality was not satisfied explaining clearly the reason for which the IE criterion cannot select the fifth PC for the model.

Concerning the MDL and EOE criteria, the main reason for which they did not retain the fifth PC is related to the number N of samples constituting the training data matrix. Indeed, their curves, which are displayed in Fig. 1, do not illustrate their convergence behaviours due to principally the sample number as well as the noise. For this reason, we have computed the selection rates for each criterion considering different values of N. Since the noise

Table 1
Selection rates of the retained PC numbers by varying the sample number N constituting the data of the numerical example 1. The rates computed using 1000 simulated training data matrices for each value of N.

		MDL	IE	EOE	VRE_{cov}	VRE_{cor}	IUV
Data including white noise ($\sigma = 0.5$)	2 PCs					100%	
	3 PCs				100%		
	4 PCs	Fig. 2(a)	100%	Fig. 2(b)			
	5 PCs	Fig. 2(a)		Fig. 2(b)			100%
Data including coloured noise	2 PCs					100%	
	3 PCs				100%		
	4 PCs			Fig. 2(c)			
	5 PCs			Fig. 2(c)			100%

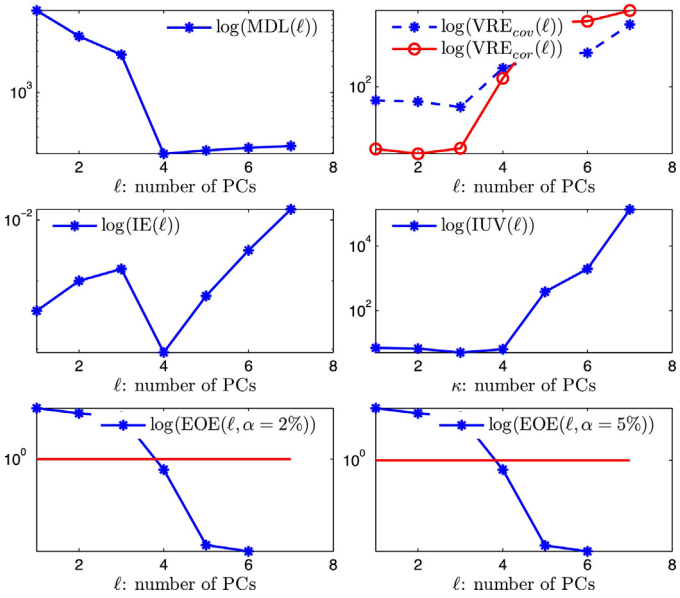

Fig. 1. Results of the studied criteria for a given simulation of process data including a white noise with identical variance $\sigma^2 = 0.25$. $N = 2000$ samples issued from example 1.

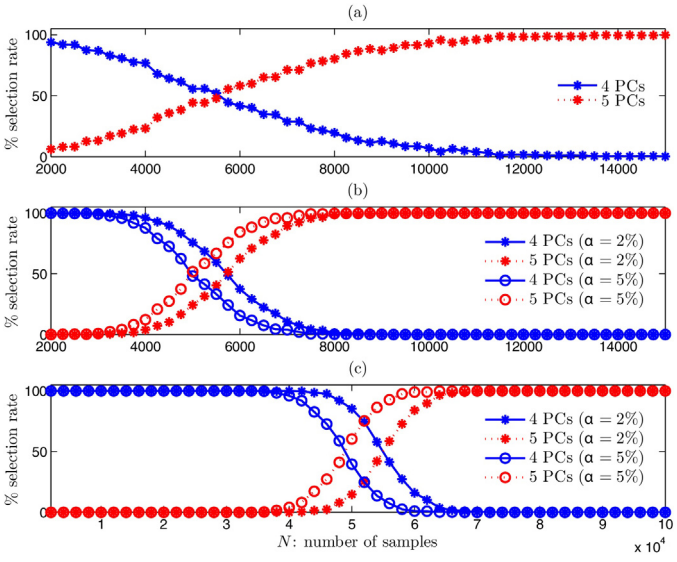

Fig. 2. % Selection rates of the PC numbers in the numerical example 1 based on: (a) MDL criterion using data including a white noise ($\sigma^2 = 0.25$), (b) and (c) EOE criterion using data including a white noise ($\sigma^2 = 0.25$) and data including a coloured noise, respectively.

consists of random signals, one simulation of the process cannot provide data that usually express the consistency and effectiveness of the criteria. For this reason, 1000 data matrices have simulated for each value of N in order to compute the selection rates.

The investigation of the rows corresponding to the white noise case given in Table 1, shows that the VRE_{cor}, VRE_{cov}, IE and IUV criteria maintained the same results displayed in Fig. 1 because the eigenvalues remain insensitive to the increase of the quantity of samples. However, the consistencies of the MDL and EOE criteria are much related to N. From Fig. 2(a), the MDL converges to a choice of 5 PCs for a sample number larger than 14,000. According to Fig. 2(b) and considering α equals to either 2% or 5%, both behaviours of the EOE test and MDL are similar. Nevertheless, the choice convergence based on the EOE is much faster than that of the MDL criterion. This latter retains 5 PCs using a number N larger than 8000 observations. By comparing both MDL and EOE criteria with the proposed IUV, this latter has correctly selected the appropriate number of PCs based on a far lower sample number N.

5.1.2. Case of data including a coloured noise

This case discusses and compares the results of the VRE, EOE and IUV criteria for process modelling procedure based on data including a coloured noise. The other criteria such as the MDL and IE cannot be applied with this noise category. We keep the same numerical process example to generate noise-free data. In the case of a coloured noise, the diagonal elements of the error covariance matrix, which represent the noise variances, are non-identical, i.e. $\mathbf{S}_e \neq \sigma^2\mathbf{I}_m$. These variances are randomly generated into the interval [0.02 0.8]. Their values are given below:

$$\mathrm{diag}(\mathbf{S}_e) = [0.65 \quad 0.1 \quad 0.25 \quad 0.4 \quad 0.3 \quad 0.04 \quad 0.8 \quad 0.2]^T \tag{87}$$

From Fig. 3 and using a training data matrix consisting of 2000 samples, the VRE criterion applied on covariance and correlation matrices shows that the best reconstruction is obtained with 3 and 2 PCs, respectively. The EOE test selects 4 PCs based on a statistical confidence limit computed using either $\alpha = 2\%$ or $\alpha = 5\%$. The

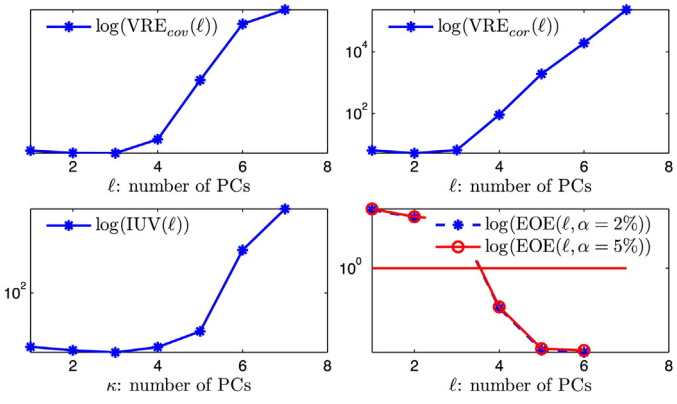

Fig. 3. Results of the VRE, IUV and EOE criteria for a given simulation of process data including a coloured noise. $N = 2000$ samples issued from example 1.

proposed IUV shows that its minimum is reached for $\kappa = 3$ PCs. By applying Eq. (58), this criterion indicates that the optimal size of the PCA model is consisted of 5 PCs, which corresponds to the correct number of source signals.

To confirm these results and similarly to the previous case of the white noise, Table 1 provides the selection rates for the studied criteria computed by varying the sample number N. In all the simulated data matrices, both VRE and IUV criteria gives the same results displayed in Fig. 3. To illustrate the EOE test behaviour, we should refer to Fig. 2(c). This criterion seems to be consistent when considering a large number of training observations. Using a statistical confidence limit calculated with either $\alpha = 2\%$ or $\alpha = 5\%$, this test may ensure the selection of 5 PCs when the used sample number N far exceeds 50,000. The advantage on the use of the IUV criterion lies in its consistency based on a relatively smaller number of training samples.

5.2. Numerical example 2

In many real industrial processes, the number of recorded variables are often much higher. The question that arises is whether the proposed selection criteria still perform better in higher dimensional cases? For this reason, we try to evaluate the ability of the proposed IUV criterion in the determination of the correct dimension of the empirical PCA model when augmenting the number of the recorded and even the source signals. Eight new recorded variables and two new source signals will be added in this second numerical example. In the absence of noise, all the 16 recorded variables depend on each other using the following mixing matrix:

$$\begin{bmatrix} \overset{\circ}{x}_1(k) \\ \overset{\circ}{x}_2(k) \\ \overset{\circ}{x}_3(k) \\ \overset{\circ}{x}_4(k) \\ \overset{\circ}{x}_5(k) \\ \overset{\circ}{x}_6(k) \\ \overset{\circ}{x}_7(k) \\ \overset{\circ}{x}_8(k) \\ \overset{\circ}{x}_9(k) \\ \overset{\circ}{x}_{10}(k) \\ \overset{\circ}{x}_{11}(k) \\ \overset{\circ}{x}_{12}(k) \\ \overset{\circ}{x}_{13}(k) \\ \overset{\circ}{x}_{14}(k) \\ \overset{\circ}{x}_{15}(k) \\ \overset{\circ}{x}_{16}(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.03 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 1 \\ -0.5 & 0 & 1 & 2 & 0 & -2 & 0.5 \\ 1 & 1 & 2 & 0 & 1 & 0 & 1 \\ -0.5 & 0 & 2 & 1 & 0 & -3 & 2 \\ 1 & 0 & -1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} s_1(k) \\ s_2(k) \\ s_3(k) \\ s_4(k) \\ s_5(k) \\ s_6(k) \\ s_7(k) \end{bmatrix} \quad (88)$$

where

$$\begin{aligned} s_6 &\sim \mathcal{N}(0,1) \\ s_7 &\sim \mathcal{N}(0,1) \end{aligned} \quad (89)$$

The source signals s_i for $i \in \{1, \ldots, 5\}$ are given in Eq. (82). Obviously, s_i for $i \in \{5, \ldots, 7\}$ are normally distributed and statistically independent. White and coloured noises will be considered to generate the studied training process data. Evidently, the assumptions about the generation of noisy data are still holding. In this example, it is expected that the criteria select 7 PCs among 16 PCs since there are 7 sources signals.

5.2.1. Case of data including a white noise

Since the generated training data included a white noise, all the studied criteria such as MDL, IE, EOE, VRE and IUV can be compared together. The identical variance of the white noise is equal to 0.25. The investigation of the results displayed in Fig. 4 shows that the MDL, IE and EOE criteria denote 6 PCs as an optimal number to build the model. These results are obtained by considering $N = 2000$ training samples. The VRE criterion suggests 4 and 6 PCs by using the correlation and the covariance matrices, respectively. From the same figure, the proposed IUV criterion retains 7 PCs by respecting its principle to determine the selected PCs number.

Due to the randomness of the measurement noise generation and the considered size of training data, the simulations have repeated several times. It was noticed that the IE, VRE and IUV criteria maintained the results displayed in Fig. 4. However, the MDL and EOE behaviours depended on the number of training samples. The MDL criterion has selected 7 PCs when increasing N to more than 6000 samples. The EOE test has retained 7 PCs based on a required number of N that is significantly higher than 7000 samples. Otherwise, this criterion found that the optimal PCA model is constituted of 6 PCs. Indeed, this case study highlights the robustness of the proposed IUV with respect to the higher dimensional of systems. Again, this criterion proves its ability in estimating the correct number of source signals using a limited number of training samples.

5.2.2. Case of data including a coloured noise

Recalling that in the case of data with coloured noise, only three criteria can be considered, which are the VRE, EOE and IUV. The noise variances that constitute the diagonal elements of the error covariance matrix are non-identical and have taken the following values:

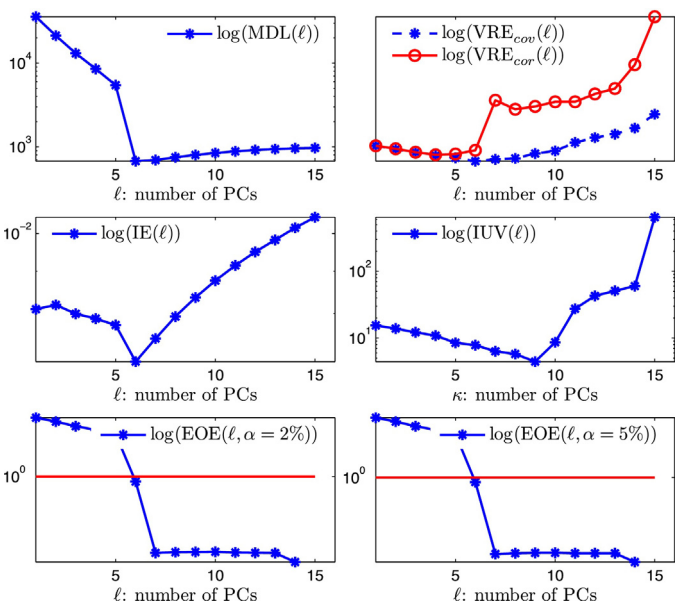

Fig. 4. Results of the studied criteria for a given simulation of process data including a white noise with identical variance $\sigma^2 = 0.25$. $N = 2000$ samples issued from example 2.

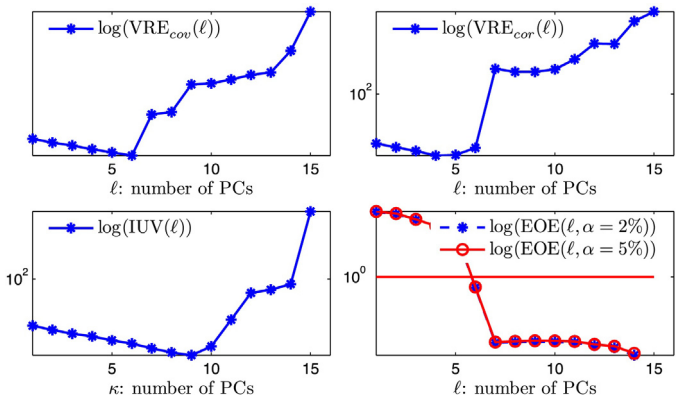

Fig. 5. Results of the VRE, IUV and EOE criteria for a given simulation of process data including a coloured noise. $N = 2000$ samples issued from example 2.

$$\text{diag}(\mathbf{S}_e) = [0.65 \quad 0.1 \quad 0.25 \quad 0.4 \quad 0.3 \quad 0.04 \quad 0.8 \quad 0.2 \quad 0.35 \quad 0.1 \quad 0.2 \quad 0.05 \quad 0.5 \quad 0.03 \quad 0.01 \quad 0.02]^T \tag{90}$$

The results of the studied criteria, using 2000 training samples, are illustrated in Fig. 5. As can be seen, the VRE criterion designates 4 and 6 PCs to build the optimal PCA model according to the correlation and the covariance matrices, respectively. The EOE test selects 6 PCs using a confidence limit that is computed with either $\alpha = 2\%$ or $\alpha = 5\%$. The minimum of the proposed IUV criterion is obtained for $\kappa = 9$ PCs. By referring to Eq. (58), this criterion shows that the optimal empirical model consists of 7 PCs.

To address the problem relies on the randomness generation of noise and the number of training samples, several simulations of data and calculation of these criteria were achieved. It is noted that both VRE and IUV criteria maintained the results given in Fig. 5. The only criterion that has an unstable behaviour is the EOE test. This one has ensured the selection of 7 PCs based on a training sample number N that is significantly higher than 25,000. Once again, this example has illustrated the robustness as well as effectiveness of the proposed IUV criterion compared to the other criteria.

5.3. CSTR with feedback control

The analysed selection criteria are applied on data collected from a nonisothermal continuous stirred tank reactor (CSTR) process. The process model and simulation conditions are similar to those provided by Yoon and MacGregor [39]. The diagram flow of the studied CSTR process is illustrated by Fig. 6. The process has one feed stream that is the merged stream of the solvent and the reactant A, one product stream, and a cooling water flow through the coils. The reactant A premixed with the solvent is converted to product B by respecting a first order reaction and assuming that the tank is perfectly mixed and the physical properties are constant. The dynamic behaviour of the CSTR process is modelled considering the mass balance of reactant A and the total energy balance of the reacting system as follows:

$$\frac{\partial C_A}{\partial t} = \frac{F}{V}(C_{A0} - C_A) - \gamma \tag{91}$$

and

$$\frac{\partial T}{\partial t} = \frac{F}{V}(T_0 - T) - \frac{UA}{\rho C_P V}(T - T_C) - \frac{\Delta H_{\text{rxn}}}{\rho C_P}\gamma \tag{92}$$

where V, ρ and C_P are the volume in the tank, the density in the inlet stream and the specific heat capacity of the merged reactant, respectively. γ and ΔH_{rxn} designate the reaction rate and the reaction heat, respectively. UA is the heat transfer coefficient.

The process is monitored measuring nine variables that are the cooling water temperature (T_C), the cooling water flow (F_C), the inlet temperature of the merged reactant (T_0), the inlet concentration of the reactant (C_{AA}), the reactant flow (F_A), the inlet concentration of the solvent (C_{AS}), the solvent flow (F_S), the outlet concentration (C_A), and the outlet temperature (T).

Evidently, the inlet flow rate of the merged reactant is the sum of the feed streams of the reactant and the solvent, i.e. $F = F_A + F_S$. The inlet concentration of the merged reactant (C_{A0}) is also obtained using these two feed streams as follows:

$$C_{A0}F = C_{AA}F_A + C_{AS}F_S \tag{93}$$

Both inlet concentrations (C_{AS}) and (C_{AA}), the inlet temperature (T_0), the solvent flow (F_S), and the cooling water temperature (T_C) are considered as measured process disturbances. Two unmeasured stochastic state disturbances (a_1) and (a_2) arise from poisoning of the reaction and from fouling of the cooling coils are

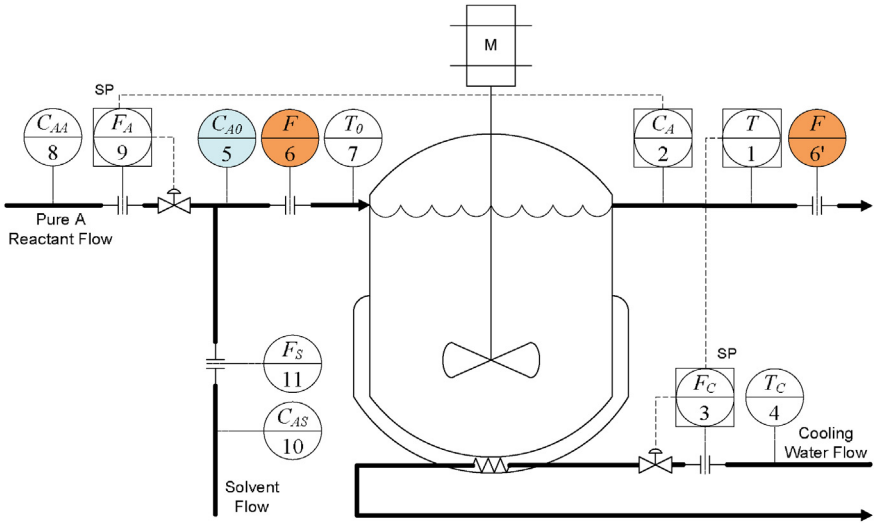

Fig. 6. Flow diagram of the continuous stirred tank reactor process.

Table 2
Simulation information of the CSTR process.

CSTR parameters	
$V = 1\,(m^3)$	$\rho = 10^6\,(g/m^3)$
$E/R = 8330.1\,(K)$	$C_P = 1\,(cal/gK)$
$\Delta H_{rxn} = -1.3 \times 10^7\,(cal/kmol)$	$k_0 = 10^{10}\,(min^{-1})$
$a = 1.678 \times 10^6\,(cal/minK)$	$b = 0.5$
Initial conditions	
$T_0 = 370.0\,(K)$	$T_C = 365.0\,(K)$
$F_C = 15\,(m^3/min)$	$T = 368.25\,(K)$
$F_S = 0.9\,(m^3/min)$	$F_A = 0.1\,(m^3/min)$
$C_A = 0.8\,(kmol/m^3)$	$C_{AS} = 0.1\,(kmol/m^3)$
$C_{AA} = 19.1\,(kmol/m^3)$	

Variances of the measurement noises and disturbances and AR coefficients

	σ_m^2	σ_e^2	ϕ
T	4×10^{-4}		
C_A	2.5×10^{-5}		
F_C	1.0×10^{-2}		
T_C	2.5×10^{-3}	0.475×10^{-1}	0.9
T_0	2.5×10^{-3}	0.475×10^{-1}	0.9
C_{AA}	1.0×10^{-2}	0.475×10^{-1}	0.9
F_A	4×10^{-6}		
C_{AS}	2.5×10^{-5}	1.875×10^{-3}	0.5
F_S	4.0×10^{-6}	0.19×10^{-2}	0.9
a_1		0.19×10^{-2}	0.9
a_2		0.0975×10^{-2}	0.95
PI temperature controller parameters			
$P(T) = -1.5$		$I(T) = -5.0$	

introduced as pre-multipliers in the reaction rate and the heat transfer coefficient, respectively as follows:

$$\gamma = a_1 k_0 \exp\left(-\frac{E}{RT}\right) C_A \tag{94}$$

and

$$UA = a_2 a F_C^b \tag{95}$$

All these measured and unmeasured stochastic disturbances are modelled as first order autoregressive processes. Note that an autoregressive signal can be obtained by filtering a white process noise. In addition, Gaussian measurement noises are added to all the process variable data. Table 2 provides the CSTR process parameters, the initial conditions, the variances of the measurement noises (σ_m^2), the variances of the process noises (σ_e^2) and their autoregressive coefficients (ϕ). These ones are selected to be the same as in Yoon and MacGregor [39].

From Fig. 6, the flows of the reactant (F_A) and the coolant water (F_C) are used to control the outlet concentration (C_A) and the outlet temperature (T), respectively. In this application, only the temperature controller is active. The nine monitored variables, that constitute the sample vector **x**, are sampled every minute. Two training data matrices are collected in order to model the CSTR process. The first data matrix consists of 2160 samples that correspond to three days of normal operating. The second one includes 5040 observations representing a continuous normal operation during a full week.

Indeed, a coloured noise was included to the process data since it consists of Gaussian signals that have different variances (see Table 2). Therefore, only the VRE, EOE and IUV criteria should be applied to model the CSTR process. Figs. 7 and 8 display the criteria results applied on the first and second training data matrices, respectively. The investigation of these figures shows similar results despite the different sizes and backgrounds of the collected training data. The VRE criterion selects 1 and 2 PCs based on the covariance and correlation matrices, respectively. The EOE test estimates 7 PCs using a confidence limit computed with either $\alpha = 2\%$ or $\alpha = 5\%$. The minimum of the proposed IUV criterion corresponds

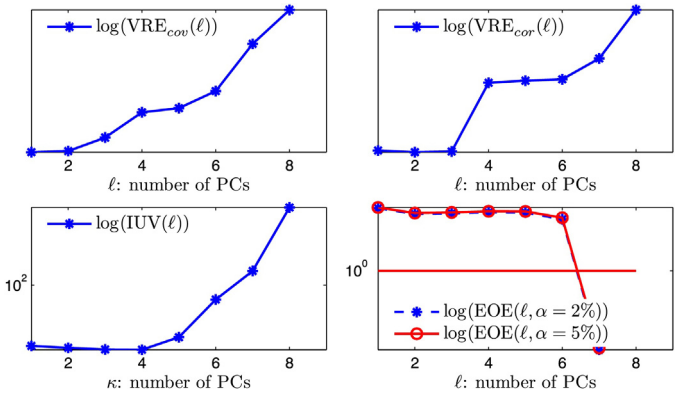

Fig. 7. Results of the VRE, IUV and EOE criteria for a given simulation of process data including a coloured noise and consisting of $N = 2160$ min (three days).

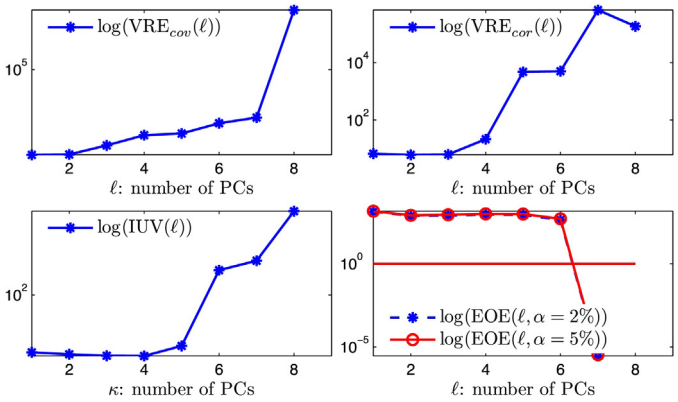

Fig. 8. Results of the VRE, IUV and EOE criteria for a given simulation of process data including a coloured noise and consisting of $N = 5040$ min (one week).

to $\kappa = 4$ PCs. According to Eq. (58), this criterion retains an optimal empirical model that consists of 5 PCs.

Based on this CSTR example, we can elect the IUV criterion as the most reasonable one that may correctly select the optimal model. The investigation of the eigenvalues and eigenvectors shows that the third, fourth and fifth PCs correspond to independent and quasi-independent recorded variables. This proves our reasoning about the limitation of the VRE criterion. As can be seen, the EOE test has overestimated the PCA model, which seems inconsistent with its property. This may be due to a biased estimator of noises resulting bad estimations.

6. Conclusions

This paper has presented our contributions to the topic of PCA-based process modelling using an improved variance of reconstruction error. Due to the diversity of the selection criteria in the literature and the remarkable divergence of their results, we have motivated the precision importance in choosing the appropriate structure for the PCA model. Insofar as this choice is not optimal, the established model does not adequately describe its real system.

Five selection criteria such as the MDL, IE, EOE, VRE and a new proposed IUV have been considered to provide fundamental analyses as well as comparisons of their performance and robustness versus the measurement noise, the presence of independent and quasi-independent process variables and the quantity of training samples. Through theoretical proofs and three examples, we have confirmed the weaknesses in using these criteria. The IE often abandons some significant PCs in case of large noise variance. By using

either covariance matrix or correlation matrix, the VRE criterion cannot cross and therefore, retain the PCs corresponding to the weakly correlated or even independent process variables. On the other hand, the MDL and EOE criteria seem more suitable for these constraints. However, the EOE test depends relatively on a significance level used to compute its statistical confidence limit. Indeed, the performance of the EOE and MDL criteria in estimating the correct number of PCs is similar and strongly linked to the considered number of training samples. The consistency of these two criteria degrades when the noise variance increases. However, it improves by increasing exponentially the number of training observations. In practice, the noise variance is usually unknown and the samples are available only on a limited horizon. Faced with these difficulties and limitations, the proposed IUV criterion shows promising consistency as well as a reasonable robustness. Through the simulation examples, this criterion has remedied the limitations of the other criteria and overcoming the imposed constraints as the measurement noise, the presence of independent and quasi-independent process variables and the training sample size.

Appendix A. Proof of the limitation of the VRE criterion

Let's suppose that all the m process variables are linearly correlated. $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$ is a correlation matrix if the process data are scaled to zero-means and unit-variances, i.e. $\boldsymbol{\xi}_j^T \boldsymbol{\Sigma} \boldsymbol{\xi}_j = 1 \forall (1 \le j \le m)$. Considering data that include noise with non-zero variances, we suppose that there may be q PCs ensuring the minimum of the VRE criterion:

$$q = \arg\min_{\ell} \left\{ \mathrm{VRE}(\ell) \right\} \tag{A.1}$$

Since the optimal PCA model is spanned by the first q PCs, an appropriate partition of the loading matrix $\mathbf{P}$ can yield the following:

$$\mathbf{P} = \begin{bmatrix} \hat{\mathbf{U}} & \tilde{\mathbf{U}} \end{bmatrix} \tag{A.2}$$

where $\hat{\mathbf{U}} \in \mathbb{R}^{m \times q}$ and $\tilde{\mathbf{U}} \in \mathbb{R}^{m \times (m-q)}$ consist of the first q and the latest $(m-q)$ eigenvectors of $\mathbf{P}$, respectively.

Let's assume that a new independent or quasi-independent process variable has included into the data matrix $\mathbf{X}$. This allows to compute a new correlation matrix $\mathbf{M} \in \mathbb{R}^{(m+1)\times(m+1)}$ that may have the following form:

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}^T & 1 \end{bmatrix} = \mathrm{VDV}^T \tag{A.3}$$

where $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is a vector whose the Euclidean norm is close to zero since the new added process variable is nearly independent. The components of this vector represent the correlation coefficients of the added variable with the other process variables. $\mathbf{V}$ and $\mathbf{D}$ are the eigenvector and eigenvalue matrices of $\mathbf{M}$, respectively.

When considering the independence of the added variable, i.e. $\|\epsilon\| \approx 0$, the matrix $\mathbf{V}$ can be approximately written in function of $\mathbf{P}$ by applying the appropriate subdivisions as follows:

$$\mathbf{V} \approx \begin{bmatrix} \hat{\mathbf{U}} & \boldsymbol{\phi}_2 & \tilde{\mathbf{U}} \\ \boldsymbol{\phi}_1^T & \omega & \boldsymbol{\phi}_3^T \end{bmatrix} \tag{A.4}$$

where the Euclidean norms of the vectors $\boldsymbol{\phi}_1 \in \mathbb{R}^q$, $\boldsymbol{\phi}_2 \in \mathbb{R}^m$ and $\boldsymbol{\phi}_3 \in \mathbb{R}^{(m-q)}$ tend to zero. Since $\mathbf{V}$ is an orthonormal matrix, ω is a scalar such that $|\omega|$ is close to unity. Further, $\|\boldsymbol{\phi}_1\| \ll |\omega|$, $\|\boldsymbol{\phi}_2\| \ll |\omega|$ and $\|\boldsymbol{\phi}_3\| \ll |\omega|$.

Let F and $\hbar$ denote the VRE criterion and the new selected number of PCs using the new process data, respectively. The expression of F is the following:

$$F(\hbar) = \sum_{i=1}^{m+1} \frac{\boldsymbol{\xi}_i^T \tilde{\mathbf{V}} \tilde{\mathbf{V}}^T \mathbf{M} \tilde{\mathbf{V}} \tilde{\mathbf{V}}^T \boldsymbol{\xi}_i}{\left(\boldsymbol{\xi}_i^T \tilde{\mathbf{V}} \tilde{\mathbf{V}}^T \boldsymbol{\xi}_i\right)^2} \tag{A.5}$$

where $\tilde{\mathbf{V}} \in \mathbb{R}^{(m+1)\times(m+1-\hbar)}$ consists of the last $(m+1-\hbar)$ eigenvectors of the matrix $\mathbf{V}$.

In order to investigate the behaviour of the F criterion using the proposed approximations, three cases can be illustrated:

A.1. Case of $\hbar \le q$

This case involves that the matrix $\mathbf{V}$ should still undergo a partition as follows:

$$\mathbf{V} \approx \begin{bmatrix} \hat{\mathbf{U}}_1 & \hat{\mathbf{U}}_2 & \boldsymbol{\phi}_2 & \tilde{\mathbf{U}} \\ \boldsymbol{\phi}_{11}^T & \boldsymbol{\phi}_{12}^T & \omega & \boldsymbol{\phi}_3^T \end{bmatrix} \tag{A.6}$$

where $\hat{\mathbf{U}}_1 \in \mathbb{R}^{m\times\hbar}$ and $\hat{\mathbf{U}}_2 \in \mathbb{R}^{m\times(q-\hbar)}$ are composed by the first $\hbar$ and the last $(q-\hbar)$ eigenvectors of the matrix $\hat{\mathbf{U}}$, respectively. The vectors $\boldsymbol{\phi}_{11} \in \mathbb{R}^{\hbar}$ and $\boldsymbol{\phi}_{12} \in \mathbb{R}^{(q-\hbar)}$ are constituted by the first $\hbar$ and the last $(q-\hbar)$ components of the vector $\boldsymbol{\phi}_1$, respectively.

This partition enables to express the matrix $\tilde{\mathbf{V}}$ in the following form:

$$\tilde{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{U}}_2 & \boldsymbol{\phi}_2 & \tilde{\mathbf{U}} \\ \boldsymbol{\phi}_{12}^T & \omega & \boldsymbol{\phi}_3^T \end{bmatrix} \tag{A.7}$$

Considering the approximations followed Eq. (A.4), we show that:

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T & \boldsymbol{\theta} \\ \boldsymbol{\theta}^T & \omega^2 \end{bmatrix} \tag{A.8}$$

and

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \mathbf{M} \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T \boldsymbol{\Sigma} \hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \boldsymbol{\Sigma} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T & \boldsymbol{\psi} \\ \boldsymbol{\psi}^T & \omega^4 \end{bmatrix} \tag{A.9}$$

where $\boldsymbol{\theta} \in \mathbb{R}^m$ and $\boldsymbol{\psi} \in \mathbb{R}^m$. From Eqs. (A.5), (A.8) and (A.9), the F criterion can be expressed as follows:

$$F(\hbar \le q) \approx 1 + \sum_{i=1}^{m} \frac{\boldsymbol{\xi}_i^T \left[\hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T \boldsymbol{\Sigma} \hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \boldsymbol{\Sigma} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\right] \boldsymbol{\xi}_i}{\left(\boldsymbol{\xi}_i^T \left[\hat{\mathbf{U}}_2\hat{\mathbf{U}}_2^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\right] \boldsymbol{\xi}_i\right)^2} \tag{A.10}$$

When $\ell = \hbar$, the residual matrix $\tilde{\mathbf{P}}$ consists of the last $(q-\hbar)$ eigenvectors of $\hat{\mathbf{U}}$ and all the eigenvectors of $\tilde{\mathbf{U}}$:

$$\tilde{\mathbf{P}} = \begin{bmatrix} \hat{\mathbf{U}}_2 & \tilde{\mathbf{U}} \end{bmatrix} \tag{A.11}$$

Since $\hat{\mathbf{U}}_2$ and $\tilde{\mathbf{U}}$ are orthogonal matrices, we can show that the second term in the right side of Eq. (A.10) represents the expression of the classical VRE criterion for $\ell = \hbar \le q$:

$$F(\hbar \le q) \approx 1 + \sum_{i=1}^{m} \frac{\boldsymbol{\xi}_i^T \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \boldsymbol{\Sigma} \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \boldsymbol{\xi}_i}{\left(\boldsymbol{\xi}_i^T \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T \boldsymbol{\xi}_i\right)^2} = 1 + \sum_{j=1}^{m} \frac{\boldsymbol{\xi}_j^T \tilde{\mathbf{C}} \boldsymbol{\Sigma} \tilde{\mathbf{C}} \boldsymbol{\xi}_j}{\left(\boldsymbol{\xi}_j^T \tilde{\mathbf{C}} \boldsymbol{\xi}_j\right)^2} \tag{A.12}$$

hence

$$F(\hbar \le q) \approx 1 + \mathrm{VRE}(\ell = \hbar) \tag{A.13}$$

A.2. Case of $\hbar = q + 1$

From Eq. (A.7), the matrices $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$ will have the following forms, respectively:

$$\tilde{\mathbf{V}} = \begin{bmatrix} \tilde{\mathbf{U}} \\ \boldsymbol{\phi}_3^T \end{bmatrix} \tag{A.14}$$

and

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T & \tilde{\mathbf{U}}\boldsymbol{\phi}_3 \\ \boldsymbol{\phi}_3^T\tilde{\mathbf{U}}^T & \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 \end{bmatrix} \tag{A.15}$$

Since $\tilde{\mathbf{U}}$ is an orthonormal matrix, i.e. $\tilde{\mathbf{U}}^T\tilde{\mathbf{U}} = \mathbf{I}_{(m-q)}$, we can then deduce the following:

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\boldsymbol{\phi}_3\boldsymbol{\phi}_3^T\tilde{\mathbf{U}}^T & \tilde{\mathbf{U}}\boldsymbol{\phi}_3 + \tilde{\mathbf{U}}\boldsymbol{\phi}_3\boldsymbol{\phi}_3^T\tilde{\mathbf{U}}^T \\ \boldsymbol{\phi}_3^T\tilde{\mathbf{U}}^T + \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3\boldsymbol{\phi}_3^T\tilde{\mathbf{U}}^T & \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 + (\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^2 \end{bmatrix} \tag{A.16}$$

Notice that $\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T$ should be an idempotent matrix, i.e.

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T = \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \tag{A.17}$$

The identification by analogy between the terms of Eqs. (A.15) and (A.16) leads to conclude that $\tilde{\mathbf{U}}\boldsymbol{\phi}_3$ is a zero vector. In addition, $\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3$ is a scalar that is close to zero. This implies that $(\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^2$ is perfectly negligible compared to $\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3$. Further, $\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 \approx \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 + (\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^2$. These deductions may allow to write:

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\mathbf{M}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\boldsymbol{\Sigma}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T & \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\boldsymbol{\epsilon}\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 \\ \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3\boldsymbol{\epsilon}^T\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T & \boldsymbol{\phi}_3^T\boldsymbol{\phi}_3 + (\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^2 \end{bmatrix} \tag{A.18}$$

From Eqs. (A.5), (A.15) and (A.18), we can deduce that:

$$F(\hbar = q+1) \approx 1 + \frac{1}{\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3} + \sum_{i=1}^{m} \frac{\boldsymbol{\xi}_i^T\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\boldsymbol{\Sigma}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\boldsymbol{\xi}_i}{(\boldsymbol{\xi}_i^T\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\boldsymbol{\xi}_i)^2} \tag{A.19}$$

It is clear that the third term in the right side of Eq. (A.19) corresponds to the expression of the VRE criterion for $\ell = q$ PCs. The previous equation can be rewritten as follows:

$$F(\hbar = q+1) \approx 1 + \frac{1}{\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3} + \mathrm{VRE}(\ell = q) \tag{A.20}$$

Notice that $(\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^{-1}$ tends to infinity.

A.3. Case of $\hbar > q + 1$

This case is illustrated by a residual matrix $\tilde{\mathbf{V}}$ which is formed by the latest $(m+1-\hbar)$ columns of the matrix $\mathbf{V}$ given by Eq. (A.4). An appropriate partition of its last block yields the following:

$$\begin{bmatrix} \tilde{\mathbf{U}} \\ \boldsymbol{\phi}_3^T \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{U}}_1 & \tilde{\mathbf{U}}_2 \\ \boldsymbol{\phi}_{13}^T & \boldsymbol{\phi}_{23}^T \end{bmatrix} \tag{A.21}$$

This enables to deduce that:

$$\tilde{\mathbf{V}} = \begin{bmatrix} \tilde{\mathbf{U}}_2 \\ \boldsymbol{\phi}_{23}^T \end{bmatrix} \tag{A.22}$$

where $\tilde{\mathbf{U}}_2 \in \mathbb{R}^{m\times(m+1-\hbar)}$ consists of the last $(m+1-\hbar)$ eigenvectors of $\mathbf{P}$. The vector $\boldsymbol{\phi}_{23} \in \mathbb{R}^{(m+1-\hbar)}$ is constituted by the latest $(m+1-\hbar)$ components of the vector $\boldsymbol{\phi}_3$. By applying the necessary approximations as in the second case, we can show that:

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T & \tilde{\mathbf{U}}_2\boldsymbol{\phi}_{23} \\ \boldsymbol{\phi}_{23}^T\tilde{\mathbf{U}}_2^T & \boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23} \end{bmatrix} \tag{A.23}$$

and

$$\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T\mathbf{M}\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \approx \begin{bmatrix} \tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T\boldsymbol{\Sigma}\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T & \tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T\boldsymbol{\epsilon}\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23} \\ \boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23}\boldsymbol{\epsilon}^T\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T & \boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23} + (\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23})^2 \end{bmatrix} \tag{A.24}$$

The F criterion can then be written as follows:

$$F(\hbar > q+1) \approx 1 + \frac{1}{\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23}} + \sum_{i=1}^{m} \frac{\boldsymbol{\xi}_i^T\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T\boldsymbol{\Sigma}\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T\boldsymbol{\xi}_i}{(\boldsymbol{\xi}_i^T\tilde{\mathbf{U}}_2\tilde{\mathbf{U}}_2^T\boldsymbol{\xi}_i)^2} \tag{A.25}$$

The third term of the right side of Eq. (A.25) corresponds to the classical VRE criterion for $\ell = (\hbar - 1)$ PCs, whence

$$F(\hbar > q+1) \approx 1 + \frac{1}{\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23}} + \mathrm{VRE}(\ell = \hbar - 1) \tag{A.26}$$

where $(\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23})^{-1}$ is monotonically increasing with respect to $\hbar$ and tends to infinity.

As a result, the combination of Eqs. (A.13), (A.20) and (A.26) shows that the F criterion can be expressed as follows:

$$F(\hbar) \approx 1 + \begin{cases} \mathrm{VRE}(\hbar) & \text{if} \quad \hbar \le q \\ (\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^{-1} + \mathrm{VRE}(q) & \text{if} \quad \hbar = q+1 \\ (\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23})^{-1} + \mathrm{VRE}(\hbar - 1) & \text{if} \quad \hbar > q+1 \end{cases} \tag{A.27}$$

From Eq. (A.21), it is clear that $(\boldsymbol{\phi}_{23}^T\boldsymbol{\phi}_{23})^{-1} \ge (\boldsymbol{\phi}_3^T\boldsymbol{\phi}_3)^{-1}$. Using Eq. (A.1) from which we have assumed that the minimum of the VRE is reached at q PCs, we show therefore that:

$$\underset{\hbar}{\mathrm{argmin}}\left\{F(\hbar)\right\} = \underset{\ell}{\mathrm{argmin}}\left\{\mathrm{VRE}(\ell)\right\} = q \tag{A.28}$$

References

[1] T. Kourti, Process analysis and abnormal situation detection: from theory to practice, IEEE Control Syst. 22 (2002) 10–25.
[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, A review of process fault detection and diagnosis. Part I: Quantitative model-based methods, Comput. Chem. Eng. 27 (2003) 293–311.
[3] S.J. Qin, Statistical process monitoring: basics and beyond, J. Chemom. 17 (2003) 480–502.
[4] M.-F. Harkat, G. Mourot, J. Ragot, An improved PCA scheme for sensor FDI: application to an air quality monitoring network, J. Process Control 16 (2006) 625–634.
[5] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE J. 40 (1994) 1361–1375.
[6] M. Wasim, R.G. Brereton, Determination of the number of significant components in liquid chromatography nuclear magnetic resonance spectroscopy, Chemom. Intell. Lab. Syst. 72 (2004) 133–151.
[7] M. Tamura, S. Tsujita, A study on the number of principal components and sensitivity of fault detection using PCA, Comput. Chem. Eng. 31 (2007) 1035–1046.
[8] J. Toft, Evolutionary rank analysis applied to multidetectional chromatographic structures, Chemom. Intell. Lab. Syst. 29 (1995) 189–212.
[9] L. Ferré, Selection of components in principal component analysis: a comparison of methods, Comput. Stat. Data Anal. 19 (1995) 669–682.
[10] S. Valle, W. Li, S.J. Qin, Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods, Ind. Eng. Chem. Res. 38 (1999) 4389–4401.
[11] I.T. Jolliffe, Principal Component Analysis, Springer Series in Statistics, Springer, New York, 2002.
[12] U. Kruger, L. Xie, Advances in Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control, Statistics in Practice, Wiley, Chichester, 2012.
[13] J.E. Jackson, A User's Guide to Principal Components, Wiley Series in Probability and Statistics, Wiley, New York, 2002.
[14] M. Wax, T. Kailath, Detection of signals by information theoretic criteria, IEEE Trans. Acoust. Speech Signal Process. 33 (1985) 387–392.

[15] R.B. Cattell, The scree test for the number of factors, Multivar. Behav. Res. 1 (1966) 245–276.
[16] H.F. Kaiser, A note on Guttman's lower bound for the number of common factors, Br. J. Stat. Psychol. 14 (1961) 1–2.
[17] E.R. Malinowski, Determination of the number of factors and the experimental error in a data matrix, Anal. Chem. 49 (1977) 612–617.
[18] R.I. Shrager, R.W. Hendler, Titration of individual components in a mixture with resolution of difference spectra, pKs, and redox transitions, Anal. Chem. 54 (1982) 1147–1152.
[19] M.S. Bartlett, A note on the multiplying factors for various χ^2 approximations, J. R. Stat. Soc. Ser. B Methodol. 16 (1954) 296–298.
[20] D.N. Lawley, Tests of significance for the latent roots of covariance and correlation matrices, Biometrika 43 (1956) 128–136.
[21] T. Feital, U. Kruger, L. Xie, U. Schubert, E.L. Lima, J.C. Pinto, A unified statistical framework for monitoring multivariate systems with unknown source and error signals, Chemom. Intell. Lab. Syst. 104 (2010) 223–232.
[22] S. Wold, Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20 (1978) 397–405.
[23] H.T. Eastment, W.J. Krzanowski, Cross-validatory choice of the number of components from a principal component analysis, Technometrics 24 (1982) 73–77.
[24] P. Besse, PCA stability and choice of dimensionality, Stat. Prob. Lett. 13 (1992) 405–410.
[25] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory, Tsahkadsor, Armenian SSR, 1973, pp. 267–281.
[26] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[27] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[28] H. Wang, H. Zhou, B. Hang, Number selection of principal components with optimized process monitoring performance, in: 43th IEEE Conference on Decision and Control, Paradise Island, Bahamas, 2004, pp. 4726–4731.
[29] R. Dunia, S.J. Qin, A unified geometric approach to process and sensor fault identification and reconstruction: the unidimensional fault case, Comput. Chem. Eng. 22 (1998) 927–943.
[30] S.J. Qin, R. Dunia, Determining the number of principal components for best reconstruction, J. Process Control 10 (2000) 245–250.
[31] S.J. Qin, Survey on data-driven industrial process monitoring and diagnosis, Annu. Rev. Control 36 (2012) 220–234.
[32] W. Li, S.J. Qin, Consistent dynamic PCA based on errors-in-variables subspace identification, J. Process Control 11 (2001) 661–678.
[33] T.W. Anderson, Asymptotic theory for principal component analysis, Ann. Math. Stat. 34 (1963) 122–148.
[34] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley Series in Probability and Statistics, Wiley, New York, 2003.
[35] S. Narasimhan, S.L. Shah, Model identification and error covariance matrix estimation from noisy data using PCA, Control Eng. Pract. 16 (2008) 146–155.
[36] B. Mnassri, E.M. El Adel, B. Ananou, M. Ouladsine, A generalized variance of reconstruction error criterion for determining the optimum number of principal components, in: 18th IEEE Mediterranean Conference on Control and Automation, Marrakech, Morocco, 2010, pp. 868–873.
[37] B. Mnassri, E.M. El Adel, B. Ananou, M. Ouladsine, Selection of the number of principal components based on the fault reconstruction approach applied to a new combined index, in: 49th IEEE Conference on Decision and Control, Atlanta, GA, USA, 2010, pp. 3307–3312.
[38] B. Mnassri, E.M. El Adel, M. Ouladsine, Inverse-variance weighting PCA-based VRE criterion to select the optimal number of PCs, in: 18th IFAC World Congress, Milano, Italy, 2011, pp. 2851–2856.
[39] S. Yoon, J.F. MacGregor, Fault diagnosis with multivariate statistical models part I: using steady state fault signatures, J. Process Control 11 (2001) 387–400.