

# New PCA-based scheme for process fault detection and identification. Application to the Tennessee Eastman Process

Mohamed Guerfel<sup>1,2\*</sup>, Anissa Ben Aicha<sup>3,\*\*</sup>, Kamel Belkhiria<sup>3,\*\*\*</sup>, Hassani Messaoud<sup>2\*\*\*\*</sup>

<sup>1</sup> Higher Institute of Applied Science and technology of Sousse (ISSATSo), University of Sousse, Tunisia

<sup>2</sup> LARATSI, National Engineering School of Monastir (ENIM), University of Monastir, Tunisia

<sup>3</sup> LAS2E, National Engineering School of Monastir (ENIM), University of Monastir, Tunisia

**Abstract.** This paper proposes a new Principal Component Analysis (PCA) scheme to perform fault detection and identification (FDI) for systems affected by process faults. In this scheme, a new modeling method which maximizes the model sensitivity to a certain process fault type is proposed. This method uses normal operating or known faulty data to build the PCA model and other faulty data to fix its structure. A new structuration method is proposed to identify the process fault. This method computes the common angles between the residual subspaces of the different modes. It generates a reduced set of detection indices that are sensitive to certain process fault and insensitive to others. The proposed FDI scheme is successfully applied to the Tennessee Eastman Process (TEP) supposedly affected by several process faults.

**Key words:** PCA Modeling, Process Fault, FDI, Principal Angles, Structured Residuals, Tennessee Eastman Process

## 1. INTRODUCTION

The use of data based FDI methods are becoming increasingly valuable assets to overcome the growth of complexity in high scale industrial processes [1, 2, 3]. These methods excel at extracting valuable insights from historical data and computing relationships between its variables. Such insights play a crucial role in avoiding undesirable events by enabling timely and accurate decision-making which ensure smooth and reliable industrial process operations [4, 5].

PCA gained popularity as a data-driven technique in FDI due to its simplicity, efficiency in detecting abnormal operating conditions and aptitude in identifying malfunctioning components [6]. It finds successful applications in various industrial fields [7, 1] and transforms the original process variables into a new ones, linear combinations of the original, occupying a lower dimensional subspace.

Industrial operations often undergo frequent alterations attributed to factors such as setpoint adjustments, fluctuations in raw materials, variations in feed material composition, equipment aging and process modifications [8, 9]. These alterations lead to changes in process parameters, issues with actuators, and sensor degradation. Parameter changes in operations arise from either changing operating conditions or process faults [3, 10] which represent the studied fault type in this paper.

PCA-based FDI methods have been extensively studied in processes affected by actuator and sensor faults [11, 7]. However, the focus on the application of PCA to recognize different process varying operating modes and identify its parameter changes is a novelty in the existing research [6, 12]. The exploration of PCA's potential in handling such scenarios holds

promise for enhancing fault detection and diagnosis capabilities in real industrial processes [13, 11, 14].

This paper introduces a novel PCA-based approach for detecting and identifying alterations in process parameters. An outlined modeling technique designed to maximize the detection of certain faulty process data is proposed. A new strategy of structuration of the detection indices is presented. It revolves around investigating the angles among residual space vectors from distinct process operating modes. It involves splitting the residual space vectors of each mode into two subsets. The subset common to the studied modes aids in generating indices that are sensitive to specific modes while remaining insensitive to others. The whole proposed PCA-based scheme for modeling and process fault identification is successfully applied to the Tennessee Eastman Process [15, 16].

The organization of the paper is as following : section 2 presents the principle, the existing and the proposed modeling method in PCA. Section 3 presents the principal angles theory and the proposed structuring method. Section 4 illustrates a study of the TEP and demonstrates the proposed modeling and structuring method efficiency in detection and identification of different operating modes. Conclusions are given in Section 5.

## 2. PRINCIPAL COMPONENT ANALYSIS

This section presents the formulation of PCA, some of its most used modeling techniques and a proposed method of modeling based on maximizing the detection of a known process fault.

### 2.1. PCA formulation

In order to perform a PCA, all the inputs and outputs observed in the instant  $k$  are regrouped into column data vector  $z^r(k)$  :

$$z^r(k) = [z_1^r(k) \ z_2^r(k) \ \dots \ z_m^r(k)]^t \quad (1)$$

After that,  $z^r(k) \in \mathbb{R}^m$  is scaled to zero mean and unity variance in order to obtain the vector  $z(k) = \begin{bmatrix} z_1(k) & \dots & z_m(k) \end{bmatrix}^t$ .

\*e-mail: guerfel\_mohamed@yahoo.fr

\*\*e-mail: benaicha.anissa@gmail.com

\*\*\*e-mail: kamelbenothman@yahoo.fr

\*\*\*\*e-mail: hassani.messaoud@enim.rnu.tn

The data matrix  $Z$  resulting from the juxtaposition of  $z(k)$  in different instants is written :

$$Z = \begin{bmatrix} z(k) & \dots & z(k+N-1) \end{bmatrix}^T \quad (2)$$

The subscript  $N$  designates the number of observations used in the construction of the matrix  $Z$ .

The correlation matrix  $\Sigma \in \mathbb{R}^{m \times m}$  identifies the relationships among process variables. It is obtained from  $Z$  by using :

$$\Sigma = \frac{1}{N-1} \cdot Z^T Z \quad (3)$$

This matrix can always be decomposed as following :

$$\Sigma = P \Lambda P^t \quad ; \quad P P^t = P^t P = I_m \quad (4)$$

The notation  $\Lambda = \text{diag}(\lambda_1 \dots \lambda_m)$  designates the diagonal matrix of eigenvalues.

The matrix  $\Sigma$  is symmetric and semidefinite positive so its eigenvalues  $\lambda_q$  are positive reals and can always be arranged according to decreasing magnitude  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ .

The notation  $P \in \mathbb{R}^{m \times m}$  represents the matrix containing the eigenvector  $p_q \in \mathbb{R}^m$ .

Modeling a process via static PCA consists in seeking an optimal linear transformation (with respect to a variance criterion) of the original data matrix  $Z$  into a new one called  $T$  and defined as follows :

$$T = Z \times P \quad ; \quad T = \begin{bmatrix} t_1 & \dots & t_m \end{bmatrix} \in \mathbb{R}^{N \times m} \quad (5)$$

The vectors  $t_q \in \mathbb{R}^N$ ,  $q \in \{1, \dots, m\}$ , called principal components are uncorrelated and arranged in the decreasing variance order [17]. This guarantees that all the process variations are contained in the few first principal components, thus permitting the reduction of variables number for process description. It was demonstrated [17] that the column vectors  $p_q$  of  $P$  represent the eigenvectors associated to the eigenvalues  $\lambda_q$  obtained from the diagonalization of  $\Sigma$ . The matrices  $\Lambda$ ,  $P$  and  $T$  can always be partitioned as follows :

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix}, \quad P = \begin{bmatrix} \hat{P} & \tilde{P} \end{bmatrix}, \quad T = \begin{bmatrix} \hat{T} & \tilde{T} \end{bmatrix} \quad (6)$$

The matrices  $\hat{\Lambda} \in \mathbb{R}^{\ell \times \ell}$ ,  $\hat{P} \in \mathbb{R}^{m \times \ell}$  and  $\hat{T} \in \mathbb{R}^{N \times \ell}$  are associated to the modeled part of process variance. Whereas  $\tilde{\Lambda} \in \mathbb{R}^{(m-\ell) \times (m-\ell)}$ ,  $\tilde{P} \in \mathbb{R}^{m \times (m-\ell)}$  and  $\tilde{T} \in \mathbb{R}^{N \times (m-\ell)}$  are linked to the non modeled part.

The data matrix can be decomposed in the following form:

$$Z = \hat{Z} + E \quad \text{with} \quad \hat{Z} = Z \hat{C} \quad ; \quad E = Z \tilde{C} \quad (7)$$

The matrices  $\hat{C} = \hat{P} \hat{P}^t$  and  $\tilde{C} = I_m - \hat{C}$  form the static PCA model of the process [17]. The matrices  $\hat{Z}$  and  $E$  represent, respectively, the modeled and the non modeled variations of  $Z$  from  $\ell$  components ( $\ell < m$ ). The first  $\ell$  eigenvectors forming the matrix  $\hat{P}$  constitute the representation space whereas the last  $(m - \ell)$  eigenvectors forming the matrix  $\tilde{P}$  constitute the residual space [18].

## 2.2. PCA modeling

The identification of the PCA model thus consists of estimating its parameters by an eigenvalue/eigenvector decomposition of the matrix  $\Sigma$  and determining its structural parameter which is the number of principal components  $\ell$  to retain [4].

**2.2.1. Heuristic methods** These methods are often ambiguous and do not present an optimum with regards to the retained number  $\ell$  [19].

**CPV** One chooses a percentage of the variance to retain in the model and keeps the first  $\ell$  components for which the cumulative percentage of variance  $CPV(\ell)$  is exceeded.

$$CPV(\ell) = 100 \left( \frac{\sum_{j=1}^{\ell} \lambda_j}{\sum_{q=1}^m \lambda_q} \right) \% \quad (8)$$

**Eigenvalues mean** This method chooses the components whose the eigenvalue is superior to the arithmetical mean of all values (which is equal to 1).

**Catell test** The Catell test also known as elbow method looks for the "elbow" in the screeplot and selects all components whose eigenvalues are just before the curve flattens out.

**2.2.2. Reconstruction error based method** Some works proposed to fix  $\ell$  via the minimization of the variance of the reconstruction error of the variables called VRE [18]. The non reconstructed variance of the  $j^{\text{th}}$  component of  $z$  is noted  $\rho_j$  :

$$\rho_j = \frac{\tilde{\xi}_j^t \tilde{\Sigma} \tilde{\xi}_j}{\left( \tilde{\xi}_j^t \tilde{\xi}_j \right)^2} \quad (9)$$

Where  $\tilde{\Sigma} = \tilde{C} \Sigma \tilde{C}$  and  $\tilde{\xi}_j = \tilde{C} \xi_j$ . The vector  $\xi_j$  designates the  $j^{\text{th}}$  column of the identity matrix  $I_m \in \mathbb{R}^{m \times m}$ .

It is noted that  $\rho_j$  depends on  $\tilde{C}$  which in its turn depends on  $\ell$  used to fix the PCA model structure [8]. The number  $\ell$  resulting from the VRE criterion will be the one that minimizes the sum  $\Phi$  of the error variance of all the components of  $z$ .

$$\Phi = \min_{\ell} \sum_{j=1}^m \rho_j(\ell) \quad (10)$$

The VRE criterion is convex, it has always a minimum but underestimates the exact  $\ell$  to retain in real applications [19].

**2.2.3. Proposed Modeling Method** All the methods aiming at the determination of  $\ell$  seek to find its theoretical or exact value called  $(\ell_{\text{th}})$  which represents the theoretical number of linear or quasi-linear relations existing between the different components of  $z(k)$  [4]. In the case of models built for diagnosis purposes,  $\ell$  can be different from  $\ell_{\text{th}}$  provided that the resulting PCA model can detect changes [19, 20].

The proposed modeling method uses two dataset matrices  $Z_0 \in \mathbb{R}^{a \times m}$  to obtain PCA model and  $Z_1 \in \mathbb{R}^{b \times m}$  to fix its structure. It chooses  $\ell$  that maximizes the mean value of the ratio  $\frac{SPE}{\delta^2}$  computed for all the data contained in  $Z_1$ .

It is noted that the squared prediction error called  $SPE$ , depends on the number of retained principal components  $\ell$ .

For every sample  $k$ , its expression is given as follows :

$$SPE(k) = \sum_{h=m-\ell+1}^m t_h^2(k) = \left\| \tilde{P}_0^T z_1(k) \right\|^2 \quad (11)$$

The quantity  $\delta^2$  designates the detection threshold which can be computed in a theoretical manner [17] or by training on a real dataset [19].

The algorithm of the proposed modeling method is given in 1. The used symbol "o" represents the hadamard product, a mathematical operation that generates a new matrix by multiplying corresponding elements of two matrices element-wise.

The flow chart of the proposed method is given in figure 1.

### Algorithm 1 Proposed Modeling Method

**Require:** Data matrices  $Z_0^r, Z_1^r$

**Ensure:** Principal components number  $\ell$

- 1:  $j \leftarrow 0$  and  $SPE_0 \leftarrow 0$
- 2: Compute  $m_0, \sigma_0, Z_0, \Lambda_0$  and  $P_0$
- 3: Center, reduce  $Z_1^r$  with  $m_0, \sigma_0$ , obtain  $Z_1$
- 4: Compute  $T_1 = Z_1 \times P_0$  and  $D_1 = T_1 \circ T_1$
- 5: **for**  $i=1:m$  **do**
- 6:  $V_1(i) = b^{-1} \times \text{sum } M_1(:, i)$
- 7: **end for**
- 8: **while**  $j < m$  **do**
- 9:  $SPE_{j+1} = \text{sum } D_1(m-j:m)$
- 10: **if**  $\frac{SPE_{j+1}}{\delta_{(m-j+1), \alpha}^2} > \frac{SPE_j}{\delta_{(m-j), \alpha}^2}$  **then**
- 11:  $\ell \leftarrow (m-j-1)$
- 12: **else**
- 13:  $\ell \leftarrow \ell$
- 14: **end if**
- 15:  $j \leftarrow (j+1)$
- 16: **end while**

The limitation of this approach stems from the requirement of prior knowledge about the process fault, ensuring the selection of components to retain in the PCA model. Conversely, with insight into the faulty mode, the method becomes appealing by identifying the most responsive model to the process fault [19].

### 3. IDENTIFICATION OF CHANGES IN PROCESS OPERATIONS

Industrial processes have three types of operating modes : the steady state mode, can be simple or multiple and represents a correct operating mode. The transient state mode that regroup the grade transition and the start-up mode. It is characterized by a nonlinear behavior [5, 21]. The faulty mode, can be simple or multiple and represents an undesirable mode [21, 22]. It can be caused by a sensor/actuator fault or by a modification in process parameters known as process fault [23, 10]. In this work, the transient stage is supposed to be of very short duration, and as such, its identification and monitoring will not be considered. From a practical standpoint, it is crucial to distinguish between changes in the operating mode and process faults, even though both have a similar impact on the covariance structure [3, 12]. Therefore, a method is required to extract features from the covariance change pattern [14]. In this context, one propose utilizing the angles between subspaces to identify common and distinct subspaces for each operating mode, whether faulty or not. These different spaces are then

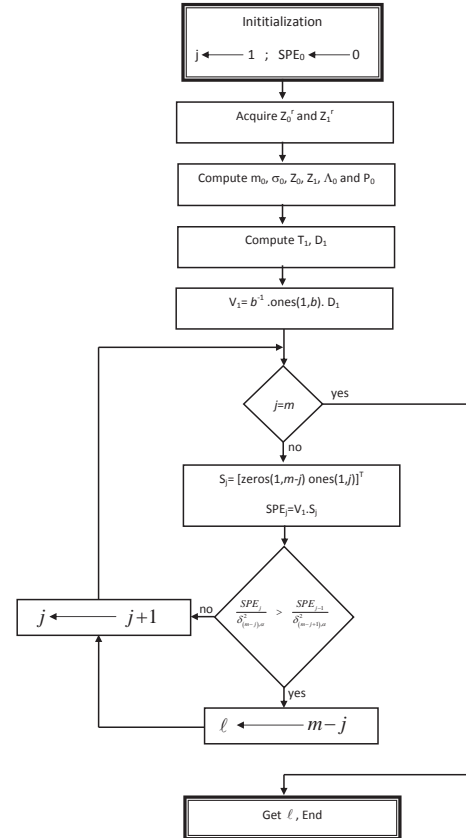


Fig. 1. Flow chart of the proposed modeling method

employed to compute a structured detection index that is sensitive to certain modes and insensitive to others. The number of detection indices needed in this work is lesser than the one needed in the classical change identification that require a separate PCA model for each type of operating mode [20, 9, 24]. This section is divided into two parts, the first one presents the angles between the subspaces and the second one explains the proposed structuration method.

#### 3.1. Angles between subspaces

Let  $A_1 \in \mathbb{R}^{m \times p}$  and  $A_2 \in \mathbb{R}^{m \times r}$  be two matrices which columns form orthonormal basis of two subspaces  $S_1$  and  $S_2$ . One assumes in the following that  $p \leq r$ . In this case, it exists  $p$  principal angles, called  $\theta_s, s \in \{1, \dots, p\}$ , between  $S_1$  and  $S_2$ . The angles  $\theta_1, \theta_p$  represents the min and max values [25]. Figure 2 illustrates the principal angles between two plans  $S_1$  et  $S_2$ . The min angle  $\theta_1$  is nul and the max is  $\theta_2$ . The vectors  $(u_1, u_2)$  and  $(v_1, v_2)$  are the directors of the two subspaces.

The computation of principal angles and director vectors is in-

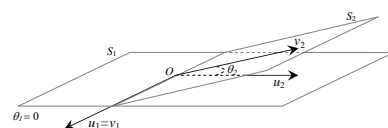


Fig. 2. Angle  $\theta_2$  between the subspaces  $S_1$  and  $S_2$

sured with a singular value decomposition (SVD) of the  $A_2^T A_1$

matrix [25]:

$$A_2^T A_1 = L S F^T \quad (12)$$

All the columns of  $L \in \mathbb{R}^{r \times p}$  and  $F \in \mathbb{R}^{p \times p}$  are orthonormal. The matrices  $U$  et  $V$  containing the director vectors of  $S_1$  et  $S_2$  are given :

$$U = A_1 F \quad ; \quad V = A_2 L \quad (13)$$

The matrix  $S$  is diagonal and written as follows :

$$S = \text{diag} \left( \rho_1 \quad \dots \quad \rho_p \right) \quad (14)$$

with  $1 \geq \rho_1 \geq \dots \geq \rho_p$ .

The value of the  $s^{\text{th}}$  angle  $\theta_s$  can be computed :

$$\theta_s = (\widehat{u_s, v_s}) = \arccos(\rho_s) \quad (15)$$

The vectors  $u_s$  et  $v_s$  designates the  $s^{\text{th}}$  columns of the matrices  $U$  and  $V$  [22].

### 3.2. Proposed structuration method

Structured residuals, selectively sensitive to specific subsets of faults from the complete PCA model, can be generated using concepts derived from parity relations [20, 13]. This generation process is feasible only in cases involving simple or multiple sensor faults. However, it is not achievable when dealing with process faults that propagate across all variables. In such instances, residual structuring is substituted with a technique relying on partial PCA models [24]. By generating a fault detection index for each mode, it becomes possible to identify distinct faulty modes effectively [12]. In this part, a novel method is proposed to generate a reduced set of structured fault detection indices for systems affected by process faults. These indices, whose quantity is fewer than the number of system modes, enable the accurate identification of known faulty modes and the recognition of a new unidentified mode. Suppose that a process have  $M_i$  known operating modes,  $i \in \{0 \dots n-2\}$  and an unknown mode called  $M_{n-1}$ . Here one assumes that  $n$  is even, otherwise it have to be replaced with  $n+1$ .

Let  $X_g \in \mathbb{R}^{\frac{n}{2}}$  be the mode vector defined as follows :

$$X_g = \left( g \quad g+1 \quad \dots \quad g + \frac{n}{2} - 1 \right)^T \quad (16)$$

The subscript  $g = 0, \dots, \frac{n}{2} - 1$ , is the mode vector number.

One defines the mode state  $In_{X_g}$  as a boolean variable equal to zero if the current operating mode is equal to any of the components of  $X_g$  and equal to one otherwise. From this mode state and the different operating modes  $M_i$ , one can construct a localization table (Lt) where the rows are constituted by  $In_{X_g}$  and the columns correspond to  $M_i$ . Table 1 presents an example of localization table for process with four operating modes.

The mode state  $In_{X_g}$  can be linked to a detection index denoted as  $SPE_{X_g}$ . This index will remain below a specified threshold  $\delta_{X_g}^2$  when the current operating mode matches any of the components of  $X_g$  ; otherwise, it will surpass its threshold. For every sample  $k$ , the index  $SPE_{X_g}$  can be computed as follows :

$$SPE_{X_g}(k) = \left\| \tilde{P}_{X_g}^T z(k) \right\|^2 \quad (17)$$

**Table 1.** Localization table in case of four actif modes

| mode       | $M_0$ | $M_1$ | $M_2$ | $M_3$ |
|------------|-------|-------|-------|-------|
| $In_{X_0}$ | 0     | 0     | 1     | 1     |
| $In_{X_1}$ | 1     | 0     | 0     | 1     |

The number of the structured indices  $SPE_{X_g}$ , is equal to  $\frac{n}{2}$ .

The row vectors of  $\tilde{P}_{X_g}$  represents the vector of residual spaces common to all the modes which components form the mode vector  $X_g$ . Let's call  $X_g^{[r]} \in \mathbb{R}^r$  a vector formed with the first  $r$  components of  $X_g$ . It is noted that for  $r = \frac{n}{2}$ , one have  $X_g^{[\frac{n}{2}]} = X_g$ .

In order to obtain  $\tilde{P}_{X_g}$ , one have to compute the different matrices  $\tilde{P}_{X_g^{[r]}}$ ,  $r = 2, \dots, \frac{n}{2}$ . Each matrix  $\tilde{P}_{X_g^{[r]}}$  is obtained in a recursive manner from  $\tilde{P}_{X_g^{[r-1]}}$  as following :

In a first time, one computes  $R_{X_g^{[r]}}$  and perform an SVD on it to obtain the matrices  $U_{X_g^{[r]}}$ ,  $V_{X_g^{[r]}}$  and  $S_{X_g^{[r]}}$  like follows :

$$R_{X_g^{[r]}} = \tilde{P}_{(g+r-1)}^T \cdot \tilde{P}_{X_g^{[r-1]}} = U_{X_g^{[r]}} S_{X_g^{[r]}} V_{X_g^{[r]}}^T \quad (18)$$

The matrix  $\tilde{P}_{(g+r-1)}^T$  designates the residual matrix of the  $(g+r-1)^{\text{th}}$  operating mode.

After that, one calculates the matrix  $\tilde{Q}_{X_g^{[r]}}$  as following :

$$\tilde{Q}_{X_g^{[r]}} = \tilde{P}_{X_g^{[r-1]}} \cdot V_{X_g^{[r]}} \quad (19)$$

The matrix  $\tilde{P}_{X_g^{[r]}}$  is formed with the first  $q$  vectors of  $\tilde{Q}_{X_g^{[r]}}$  corresponding to the  $q$  unity eigenvalues of  $S_{X_g^{[r]}}$ .

After the computation of  $SPE_{X_g}(k)$ , a boolean indicator  $In_{X_g}(k)$  is build in the following way :

$$In_{X_g}(k) = \begin{cases} 0 & \text{if } SPE_{X_g}(k) < \delta_{X_g}^2 \\ 1 & \text{if } SPE_{X_g}(k) > \delta_{X_g}^2 \end{cases} \quad (20)$$

The term  $\delta_{X_g}^2$  designates the detection threshold of  $SPE_{X_g}$ . The localization procedure is carried out by comparing the values of all  $In_{X_g}(k)$  to the columns of (Lt).

In order to ensure the obtention of the whole structured residuals, the vectorial space resulting from each matrix  $\tilde{P}_{X_g}$  has to be different. This can be ensured if the principal angles between the different matrices  $\tilde{P}_{X_g}$  are not null.

The algorithm 2 illustrates the different steps used to compute the structured residuals and their use in the identification of known faulty operating mode and the recognition of a new unknown mode. The notations  $m_g$  and  $\sigma_g$  designate the mean and standard deviation vectors of the  $g^{\text{th}}$  mode data matrix called  $Z_g^r$ . They will be used to build the normalized  $g^{\text{th}}$  mode data matrix  $Z_g$ . The flow chart presented in figure 3 shows a graphical visualization of the proposed structuration method.

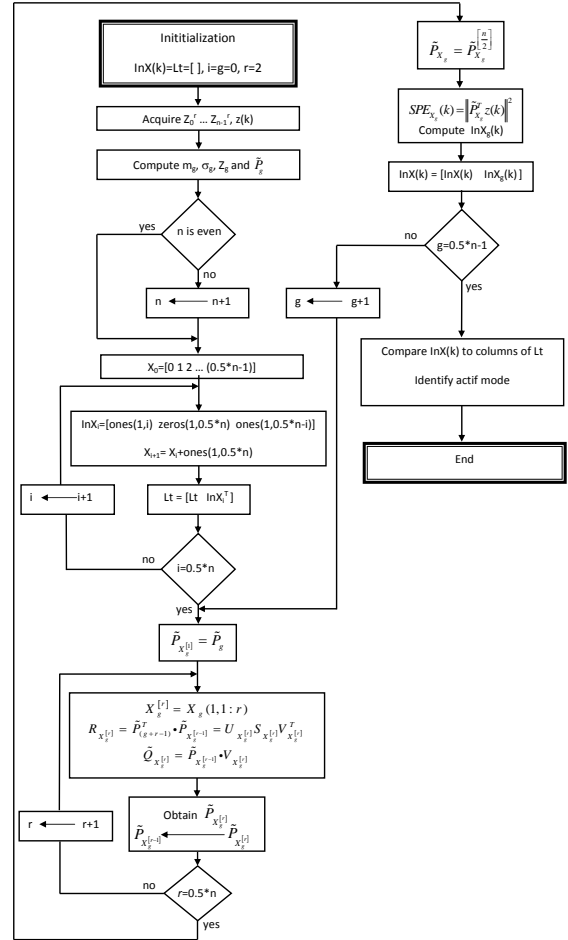
## 4. APPLICATION TO THE TENNESSEE EASTMAN PROCESS

The Tennessee Eastman Process (TEP), illustrated in Figure 4, is a chemical process created by the Eastman Chemical Company. It's an emulation of an actual industrial process and

**Algorithm 2** Proposed Structuration Method**Require:** Data  $Z_0^r, \dots, Z_{n-1}^r, z(k)$ **Ensure:** Indices  $SPE_{X_0}(k), \dots, SPE_{X_{\frac{n}{2}-1}}(k)$  and  $In_{X_0}(k), \dots,$ 

$In_{X_{\frac{n}{2}-1}}(k)$

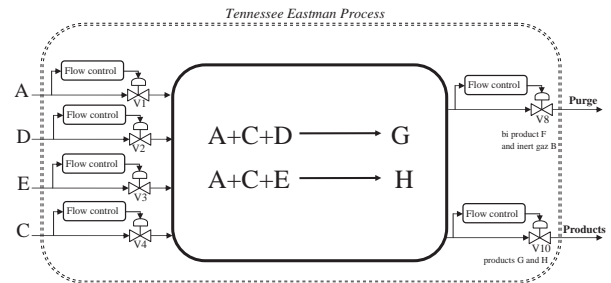
- 1: **if**  $n$  is even **then**
- 2:      $n \leftarrow n$
- 3: **else**
- 4:      $n \leftarrow n + 1$
- 5: **end if**
- 6: Compute  $m_g, \sigma_g, Z_g$  and  $\tilde{P}_g$
- 7: Build  $X_0 = (012 \dots \frac{n}{2} - 1)$
- 8: **for**  $i=0: \frac{n}{2} - 1$  **do**
- 9:      $X_{i+1} = X_i + \text{ones}(1, \frac{n}{2})$
- 10:     $In_{X_i} = [\text{ones}(1, i) \text{ zeros}(1, \frac{n}{2}) \text{ ones}(1, \frac{n}{2} - i)]$
- 11: **end for**
- 12: Build  $Lt = [(In_{X_0})^T (In_{X_1})^T \dots (In_{X_{\frac{n}{2}-1})^T]^T$
- 13:  $InX = []$
- 14: **for**  $g=0: \frac{n}{2} - 1$  **do**
- 15:      $\tilde{P}_{X_g^{[1]}} = \tilde{P}_g$
- 16:     **for**  $r=2: \frac{n}{2}$  **do**
- 17:          $X_g^{[r]} = X_g(1, 1 : r)$
- 18:          $R_{X_g^{[r]}} = \tilde{P}_{X_g^{[r-1]}}^T \times \tilde{P}_{X_g^{[r-1]}} = U_{X_g^{[r]}} S_{X_g^{[r]}} V_{X_g^{[r]}}^T$
- 19:          $Q_{X_g^{[r]}} = \tilde{P}_{X_g^{[r-1]}} \times V_{X_g^{[r]}}$
- 20:         Obtain  $P_{X_g^{[r]}}$
- 21:         Set  $\tilde{P}_{X_g^{[r-1]}} \leftarrow \tilde{P}_{X_g^{[r]}}$
- 22:     **end for**
- 23:      $\tilde{P}_{X_g} \leftarrow \tilde{P}_{X_g^{[\frac{n}{2}]}}$
- 24:     Compute  $SPE_{X_g}(k)$  and  $In_{X_g}(k)$
- 25:      $InX = [InX; In_{X_g}(k)]$
- 26: **end for**
- 27: Compare  $InX$  to columns of  $Lt$ , identify current mode.

**Fig. 3.** Flow chart of the proposed structuration method

has become a cornerstone for the process control and monitoring community for benchmarking diverse methodologies [26, 15, 27, 14, 16]. This process comprise four primary inputs, an output, and a purge. The TEP yields two products (G and H) alongside two unwanted residuals: byproduct F and inert gas B. This outcome stems from the interaction of four reactants: A, C, D, and E. Due to confidentiality, the exact nature of reactants, residuals, and products remains undisclosed.

**4.1. Process presentation**

The TEP has a total of 53 variables : 41 measured and 12 manipulated. It comprises five key units : the reactor, product condenser, vapor-liquid separator, recycle compressor, and product stripper. As shown in figure 5, the main units are interconnected by different flows shown by numbered red pentagrams. The Reactants A, D, and E enter the exothermic two-phase reactor through flow 6 from the upstream. The reactor's output stream, a mix of products and unreacted feed, flows into a condenser through flow 7 and then into a vapor-liquid unit via flow

**Fig. 4.** General sketch of the TEP

13. Condensed elements are directed to the product stripper through flow 10, where unreacted reactants are separated from products G and H. Non-condensed components are recycled via flow 5 to the reactor feed, facilitated by the recycle compressor through flow 8. The final product is obtained via flow 11, while the byproduct is collected through the purge at flow 9. Flows 12, 11b, and 13b serve to cool the various units.

The TEP is open loop unstable, it is necessary to adopt an adequate control structure to ensure its correct operations [27]. Two strategies can be used in order to control a process : the centralized and the decentralized one. The centralized strategy controls only the outputs and thus considers the process as a unique bloc. The decentralized strategy split-up the process

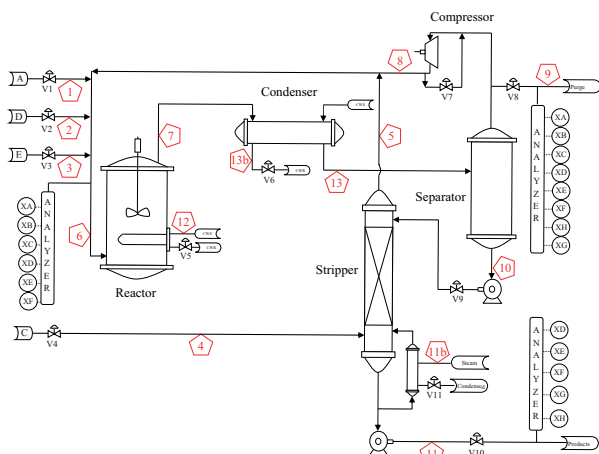


Fig. 5. Detailed layout of the TEP

and controls each of its parts separately. This strategy is more efficient than the centralized approach [28]. It will be adopted afterwards and contains 19 PI regulation loops [29]. The evolution of some measured and manipulated variables in closed loop for a sampling time equal to 3 minutes is shown in figure 6.

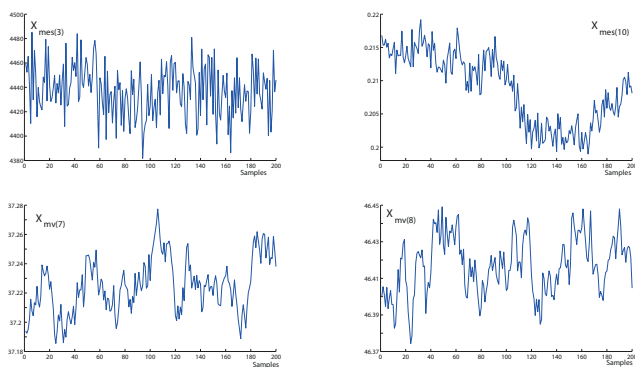


Fig. 6. Evolution of some TEP variables

#### 4.2. Process modeling

Many works provides some interesting datasets of different TEP operating modes, one can find some in [26, 27, 30]. In this work, the closed loop process, described in [29], is simulated in order to obtain the reference operating mode datasets. The source code of the closed loop process can be found in [30]. The datasets of the faulty modes used here are directly downloaded also from [30]. One assumes afterwards that the TEP possesses three distinct known operating modes : the reference mode, the  $Idv_1$  and the  $Idv_{13}$  modes. The reference mode is characterized by a production feed equal to 14076 kg/h with a mass ratio of 50/50 of reactants G and H. The modes  $Idv_1$  and  $Idv_{13}$  represents faulty operating modes. The  $Idv_1$  is characterized by variation in the composition of flow 4. The  $Idv_{13}$  denotes a drift in the process chemical reactions kinetics. One supposes the TEP have another faulty mode  $Idv_{12}$  defined by a random variation in the temperature of the flow 12 of the reactor cooling. Its data will be used in the fixation of different

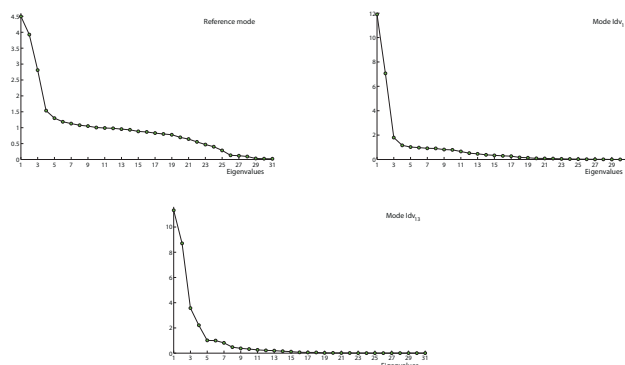
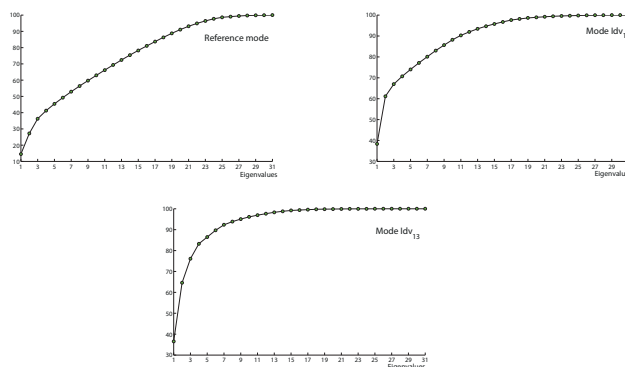
PCA model structure by the method in subsection 2.2.3.

**4.2.1. Choice of modeling variables** The choice of the variables used to build the PCA model have a significant impact on its sensitivity for process changes detection [2, 28]. After several heuristic tests, a vector  $z^r(k)$  containing 31 variables  $\{z_1^r(k), \dots, z_{31}^r(k)\}$  is chosen to build the different TEP models. This vector, see table 2, contains 22 measured and 9 manipulated variables.

**4.2.2. Fixation of the PCA model Structure** One thousand input/output samples are used for each of the three operating modes to fix the structures of the different PCA models. For each mode, one thousand values of  $z^r(k)$  are centered and reduced to obtain  $z(k)$  used to build the data matrices  $Z_{N_{ref}}$ ,  $Z_{N_{Idv1}}$  and  $Z_{N_{Idv13}}$  using the equation (2). The eigenvector and eigenvalue matrices of each mode are then obtained using (4).

The figure 7 shows the scree plot of the correlation matrices of different modes. Unlike the screeplots of modes  $Idv_1$  and  $Idv_{13}$ , all the eigenvalues of the reference mode are close to one and there's not a clear "elbow" permitting the right choice of  $\ell$  by the catell test. The figure 8 shows the increase of the pourcentage CPV according to  $\ell$  for all three operating modes, the choosen pourcentage to retain is equal to 97%.

Figure 9 represents the variation of VRE criterion according

Fig. 7. Scree plot of modes of reference,  $Idv_1$  and  $Idv_{13}$ Fig. 8. CPV for modes of reference,  $Idv_1$  and  $Idv_{13}$ 

to retained number  $\ell$  for the three process operating modes.

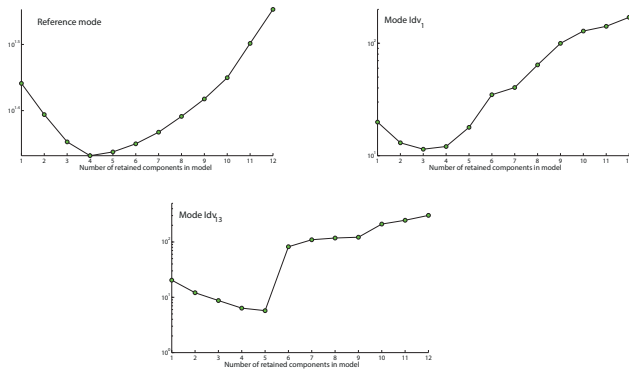
## New PCA-based scheme for process FDI. Application to TEP

**Table 2.** Variables used in construction of measurement vector  $z^r(k)$ 

| Variables  | Measured                                   | Variables  | Manipulated              |
|------------|--|------------|--------------------------|
| $z_1^r$    | A feed rate (flow 1)                       | $z_{23}^r$ | D feed rate (flow 2)     |
| $z_2^r$    | D feed rate (flow 2)                       | $z_{24}^r$ | E feed rate (flow 3)     |
| $z_3^r$    | E feed rate (flow 3)                       | $z_{25}^r$ | A feed rate (flow 1)     |
| $z_4^r$    | Total feed rate (flow 4)                   | $z_{26}^r$ | Total feed rate (flow 4) |
| $z_5^r$    | Recycle flow (flow 8)                      | $z_{27}^r$ | Output purge valve       |
| $z_6^r$    | Reactor feed rate (flow 6)                 | $z_{28}^r$ | Output separator valve   |
| $z_7^r$    | Reactor pressure                           | $z_{29}^r$ | Output stripper valve    |
| $z_8^r$    | Reactor liquid level                       | $z_{30}^r$ | Reactor cooling Valve    |
| $z_9^r$    | Reactor Temperature                        | $z_{31}^r$ | Output condenser Valve   |
| $z_{10}^r$ | Purge rate (flow 9)                        |            |                          |
| $z_{11}^r$ | Separator temperature                      |            |                          |
| $z_{12}^r$ | Separator liquid level                     |            |                          |
| $z_{13}^r$ | Separator pression                         |            |                          |
| $z_{14}^r$ | Separator underflow (flow 10)              |            |                          |
| $z_{15}^r$ | Stripper liquid level                      |            |                          |
| $z_{16}^r$ | Stripper pression                          |            |                          |
| $z_{17}^r$ | Stripper underflow (flow 11)               |            |                          |
| $z_{18}^r$ | Stripper temperature                       |            |                          |
| $z_{19}^r$ | Stripper cooling steam flow                |            |                          |
| $z_{20}^r$ | Compressor work                            |            |                          |
| $z_{21}^r$ | Reactor cooling water outlet temperature   |            |                          |
| $z_{22}^r$ | Separator cooling water outlet temperature |            |                          |

For clarity reasons, only the first twelve principal components are shown on the abscissa axis. Contrary to heuristic criteria, the VRE is convex and that it admits a single minimum.

Table 3 presents the number  $\ell$  obtained from the application of


**Fig. 9.** Evolution of VRE criterion regarding retained  $\ell$  for modes of reference,  $Idv_1$  and  $Idv_{13}$ 

all modeling methods presented in section 2.2. The choice of the structure of the PCA model to be retained from the heuristic methods is ambiguous because of their non convexity. The number  $\ell$  retained via the VRE criterion does not capture a good pourcentage of the process variance in the case of the reference and  $Idv_1$  modes. The proposed method does not only maximize the detection of the mode  $Idv_{12}$  but also capture a good pourcentage of process variance in all the three known modes.

### 4.3. Detection and identification of process changes

One supposes that the TEP have three different operating modes with known models : the reference mode  $M_0$ , and the faulty modes  $Idv_1$  and  $Idv_{13}$  called  $M_1$  et  $M_2$ . One supposes that the TEP have also an unknown operating mode : the  $Idv_{12}$  mode  $M_3$ . Two mode vectors  $X_0 = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$  and  $X_1 = \begin{pmatrix} 1 & 2 \end{pmatrix}^T$  are built to carry out the identification procedure using algorithm 2. The residual matrices  $\tilde{P}_{X_0}$  and  $\tilde{P}_{X_1}$  are built in order to identify the different operating process modes. The TEP is simulated during 200 samples following the illustrated timeline shown in figure 10. The transition period between the different modes is equal to zero. From the residual matrices and the current vector  $z(k)$ , one can build  $SPE_{X_0}(k)$  and  $SPE_{X_1}(k)$  as well as the mode state variables  $In_{X_0}(k)$  and  $In_{X_1}(k)$ .

Figure 11 portrays the evolution of indices  $SPE_{X_0}(k)$  and


**Fig. 10.** Chronology of modes of TEP

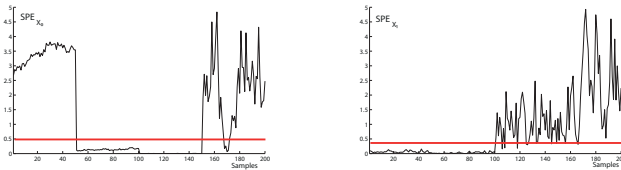
$SPE_{X_1}(k)$ , along with their respective thresholds in the simulation used to test the performance of the proposed scheme. The  $SPE_{X_0}(k)$  index exceeds its threshold within intervals  $[0, 50]$  (mode  $M_2$ ) and  $[151, 200]$  (mode  $M_3$ ), while staying below the threshold during  $[51, 150]$  corresponding to modes  $M_1$  and  $M_0$ . On the other hand, the  $SPE_{X_1}(k)$  index remains below its threshold in  $[0, 100]$  (modes  $M_2, M_1$ ), but surpasses it within

**Table 3.** Retained  $\ell$  and pourcentage of conserved variance for every mode by different modeling methods

|                         | Reference Mode |               | Mode $Idv_1$ |               | Mode $Idv_{13}$ |               |
|-------------------------|----------------|---------------|--------------|---------------|-----------------|---------------|
|                         | $\ell$         | % of variance | $\ell$       | % of variance | $\ell$          | % of variance |
| Eigenvalues mean method | 10             | 62.96 %       | 5            | 74 %          | 5               | 86.48 %       |
| Catell test method      | 4              | 41.2 %        | 3            | 66.97 %       | 5               | 86.48 %       |
| CPV Method              | 24             | 97.72 %       | 17           | 97.62 %       | 12              | 97.65 %       |
| VRE Method              | 4              | 41.2 %        | 3            | 66.97 %       | 5               | 86.48 %       |
| Proposed Method         | 24             | 97.72 %       | 18           | 98.18 %       | 12              | 97.65 %       |

[101,200] (modes  $M_0, M_3$ ). Sporadic false alarms emerge in these indices trends due to process non-linearity and noise.

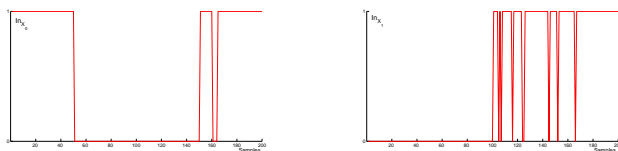
The different modes identification is realized via the mode



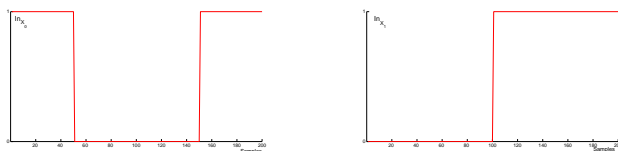
**Fig. 11.** Variation of  $SPE_{X_0}(k)$  and  $SPE_{X_1}(k)$

state variables  $In_{X_0}(k)$  and  $In_{X_1}(k)$ . Those boolean variables give some wrong detections due to uncertainties and measurement noise as shown in figure 12.

In order to fix this problem, a persistence condition is used. In this case, a variable alteration is declared when its value undergoes a change and remains constant for five consecutive sampling periods. The value "five" is to be adopted by learning according to the treated example. The evolution of the mode state variables after the application of the persistence condition is shown in figure 13. The identification of the operating mode is correctly achieved by comparing the values of the persistent mode state variables to the columns of Table 1.



**Fig. 12.** Variation of  $In_{X_0}(k)$  and  $In_{X_1}(k)$  before persistence



**Fig. 13.** Variation of  $In_{X_0}(k)$  and  $In_{X_1}(k)$  after persistence

## 5. CONCLUSION

This paper presents a novel PCA-based approach for process fault detection and localization, divided into three main parts.

The first part introduces PCA fundamentals and various methods for determining its model structure, including a new technique enhancing sensitivity to specific faults. Although it requires prior knowledge of faulty data, it excels at creating sensitive models without the use of large faulty dataset. The second part presents a unique method for structuring indices tailored to systems with process faults, reducing the required number of indices by half compared to traditional approaches. Operating mode identification is based on the use of a boolean variable derived from the computed indices and a localization table. The third part delves into the TE Process, a literature review to select control structures and variables is performed. The proposed modeling and structuration methods are successfully applied to the TE process. This paper has exclusively concentrated on process changes without considering transient states. The inclusion of these states in process operations worth further investigations.

## REFERENCES

- [1] P. Bielenica, J. Widzinska, A. Łukaszewski, Nogal, and P. Łukaszewski, "Decentralized fault location, isolation and self restoration (flisr) logic implementation using iec 61850 goose signals," *Bulletin of the Polish Academy of Sciences Technical Sciences*, vol. 70(5), p. e143101, 2022.
- [2] F. Pimentel Peres and F. Sanson Fogliatto, "Variable selection methods in multivariate statistical process control: A systematic literature review," *Computers & Industrial Engineering*, vol. 115, pp. 603–619, 2018.
- [3] M. Quinones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multi-mode continuous processes: A review," *Chemometrics and Intelligent Laboratory Systems*, vol. 189, pp. 56–71, 2019.
- [4] Y. Dong and S. Qin, "A novel dynamic pca algorithm for dynamic data modeling and process monitoring," *Journal of Process Control*, vol. 67, pp. 1–11, 2018.
- [5] W. Jakowluk and S. Jaszczak, "Cascade tanks system identification for robust predictive control," *Bulletin of the Polish Academy of Sciences Technical Sciences*, vol. 70(6), p. e143646, 2022.
- [6] D. Garcia-Alvarez, M. Fuente, and G. Sainz, "Fault detection and isolation in transient states using principal component analysis," *Journal of Process Control*, vol. 22, pp. 551–563, 2012.



- [7] W. Li, M. Peng, and Q. Wang, "Improved pca method for sensor fault detection and isolation in a nuclear power plant," *Nuclear Engineering and Technology*, vol. 51, pp. 146–154, 2019.
- [8] L. M. Elshenawy and T. Mahmoud, "Fault diagnosis of time-varying processes using modified reconstruction-based contributions," *Journal of Process Control*, vol. 70, pp. 12–23, 2018.
- [9] D. Garcia-Alvarez, A. Bregon, B. Puildo, and A.-G. C.J., "Integrating pca and structural model decomposition to improve fault monitoring and diagnosis with varying operation points," *Engineering Applications of Artificial Intelligence*, vol. 122, pp. 1–14, 2023.
- [10] H. Lahdhiri and O. Taouali, "Reduced rank kpca based on glrt chart for sensor fault detection in nonlinear chemical process," *Measurement*, vol. 169, p. 108342, 2021.
- [11] B. Malluhi, H. Nounou, and M. Nounou, "Enhanced multiscale principal component analysis for improved sensor fault detection and isolation," *Sensors*, vol. 22, p. 5564, 2022.
- [12] J. Zhang, D. Zhou, and M. Chen, "Monitoring multi-mode processes: A modified pca algorithm with continual learning ability," *Journal of Process Control*, vol. 103, pp. 76–86, 2021.
- [13] D. Jung and E. Frisk, "Residual selection for fault detection and isolation using convex optimization," *Automatica*, vol. 97, pp. 143–149, 2018.
- [14] J. Wang, W. Ge, J. Zhou, H. Wu, and Q. Jin, "Fault isolation based on residual evaluation and contribution analysis," *Journal of the Franklin Institute*, vol. 354(6), pp. 2591–2612, 2017.
- [15] M. Harkat, M. Mansouri, M. Nounou, and H. Nounou, "Fault detection of uncertain nonlinear process using interval-valued data-driven approach," *Chemical Engineering Science*, vol. 205, pp. 36–45, 2019.
- [16] H. Zhao, J. Liu, W. Dong, X. Sun, and Y. Ji, "An improved case-based reasoning method and its application on fault diagnosis of tennessee eastman process," *Neurocomputing*, vol. 219, pp. 39–49, 2017.
- [17] I. T. Jolliffe, *Principal component analysis*. New York: Springer-Verlag, 2003.
- [18] S. Qin and R. Dunia, "Determining the number of principal components for best reconstruction," *Journal of Process Control*, vol. 10, pp. 245–250, 2000.
- [19] M. Guerfel, *Systems diagnosis via data analysis and without behavior model a priori*. Tunisia: PhD Thesis of National Engineering School of Tunis, 2012.
- [20] J. Gertler and J. Cao, "Design of optimal structured residuals from partial principal component models for fault diagnosis in linear systems," *Journal of Process Control*, vol. 15, pp. 585–603, 2005.
- [21] X. Xu, L. Xie, and S. Wang, "Multimode process monitoring with pca mixture model," *Computers and Electrical Engineering*, vol. 40, pp. 2101–2112, 2014.
- [22] C. Zhao, W. Wang, Y. Qin, and F. Gao, "Comprehensive subspace decomposition with analysis of between-mode relative changes for multimode process monitoring," *Industrial Engineering Chemical Research*, vol. 54, pp. 3154–3166, 2015.
- [23] M. Harkat, M. Mansouri, M. Nounou, and H. Nounou, "Enhanced data validation strategy of air quality monitoring network," *Environmental Research*, vol. 160, pp. 183–194, 2018.
- [24] Y. Huang, J. Gertler, and T. McAvoy, "Sensor and actuator fault isolation by structured partial pca with nonlinear extensions," *Journal of Process Control*, vol. 10, pp. 459–469, 2000.
- [25] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, Maryland: Johns Hopkins University Press, 2013.
- [26] A. Bathelt, N. Ricker, and M. Jelali, "Revision of the tennessee eastman process model," *IFAC-PapersOnLine*, vol. 48(8), pp. 309–314, 2015.
- [27] C. Reinartz, M. Kulahci, and O. Ravn, "An extended tennessee eastman simulation dataset for fault-detection and decision support systems," *Computers & Chemical Engineering*, vol. 149, p. 107281, 2021.
- [28] S. Yin, S. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process," *Journal of Process Control*, vol. 22, pp. 1567–1581, 2012.
- [29] N. Ricker, "Decentralized control of the tennessee eastman challenge process," *Journal of Process Control*, vol. 4, pp. 205–221, 1996.
- [30] N. Ricker, "Tennessee eastman challenge archive," 2023, available at <https://depts.washington.edu/control/LARRY/TE/download.html>.