International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 8, No. 1, 22-38

Fault Detection Using an Adapted Interval PCA Approach

Anis El Maharat[1], Chouaib Chakour[2], and Azzedine Hamza[1]

[1] Department of Electronics and Communications, Kasdi Merbah University, Ouargla, Algeria
[2] Electrical Engineering Department, University 20 August 1955, Skikda, Algeria.
elmaharat.anis@univ-ouargla.dz

Abstract. Principal Component Analysis (PCA) is a commonly employed technique in industrial systems for process monitoring and fault diagnosis, owing to its capability to efficiently process large datasets. Traditionally, it is applied to single-valued variables, where critical information can be lost in real scenarios with data uncertainties. Intervalvalued PCA methods like Symbolic Covariance PCA (SCPCA) and Complete Information PCA (CIPCA) have been developed to enhance fault detection by incorporating data uncertainties in the PCA model. This paper presents a novel adaptation of SCPCA for detecting incertain sensor faults, marking the first correct implementation of SCPCA for fault detection and isolation (FDI). It aims to compare the performance of the NewSCPCA with that of CIPCA, evaluating its reliability and accuracy in detecting sensor faults in a greenhouse prototype system.

Keywords: Fault Detection, · Symbolic Covariance PCA · Complete Information PCA · Interval Data.

Received: 19 December 2024 | Revised: 29 December 2024 | Accepted: 21 April 2025
DOI: 10.53508/ijiam.1603442

1 Introduction

Principal Component Analysis (PCA) is a widely adopted technique for process moni-toring and fault diagnosis in industrial systems. Its popularity stems from its ability to efficiently manage and analyze large volumes of correlated process data [1][2]. PCA simplifies the data into a lower-dimensional representation by extracting linear relationships from high-dimensional datasets without losing critical information. This reduced representation, known as the PCA model, serves as a powerful tool for detecting anomalies and diagnosing faults within the system.

Traditionally, PCA has been applied to single-valued variables, where each data point is represented by a specific value [3]. However, this simplification can lead to a significant loss of information, particularly in real-world scenarios where data is often approximate and stained with uncertainties. To address this limitation, researchers have proposed representing the measurements as intervals to include process data and its inherent uncertainties [4][5].

Extending PCA to handle interval data involves developing new algorithms capable of dealing with these uncertainties. Various interval-valued PCA approaches have been introduced, such as Vertices PCA (VPCA), Centers PCA (CPCA) [6], Midpoints-Radii PCA (MRPCA) [7], Symbolic Covariance PCA (SCPCA) [8][9], and Complete Information PCA (CIPCA) [10]. These methods improve the robustness of fault detection in uncertain and complex industrial systems, thereby preventing false detections and maintaining sensitivity to deviations [11].

In this paper, we present a new adaptation of the SCPCA approach to detect sensor faults in the presence of uncertainties. The sample covariance matrix function calculated by SCPCA does not respect the properties of a classical covariance matrix, as discussed in [12] This limitation significantly affects the model's performance in detecting sensor faults. This work proposes a new version of the SCPCA technique, which maintains the mathematical properties of classical PCA model for detecting and isolating (FDI) sensor faults. The effectiveness of the NewSCPCA technique will be evaluated by comparing its performance to the CIPCA and conventional SCPCA approaches. The comparison will focus on the reliability and accuracy of sensor fault detection in the greenhouse prototype system.

This paper is organized as follows. Section.2 provides the theoretical foundation of PCA as a multivariate statistical tool for process monitoring. In section.3, we present the modeling of interval data using new and conventional SCPCA and CIPCA approaches. Section. 4 addresses the index fault detection used in this study. Experimental results of the comparison study between these three models for fault detection of the greenhouse prototype system are given in Section. 5. Finally, conclusions are given in Section. 6.

2 Classical PCA Model

PCA aims to identify the axis that captures the most information in the process data. This is achieved by calculating uncorrelated linear combinations of the original variables [13].

We consider a high dimensional data matrix $X = \{x_1, x_2, \dots x_m\} \in \mathbb{R}^{n \times m}$, where n represents the number of samples and m represents the number of process variables collected under normal operating conditions. The process data matrix is supposed normalized to zero mean and unit variance. For score process data matrix, $T = \{t_1, t_2, \dots x_l\} \in \mathbb{R}^{n \times l}$ is the low-dimensional output matrix, which consists of n samples of l independent components. The transformation matrix $\hat{p} = \{p_1, p_2, \dots p_l\} \in \mathbb{R}^{m \times l}$ contains orthogonal vectors p_i. The projection of the original high-dimensional data X into the reduced-dimension output T is expressed by the formula:

$$T = X\hat{P} \tag{1}$$

With, $\hat{P}$ is obtained from the eigendecomposition equation of covariance matrix. The sample covariance matrix, $COV \in \mathbb{R}^{m \times m}$, is defined as:

$$COV = \frac{1}{n-1} X^T X = P^T \Lambda P, P^T P = I \tag{2}$$

Where $I \in \mathbb{R}^{m \times m}$ is an identical matrix, $P \in \mathbb{R}^{m \times m}$ loading eigenvectors of COV, and Λ is a diagonal matrix containing eigenvalues of COV. The sample correlation function between X_j and $X_{j'}$ is given as follows:

$$Cor_{jj'} = \frac{Cov_{jj'}}{\sigma_j \sigma_{j'}} \tag{3}$$

Where, the $Covjj'$ equals $Corjj'$ when data is normalized to zeros mean and unit variance.

The first l eigenvectors that allow the generation of a matrix $\hat{P}$ with the highest variances are chosen using the Variance of Reconstruction Error (VRE) method. This method is a pioneering approach for selecting the optimal number of l [13] .

3 Modeling Uncertainties as Intervals

3.1 Interval-Valued Data Description

Measurements received from sensors are never completely precise and are always subject to a margin of error, known as uncertainty. This means that they cannot be perfectly accurate. When constructing a robust monitoring model, it is crucial to take into account the uncertainties involved in order to effectively detect potential failures [4]. One way to handle uncertainties in process data is to describe the data using intervals instead of representing it with single values.

An interval-valued variable $X_j(k)$ is a type of variable that is characterized by having its values constrained within a range defined by two bounds: a minimum value and a maximum value. This can be represented mathematically as:

$$[x_j(k)] = [\underline{x}_j(k), \bar{x}_j(k)] \tag{4}$$

The interval width is determined by the sensor accuracy (Δ) provided by the manu-facturer. Construct the interval data matrix $[X]$, where each entry represents an intervalvalued sample as:

$$[X] = \begin{pmatrix} [x_1(1)] \ldots [x_j(1)] \ldots [x_m(1)] \\ \vdots \quad \vdots \quad \vdots \\ [x_1(k)] \quad [x_j(k)] \quad [x_m(k)] \\ \vdots \quad \vdots \quad \vdots \\ [x_1(n)] \ldots [x_j(n)] \ldots [x_m(n)] \end{pmatrix} \tag{5}$$

knowing that, $j = 1, \ldots, m$ is the number of sensors or variables, $k = 1, \ldots, n$ is the number of observations, and $\bar{x}_j(k) < \underline{x}_j(k)$. Similar to classical PCA, normalizing the data is necessary before applying any In-terval PCA (IPCA) approach for modeling and monitoring the process [5]. The authors in [3] provided the mean and variance values of the interval variable as follows:

$$\begin{aligned} m_j &= \frac{1}{n} \sum_{K=1}^{n} \frac{\left(x_j(k) + \bar{x}_j(k)\right)}{2}, \\ \sigma_j^2 &= \frac{1}{3n} \sum_{k=1}^{n} \left(\underline{x}_j^2(k) + \underline{x}_j(k)\bar{x}_j(k) + \bar{x}_j^2(k)\right) - (m_j)^2 \end{aligned} \tag{6}$$

Hence, the normalization process at instant k for the interval variable is performed by [14]:

$$\left[\frac{x_j(k) - m_j}{\sigma_j}, \frac{\bar{x}_j(k) - m_j}{\sigma_j}\right] \tag{7}$$

3.2 Interval PCA Model

When classical PCA is applied to interval data, it can result in information loss. This limitation necessitates extending classical PCA to the Interval case. The extension for handling interval data involves adapting only the method of calculating the covariance function. However, the structure of the covariance matrix and its eigen decomposition into eigenvalues and eigenvectors remain unchanged. Therefore, the mathematical properties that are considered in the classical model should be respected in the interval PCA. In this subsection, we will discuss the most well-known approaches to Interval PCA (IPCA), including Complete Information PCA (CIPCA) and Symbolic Covari-ance PCA (SCPCA), along with their limitations.

Problem Statement The covariance function defined by the classical PCA technique considers two mathe-matical properties. These proprieties appear when we have data normalized to zeros mean and unit variance. Supposing x_j and $x_{j'}$ two normalized variables, $Cov_{jj'}$ is the covariance function between them, is given by:

$$Cov_{jj'} = \frac{1}{n}\sum_{k=1}^{n} x_j(k)x_{j'}(k) \tag{8}$$

The first properties that should be respected include ensuring that the quantity of the $Cov_{jj'}$ is equivalent to a value that falls within the range of -1 to 1. The second one is when x_j is the same of x'_j, the covariance function $Cov_{jj'}$ become equivalent to the variance function, given by:

$$Cov_{jj'} = \sigma^2_{junit} = \frac{1}{n}\sum_{k=1}^{n}((x_j(k))^2 = 1 \tag{9}$$

The covariance function developed by SCPCA model does not allow to respect the mathematical properties of the classical PCA for the sample covariance matrix when data is normalized. Specifically, the diagonal elements of the covariance matrix do not align with the unit variance. As a result, the total variance of the principal components does not equal the sum variance of the input data, as mentioned in [12]. Additionally, the other covariance matrix elements can have values outside the range of -1 to 1. This drawback limits the application of this method for real-world applications and can lead to misdiagnosis in fault prediction cases.

Complete Information PCA One of the developed IPCA approaches is a method known as Complete Information PCA (CIPCA), which was introduced by [10]. This approach considers each interval data unit in Eq.4 as an infinitely dense point uniformly distributed within it. It defines the covariance matrix using the interval data's inner product and squared norm operator. The covariance matrix, denoted $\sum cipca$, of process data matrix $[X]$, is computed by:

$$\Sigma_{\text{cipca}} = \frac{1}{n}\begin{pmatrix} \langle[x_1],[x_1]\rangle & \langle[x_1],[x_2]\rangle & \dots & \langle[x_2],[x_m]\rangle \\ \langle[x_2],[x_1]\rangle & \langle[x_2],[x_2]\rangle & \dots & \langle[x_1],[x_m]\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle[x_m],[x_1]\rangle & \langle[x_m],[x_2]\rangle & \dots & \langle[x_m],[x_m]\rangle \end{pmatrix} \tag{10}$$

Where,
For $j \neq j'$, the inner product is given by:

$$Cov_{jj'} = \langle[x_j],[x_{j'}]\rangle = \sum_{k=1}^{n}\langle[x_j(k)],[x_{j'}(k)]\rangle \tag{11}$$

with,

$$\langle [x_j(k)], [x_{j'}(k)] \rangle = \frac{1}{4}(x_j + \bar{x}_j(x_i(k) + \bar{x}_i(k)) \tag{12}$$

For $j = j'$, the squared norm is defined as:

$$Cov_{jj'} = |||[x_j]|||^2 = \langle [x_j(k)], [x_j(k)] \rangle = \sum_{k=1}^{n} |||[x_j(k)]|||^2 \tag{13}$$

with,

$$|||[x_k(k)]|||^2 = \frac{1}{3}\Big(\underline{x}_j^2(k) + \underline{x}_j(k)\bar{x}_j(k) + \bar{x}_j^2(k)\Big) \tag{14}$$

The sample correlation function between x_j and $x_{j'}$ is defined as:

$$Cor_{jj'} = \frac{Cov_{jj'}}{\sigma_j \sigma_{j'}} \tag{15}$$

Where, when data is normalized, the $Cov_{jj'}$ and $Cor_{jj'}$ are equal.

The CIPCA method determines interval principal components $[t_k](k = 1, 2, \ldots, l)$, using a linear combination algorithm for interval-valued variables $[t_j](j = 1, 2, \ldots, m)$. This algorithm was originally developed by [15] and requires solving decomposition equation of the covariance/correlation matrix obtained. More details are given in [16].

Symbolic Covariance PCA The total use of the information contained within intervals using IPCA approaches leads to good modeling and improving the performance of the monitoring and robust diagnosis of sensor faults. Symbolic covariance PCA (SCPCA) is one of the IPCA approaches that addressed this issue, which was developed by [8,9]. This approach considers that the sample variance in Eq. 6 of interval data representation is a function of the total sum of squares (SST), and proves that the SST can be decomposed into the sum of the within variation, denoted SSW, and the between variation, denoted SSB:

$$n\sigma_j^2 = SST_j = SSW_j + SSB_j \tag{16}$$

Assuming that values within an interval are uniformly distributed across the intervals. The internal variation measured by the SSW_j can be defined as:

$$SSW_j = \frac{1}{n}\sum_{k=1}^{n} \frac{(\bar{x}_j(k) - \underline{x}_j(k))^2}{12} \tag{17}$$

And the SSB_j describes the variation of the interval midpoints is given by:

$$SSB_j = \sum_{k=1}^{n}\Big(\frac{\underline{x}_j(k) + \bar{x}_j(k)}{2} - m_j\Big)^2 \tag{18}$$

In a similar way, when $j \neq j'$, the authors in [8] extended the Eqs. 16, 17, and 18 to bivariate case. Consequently, the total sum of products SPT is the

sum of the within sum of products, SPW, and the between sum of products, SPT. The relation between sample covariance Cov and SPT is given as follows:

$$nCov_{jj'} = SPT_{jj'} = SPW_{jj'} + SPB_{jj'} \tag{19}$$

where,

$$SPW_{jj'} = \sum_{k=1}^{n} \frac{\left(\bar{x}_j(k) - \underline{x}_j(k)\right)\left(\bar{x}_{j'}(k) - \underline{x}_{j'}(k)\right)}{12} \tag{20}$$

$$SPB_{jj'} = \sum_{k=1}^{n} \left(\frac{\left(\underline{x}_j(k) + \bar{x}_j(k)\right)}{2} - m_j \right) \left(\frac{\left(\underline{x}_{j'}(k) + \bar{x}_{j'}(k)\right)}{2} - m_{j'} \right) \tag{21}$$

For Eqs. 19 20, and 21, the sample covariance function $Covjj'$ is given by:

$$Cov_{jj'} = \frac{1}{6n} \sum_{k=1}^{n} 2(\underline{x}_j(k) - m_j)(\underline{x}_{j'}(k) - m_{j'}) + (\underline{x}_j(k) - m_j)(\underline{x}_{j'}(k) - m_{j'}) + (\underline{x}_j(k) - m_j)(\underline{x}_{j'}(k) - m_{j'}) + 2(\underline{x}_j(k) - m_j)(\underline{x}_{j'}(k) - m_{j'}) \tag{22}$$

The sample covariance matrix $\Sigma scpca$is given as follows:

$$\Sigma \text{scpca} = \begin{pmatrix} Cov_{11} & & Cov_{12} & \dots & Cov_{1m} \\ Cov_{21} & & Cov_{22} & \dots & Cov_{2m} \\ \vdots & \vdots & \ddots & \vdots & \\ Cov_{m1} & Cov_{m2} & \dots & Cov_{mm} & \end{pmatrix} \tag{23}$$

The sample correlation function between two variables x_j and $x_{j'}$ is computed as:

$$Cor_{jj'} = \frac{Cov_{jj'}}{\sigma_j \sigma_{j'}} \tag{24}$$

Where, when data is normalized, the $Cov_{jj'}$ and $Cor_{jj'}$ are equivalent.

SCPCA constructs the interval principal components $[t_k](k = 1, 2, \dots, l)$ that max-imize the sample covariance matrix defined in Eq. 23, through calculating uncorre-lated linear combinations of the input data matrix $[x_J](j = 1, 2, \dots, m)$. Let $P = (p_1, p_2, \dots, p_m)$ be eigenvectors of $\Sigma scpca$, and $\Lambda = [\lambda_1, \dots, \lambda_m]$ its corresponding eigen-values. The transformation linear of higher-dimensional matrix $[X]$ into lower-dimen-sional matrix $[T]$ is performed by:

$$\begin{cases} \underline{t}_j(k) = \sum_{i=1}^{m} p_{ij} \underline{x}_i(k) \\ \bar{t}_j(k) = \sum_{i=1}^{m} p_{ij} \bar{x}_i(k) \end{cases} \tag{25}$$

The estimated interval variables for the first l components are provided by:

$$\begin{cases} \underline{x}_j(k) = \sum_{i=1}^{m} C_{ij}\underline{x}_i(k) \\ \bar{x}_j(k) = \sum_{i=1}^{m} C_{ij}\bar{x}_i(k) \end{cases} \tag{26}$$

Where, C_{ij} is the ith element of the jth column of matrix $C = \hat{P}\hat{P}^T$.

Proposed Symbolic Covariance PCA The newly adapted SCPCA method aims to overcome the limitations of the SCPCA approach when detecting sensor faults in the presence of uncertainties. The proposed method supposes the data is normalized to zero means and unit variance in Eqs. 17 and 18. While Eq. 17 remains unchanged, Eq. 18 can be rewritten as follows:

$$SSB_j = \sum_{k=1}^{n} \Big(\frac{\underline{x}_j(k) + \bar{x}_j(k)}{2} \Big)^2 \tag{27}$$

And between sum of products SPB becomes given by:

$$SPB_{jj'} = \sum_{k=1}^{n} \left(\frac{\big(\underline{x}_j(k) + \bar{x}_j(k)\big)}{2}) \right) \left(\frac{\big(\underline{x}_{j'}(k) + \bar{x}_{j'}(k)\big)}{2} \right) \tag{28}$$

Therefore, the sample covariance function $Cov_{jj'}$ in Eq. 22 becomes:

$$Cov_{jj'} = \frac{1}{n} \left(\sum_{k=1}^{n} \frac{\big(\bar{x}_j(k) - \underline{x}_j(k)\big)\big(\bar{x}_{j'}(k) - \underline{x}_{j'}(k)\big)}{12} + \sum_{k=1}^{n} \frac{\big(\underline{x}_j(k) - \bar{x}_j(k)\big)\big(\underline{x}_{j'}(k) - \bar{x}_{j'}(k)\big)}{4} \right) \tag{29}$$

Knowing that, when $j = j'$, $Cov_{jj'} = Cov_{jj} = \sigma^2_{\text{j unit}}$, and when $j \neq j'$, $-1 \leq Cov_{jj'} \leq +1$.
The sample covariance matrix $\Sigma scpca$ in Eq. 23 becomes as follows:

$$\Sigma \text{scpca} = \begin{pmatrix} \sigma^2_{\text{1 unit}} & & Cov_{12} & \dots & Cov_{1m} \\ Cov_{21} & & \sigma^2_{\text{2 unit}} & \dots & Cov_{2m} \\ \vdots & \vdots & \ddots & \vdots & \\ Cov_{m1} & Cov_{m2} & \dots & \sigma^2_{\text{m unit}} & \end{pmatrix} \tag{30}$$

Note that the sample covariance matrix contains the same values as the sample correlation matrix elements, assuming that data is normalized.

The proposed SCPCA determines the interval principal components $[t_k](k = 1, 2, \dots, l)$ as:

$$\begin{cases} \underline{t}_j(k) = \sum_{i=1}^{m} p_{ij}\underline{x}_i(k) \\ \bar{t}_j(k) = \sum_{i=1}^{m} p_{ij}\bar{x}_i(k) \end{cases} \tag{31}$$

Where, $P = (p_1, p_2, \dots p_m)$ is eigenvectors of $\Sigma scpca$, and $\Lambda = [\lambda_1, \dots, \lambda_m]$ representing their associated eigenvalues.

The interval variables for the first l components are estimated as:

$$\begin{cases} \underline{\hat{x}}_j(k) = \sum_{i=1}^{m} C_{ij} \underline{x}_i(k) \\ \bar{\hat{x}}_j(k) = \sum_{i=1}^{m} C_{ij} \bar{x}_i(k) \end{cases} \tag{32}$$

Where, C_{ij} refers to the i^{th} element of the j^{th} column of matrix $C = \hat{P}\hat{P}^T$.

4 Fault Detection

Once the PCA model representing the normal behavior of the process is established, various statistics indices are used for the monitoring phase. The most commonly used statistic for anomaly detection is the squared prediction error (SPE) [16]. At instant k, SPE is computed as follows:

$$SPE(k) = ||x(k) - \hat{x}(k)||^2 = \sum_{J=1}^{m} \big(e_j(k)\big)^2 \tag{33}$$

Normal behavior of the process must be verified: $SPE \leq \delta^2$.

Where δ^2 is threshold of SPE, which can be defined as [17]:

$$\delta_\alpha^2 = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right] \tag{34}$$

With,$\theta = \sum_{k=l+1}^{2} \lambda_k^{,} i = 1, 2, 3, h_0 = \frac{2\theta_1 \theta_3}{3\theta_2^{\frac{2}{2}}}$ and c_α represents the confidence thresh-old limit $(1 - \alpha)$ in the case of a normal distribution. When dealing with interval-valued data, the indicator SPE must be adapted to fit the interval-valued PCA model. This means extending the statistical measure SPE to appropriately handle the variability structure within intervals. The SPE for interval data is defined as follows:

$$[SPE] = \{\underline{SPE}, \overline{SPE}\} \tag{35}$$

Where,

$$\underline{SPE}(k) = ||\underline{x}(k) - \underline{\hat{x}}(k)||^2 = \sum_{J=1}^{m} \underline{bigl(e_j(k)\big)^2} \tag{36}$$

And,

$$\overline{SPE}(k) = ||\bar{x}(k) - \bar{\hat{x}}(k)||^2 = \sum_{J=1}^{m} \overline{\big(e_j(k)\big)}^2 \tag{37}$$

Where,

$\underline{SPE}$ and $\overline{SPE}$ result from applying traditional SPE to the lower and upper bounds of interval data, respectively.

The authors in [4] were introduced a new statistical index for interval fault detection called interval squared prediction error (ISPE). This index is based on the squared norm of the interval. The ISPE is calculated at instant k as follows:

$$ISPE(k) = |||[e(k)]|||^2 = \sum_{j=1}^{m} |||[e_j(k)]|||^2 \tag{38}$$

Where,

$$|||[e_j(k)]|||^2 = \frac{1}{3}\left(\underline{e}_j^2(k) + \underline{e}_j(k)\bar{e}_j(k) + \bar{e}_j^2(k)\right) \tag{39}$$

5 Application to Greenhouse Prototype System

This section outlines the greenhouse prototype system and identifies the key signals used for fault detection. It details the datasets collected, including their types and sizes, which were used to develop and validate the performance of NewSCPCA and CIPCA methods for monitoring the system.

5.1 Greenhouse Prototype Description

This greenhouse prototype was constructed and developed by [18] . It was located in M'ziraa, Biskra province, Algeria (34°4319.7" N 6°1739.2" E), an area known for its mild desert winters. The structure features a simple gable design with a single span and wooden framework, covered with 0.2 mm thick polyethylene film (shown in Figure 1). The facility serves as a nursery—a specialized greenhouse type designed to nurture seedlings until they're ready for transplanting. Inside, the nursery has a raised wooden platform housing three seedling trays (45×20 cm) filled with prepared soil.

Fig. 1: External and internal views of the greenhouse prototype.

The prototype's data acquisition hardware system used in this prototype is based on an Arduino Mega 2560 programmable board, which functions as the central collection point for all sensor measurements. These sensors are designed to gather sufficient information about the current state of the greenhouse's internal and external climate. The monitoring system employs multiple sensors strategically placed inside and outside the greenhouse. Inside, two DHT22 sensors are mounted at 0.3 meters height to measure air temperature, while their outdoors counterpart is positioned at 1.25 meters. Solar irradiation is monitored by a cost-effective pyranometer using a BPW34 silicon photodiode, installed externally at 1.4 meters. A DC motor-based anemometer is mounted outside at 1.55 meters to measure wind conditions, and a MH-RD Rain module provides precipitation alerts. For more details, the reader may refer to the work in [18].

5.2 Data Collection

Our database captures winter conditions over five successive days for the month of January 2019. The measurement system operated at a frequency of one sample per minute, resulting in a daily collection of 1440 distinct measurements.

In the CIPCA and SCPCA methods modeling process, the first 6200 samples of X were used to construct the CIPCA and SCPCA models. The remaining 1000 samples were then used for testing. The number of sensors used in this prototype is 9 sensors. We have chosen 7 variables among them. By considering an uncertainty δ_{xj} that is about 10% of the range of variation in measurements for each variable x_j, we create a new interval data matrix for the process. In this matrix, δ_{xj} serves as the radius of the data intervals, as Figure 2.

5.3 Results and Discussion

Once interval-valued data representing the prototype's behavior is established, the sample covariance matrices of the approaches that discussed in Section 3 are constructed, as shown in Tables 1, 2, 3, and 4.

Table 1 shows the sample covariance matrix of the PCA model when the data is single-valued (classical), before conversion to an interval format. The values along the diagonal are equal to one, representing unit variance. The remaining elements indicate the covariance between different variables, which are constrained between -1 and $+1$.

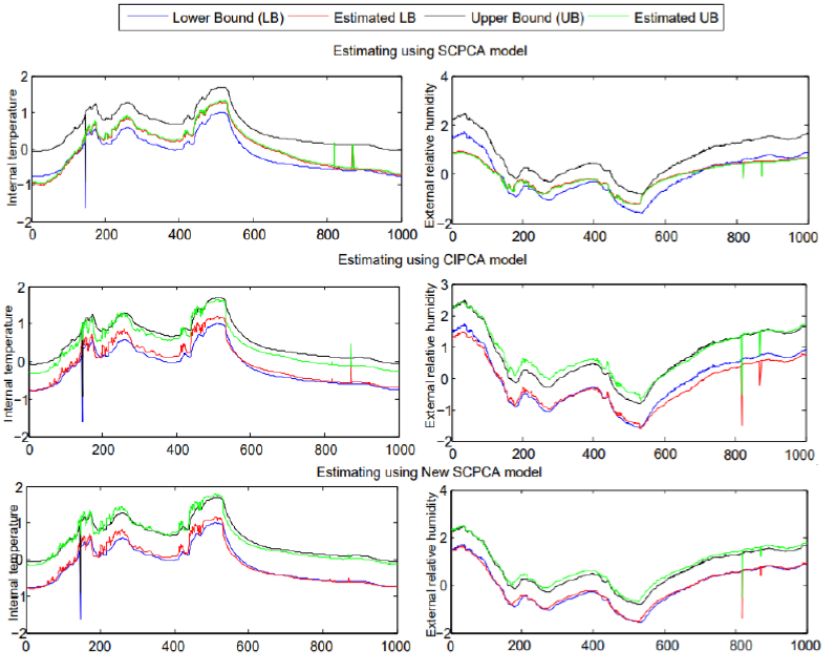

Fig. 2: Interval data of internal temperature and external relative humidity with their estimations using SCPCA, CIPCA, and NewSCPCA models.

Table 1: The sample covariance matrix of the classical PCA approach.

1.00	0.98	0.95	-0.87	-0.87	-0.88	0.91
0.99	1.00	0.96	-0.86	-0.87	-0.89	0.92
0.95	0.96	1.00	-0.81	-0.82	-0.93	0.83
-0.87	-0.86	-0.81	1.00	0.99	0.83	-0.75
-0.87	-0.87	-0.82	0.99	1.00	0.85	-0.76
-0.88	-0.89	-0.93	0.83	0.85	1.00	-0.75
0.91	0.92	0.83	-0.75	-0.76	-0.75	1.00

In the case of interval approaches, Tables 2 and 3 present the sample covariance matrices for the CIPCA and the proposed SCPCA techniques. These methods adhere to the same properties as the classical PCA model. However, the sample covariance matrix values for the conventional version of the SCPCA model, as defined in Table 4, do not observe the mathematical properties characteristic of the classical PCA model. Therefore, the sum of variance in Table 4 does not equal the total variance of input data, leading to inaccuracies in

the model quality, fault detection index performance, and threshold calculation. This limitation renders this method unreliable for diagnosing sensor faults.

Table 2: The sample covariance matrix of the NewSCPCA approach.

1.00	0.99	0.95	-0.79	-0.80	-0.75	0.91
0.99	1.00	0.96	-0.78	0.82	-0.75	0.92
0.95	0.96	1.00	-0.72	-0.73	-0.76	0.83
-0.79	-0.78	-0.72	1.00	0.99	0.84	-0.68
-0.79	-0.78	-0.73	0.99	1.00	0.85	-0.69
-0.75	-0.75	-0.76	0.84	0.85	1.00	-0.63
0.91	0.92	0.83	-0.68	-0.69	-0.63	1.00

Table 3: The sample covariance matrix of the CIPCA approach.

1.00	0.95	0.91	-0.83	-0.84	-0.81	0.88
0.95	1.00	0.91	-0.82	-0.83	-0.82	0.88
0.91	0.91	1.00	-0.77	-0.78	-0.85	0.79
-0.83	-0.82	-0.77	1.00	0.94	0.76	-0.72
-0.84	-0.83	-0.78	0.94	1.00	0.78	-0.73
-0.81	-0.82	-0.85	0.76	0.78	1.00	-0.70
0.88	0.88	0.79	-0.72	-0.730	-0.70	1.00

Table 4: The sample covariance matrix of the SCPCA approach.

1.14	1.13	1.08	-0.95	-0.96	-0.92	1.04
1.13	1.13	1.09	-0.93	-0.94	-0.92	1.05
1.08	1.09	1.13	-0.87	-0.88	-0.95	0.94
-0.95	-0.93	-0.87	1.13	1.13	0.93	-0.82
-0.96	-0.94	-0.88	1.13	1.13	0.94	-0.83
-0.92	0.92	-0.95	0.93	0.94	1.08	-0.78
-0.92	0.92	-0.95	0.93	0.94	1.08	-0.78

The model PCA for each approach is constructed by solving the eigendecomposition equation of the corresponding covariance matrix. The matrix of interval data $[X]$ is estimated by selecting the number of principal components l using the VRE method. we select $l = 2$ for two versions of SCPCA and CIPCA ap-

proaches, corresponding to the largest eigenvalues of the covariance matrix in each method, respectively.

In order to demonstrate the effectiveness of the NewSCPCA approach, we will compare its performance with that of CIPCA and conventional SCPCA methods. Figure 3 shows the monitoring performances of the Interval PCA methods when the process is operating under normal conditions.

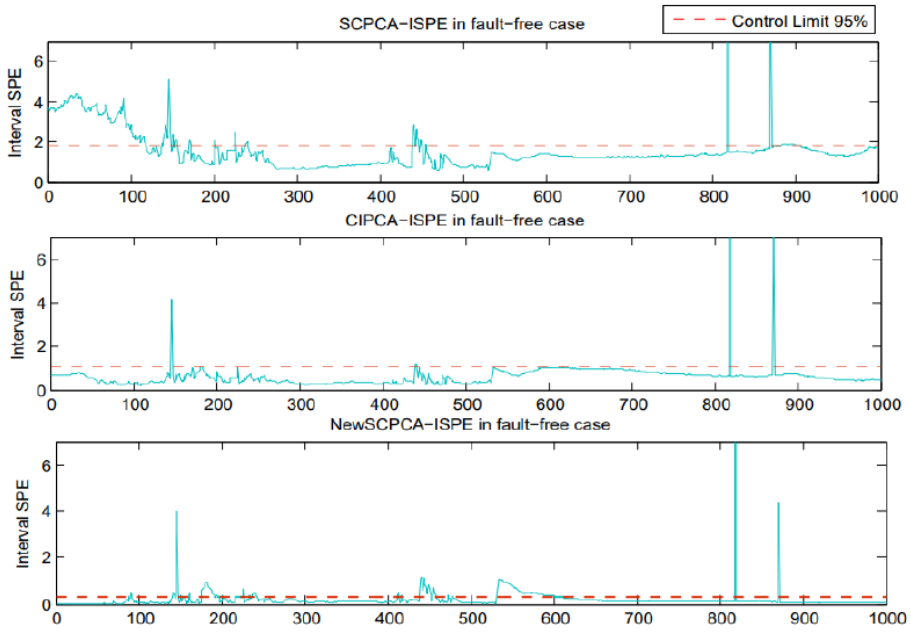

Fig. 3: Evolution of SCPCA-ISPE, CIPCA-ISPE, and NewSCPCA-ISPE under normal operation.

The detection index of the ISPE-based SCPCA approach indicates that the system is poorly modeled when compared to the ISPE-based CIPCA and the NewSCPCA methods, even though the system performs well in this simulation. Additionally, both versions of SCPCA exhibit an undesirable false alarm rate; however, the NewSCPCA is more effective than the older version in managing false alarms caused by uncertainties. Figure 4. shows the system in the case of fault, which has been injected by real fault since the instant 500 min of the testing phase. The magnitude of fault injected represents 25 % of the range of variation in the internal temperature sensor. The control limits are calculated at the confidence level of 95%. The ISPE detection index, based on the NewSCPCA and CIPCA approaches, requires only a lower-magnitude fault to continuously

detect deviations in the system. In contrast, the ISPE-based SCPCA requires a higher-magnitude fault.

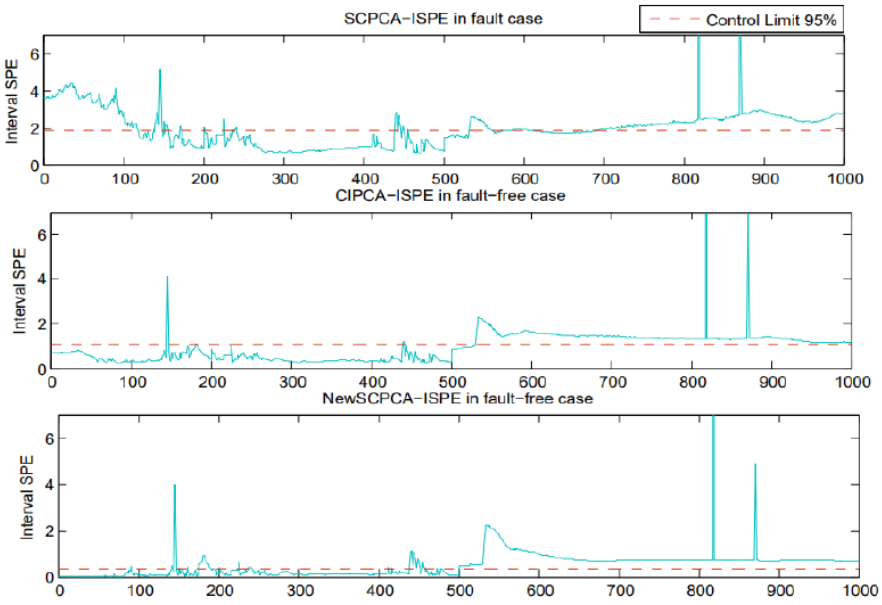

Fig. 4: Evolution of SCPCA-ISPE, CIPCA-ISPE, and NewSCPCA-ISPE in case of faulty pro-cess.

Table 5 shows a comparison between previous and NewSCPCA and CIPCA methods based on three factors: Good Detection Rate (GDR), False Alarm Rate (FAR), and Fault Detection Time Delay (DTD). Faults from f1 to f5 are successively injected into the system in steps of 5% of the variation range of the internal temperature sensor, taking into account the uncertainty defined by the model, which means the first step starts at 15%. It was observed that the fault continuously exceeded the threshold at f4 for NewS-CPCA-ISPE and CIPCA-ISPE while at f5 for SCPCA-ISPE, precisely at the values 29.70%, 29.13%, and 31.60% respectively.

The time required to indicate a fault after it occurs (DTD) is the same for both CIPCA-ISPE and SCPCA-ISPE, at an instant of 538 minutes. However, this factor differs for the NewSCPCA method, as shown in Table 4. This indicates that the new method is more sensitive to faults. The robustness of these approaches lies in their capacity to model uncertainties in the system, which are calculated using the FAR factor. Table 5 shows that the CIPCA method is the only robust technique in this study.

Table 5: GDR (%), FAR (%), and DTD (min) factors for SCPCA, CIPCA, and New SCPCA models.

Fault	**SCPCA-ISPE GDR(%)**	**CIPCA-ISPE GDR(%)**	**NewSCPCA-ISPE GDR(%)**
f_1	32.08	33.40	79.00
f_2	56.00	78.20	89.20
f_3	73.20	93.80	94.00
f_4	95.20	100.0	100.0
f_5	100.0	100.0	100.0
FAR(%)	16.40	0.60	13.00
DTD(min)	538	538	505

6 Conclusion

In this paper, we present a new adaptation of the Symbolic Covariance PCA (SCPCA) approach that addresses the limitations of the original method related to the sample covariance matrix. This proposed method leads to improved performance for detecting sensor faults amidst uncertainties. The greenhouse prototype simulation example evaluates the performance of this method by comparing it with that of Complete Information PCA (CIPCA) and the conventional version of SCPCA approaches. The application results demonstrated that the NewSCPCA is a more robust model for handling uncertainties and effectively detecting deviations in the system than the conventional SCPCA approach, but it is considered less efficient compared to the CIPCA method.
The issue discussed in this method can also be applied to other Interval PCA (IPCA) approaches that do not adhere to the mathematical properties of the sample covariance matrix of the classical PCA model, such as Mid-point and Radii PCA (MRPCA). This adjustment allows the model to use all the information contained in interval data, thereby enhancing its performance for detecting and isolating (FDI) sensor faults.

References

1. Neal B Gallagher, Barry M Wise, Stephanie Watts Butler, Daniel D White Jr, and Gabriel G Barna. Development and benchmarking of multivariate statistical process control tools for a semiconductor etch process: improving robustness through model updating. *IFAC Proceedings Volumes*, 30(9):79–84, 1997.
2. S Joe Qin. Statistical process monitoring: basics and beyond. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(8-9):480–502, 2003.
3. Lynne Billard and Edwin Diday. *Symbolic data analysis: Conceptual statistics and data mining*. John Wiley & Sons, 2012.
4. Tarek Ait-Izem, M-Faouzi Harkat, Messaoud Djeghaba, and Frédéric Kratz. On the application of interval pca to process monitoring: A robust strategy for sensor fdi with new efficient control statistics. *Journal of Process Control*, 63:29–46, 2018.

5. Chouaib Chakour, Azzedine Hamza, and Lamiaa M Elshenawy. Adaptive cipca-based fault diagnosis scheme for uncertain time-varying processes. *Neural Computing and Applications*, 33(22):15413–15432, 2021.
6. Pierre Cazes, Ahlame Chouakria, Edwin Diday, and Yves Schektman. Extension de l'analyse en composantes principales à des données de type intervalle. *Revue de Statistique appliquée*, 45(3):5–24, 1997.
7. Francesco Palumbo and Carlo N Lauro. A pca for interval-valued data based on midpoints and radii. In *New Developments in Psychometrics: Proceedings of the International Meeting of the Psychometric Society IMPS2001. Osaka, Japan, July 15–19, 2001*, pages 641–648. Springer, 2003.
8. Lynne Billard. Sample covariance functions for complex quantitative data. In *Proceedings of World IASC Conference, Yokohama, Japan*, pages 157–163, 2008.
9. Jennifer Le-Rademacher and Lynne Billard. Symbolic covariance principal component analysis and visualization for interval-valued data. *Journal of Computational and Graphical Statistics*, 21(2):413–432, 2012.
10. Huiwen Wang, Rong Guan, and Junjie Wu. Cipca: Complete-information-based principal component analysis for interval-valued data. *Neurocomputing*, 86:158–169, 2012.
11. Chouaib Chakour, Abdelhafid Benyounes, and Mahmoud Boudiaf. Diagnosis of uncertain nonlinear systems using interval kernel principal components analysis: Application to a weather station. *ISA transactions*, 83:126–141, 2018.
12. Antonio Irpino and Rosanna Verde. Basic statistics for distributional symbolic variables: a new metric-based approach. *Advances in Data Analysis and Classification*, 9:143–175, 2015.
13. Ricardo Dunia and S Joe Qin. Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 6(4):457–469, 1998.
14. Francisco de AT De Carvalho, Paula Brito, and Hans-Hermann Bock. Dynamic clustering for interval data based on l 2 distance. *Computational Statistics*, 21:231–250, 2006.
15. Ramon E Moore. *Interval analysis*. Prentice-Hall, 1966.
16. Mohamed-Faouzi Harkat, Gilles Mourot, and José Ragot. An improved pca scheme for sensor fdi: Application to an air quality monitoring network. *Journal of Process Control*, 16(6):625–634, 2006.
17. J Edward Jackson and Govind S Mudholkar. Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3):341–349, 1979.
18. Mounir Guesbaya and Hassina Megherbi. Thermal modeling and prediction of soilless greenhouse in arid region based on particle swarm optimization. experimentally validated. In *2019 International Conference on Advanced Electrical Engineering (ICAEE)*, pages 1–6. IEEE, 2019.