

## Article

# A Satellite Incipient Fault Detection Method Based on Local Optimum Projection Vector and Kullback-Leibler Divergence

Ge Zhang <sup>1,2</sup> , Qiong Yang <sup>1</sup>, Guotong Li <sup>1,2,3,\*</sup>, Jiaxing Leng <sup>1</sup> and Long Wang <sup>1</sup>

<sup>1</sup> Innovation Academy for Microsatellites of CAS, Shanghai 201203, China; zhangge1@shanghaitech.edu.cn (G.Z.); yangq@microstate.com (Q.Y.); lengjiaxing0520@163.com (J.L.); wangprchina@hotmail.com (L.W.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China

\* Correspondence: ligt@microstate.com

**Abstract:** Timely and effective detection of potential incipient faults in satellites plays an important role in improving their availability and extending their service life. In this paper, the problem of detecting incipient faults using projection vector (PV) and Kullback-Leibler (KL) divergence is studied in the context of detecting incipient faults in satellites. Under the assumption that the variables obey a multidimensional Gaussian distribution and using KL divergence to detect incipient faults, this paper models the optimum PV for detecting incipient faults as an optimization problem. It proves that the PVs obtained by principal component analysis (PCA) are not necessarily the optimum PV for detecting incipient faults. It then compares the on-line probability density function (PDF) with the reference PDF for detecting incipient faults on the local optimum PV. A numerical example and a real satellite fault case were used to assess the validity and superiority of the method proposed in this paper over conventional methods. Since the method takes into account the characteristics of the actual incipient faults, it is more adaptable to various possible incipient faults. Fault detection rates of three simulated faults and the real satellite fault are 98%, 84%, 93% and 92%, respectively.



**Citation:** Zhang, G.; Yang, Q.; Li, G.; Leng, J.; Wang, L. A Satellite Incipient Fault Detection Method Based on Local Optimum Projection Vector and Kullback-Leibler Divergence. *Appl. Sci.* **2021**, *11*, 797. <https://doi.org/10.3390/app11020797>

Received: 30 November 2020

Accepted: 10 January 2021

Published: 15 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Kullback-Leibler (KL) divergence; principal component analysis (PCA); optimum projection vector (PV); incipient fault; satellite

## 1. Introduction

Some faults may occur during the operation of satellites because of the uncertainties of the space environment, the limitations on pre-launch testing, and the imperfect design of satellites. It remains a challenge to detect and solve incipient faults and avoid further complications [1,2]. At present, the detection of faults in satellites is done by comparing the telemetry parameters with the preset thresholds [3,4]. However, this method is not suitable for detecting incipient faults because they may not change significantly during early manifestation of the fault. Compared to system signals, the amplitudes of incipient faults are small, typically from 1 to 10% [5,6] and can easily be masked by the normal variations in systems [7].

Model-based fault detection methods have the advantage of being highly intelligent and interpretable. Examples of such systems include the Livingstone2 and the HyDE fault diagnosis models developed by National Aeronautics and Space Administration (NASA) [8] as well as the Testability Engineering and Maintenance System (TEAMS), developed by Qualtech Systems Inc (QSI) [9]. All of these model-based fault detection methods require manual fault modeling by experts in the field. In addition, the accuracy of these models has a determining impact on their final performance at detecting faults. However, due to the high cost yet low yield of components used in satellites, the lifetime tests or the acquisition of actual fault samples required to build accurate models may often be too costly. Moreover, model-based fault detection methods are only capable of detecting

a narrow range of faults. This means that there may be faults that occur outside of the range detectable by the models and this may result in undetected incipient faults [10].

In recent years, data-driven fault detection methods have become a popular field of research [11–13]. These methods effectively utilize the extensive ground test data and operational data of satellites, with limited involvement of experts and do not require accurate mathematical models. Iverson et al. [14] proposed a fault detection method named the Inductive Monitoring System (IMS) and used it to detect faults in the Columbia Space Shuttle, and the TacSat-3 satellite. Schwabacher et al. [15] compared four data-driven fault detection methods containing IMS and One Class Support Vector Machine (OS-SVM). Pang et al. [16] proposed a fault detection method based on Gaussian Process Regression (GPR) and implemented it to detect faults in the periodic telemetry parameters of satellites [16]. Hundman et al. [17] proposed a spacecraft fault detection method based on Long Short-Term Memory (LSTM) networks and dynamic thresholds. Tariq et al. [18] proposed a fault detection method using a combination of multivariate convolutional LSTM and probabilistic principal component analysis (PCA) and applied it to the Korea Multi-Purpose Satellite 2 (KOMPSAT-2). Among these data-driven fault detection methods, distance-based fault detection methods such as IMS and OS-SVM are suitable for detecting fault data that are significantly different from normal data. However, these methods are less effective when the fault data and normal data mostly overlap. The fault detection methods using GPR or LSTM are suitable for detecting faults in regular (e.g., periodic) and time-continuous systems, but are ineffective when the telemetry parameters include interruptions in time.

Among the data-driven fault detection methods, methods based on PCA rely on the simplicity of the principle and on good real-time performance and have been used extensively in detecting faults in equipment [19,20]. The Hotelling's  $T^2$  and squared prediction error (SPE) statistic are usually used to measure deviations in the principal subspace, and to measure deviations in the residual subspace, respectively. However, the conventional  $T^2$  and SPE statistics are not effective at detecting incipient faults that are easily drowned out by noise [21]. Jinane et al. [21] proposed an incipient fault detection method using Kullback-Leibler (KL) divergence and PCA. On the basis of Jinane et al.'s research, Chen et al. [22] proposed an improved method. The proposed method monitors the derivations in principal and residual subspace. Youssef et al. [23] investigated a method of setting the threshold for fault detection using KL divergence for unknown distribution.

However, existing methods for detecting incipient faults based on PCA and KL divergence depend entirely on the off-line normal data to obtain the principal components. Each principal component is essentially a projection vector (PV). Once the normal data have been determined, the PVs used to detect faults also remain fixed. Unfortunately, the PVs obtained by PCA are not necessarily the optimum PV to detect the incipient fault from the view of fault detection. Fixed PVs may result in lower fault detection rates for some types of incipient faults [24,25]. To address the problem of detecting incipient faults in satellites, this paper proposes a novel incipient fault detection method based on local optimum PV (LOPV) and KL divergence. The main contributions of the present research are summarized in the following three points:

1. This paper puts forward the argument that the PVs obtained by PCA are not necessarily the optimum PV for using KL divergence to detect an incipient fault.
2. The problem of finding the optimum PV to detect the incipient fault is modeled as an optimization problem, and the KL divergence is used to detect the incipient fault on the LOPVs.
3. The application of the incipient fault detection method based on PCA and KL divergence is extended to the satellites. The effectiveness of the proposed method is proven in a real satellite fault.

This paragraph outlines the structure of this paper. Section 2 provides a brief introduction about PCA and KL divergence. In Section 3, the incipient fault detection method based on LOPV and KL divergence is presented in detail. In Section 4, the proposed method is

validated and analyzed by using a numerical example and a real satellite fault. Finally, the conclusions are presented in Section 5.

## 2. Preliminary

### 2.1. PCA

Let  $n_0, m$  be the numbers of samples and variables of a normal data matrix  $X \in R^{n_0 \times m}$ , with  $X = [x_1, \dots, x_k, \dots, x_m]$ . In general, each column of the original normal data  $X$  is standardized to eliminate the impact of different ranges of variables.

$$\bar{x}_k = \frac{x_k - E(x_k)}{\sigma(x_k)}, k \in [1, m] \tag{1}$$

$E(x_k)$  and  $\sigma(x_k)$  in Equation (1) are the mean and standard deviation of the  $k$ th column in the matrix  $X$ , respectively. The covariance matrix  $\Sigma \in R^{m \times m}$  of the standardized normal data matrix  $\bar{X}$  is then extracted and the singular value decomposition (SVD) is performed.

$$\Sigma = \frac{\bar{X}^T \bar{X}}{n_0 - 1} = V \Lambda V^T = \begin{bmatrix} P & \tilde{P} \end{bmatrix} \begin{bmatrix} \Lambda_{pc} & 0 \\ 0 & \Lambda_{res} \end{bmatrix} \begin{bmatrix} P & \tilde{P} \end{bmatrix}^T \tag{2}$$

The matrix  $V = [v_1, \dots, v_m] \in R^{m \times m}$  in Equation (2) is a set of standard orthogonal bases in  $R^m$  space. The space  $P = [v_1, \dots, v_l] \in R^{m \times l}$  consists of the first  $l$  column vectors in the matrix  $V$ , is the principal component subspace, and the space  $\tilde{P} = [v_{l+1}, \dots, v_m] \in R^{m \times (m-l)}$  which consists of the last  $m - l$  column vectors is the residual subspace [26].

The PCA based fault detection method usually uses the two statistics  $T^2$  and SPE, for fault detection. The calculations of the  $T^2$  statistic and the SPE statistic are shown in Equations (3) and (4), respectively. A fault is considered to have occurred when either the  $T^2$  or the SPE statistic has exceeded its corresponding detection threshold of  $\delta_T^2$  or  $\delta_{SPE}^2$ , respectively [27].

$$T^2 = \|\Lambda_{pc}^{-\frac{1}{2}} P^T \bar{X}^T\|^2 \leq \delta_T^2 \tag{3}$$

$$SPE = \bar{X} (I_{m \times m} - PP^T) \bar{X}^T \leq \delta_{SPE}^2 \tag{4}$$

### 2.2. KL Divergence

KL divergence is often used to measure the difference between two probability distributions in the same event space and is widely used for optimization tasks in machine learning [28,29]. The KL divergence between two continuous probability density functions (PDFs)  $f_1$  and  $f_2$  is defined as follows [30]:

$$I(f_1 \| f_2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} \tag{5}$$

Since the above definition does not satisfy symmetry and may result in  $I(f_1 \| f_2) \neq I(f_2 \| f_1)$ , an improved form of KL divergence referred to as symmetric KL divergence is presented as follows [21]:

$$KLD(f_1 \| f_2) = I(f_1 \| f_2) + I(f_2 \| f_1) \tag{6}$$

Let  $n_1, m$  be the numbers of samples and variables of on-line data matrix  $Y \in R^{n_1 \times m}$ . The main processes of the conventional incipient fault detection methods based on PCA and KL divergence are as follows [21,22]:

1. The standardized normal data  $\bar{X}$  are projected on each PV  $v_j$  and the reference PDF of the normal data is obtained after projection on  $v_j$  as  $f_j$  and the corresponding detection threshold is  $\varepsilon_j, v_j \in V, j \in [1, m]$ .

2. Each column  $y_k$  of the on-line data matrix  $Y$  is standardized with  $\bar{y}_k = \frac{y_k - E(x_k)}{\sigma(x_k)}$ ,  $k \in [1, m]$ .
3. The standardized on-line data  $\bar{Y}$  are projected on each PV  $v_j$  to obtain the PDF  $g_j$  of the on-line data.
4. Equation (6) is used to calculate the KL divergence  $KLD(f_j||g_j)$  between the PDF  $g_j$  and the reference PDF  $f_j$  for each PV  $v_j$ .
5. Whether the  $KLD(f_j||g_j)$  is greater than the corresponding detection threshold  $\epsilon_j$  is determined for each PV  $v_j$ . It is considered to be faulty when at least one of the  $m$  units of  $KLD(f_j||g_j)$  exceeds the corresponding detection threshold.

$$\begin{cases} H_0 : \forall v_j, v_j \in V, KLD(f_j||g_j) \leq \epsilon_j, \text{fault-free} \\ H_1 : \exists v_j, v_j \in V, KLD(f_j||g_j) > \epsilon_j, \text{faulty} \end{cases} \quad (7)$$

### 3. Incipient Fault Detection Method Based on LOPV and KL Divergence

For each PV  $v_j$ , if the fault detection method could detect an incipient fault with a high detection rate on the  $v_j$ , then  $v_j$  is considered sensitive to the incipient fault with a given significance level  $\alpha$ . The bigger the detection rate, the greater sensitivity to the incipient fault. It can be seen from formula (7) that the incipient fault can be well detected if there is one PV sensitive to the fault. From the perspective of fault detection, we hope to find an optimum PV that is most sensitive to the incipient fault.

#### 3.1. Optimum PV for Incipient Fault Detection

We assume that there is an optimum PV that is more sensitive to the incipient fault than the existing PVs  $v_1, v_2, \dots, v_m$  and the optimum PV is  $w \in R^m$ . To find the optimum PV, this paper assumes the  $m$  dimensional feature parameters to be monitored in their normal state, obeying the  $m$  dimensional joint Gaussian distribution  $X \in R^{n_0 \times m} \sim N(\mu_x, \Sigma_x)$ . Let the projection of  $X$  on  $w$  be  $F \in R^{n_0}$ . By the nature of the  $m$  dimensional joint Gaussian distribution,  $F$  obeys the one-dimensional Gaussian distribution  $F \sim N(\mu_x^T w, w^T \Sigma_x w)$  [26].

$$F = Xw = x_1w_1 + x_2w_2 + \dots + x_mw_m \quad (8)$$

We assumed that after the incipient fault occurs, the on-line faulty data  $Y$  still obey the  $m$  dimensional joint Gaussian distribution  $Y \in R^{n_1 \times m} \sim N(\mu_y, \Sigma_y)$ . However, due to the fault, the mean vector  $\mu_y \in R^m$  and the covariance matrix  $\Sigma_y \in R^{m \times m}$  of  $Y$  deviated and that their deviations were  $\Delta\mu$  and  $\Delta\Sigma$ , respectively.

$$\mu_y = \mu_x + \Delta\mu \quad (9)$$

$$\Sigma_y = \Sigma_x + \Delta\Sigma \quad (10)$$

The projection of  $Y$  on  $w$  is  $G = Yw$ . Then, from the nature of the joint Gaussian distribution,  $G$  obeys the one-dimensional Gaussian distribution  $G \sim N(\mu_y^T w, w^T \Sigma_y w)$ . Assume that there are two PDFs  $f$  and  $g$  for two Gaussian distribution signals  $F \sim N(\mu_1, \sigma_1^2)$  and  $G \sim N(\mu_2, \sigma_2^2)$ , respectively. The KL divergence of  $f$  and  $g$  can be expressed as follows [21]:

$$KLD(f||g) = \frac{1}{2} \left[ \frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + (\mu_1 - \mu_2)^2 \left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) - 2 \right] \quad (11)$$

The projections of the normal data matrix  $X$  and the faulty data matrix  $Y$  on  $w$  obey the one-dimensional Gaussian distribution  $F \sim N(\mu_x^T w, w^T \Sigma_x w)$  and  $G \sim N(\mu_y^T w, w^T \Sigma_y w)$ ,

respectively. Consequently, the mean and variance of  $F$  and  $G$  can be brought into Equation (11), respectively, obtaining:

$$\begin{aligned}
 KLD(f||g) &= \frac{1}{2} \left[ \frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Sigma_x w}{w^T \Sigma_y w} + (\mu_x^T w - \mu_y^T w)^2 \left( \frac{1}{w^T \Sigma_x w} + \frac{1}{w^T \Sigma_y w} \right) - 2 \right] \\
 &= \frac{1}{2} \left[ \frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Sigma_x w}{w^T \Sigma_y w} + (\Delta \mu^T w)^2 \left( \frac{1}{w^T \Sigma_x w} + \frac{1}{w^T \Sigma_y w} \right) - 2 \right]
 \end{aligned}
 \tag{12}$$

In Equation (12), the mean deviation vector  $\Delta \mu$ , normal data covariance matrix  $\Sigma_x$ , and on-line faulty data sample covariance matrix  $\Sigma_y$  can be obtained through the on-line faulty data matrix  $Y$  and normal data matrix  $X$ . Therefore, when historical normal data  $X$  and the on-line faulty data  $Y$  to be verified are obtained, only the PV  $w$  is unknown in Equation (12). In other words,  $KLD(f||g)$  is a function  $h(w)$  on the PV  $w$  as shown in Equation (13).

$$h(w) = KLD(f||g) = \frac{1}{2} \left[ \frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Sigma_x w}{w^T \Sigma_y w} + (\Delta \mu^T w)^2 \left( \frac{1}{w^T \Sigma_x w} + \frac{1}{w^T \Sigma_y w} \right) - 2 \right]
 \tag{13}$$

Sliding windows with the length of  $n_1$  are used to divide the normal data matrix  $X$  and submatrices of  $X$  can be obtained  $X_1, \dots, X_i, \dots, X_r, X_i \in R^{n_1 \times m}, i \in [1, r]$ . We project all submatrices  $X_1, \dots, X_i, \dots, X_r$  on  $w$  to obtain  $r$  PDFs  $PDF_{X_1}, \dots, PDF_{X_i}, \dots, PDF_{X_r}$ . We assume that each  $PDF_{X_i}$  obeys a one-dimensional Gaussian distribution. The KL divergence between  $PDF_{X_i}$  and  $f$  can be calculated by Equation (13).  $h_{X_i}(w) = KLD(f||PDF_{X_i}), i \in [1, r]$ . The process used to obtain vector  $h_X(w)$  is shown in Figure 1.

$$h_X(w) = [h_{X_1}(w), \dots, h_{X_i}(w), \dots, h_{X_r}(w)]
 \tag{14}$$

$$J(w) = \frac{h(w) - E(h_X(w))}{\sigma(h_X(w))}
 \tag{15}$$

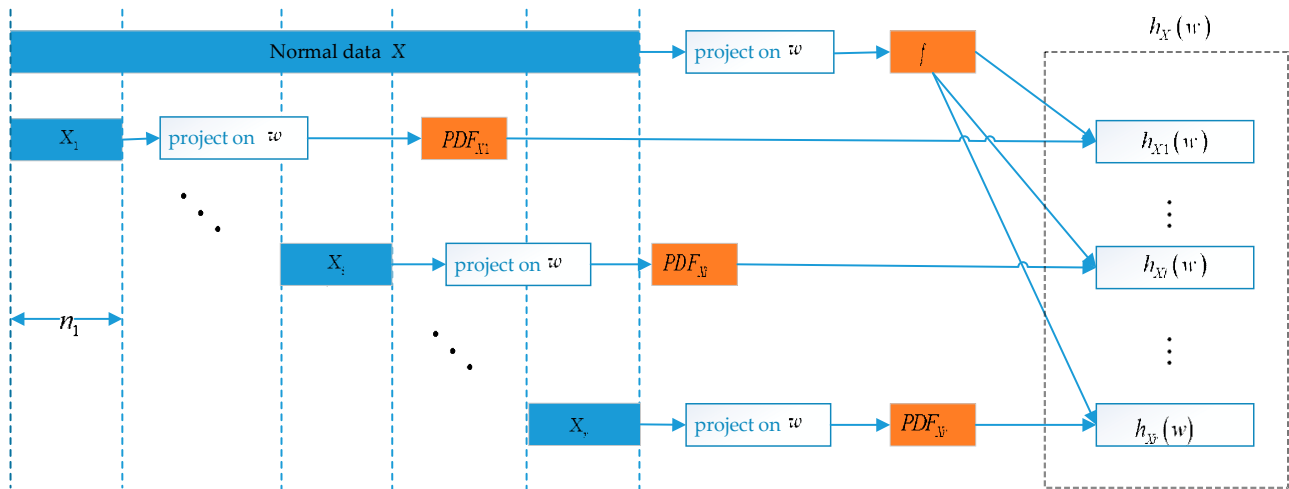


Figure 1. The process of obtaining vector  $h_X(w)$ .

The relative deviation in KL divergence of the on-line faulty data matrix  $Y$  and normal data matrix  $X$  on the PV  $w$  is defined as shown in Equation (15).  $E(h_X(w))$  and  $\sigma(h_X(w))$  in Equation (15) are the mean and standard deviation of the vector  $h_X(w)$ .  $E(h_X(w))$  and  $\sigma(h_X(w))$  reflect the center and range of the interval of variation range of the KL divergence for normal data after projection on  $w$ , respectively. Once the on-line faulty data matrix  $Y$  and normal data matrix  $X$  are given, the KL relative deviation  $J(w)$  still remains as the function about  $w$ .

With the given significance level  $\alpha$ , the matrix  $Y$  and the matrix  $X$ , The larger the value of the  $J(w)$ , the easier it is for that fault to be detected. On the contrary, if the value of the  $J(w)$  is small, it is likely to be drowned out by the noise and this leads to poor performance in detection of the fault. Thus, the problem of finding the optimum PV for fault detection can be modeled as an optimization problem, as shown in Equation (16).

$$\begin{cases} \min_w - J(w) \\ s.t. w^T w = 1; \forall w_i, -1 \leq w_i \leq 1, i \in [1, m] \end{cases} \quad (16)$$

If the PV  $w$  is directly optimized in Equation (16) so that  $h(w)$  is maximized, and the optimum solution may continue to converge on the PV corresponding to the smallest eigenvalue, thus making the optimization process meaningless. The above mentioned problem arises because the variation ranges of the eigenvalues of the different PVs are different. A small eigenvalue leads to small  $w^T \Sigma_x w$  and  $w^T \Sigma_y w$  in Equation (13), which are exactly in the denominator position, and slight noise can cause large fluctuations in the KL divergence on the corresponding PVs of the small eigenvalue. Therefore, it is necessary to standardize  $h(w)$ , and the optimization goal is to maximize the relative deviation of KL divergence,  $J(w)$ , rather than the absolute deviation of KL divergence,  $h(w)$ .

The goal of PCA is to maximize the variance of the normal data  $X$  after it is projected on the PV  $w$  and can be modeled as an optimization problem, as shown in Equation (17) [26].

$$\begin{cases} \min_w - w^T \Sigma_x w \\ s.t. w^T w = 1 \end{cases} \quad (17)$$

We supposed that  $w_p$  is the optimum solution of the optimization problem shown in formula (16),  $w_q$  is the optimum solution of the optimization problem shown in formula (17). Since the objective functions of the two optimization problems are different,  $w^T \Sigma_x w \neq J(w)$ , generally,  $w_p$  is not equal to  $w_q$ . However, solving formula (16) solves the optimization problem, and  $w_p$  is the optimum PV for using KL divergence to detect the fault. Because  $w_p \neq w_q$ ,  $w_q$  is not the optimum PV for using KL divergence to detect a fault in general. Therefore, the PVs obtained by PCA are not necessarily the optimum PV for using KL divergence to detect the fault.

Equation (16) converts the problem of finding the maximum  $J(w)$  into a standard constrained optimization problem by adding a negative sign and constrained conditions. The search for the optimum PV  $w$  that maximizes  $J(w)$  is on the surface of a hypersphere with a radius of 1. Equation (16) is a typical constrained optimization problem, which can be solved using readily available tools for solving optimization problem, such as the `fmincon` function built into MATLAB 2019b.

Since the objective function  $\min_w - J(w)$  in Equation (16) is a nonlinear function about  $w$ , the solution obtained by the iterative method is not necessarily a global optimum solution, but probably a local optimum solution [31,32]. However, considering that the optimization problem shown in Equation (16) needs to be solved when each set of on-line data arrives, finding the global optimum solution will undoubtedly increase the computational cost. Considering the timeliness, the iteration can be stopped when the optimization problem in Equation (16) converges to a local optimum solution that is better than the existing fixed PVs.

It is necessary to ensure that the local optimum solution searched for by Equation (16) is better than all the fixed PVs of  $v_1, v_2, \dots, v_m$  obtained by PCA. To achieve this, when the on-line data  $Y$  arrive, the on-line data  $Y$  are projected on each fixed PV  $v_1, v_2, \dots, v_m$  and the mean value drift vector  $\Delta\mu$ , respectively. The KL divergence relative deviations are calculated according to Equations (13)–(15) in each PV and the vector of mean value drift.

Then the vector  $v_j$  that makes the largest KL divergence relative deviation is selected as the initial iteration vector  $w_0$  of the optimization problem in Equation (16).

$$\begin{cases} w_0 = v_j \\ s.t. J(v_j) = \max\left(J(v_1), J(v_2), \dots, J(v_m), J\left(v_{m+1} = \frac{\Delta\mu}{\|\Delta\mu\|}\right)\right), j \in [1, m+1] \end{cases} \quad (18)$$

### 3.2. Detecting Incipient Faults Using LOPVs and Dynamic Thresholds

In this section, sliding windows with the length of  $n_1$  are used to extract and monitor the on-line data in real time. The on-line data extracted by the  $k^{th}$  sliding window are  $Y_k \in R^{n_1 \times m}$ . This article intends to use the idea of hypothesis testing to detect incipient faults for each  $Y_k$ , as shown in Figure 2. Let us assume that  $Y_k$  is faulty, then we test this hypothesis holds. If the hypothesis held after testing, it is considered that a fault has occurred in  $Y_k$ . Otherwise,  $Y_k$  is normal. The specific hypothesis testing steps are as follows:

1. Let  $Y = Y_k$ , we assume that  $Y_k$  is faulty. The method is used as described in Section 3.1 to find the local optimum PV  $w_k$  for fault detection between the on-line data  $Y_k$  and the historical normal data  $X$ .
2. Let the projection of  $X$  and  $Y_k$  on the local optimum PV  $w_k$  be  $Xw_k$  and  $Y_k w_k$ .
3. The KL divergence  $KLD(Xw_k || Y_k w_k)$  of  $Xw_k$  and  $Y_k w_k$  is calculated.
4. The threshold  $\epsilon_k$  of the local optimum PV  $w_k$  is set according to the given significance level  $\alpha$ .
5. If  $KLD(Xw_k || Y_k w_k) > \epsilon_k$ , then assuming that  $Y_k$  is faulty is correct. Otherwise,  $Y_k$  is normal. Let  $k = k + 1$ , the next sliding window  $Y_{k+1}$  will be tested from steps 1 to 5.

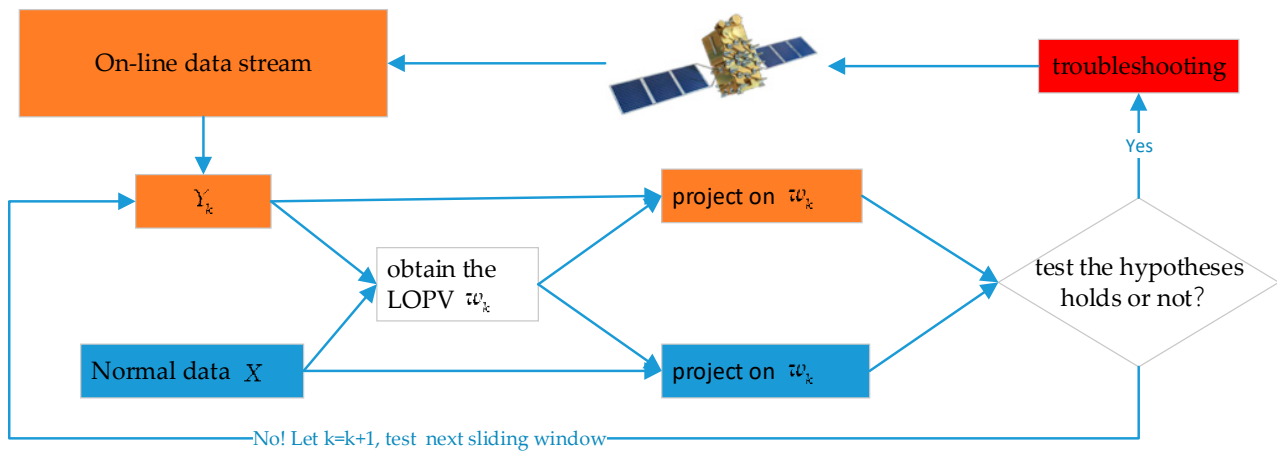


Figure 2. The idea of hypothesis testing to detect incipient faults for each  $Y_k$ .

For each on-line datum  $Y_k$ , the method proposed in this paper will find a local optimum PV  $w_k$  that is most sensitive to the fault. After  $w_k$  is obtained, we can obtain vector  $h_X(w_k)$ . According to reference [22],  $Kh_X(w_k)$  obeys the chi-square distribution with one degree of freedom,  $Kh_X(w_k) \sim \chi^2(1)$ , Where  $K$  is a large integer. Given the significance level  $\alpha$ , the method of setting the threshold  $\epsilon_k$  is shown in Equation (19). In Equation (19),  $\gamma$  is a constant. We suggest  $\gamma = E(h_X(w_k))$ , because there may be  $E(h_X(w_k)) \neq E(h_X(w_{k+1}))$ , which leads to  $\epsilon_k \neq \epsilon_{k+1}$ . Generally,  $w_k \neq w_{k+1}$ , thus,  $\epsilon_k \neq \epsilon_{k+1}$ . Therefore, dynamic thresholds are used to detect faults.

$$\epsilon_k = \frac{1}{K} \chi_\alpha^2(1) = \gamma \chi_\alpha^2(1) \quad (19)$$

### 3.3. The Complete Incipient Fault Detection Process

The complete incipient fault detection method based on LOPV and KL divergence consists of two parts: the preprocessing process and the monitoring process:

---

### The preprocessing process:

---

1. Z-score normalization is performed on the normal data matrix  $X$  to obtain  $\bar{X}$ .
  2. The reference mean vector  $\mu_x$  and the reference covariance matrix  $\Sigma_x$  for the normal data matrix  $\bar{X}$  are calculated.
  3. PCA is performed on the normal matrix  $\bar{X}$  and the matrix  $V$  of all  $m$  PVs is retained.
  4. Data matrices  $\bar{X}_1, \dots, \bar{X}_i, \dots, \bar{X}_r$  ( $\bar{X}_i \in \bar{X}, i \in [1, r]$ ) are extracted from the normal matrix  $\bar{X}$  using sliding windows with the length of  $n_1$ . The mean vector  $\mu_{X_i}$  and the covariance matrix  $\Sigma_{X_i}$  for each  $\bar{X}_i$  are calculated and stored.
- 

### The monitoring process:

---

1. The on – line data matrix  $Y_k$  is extracted using a sliding window with the length of  $n_1$ .
  2. Z – score normalization is performed on the on – line data matrix  $Y_k$  to obtain  $\bar{Y}_k$ .
  3. The mean vector  $\mu_{Y_k}$  and the covariance matrix  $\Sigma_{Y_k}$  of the matrix  $\bar{Y}_k$  are calculated.
  4. The vector of mean drift  $\Delta\mu_{Y_k} / \|\Delta\mu_{Y_k}\|$  between  $\bar{Y}_k$  and  $\bar{X}$  is calculated.
  5. Equation (18) is used to find the initial vector of the iteration of the optimization problem  $w_0$ .
  6. The optimization problem in Equation (16) is solved to obtain the local optimum PV  $w_k$  of the online data matrix  $\bar{Y}_k$ .
  7. Equation (13) and Equation (14) are used to obtain the vector  $h_X(w_k)$ .
  8. The fault detection threshold  $\varepsilon_k$  is determined using Equation (19).
  9. The KL divergence  $h(w_k)$  of  $\bar{X}$  and  $\bar{Y}_k$  after being projected on the local optimum PV  $w_k$  is calculated by using Equation (13).
  10. If  $h(w_k) > \varepsilon_k$  was, then assuming that  $Y_k$  is faulty is correct. Otherwise,  $Y_k$  is normal. Let  $k = k + 1$ , the next sliding window  $Y_{k+1}$  will be tested from steps 1 to 10.
- 

It should be noted that the above incipient fault detection method based on LOPV and KL divergence is suitable for systems with high requirements for incipient fault detection and sufficient computational resources. When the system has insufficient computational resources, the sliding window interval can be increased to reduce the frequency of solving the optimization problem. For example, the optimization problem can be solved every hour or daily. The LOPV obtained by solving the optimization problem can be added to the matrix  $V$  which contains  $m$  PVs obtained by the PCA method,  $V = V \cup \text{LOPV}$ . Then, the new  $V$  that contains  $m + 1$  PVs is used to detect incipient faults.

Another method to reduce the computational requirements is to set detection thresholds lower for existing fault detection methods based on PCA and KL divergence, so that the optimization problem is solved when a sliding window of on-line data exceeds the detection threshold of one PV. Then,  $V = V \cup \text{LOPV}$ , increase the detection thresholds and use the new  $V$  to detect faults. Users can flexibly select the sliding window interval or detection thresholds based on actual monitoring needs and resource constraints.

## 4. Results and Analysis

### 4.1. Numerical Example

The effectiveness of the proposed method is verified through a numerical example which includes three incipient faults. The system is modeled by

$$\begin{aligned}
 x_1 &= s_1 + s_2 + e_1 \\
 x_2 &= s_2 - s_3 + f_1 + e_2 \\
 x_3 &= s_1 - (1 + f_2)s_4 + e_3 \\
 x_4 &= (1 + f_3)x_1 + x_3 + e_4
 \end{aligned} \tag{20}$$

In Equation (20),  $[s_1, s_2, s_3, s_4]^T$  and  $[e_1, e_2, e_3, e_4]^T$  are Gaussian-distributed source signals and noises. All the source signals obey the standard normal distribution  $N(0, 1)$ , and  $f_1, f_2$  and  $f_3$  are three incipient faults, respectively, and  $[x_1, x_2, x_3, x_4]^T$  are the four variables that need be monitored [22]. Three incipient faults were simulated in this numerical



experiment. Fault  $f_1$  is the offset fault, and fault  $f_2$  and fault  $f_3$  are the gain faults. The three faults were inserted as follows:

$$\begin{aligned} f_1 &= \begin{cases} 0 & 1st - 30000th \\ 0.35 & 30001st - 60000th \end{cases} \\ f_2 &= \begin{cases} 0 & 1st - 30000th \\ 0.25 & 30001st - 60000th \end{cases} \\ f_3 &= \begin{cases} 0 & 1st - 30000th \\ 0.05 & 30001st - 60000th \end{cases} \end{aligned} \quad (21)$$

The fault detection rate (FDR) and false alarm rate (FAR) are calculated as follows:

$$FDR = \text{prob}\{h(w_k) > \varepsilon_k | H_1\} \quad (22)$$

$$FAR = \text{prob}\{h(w_k) > \varepsilon_k | H_0\} \quad (23)$$

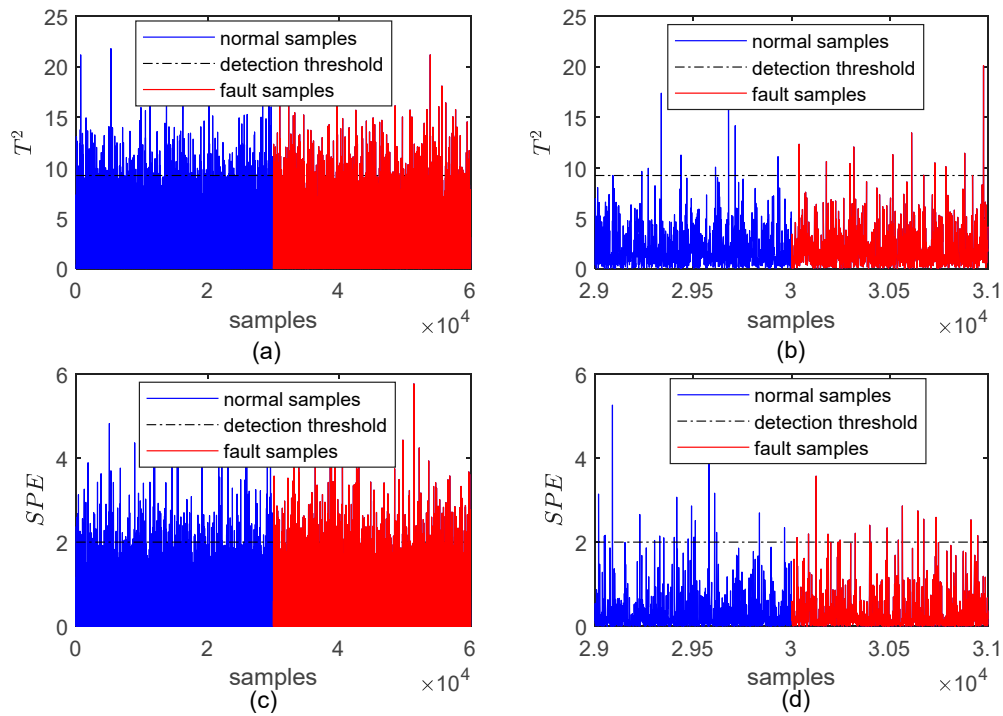
The other parameters of the numerical simulation experiment were set as follows. The total number of samples were 120,000, of which 60,000 were historical normal samples and 60,000 were online samples for testing. The sliding window length for both the historical normal data and the on-line test data in the experiment was 300. After using the sliding windows, a total of 200 windows of on-line data were obtained, of which the first 100 windows are normal windows, and the last 100 windows are fault windows. The default Signal-to-Noise Ratio (SNR) was 20 dB.

For the purpose of comparison with other PCA-based fault detection methods, the proposed method was compared with the conventional  $T^2$  statistic [33] (PCA +  $T^2$ ), the SPE statistic [33] (PCA + SPE), the method of using KL divergence to monitor the PVs in the principal subspace [21] (PCA + KLD1), and the method of using KL divergence to monitor the PVs both in the principal subspace and the residual subspace [22] (PCA + KLD2). The principal subspace was selected with a cumulative variance contribution of more than 90%. The confidence level for the  $T^2$  statistic and the SPE statistic were both set at 0.99. The significance level for the PCA + KLD1 method and the PCA + KLD2 method were both set at 0.05. The significance level for proposed method was set at 0.01.

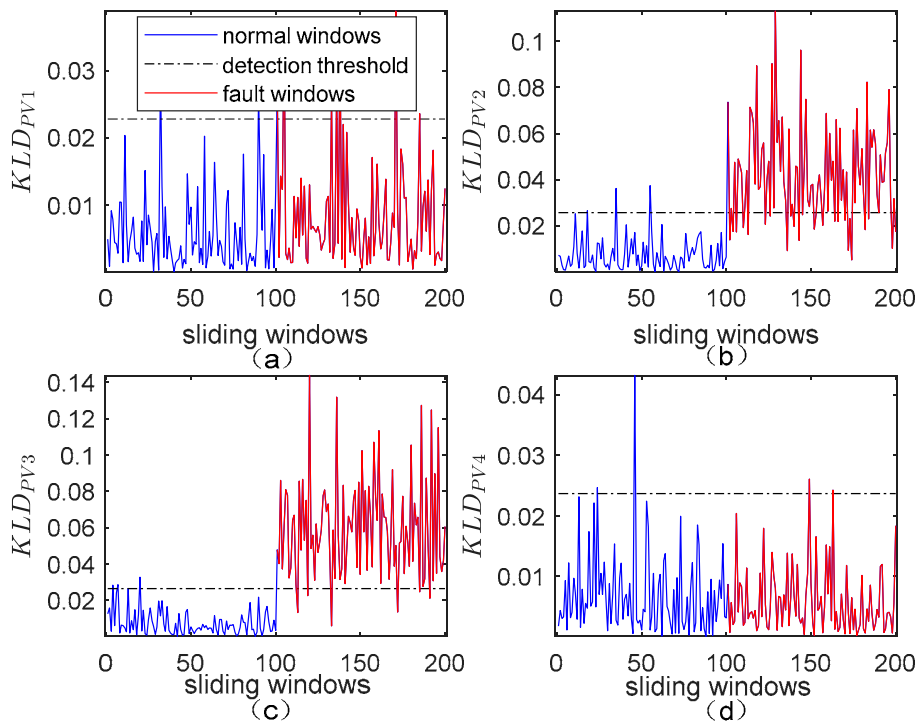
The results of the  $T^2$  statistic and the SPE statistic for the incipient fault  $f_1$  are shown in Figure 3. It can be seen from Figure 3 that the  $T^2$  statistic and SPE statistic did not change significantly before and after insertion of the fault  $f_1$ . After  $f_1$  occurred, most of the online fault samples were still within the detection threshold. The  $T^2$  statistic and the SPE statistic were equally ineffective in detecting the incipient faults  $f_2$  and  $f_3$ .

The results of detecting the incipient fault  $f_1$  by using the PCA + KLD1 method and the PCA + KLD2 method are shown in Figure 4. A comparison of Figure 4a–d shows that the second and third PV were sensitive to the incipient fault  $f_1$ . The remaining two PVs were less effective at detecting  $f_1$ . The detection result of the proposed method for the incipient fault  $f_1$  is shown in Figure 5. It can be seen from Figure 5 that the proposed method also has a good detection effect for the fault  $f_1$ .

The results of detecting the incipient fault  $f_2$  by using the PCA + KLD1 method and the PCA + KLD2 method are shown in Figure 6. Comparing Figure 6a–d shows that only the third PV of the four PVs obtained by PCA was sensitive to  $f_2$ . However, there are still about 30% of the fault windows below the fault detection threshold. The result of detecting the fault  $f_2$  by using the proposed method is shown in Figure 7. It can be seen from Figure 7 that only about 12% of the fault windows were below the fault detection threshold, which is better than the results of PVs obtained by PCA.



**Figure 3.** Results of detecting fault  $f_1$  using the  $T^2$  and squared prediction error (SPE) statistics: (a)  $T^2$  statistic result for all on-line samples; (b) the result for the  $T^2$  statistic at 1000 samples before and after insertion of the fault; (c) SPE statistic result for all on-line samples; (d) the result for the SPE statistic at 1000 samples before and after insertion of the fault.



**Figure 4.** The results of detecting the incipient fault  $f_1$  by using the principal component analysis (PCA) + Kullback-Leibler divergence (KLD)1 method and the PCA + KLD2 method: (a) the result of first projection vector (PV) for the fault  $f_1$ ; (b) the result of second PV for the fault  $f_1$ ; (c) the result of third PV for the fault  $f_1$ ; (d) the result of fourth PV for the fault  $f_1$ .

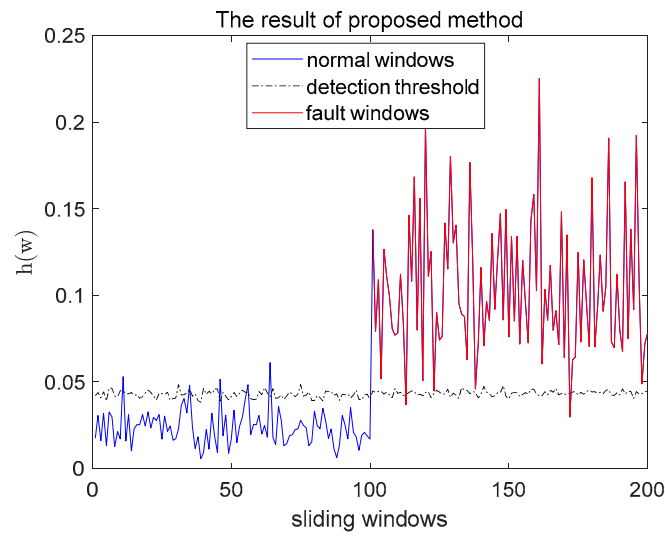


Figure 5. The detection result of the proposed method for the incipient fault  $f_1$ .

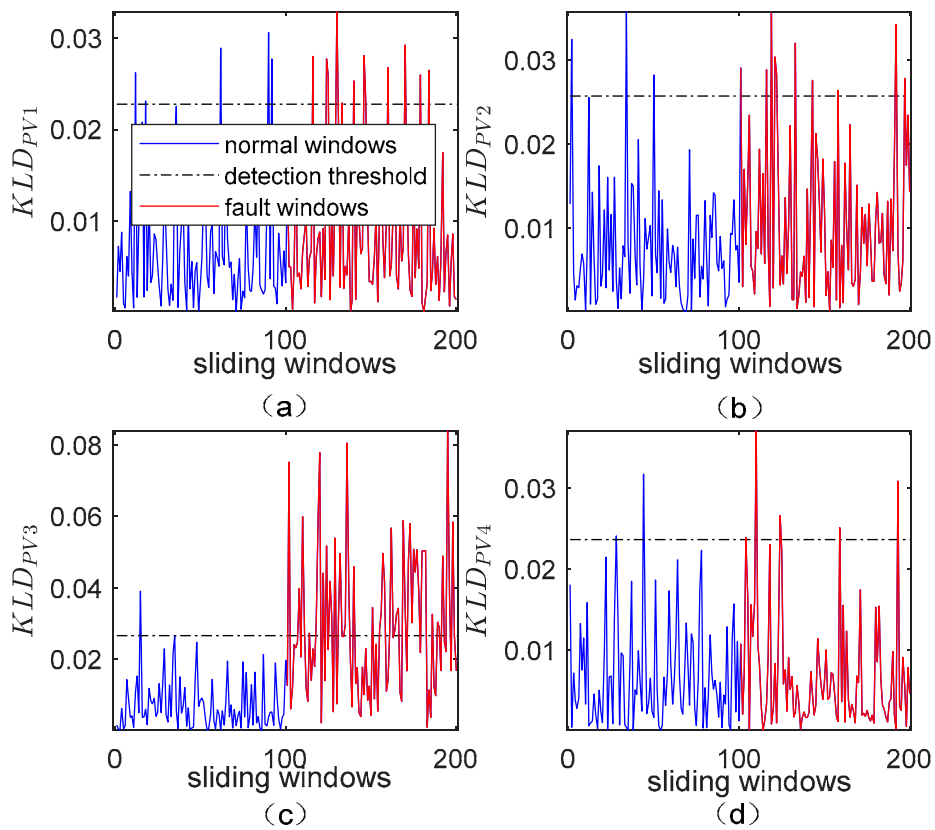


Figure 6. The results of detecting the incipient fault  $f_2$  by using the PCA + KLD1 method and the PCA + KLD2 method: (a) the result of first PV for the fault  $f_2$ ; (b) the result of second PV for the fault  $f_2$ ; (c) the result of third PV for the fault  $f_2$ ; (d) the result of fourth PV for the fault  $f_2$ .

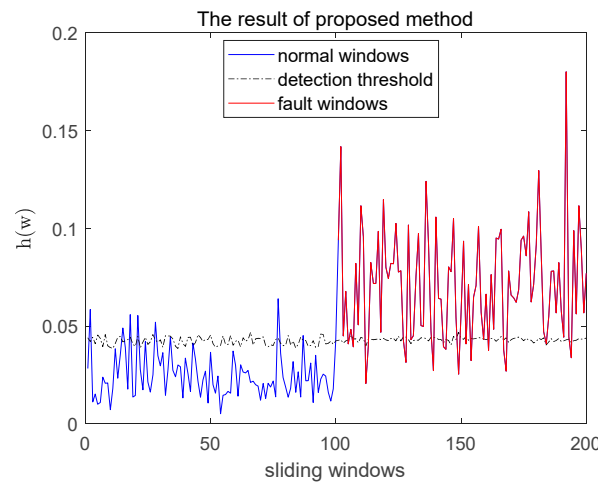


Figure 7. The detection result of the proposed method for the incipient fault  $f_2$ .

Figure 8 shows the results of detecting the incipient fault  $f_3$  by using the PCA + KLD1 method and the PCA + KLD2 method. Both the PCA + KLD1 method and the PCA + KLD2 method have poor detection results for  $f_3$ . There are still about 60% of the fault windows below the fault detection threshold. Although we could increase the detection rate by decreasing the detection threshold, because the fault windows and the normal windows are not clearly separated, the false alarm rate will also increase. It can be seen from Figure 9 that the proposed method has a very good detection effect on  $f_3$ , and the fault detection rate is 98% with similar false alarm rate as the PCA + KLD2 method.

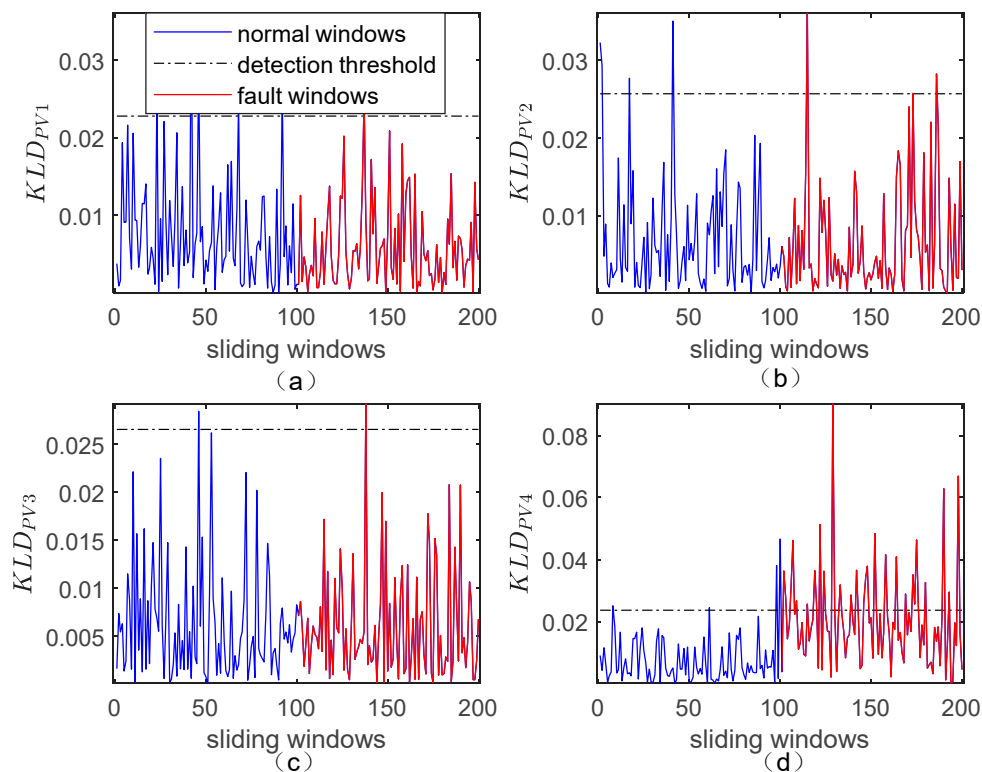


Figure 8. The results of detecting the incipient fault  $f_3$  by using the PCA + KLD1 method and the PCA + KLD2 method: (a) the result of first PV for the fault  $f_3$ ; (b) the result of second PV for the fault  $f_3$ ; (c) the result of third PV for the fault  $f_3$ ; (d) the result of fourth PV for the fault  $f_3$ .

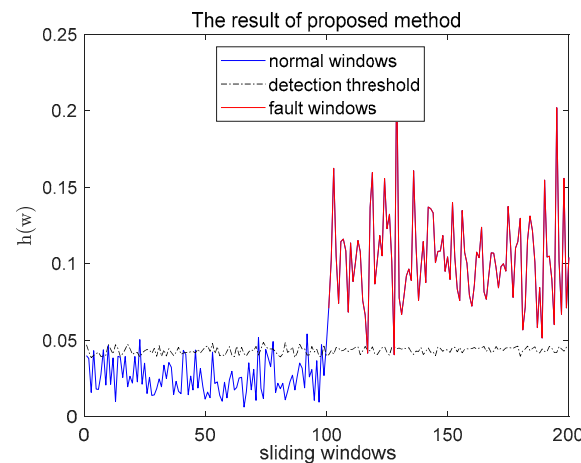


Figure 9. The detection result of the proposed method for the incipient fault  $f_3$ .

The three incipient faults were simulated randomly for a total of 100 times and then the average values of the fault detection results were calculated and are summarized in Table 1. In Table 1, the traditional fault detection methods using PCA and  $T^2$  or SPE statistics were very poor at detecting the three incipient faults. The PCA + KLD1, and PCA + KLD2 methods using fixed PVs and KL divergence greatly improved the detection of the three incipient faults compared to the  $T^2$  and SPE statistics. However, under the condition of similar false alarm rate, the average fault detection rate of the proposed method in this paper for the three faults was 60% and 30% higher than the PCA + KLD1, and PCA + KLD2 methods with fixed PVs, respectively.

Table 1. Comparison of the performance of five different methods at detecting the three faults.

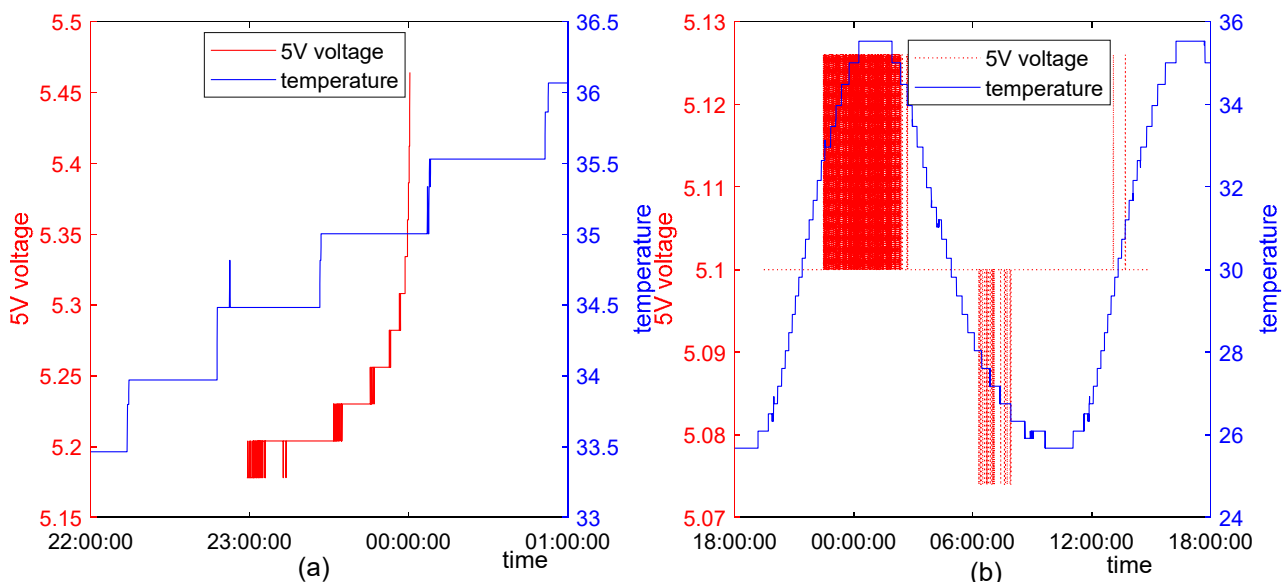
Faults	PCA+ $T^2$		PCA + SPE		PCA + KLD1		PCA + KLD2		Proposed Method	
	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
$f_1$	1.09%	1.01%	1.20%	1.01%	79.63%	4.87%	98.04%	9.50%	<b>98.11%</b>	8.10%
$f_2$	1.45%	0.98%	2.13%	1.02%	13.83%	4.63%	62.92%	9.01%	<b>84.77%</b>	7.48%
$f_3$	1.04%	0.99%	1.02%	1.01%	4.38%	3.70%	35.17%	9.20%	<b>98.33%</b>	8.05%
Average value	1.19%	1.00%	1.45%	1.01%	32.61%	4.73%	65.38%	9.24%	<b>93.74%</b>	7.88%

In Table 1, the proposed method and the PCA + KLD2 method had similar effects for the incipient fault  $f_1$ , while the detection rate of the incipient fault  $f_3$  was improved by up to 60%. After analysis, we found that the different improvements in the detection rate of faults was related to the characteristics of the actual occurrence of the incipient faults. When the occurrence of the incipient fault is in the detection sensitive region of a fixed PV obtained by the PCA method, the method of PCA + KLD2 can obtain good results. For example, the incipient fault  $f_1$  is both located in the sensitive region of the second and the third PV. Consequently, the detection rate of the PCA + KLD2 method was more than 98% for the incipient fault  $f_1$ . However, when the incipient fault is located in regions where the existing PVs are insensitive (e.g., the incipient fault  $f_3$ ), the detection rate of the proposed method is improved by 60% compared with the traditional PCA + KLD2 method which uses fixed PVs. Combining the results of detecting the above three incipient faults, with the given significance level  $\alpha$ , the PCA + KLD2 method is only sensitive to a part of the fault space. In contrast, the method proposed in this paper covers more of the fault space by finding the optimum PV for detecting incipient faults detection and is more adaptable to various possible incipient faults than the fixed PV fault detection method because it takes into account the characteristics of the actual incipient faults.

## 4.2. Satellite Spread-Spectrum Transponder Fault Case

### 4.2.1. The Phenomena and Causes of the Fault

On 3 October 2018, the downlink telemetry fault of a satellite spread spectrum transponder occurred. The ground telemetry, track and command (TT&C) station could not accept the satellite's downlink telemetry signal, and the satellite telemetry remote communication returned to normal after enabling the backup spread spectrum transponder. After the incident, the data transmitted about the spread spectrum transponder telemetry showed that the spread spectrum transponder was in a shutdown state; the 5 V power telemetry jumped to 0 V; the solid-state amplifier current jumped from 0.715 A to 0.13 A; the solid-state amplifier power jumped from 11.19 w to 0.13 w; the temperature data showed the occurrence of the fault in the high-temperature region; the 12 V power telemetry slightly decreased; and other telemetry parameters were in the normal range. The spread spectrum transponder 5 V power telemetry performance before the jump to 0 is illustrated in Figure 10a. A similar exponential increase in the 5 V voltage before the fault and the eventual power failure is illustrated in Figure 10a.



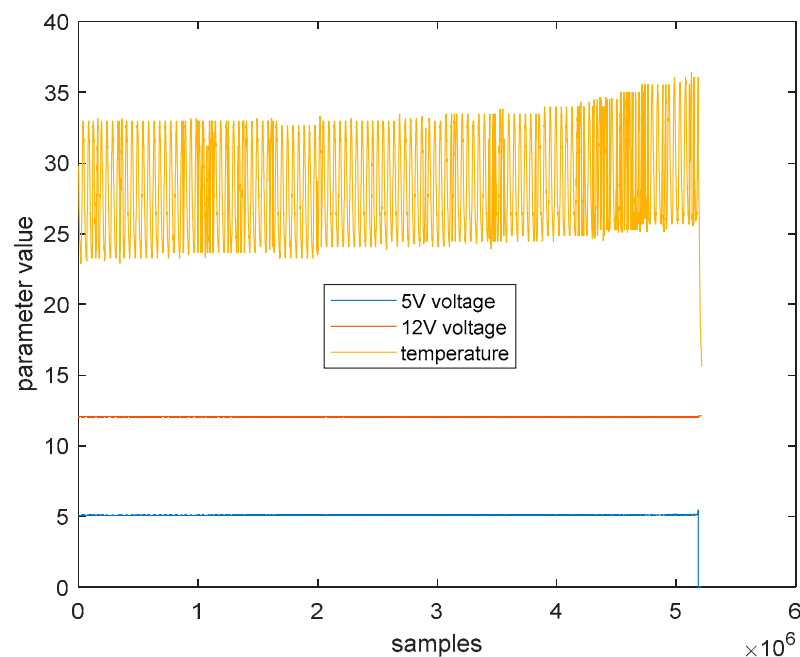
**Figure 10.** The phenomena of the spread-spectrum transponder fault: (a) the spread spectrum transponder telemetry performance before failure; (b) the relationship of the 5 V voltage and the temperature.

The post-fault analysis located the fault as an anomaly in the DC–DC module of the spread spectrum transponder power module. This module is responsible for converting the voltage of the 42 V primary power supply to generate 5 V and 12 V power and then supplying power to the spread spectrum receiver and the spread spectrum transmitter. The analysis of the telemetry data of the occurrence of the fault shows that both the 5V and 12 V voltage fluctuations of the spread spectrum transponder are correlated with the temperature fluctuations of the spread spectrum transponder. As shown in Figure 10b, periodic fluctuations in the spread spectrum transponder temperature occurred because the satellite-Z surface where the stand-alone transmitter is located fluctuates periodically due to the influence of an external heat source. The 5 V voltage telemetry also fluctuated with the temperature before the fault. The 5 V voltage fluctuation to higher values mainly occurred in the high temperature region, while the 5 V voltage fluctuation to lower values mainly occurred in the low temperature region.

### 4.2.2. Results and Comparison of Detecting the Fault

We used a total of 5,215,000 samples of real telemetry data obtained from satellite measurements and control systems from 0:0:0 on 1 July 2018 to 8:05:49 on 3 October

2018. Each sample included three telemetry parameters: the spread-spectrum transponder temperature; the 5 V voltage of the spread-spectrum transponder; and the 12 V voltage of the spread-spectrum transponder, as shown in Figure 11. Due to the influence of the variation of satellite-Z's angle to the sun, the temperature of the spread spectrum transponder drifts slowly with time. Consequently, in the experiment the temperature telemetry data were de-drifted. The sampling rate of all data was 1 Hz, and the satellite orbital period was 46,468 s. However, the data were discontinuous in time, which is due to the influence of the visible arc segment of the satellite and the ground station measurement and control resources. This resulted in telemetry data not being transmitted downward during part of the time.



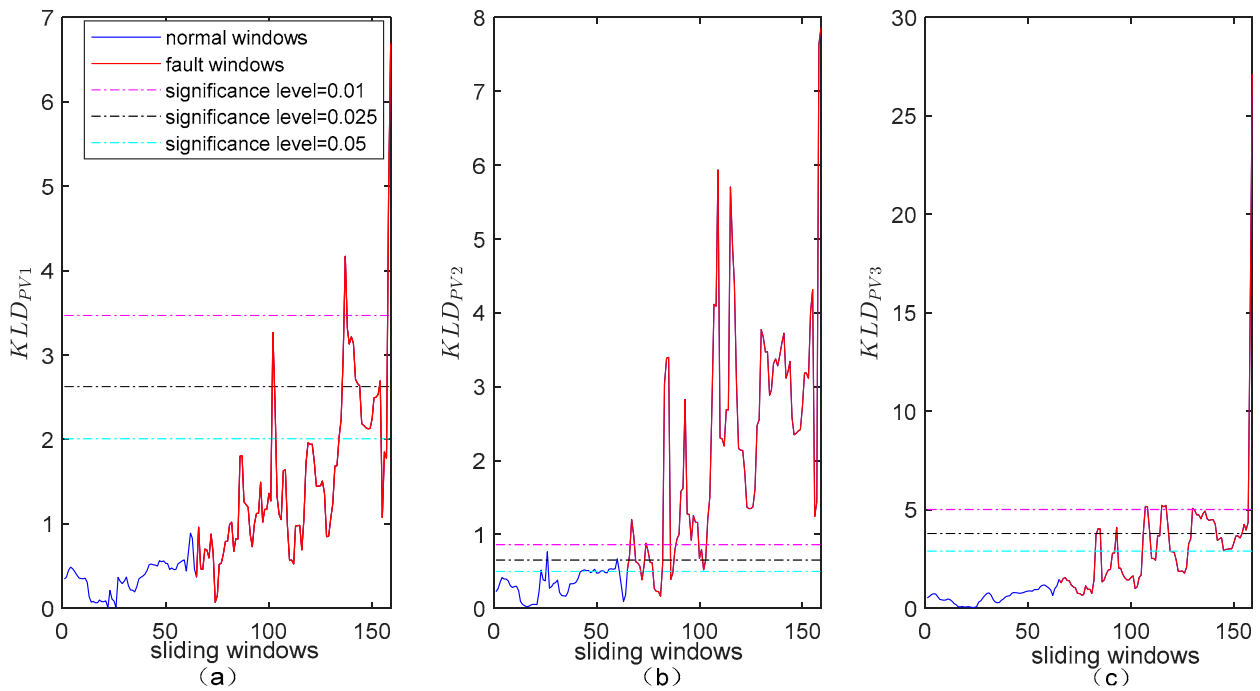
**Figure 11.** Raw telemetry data of the spread spectrum transponder.

In this article, the satellite orbital period was used as the length of the sliding window. The sliding window interval was 10,000 s, and a total of 539 sliding windows were retained with more than 20,000 samples within each sliding window. The first 380 sliding windows were selected as normal windows, and the last 159 sliding windows were used as on-line test windows. The first 64 windows in the on-line test windows were the normal windows, and the windows after that were the fault windows. The significance level of the PCA + KLD1 method and PCA + KLD2 method participating in the comparison were set to 0.05, 0.025 and 0.01, respectively, and significance level of the proposed method was set to 0.01.

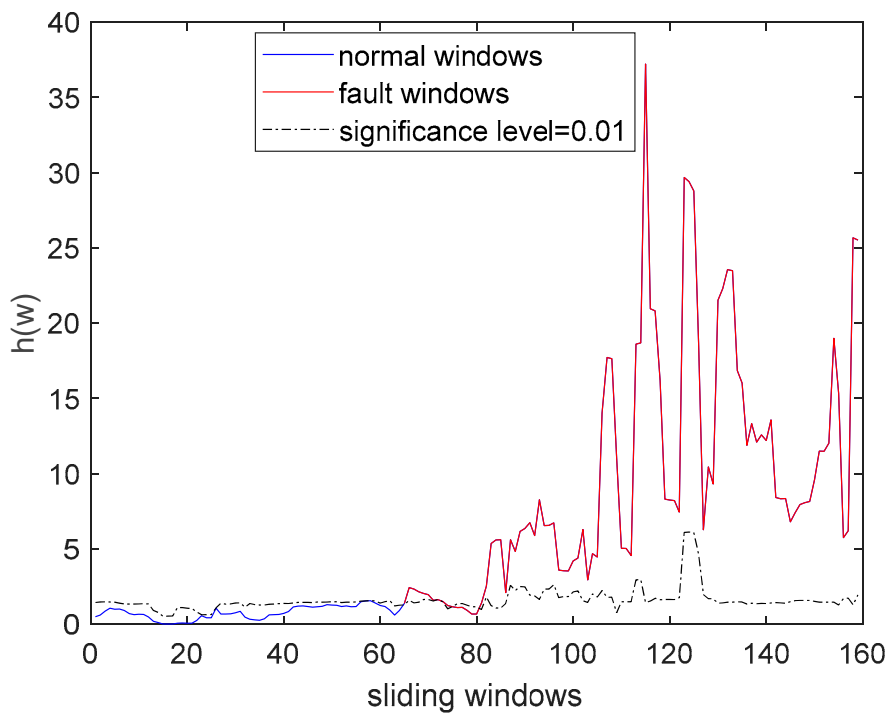
Figure 12a–c shows the results of the spread spectrum transponder fault detection on the three PVs obtained from the PCA method. It can be seen from Figure 12 that only the second PV was sensitive to the fault. With a significance level of 0.01, there were still 23% of the fault windows located below the detection threshold. With a significance level of 0.025, 15% of the fault windows were still located below the detection threshold and a normal window was considered to be a fault window.

In comparison, Figure 13 shows the results of the proposed method for the spread-spectrum transponder fault. It is evident that the detection results on the local optimum PVs were better than those on the fixed PVs. Most of the fault windows were located above the detection threshold, with a high fault detection rate of 92.63% and a false alarm rate of 0%. With the significance level of 0.01, the PCA + KLD2 method can also obtain a 92.63% fault detection rate. However, 23.44% of the normal windows were considered to be fault windows. The fault detection results were calculated and are summarized in Table 2. This

real case of detecting the fault in a satellite's spread spectrum transponder further verifies the effectiveness and superiority of the method proposed in this paper.



**Figure 12.** The results of detecting the spread-spectrum transponder fault by using the PCA + KLD1 method and the PCA + KLD2 method at different significance levels: (a) the result of first PV for the fault; (b) the result of second PV for the fault; (c) the result of third PV for the fault.



**Figure 13.** The detection result of the proposed method for the spread-spectrum transponder fault.



**Table 2.** Comparison of the performance of three different methods at detecting the spread-spectrum transponder fault.

Methods	PCA + KLD1			PCA + KLD2			Proposed Method
	0.05	0.025	0.01	0.05	0.025	0.01	0.01
FDR	76.84%	84.21%	92.63%	76.84%	84.21%	92.63%	92.63%
FAR	0%	3.13%	23.44%	0%	3.13%	23.44%	0%

## 5. Conclusions

The efficient detection and solving of incipient faults will effectively reduce the losses and the hazards they cause. Under the assumption that the variables obey a multidimensional Gaussian distribution, this paper models the optimum PV for detecting incipient faults as an optimization problem. Then, the validity of the proposed method is assessed using a numerical simulation example and an actual satellite fault case. Two different mitigation schemes are proposed to address the computational problem of the proposed method, in which the user can flexibly choose the sliding window interval or the alarm threshold according to the actual monitoring requirements and resource constraints. The method only has been proved in three faults. If the system was non-Gaussian or nonlinear, the detection effect of the proposed method might decrease. The future works related with improvement of the method can come from time efficiency and expansion to nonlinear systems. Although the proposed method is based on detecting incipient faults in satellites, it is not specific to this application of detecting faults in satellites and can be extended to other applications where multivariate statistical analysis can be used to detect incipient faults.

**Author Contributions:** Conceptualization, G.Z. and Q.Y.; software, G.Z.; validation, G.Z., Q.Y. and G.L.; formal analysis, G.Z.; investigation, G.Z.; resources, Q.Y. and J.L.; data curation, Q.Y. and J.L.; writing—original draft preparation, G.Z.; writing—review and editing, Q.Y., G.L. and L.W.; visualization, L.W.; supervision, G.L.; project administration, G.L.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Beidou Navigation In-orbit Support System, grant number JKBDZGDH01.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available on request due to restrictions eg privacy or ethical. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

FAR	false alarm rate
FDR	fault detection rate
GPR	Gaussian Process Regression
IMS	Inductive Monitoring System
KL	Kullback–Leibler
KLD	KL divergence
LOPV	local optimum PV
LSTM	Long Short-Term Memory
OS-SVM	One Class Support Vector Machine
PCA	principal component analysis
PDF	probability density function
PV	projection vector

## Mathematical Notations

Symbol	Size	Description
$X$	$n_0 \times m$	Normal data matrix.
$n_0$	$1 \times 1$	Number of normal samples
$m$	$1 \times 1$	Number of variables of a system
$\bar{X}$	$n_0 \times m$	Normal data matrix after standardized.
$\Sigma$	$m \times m$	Covariance matrix of $\bar{X}$
$V$	$m \times m$	Eigenvector matrix of $\bar{X}$ , each column of the matrix $V$ is a projection vector
$n_1$	$1 \times 1$	Number of on-line samples
$Y$	$n_1 \times m$	On-line data matrix.
$v_j$	$m \times 1$	$j$ th column of the matrix $V$
$\varepsilon_j$	$1 \times 1$	Detection threshold of the PV $v_j$
$\alpha$		Significance level
$\mu_x$	$m \times 1$	Mean vector of the joint Gaussian distribution that the normal data obeyed
$\Sigma_x$	$m \times m$	Covariance matrix of the joint Gaussian distribution that the normal data obeyed
$w$	$m \times 1$	Optimum PV
$\mu_y$	$m \times 1$	Mean vector of the joint Gaussian distribution that the on-line data obeyed
$\Sigma_y$	$m \times m$	Covariance matrix of the joint Gaussian distribution that the on-line data obeyed
$\Delta\mu$	$m \times 1$	Mean deviation vector of $\mu_x$ and $\mu_y$
$h(w)$	$1 \times 1$	KL divergence of the projections of the normal data matrix $X$ and the on-line data matrix $Y$ on $w$
$h_X(w)$	$r \times 1$	Vector of the KL divergences of the projections between all submatrices and the normal data matrix $X$ on $w$
$J(w)$	$1 \times 1$	Relative KL divergence of the projection of the normal data matrix $X$ and the on-line data matrix $Y$ on $w$
$w_0$	$m \times 1$	Initial iteration vector
$Y_k$	$n_1 \times m$	On-line data extracted by the $k$ th sliding window.
$w_k$	$m \times 1$	Local optimum PV of the on-line data $Y_k$ and the normal data $X$
$\varepsilon_k$	$1 \times 1$	Detection threshold of the local optimal PV $w_k$

## References

- Jiang, H.; Zhang, K.; Wang, J.; Wang, X.; Huang, P. Anomaly Detection and Identification in Satellite Telemetry Data Based on Pseudo-Period. *Appl. Sci.* **2019**, *10*, 103. [\[CrossRef\]](#)
- Ezhilarasu, C.M.; Skaf, Z.; Jennions, I.K. The application of reasoning to aerospace Integrated Vehicle Health Management (IVHM): Challenges and opportunities. *Prog. Aerosp. Sci.* **2019**, *105*, 60–73. [\[CrossRef\]](#)
- Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 1–58. [\[CrossRef\]](#)
- Tipaldi, M.; Bruenjes, B. Survey on Fault Detection, Isolation, and Recovery Strategies in the Space Domain. *J. Aerosp. Inf. Syst.* **2015**, *12*, 235–256. [\[CrossRef\]](#)
- Youssef, A.; Delpha, C.; Diallo, D. Enhancement of incipient fault detection and estimation using the multivariate Kullback-Leibler Divergence. In Proceedings of the 2016 24th European Signal Processing Conference (EUSIPCO), Budapest, Hungary, 28 August–2 September 2016; pp. 1408–1412.
- Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback-Leibler Divergence Measure for Multivariate Skew-Normal Distributions. *Entropy* **2012**, *14*, 1606–1626. [\[CrossRef\]](#)
- Chen, H.; Jiang, B.; Lu, N.; Mao, Z. Deep PCA Based Real-Time Incipient Fault Detection and Diagnosis Methodology for Electrical Drive in High-Speed Trains. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4819–4830. [\[CrossRef\]](#)
- Narasimhan, S.; Brownston, L. HyDE—A General Framework for Stochastic and Hybrid Model-based Diagnosis. *Proc. DX* **2007**, *7*, 162–169.
- Deb, S.; Pattipati, K.R.; Shrestha, R. QSI's Integrated Diagnostics Toolset. In Proceedings of the 1997 IEEE Autotestcon Proceedings AUTOTESTCON'97, Anaheim, CA, USA, 22–25 September 1997; pp. 408–421.
- Kurien, J.; R-Moreno, M.D. Costs and Benefits of Model-based Diagnosis. In Proceedings of the 2008 IEEE Aerospace Conference, Big Sky, MT, USA, 1–8 March 2008; pp. 1–14.
- Luo, B.; Wang, H.; Liu, H.; Li, B.; Peng, F. Early Fault Detection of Machine Tools Based on Deep Learning and Dynamic Identification. *IEEE Trans. Ind. Electron.* **2019**, *66*, 509–518. [\[CrossRef\]](#)
- Duncan Imbassahy, D.W.; Costa Marques, H.; Conceição Rocha, G.; Martinetti, A. Empowering Predictive Maintenance: A Hybrid Method to Diagnose Abnormal Situations. *Appl. Sci.* **2020**, *10*, 6929. [\[CrossRef\]](#)
- Pang, J.; Liu, D.; Peng, Y.; Peng, X. Optimize the Coverage Probability of Prediction Interval for Anomaly Detection of Sensor-Based Monitoring Series. *Sensors* **2018**, *18*, 967. [\[CrossRef\]](#)
- David, I.; Rodney, M.; Mark, S.; Spirkovska, L. General Purpose Data-Driven System Monitoring for Space Operations—AIAA Infotech@Aerospace Conference (AIAA). In Proceedings of the Aiaa Infotech, Seattle, WA, USA, 6–9 April 2009; pp. 26–44.

15. Schwabacher, M.; Oza, N.; Matthews, B. Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring. *J. Aeros. Comp. Inf. Com.* **2009**, *6*, 464–482. [[CrossRef](#)]
16. Pang, J.; Liu, D.; Peng, Y.; Peng, X. Anomaly detection based on uncertainty fusion for univariate monitoring series. *Measurement* **2017**, *95*, 280–292. [[CrossRef](#)]
17. Hundman, K.; Constantinou, V.; Laporte, C.; Colwell, I.; Soderstrom, T. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 387–395.
18. Tariq, S.; Lee, S.; Shin, Y.; Lee, M.S.; Jung, O.; Chung, D.; Woo, S.S. Detecting Anomalies in Space using Multivariate Convolutional LSTM with Mixtures of Probabilistic PCA. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2123–2133.
19. Yin, S.; Ding, S.X.; Xie, X.; Luo, H. A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6418–6428. [[CrossRef](#)]
20. Li, X.; Zhang, X.; Zhang, P.; Zhu, G. Fault Data Detection of Traffic Detector Based on Wavelet Packet in the Residual Subspace Associated with PCA. *Appl. Sci.* **2019**, *9*, 3491. [[CrossRef](#)]
21. Harmouche, J.; Delpha, C.; Diallo, D. Incipient fault detection and diagnosis based on Kullback-Leibler divergence using Principal Component Analysis: Part I. *Signal Process.* **2014**, *94*, 278–287. [[CrossRef](#)]
22. Chen, H.; Jiang, B.; Lu, N. An improved incipient fault detection method based on Kullback-Leibler divergence. *ISA Trans.* **2018**, *79*, 127–136. [[CrossRef](#)]
23. Youssef, A.; Delpha, C.; Diallo, D. An optimal fault detection threshold for early detection using Kullback-Leibler Divergence for unknown distribution data. *Signal Process.* **2016**, *120*, 266–279. [[CrossRef](#)]
24. Song, B.; Zhou, X.; Tan, S.; Shi, H.; Zhao, B.; Wang, M. Process Monitoring via Key Principal Components and Local Information Based Weights. *IEEE Access* **2019**, *7*, 15357–15366. [[CrossRef](#)]
25. Luo, L.; Bao, S.; Mao, J. Adaptive Selection of Latent Variables for Process Monitoring. *Ind. Eng. Chem. Res.* **2019**, *58*, 9075–9086. [[CrossRef](#)]
26. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
27. Cherry, G.A.; Qin, S.J. Multiblock Principal Component Analysis Based on a Combined Index for Semiconductor Fault Detection and Diagnosis. *IEEE Trans. Semicond. Manuf.* **2006**, *19*, 159–172. [[CrossRef](#)]
28. Doersch, C. Tutorial on Variational Autoencoders. *arXiv* **2016**, arXiv:1606.05908.
29. Huang, Y.; Zhang, Y.; Chambers, J.A. A Novel Kullback-Leibler Divergence Minimization-Based Adaptive Student's t-Filter. *IEEE Trans. Signal Process.* **2019**, *67*, 5417–5432. [[CrossRef](#)]
30. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat. Mar.* **1951**, *22*, 79–86. [[CrossRef](#)]
31. Byrd, R.H.; Gilbert, J.C.; Nocedal, J. A trust region method based on interior point techniques for nonlinear programming. *Math. Program.* **2000**, *89*, 149–185. [[CrossRef](#)]
32. Coleman, T.F.; Li, Y. On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. *Math. Program.* **1994**, *67*, 189–224. [[CrossRef](#)]
33. Nassar, B.; Hussein, W.; Mokhtar, M. Space Telemetry Anomaly Detection Based on Statistical PCA Algorithm. In Proceedings of the ICSSC 2015: International Conference on Satellite and Space Communications, Paris, France, 27–28 August 2015; pp. 637–645.