A Satellite Incipient Fault Detection Method Based on Decomposed Kullback–Leibler Divergence

Ge Zhang [1,2], Qiong Yang [1], Guotong Li [1,2,3,*], Jiaxing Leng [1] and Mubiao Yan [1,2]

1   Innovation Academy for Microsatellites of CAS, Shanghai 201203, China; zhangge1@shanghaitech.edu.cn (G.Z.); yangq@microsate.com (Q.Y.); lengjiaxing0520@163.com (J.L.); yanmb@microsate.com (M.Y.)
2   University of Chinese Academy of Sciences, Beijing 100049, China
3   School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
*   Correspondence: ligt@microsate.com

Abstract: Detection of faults at the incipient stage is critical to improving the availability and continuity of satellite services. The application of a local optimum projection vector and the Kullback–Leibler (KL) divergence can improve the detection rate of incipient faults. However, this suffers from the problem of high time complexity. We propose decomposing the KL divergence in the original optimization model and applying the property of the generalized Rayleigh quotient to reduce time complexity. Additionally, we establish two distribution models for subfunctions $F_1(w)$ and $F_3(w)$ to detect the slight anomalous behavior of the mean and covariance. The effectiveness of the proposed method was verified through a numerical simulation case and a real satellite fault case. The results demonstrate the advantages of low computational complexity and high sensitivity to incipient faults.

Keywords: Kullback–Leibler (KL) divergence; fault detection; condition monitoring; incipient fault; generalized Rayleigh quotient (GRQ); optimum projection vector (PV)

1. Introduction

Due to the vigorous development of the space industry, the number of satellites in orbit has increased to meet various needs, such as navigation [1], communication [2], meteorology [3], and earth observation [4]. However, satellites face the risk of abnormalities or experience failure because of high-energy particles in space, electrostatic discharge, and cycle temperature [5–7]. Because serious faults may occur due to the continuous deterioration of incipient faults [8], timely and accurate detection of incipient faults can reserve sufficient processing time for satellite operation and maintenance system, which is of great significance to guarantee the availability and continuity of satellite services [9].

During the past three decades, the problem of satellite fault detection has been extensively studied in various studies [10–13]. In traditional satellite fault detection methods, such as threshold-based methods [14,15] and model-based methods [16,17], the thresholds or the models required for fault detection necessitate manual setting. Therefore, the performance of these fault detection methods heavily relies on the experience of experts [18]. In recent years, data-driven fault detection methods have eliminated this heavy dependence on expert experience and become a popular research field [19–22]. These methods establish normal models based on satellite normal historical data, and then compare the online data with the normal models to assess whether the online data is faulty. However, the methods proposed in the existing literature are mainly applied to serious faults, and an extremely small amount of research and application relates to incipient faults of satellites. The amplitudes of incipient faults are small compared to system signals, usually ranging from 1% to 10% [23], which are easily masked by normal system variations [24]. Therefore, satellite incipient fault detection is a daunting task [25].

Ji et al. [26] found that the introduction of smoothing technology can improve the detection rate of incipient faults. Jinane et al. [27] proposed an incipient fault detection method based on principal component analysis (PCA) and the KL divergence, but this method only considered the incipient faults in the principal component subspace. Chen et al. [23] proposed an improved method that monitors anomalous behaviors in principal and residual subspaces. Gautam et al. [28] presented a sensor incipient fault detection method based on a Kalman Filter and the KL divergence. Deng et al. [29] combined two-step localized kernel PCA with the KL divergence for nonlinear system incipient fault monitoring. Zhang et al. [30] proposed that the principal components obtained by PCA are not necessarily the optimum projection vector (PV) for detecting incipient faults. Furthermore, the problem of finding the optimum PV was modeled as an optimization model. Using local optimum PV in real time makes the method more sensitive to incipient faults, but it also raises the problem of high computational complexity. For this reason, this paper proposes a new incipient fault detection method with lower computational complexity by decomposing the KL divergence. The main contributions of this work are summarized as follows:

1. We analyzed the necessity and feasibility of decomposing the KL divergence in the optimization model.
2. We constructed two distribution models for subfunctions $F_1(w)$ and $F_3(w)$.
3. The effectiveness of the proposed method was verified through a numerical case and a real satellite fault case.

This paper is organized as follows. The generalized Rayleigh quotient (GRQ) and original optimization model are introduced in Section 2. The fault detection method based on the decomposed KL divergence is presented in detail in Section 3. In Section 4, the proposed method is illustrated and analyzed through two cases. Finally, conclusions are given in Section 5.

2. Preliminary

In this section, we introduce the definition and property of the generalized Rayleigh quotient and note the problem of original optimization model.

2.1. Generalized Rayleigh Quotient (GRQ)

The GRQ is defined as follows [31]:

$$R(A, B, x) = \frac{x^T A x}{x^T B x} \tag{1}$$

where x is a non-zero vector, A is a symmetric matrix, and B is a positive definite symmetric matrix. The GRQ has a critical property that the maximum value of $R(A, B, x)$ is equal to the maximum eigenvalue of matrix $B^{-1}A$ [32]; that is, $R(A, B, x) \leq \lambda_{\max}$, where $\lambda_{\max}$ is the maximum eigenvalue of the matrix $B^{-1}A$. In addition, the optimum vector x which maximizes $R(A, B, x)$ is the eigenvector corresponding to the maximum eigenvalue [32].

The sum of two GRQs is defined as follows [33]:

$$R(A_1, B_1, A_2, B_2, x) = \frac{x^T A_1 x}{x^T B_1 x} + \frac{x^T A_2 x}{x^T B_2 x} \tag{2}$$

where x is a non-zero vector, both A_1 and A_2 are symmetric matrices, and both B_1 and B_2 are positive definite symmetric matrices.

Because iteration is not required, the maximum value of a single GRQ can be quickly obtained by directly applying the property of the GRQ. Regarding the maximum value of the sum of two GRQs, according to Reference [33], the time complexity of maximizing the sum of two GRQs is NP-hard. Prominently, accurate algorithms cannot solve large instances of such a problem, and approximate algorithms are necessary.

2.2. Original Optimization Model

Under the assumption that the data obey a multidimensional Gaussian distribution, and using the KL divergence to detect incipient faults, the problem of finding the optimum projection vector (PV) is modeled as follows [30]:

$$\begin{cases} \min_w - h(w) \\ s.t.w^T w = 1; \forall w_i, -1 \leq w_i \leq 1, i \in [1, m] \end{cases} \tag{3}$$

$$h(w) = \frac{1}{2}\left[\frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Sigma_x w}{w^T \Sigma_y w} + \left(\Delta\mu^T w\right)^2 \left(\frac{1}{w^T \Sigma_x w} + \frac{1}{w^T \Sigma_y w} \right) - 2 \right] \tag{4}$$

In Equations (3) and (4), w is a PV, $h(w)$ is the KL divergence of the projections of normal historical data X and online data Y. Both the normal historical data and the online data obey m dimensional joint Gaussian distributions, $X \sim N(\mu_x, \Sigma_x), Y \sim N(\mu_y, \Sigma_y)$ [30]. Let $\Delta\mu = \mu_y - \mu_x, \Delta\Sigma = \Sigma_y - \Sigma_x$; the KL divergence $h(w)$ can be expressed as the sum of two GRQs, as shown in Equation (5):

$$\begin{aligned} h(w) &= \frac{1}{2}\left[\frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Sigma_x w}{w^T \Sigma_y w} + \left(\Delta\mu^T w\right)^2 \left(\frac{1}{w^T \Sigma_x w} + \frac{1}{w^T \Sigma_y w} \right) - 2 \right] \\ &= \frac{1}{2}\left[\frac{w^T \Sigma_y w}{w^T \Sigma_x w} + \frac{w^T \Delta\mu\Delta\mu^T w}{w^T \Sigma_x w} \right] + \frac{1}{2}\left[\frac{w^T \Sigma_x w}{w^T \Sigma_y w} + \frac{w^T \Delta\mu\Delta\mu^T w}{w^T \Sigma_y w} \right] - 1 \\ &= \frac{1}{2}\left[\frac{w^T \left(\Sigma_y + \Delta\mu\Delta\mu^T\right) w}{w^T \Sigma_x w} + \frac{w^T \left(\Sigma_x + \Delta\mu\Delta\mu^T\right) w}{w^T \Sigma_y w} \right] - 1 \\ &= \frac{1}{2}\left(\frac{w^T A_1 w}{w^T B_1 w} + \frac{w^T A_2 w}{w^T B_2 w} \right) - 1 \end{aligned} \tag{5}$$

where $A_1 = \Sigma_y + \Delta\mu\Delta\mu^T, A_2 = \Sigma_x + \Delta\mu\Delta\mu^T, B_1 = \Sigma_x, B_2 = \Sigma_y$. According to the property of the covariance matrix, both Σ_x and Σ_y in Equation (5) are non-negative symmetric matrices. This paper considers only the case that both of the matrices are positive definite symmetric matrices to satisfy the condition of the GRQ. If the influence of the coefficient 0.5 and the constant -1 is ignored, Equation (3) can be equally expressed as the maximization of the sum of two GRQs:

$$\begin{cases} \max_w \frac{w^T A_1 w}{w^T B_1 w} + \frac{w^T A_2 w}{w^T B_2 w} \\ s.t.w^T w = 1; \forall w_i, -1 \leq w_i \leq 1, i \in [1, m] \end{cases} \tag{6}$$

As stated in Section 2.1, the time complexity of solving the optimization problem in Equation (6) is NP-hard. Similarly, the optimization problem in Equation (3) is NP-hard. In Reference [30], a ready-made optimization solution tool (the fmincon function in MATLAB) is used to solve the optimization problem. However, this method can only obtain the local optimum PV, rather than the global optimum PV. Additionally, with the gradual increase in the number of variables to be monitored, the time complexity of iteration becomes more prominent. Therefore, this study aimed to determine an approximate algorithm with lower time complexity.

3. Incipient Fault-Detection Method Based on Decomposed KL Divergence

In this section, we propose the idea of decomposing the KL divergence and built two distribution models to detect incipient faults.

3.1. Decomposed KL Divergence

As stated in Section 2.1, the maximum value of a single GRQ can be quickly obtained by applying the property of the GRQ. Therefore, this paper attempts to decompose $h(w)$ to reduce time complexity. Specifically, we attempt to decompose $h(w)$ into the sum of multiple GRQs, and then calculate the maximum value and the optimum PV of each GRQ.

Under the guidance of this idea, the KL divergence $h(w)$ can be decomposed into the sum of four GRQs, as expressed in Equations (8)–(11):

$$h(w) = \frac{1}{2}(F_1(w) + F_2(w) + F_3(w) + F_4(w)) - 1 \tag{7}$$

$$F_1(w) = \frac{w^T \Sigma_y w}{w^T \Sigma_x w} \tag{8}$$

$$F_2(w) = \frac{w^T \Sigma_x w}{w^T \Sigma_y w} \tag{9}$$

$$F_3(w) = \frac{w^T \Delta\mu\Delta\mu^T w}{w^T \Sigma_x w} \tag{10}$$

$$F_4(w) = \frac{w^T \Delta\mu\Delta\mu^T w}{w^T \Sigma_y w} \tag{11}$$

where $F_1(w)$, $F_2(w)$, $F_3(w)$, and $F_4(w)$ are collectively referred to as the subfunctions of $h(w)$. In Equations (8)–(11), both Σ_x and Σ_y are positive definite symmetric matrices, so that each subfunction of $h(w)$ satisfies the form of the GRQ. Therefore, we can obtain the maximum value and optimum PV of each subfunction using the property of the GRQ.

Clearly, the maximization of each subfunction may not maximize the original function. For instance, we can find a PV w_1 that maximizes $F_1(w)$, but w_1 does not necessarily maximize $h(w)$. In this case, what is the point of decomposing $h(w)$? According to reference [30], the ultimate goal of maximizing $h(w)$ is to determine the PV w that is most sensitive to the incipient fault; that is, our ultimate goal is to detect the incipient fault. From the aspect of fault detection, although the PV obtained by maximizing the subfunction may not be optimal for the original function, the PV has its own value if it can detect the fault and be obtained in a fast manner.

Which subfunctions of $h(w)$ are effective and can be solved quickly? After analysis, two subfunctions $F_1(w)$ and $F_3(w)$ are selected. According to Equation (8) and the property of GRQ, the maximum value of $F_1(w)$ is the maximum eigenvalue of matrix $\Sigma_x^{-1}\Sigma_y$. Furthermore, the optimum PV of $F_1(w)$ is the eigenvector corresponding to the maximum eigenvalue. Similarly, According to Equation (10) and the property of the GRQ, the optimum PV of $F_3(w)$ is the eigenvector corresponding to the maximum eigenvalue of matrix $\Sigma_x^{-1}\Delta\mu\Delta\mu^T$. Let the optimum PVs of $F_1(w)$ and $F_3(w)$ be w_{F1} and w_{F3}, respectively.

3.2. Construction of Fault Detection Models

The optimum PVs w_{F1} and w_{F3} only provide two optimal perspectives of observation which the on-line data and the normal historical data are the easiest to distinguish that can be most easily distinguished by the online data and the normal historical data. We still lack some measurement indices to test whether a fault has occurred in the on-line data Y. This section uses $F_1(w)$ and $F_3(w)$ as the deviation measurement indices. Due to noise, both $F_1(w)$ and $F_3(w)$ fluctuate in their normal ranges when there is no fault in Y. However, $F_1(w)$ or $F_3(w)$ are outside of the normal ranges when a fault occurs in Y.

The normal ranges of $F_1(w)$ or $F_3(w)$ are the key to fault detection. To obtain them, we assume that the normal historical data X and the online data Y obey two m-dimensional joint Gaussian distributions, $X \sim N(\mu_x, \Sigma_x)$ and $Y \sim N(\mu_y, \Sigma_y)$, respectively. Denote the projections of X and Y onto the vector w_{F1} as p_{F1} and q_{F1}, respectively. According to the property of m-dimensional joint Gaussian distribution, p_{F1} and q_{F1} obey one-dimensional Gaussian distributions $p_{F1} \sim N(w_{F1}^T\mu_x, w_{F1}^T\Sigma_x w_{F1})$ and $q_{F1} \sim N(w_{F1}^T\mu_y, w_{F1}^T\Sigma_y w_{F1})$, respectively [34]. The relationship of $F_1(w)$, w_{F1}, Σ_x, and Σ_y is presented in Equation (12):

$$F_1(w) = \frac{w_{F1}^T \Sigma_y w_{F1}}{w_{F1}^T \Sigma_x w_{F1}} \tag{12}$$

Denote the projections of X and Y onto the vector w_{F3} as p_{F3} and q_{F3}, respectively. Similarly, according to the property of m-dimensional joint Gaussian distribution, p_{F3} and q_{F3} obey one-dimensional Gaussian distributions $p_{F3} \sim N\left(w_{F3}{}^T \mu_x, w_{F3}{}^T \Sigma_x w_{F3}\right)$ and $q_{F3} \sim N\left(w_{F3}{}^T \mu_y, w_{F3}{}^T \Sigma_y w_{F3}\right)$, respectively [34]. The relationship of $F_3(w)$, w_{F3}, Σ_x and $\Delta\mu$ is presented in Equation (13):

$$F_3(w) = \frac{w_{F3}{}^T \Delta\mu \Delta\mu^T w_{F3}}{w_{F3}{}^T \Sigma_x w_{F3}} \tag{13}$$

Because the normal historical data X are obtained before fault detection, and the optimum PVs w_{F1} and w_{F3} are obtainable from Section 3.1, it can be considered that Σ_x, μ_x, w_{F1}, and w_{F3} in Equations (12) and (13) are known and invariable. Furthermore, the mean offset vector $\Delta\mu$ and the covariance matrix Σ_y related to Y are unknown and variable. Because Σ_x, w_{F1}, and w_{F3} are known, we can assume $w_{F1}^T \Sigma_x w_{F1} = c_{F1}$ and $w_{F3}^T \Sigma_x w_{F3} = c_{F3}$, where both c_{F1} and c_{F3} are constants. Hence, Equations (14) and (15) can be obtained:

$$F_1(w) = \frac{w_{F1}{}^T \Sigma_y w_{F1}}{c_{F1}} \tag{14}$$

$$F_3(w) = \frac{w_{F3}{}^T \Delta\mu \Delta\mu^T w_{F3}}{c_{F3}} \tag{15}$$

To obtain the normal ranges of $F_1(w)$ or $F_3(w)$, it is supposed that the fault-free online data Y are obtained by sampling the joint Gaussian distribution obeyed by X. Because p_{F1} and q_{F1} are the projections of X and Y onto the vector w_{F1}, respectively, we can consider that q_{F1} is obtained by sampling the one-dimensional Gaussian distribution obeyed by p_{F1}. Similarly, we can consider that q_{F3} is obtained by sampling the one-dimensional Gaussian distribution obeyed by p_{F3}.

Assume that f obeys a one-dimensional Gaussian distribution $N\left(\mu, \sigma^2\right)$. Let g denote the sample set of f, $\overline{\mu}$ denote the sample mean of g, S^2 denote the sample variance of g, and n_1 denote the sample number of g. Thus, $\overline{\mu}$ satisfies [35]:

$$\overline{\mu} \sim N\left(\mu, \frac{\sigma^2}{n_1}\right) \tag{16}$$

S^2 satisfies [35]:

$$\frac{(n_1 - 1)S^2}{\sigma^2} \sim \chi^2(n_1 - 1) \tag{17}$$

Let $f = p_{F1}$ and $g = q_{F1}$, then the variances of p_{F1} and q_{F1} are substituted into Equation (17). We can obtain:

$$\frac{(n_1 - 1)w_{F1}{}^T \Sigma_y w_{F1}}{w_{F1}{}^T \Sigma_x w_{F1}} \sim \chi^2(n_1 - 1) \tag{18}$$

Because $w_{F1}^T \Sigma_x w_{F1} = c_{F1}$, we can obtain:

$$\frac{(n_1 - 1)w_{F1}{}^T \Sigma_y w_{F1}}{c_{F1}} \sim \chi^2(n_1 - 1) \tag{19}$$

Comparing Equation (14) with Equation (19), we can obtain:

$$(n_1 - 1)F_1(w) \sim \chi^2(n_1 - 1) \tag{20}$$

Therefore, the subfunction $F_1(w)$ multiplied by a constant $n_1 - 1$ obeys a chi-square distribution with $n_1 - 1$ degrees of freedom when there is no fault in Y.

Let $f = p_{F3}$ and $g = q_{F3}$, then the mean and variance of q_{F3} and the mean of p_{F3} are substituted into Equation (16). We can obtain:

$$w_{F3}^T \mu_y \sim N\left(w_{F3}^T \mu_x, \frac{w_{F3}^T \Sigma_x w_{F3}}{n_1} \right) \tag{21}$$

Because μ_x, Σ_x, and w_{F3} are all known, we can suppose $w_{F3}^T \mu_x = c_3$, where c_3 is a constant. According to the property of the one-dimensional Gaussian distribution, $w_{F3}^T \mu_y - c_3$ still obeys the one-dimensional Gaussian distribution, as shown in Equation (22):

$$w_{F3}^T \mu_y - w_{F3}^T \mu_x = w_{F3}^T \mu_y - c_3 \sim N\left(0, \frac{w_{F3}^T \Sigma_x w_{F3}}{n_1} \right) \tag{22}$$

Since $\Delta \mu = \mu_y - \mu_x$ and $w_{F3}^T \Sigma_x w_{F3} = c_{F3}$, we can obtain:

$$w_{F3}^T \Delta \mu_k \sim N\left(0, \frac{c_{F3}}{n_1} \right) \tag{23}$$

Normalize $w_{F3}^T \Delta \mu_k$ and we can obtain:

$$\sqrt{\frac{n_1}{c_{F3}}} w_{F3}^T \Delta \mu_k \sim N(0, 1) \tag{24}$$

Furthermore, we can obtain Equation (25) from the relationship between the standard normal distribution and the chi-square distribution:

$$\frac{n_1 w_{F3}^T \Delta \mu_k \Delta \mu_k^T w_{F3}}{c_{F3}} \sim \chi^2(1) \tag{25}$$

Comparing Equation (15) with Equation (25), we can obtain:

$$n_1 F_3(w) \sim \chi^2(1) \tag{26}$$

Therefore, the subfunction $F_3(w)$ multiplied by a constant n_1 obeys a chi-square distribution with one degree of freedom when there is no fault in Y.

In summary, $(n_1 - 1) F_1(w)$ and $(n_1) F_3(w)$ obey chi-square distributions with $n_1 - 1$ and one degree of freedom, respectively. Thus, the chi-square test is applicable to verify whether a fault occurs in Y. Given a significance level α, the fault detection thresholds of $(n_1 - 1) F_1(w)$ and $n_1 F_3(w)$ are obtainable from the chi-square test. Denote the fault detection thresholds of $(n_1 - 1) F_1(w)$ and $n_1 F_3(w)$ as ε_{F1} and ε_{F3}, respectively. In this case, two fault detection models are established as follows:

$$\begin{cases} H_0 : F_1(w) \leq \frac{\varepsilon_{F1}}{n_1 - 1}, fault - free \\ H_1 : F_1(w) > \frac{\varepsilon_{F1}}{n_1 - 1}, faulty \end{cases} \tag{27}$$

$$\begin{cases} H_0 : F_3(w) \leq \frac{\varepsilon_{F3}}{n_1}, fault - free \\ H_1 : F_3(w) > \frac{\varepsilon_{F3}}{n_1}, faulty \end{cases} \tag{28}$$

The reason for selecting the subfunctions $F_1(w)$ or $F_3(w)$ is the coverage of detectable faults. It can be seen from Equation (4) that $h(w)$ is a function of w, Σ_x, Σ_y, and $\Delta \mu$. Because the normal historical data X and the PV w are determined, both w and Σ_x are known, whereas $\Delta \mu$ and $\Delta \Sigma$, which are related to the online data, are unknown. Thus, $h(w)$ is a function of $\Delta \mu$ and $\Delta \Sigma$.

Due to noise, both $\Delta \mu$ and $\Delta \Sigma$ fluctuate within their normal ranges. However, $\Delta \mu$ or $\Delta \Sigma$ are outside of the acceptable range when the online data is faulty. Because $h(w)$ is a

function of $\Delta\mu$ and $\Delta\Sigma$, the abnormal change in $\Delta\mu$ or $\Delta\Sigma$ will further position $h(w)$ outside of the acceptable range. Therefore, the abnormal change in $\Delta\mu$ or $\Delta\Sigma$ can be detected by $h(w)$. It can be seen from Equation (8) and $\Delta\Sigma = \Sigma_y - \Sigma_x$ that $F_1(w)$ is a function of $\Delta\Sigma$; thus, the fault caused by the abnormal change in $\Delta\Sigma$ can be detected by $F_1(w)$. Similarly, the fault caused by the abnormal change in $\Delta\mu$ can be detected by $F_3(w)$ from Equation (10). Therefore, the combination of $F_1(w)$ and $F_3(w)$ can cover the majority of faults that can be detected by $h(w)$.

Why are the other two subfunctions $F_2(w)$ and $F_4(w)$ not chosen to detect faults? Comparing Equation (8) with Equation (10), $F_1(w)$ and $F_2(w)$ are reciprocal to each other. Therefore, we can detect the abnormal change in $\Delta\Sigma$ by taking either of them. The expressions of $F_3(w)$ and $F_4(w)$ differ only in the denominator. After experimental verification, the fault detection ability of $F_4(w)$ is similar to that of $F_3(w)$. Thus, only one of $F_3(w)$ and $F_4(w)$ needs to be selected to detect the abnormal change in $\Delta\mu$

3.3. Overall Fault Detection Process

We intend to use sliding windows to extract and monitor the online data in real time. Let the online data extracted by the *kth* sliding window be Y_k. The pseudocode and the flow chart of the proposed method are shown as follows:

1. Z-score normalization is performed for each parameter of the normal historical data X, and $\overline{X}$ is obtained.
2. The online data Y_k are extracted by a sliding window with the length of n_1.
3. The on-line data Y_k are normalized by Z-score to obtain $\overline{Y}_k$.
4. Two optimum PVs w_{F1} and w_{F3} between $\overline{X}$ and $\overline{Y}_k$ are obtained by using the property of the GRQ, as stated in Section 3.1.
5. Two fault detection thresholds ε_{F1} and ε_{F3} are set by using the chi-square test with a significance level α.
6. Equations (12) and (13) are used to calculate the actual values $F_1(w)$ and $F_3(w)$ of $\overline{X}$ and $\overline{Y}_k$.
7. The potential existence of a fault in Y_k is tested according to Equations (27) and (28). If at least one of two fault detection models detect fault, the online data Y_k can be considered to be faulty. Otherwise, Y_k is normal. Let $k = k + 1$; the online data of the next sliding window Y_k is tested from steps 2 to 7.

As can be seen from Figure 1, for each sliding window Y_k, we can use the property of the GRQ to obtain the optimum PVs w_{F1} and w_{F3} between X and Y_k. Because the online data Y_k may vary from different windows, w_{F1} and w_{F3} may not be the same for each window; that is, the optimum PVs adjust the online data in real time, which makes the proposed method more adaptable to potential faults.

We suppose that the system model includes n monitored variables and the length of sliding windows is n_1. The computation cost of Z-score normalization for Y_k is $O(nn_1)$. The computation cost of obtaining the mean vector and the covariance matrix of $\overline{Y}_k$ is $O(n_1)$ and $O(n^2 n_1)$, respectively. The computation cost of obtaining the inverse matrix of Σ_x is $O(n^3)$. The computation cost of obtaining $\Sigma_x^{-1}\Sigma_y$ is $O(n^3)$. Similarly, the computation cost of obtaining $\Sigma_x^{-1}\Delta\mu\Delta\mu^T$ is $O(n^3)$. The computation cost of obtaining both the maximum eigenvalue and the eigenvector of $\Sigma_x^{-1}\Sigma_y$ and $\Sigma_x^{-1}\Delta\mu\Delta\mu^T$ is $O(n^3)$. Combining all the computation cost parts above, we can get the overall computation cost of obtaining two optimum projection vectors for each window as $O(n^3)$.
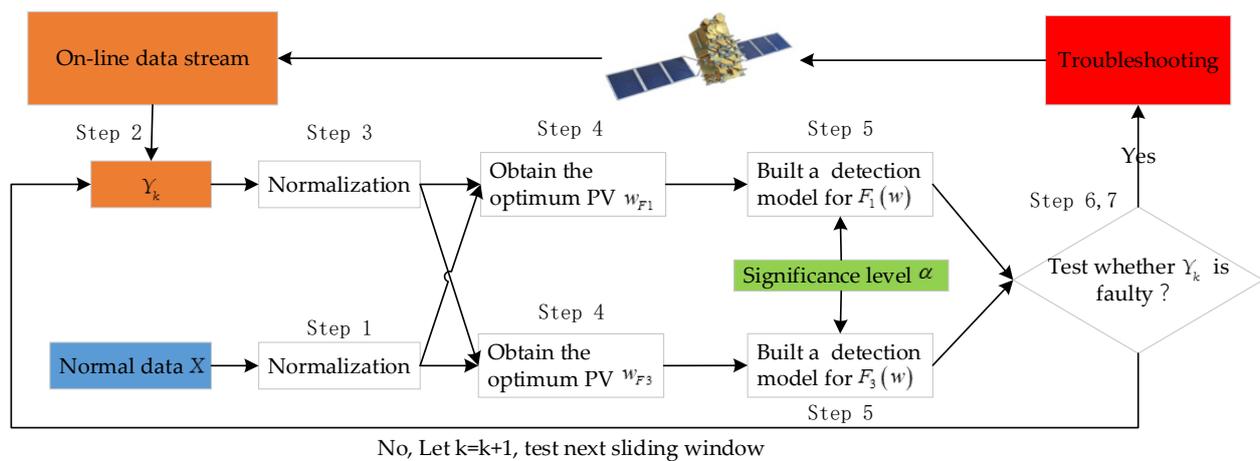
Figure 1. The flow chart of the proposed method.

4. Results and Analysis

In this section, we use a numerical case and a real satellite fault case to assess the effectiveness of the proposed method.

4.1. Numerical Case

In this subsection, a numerical simulation case, which includes three incipient faults, is provided to verify the correctness and effectiveness of the proposed method. The system model is as shown in Equation (29):

$$
\begin{aligned}
x_1 &= s_1 + s_2 + f_1 + e_1 \\
x_2 &= s_1 - s_5 + e_2 \\
x_3 &= (1 + f_3)(s_2 - s_3) + e_3 \\
x_4 &= s_1 - (1 + f_2)s_4 + e_4 \\
x_5 &= s_1 + s_3 + (1 + f_2)s_4 + e_5
\end{aligned}
\tag{29}
$$

In Equation (29), $[x_1, x_2, x_3, x_4, x_5]^T$ are five monitored variables, $[s_1, s_2, s_3, s_4, s_5]^T$ are five signal sources, $[e_1, e_2, e_3, e_4, e_5]^T$ are five noise sources, and $[f_1, f_2, f_3]^T$ are three incipient fault sources. All the signal sources and the noise sources are independent of each other and obey the standard normal distribution $N(0, 1)$.

The experimental parameters of the numerical case were set as follows. The number of each of normal historical samples and online samples was 60,000. The values of the fault sources before and after injecting faults were $[0, 0, 0]^T$ and $[0.09, 0.20, 0.09]^T$, respectively. All the incipient faults were injected at the moment of 30,001 and did not occur simultaneously. The fault types of f_1, f_2, and f_3 were offset fault, gain fault, and gain fault, respectively. Both the length and interval of sliding windows were 300 for all data in the experiment. A total of 200 windows were obtained from the online data after using sliding windows. The first 100 of the 200 windows were normal windows, whereas the last 100 were fault windows. The default signal-to-noise ratio (SNR) was set as 20 dB [30]. The simulation hardware platform was a desktop computer (CPU: Intel core i5 − 10400, RAM: DDR4/2666/16G) and the software was MATLAB 2019b.

The compared fault-detection methods included using PCA and the T^2 statistic [36] (PCA + T^2), PCA and the squared prediction error statistic [36] (PCA + SPE), PCA and the KL divergence [23] (PCA + KLD), and the method based on the local optimum PV and the KL divergence [30] (LOPVKLD). Because of the poor effect of directly monitoring the original variables, the methods of PCA + T^2 and PCA + SPE in this experiment monitored the means and variances of the original variables. The principal subspace was selected

with a cumulative variance contribution of more than 90%. The confidence levels for the PCA + T^2 method and the PCA + SPE method were both set at 0.95. The significance levels for the PCA + KLD method and the LOPVKLD method were 0.05 and 0.01, respectively. The significance levels of the subfunctions $F_1(w)$ and $F_3(w)$ proposed in this paper were 0.0005 and 0.001, respectively. Three evaluation indexes—fault detection rate (FDR), false alarm rate (FAR), and the time consumption of finding the optimum PV for each window (time consumption)—were chosen as the indexes for evaluating the fault detection results. For the purpose of conciseness, only the fault detection result of the PCA + KLD method of the principal component that was most sensitive to the fault is presented, whereas the other, relatively poor results are not displayed.

The detection results of five fault-detection methods for the incipient fault f_1 are shown in Figure 2. As can be seen from Figure 2, both the PCA + T^2 method and the PCA + SPE method failed to detect f_1 because most of the fault windows were still within the detection threshold. Conversely, both the PCA + KLD method and the LOPVKLD method successfully detected f_1. As stated in Section 3.2, the subfunctions $F_1(w)$ and $F_3(w)$ can detect the fault that causes the abnormal change in $\Delta\Sigma$ and $\Delta\mu$, respectively. Because f_1 is the offset fault that can cause the abnormal change in $\Delta\mu$, the fault f_1 can be successfully detected by the subfunction $F_3(w)$ rather than the subfunction $F_1(w)$.
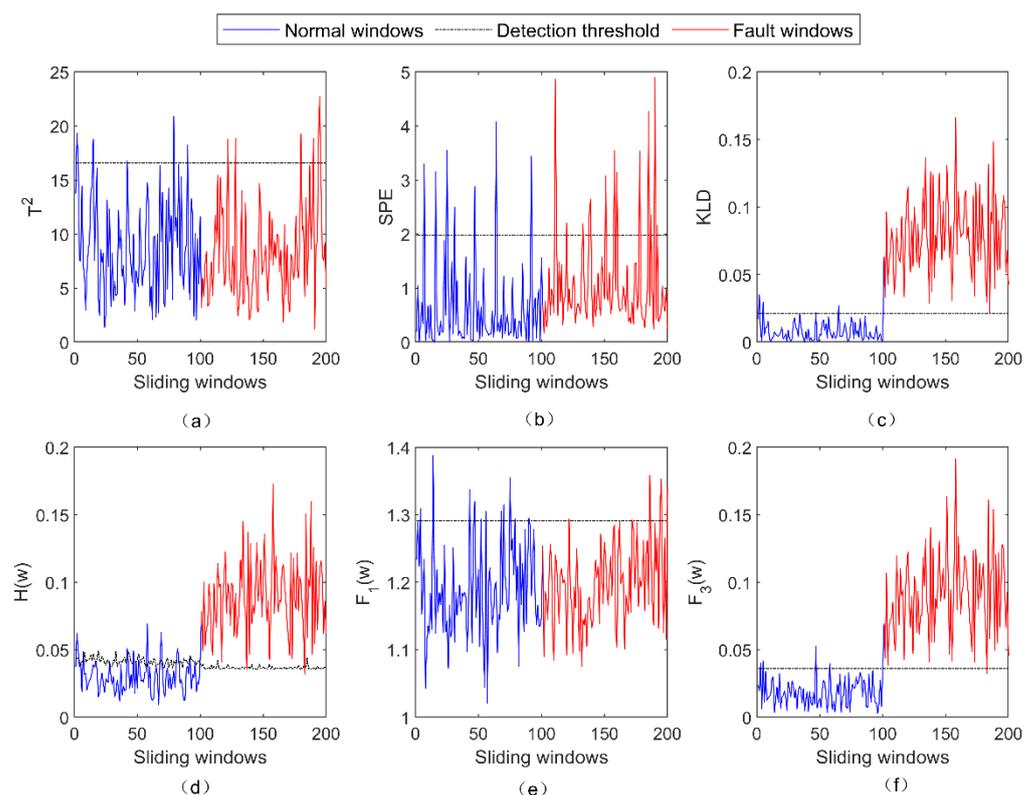

Figure 2. The detection results of five fault-detection methods for the fault f_1. (**a**) The result of PCA + T^2 for f_1; (**b**) the result of PCA + SPE for f_1; (**c**) The result of PCA + KLD for f_1; (**d**) the result of LOPVKLD for f_1; (**e**) the result of $F_1(w)$ for f_1; (**f**) the result of $F_3(w)$ for f_1.

The detection results of five fault-detection methods for the incipient fault f_2 are presented in Figure 3. As shown, the PCA + SPE method still fails to detect f_2. Both the PCA + T^2 method and the PCA + KLD method have relatively poor detection results for f_2. Due to the application of the local optimum PV, the LOPVKLD method has a better detection result for f_2. Because f_2 is the gain fault which can cause the abnormal change in $\Delta\Sigma$, f_2 can be successfully detected by the subfunction $F_1(w)$ rather than the subfunction $F_3(w)$.
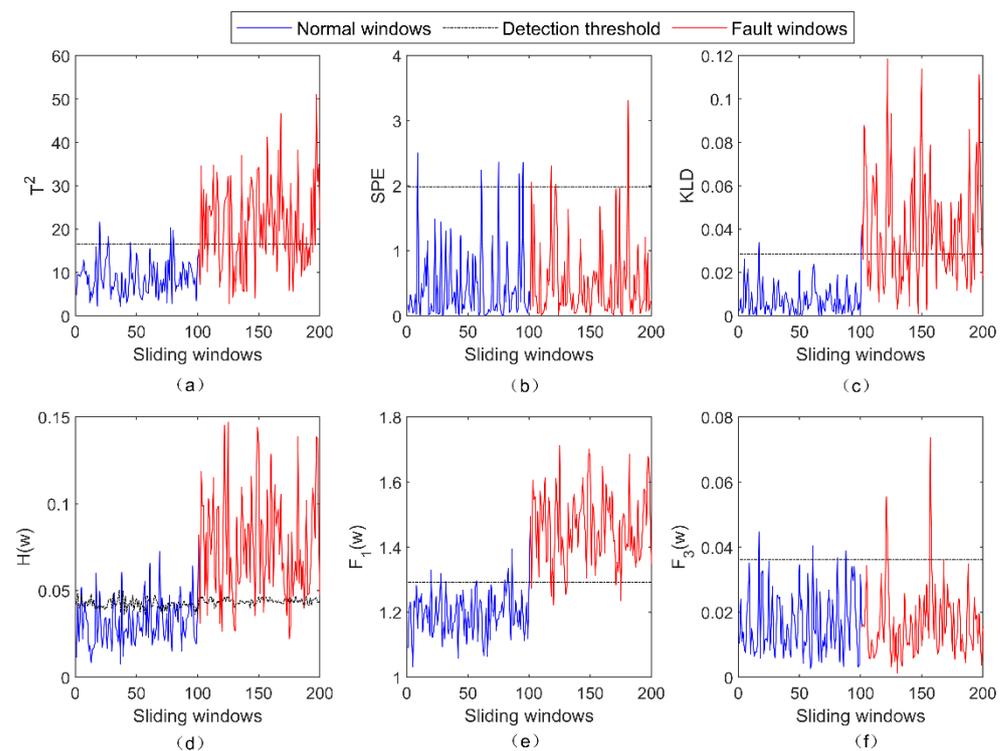
Figure 3. The detection results of five fault-detection methods for the fault f_2. (**a**) The result of PCA + T^2 for f_2; (**b**) the result of PCA + SPE for f_2; (**c**) the result of PCA + KLD for f_2; (**d**) the result of LOPVKLD for f_2; (**e**) the result of $F_1(w)$ for f_2; (**f**) the result of $F_3(w)$ for f_2.

As can be seen from Figure 4, three fault-detection methods—PCA + T^2, PCA + SPE and PCA + KLD—are ineffective in detecting the fault f_3, because most of the result values of these methods are still under the detection threshold. It can be seen from Figure 3d,f that the LOPVKLD method and the subfunction $F_1(w)$ are effective at detecting f_3. As f_3 is the gain fault, the subfunction $F_3(w)$ fails to detect f_3.

Considering the randomness of the signal sources and the noise sources in the numerical case, we simulated the three incipient faults 100 times and then derived the average of the fault detection results, as presented in Table 1.

It can be seen from Reference [30] that the PCA + T^2 and the PCA + SPE methods are ineffective in detecting incipient faults when the original variables are monitored. As can be seen from Table 1, the fault detection rates of these two methods increase, particularly the fault detection rate for f_2. The reason for the improvement in these two methods is that the extraction of the means and variances of the variables can be considered as smoothing the variables. Although the means and variances of the variables are monitored, the detection results of these two methods are inferior to those of the PCA + KLD method. Due to the usage of constant PVs, the PCA + KLD method is effective at detecting f_1 and f_2, but has poor detection results for f_3.

Because of the application of the local optimum PV, the LOPVKLD method is sensitive to all three incipient faults. However, as stated in Section 2.2, the LOPVKLD method has the disadvantage of high computation complexity. As can be seen from Table 1, the LOPVKLD method requires a long duration (about 70 ms) to obtain the optimum PV. By contrast, the duration to obtain the optimum PV for each subfunction is less than 25 μs, three orders of magnitude faster than the LOPVKLD method. Because finding the optimum PV is not required, the PCA + T^2, PCA + SPE, and PCA + KLD methods have lower computation complexity than the proposed method. However, the detection results of these methods are not as good as those of the proposed method, particularly the detection result for f_3. Because the subfunctions $F_1(w)$ and $F_3(w)$ can detect the faults caused by the abnormal change in

$\Delta\Sigma$ and $\Delta\mu$, respectively, the three faults can be successfully detected by $F_3(w)$, $F_1(w)$, and $F_1(w)$, respectively.
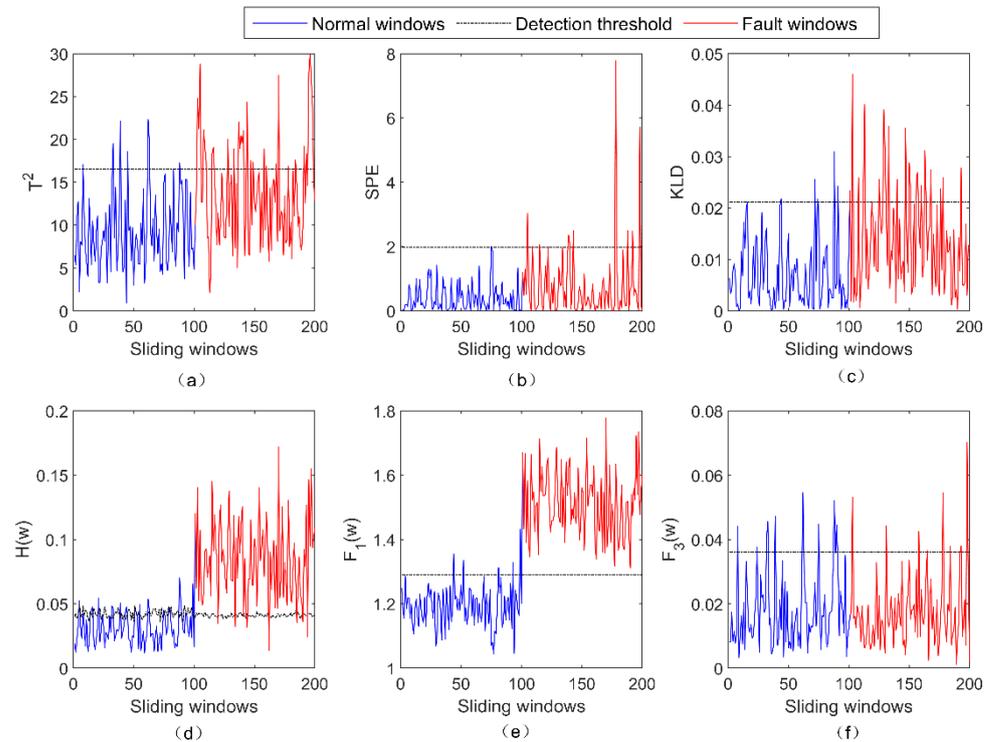

Figure 4. The detection results of five fault-detection methods for the fault f_3. (**a**) The result of PCA + T^2 for f_3; (**b**) the result of PCA + SPE for f_3; (**c**) the result of PCA + KLD for f_3; (**d**) the result of LOPVKLD for f_3; (**e**) the result of $F_1(w)$ for f_3; (**f**) the result of $F_3(w)$ for f_3.

Table 1. Comparison of fault detection performance for the three incipient faults.

Faults	Evaluation Indexes	PCA + T^2	PCA + SPE	PCA + KLD	LOPVKLD	Proposed Method	
						$F_1(w)$	$F_3(w)$
f_1	FDR (%)	5.76	17.02	97.41	94.63	7.41	96.67
	FAR (%)	4.49	7.67	11.90	15.76	8.5	5.56
	Time consumption	0 (μs)	0 (μs)	0 (μs)	68.5 (ms)	18.42 (μs)	24.26 (μs)
f_2	FDR (%)	58.46	25.96	79.36	89.08	95.99	8.41
	FAR (%)	4.41	8.05	11.58	14.84	7.17	5.80
	Time consumption	0 (μs)	0 (μs)	0 (μs)	70.8 (ms)	18.20 (μs)	23.75 (μs)
f_3	FDR (%)	27.56	20.87	30.37	90.91	97.81	7.25
	FAR (%)	4.61	7.68	11.50	15.82	7.53	5.88
	Time consumption	0 (μs)	0 (μs)	0 (μs)	71.7 (ms)	18.31 (μs)	23.99 (μs)

The reason for the sensitivity of the proposed method to incipient faults, from the perspective of optimum PV, is explained in this paper. The projection process can be regarded as a weighted sum process, as presented in Equation (30):

$$w^T X = w_1 x_1 + w_2 x_2 + \cdots + w_5 x_5 \tag{30}$$

In Equation (30), w is an optimum PV and can be considered to be a weight coefficient vector and X is the vector which includes five monitored variables. For the purpose of presentation, all the optimum PVs in the numerical case were normalized (the moduli of the vectors were set to 1) and the absolute value was taken. The optimum PVs obtained using the LOPVKLD method, the subfunction $F_1(w)$, and the subfunction $F_3(w)$ before

and after insertion of the faults f_1 and f_3 are shown in Figure 5a–f, respectively. In each subfigure of Figure 5, the first 100 windows were the normal windows, whereas the last 100 windows were the fault windows.
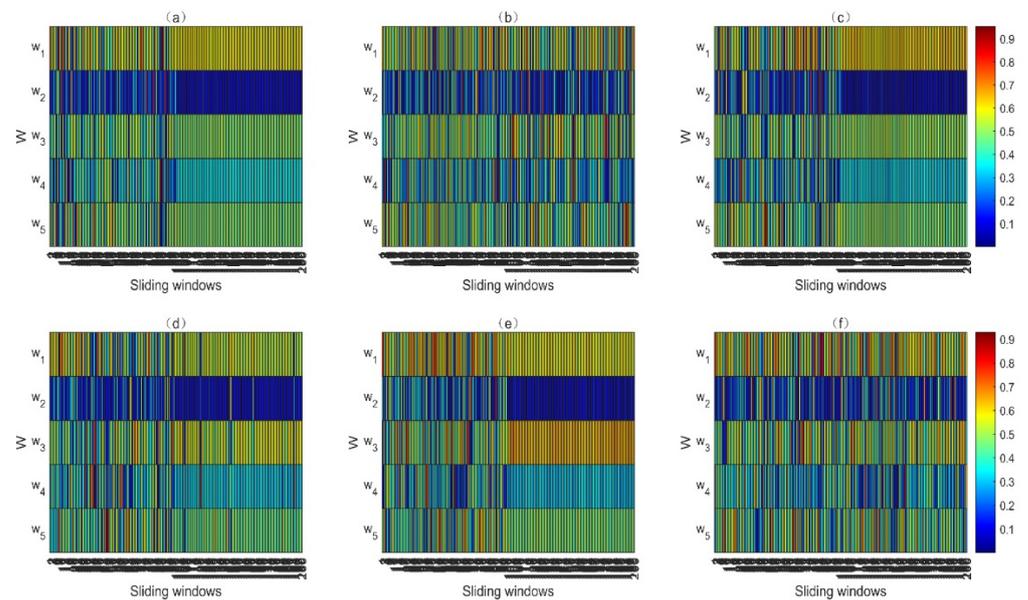

Figure 5. Comparison of the optimum PVs for different faults. (**a**) The optimum PVs of LOPVKLD for f_1; (**b**) the optimum PVs of $F_1(w)$ for f_1; (**c**) the optimum PVs of $F_3(w)$ for f_1; (**d**) the optimum PVs of LOPVKLD for f_3; (**e**) the optimum PVs of $F_1(w)$ for f_3; (**f**) the optimum PVs of $F_3(w)$ for f_3.

Due to the enlargement of the faulty variables, the fault is easier to expose and the detection ability is improved. It can be seen from Equation (29) that the fault f_1 was added to the variable x_1. As can be seen from Figure 5a–c, both the LOPVKLD method and the subfunction $F_3(w)$ enlarged the weight of faulty variable x_1 after the fault f_1 occurred. As shown in Figure 5d–f, because the fault variable of the fault f_3 is x_3, both the LOPVKLD method and the subfunction $F_1(w)$ enlarged the weight of faulty variable x_3 after the fault f_3 occurred. In addition, because iteration is not needed, the computation complexity of the proposed method is less than that of the LOPVKLD method. In summary, the proposed method not only retains the advantage of being more sensitive to possible incipient faults, but also alleviates the disadvantage of high computational complexity.

4.2. Real Satellite Fault Case

On 16 March 2021, key telemetry parameters of a satellite payload abnormally fluctuated. Figure 6 presents the phenomena of a telemetry parameter fluctuation related to the fault. In this case, the development of the fault experienced three stages. In the first stage, the variance of the telemetry parameter increased slightly and lasted around 50 days. With the further deterioration of the fault, the mean and variance of the telemetry parameter significantly fluctuated in the second stage. The fault lasted around 70 days in this stage. As the fault developed to the third stage, the mean and variance of the telemetry parameter seriously deviated from the normal fluctuation range. Because the current fault detection system adopts the method based on a threshold, the system cannot detect the fault until it develops to the third stage. If the fault was successfully detected at the beginning of the first stage, it could be found about four months earlier. Thus, the research objective of this paper is to detect the incipient fault from the first stage.
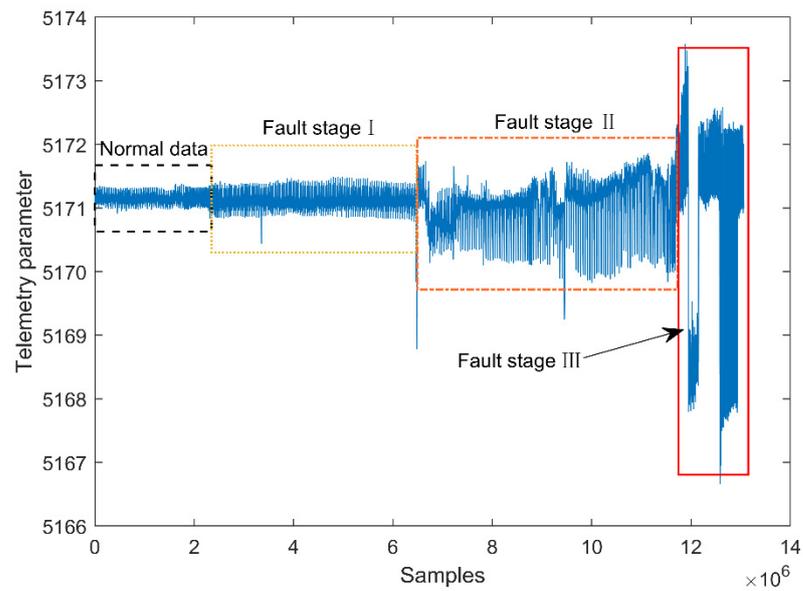
Figure 6. The phenomena of the fault parameter fluctuation.

In this study, a total of 13,066,123 samples were collected and arranged from the satellite measurement and control system from 7:35:34 on 15 November 2020 to 16:27:52 on 16 May 2021. Two telemetry parameters related to the fault were selected, as presented in Figure 7. For the reason of confidentiality, the true telemetry parameter names are hidden. The sampling rate of the telemetry data in Figure 7 was 1 Hz. Due to the constraints of the satellite's visible arc and the ground station measurement and control resources, some telemetry data were not transmitted; that is, the telemetry data were discontinuous in time.
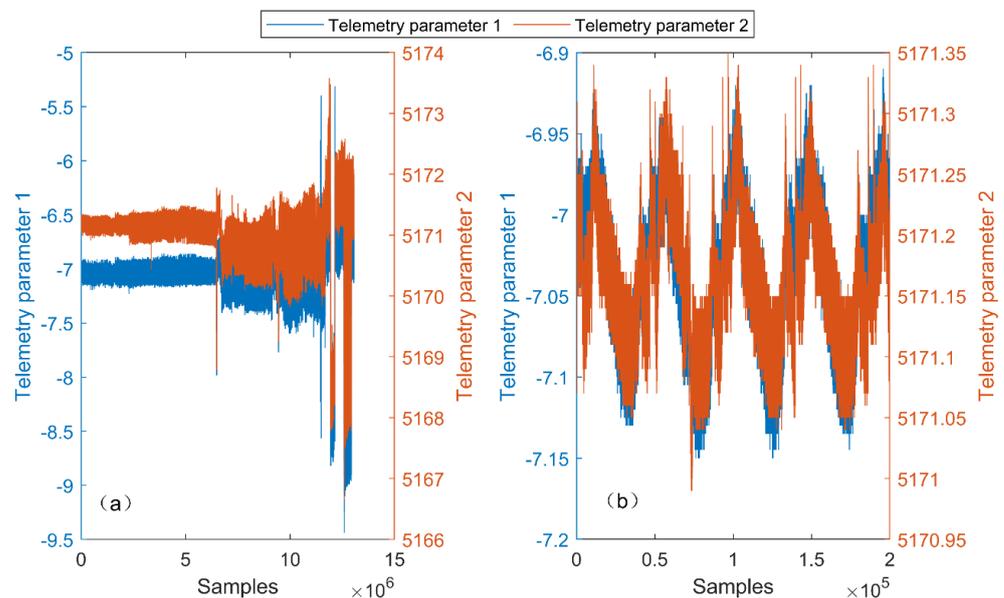

Figure 7. The phenomena of the selected fault parameters. (**a**) All the data of the parameters; (**b**) the periodic phenomenon of the parameters.

As indicated in Figure 7b, the parameters show a periodicity, and the period is consistent with the satellite orbital period (46, 468 s). For this reason, in this study, we took the satellite orbital period as the length of the sliding window, set the interval of the sliding window as 10,000, and retained the sliding windows comprising more than 40,000 samples as effective windows. A total of 524 effective windows were obtained

from the first 6,246,451 samples after being extracted by sliding windows. The samples of the first 100 effective windows were selected as the normal historical data. The last 424 effective windows were selected as the online data for testing. Among the 424 windows for testing, the first 72 windows were normal windows, whereas the last 354 windows were fault windows.

Furthermore, it can be seen from Figure 7b that the telemetry parameters do not obey Gaussian distributions; thus, the fault detection threshold set by the chi-square test may not be appropriate, and the normal historical data must be used to assist in setting the threshold. As stated in Section 3.2, the subfunction $F_1(w)$ multiplied by the constant $n_1 - 1$ obeys a chi-square distribution with $n_1 - 1$ degrees of freedom. In this case, the length of the sliding window n_1 was 46,468. The degrees of freedom were sufficiently high that the subfunction $F_1(w)$ could be considered to obey a normal distribution; that is, the 3σ method could be used in this case to test whether there is a fault in $F_1(w)$.

Let X be the normal historical data, which include the date of 100 normal windows. We assume that the ith normal window data is X_i. We set X_i as the online data and then use the property of the GRQ to obtain the optimum PV w_{F1_i} between X and X_i. Let $Y = X_i$, $w = w_{F1\_i}$; we can obtain the value of $F_{1\_i}(w)$ from Equation (8). Furthermore, we can obtain a vector $F_{1\_X}(w)$ from 100 normal windows. The process of obtaining the vector $F_{1\_X}(w)$ is shown in Figure 8.
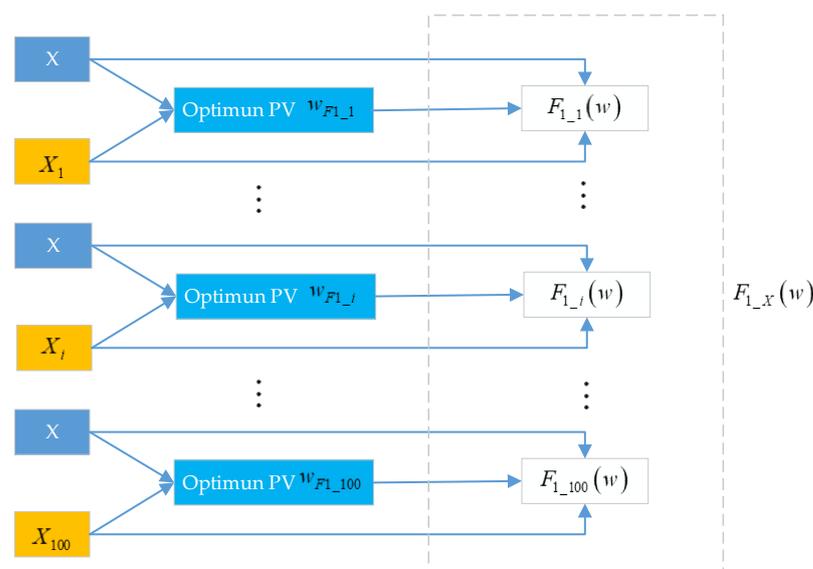

Figure 8. The process of obtaining vector $F_{1\_X}(w)$.

Let M_1 and S_1 be the mean and the standard deviation of the vector $F_{1\_X}(w)$, respectively. The fault-detection method of the subfunction $F_1(w)$ is presented as follows:

$$\begin{cases} H_0 : M_1 - 3S_1 \leq F_1(w) \leq M_1 + 3S_1, fault - free \\ H_1 : M_1 - 3S_1 > F_1(w) | F_1(w) > M_1 + 3S_1, faulty \end{cases} \tag{31}$$

It can be seen from Section 3.2 that the subfunction $F_3(w)$ multiplied by the constant n_1 obeys a chi-square distribution with one degree of freedom. Therefore, we refer to the method in Reference [30] to set the threshold. Let $F_{3\_X}(w)$ be the set of 100 $F_3(w)$ values of 100 normal windows. The process of obtaining the vector $F_{3\_X}(w)$ is similar to that of the vector $F_{1\_X}(w)$. The difference between these two processes is that we use the property of the GRQ to obtain the optimum PV w_{F3_i} and then obtain the value of $F_{3\_i}(w)$ from

Equation (10). Let M_3 be the mean of the vector $F_{3\_X}(w)$. The fault-detection method of the subfunction $F_3(w)$ is presented as follows:

$$\begin{cases} H_0 : F_3(w) \leq M_3\chi_\alpha^2(1), fault - free \\ \quad H_1 : F_3(w) > M_3\chi_\alpha^2(1), faulty \end{cases} \tag{32}$$

where $\chi_\alpha^2(1)$ is the threshold of the chi-square distribution with one degree of freedom with a given significance level α.

In this real satellite fault case, both the PCA + T^2 and the PCA + SPE methods still monitored the means and variances of the telemetry parameters. The experimental parameters of these two methods were the same as those presented in Section 4.1. The significance levels of the PCA + KLD method were set to 0.05 and 0.01, respectively. The significance levels of the LOPVKLD method were set to 0.05 and 0.01, respectively. The threshold of $F_1(w)$ was set by the 3δ method, and the significance level of $F_3(w)$ was 0.01. The detection results and evaluation indexes of these five methods for the real satellite fault are shown in Figure 9 and Table 2, respectively.
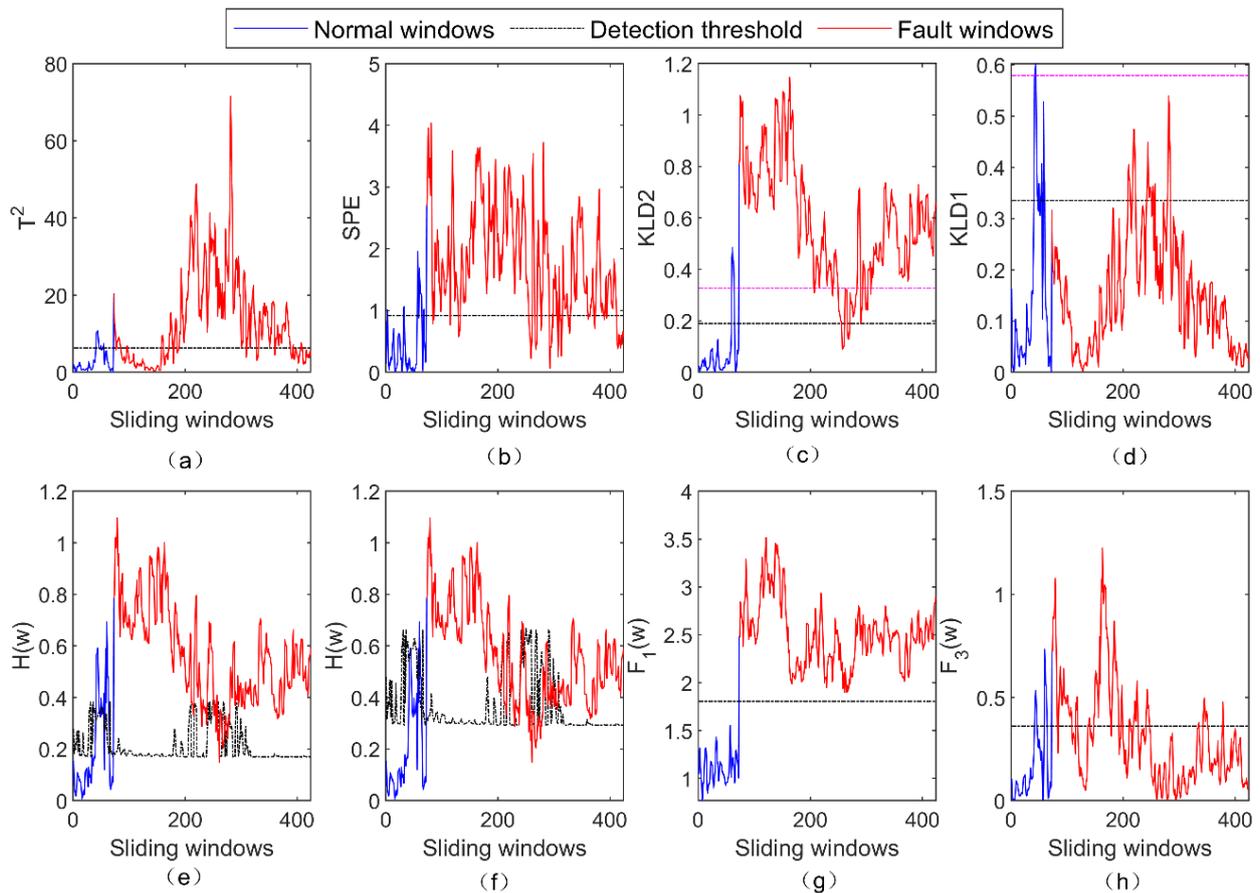

Figure 9. The detection result of five methods for the real satellite fault. (**a**) The result of PCA + T^2 for the fault; (**b**) the result of PCA + SPE for the fault; (**c**) the result of second principal component of PCA + KLD for the fault; (**d**) the result of first principal component of PCA + KLD for the fault; (**e**) the result of LOPVKLD with the significance level of 0.05; (**f**) the result of LOPVKLD with the significance level of 0.01; (**g**) the result of $F_1(w)$ for the fault; (**h**) the result of $F_3(w)$ for the fault.

It can be seen from Figure 9a that the PCA + T^2 method has a poor detection result for the real satellite fault, particularly the fault windows between Nos. 100 and 200. Compared to Figure 9a, the detection result of the PCA + SPE method in Figure 9b is significantly improved. However, some fault windows around Nos. 250 to 300 are below the fault detection threshold. Figure 9c,d presents the detection results of the two principal

components of the PCA + KLD method for the real satellite fault. In Figure 9c,d, the detection thresholds of significance levels of 0.05 and 0.01 are represented by the black dashed line and the magenta dashed line, respectively. Figure 8e,f illustrates the fault detection results of the LOPVKLD method with the significance levels of 0.05 and 0.01, respectively. According to Figure 9c,f, the fault detection rates of the PCA + KLD and the LOPVKLD methods are higher than 95% with the significance level of 0.05. However, the false alarm rates of both these methods are higher than 25% at this significance level. At significance levels of 0.01, the false alarm rates of these two methods are around 12%, but the fault detection rates decrease by around 10%. As a comparison, the fault detection and false alarm rates of the subfunction $F_1(w)$ are 100% and 0%, respectively. The false alarm of the proposed method comes from the subfunction $F_3(w)$. It can be seen from Figure 9 and Table 2 that the false alarm rate of the proposed method is 13.89%. The effectiveness and superiority of the proposed method is further verified by the real satellite case.

Table 2. The evaluation indexes of five fault methods for the real satellite fault.

Evaluation Indexes	PCA + T^2	PCA + SPE	PCA + KLD		LOPVKLD		Proposed Method	
			$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$F_1(w)$	$F_3(w)$
FDR (%)	63.46	83.85	97.16	85.65	95.17	85.51	100	32.95
FAR (%)	14.08	16.9	25	11.11	26.39	12.50	0	13.89

5. Conclusions

In this paper, we propose a new and fast method to detect incipient faults of satellites. We decompose the KL divergence and use the property of the generalized Rayleigh quotient to obtain the optimum projection vector. Under the assumption that the variables obey a multidimensional Gaussian distribution, the distributions of the subfunctions $F_1(w)$ and $F_3(w)$ are presented and verified. To address non-Gaussian satellite telemetry parameters, we use the normal historical data to assist in setting the threshold. The proposed method is a linear method. Future work may focus on developing a nonlinear fault-detection method.

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

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

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Description
KL	Kullback–Leibler
PCA	principal component analysis
PV	projection vector
GRQ	generalized Rayleigh quotient
FDR	fault detection rate
FAR	false alarm rate

References

1. Yang, Y.; Mao, Y.; Sun, B. Basic performance and future developments of BeiDou global navigation satellite system. *Satell. Navig.* **2020**, *1*, 1. [CrossRef]
2. Wang, J.; Chen, H.; Zhu, Z. Modeling research of satellite-to-ground quantum key distribution constellations. *Acta Astronaut.* **2021**, *180*, 470–481. [CrossRef]
3. Chen, H.; Yong, B.; Shen, Y.; Liu, J.; Hong, Y.; Zhang, J. Comparison analysis of six purely satellite-derived global precipitation estimates. *J. Hydrol.* **2020**, *581*, 124376. [CrossRef]
4. Burke, M.; Driscoll, A.; Lobell, D.B.; Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **2021**, *371*, 6535. [CrossRef] [PubMed]
5. Ezhilarasu, C.M.; Skaf, Z.; Jennions, I.K. The application of reasoning to aerospace Integrated Vehicle Health Management (IVHM): Challenges and opportunities. *Prog. Aeronaut. Sci.* **2019**, *105*, 60–73. [CrossRef]
6. Tafazoli, M. A study of on-orbit spacecraft failures. *Acta Astronaut.* **2009**, *64*, 195–205. [CrossRef]
7. Li, E.-H.; Li, Y.-Z.; Li, T.-T.; Li, J.-X.; Zhai, Z.-Z.; Li, T. Intelligent analysis algorithm for satellite health under time-varying and extremely high thermal loads. *Entropy* **2019**, *21*, 983. [CrossRef]
8. Safaeipour, H.; Forouzanfar, M.; Casavola, A. A survey and classification of incipient fault diagnosis approaches. *J. Process Control* **2021**, *97*, 1–16. [CrossRef]
9. Peng, Z.; Lu, Y.; Miller, A.; Zhao, T.; Johnson, C. Formal specification and quantitative analysis of a constellation of navigation satellites. *Qual. Reliab. Eng. Int.* **2016**, *32*, 345–361. [CrossRef]
10. Cayrac, D.; Dubois, D.; Prade, H. Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. *IEEE Trans. Fuzzy Syst.* **1996**, *4*, 251–269. [CrossRef]
11. Chen, R.H.; Ng, H.K.; Speyer, J.L.; Guntur, L.S.; Carpenter, R. Health monitoring of a satellite system. *J. Guid. Control. Dynam.* **2006**, *29*, 593–605. [CrossRef]
12. 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]
13. Pang, J.; Liu, D.; Peng, Y.; Peng, X. Collective anomalies detection for sensing series of spacecraft telemetry with the fusion of probability prediction and Markov chain model. *Sensors* **2019**, *19*, 722. [CrossRef]
14. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Comput. Surv.* **2009**, *41*, 1–58. [CrossRef]
15. Verzola, I.; Donati, A.; Martinez, J.; Schubert, M.; Somodi, L. Project Sibyl: A Novelty Detection System for Human Spaceflight Operations. In Proceedings of the 14th International Conference on Space Operations, Daejeon, Korea, 16–20 May 2016; p. 2405.
16. Hayden, S.; Sweet, A.; Christa, S. Livingstone model-based diagnosis of Earth Observing One. In Proceedings of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA, 20–22 September 2004; p. 6225.
17. Deb, S.; Pattipati, K.R.; Shrestha, R. QSI's integrated diagnostics toolset. In Proceedings of the 1997 IEEE Autotestcon Proceedings AUTOTESTCON'97. IEEE Systems Readiness Technology Conference. Systems Readiness Supporting Global Needs and Awareness in the 21st Century, Anaheim, CA, USA, 22–25 September 1997; pp. 408–421.
18. Cheng, C.; Wang, J.; Chen, H.; Chen, Z.; Luo, H.; Xie, P. A review of intelligent fault diagnosis for high-speed trains: Qualitative approaches. *Entropy* **2021**, *23*, 1. [CrossRef] [PubMed]
19. Muthusamy, V.; Kumar, K.D. A novel data-driven method for fault detection and isolation of control moment gyroscopes onboard satellites. *Acta Astronaut.* **2021**, *180*, 604–621. [CrossRef]
20. Ibrahim, S.K.; Ahmed, A.; Zeidan, M.A.E.; Ziedan, I.E. Machine learning techniques for satellite fault diagnosis. *Ain. Shams. Eng. J.* **2020**, *11*, 45–56. [CrossRef]
21. Pang, J.; Liu, D.; Peng, Y.; Peng, X. Anomaly detection based on uncertainty fusion for univariate monitoring series. *Measurement* **2017**, *95*, 280–292. [CrossRef]
22. 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.
23. Chen, H.; Jiang, B.; Lu, N. An improved incipient fault detection method based on Kullback–Leibler divergence. *ISA Trans.* **2018**, *79*, 127–136. [CrossRef]
24. 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]
25. Shang, J.; Chen, M.; Ji, H.; Zhou, D. Recursive transformed component statistical analysis for incipient fault detection. *Automatica* **2017**, *80*, 313–327. [CrossRef]
26. Ji, H.; He, X.; Shang, J.; Zhou, D. Incipient fault detection with smoothing techniques in statistical process monitoring. *Control Eng. Pract.* **2017**, *62*, 11–21. [CrossRef]
27. 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]
28. Gautam, S.; Tamboli, P.K.; Patankar, V.H.; Roy, K.; Duttagupta, S.P. Sensors Incipient Fault Detection and Isolation Using Kalman Filter and Kullback–Leibler Divergence. *IEEE Trans. Nucl. Sci.* **2019**, *66*, 782–794. [CrossRef]
29. Deng, X.; Cai, P.; Cao, Y.; Wang, P. Two-step localized kernel principal component analysis based incipient fault diagnosis for nonlinear industrial processes. *Ind. Eng. Chem. Res.* **2020**, *59*, 5956–5968. [CrossRef]

30. 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. [CrossRef]
31. Hart, P.E.; Stork, D.G.; Duda, R.O. *Pattern Classification*; Wiley Hoboken: Hoboken, NJ, USA, 2000.
32. Watkins, D.S. *Fundamentals of Matrix Computations*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2002.
33. Wang, L.-F.; Xia, Y. A linear-time algorithm for globally maximizing the sum of a generalized rayleigh quotient and a quadratic form on the unit sphere. *SIAM J. Optim.* **2019**, *29*, 1844–1869. [CrossRef]
34. Jolliffe, I. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
35. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Duxbury Press: Pacific Grove, CA, USA, 2002.
36. Nassar, B.; Hussein, W.; Mokhtar, M. Space telemetry anomaly detection based on statistical PCA algorithm. In Proceedings of the International Journal of Electronics and Communication Engineering, Paris, France, 27–28 August 2015; pp. 637–645.