IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 20, 2018, accepted July 23, 2018, date of publication August 3, 2018, date of current version August 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2862940

A Comparison of Different Statistics for Detecting Multiplicative Faults in Multivariate Statistics-Based Fault Detection Approaches

KAI ZHANG[1], KAIXIANG PENG[1], (Member, IEEE), AND YURI A. W. SHARDT[2], (Member, IEEE)

[1]Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[2]Department of Automation Engineering, Institute of Automation and Systems Engineering, Technical University of Ilmenau, 98684 Ilmenau, Germany

Corresponding author: Kai Zhang (kaizhang@ustb.edu.cn)

The work of K. Zhang and K. Peng was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0306403, in part by the Natural and Science Foundation of China (NSFC) under Grant 61473033, and the China Postdoctoral Science Foundation under Grant 2017M620616. The work of Y. A. W. Shardt was supported in part by the National Science Foundation of China under Grant 61673053 and in part by the Beijing Natural Science Foundation under Grant 4162041.

ABSTRACT The explosion of different fault detection (FD) statistics in multivariate statistics-based FD approaches has meant that the practitioner is faced with the unenviable job of determining which to use in a given circumstance. Moreover, compared to extensive investigations on additive faults, the performance of commonly used FD statistics for detecting multiplicative faults has not been holistically evaluated. Therefore, this paper seeks to investigate the different statistics that can be applied to detect multiplicative faults in order to provide users and practitioners in the FD field with guidance to select an appropriate method. The considered statistics are broadly classified into three groups: traditional methods (*e.g.* T^2-statistic) and their extensions; the Wishart distribution-based methods; and those methods created in the information and communication fields to describe the measurement variance and covariance (*e.g.* Kullback-Leibler divergence). These three groups are compared by considering the required probability distributions, interconnections, and detection performance for multiplicative faults. Using simulated data from numerical examples and the Tennessee Eastman benchmark process, the theoretical results are validated, and the applicability of multivariate statistics-based FD methods incorporating all considered statistics for detecting multiplicative faults is examined at the end of this paper.

INDEX TERMS Fault detection statistics, multiplicative fault, fault detection rate, multivariate statistics-based fault detection.

I. INTRODUCTION

Since the scale and degree of automation of industrial processes have increased, it nowadays become necessary to design fault detection (FD) methods to guarantee the safety and stability of process operations. Date back to 1950s and intensively investigated since 1990s, the multivariate statistics-based FD methods have nowadays become an active area of research in the data-driven FD field due to its simplicity and high-efficiency to handle high-dimensional process data [1]–[3]. Proposed in a chemistry background [4], [5], it has currently been recognized and extended to a wider scope of industry, including polymers, microelectronics manufacturing, iron and steel, and pharmaceutical processes [2], [6]–[8]. In this field, recent work reveals that such methods are comprised of two important techniques: multivariate data analysis (MDA) methods and FD statistics [9], [11], [12]. They have established a unified framework as shown in Fig. 1 that the process data are first modeled using MDA methods, then the modeled data are sent to appropriate FD statistics to examine whether the process is operating properly [9]. Compared with directly sending data to FD statistics, multivariate statistics-based FD methods can significantly improve the FD performance and resolve the specific problem met in real industry [1], [3], [5], [9]. Note that MDA can be referred to as the residual generator, which is commonly used in model-based FD methods [9].

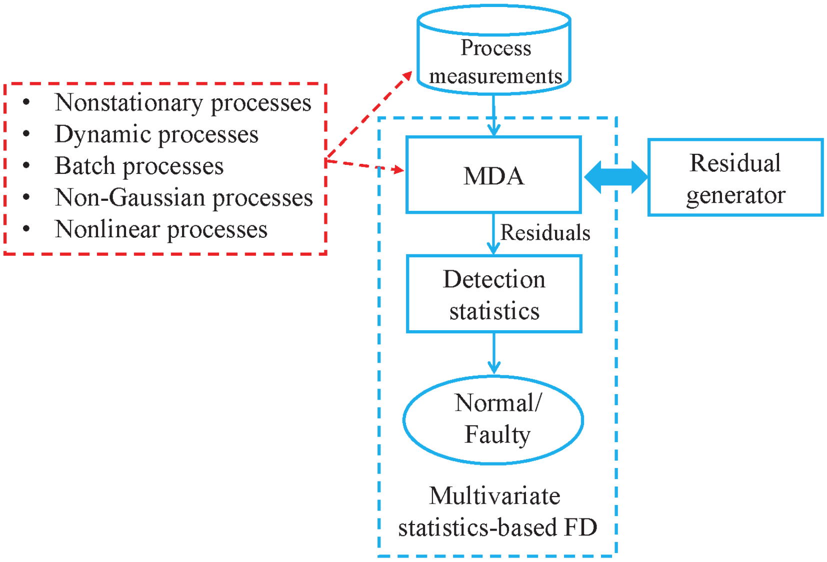

FIGURE 1. **A schematic of multivariate statistics-based FD methods.**

The most popular MDA methods are principal component analysis (PCA), partial least squares (PLS), and canonical correlation analysis (CCA) [9], [13]–[16]. The core of PCA is reducing the dimension of process data such that significant variations of the process data are kept in low-dimensional scores. Monitoring scores instead of the original data can, on one hand, save the computational resource caused by FD statistics, on the other hand, improve the detecting performance compared to directly using high-dimensional data [1]. As the foundation of chemometrics, PLS was proposed to address the numerical problem when implementing least squares for liner regression [17]. PLS-based FD methods were developed to resolve the quality-related process monitoring problem, that is, using the readily available process data to monitor the hard-to-measure or infrequent quality index data [2]. CCA was proposed to examine the correlation between two data sets, which, thus, has commonly been used to monitor the relationship between process inputs and outputs [8]–[10]. Other alterative MDA methods have also drawn much attention, such as independent component analysis and factor analysis. Moreover, as shown in Fig. 1, they can be extended to model processes with other characteristics like multibatch, dynamics and nonstationary etc. Although previous work has tended to focus on the development and modification of classic MDA methods [19], [20], it is the case that only FD statistics can directly affect the detection performance. In the statistical FD framework, the FD statistic plays the central role in identifying potential faults in industrial processes [9]. Depending on the normal process situation where the data are assumed to be normally distributed, a statistic using a specific probability distribution is chosen and appropriate thresholds, either one (upper) or two (upper and lower), can be determined based on the distribution properties. Then, faulty or normal operating performance can be observed by comparing the statistic against the thresholds. In the previous literature, different statistics have been developed for detecting additive faults. The most commonly used methods are T^2- and Q-statistics [1], [3], [5], [18], and the assumed probability functions are χ^2 or F-distributions [1], [3], [9]. Since the additive fault model was introduced to FD field, it is straightforward to evaluate the detectability of different FD statistics. When dealing with additive faults, it was found that T^2- and Q-statistics can detect undesired changes in the mean values of process data using only the time series samples. Furthermore, T^2, in terms of the fault detection rate (FDR), performs better than Q [9]. Unlike the mean value change case, the covariance change case, which can be well described using a multiplicative fault model, will change the elements in the covariance matrix [9], [12], [21], [22]. The performance of T^2 and Q for detecting the variance change was evaluated, where it was shown that they have inherent weakness to deal with this type of change [23]. The weakness can be due to the different expressions of calculating the FDR index compared to detecting additive faults. Similarly, in [12], the typical multivariate statistics-based FD methods incorporating the two statistics and typical MDA methods for detecting multiplicative faults are evaluated, in which the expected detection delay (EDD) evaluating index was taken [11] which is similar to the average run length index widely used in statistical quality control field [24]. The poor performance with large EDD for multiplicative faults was shown due to the bad detectability offered by the two statistics. Therefore, to detect well this type of process changes, more efficient FD statistics that can be developed based on a sequential of samples covered by a moving window-based approach and using some advanced distributions are needed [25].

In statistics, the Wishart distribution was created based on the Wishart matrix to describe the property of covariance matrix. Thus, using this distribution, some relevant FD statistics can be designed to track the process covariance changes. In [26], a FD statistic using the 2-norm of the Wishart matrix was developed. Similarly, applying other metrics to the Wishart matrix would allow the development of new statistics [27]–[29]. Furthermore, some FD statistics that were not originally proposed for multiplicative faults would have an improvement to detect this type of fault, such as the statistical local approach [30]–[35]. Some useful tools in communication field such as entropy [25] and Kullback-Leibler (KL) divergence [36]–[39], [41] have been reported to be useful in dealing with this type of change. For example, in [36], the KL-based statistic was developed and compared with T^2 and statistical local approach using the additive fault model. In [25], multivariate control charts for monitoring covariance matrix were investigated. In [43], basic FD statistics for detecting multiplicative faults were briefly reviewed. However, there is still no work focused on reviewing and comparing in detail these methods in order to determine the interconnections between different statistics. In addition, the problem of using FDR to evaluate the performance of them and providing the users and practitioners in FD field with information about which methods are to be used in which conditions. It should be noted that the false alarm rate (FAR) index is also commonly used in real applications. For statistics developed based on explicit probability distributions, the FAR value can appropriate a fixed value, thus will not be considered in this paper. Also note that rather than above methods proposed in the statistics framework, there are other methods in the literature, for example, the dissimilarity analysis-based methods [44], [45], data description-based methods, *e.g.* support vector data description [14], local outlier factor [46] and k nearest neighbor [14]. These methods have fell out of the scope of this paper, thus will not be deeply compared herein.

Therefore, motivated by above observations, the objectives of this paper are:

- to review the available FD statistics for detecting multiplicative faults and classify them according to their properties,
- to compare the methods, including their links and detection performance,
- to assess, using a numerical example, their detection performance based on the FDR index, and
- to perform simulation studies using the Tennessee Eastman (TE) benchmark process to show their applicability.

Notation: Let $\mathbb{R}^m$ be the m-dimensional Euclidean space; $\mathbb{R}^{n\times m}$ be the set of $n \times m$ real matrices; I_m be the $m \times m$-dimensional unit matrix; $tr(\cdot)$ represent the trace of a matrix; $x \sim \mathcal{N}_m(\mu_x, \Sigma_x)$ represent a m-variate normally distributed vector x with mean vector μ_x and covariance matrix Σ_x; $S \sim \mathcal{W}_m(\Sigma, n)$ represent Wishart distribution with covariance $\Sigma \in \mathbb{R}^{m\times m}$ and degree of freedom n; $E(\cdot)$ and $Var(\cdot)$ denote the mean and variance values; χ^2_m be the χ^2-distribution with m degrees of freedom; $\chi^2_m(\delta)$ be the noncentral χ^2 distribution with m degrees of freedom and noncentrality parameter δ. Let $prob\left(\chi^2 > \chi^2_{m,\alpha}\right) = \alpha$ represent the probability that $\chi^2 > \chi^2_{m,\alpha}$ equals α.

II. BACKGROUND AND MOTIVATION

In the model-based FD framework, the key step involves the generation of residual signals that, during normal operation, only contain system disturbances, but during process upsets contain both the system disturbances and fault information. The design of these residual generators has been the focus of much research. It can be noted that from this perspective, all MDA methods used in the multivariate statistics-based FD field can be referred as to different data-driven residual generators, for example, PLS generates residuals reflecting quality-related information.

A. BASIC FD STATISTICS

In the model-based FD framework, the key step is to generate residual signals that contain only system disturbances in normal operations and fault information as well as disturbances when a fault has occurred. Thus, significant effort has been made to design different residual generators. From this viewpoint, it can be assumed that MDA methods used in MSPM field are equivalent to different data-based residual generators, for example, PLS generates residuals reflecting the quality-related information.

Let $y \in \mathbb{R}^m$ be the residual data including m correlated signals. When the process operates under normal conditions, it is assumed that $y \sim \mathcal{N}_m(0, \Sigma_y)$, where Σ_y is the covariance matrix of y. Often Σ_y is unknown and must be estimated using historical process data, $y_1, ..., y_N$, using $\Sigma_y = \frac{1}{N-1}\sum_{i=1}^{N} y_i y_i^T$ [11]. On the other hand, when abnormal changes occur in the process, it can be assumed that $y_f \sim \mathcal{N}_m(\mu_f, \Sigma_f)$, where $\mu_f \neq 0$, $\Sigma_f \neq \Sigma_y$, or both. Note that faults that only lead to $\mu_f \neq 0$ are generally called additive faults [9]. They can be expressed as $y_f = y + f$ and give $y_f \sim \mathcal{N}_m(f, \Sigma_y)$. Those process changes that lead to $\Sigma_f \neq \Sigma_y$ are called multiplicative faults, which can be expressed as $y_f = My$ with $y_f \sim \mathcal{N}_m(0, M\Sigma_y M^T)$. It should be further noted that the mean vector of y is assumed to be unchanged for multiplicative faults.

The general fault detection procedure requires defining a FD statistic, J, and a corresponding threshold, J_{th}, or two thresholds, $J_{th,1}$ (upper) and $J_{th,2}$ (lower). Based on the relationship between the FD statistic and the threshold, two different cases can be determined:

- **Case I**: If $J > J_{th}$, or $J > J_{th,1}$ or $J < J_{th,2}$ then a fault is assumed to have occurred.
- **Case II**: If $J \leq J_{th}$, or $J_{th,1} \leq J \leq J_{th,2}$ then it is assumed that the process is operating properly.

Since additive faults have a large and immediate impact on process safety, a wealth of different methods and approaches have been proposed in the literature to handle such circumstances. The most commonly used FD statistics are T^2- and Q-statistics:

$$\begin{aligned} J_{T^2} &= y^T \Sigma_y^{-1} y \sim \chi_m^2 \\ J_Q &= y^T y \sim g\chi_h^2 \end{aligned} \tag{1}$$

where $g = tr\left(\Sigma_y^2\right)/tr\left(\Sigma_y\right)$, $h = tr^2\left(\Sigma_y\right)/tr\left(\Sigma_y^2\right)$. Note that Eq. (1) holds based on the assumption that the training data size is sufficiently large such that the estimated Σ_y approximates the real one. In general, J_{T^2} is used when the covariance matrix, Σ_y, is numerically stable, while J_Q, is used when the covariance matrix is not stable, that is, it may be difficult to compute Σ_y^{-1}.

B. PERFORMANCE EVALUATION FOR FD STATISTICS

Different metrics are used to evaluate the performance of a proposed fault detection method. The most common metric is *FDR*, which measures how accurately the proposed method can detect a given fault [11]. Mathematically, it can be written as [11]

$$\text{FDR} = \int_{J_{th}}^{\infty} f_{J_f}(x)dx \tag{2}$$

where $f_{J,f}$ is the probability density function (PDF) of J for detecting a specific fault. For additive faults, the resulting distribution for $f_{J,f}$ is a noncentral χ^2 distributed $f_{J_f}(x)$ [12]. Furthermore, it can be noted that any $f \neq 0$ can be tracked by J_{T^2} [12].

However, for multiplicative faults the behaviour of standard T^2-statistic is different. Before going into a detailed explanation, it is helpful to first consider a simple two-dimensional numerical example that shows the difference between the performance of the J_{T^2} statistic for additive and multiplicative faults. Let the normal covariance matrix be given as $\Sigma_y = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ Assume that the additive fault can be described as $f = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \times 1.5$ while the multiplicative fault can be described as $M = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$.

Fig. 2 shows the distribution of the statistic as well as the computed FDR for both the additive and multiplicative faults. The FDR is obtained by integrating $f_{J_f}(x)$ from J_{th}, which is shown as the vertical dashed line, to infinity. First, it can be seen from Fig. 2 (top, left) that once an additive fault occurs, the distribution of the FD statistic changes from a central χ^2-distribution (blue) to a noncentral one (green), while the threshold stays constant. On the other hand, for a multiplicative fault, the threshold shifts when a fault occurs. This shift can be obtained by dividing the original threshold by the magnitude of the fault. This is shown in Fig. 2 (top, right) by the movement of the dashed vertical line to the left, that is, to $J_{th}/3^2$ in this case. The FDR would then be computed by integrating between this new value of $J_{th}/3^2$ and infinity. In order to understand the impact that this shift has on the overall performance of the given method, additive and multiplicative faults with different magnitudes were simulated and the FDR values were computed. From Fig. 2 (bottom, left), it can be seen that for additive faults, above a given fault magnitude f, FDR will equal 1. However, for multiplicative faults, as seen in Fig. 2 (bottom, right), as the fault magnitude increase, M, the FDR will approach 1, but never reach it. This is a consequence of the fact that the threshold will approach zero, but not reach it. Thus, it can be seen that traditional FD statistics do not perform well for multiplicative faults. Thus, this motivates an investigation of other statistics that can improve the FD performance for multiplicative faults.

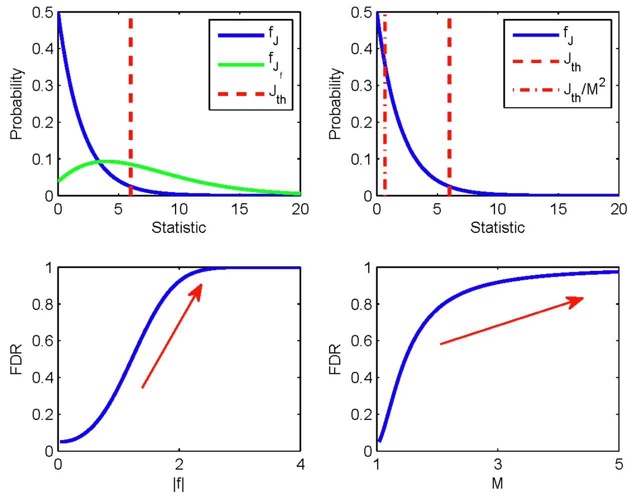

FIGURE 2. Comparisons between detecting additive and multiplicative faults.

Remark 1: In this paper, we pay more attention on investigating from the theoretical viewpoint the detectability of different statistics to detect multiplicative faults. Thus, it is assumed in the whole paper that only multiplicative faults have occurred, i.e. $y_f = My$. However, from a practical perspective, it is unknown whether the occurred fault is multiplicative or additive or both. In most practical cases, the occurrence of a fault can cause both types of changes to y, i.e. $y_f = My + f$. To prejudge the property of a potential fault, on can refer to [47].

III. SURVEY OF FD STATISTICS FOR MULTIPLICATIVE FAULTS

In this section, the available FD statistics for multiplicative faults can be classified, based on their area of application and underlying statistical assumptions, into three broad groups: modified T^2- and Q-statistic-based methods, Wishart distribution-based methods, and information theory-based methods.

A. MODIFIED T^2- AND Q-STATISTIC-BASED METHODS

Eq. (1) shows that J_{T^2} and J_Q are calculated using only the time series data of y. One method commonly used to improve the performance of such statistics for multiplicative faults is to consider not only the current sample, but also past samples from a given time horizon. Two common extensions, the cumulative T^2 and Q-statistics and the statistical local approach are considered.

1) CUMULATIVE T^2- AND Q-STATISTICS

Since it is commonly assumed that y is independent and identically distributed (i.i.d.), a simple extension of T^2 and Q is to incorporate a series of y samples to define a cumulative T^2-statistic as

$$J_{T_n^2} = \sum_{i=1}^{n} J_{T_i^2} = \sum_{i=1}^{n} y_i^T \Sigma_y^{-1} y_i \sim \chi^2_{nm} \tag{3}$$

Since $\forall i, y_i$ is i.i.d, $J_{T_n^2}$ follows a χ^2_{nm} distribution. Let $J_{th,T_n^2} = \chi^2_{nm,\alpha}$, the condition for Case I is determined by $J_{T_n^2} > J_{th,T_n^2}$. Similarly, the cumulative extension of the Q statistic can be developed identically to the T^2 case. Practically, we can chose an appropriate value for n and calculate $J_{T_n^2}$ at the current time instance i using y data within the interval from $i-n+1$ to i. Note that another extension of the T^2-statistic that takes the form of $\bar{y}^T \Sigma_y^T \bar{y} \sim \frac{1}{n}\chi^2_m$ with $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ has been recently proposed [51]. However, since it cannot be applied to multiplicative faults, it will not be considered.

2) STATISTICAL LOCAL APPROACH

The statistical local approach was proposed in model-based FD field for detecting changes in system model parameters. It has been incorporated into various methods, such as PCA, PLS, and CCA [31], [32], [35] as well as some machine learning techniques like support vector and kernel-based methods [33], [34], to detect incipient or small faults. The key contribution of this method is that it detects changes in Σ_y by taking into account the eigenvalues and eigenvectors of the system.

The statistical local approach requires a series of manipulations of the data in order to obtain appropriate thresholds and metrics. The first step is to perform a singular value decomposition on Σ_y to give $\Sigma_y = P\Lambda P^T$ with $P = [p_1, \ldots, p_m] \in \mathbb{R}^{m\times m}$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_m)$. Then, define a new variable, $l_i \triangleq z_i^T z_i - \lambda_i^2$ with $z_i = p_i^T y$ for $i = 1, \ldots, m$, which form an m-dimensional vector $L = [l_1, \ldots, l_m]^T$. It should be noted that $E(l_i) = 0$ and $E(L) = 0$. Using K samples of y, define a new random variable, ϕ as

$$\phi(l_i, K) \triangleq \frac{1}{\sqrt{K}} \sum_{j=1}^{K} l_i \tag{4}$$

Based on the central limit theorem, it follows that

$$\lim_{K\to\infty} \psi_K = [\phi(l_1, K), \ldots, \phi(l_m, K)]^T \sim \mathcal{N}_m\left(0, \Sigma_\psi\right) \tag{5}$$

where $\Sigma_\psi = \lim_{K\to\infty} \left(1/K\right) \sum_{j=1}^{K}\sum_{j=1}^{K} E\left(\psi_K \psi_K^T\right)$.

Practically, a small k_0 can be used, which gives $\psi_{k_0} = [\phi(l_1, k_0), \ldots, \phi(l_m, k_0)]^T$. Finally, based on the T^2-statistic, the statistical local approach defines its FD statistic as [30], [31]:

$$J_{\mathcal{L}} = \psi_{k_0}^T \Sigma_\psi^{-1} \psi_{k_0} \sim \chi^2_m \tag{6}$$

This statistic can be used to detect those multiplicative faults that affect the eigenvalues of Σ_y [30], [31]. Let $J_{th,\mathcal{L}} = \chi^2_{m,\alpha}$. Then, the condition for Case I is given as $J_{\mathcal{L}} > J_{th,\mathcal{L}}$. Note that $J_{\mathcal{L}}$ can also been determined based on the idea of the Q-statistic.

A brief summary of the steps required to compute this statistic is:

- ***S1:*** Calculate L using y, P, and Λ;
- ***S2:*** Select k_0 and calculate Σ_ψ using L;
- ***S3:*** Obtain $J_{\mathcal{L}}$ using the online L and Σ_ψ.

B. WISHART DISTRIBUTION-BASED METHODS

Let $S = \sum_{i=1}^{n} y_i y_i^T$ be the Wishart matrix. It follows that $S \sim \mathcal{W}_m\left(\Sigma_y, n\right)$. Given a sufficiently large n, S/n will approximate Σ_y. Therefore, changes in the elements of S can approximately mirror changes occurring in Σ_y. Using the properties of the Wishart distribution, three different metrics, the 2-norm, trace, and determinant, can be applied to S to develop the following FD statistics to deal with process changes affecting Σ_y. The Wishart distribution-based methods can be summarized as:

- ***S1:*** Obtain the Wishart matrix S;
- ***S2:*** Calculate the metrics for measuring S, and form the appropriate statistic;
- ***S3:*** Compare the statistics to the corresponding thresholds.

1) $||S||_2$-BASED METHOD

Let $\bar{y} = \Sigma_y^{-1/2} y$, which implies $\bar{y} \sim \mathcal{N}_m(0, I_m)$. As well, let $\bar{S} = \sum_{i=1}^{n} \bar{y}_i \bar{y}_i^T$, which gives $\bar{S} \sim \mathcal{W}_m(I_m, n)$. Since $\gamma = \left\| \frac{\bar{S}}{m} \right\|_2$ and $\frac{n}{m} \geq 1$, it follows that $\frac{\gamma - \mu_{nm}}{\sigma_{nm}} \sim TW_1$ [26], where TW_1 is the Tracy-Wisdom distribution [27], and $\mu_{nm} = \frac{1}{n}\left(\sqrt{n-1/2} + \sqrt{m-1/2}\right)^2$ and $\sigma_{nm} = \sqrt{\mu_{nm}/n}\left(1/\sqrt{n-1/2} + 1/\sqrt{m-1/2}\right)^{1/3}$.

Using these results, the FD statistic becomes

$$J_\gamma = \frac{\gamma - \mu_{nm}}{\sigma_{nm}} \tag{7}$$

It can be designed to detect changes in γ, which implies changes that occur in S. The upper and lower thresholds, J_{th,γ_1} and J_{th,γ_2}, are determined using TW_{1,α_1} and TW_{1,α_2} by properly selecting α_1 and α_2 [26]. The condition for Case I are either $J_\gamma > J_{th,\gamma_1}$ or $J_\gamma < J_{th,\gamma_2}$.

2) $tr(S)$-BASED METHOD

Let $\mathcal{T} = tr(S)$. Its distribution can be approximated by $\mathcal{T} \sim \sum_{i=1}^{m} \lambda_i \chi^2_n$ [28], where λ_i is the i^{th} eigenvalue of Σ_y. Then, based on above results, the FD statistic can be defined as

$$J_{\mathcal{T}} = tr(S) \tag{8}$$

The threshold, $J_{th,\mathcal{T}}$, is computed as

$$J_{th,\mathcal{T}} = \sum_{i=1}^{m} \lambda_i \chi^2_{n,\alpha} \tag{9}$$

The condition for Case I is then $J_\gamma > J_{th}$.

3) $det(S)$-BASED METHOD

Let $\mathcal{D}$ be $\left|\frac{S}{n}\right|$. It has been proven that $(2n)^m \frac{\mathcal{D}}{|\Sigma_y|} \sim \prod_{i=0}^{m-1} \chi^2_{2(n-i)}$ [29]. Then, the FD statistic can be defined as

$$J_{\mathcal{D}} = (2n)^m \frac{\mathcal{D}}{|\Sigma_y|} \tag{10}$$

This statistic can be used to track those changes in Σ_y that increase the determinant of S. The threshold is determined as

$$J_{th,\mathcal{D}} = \prod_{i=0}^{m-1} \chi^2_{2(n-i),\alpha} \tag{11}$$

The condition for Case I is $J_{\mathcal{D}} > J_{th,\mathcal{D}}$.

C. INFORMATION THEORY-BASED METHODS

It is possible to use some of the methods in information theory for measuring the characteristics of communication signals for detecting multiplicative faults. The two most commonly used such methods are the conditional entropy and the Kullback-Leibler divergence-based methods. The information theory-based methods can be summarized as:

- ***S1:*** Offline train σ_y^2 or σ_z^2;
- ***S2:*** Online estimate the required matrices and calculate $J_{\mathcal{C}}$ and $J_{\mathcal{K}}$;
- ***S3:*** Compare the statistics to the corresponding thresholds.

Finally, it should be noted that there are other metrics and concepts, such as mutual information, that are of interest for designing FD statistics [49]. However, at present, there does not exist any information about the associated probability distributions. Instead, the thresholds must be computed using kernel density estimation, which is an empirical method that does not provide any information. Thus, these methods are not considered in this paper.

1) CONDITIONAL ENTROPY-BASED APPROACH

The concept of entropy was first proposed in the statistical thermodynamics field. It was brought to the information theory as a logarithmic measure of the number of states with significant probability of being occupied. Given $y \sim \mathcal{N}_m(0, \Sigma_y)$, the entropy of y is

$$H(y) = \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \ln|\Sigma_y| \tag{12}$$

Guerrero-Cusumano [40] proposed an alternative expression for $H(y)$:

$$H(y) = \frac{1}{2} m \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^{m} \sigma_i^2 - T(y) \tag{13}$$

where $T(y)$ is called the mutual information of y and σ_i^2 is the variance of y_i. Assume that the correlation measured by $T(y)$ is known. Let

$$\Delta = H(y) - \hat{H}(y) = \sum_{i=1}^{m} \ln\left(\frac{s_i^2}{\sigma_i^2}\right) \tag{14}$$

which measures the difference between the estimated and theoretical entropy, where s_i^2 is the estimate of σ_i^2 using n online y data. Let

$$\mathcal{C} = \sqrt{\frac{n-1}{2m}} \sum_{i=1}^{m} \ln\left(\frac{s_i^2}{\sigma_i^2}\right) \tag{15}$$

Arthur *et al.* [25] showed that $\mathcal{C}$ is distributed asymptotically as a univariate standard normal distribution. Thus, $J_{\mathcal{C}}$ can be taken as a FD statistic to detect multiplicative faults affecting the entropy. The upper threshold $J_{th,\mathcal{C}_1}$ and lower threshold $J_{th,\mathcal{C}_2}$ are [25]:

$$\begin{aligned} J_{th,C_1} &= m\sqrt{\frac{2(n-1)}{m}}\left[G'\left(\frac{n-1}{2}\right) - \ln\left(\frac{n-1}{2}\right)\right] \\ &\quad + g_{\alpha/2} n \sqrt{mG''\left(\frac{n-1}{2}\right) + \frac{2}{n-1} tr(P_0 - I_m)^2} \\ J_{th,C_2} &= m\sqrt{\frac{2(n-1)}{m}}\left[G'\left(\frac{n-1}{2}\right) - \ln\left(\frac{n-1}{2}\right)\right] \\ &\quad - g_{\alpha/2} n \sqrt{mG''\left(\frac{n-1}{2}\right) + \frac{2}{n-1} tr(P_0 - I_m)^2} \end{aligned}$$

where G' and G'' are the first and second derivative of the natural logarithm of the Γ-function and g_α is the $1-\alpha$ quantile of $\mathcal{N}(0, 1)$. The condition for Case I can be given as either $J_{\mathcal{C}} > J_{th,\mathcal{C}_1}$ or $J_{\mathcal{C}} < J_{th,\mathcal{C}_2}$.

2) KULLBACK-LEIBLER DIVERGENCE

In probability and statistics theory, Kullback-Leibler (KL) divergence is a measure of the nonsymmetric difference between two probability distributions. Since $z = P^T y$, which satisfies $z \sim \mathcal{N}_m(0, \Sigma_z)$ with Σ_z being diagonal. Suppose that we have obtained $\mathcal{N}_m(0, \Sigma_z)$ offline. As well, we have estimated $\tilde{z} \sim \mathcal{N}_m(0, \tilde{\Sigma}_z)$ using n online z data. Using the KL divergence, it can be shown that

$$KL\left(\tilde{f}_z(z), f_z(z)\right) = \frac{1}{2}\left\{tr\left(\Sigma_z^{-1}\tilde{\Sigma}_z\right) - m\right\} + \frac{1}{2}\ln\left(\frac{\det(\Sigma_z)}{\det\left(\tilde{\Sigma}_z\right)}\right) \tag{16}$$

where $f_z(\cdot)$ and $\tilde{f}_z(\cdot)$ are the actual and online estimated PDFs of z. It was further proven in [36] that, given a sufficiently large n,

$$J_{\mathcal{K}} = 2nKL(\tilde{f}_z(z), f_z(z)) \sim \chi^2_m \tag{17}$$

is satisfied, where [13], [37]

$$KL\left(\hat{f}_z(z), f_z(z)\right) = \frac{1}{2}\sum_{i=1}^{m}\left\{\ln\left(\frac{\sigma_{z,i}^2}{s_{z,i}^2}\right) + \frac{s_{z,i}^2}{\sigma_{z,i}^2} - 1\right\} \tag{18}$$

TABLE 1. Comparison of FD statistics using χ^2 distribution.

Name	Expression	Threshold	Condition for Case I
T^2-statistic	$J_{T^2} = y^T \Sigma_y^{-1} y$	$a = 1, b = m$	$J_{T_f^2} > J_{th,T_f^2}$
Cumulative T^2-statistic	$J_{T_n^2} = \sum_{i=1}^n y_i^T \Sigma_y^{-1} y_i$	$a = 1, b = mn$	$J_{T_n^2} > J_{th,T_n^2}$
Q-statistic	$J_Q = y^T y$	$a = g, b = h$	$J_Q > J_{th,Q}$
Cumulative Q-statistic	$J_{Q_n} = \sum_{i=1}^n y_i^T y_i$	$a = g, b = nh$	$J_{Q_n} > J_{th,Q_n}$
Statistical local approach	$J_{\mathcal{L}} = \psi_{k_0}^T \Sigma_\psi^{-1} \psi_{k_0}$	$a = 1, b = m$	$J_{\mathcal{L}} > J_{th,\mathcal{L}}$
Trace of the Wishart matrix	$J_{\mathcal{T}} = tr(S)$	$a = tr(\Sigma_y), b = n$	$J_{\mathcal{T}} > J_{th,\mathcal{T}}$
KL divergence	$J_{\mathcal{K}} = 2nKL(\tilde{f}(z), f(z))$	$a = 1, b = m$	$J_{\mathcal{K}} > J_{th,\mathcal{K}}$

TABLE 2. Parameters involved in J_{T^2} and J_Q for detecting multiplicative faults.

FD Statistics	c	d	ζ
J_{T^2}	$\frac{tr\left(\Sigma_y^{-1} M \Sigma_y M \Sigma_y^{-1} M \Sigma_y M\right)}{tr\left(\Sigma_y^{-1} M \Sigma_y M\right)}$	$\frac{tr^2\left(\Sigma_y^{-1} M \Sigma_y M\right)}{tr\left(\Sigma_y^{-1} M \Sigma_y M \Sigma_y^{-1} M \Sigma_y M\right)}$	$\chi^2_{m,\alpha} \big/ c$
J_Q	$\frac{tr(M\Sigma_y M \Sigma_y)}{tr(M\Sigma_y)}$	$\frac{tr^2(\Sigma_y M)}{tr(M\Sigma_y M\Sigma_y)}$	$g\chi^2_{h,\alpha} \big/ c$

Thus, the relevant FD statistic is $J_{\mathcal{K}} = 2nKL(\tilde{f}_z(z), f_z(z))$, which can be used to track changes in Σ_y. The associated threshold is $J_{th,\mathcal{K}} = \chi^2_{m,\alpha}$. The condition for Case I is $J_{\mathcal{K}} > J_{th,\mathcal{K}}$.

IV. COMPARISON AND PERFORMANCE EVALUATION OF FD STATISTICS FOR MULTIPLICATIVE FAULTS

This section will compare the different FD statistics, as well as their performance for detecting multiplicative faults.

A. GENERAL COMPARISONS

Before considering the performance of the different statistics, it is helpful to consider how the different statistics can be summarised. Firstly, it can be noted that many of the statistics have been developed based on the χ^2 distribution. In fact, the thresholds for such statistics can be written as $J_{th} = a\chi^2_{b,\alpha}$, where a and b are constants that depend on the specific method. A summary of the different statistics and their values of a and b is shown in Table 1. It should be noted that, although J_{T^2}, $J_{\mathcal{L}}$ and $J_{\mathcal{K}}$ use different formulae to calculate the statistic, they all have the same χ^2 distribution for normal operating data. This implies that they will have the same threshold. Furthermore, it can be ascertained that based on the properties of the χ^2, the detection signal will always be positive for such statistics.

Note that not all methods can be summarized using this formulation. Some, such as the third Wishart-based method, require the use and construction of complex additional parameters and distributions. Furthermore, the statistics based on 2-norm of the Wishart matrix S and conditional entropy do not assume a χ^2-related distribution and cannot be guaranteed to be positive. Thus, J must be compared against upper and lower thresholds given as J_{th1} and J_{th2} to check the condition for Case I.

B. PERFORMANCE EVALUATION

This section will evaluate the performance of the different statistics for detecting multiplicative faults. The three key evaluations will be considered: comparison of the T^2 and Q-based statistics, comparison of the Wishart statistics, and a comparison of the KL-based method with the T^2-based methods.

1) EVALUATION OF THE DIFFERENT T^2 AND Q-BASED STATISTICS

As has been mentioned, when detecting multiplicative faults of the form $y_f = My$, the standard J_{T^2} and J_Q can be rewritten for the faulty case as:

$$\begin{aligned} J_{T_f^2} &= y_f^T \Sigma_y^{-1} y_f \\ J_{Q_f} &= y_f^T y_f \end{aligned} \tag{19}$$

Since the thresholds for these statistics can be written in compact form as $c\chi^2_{d,\alpha}$, with the values read from Table 2, it is possible to compactly analyse the performance of these methods using the FDR. From Eq. (2), the compact expression for the FDR is

$$FDR = prob\left(\chi^2_d > \zeta\right) = \int_\zeta^\infty f_{\chi^2_d}(x)dx \tag{20}$$

where ζ is computed based on the formulae in Table 2. Thus, FDR performance for $J_{T_n^2}$ is computed as

$$FDR_{T_n^2} = prob\left(\chi^2_{nd} > \chi^2_{nm,\alpha} \big/ c\right) \tag{21}$$

With these definitions, it is now possible to state some results regarding the performance between the standard and cumulative T^2-statistics. Firstly, it can be obtained that when d is close to m,

$$\begin{aligned} FDR_{T^2} &= prob\left(\chi^2_d > \chi^2_{d,\alpha} \big/ c\right) \\ &< prob\left(\chi^2_{nd} > \chi^2_{nd,\alpha} \big/ c\right) = FDR_{T_n^2} \end{aligned} \tag{22}$$

thus, the cumulative T^2-statistic is better than the plain T^2-statistic. For the case that c is sufficiently large, $\chi^2_{d,\alpha} \big/ c$ approximates $\chi^2_{nd,\alpha} \big/ c$, which also guarantees Eq. (22). Similarly, when $c \gg g$, $FDR_{Q_n} \geq FDR_Q$.

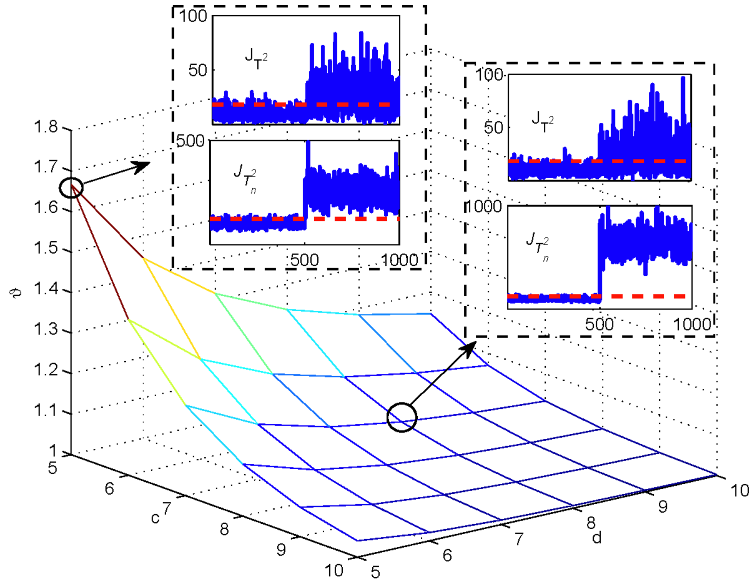

FIGURE 3. Performance of ϑ for different c and d, $m = 10$, $n = 10$.

Since there exist regions where one or the other form of the FD statistic performs better, it is important to understand how the performance changes in each of the regions. In order to accurately measure the difference, define the ratio of the FDR for the two statistics as the performance index $\vartheta = FDR_{T_n^2} \big/ FDR_{T^2}$. When $\vartheta > 1$, this implies that $J_{T_n^2}$ performs better than J_{T^2}.

In order to further investigate the performance differences, a series of simulations was performed for different values of c and d between 5 and 10 and the results compiled. Fig. 3 shows ϑ as a function of c and d. It can be observed that in all cases, ϑ is larger than 1, which means $J_{T_n^2}$ gives a better performance. In Section 2, it was shown that FDR_{T^2} approaches 1 but cannot theoretically reach 1 when detecting multiplicative faults, even if they gets large enough. Using $J_{T_n^2}$, can in some cases can overcome this disadvantage. This improvement can be observed from the two subfigures in Fig. 3. The first one corresponds to a fault with $c = d = 5$. It can be clearly seen that $FDR_{T_n^2} = 1$ compared to $FDR_{T^2} < 1$. Considering another fault with $c = d = 8$, the second subfigure shows that FDR_{T^2} has increased, but is still smaller than 1. On the other hand, using $J_{T_n^2}$ guarantees $FDR_{T_n^2} = 1$. Fig. 4 examines the performance of ϑ corresponding to different n, where the fault with $c = 5$ and $d = 5$ is used. It can be found that as n increases, ϑ likewise increases and finally approaches 1.65. Since, a large n requires extensive computation efforts, $n = 8$ or 9 can be chosen for detecting this fault. In other cases, $J_{T_n^2}$ may perform worse than J_{T^2}. For example, let $c = 3$, $d = 3$ and $n = 5$, it leads to $FDR_{T^2} = 0.1067 > 0.0953 = FDR_{T_n^2}$. Likewise, Fig. 5 shows the value of the ϑ performance index for different n. It can be found that $\vartheta < 1$ for all choices of n. As n increases, ϑ decreases. This shows that when

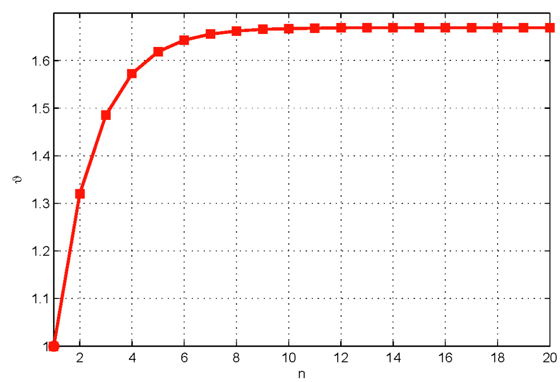

FIGURE 4. Performance of ϑ for different n, $m = 10$, $c = 5$ and $d = 5$.

$d \ll m$ and c is not big enough, $J_{T_n^2}$ cannot improve the FDR performance even if a large n is chosen. Note that it is difficult to find a condition explicitly distinguishing the use of $J_{T_n^2}$ and J_{T^2}. Finally, it should be mentioned that in cases of improving the FDR performance, $J_{T_n^2}$ and J_{Q_n} cannot increase the probability of false detection alarms.

2) EVALUATION OF THE WISHART DISTRIBUTION-BASED STATISTICS

The methods based on the Wishart distribution can detect multiplicative faults that affect the Wishart matrix S. The following results present some key observations regarding the different Wishart-based statistics and their relationships with other statistics.

Theorem 1: $J_{\mathcal{T}}$, which takes the trace of S, can only track faults that increase the variance of y.

Proof: This theorem will be proved by examining the performance of the method for both the normal and

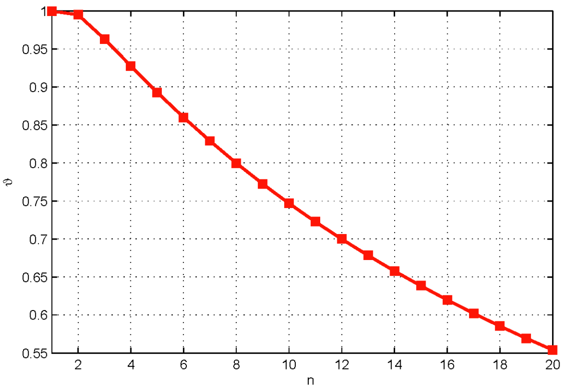

FIGURE 5. Performance of ϑ for different n, $m = 10$, $c = 3$ and $d = 3$.

faulty cases. A multiplicative fault, $y_f \sim \mathcal{N}_m\left(0, M\Sigma_y M^T\right)$ implies that the faulty S_f follows $\mathcal{W}_m(M\Sigma_y M^T, n)$. Thus, $J_{\mathcal{T}}$ becomes $J_{\mathcal{T},f} \sim \sum_{i=1}^{m} \lambda_{f,i}\chi_n^2$. This implies that only if $\sum_{i=1}^{m} \lambda_{f,i} = tr\left(\Sigma_{y,f}\right) > tr\left(\Sigma_y\right) = \sum_{i=1}^{m} \lambda_i$, the fault can be detected by $J_{\mathcal{T}}$. Thus, those faults that increase the variance can be tracked by this statistic. ■

Theorem 2: $J_{\mathcal{T}}$ provides a more conservative threshold than J_{Q_n}.

Proof: Firstly, note that both $J_{\mathcal{T}}$ and J_{Q_n} have the same formulation, but different estimates for the distribution, that is, J_{Q_n} is based on the Q-statistic with a distribution approximated using the scaled χ^2 distribution, while $J_{\mathcal{T}}$ is based on the Wishart distribution. The mean value of the two statistics is

$$E(J_{Q_n}) = ngh = n\sum_{i=1}^{m} \lambda_i = E(J_{\mathcal{T}}) \tag{23}$$

The variance is

$$\begin{aligned} Var\left(J_{Q_n}\right) &= 2ng^2h = 2n\sum_{i=1}^{m} \lambda_i^2 \\ &\leq Var\left(J_{\mathcal{T}}\right) = 2n\Big(\sum_{i=1}^{m} \lambda_i\Big)^2 \end{aligned} \tag{24}$$

Thus, since the means are the same, but the variance for $J_{\mathcal{T}}$ is larger, it can be concluded that $J_{\mathcal{T}}$ provides a more conservative threshold. ■

Theorem 3: $J_{T_n^2}$ can also be interpreted using the Wishart distribution, and $J_{\mathcal{T}}$ is a special case of $J_{T_n^2}$ using I_m instead of Σ_y^{-1}.

Proof: In order to prove the result, it will be necessary to show that the two statistics have a similar form. From Eq. (3), we have that

$$E(J_{T_n^2}) = E(\chi_{nm}^2) = nm \tag{25}$$

and

$$Var(J_{T_n^2}) = Var(\chi_{nm}^2) = 2nm \tag{26}$$

Now, note that $J_{T_n^2}$ can be alternatively expressed as

$$J_{T_n^2} = tr(S\Sigma_y^{-1}) = tr(\Sigma_y^{-1/2}S\Sigma_y^{-1/2}) \tag{27}$$

Since $S \sim \mathcal{W}_m(\Sigma_y, n)$ and $\Sigma_y^{-1/2}S\Sigma_y^{-1/2}$ follows $\mathcal{W}_m(I_m, n)$ [42], it follows that

$$\begin{aligned} E(tr(\Sigma_y^{-1/2}S\Sigma_y^{-1/2})) &= tr(E(\Sigma_y^{-1/2}S\Sigma_y^{-1/2})) \\ &= tr(E(W_m(I_m, n))) = nm \end{aligned} \tag{28}$$

and

$$\begin{aligned} Var(tr(\Sigma_y^{-1/2}S\Sigma_y^{-1/2})) &= tr(Var(\Sigma_y^{-1/2}S\Sigma_y^{-1/2})) \\ &= tr\left(Var(W_m(I_m, n))\right) = 2nm \end{aligned} \tag{29}$$

Comparing Eqs. (25) and (26) with Eqs. (28) and (29) implies that $J_{T_n^2}$ can be understood using the Wishart distribution. Let $\Sigma_y^{-1} = I_m$, $J_{T_n^2}$ reduces to $J_{\mathcal{T}}$. ■

Finally, we can note that J_γ is preferred for detecting faults that lead to changes in the 2-norm of S. Such faults can either increase or decrease $\|S\|_2$ and still be tracked by J_γ. Similarly, $J_{\mathcal{D}}$ is designed to deal with faults that increase the determinant of S. Since $|S| = \prod_i \lambda_{s,i}$, where $\lambda_{s,i}$ is the eigenvalue of S, it can be seen that faults that affect J_γ will also affect $J_{\mathcal{D}}$. Thus, based on the above results, it is recommended that these three statistics be used together in order to achieve better FD performance.

3) EVALUATION OF $J_{\mathcal{L}}$ AND A COMPARISON WITH J_{T^2}

This section will show how $J_{\mathcal{L}}$ can improve the performance over J_{T^2}. Theorem IV-B.3 shows that $J_{\mathcal{L}}$ transfers the detection of multiplicative faults to detecting additive faults.

Theorem 4: The FDR for $J_{\mathcal{L}}$ can be obtained using a noncentral χ^2 distribution.

Proof: A multiplicative fault $y_f = My$ gives $z_f = P^TMy$, such that $E\left(l_{f,i}\right) = E\left(z_{f,i}^Tz_{f,i} - \lambda_i^2\right) \neq 0$. Thus, we have

$$\begin{aligned} \lim_{K\to\infty} \psi_{K,f} &= \left[\phi\left(l_{1,f}, K\right), \ldots, \phi\left(l_{m,f}, K\right)\right] \\ &\sim \mathcal{N}_m\left(\Delta\psi, \Sigma_{\psi,f}\right) \end{aligned} \tag{30}$$

where $\Delta\psi \neq 0$. In this case, according to [48], $J_{\mathcal{L},f} = \psi_{K,f}^T\Sigma_\psi^{-1}\psi_{K,f}$ follows a scaled noncentral χ^2 distribution defined as

$$J_{\mathcal{L},f} \sim \sum_{i=1}^{m} c_i\chi_{h_i}^2\left(\delta_i\right) \tag{31}$$

where the details on the computation of c_i and h_i can be found in [48]. Therefore, we can use a noncentral $\chi_h^2(\delta)$ distribution to approximate the above distribution. Thus, $FDR_{\mathcal{L}} = prob\left(\chi_h^2(\delta) > \chi_{m,\alpha}^2\right)$ is obtained, which establishes the theorem. ■

Based on Theorem 4, the following proposition is given to show that the performance given by $J_{\mathcal{L}}$ is better compared to J_{T^2}.

Proposition 1: For the case with an individual y, $FDR_{\mathcal{L}}$ is obtained using the noncentral χ_1^2 distribution. Given that $M \geq 1$, it further leads to $FDR_{\mathcal{L}} \geq FDR_{T^2}$.

Proof: In this case, $y_f = My$ with M being a scalar. Using J_{T^2} gives

$$J_{T_f^2} = \frac{M^2y^2}{\sigma^2} \sim M^2\chi_1^2 \tag{32}$$

Then, the FDR for J_{T^2} is calculated as:

$$FDR_{T^2} = prob\left(\chi_1^2 > \frac{\chi_{1,\alpha}^2}{M^2}\right) \tag{33}$$

It can be observed that when $M < 1$, $FDR_{T^2} < \alpha$, such faults cannot be detected by J_{T^2}. Considering $J_{\mathcal{L}}$, it can be rewritten as $J_{\mathcal{L}} = \frac{\psi^2}{\sigma_\psi^2}$, where $\psi = \frac{1}{\sqrt{k_0}}\sum_{i=1}^{k_0}\left(y_i^2 - \sigma^2\right)$, $\sigma_\psi^2 = var\left(y^2 - \sigma^2\right) = 2\sigma^4$, and σ^2 is the variance of y. In the presence of multiplicative faults, $J_{\mathcal{L}}$ changes to

$$J_{\mathcal{L},f} = \frac{\psi_f^2}{\sigma_\psi^2} = \frac{1}{k_0}\sum_{i=1}^{k_0}\frac{\left(M^2y_i^2 - \sigma^2\right)^2}{2\sigma^4} \tag{34}$$

As $E\left(\psi_f\right) = \sqrt{k_0}\left(M^2 - 1\right)\sigma^2$ and $Var\left(\psi_f\right) = 2M^4\sigma^4$, then it holds for $J_{\mathcal{L},f}$ that

$$\frac{J_{\mathcal{L},f}}{M^4} \sim \chi_1^2\left(\delta\right) \to J_{\mathcal{L},f} \sim M^4\chi_1^2\left(\delta\right) \tag{35}$$

where $\delta = \frac{k_0\left(M^2-1\right)^2\sigma^4}{2M^4\sigma^4} = \frac{k_0\left(M^2-1\right)^2}{2M^4}$ is the noncentrality parameter. Likewise, the FDR for $J_{\mathcal{L}}$ is calculated as

$$FDR_{\mathcal{L}} = prob\left(\chi_1^2\left(\delta\right) > \frac{\chi_{1,\alpha}^2}{M^4}\right) \tag{36}$$

Given $M \geq 1$, it holds that

$$\begin{aligned} FDR_L = prob\left(\chi_1^2\left(\delta\right) > \frac{\chi_{1,\alpha}^2}{M^4}\right) &\geq prob\left(\chi_1^2 > \frac{\chi_{1,\alpha}^2}{M^4}\right) \\ \geq prob\left(\chi_1^2 > \frac{\chi_{1,\alpha}^2}{M^2}\right) &= FDR_{T^2} \end{aligned} \tag{37}$$

■

A numerical example is used to demonstrate the above results. Before evaluating their performance using Eqs. (33) and (36), we first examine the property of δ. Let $k_0 = 50$, the trajectory of δ along with M is shown in Fig. 6. It can be seen that the minimum value, namely $\delta = 0$ occurs at $M = 1$, otherwise, it increases as M increases to infinity or decreases to zero. Compared with the results in Fig. 2, it can be observed that the formula calculating FDR for $J_{\mathcal{L}}$ can be described using the noncentral χ^2 distribution, which is equivalent to using T^2 and Q statistics to detect additive faults. To show the improvement, define an index, υ as $\upsilon = \frac{E(J_f)}{\chi_{1,\alpha}^2}$ to examine their detectability. υ measures the gap between the FD statistic value over the threshold and the larger υ is, the better detectability the FD statistic gives. Based on Eq. (32), $E\left(J_{T_f^2}\right) = M^2$ is obtained, and taking the expectation of $J_{\mathcal{L},f}$ gives $E\left(J_{\mathcal{L},f}\right) = \frac{(k_0+2)M^4 - 2k_0M^2 + 1}{2}$. Fig. 7 shows the results. With M varying from 1 to 50, it can be found that $J_{\mathcal{L}}$ always offers a larger υ, which implies $J_{\mathcal{L}}$ performs better in all these cases. In terms of FDR, it can be found that $J_{\mathcal{L}}$ gives a larger index. This can be verified as shown in Fig. 8, where the FDR values given by $J_{\mathcal{L}}$ are higher than those by J_{T^2}. Further note that as M increases, $J_{\mathcal{L}}$ can rapidly reach $FDR = 1$, while J_{T^2} cannot, even though M is sufficiently large.

For faults decreasing the variances of y, according to Eq. (35), δ will likewise increase, which will increase the FDR. However, from Eq. (35), these faults results in the increase of $\chi_{1,\alpha}^2 / M^4$, which will decrease the FDR. The final performance to this type of fault can be affected by these two factors. Fig. 9 shows the performance of $J_{\mathcal{L}}$ by setting

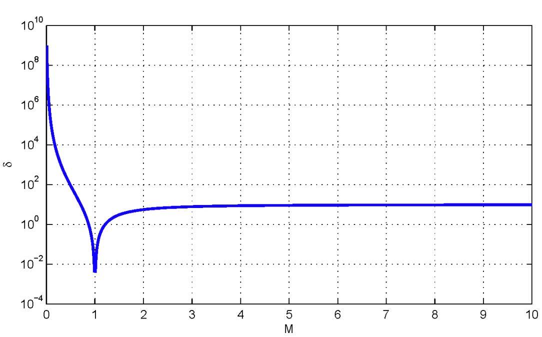

FIGURE 6. Trajectory of δ along with M ($k_0 = 50$).

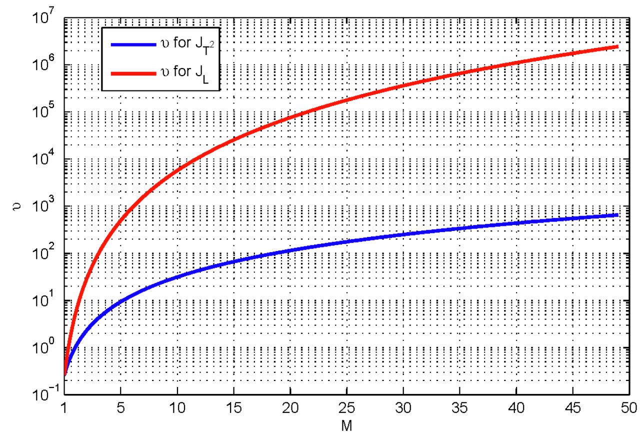

FIGURE 7. Performance of υ for J_{T^2} and $J_{\mathcal{L}}$. ($n = 50$).

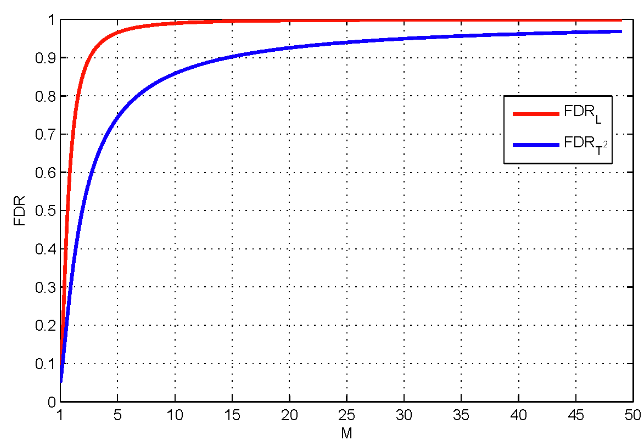

FIGURE 8. Performance of FDR for J_{T^2} and $J_{\mathcal{L}}$ ($n = 50$).

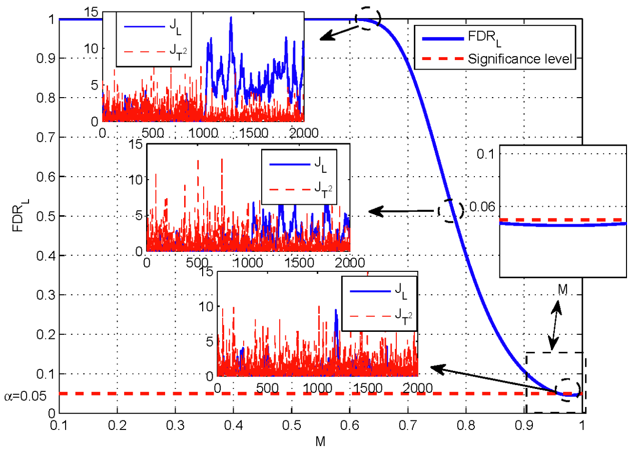

FIGURE 9. Performance of FDR for detecting multiplicative faults decreasing measurement variances ($n = 50$).

$0.01 < M <= 1$. It can be observed that the trend of FDR increases with the decrease of M. For the case when M is close to 1, the second factor impacts more than the first, thus the FDR is smaller than α. In other cases, $J_{\mathcal{L}}$ shows good performance, and at around $M = 0.6$, $J_{\mathcal{L}}$ can completely detect this fault.

4) COMPARISON BETWEEN $J_{\mathcal{L}}$ AND $J_{\mathcal{K}}$

This section investigates the equivalence between $J_{\mathcal{L}}$ and $J_{\mathcal{K}}$ in the fault-free case. Theorem 5, which follows, shows the relationship between $J_{\mathcal{L}}$ and $J_{\mathcal{K}}$.

Theorem 5: Given a large number of y data, $J_{\mathcal{K}}$ and $J_{\mathcal{L}}$ have the same formula to calculate.

Proof: For the case with an individual y, $J_{\mathcal{K}}$ is rewritten as:

$$J_{\mathcal{K}} = n\left(\ln\left(\frac{1}{x}\right) + x - 1\right) \tag{38}$$

where $x = \frac{s^2}{\sigma^2}$. Note that a sufficient large n can lead to $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 \to \sigma^2$, namely $x \to 1$. In this case, $J_{\mathcal{K}}$ can be approximated using the second order Taylor series at the point $x = 1$, that is

$$J_{\mathcal{K}} \approx n\left(\frac{1}{2}(x-1)^2\right) \tag{39}$$

Then, plugging $x = \frac{s^2}{\sigma^2}$ into Eq (39), it can be obtained as:

$$\begin{aligned} J_{\mathcal{K}} &= n\left(\frac{1}{2}\left(\frac{s^2}{\sigma^2}-1\right)^2\right) = n\left(\frac{\left(s^2-\sigma^2\right)^2}{2\sigma^4}\right) \\ &= n\left(\frac{\left(\frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \sigma^2\right)^2}{2\sigma^4}\right) \\ &= \frac{n}{(n-1)^2}\left(\frac{\left(\sum_{i=1}^{n}(y_i^2-\sigma^2)\right)^2}{2\sigma^4}\right) \\ &\to \frac{1}{n}\left(\frac{\left(\sum_{i=1}^{n}(y_i^2-\sigma^2)\right)^2}{2\sigma^4}\right) \\ &\overset{k_0 \to n}{\Leftrightarrow} J_{\mathcal{L}} \end{aligned} \tag{40}$$

That is to say, if k_0 is chosen so big as n, $J_{\mathcal{K}}$ will be equivalent to $J_{\mathcal{L}}$. For the case with multiple variables in y, $z_i = p_i^T y$ with $i = 1, \ldots, m$, is sent to these two statistics. Since $\forall i$, z_i are mutually independent, the above results can be simply extended to the multiple y case. ■

This result explains why in Table 1 that $J_{\mathcal{K}}$ and $J_{\mathcal{L}}$ are assumed to follow the same distribution.

C. ADDITIONAL OBSERVATIONS

1) COMPUTATIONAL REQUIREMENTS

Although many of the methods only require a small amount of data, $J_{\mathcal{C}}$, $J_{\mathcal{K}}$ and $J_{\mathcal{L}}$ are calculated using a large amount of online y data, which implies that they require more memory. However, the advantage of these methods is that they can detect faults that increase or decrease the variance.

As well, it can be noted that, for computing $J_{\mathcal{K}}$, the residual signal y should be first converted to z using $z = P^T y$. Such handling can not only increase the computational complexity, but also lead to a loss of fault information. Similarly, this paper does also not consider how changes in P affect the detection performance of $J_{\mathcal{L}}$. However, as shown above, compared with $J_{\mathcal{C}}$ and $J_{\mathcal{K}}$, $J_{\mathcal{L}}$ detects multiplicative faults in an efficient way. Finally, a moving window-based method can be incorporated with $J_{\mathcal{L}}$ to improve its applicability.

2) OTHER PERFORMANCE METRICS

In addition to the FDR index, another commonly used index to evaluate the performance of FD methods is DD. Compared to J_{T^2} and J_Q, all considered statistics are calculated using more than one y data. Even with the moving window-based methods, there exist cases such that some samples within the window are fault-free. For such cases, implementing these statistics may cause a larger detection delay. For example, Fig. 10 shows an comparison of using $J_{\mathcal{L}}$ and J_{T^2} to detect a fault with a change in σ^2 from 4 to 9. It can be seen that the improvements in the FDR performance are gained at the cost of causing a larger detection delay. This brings new challenges to investigating how to balance the two important indices in practical applications. Additional information regarding this issue can be found in [51] and [52].

It can be noted that, although the theoretical research shows that the considered statistics themselves are capable of detecting multiplicative faults, in practice in multivariate statistics-based FD approaches they must be incorporated

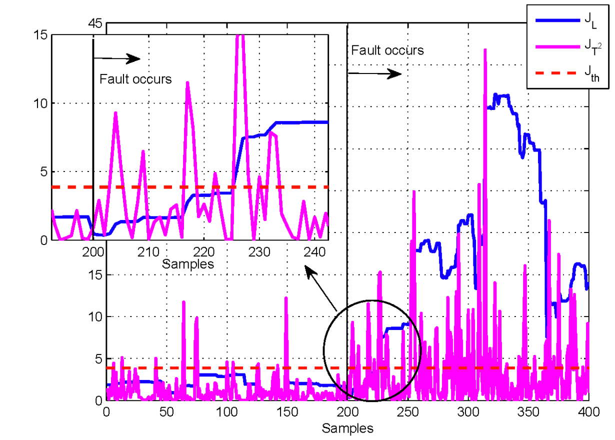

FIGURE 10. Detection delay caused by the statistical local method ($M = 1.5$, $n = 50$).

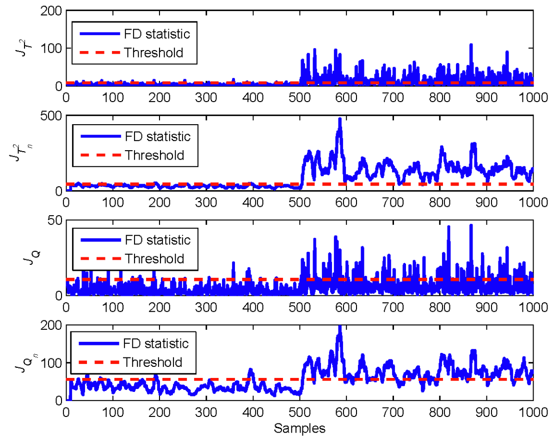

FIGURE 11. Performance of J_{T^2}, $J_{T_n^2}$, J_Q and J_{Q_n} for Scenario 1.

with MDA methods to achieve a better FD or monitoring performance, of which MDA methods are responsible for generating residual signals that contain valuable faulty information. How to properly modify MDA methods such that they can be linked well with FD statistics also plays an important role in FD approaches. Additional information can be found in [31], [32], and [34]–[36].

V. SIMULATIONS AND INDUSTRIAL BENCHMARK STUDY

This section consists of case studies using a numerical model and the TE benchmark process.

A. CASE STUDY ON NUMERICAL EXAMPLES

To evaluate the overall performance of the considered statistics, two numerical multiplicative fault scenarios are first implemented:

- Scenario 1: $\Sigma_y = \begin{bmatrix} 2 & 0.5 & 0.4 \\ 0.5 & 1 & 0.3 \\ 0.4 & 0.3 & \boxed{0.5} \end{bmatrix} \rightarrow \Sigma_f = \begin{bmatrix} 2 & 0.5 & 0.4 \\ 0.5 & 1 & 0.3 \\ 0.4 & 0.3 & \boxed{5} \end{bmatrix}$
- Scenario 2: $\Sigma_y = \begin{bmatrix} 2 & 0.5 & 0.4 \\ 0.5 & 1 & 0.3 \\ 0.4 & 0.3 & \boxed{0.5} \end{bmatrix} \rightarrow \Sigma_f = \begin{bmatrix} 2 & 0.5 & 0.4 \\ 0.5 & 1 & 0.3 \\ 0.4 & 0.3 & \boxed{0.1} \end{bmatrix}$

Using Σ_y generate 1,000 fault-free samples, then the required parameters for some of the statistics can be trained. For Scenario 1, 1,000 samples are generated with the fault occurring from 501 to 1000. Fig. 11 shows the performance of $J_{\mathcal{T}^2}$, $J_{\mathcal{T}_n^2}$, J_Q and J_{Q_n}. It can be observed that $J_{\mathcal{T}_n^2}$ and J_{Q_n} have significantly improved the detection performance compared to $J_{\mathcal{T}^2}$ and J_Q. Fig. 12 shows the performance of J_γ, $J_\mathcal{T}$ and $J_\mathcal{D}$, where all can successfully detect this fault, and perform better than $J_{\mathcal{T}^2}$. Of them, J_γ gives the best performance,

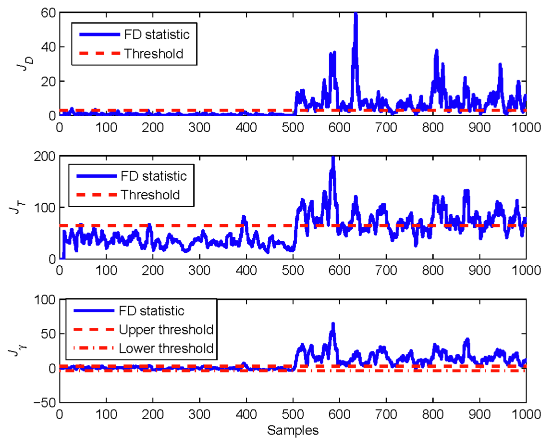

FIGURE 12. Performance of J_γ, $J_\mathcal{T}$ and $J_\mathcal{D}$ for Scenario 1.

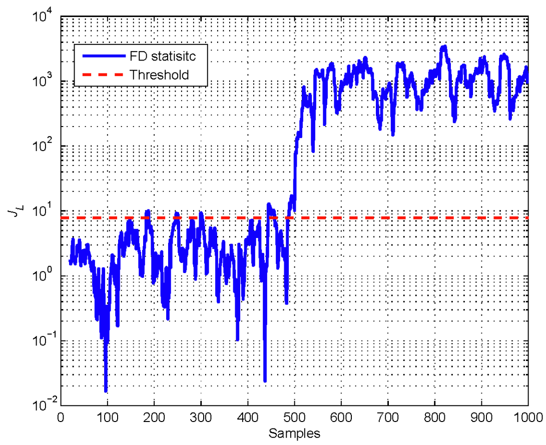

FIGURE 13. Performance of $J_\mathcal{L}$ for Scenario 1.

which is followed by $J_\mathcal{D}$ and $J_\mathcal{T}$. Note that J_{Q_n} and $J_\mathcal{T}$ are equivalent, while J_{th,Q_n} is smaller than $J_{th,\mathcal{T}}$. This is consistent with Theorem 2. From Fig. 13, $J_\mathcal{L}$ gives the best performance for this fault.

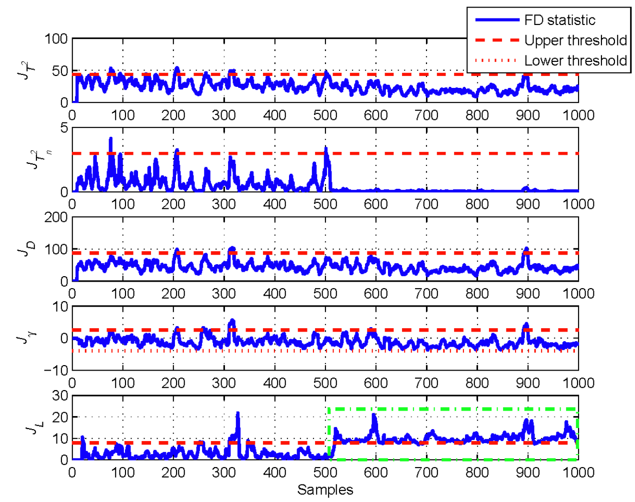

FIGURE 14. **Performance of $J_{T^2}, J_{T_n^2}, J_{\mathcal{D}}, J_{\gamma}$ and $J_{\mathcal{L}}$ for Scenario 2.**

Scenario 2 simulates a fault that decreases the variance of y. The detection results given by J_{T^2}, $J_{T_n^2}$, $J_{\mathcal{D}}$, J_{γ} and $J_{\mathcal{L}}$ are shown in Fig. 14. It can be observed that the fault can only be tracked by $J_{\mathcal{L}}$. Thus, for such type of fault, $J_{\mathcal{L}}$ is suggested.

B. CASE STUDY ON THE TE BENCHMARK PROCESS

The TE process has been widely used as an industrial benchmark to evaluate the performance of different methods in process control and monitoring fields [50]. It simulates a real chemical plant that produces the liquid products G and H with five gaseous reactants: A-E. The detailed piping and instrumentation diagram of TE process is shown in Fig. 15. The five reactants are fed into the reactor, where products G and H as well as by-product F are produced. Then, the products go through the condenser and reach the separator, where the uncondensed part is compressed back to the reactor by the compressor, the condensed part will be shifted to the stripper where the remaining reactants are removed, then the purified liquid products G an H are generated. TE process defines 22 process variables and 12 manipulated variables as summarized in Table 3, where the variable tags are consistent with their original definitions in [50]. The process variables mainly record the running environment in different units of the process, and the manipulated variables are chosen to manipulate the amount of input reactants, velocity of flow, and agitator speed, such that, given control laws, the process can operate as desired performance. In addition to the two types of variables, TE process also provides another type of variables to measure the component of five reactants and two products in different units, of which the most important two are the component of G and H at the exit of the process. They have often been called as quality variables that should be specially concerned. However, these variables are observed with a time delay varying from 6 to 15 min compared to process variables. For more detailed descriptions, one can refer to [50] and [53]. The MATLAB simulator is downloadable from a publically available website.[1] The simulator allows a simple setting of operating conditions, measurement noises, sampling time, and simulations of different faults, which makes it convenient to support the validation and demonstration of different multivariate statistics-based FD methods and FD statistics in the statistical framework.

Fig. 15 shows the proposed simulation platform for examining the performance of the considered statistics to detect multiplicative faults. It can be seen that these statistics are connected with three basic MDA methods: PCA, CCA and PLS respectively, and they together accomplish three kinds of tasks covered by multivariate statistics-based FD approaches, that is, the system overall operating performance monitoring, subsystem fault detection, and quality index monitoring. Of these, the operating performance of the whole system is monitored by implementing PCA-based methods built on all process and manipulated variable data. The performance of a subsystem, such as the reactor, is monitored by performing CCA-based methods on manipulated and process variables of the concerned subsystem as shown in Table 3. The key quality index reflected by components variables G and H is monitored using PLS-based methods developed by a mapping between process and manipulated variables and quality variables. Within the three methods, residual signals from MDA methods are sent to FD statistics to determine whether they behave well.

The operating conditions are: the sampling time is 3 min and the process is run for 100 h in the fault-free cases with 2,000 samples obtained for developing FD methods, and 1,000 h in faulty cases with faults occurring at 501 h. In practical applications, to evaluate the performance of different methods, the numerically estimated FDR index is often implemented as [53] :

$$FDR = \frac{Number\ of\ samples\ (J > J_{th} | faulty)}{total\ faulty\ samples} \tag{41}$$

This benchmark has 20 predefined process faults, of which # 8 to #12 are multiplicative faults. Take # 11 fault for example, we first perform PCA-based FD methods to detect whether this fault can affect the overall operating performance. The results are shown in the first row of Table 4. It is well-known that traditional PCA-based methods both use J_{T^2} and J_Q to respectively detect the two subspaces spanned by PCA. The results given by them are shown for comparison. Furthermore, the performance of $J_{T_n^2}$ and J_{Q_n} is recorded. It can be observed that they perform better than J_{T^2} and J_Q. Next, we replace J_{T^2} with other examined statistics. It is shown that $J_{\mathcal{L}}$ can completely detect this fault with $FDR = 1$. Others, except for $J_{\mathcal{K}}$, have also mildly improved the detection performance. For the purpose of real-time tracking the state of the reactor system, CCA-based FD methods can be developed using manipulated variables and measurement variables from the reactor. The detection results are shown in the second row of Table 4. Similar to the PCA case, using $J_{T_n^2}$ shows

[1] http://depts.washington.edu/control/LARRY/TE/download.html

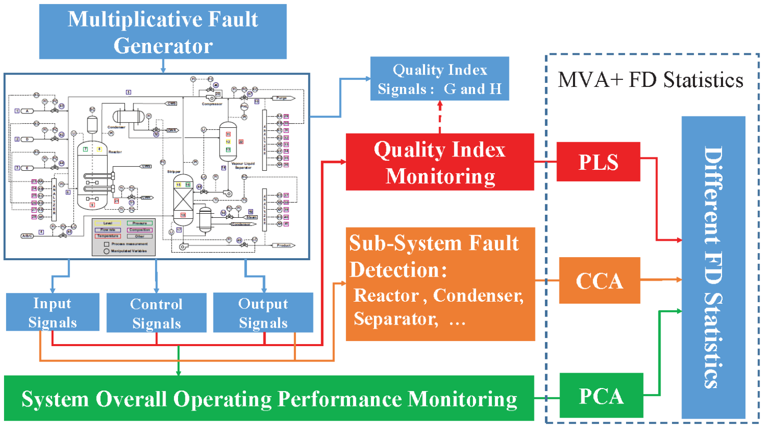

FIGURE 15. TE benchmark process test platform.

TABLE 3. Process and manipulated variables of TE process.

Block	Description	Variable tag	
		Process	Manipulated
Feeds	A feed (stream 1)	XMEAS(1)	XMV(3)
	D feed (stream 2)	XMEAS(2)	XMV(1)
	E feed (stream 3)	XMEAS(3)	XMV(2)
	A and C feed (stream 4)	XMEAS(4)	XMV(4)
Reactor	Reactor feed rate (stream 6)	XMEAS(6)	—
	Reactor pressure	XMEAS(7)	—
	Reactor level	XMEAS(8)	—
	Reactor temperature	XMEAS(9)	—
	Reactor cooling water outlet temperature	XMEAS(21)	—
	Reactor cooling water flow	—	XMV(10)
Condenser & Separator	Separator temperature	XMEAS(11)	—
	Separator level	XMEAS(12)	—
	Separator pressure	XMEAS(13)	—
	Separator underflow (stream 10)	XMEAS(14)	XMV(7)
	Condenser cooling water outlet temperature	XMEAS(22)	—
	Condenser cooling water flow	—	XMV(11)
Stripper	Stripper level	XMEAS(15)	—
	Stripper pressure	XMEAS(16)	—
	Stripper underflow (stream 11)	XMEAS(17)	XMV(8)
	Stripper temperature	XMEAS(18)	—
	Stripper steam flow	XMEAS(19)	XMV(9)
Miscellaneous	Agitator speed	—	XMV(12)
	Compressor work	XMEAS(20)	—
	Compressor recycle valve	—	XMV(5)
	Recycle flow (stream 8)	XMEAS(5)	—
	Purge rate (stream 9)	XMEAS(10)	—
	Purge valve (stream 9)	—	XMV(6)

TABLE 4. FDR performance of different FD methods for detecting #11 fault.

MDA	J_{T^2}	$J_{T^2_n}(n=8)$	J_Q	$J_{Q_n}(n=8)$	$J_{\mathcal{L}}(k_0=10)$	$J_{\mathcal{K}}(K=25)$
PCA	0.864	0.899	0.916	0.945	1.000	0.723
CCA	0.678	0.832	—	—	0.956	0.890
PLS	0.751	0.876	0.853	0.902	0.975	0.776

	$J_{\mathcal{D}}(n=10)$	$J_{\mathcal{T}}(n=10)$	$J_\gamma(n=10)$
	0.958	0.986	0.995
	0.342	0.972	0.837
	0.449	0.920	0.915

better monitoring performance in the presence of #11 fault. It can be further observed that $J_{\mathcal{L}}$ and $J_{\mathcal{T}}$ are significantly better compared with $J_{\mathcal{K}}$ and $J_{\mathcal{D}}$. Note that $J_{\mathcal{D}}$ performs poorly for this fault. When the performance of quality indices, namely components G and H in this process, is particularly concerned, PLS-based methods can be performed. Like PCA-based methods, J_{T^2} and J_Q are both used for monitoring quality-relevant and -irrelevant subspaces defined by PLS.

Their detection results are shown in the last row of Table 4. As well, comparisons with other statistics are made, where it can be found that, similar to the CCA case, $J_{\mathcal{L}}$ and $J_{\mathcal{T}}$ provide the best results and $J_{\mathcal{D}}$ is insensitive to this fault. Based on the above discussions, it could be concluded that the examined FD statistics are capable of handling multiplicative faults, of which $J_{\mathcal{L}}$ shows better improvements compared to the other statistics. The performance of Wishart distribution-based statistics would be enhanced if they could be used together.

VI. CONCLUSIONS

In this paper, fault detection statistics commonly encountered in multivariate statistics-based FD field including J_{T^2}, J_Q as well as their extensions $J_{T_n^2}$ and J_{Q_n} and statistical local method, $J_{\mathcal{L}}$; Wishart distribution-based methods: $J_{\mathcal{Y}}$, $J_{\mathcal{T}}$ and $J_{\mathcal{D}}$; and the information theory-based methods: $J_{\mathcal{C}}$ and $J_{\mathcal{K}}$ were investigated for use in detecting multiplicative faults. The different statistics were compared and their interrelationships were demonstrated. The detection performance for multiplicative faults was evaluated using the FDR index. The results show that:

- $J_{T_n^2}$ and J_{Q_n} can improve the performance of J_{T^2} and J_Q for detecting some multiplicative faults.
- $J_{\mathcal{L}}$ converts the detection of multiplicative faults to the form of detecting additive faults, which can improve the detectability. It was also found that $J_{\mathcal{L}}$ behaves well for faults that decrease the variance.
- The Wishart distribution-based methods are developed by applying different metrics, such as trace and determinant, to measure the Wishart matrix. Using these statistics together would improve the detection performance.
- In the case that a large number of online samples is used, the formulae for $J_{\mathcal{C}}$ and $J_{\mathcal{K}}$ are equivalent.

These results imply that the correct statistic depends on the problem formulation, as well as the available information.

Future work seeks to analyse a boarder range of fault detection statistics, *e.g.* data description and dissimilarity analysis-based methods, to handle multiplicative faults and investigate in depth how they can, not only improve the detectability, but also provide an acceptable EDD. As well, it is suggested that additional work will be done in understanding and using more advanced, such as machine learning, which although actively used in FD field, are yet not well studied.

REFERENCES

[1] S. J. Qin, "Statistical process monitoring: Basics and beyond," *J. Chemometrics*, vol. 17, nos. 8–9, pp. 480–502, 2003.

[2] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.

[3] J. E. Jackson, *A User Guide to Principal Components*. New York, NY, USA: Wiley, 1991.

[4] J. Kresta, J. F. MacGregor, and T. E. Marlin, "Multivariate statistical monitoring of process operating performance," *Can. J. Chem. Eng.*, vol. 69, no. 1, pp. 35–47, 1991.

[5] J. F. MacGregor, C. Jaeckle, C. C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE J.*, vol. 40, no. 5, pp. 826–838, 1994.

[6] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 12–24, 2008.

[7] Q. P. He and J. Wang, "Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes," *AIChE J.*, vol. 57, no. 1, pp. 107–121, 2011.

[8] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Eng. Pract.*, vol. 46, pp. 51–58, Jan. 2016.

[9] S. X. Ding, *Data Driven Design of Fault Diagnosis and Fault Tolerant Control Systems*. London, U.K.: Springer, 2014.

[10] Y. Liu, L. Bin, X. Zhao, and X. Min, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018.

[11] K. Zhang, H. Hao, Z. Chen, S. X. Ding, and K. Peng, "A comparison and evaluation of key performance indicator-based multivariate statistics process monitoring approaches," *J. Process Control*, vol. 33, pp. 112–126, Sep. 2015.

[12] K. Zhang, *Performance Assessment for Process Monitoring and Fault Detection Methods*. Wiesbaden, Germany: Springer Nature, 2016.

[13] L. H. Chiang and R. D. Braatz, "Process monitoring using causal map and multivariate statistics: Fault detection and identification," *Chemometrics Intell. Lab. Syst.*, vol. 65, no. 2, pp. 159–178, 2003.

[14] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.

[15] Y. Liu, Y. Pan, Z. Sun, and D. Huang, "Statistical monitoring of wastewater treatment plants using variational Bayesian PCA," *Ind. Eng. Chem. Res.*, vol. 58, no. 8, pp. 3272–3282, 2014.

[16] Q. Jiang and X. Yan, "Plant-wide process monitoring based on mutual information-multiblock principal component analysis," *ISA Trans.*, vol. 53, no. 5, pp. 1516–1527, Sep. 2014.

[17] H. Wold, "Path models with latent variables: The NIPALS approach. Quantitative sociology," *International Perspectives on Mathematical and Statistical Modeling*. New York, NY, USA: Academic, 1975.

[18] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, no. 3, pp. 341–349, 1979.

[19] S. J. Qin and Y. Y. Zheng, "Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures," *AIChE J.*, vol. 59, no. 2, pp. 496–504, Feb. 2013.

[20] D. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE J.*, vol. 56, no. 1, pp. 168–178, 2010.

[21] H. Y. Hao, K. Zhang, S. X. Ding, Z. W. Chen, and Y. G. Lei, "A data-driven multiplicative fault diagnosis approach for automation processes," *ISA Trans.*, vol. 53, no. 3, pp. 1436–1445, 2014.

[22] H. P. Huang, C. C. Li, and J. C. Jeng, "Multiple multiplicative fault diagnosis for dynamic processes via parameter similarity measures," *Ind. Eng. Chem. Res.*, vol. 46, no. 13, pp. 4517–4530, 2007.

[23] K. Zhang, S. X. Ding, Y. A. W. Shardt, Z. Chen, and K. Peng, "Assessment of T_2- and Q-statistics for detecting additive and multiplicative faults in multivariate statistical process monitoring," *J. Franklin Inst.*, vol. 354, no. 2, pp. 668–688, 2016.

[24] N. A. Adnana, I. Izadi, and T. Chen, "On expected detection delays for alarm systems with deadbands and delay-timers," *J. Process Control*, vol. 9, no. 9, pp. 1318–1331, 2011.

[25] B. Arthur, K. J. D. Lin, and R. N. McGrath, "Multivariate control charts for monitoring covariance matrix: A review," *Qual. Technol. Quant. Manage.*, vol. 3, no. 4, pp. 415–436, 2006.

[26] C. Hajiyev, "Tracy—Widom distribution based fault detection approach: Application to aircraft sensor/actuator fault detection," *ISA Trans.*, vol. 51, no. 1, pp. 189–197, 2012.

[27] C. A. Tracy and H. Widom, "The distribution of the largest eigenvalue in the gaussian ensembles: ß =1, 2, 4," in *Calogero—Moser—Sutherland Models*, J. F. van Diejen and L. Vinet, Eds. New York, NY, USA: Springer, 2000, pp. 461–472.

[28] S. Kourouklis and P. G. Moschopoulos, "On the distribution of the trace of a noncentral Wishart matrix," *Metron*, vol. 43, no. 2, pp. 85–92, 1985.

[29] N. R. Goodman, "The distribution of the determinant of a complex Wishart distributed matrix," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 178–180, 1963.

[30] M. Basseville, *Detection of Abrupt Changes: Theory and Application*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[31] U. Kruger, S. Kumar, and T. Littler, "Improved principal component monitoring using the local approach," *Automatica*, vol. 43, no. 9, pp. 1532–1542, 2007.
[32] U. Kruger and G. Dimitriadis, "Diagnosis of process faults in chemical systems using a local partial least squares approach," *AIChE J.*, vol. 54, no. 10, pp. 2581–2596, 2008.
[33] X. Wang, U. Kruger, G. W. Irwin, G. McCullough, and N. McDowell, "Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 1, pp. 122–129, Jan. 2008.
[34] Z. Ge, C. Yang, and Z. Song, "Improved kernel PCA-based monitoring approach for nonlinear processes," *Chem. Eng. Sci.*, vol. 64, no. 9, pp. 2245–2255, May 2009.
[35] Z. Chen, K. Zhang, S. X. Ding, Y. A. W. Shardt, and Z. Hu, "Improved canonical correlation analysis-based fault detection methods for industrial processes," *J. Process Control*, vol. 41, pp. 26–34, May 2016.
[36] J. Zeng, U. Kruger, J. Geluk, X. Wang, and L. Xie, "Detecting abnormal situations using the Kullback–Leibler divergence," *Automatica*, vol. 50, no. 11, pp. 2777–2786, 2014.
[37] L. Xie, J. Zeng, U. Kruger, X. Wang, and J. Geluk, "Fault detection in dynamic systems using the Kullback–Leibler divergence," *Control Eng. Pract.*, vol. 43, pp. 39–48, Oct. 2015.
[38] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part I," *Singal Process*, vol. 109, pp. 334–344, Jan. 2015.
[39] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part II," *Singal Process*, vol. 109, pp. 278–287, Apr. 2015.
[40] J. Guerrero, "Multivariate mutual information sampling distribution with applications," *Commun. Stat. Theory Methods*, vol. 23, no. 5, pp. 1319–1339, 1994.
[41] F. Harrou, Y. Sun, and M. Madakyaru, "Kullback–Leibler distance-based enhanced detection of incipient anomalies," *J. Loss Prevention Process Ind.*, vol. 44, pp. 73–87, Nov. 2016.
[42] C. R. Rao, *Linear Statistical Inference and its Applications*. Hoboken, NJ, USA: Wiley, 1965.
[43] K. Zhang, Y. A. W. Shardt, Z. Chen, S. X. Ding, and K. Peng, "A brief survey of different statistics for detecting multiplicative faults in multivariate statistical process monitoring," in *Proc. IEEE CDC*, Las Vegas, NV, USA, Dec. 2016, pp. 2152–2157.
[44] C. Zhao, F. Wang, and M. Jia, "Dissimilarity analysis based batch process monitoring using moving windows," *AIChE J.*, vol. 53, no. 5, pp. 1267–1277, 2007.
[45] C. H. Zhao and F. R. Gao, "A sparse dissimilarity analysis algorithm for incipient fault isolation with no priori fault information," *Control Eng. Pract.*, vol. 65, pp. 70–82, Aug. 2017.
[46] Y. Ma, H. Shi, H. Ma, and M. Wang, "Dynamic process monitoring using adaptive local outlier factor," *Chemometrics Intell. Lab. Syst.*, vol. 127, pp. 89–101, Aug. 2013.
[47] Y. Wang and C. Zhao, "Probabilistic fault diagnosis method based on the combination of nest-loop fisher discriminant analysis and analysis of relative changes," *Control Eng. Pract.*, vol. 68, pp. 32–45, Nov. 2017.
[48] H. Liu, Y. Tang, and H. H. Zhang, "A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables," *Comput. Statist. Data Anal.*, vol. 53, no. 4, pp. 853–856, 2009.
[49] M. M. Rashid and J. Yu, "A new dissimilarity method integrating multidimensional mutual information and independent component analysis for non-Gaussian dynamic process monitoring," *Chemometrics Intell. Lab. Syst.*, vol. 115, pp. 44–58, Jun. 2012.
[50] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
[51] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Eng. Pract.*, vol. 62, pp. 11–21, May 2017.
[52] N. A. Adnana, Y. Cheng, I. Izadib, and T. W. Chen, "Study of generalized delay-timers in alarm configuration," *J. Process Control*, vol. 23, no. 3, pp. 382–395, 2013.
[53] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

KAI ZHANG received the B.E. degree from Shandong University in 2009, the M.E. degree from the University of Science and Technology Beijing in 2012, and the Ph.D. degree from the Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, in 2016. He currently holds a post-doctoral aposition with the University of Science and Technology Beijing. His research interest covers process monitoring and fault diagnosis for industrial processes.

KAIXIANG PENG (M'15) received the B.E. degree in automation and the M.E. and Ph.D. degrees from the Research Institute of Automatic Control, University of Science and Technology Beijing, Beijing, China, in 1995, 2002, and 2007, respectively. He is currently a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests are fault diagnosis, prognosis, and maintenance of complex industrial processes, modeling and control for complex industrial processes, and control system design for the rolling process.

YURI A. W. SHARDT (M'03) received the B.Sc. degree in chemical engineering and the Ph.D. degree in process engineering from the University of Alberta, Canada, in 2008 and 2012, respectively. After his doctorate, he worked in industry for a year, focusing on the development and application of control methods to novel chemical processes. He was an Assistant Professor with the University of Waterloo, Canada. He is currently a Professor and the Chair of the Department of Automation Engineering (Fachgebiet Automatisierungstechnik), Technical University of Ilmenau, Germany. His research interests include soft sensors, system identification, fault management, and holistic control. He was an Alexander von Humboldt Fellow with the University of Duisburg-Essen, Germany.