

A Review of Statistical-Based Fault Detection and Diagnosis with Probabilistic Models

Yanting Zhu ¹ , Shunyi Zhao ¹ , Yuxuan Zhang ², Chengxi Zhang ^{1,*}  and Jin Wu ³ 

¹ Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China; 7231905006@stu.jiangnan.edu.cn (Y.Z.); shunyi@jiangnan.edu.cn (S.Z.)

² Department of Management Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; yuxuan.zhang@nuaa.edu.cn

³ Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong 999077, China; jwucp@connect.ust.hk

* Correspondence: dongfangxy@163.com

Abstract: As industrial processes grow increasingly complex, fault identification becomes challenging, and even minor errors can significantly impact both productivity and system safety. Fault detection and diagnosis (FDD) has emerged as a crucial strategy for maintaining system reliability and safety through condition monitoring and abnormality recovery to manage this challenge. Statistical-based FDD methods that rely on large-scale process data and their features have been developed for detecting faults. This paper overviews recent investigations and developments in statistical-based FDD methods, focusing on probabilistic models. The theoretical background of these models is presented, including Bayesian learning and maximum likelihood. We then discuss various techniques and methodologies, e.g., probabilistic principal component analysis (PPCA), probabilistic partial least squares (PPLS), probabilistic independent component analysis (PICA), probabilistic canonical correlation analysis (PCCA), and probabilistic Fisher discriminant analysis (PFDA). Several test statistics are analyzed to evaluate the discussed methods. In industrial processes, these methods require complex matrix operation and cost computational load. Finally, we discuss the current challenges and future trends in FDD.

Keywords: statistical framework; fault detection and diagnosis; probabilistic models



Citation: Zhu, Y.; Zhao, S.; Zhang, Y.; Zhang, C.; Wu, J. A Review of Statistical-Based Fault Detection and Diagnosis with Probabilistic Models. *Symmetry* **2024**, *16*, 455. <https://doi.org/10.3390/sym16040455>

Academic Editors: Sergei D. Odintsov and Marek T. Malinowski

Received: 3 January 2024

Revised: 8 March 2024

Accepted: 11 March 2024

Published: 8 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern industry has brought about more complex and high-dimensional industrial processes. There is less tolerance for potential safety hazards, which means performance degradation and productivity drawdown. FDD is a significant task to ensure product quality and process reliability in modern industrial systems. Traditional FDD methods are based on experiences and have met challenges with the expansion of plant scale and large numbers of process variables. Methods based on statistical analysis become a trend in industry applications. Recently, the probabilistic model based on statistical methods broadened the industrial application in cases of high dimensionality, non-Gaussian distribution, nonlinear relationships, and time-varying variables. This article aims to overview the statistical analysis of FDD methods, especially under the probabilistic framework.

1.1. Background

Compared to passive fault-tolerant control, taking fault into account as a system perturbation, FDD is an active strategy for detecting and identifying potential abnormalities and faults, providing early warning, and recommending corrective actions to prevent failure occurrence. Compared to prognostics, dealing with fault prediction before it occurs, diagnostics is a posterior event analysis and it is required after occurring a fault. In harsh working environments, such as extreme temperatures, high pressure, and underwater,

sensors are prone to faults, while the sensor is an essential component of data acquisition systems, sensor faults, including incipient failure and abrupt failure, will affect the accuracy, stability, and reliability. It has become essential for industrial applications, especially for engineering systems such as mechanical engineering [1–4], electric vehicle dynamics [5,6], power electronic systems [7–10], electric machines [11–14], and wind energy conversion systems [15–18]. The task of FDD is to spot process abnormalities promptly and identify their early causes [19]. The elements of a general structure for fault diagnosis system and control system are shown in Figure 1. It demonstrates different components in a control loop, and failures could exist in actuators, dynamic plants, sensors, and feedback controllers.

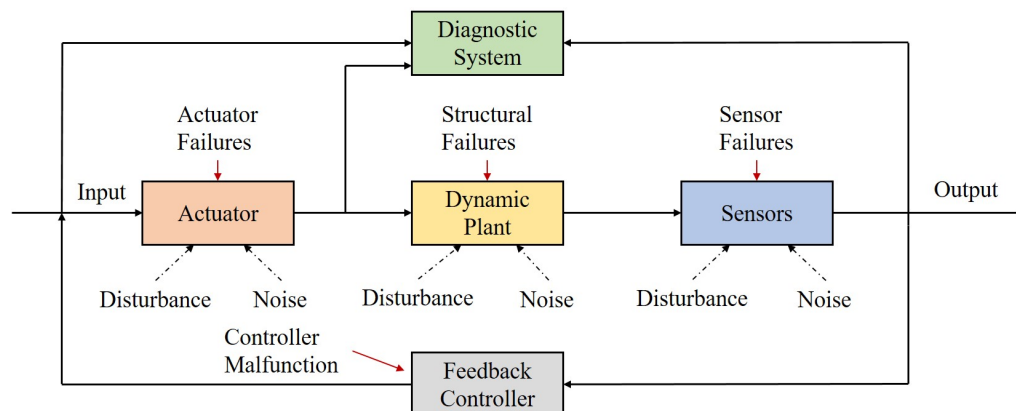


Figure 1. A fault-diagnosis system is targeted for detecting failures from collected information and improving robustness and accuracy.

1.2. Evolution of Fault Detection and Diagnosis

Generally, approaches to detecting and diagnosing faults can be divided into three categories, as shown in Figure 2.

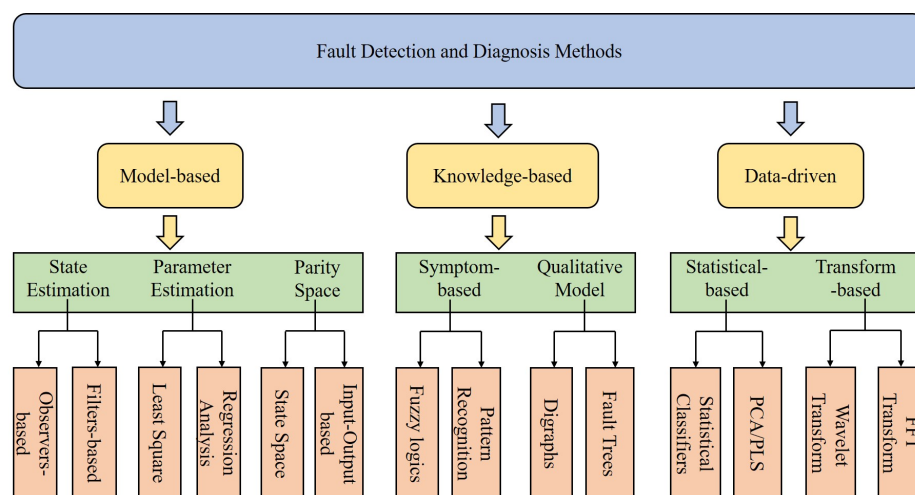


Figure 2. Classification of fault diagnosis methods.

1.2.1. Model-Based Methods

Model-based methods include state estimation, parameter estimation, and parity space. The model-based methods utilize physical and mathematical knowledge and they possess the explainability for making decisions in a transparent way rather than in a black box [20–22]. However, the noisy operation environment hinders physics-based modeling and degrades the accuracy of complex dynamic modeling [23,24].

1.2.2. Knowledge-Based Methods

Knowledge-based methods include symptom-based methods and qualitative methods. The knowledge-based methods are usually implemented by experts, and the fault diagnosis relies on the accumulation of prior information and logical reference [25]. These methods are efficient within the scope of existing knowledge and struggle to tackle unexpected failures.

1.2.3. Data-Driven Methods

Data-driven methods include statistical-based methods and transform-based methods. Compared to model-based methods and knowledge-based methods, data-driven approaches are efficient when confronted with high-dimensional data. They require a sufficient quantity of data and enhance accuracy by extracting information or features from large-scale datasets [26,27]. Yet the lack of an accurate mathematical model makes it inadequate for detecting incipient faults.

The Internet of Things era revolutionizes industrial processes by collecting a large amount of information via a network of terminals, leading to a data explosion and the rocketing complexity of model construction. Troubleshooting highly complex systems involves multiple processes and multiple anomalies, making conventional physics-based deterministic modeling more challenging. Moreover, implementing the knowledge-based FDD methods relies heavily on expertise or prior information, which is time-consuming and labor-intensive especially when dealing with a high-dimensional process. Data-driven approaches are targeted for addressing large-scale data and extracting features from data.

The statistical approach is a branch of data-driven schemes, and extracts the process information from measurement or observation. Its significant advantage is that it can tackle many highly correlated variables without complex mathematical forms or costly design efforts. Statistical methods [28] have gained popularity in practical applications for their ability to directly analyze input and output data, particularly PCA [29–33] and PLS [34–37]. PCA, PLS, and independent component analysis (ICA) are traditional statistical analysis methods based on deterministic models. The underlying theoretical foundation of traditional multivariate analysis is linear algebra.

In practice, industrial process data are polluted by missing data and outliers, which can significantly influence the accuracy of features and control thresholds. The probabilistic extensions of traditional statistical methods employ distributions to describe states to enhance their ability to process sampling data with disturbances, outliers, and missing values. In addition, the probabilistic form of statistical analysis can employ non-linear data and thus can be applied in industrial processes. Recently, the probabilistic counterparts of PCA [38] and factor analysis (FA) [39] have been generated for fault diagnosis. Furthermore, the extensions of their mixture form with multiple operation modes have been generalized [40].

The study on statistical-based fault diagnosis methods draws remarkable research attention. Due to their similarity in collecting, processing, and extracting information from data, these strategies are viewed as statistical-based methods. In this context, these schemes are mainly characterized as follows:

- (1) Without a complex model construction, a statistical-based FDD design can extract the information and make decisions directly on the sampling data.
- (2) These strategies are designed to address FDD in static or dynamic systems in a stable state with the flexible application of statistical tests and their mixed indices.

1.3. Motivation and Contribution

Statistical-based fault diagnosis has attracted attention in industrial applications and the academic community. The sensor technology gives rise to a data explosion, and the data quality significantly impacts the modeling of the process and thus influences the performance of fault diagnosis. Probabilistic extensions of conventional statistical methods spring up due to their robustness and advantages in treating outliers, disturbance, and missing values. There is no comprehensive review of statistical FDD methods under a

probabilistic framework. Therefore, statistical methods with a probabilistic model are an unavoidable element that needs to be addressed in industrial applications and the academic community. Different from other reviews on FDD [41–43], this review focuses on detailed explanations to let readers understand the principle of each method and save some time searching a lot of references. The purpose of this review is to provide theoretical background and recent application instances of probabilistic-based statistical fault diagnosis.

1.4. Organization of This Paper

The remainder of this paper is organized as follows. The theoretical background of the probabilistic model is presented in Section 2. Section 3 gives a brief overview of the probabilistic extensions of statistical methods and their practical applications. The challenges and perspectives are demonstrated in Section 4. Conclusions are finally drawn in Section 5.

2. Theoretic Background

Maximum likelihood estimation (MLE) and Bayesian theory are employed when the probabilistic model is introduced into statistical methods. This section will briefly introduce the principles of MLE and Bayesian inference.

2.1. Maximum Likelihood Estimation

In statistics, MLE estimates the parameters of an assumed probability distribution with observation measurement and maximizes the probability [44]. MLE is generated to find the probability density function (PDF) that is most likely to produce a data sample given the observation. The data $\mathbf{y} = (y_1, y_2, \dots, y_m)$ is a random sample from an unknown population. In practice, the model usually involves abundant parameters, and the likelihood function is probably nonlinear, making it difficult to obtain an analytic solution. A nonlinear model is established to estimate the remaining useful life of a system, where the unknown parameters are estimated with the help of MLE [45].

The $\mathbf{v} = (v_1, v_2, \dots, v_k)$ is a vector defined on a multi-dimensional parameter space. $p(\mathbf{y}|\mathbf{v})$ denotes the probability of y given v . The likelihood function is defined by [46]

$$L(\mathbf{v}|\mathbf{y}) = p(\mathbf{y}|\mathbf{v}). \quad (1)$$

The MLE estimate is obtained by maximizing the log-likelihood function. Assuming that the log-likelihood $\ln L(\mathbf{v}|\mathbf{y})$ is differentiable, if $\mathbf{v}_{MLE}$ exists, it must satisfy Equation (2):

$$\frac{\partial \log L(\mathbf{v}|\mathbf{y})}{\partial v_i} = 0. \quad (2)$$

$L(\mathbf{v}|\mathbf{y})$ and $p(\mathbf{y}|\mathbf{v})$ are defined on different axes, that $p(\mathbf{y}|\mathbf{v})$ is a function of the data given parameters, defined on the data scale, and $L(\mathbf{v}|\mathbf{y})$ is defined on the parameter scale.

2.2. Bayesian Learning

Bayesian learning provides a rigorous framework for complex nonlinear systems whose internal state variables are inaccessible to direct measurement. Given a general discrete-time state estimation equation and measurement function,

$$\mathbf{x}_k = g(\mathbf{x}_{k-1}) + \mathbf{e}_k, \quad (3)$$

$$\mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{v}_k, \quad (4)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state at time step k , $\mathbf{e}_k \in \mathbb{R}^n$ is the process noise, and $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes the transition function. $\mathbf{y}_k \in \mathbb{R}^p$ is the measurement, $\mathbf{v}_k \in \mathbb{R}^p$ is the measurement noise, and $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$. $\mathbf{e}_k$ and $\mathbf{v}_k$ are independent.

The Bayesian learning is to recursively estimate the PDF of $\mathbf{x}_k$ given measurements $\mathbf{y}_k$. The initial density is determined beforehand, and $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ denotes the transition

probability density. The inference of the state $\mathbf{x}_k$ relies on the marginal density $p(\mathbf{x}_k|\mathbf{y}_{1:k})$. The predictive function of $\mathbf{x}_k$ at step k is estimated by [47]

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}. \quad (5)$$

Then, the marginal filtering density is computed

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})}, \quad (6)$$

where $p(\mathbf{y}_k|\mathbf{y}_{1:k-1})$ is the normalizing parameter.

A Bayesian network is a probabilistic graphical model that illustrates the relationships between variables.

$$p(\mathbf{a}, \mathbf{b}) = p(\mathbf{a}|\mathbf{b}) \cdot p(\mathbf{b}), \quad (7)$$

and the joint probability distribution for a Bayesian network with nodes $\mathbf{a} = \{a_1, \dots, a_n\}$ is given by

$$p(\mathbf{a}) = \prod_{i=1}^n p(a_i|\text{parents}(a_i)), \quad (8)$$

where $\text{parents}(a_i)$ is the parent set of node a_i .

Research based on Bayesian learning has attracted huge attention in the field of fault diagnosis. Zhao proposed advanced Bayesian estimation algorithms to monitor the faulty sensors [48], then improved algorithms for online ability and correlated signals in nonlinear processes, respectively, [49,50]. Bayesian network is adopted in fault diagnosis [51–61]. Multivariate statistical analysis has been combined with Bayesian inference for fault detection and isolation [62]. A Bayesian maximum likelihood classifier is validated as accurate for induction machine and stator short circuit fault diagnosis [63]. The probabilistic Bayesian deep learning framework exploits the risk-aware model to identify unknown faults and enhance the trustworthiness of the diagnostic results [64,65].

3. Probabilistic Statistical-Based Approaches

This section discusses different kinds of static statistical-based approaches. Four probabilistic models applied in the field of FDD are illustrated, including probabilistic PCA, probabilistic PLS, probabilistic ICA, probabilistic canonical correlation analysis (CCA), and probabilistic Fisher discriminant analysis (FDA). PCA extracts the principal components that are retained to explain the majority of the variability in the data by maximizing the variance. Compared to PCA, the FDA maximizes the separation among classes while minimizing the separation between classes. The components after PCA decomposition are orthogonal and therefore irrelevant, but independence is not guaranteed. Compared to PCA, ICA can find the original components in the observed mixtures, and it is a linear transformation of the data in the original feature space. PCA involves only one set of variables, while CCA extends to the interdependence between two sets of variables, measuring the correlation between the two sets of variables. The probabilistic approaches of traditional statistical analysis are shown in Figure 3. They differ in the variable distribution, the application scenario, and whether the dataset is labeled. PPCA and PICA have the same characteristics, they can deal with non-Gaussian distribution data and stationary processes. PFDA and PPLS are supervised methods. PCCA excels in other methods of tackling dynamic processes. The probabilistic extensions of traditional multivariate statistical analysis still retain original characteristics and broaden the range of applications.

Properties for Different Approaches	Methods				
	PPCA	PICA	PPLS	PCVA	PFDA
Non-Gaussian Distributed Variables	*	*			*
Gaussian Distributed Variables	*	*	*	*	*
Stationary Process	*	*	*	*	*
Dynamic Issues				*	
Supervised Learning			*		*
Unsupervised Learning	*	*		*	

Figure 3. Probabilistic approaches for processing monitoring.

3.1. Probabilistic Principal Component Analysis

PCA is a technique targeted for dimensionality reduction, and it has wide applications, including data compression, image processing, data analysis, and pattern recognition [66–69]. The probabilistic derivation of PCA is given by

$$\mathbf{y} = \mathbf{W}\mathbf{t} + \mathbf{m} + \boldsymbol{\epsilon}, \quad (9)$$

where $\mathbf{y} \in \mathbb{R}^d$, independent unobservable variable $\mathbf{t} \in \mathbb{R}^q \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $q < d$. The transition matrix is $\mathbf{W} \in \mathbb{R}^{d \times q}$, and the vector is $\mathbf{m} \neq \mathbf{0}$. The key assumption for PPCA is that the noise in this probability model is likewise Gaussian $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and the covariance $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}$ is constrained to be a diagonal matrix, so that $\mathbf{y}$ are conditionally independent given the values of $\mathbf{t}$. The conditional PDF of $\mathbf{y}$ and the marginal PDF of $\mathbf{t}$ obtained by integration are implied by

$$\mathbf{y}|\mathbf{t} \sim \mathcal{N}(\mathbf{W}\mathbf{t} + \mathbf{m}, \sigma^2 \mathbf{I}), \quad (10)$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{m}, \mathbf{A}), \quad (11)$$

where $\mathbf{A} = \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}$, then the log-likelihood function is

$$\mathcal{L} = -\frac{N}{2} \{\ln |\mathbf{A}| + \text{tr}(\mathbf{A}^{-1} \mathbf{S}) + d \ln(2\pi)\}, \quad (12)$$

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_n - \mathbf{m})(\mathbf{y}_n - \mathbf{m})^T. \quad (13)$$

Estimates for $\mathbf{W}$ and σ^2 is obtained by iteratively maximizing Equation (12) by employing the EM algorithm [70].

$$\mathbf{W}_{ML} = \mathbf{U}_q (\boldsymbol{\Lambda}_q - \sigma^2 \mathbf{I})^{\frac{1}{2}} \mathbf{R}, \quad (14)$$

$$\sigma_{ML}^2 = \frac{1}{d-q} \sum_{j=q+1}^d \lambda_j, \quad (15)$$

where the first q th column vectors in $\mathbf{U}_q \in \mathbb{R}^{d \times q}$ are the q th principal eigenvectors of $\mathbf{S}$, with corresponding eigenvalues $\lambda_1, \dots, \lambda_q$ in the diagonal matrix $\boldsymbol{\Lambda}_q \in \mathbb{R}^{q \times q}$, and $\mathbf{R} \in \mathbb{R}^{q \times q}$ is an arbitrary orthogonal matrix. Note that $\mathbf{U}_q$ and $\boldsymbol{\Lambda}_q$ can be obtained by performing the singular value decomposition on $\mathbf{S}$.

The conditional distribution of the latent variable $\mathbf{t}$ given the observed $\mathbf{y}$ is obtained with the help of Bayesian inference

$$\mathbf{t}|\mathbf{y} \sim \mathcal{N}(\mathbf{B}^{-1} \mathbf{W}^T (\mathbf{y} - \mathbf{m}), \sigma^2 \mathbf{B}^{-1}), \quad (16)$$

where $\mathbf{B} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}$. From Equation (16), the point-wise technique can be employed when the conditional distribution is generated

$$\langle \mathbf{t} | \mathbf{y} \rangle = \mathbf{B}^{-1} \mathbf{W}_{ML}^T (\mathbf{y} - \mathbf{m}). \quad (17)$$

Then, the high-dimensional observed data $\mathbf{y}$ can be condensed into a new distribution $\mathbf{t}$ which satisfies Gaussian.

The probabilistic PCA model has abundant modifications and extensions to be applied for fault diagnosis. Choi et al. proposed a fault detection scheme based on a maximum-likelihood PCA mixture model [40]. To address the challenge of separating several factors that together cause a failure to occur, probabilistic PCA was created [71]. An aligned mixture probabilistic PCA is proposed by Yang for fault detection of multimode chemical processes [72]. Then, a reconstruction-based multivariate contribution analysis is applied to the PPCA mixture model for fault isolation [73].

The robust version of probabilistic PCA is modified to deal with outliers and missing data during the modeling stage [74]. In addition, a variational inference process based on the Bayesian PCA model structure provides the foundation of a defect reconstruction method [75]. Additionally, the hidden Markov model framework temporally extends the static mixture probabilistic PCA model-based classifier to the dynamic form [76,77]. It was suggested to use a hybrid framework that takes into account moving window PCA and Bayesian networks to cope with barely accessible data in a fault state [78].

By thoroughly analyzing the principle and implementation of probabilistic PCA, the significant advantages can be summarized as follows.

- (1) Enhanced Robustness: In practical applications, disturbances are unavoidable in a complex working environment. Probabilistic PCA disposes of the problem that sampling data are mixed with outliers and missing values by using distribution modeling of these data and enhancing robustness.
- (2) Increased Complex Data: The introduced latent variables enable probabilistic PCA to process non-linear data, improving the performance of dimensional reduction.
- (3) Probability Inference: Probabilistic PCA is a dimensionality reduction method based on probability models. It provides quantitative information on uncertainty and probabilistic inferences to obtain more accuracy and effectiveness. Ultimately, the ability to interpret data is substantially intensified.

3.2. Probabilistic Partial Least Squares

The core of the probabilistic PLS model is to use a part of the latent variables to explain the observed data set. The probabilistic PLS model is formulated by [79]

$$\mathbf{x} = \mathbf{m}_x + \mathbf{J} \mathbf{t}^s + \mathbf{Q} \mathbf{t}^b + \epsilon_x, \quad (18)$$

$$\mathbf{y} = \mathbf{m}_y + \mathbf{K} \mathbf{t}^s + \epsilon_y, \quad (19)$$

where $\mathbf{J} \in \mathbb{R}^{m \times q_s}$, $\mathbf{K} \in \mathbb{R}^{r \times q_s}$ and $\mathbf{Q} \in \mathbb{R}^{m \times q_b}$; $\mathbf{t}^s \in \mathbb{R}^{q_s \times 1}$, $\mathbf{t}^b \in \mathbb{R}^{q_b \times 1}$; $\mathbf{m}_x$ and $\mathbf{m}_y$ are the mean of $\mathbf{x}$ and $\mathbf{y}$; and ϵ_x and ϵ_y denotes measurement noises of $\mathbf{x}$ and $\mathbf{y}$, respectively.

In the probabilistic PLS model, $\mathbf{t}^s \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{t}^b \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\epsilon_x \sim \mathcal{N}(\mathbf{0}, \Sigma_x)$, and $\epsilon_y \sim \mathcal{N}(\mathbf{0}, \Sigma_y)$. Different from the probabilistic PCA model that assumes the error covariance matrix to be a diagonal matrix with a constant value, different noise variances have been assumed for different variables $\Sigma_x = \text{diag}\{\sigma_{x,u}^2\}_{u=1,2,\dots,m}$ and $\Sigma_y = \text{diag}\{\sigma_{y,v}^2\}_{v=1,2,\dots,r}$. Given data sets $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T \in \mathbb{R}^{n \times m}$ and $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T \in \mathbb{R}^{n \times r}$, the log-likelihood function can be determined by

$$\begin{aligned} & L(\mathbf{X}, \mathbf{Y} | \mathbf{m}_x, \mathbf{m}_y, \mathbf{J}, \mathbf{Q}, \mathbf{K}, \Sigma_x, \Sigma_y) \\ &= \ln \prod_{i=1}^n p(\mathbf{x}_i, \mathbf{y}_i | \mathbf{m}_x, \mathbf{m}_y, \mathbf{J}, \mathbf{Q}, \mathbf{K}, \Sigma_x, \Sigma_y). \end{aligned} \quad (20)$$

The optimal values of parameters are determined by the EM algorithm:

$$\mathbf{J}^{new} = \left[\sum_{i=1}^n \mathbf{x}_i E(\mathbf{t}_i^s | [\mathbf{x}_i])^T \right] \left[\sum_{i=1}^n E(\mathbf{t}_i^s \mathbf{t}_i^{sT} | [\mathbf{x}_i]) \right]^{-1}, \quad (21)$$

$$\mathbf{Q}^{new} = \left[\sum_{i=1}^n \mathbf{x}_i E(\mathbf{t}_i^b | [\mathbf{x}_i])^T \right] \left[\sum_{i=1}^n E(\mathbf{t}_i^b \mathbf{t}_i^{bT} | [\mathbf{x}_i]) \right]^{-1}, \quad (22)$$

$$\mathbf{K}^{new} = \left[\sum_{i=1}^n \mathbf{y}_i E(\mathbf{t}_i^s | [\mathbf{y}_i])^T \right] \left[\sum_{i=1}^n E(\mathbf{t}_i^s \mathbf{t}_i^{sT} | [\mathbf{y}_i]) \right]^{-1}, \quad (23)$$

$$\Sigma_x^{new} = \frac{1}{n} \text{diag} \left\{ \sum_{i=1}^n [\mathbf{x}_i \mathbf{x}_i^T - \mathbf{J}^{new} \mathbf{Q}^{new} E(\mathbf{t}_i^s | [\mathbf{x}_i]) \mathbf{y}_i^T] \right\}, \quad (24)$$

$$\Sigma_y^{new} = \frac{1}{n} \text{diag} \left\{ \sum_{i=1}^n [\mathbf{y}_i \mathbf{y}_i^T - \mathbf{K}^{new} E(\mathbf{t}_i^s | [\mathbf{y}_i]) \mathbf{y}_i^T] \right\}. \quad (25)$$

The supervised model PPLS builds a regression model between two sets of variables. For further applications, the probabilistic PLS model has been modified. On the basis of this, the validity of the classification of an unknown item is assessed [80]. Zheng adapted the probabilistic PLS model to the semi-supervised version for the creation of soft sensors [81]. Data-driven fault identification and diagnosis techniques are proposed based on a novel locally weighted probabilistic kernel PLS [82]. Botella described an improvement to discriminant partial least squares that use the kernel trick and Bayes rule to implement data classification [83]. To further decompose the PPLS model, a concurrent probabilistic PLS approach is suggested, and monitoring statistics are created for assessment [84].

Compared with traditional PLS, probabilistic PLS treats independent and dependent variables as random variables and assumes that they satisfy the Gaussian distribution. The probabilistic model is capable of dealing with disturbance, outliers, and missing data and then improving the stability and prediction accuracy of the model.

3.3. Probabilistic Independent Component Analysis

ICA separates the dataset into linear combinations of statistically independent non-Gaussian sources. It is a significant application of the blind source separation method. The probabilistic ICA model is [85]

$$\mathbf{x}_n = \mathbf{A} \mathbf{s}_n + \epsilon_n, \quad (26)$$

where $\mathbf{x}_n$ is the observation, $\mathbf{s}_n$ is a statistically independent non-Gaussian source, $\epsilon \sim \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I})$ denotes the noise vector, and $\mathbf{A}$ represent a linear transformation. The likelihood can be given by

$$p(\mathbf{x}_n | \mathbf{s}_n, \mathbf{A}, \beta) = \sqrt{\frac{\beta}{2\pi}} e^{\left(\frac{\beta}{2}\right)(\mathbf{x}_n - \mathbf{A} \mathbf{s}_n)^T (\mathbf{x}_n - \mathbf{A} \mathbf{s}_n)}. \quad (27)$$

Due to the adaptive tails in the Student's t probability model, the Student's t distribution can approach the distribution of non-Gaussian sources. The sources used by the Student's t might be described as

$$\begin{aligned} p(s_n^j) &= t(s_n^j | 0, \sigma_j^2, \nu_j) \\ &= \frac{\Gamma(\frac{\nu_j+1}{2})}{\Gamma(\frac{\nu_j}{2}) \sqrt{\nu_j \pi \sigma_j^2}} \left(1 + \frac{(s_n^j)^2}{\nu_j \sigma_j^2} \right)^{-\frac{(\nu_j+1)}{2}} \\ &= \int_0^\infty \mathcal{N}(s_n^j | 0, (u_n^j)^{-1} \sigma_j^2) \text{Ga}(u_n^j | a_n^j, b_n^j) du_n^j, \end{aligned} \quad (28)$$

where $Ga(\cdot)$ represent the Gamma distribution. To estimate the non-Gaussian variables, one can use the variational Bayesian EM approach. Defining $\mathbf{F} \sim \{\mathbf{S}, \mathbf{U}\}$ for latent variables, the log-likelihood is given by

$$\log p(\mathbf{X}|\boldsymbol{\theta}) = \log \frac{p(\mathbf{F}, \mathbf{X}|\boldsymbol{\theta})}{p(\mathbf{F}|\mathbf{X}, \boldsymbol{\theta})}. \quad (29)$$

By introducing an auxiliary distribution $q(\mathbf{F})$ as the approximation distribution, we have

$$\log \frac{p(\mathbf{F}, \mathbf{X}|\boldsymbol{\theta})}{p(\mathbf{F}|\mathbf{X}, \boldsymbol{\theta})} \quad (30)$$

$$= \log \int q(\mathbf{F}) \frac{p(\mathbf{F}, \mathbf{X}|\boldsymbol{\theta})q(\mathbf{F})}{p(\mathbf{F}|\mathbf{X}, \boldsymbol{\theta})q(\mathbf{F})} d\mathbf{F} \quad (31)$$

$$\geq \int q(\mathbf{F}) \left\{ \log \frac{p(\mathbf{F}, \mathbf{X}|\boldsymbol{\theta})}{q(\mathbf{F})} - \log \frac{p(\mathbf{F}|\mathbf{X}, \boldsymbol{\theta})}{q(\mathbf{F})} \right\} d\mathbf{F} \quad (32)$$

$$= F(q(\mathbf{F}), \boldsymbol{\theta}) + KL(q||p) \quad (33)$$

$$\geq F(q(\mathbf{F}), \boldsymbol{\theta}). \quad (34)$$

For variational Bayesian, the latent distributions are assumed as independent $q(\mathbf{F}) \approx q(\mathbf{S})q(\mathbf{U})$. Then, taking the derivative of the lower bound with respect to $q(\mathbf{S})$, $q(\mathbf{U})$, $q(\mathbf{A})$, $q(\beta)$, and $q(v_j)$.

$$\frac{\partial F(q(\mathbf{F}), \boldsymbol{\theta})}{\partial q(\mathbf{S})} = -\frac{\beta}{2}(\mathbf{x}_n^T \mathbf{x}_n - 2\mathbf{A}^T \mathbf{x}_n \mathbf{s}_n + \mathbf{A}^T \mathbf{A} \mathbf{s}_n^2) - \frac{1}{2} \text{diag}(\langle \mathbf{u}_n \rangle) \mathbf{s}_n^2, \quad (35)$$

where $\langle \mathbf{u}_n \rangle = [\langle u_n^1 \rangle, \langle u_n^2 \rangle, \dots, \langle u_n^d \rangle]^T$, on the basis of conjugate exponential distribution, $q(\mathbf{S}) \sim \prod_{n=1}^N (\mathbf{s}_n | \bar{\mathbf{s}}_n, \bar{\boldsymbol{\Sigma}}_s^n)$, the parameters can be obtained as

$$\bar{\boldsymbol{\Sigma}}_s^n = [\text{diag}(\langle \mathbf{u}_n \rangle) + \beta \mathbf{A}^T \mathbf{A}]^{-1}, \quad (36)$$

$$\bar{\mathbf{s}}_n = \beta \bar{\boldsymbol{\Sigma}}_s^n \mathbf{A}^T \mathbf{x}_n. \quad (37)$$

Similarly, by defining $q(\mathbf{U}) = \prod_{j=1}^d \prod_{n=1}^N Ga(u_n^j | \bar{a}_{u_n^j}, \bar{b}_{u_n^j})$, we obtain

$$\frac{\partial F(q(\mathbf{F}), \boldsymbol{\theta})}{\partial q(\mathbf{U})} = \left(\frac{v_j - 1}{2} \right) \log u_n^j - \frac{1}{2} \left(\langle s_n^j \rangle^2 + v_j \sigma_j^2 \right) u_n^j. \quad (38)$$

We have

$$\bar{a}_{u_n^j} = v_j + 1, \quad (39)$$

$$\bar{b}_{u_n^j} = \frac{\langle s_n^j \rangle^2 + v_j \sigma_j^2}{v_j + 1}. \quad (40)$$

Other parameters can be derived by taking the differentiation concerning $\mathbf{A}$, β , and v_j

$$\mathbf{A} = \left[\sum_{n=1}^N \langle s_n \rangle \mathbf{x}_n^T \right] \left[\sum_{n=1}^N \langle s_n s_n^T \rangle \right]^{-1}, \quad (41)$$

$$\beta = ND \left(\sum_{n=1}^N \sum_{i=1}^D \langle (x_n^i - \mathbf{a}_i^T \mathbf{s}_n)^2 \rangle \right)^{-1}, \quad (42)$$

where x_n^i is the i th element of observation and $\mathbf{a}_i$ is the i th column of $\mathbf{A}$, while the degree of freedom v_j can be induced by solving the following nonlinear formula:

$$1 + \frac{N}{2} \log \frac{v_j}{2} - \psi\left(\frac{v_j}{2}\right) + \log \sigma_j^2 + \frac{1}{2} \sum_{n=1}^N (\langle \log u_n^j - \sigma_j^2 \langle u_n^j \rangle \rangle) = 0. \quad (43)$$

The expectations involved in Equation (43) are given by

$$\langle u_n^j \rangle = (\bar{b}_{u_n^j})^{-1}, \quad (44)$$

$$\langle s_n^j \rangle^2 = \langle \mathbf{s}_n \mathbf{s}_n^T \rangle_{jj} = (\bar{\mathbf{s}}_n \bar{\mathbf{s}}_n^T + \bar{\Sigma}_s^n)_{jj}, \quad (45)$$

$$\langle \log u_n^j \rangle = \psi(v_j) - \log \frac{v_j}{2} - \log \bar{b}_{u_n^j}. \quad (46)$$

Traditional ICA methods are limited in their ability to process non-Gaussian distributed signals, resulting in inaccurate decomposition results. Probabilistic ICA uses probability distributions to solve the above-mentioned problem. Probabilistic ICA can better process non-Gaussian signals because of its ability to model different probability distributions. In addition, probabilistic ICA employs the variational Bayesian method to estimate the uncertainty of the separation variable and improve the robustness and interpretability of the model.

3.4. Probabilistic Canonical Correlation Analysis

Given two random vectors, canonical correlation analysis (CCA) is concerned with finding projections such that the components within one set of projections are correlated with components in the other set. The probabilistic extension of CCA is given by [86].

$$\mathbf{x}_1 = \mathbf{W}_1 \mathbf{z} + \mathbf{m}_1 + \epsilon_1, \quad (47)$$

$$\mathbf{x}_2 = \mathbf{W}_2 \mathbf{z} + \mathbf{m}_2 + \epsilon_2, \quad (48)$$

where variables $\mathbf{x}_1 \in \mathbb{R}^{m_1}$, $\mathbf{x}_2 \in \mathbb{R}^{m_2}$, the latent variables $\mathbf{z} \in \mathbb{R}^d \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\min\{m_1, m_2\} \geq d \geq 1$. Then, the conditional distribution is supported by

$$\mathbf{x}_1 | \mathbf{z} \sim \mathcal{N}(\mathbf{W}_1 \mathbf{z} + \mathbf{m}_1, \Psi_1), \quad (49)$$

$$\mathbf{x}_2 | \mathbf{z} \sim \mathcal{N}(\mathbf{W}_2 \mathbf{z} + \mathbf{m}_2, \Psi_2), \quad (50)$$

The parameter set $\Theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{m}_1, \mathbf{m}_2, \Psi_1, \Psi_2\}$ can be determined by maximizing the likelihood. After implementing the EM algorithm, the optimal solution of parameters is given by

$$\hat{\mu}_1 = \hat{\mathbf{m}}_1, \quad (51)$$

$$\hat{\mu}_2 = \hat{\mathbf{m}}_2, \quad (52)$$

$$\hat{\mathbf{W}}_1 = \tilde{\Sigma}_{11} \mathbf{U}_{1d} \mathbf{M}_1, \quad (53)$$

$$\hat{\mathbf{W}}_2 = \tilde{\Sigma}_{22} \mathbf{U}_{2d} \mathbf{M}_2, \quad (54)$$

$$\hat{\Psi}_1 = \tilde{\Sigma}_{11} - \hat{\mathbf{W}}_1 \hat{\mathbf{W}}_1^T, \quad (55)$$

$$\hat{\Psi}_2 = \tilde{\Sigma}_{22} - \hat{\mathbf{W}}_2 \hat{\mathbf{W}}_2^T, \quad (56)$$

where $\mathbf{M}_1, \mathbf{M}_2 \in \mathbb{R}^{d \times d}$ satisfy $\mathbf{M}_1 \mathbf{M}_2^T = \mathbf{J}_d$, $\mathbf{J}_d$ is the diagonal matrix of the first d canonical correlations, $\mathbf{U}_{1d}$ and $\mathbf{U}_{2d}$ are the first d canonical directions, $\hat{\mathbf{m}}$ denotes the sample mean, and $\tilde{\Sigma} = \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} \end{pmatrix}$ denotes sample covariance matrix obtained from data $\mathbf{x}_1^j, \mathbf{x}_2^j$ ($j = 1, 2, \dots, d$). The log-likelihood is given by

$$L \propto \frac{n}{2} \log |\Sigma| + \frac{1}{2} \text{tr} \left(\Sigma^{-1} (\mathbf{x}_j - \mathbf{m})(\mathbf{x}_j - \mathbf{m})^T \right), \quad (57)$$

$$\text{where } \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \mathbf{W}_1 \mathbf{W}_1^T + \Psi_1 & \mathbf{W}_1 \mathbf{W}_2^T \\ \mathbf{W}_2 \mathbf{W}_1^T & \mathbf{W}_2 \mathbf{W}_2^T + \Psi_2 \end{pmatrix}.$$

According to Bayesian inference, the posterior expectations and variances of $\mathbf{z}$ given $\mathbf{x}_1$ and $\mathbf{x}_2$ are

$$\text{var}(\mathbf{z}|\mathbf{x}_1) = \mathbf{I} - \mathbf{M}_1 \mathbf{M}_1^T, \quad (58)$$

$$\text{var}(\mathbf{z}|\mathbf{x}_2) = \mathbf{I} - \mathbf{M}_2 \mathbf{M}_2^T, \quad (59)$$

$$E(\mathbf{z}|\mathbf{x}_1) = \mathbf{M}_1^T \mathbf{U}_{1d}^T (\mathbf{x}_1 - \hat{\mathbf{m}}_1), \quad (60)$$

$$E(\mathbf{z}|\mathbf{x}_2) = \mathbf{M}_2^T \mathbf{U}_{2d}^T (\mathbf{x}_2 - \hat{\mathbf{m}}_2), \quad (61)$$

$$\text{var}(\mathbf{z}|\mathbf{x}_1, \mathbf{x}_2) = \mathbf{I} - \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix}^T \begin{pmatrix} (\mathbf{I} - \mathbf{J}_d^2)^{-1} & (\mathbf{I} - \mathbf{J}_d^2)^{-1} \mathbf{J}_d \\ (\mathbf{I} - \mathbf{J}_d^2)^{-1} \mathbf{J}_d & (\mathbf{I} - \mathbf{J}_d^2)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix}, \quad (62)$$

$$E(\mathbf{z}|\mathbf{x}_1, \mathbf{x}_2) = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \end{pmatrix}^T \begin{pmatrix} (\mathbf{I} - \mathbf{J}_d^2)^{-1} & (\mathbf{I} - \mathbf{J}_d^2)^{-1} \mathbf{J}_d \\ (\mathbf{I} - \mathbf{J}_d^2)^{-1} \mathbf{J}_d & (\mathbf{I} - \mathbf{J}_d^2)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{U}_{1d}^T (\mathbf{x}_1 - \hat{\mathbf{m}}_1) \\ \mathbf{U}_{2d}^T (\mathbf{x}_2 - \hat{\mathbf{m}}_2) \end{pmatrix}. \quad (63)$$

Unlike the sensitivity to noise and missing data in traditional CCA, the probabilistic CCA employs probabilistic models to describe the data generation process, which can naturally deal with noise and missing data, thus securing robustness.

3.5. Probabilistic Fisher Discriminant Analysis

FDA attempts to characterize or distinguish between two classes of objects by using a linear combination of features. Many machine learning and pattern recognition applications use this strategy [87–90]. To identify mixed errors, FDA was integrated with a hybrid kernel extreme learning machine [91]. The criterion of Fisher discriminant is [92]

$$J(U) = \text{tr}((U^T \mathbf{S}_W U)^{-1} U^T \mathbf{S}_B U), \quad (64)$$

$$\mathbf{S}_W = \frac{1}{n} \sum_{k=1}^K \sum_{y_i \in C_k} (y_i - m_k)(y_i - m_k)^T, \quad (65)$$

$$\mathbf{S}_B = \frac{1}{n} \sum_{k=1}^K n_k (m_k - \bar{y})(m_k - \bar{y})^T, \quad (66)$$

where $\mathbf{S}_W$ is the covariance matrix within classes, and $\mathbf{S}_B$ is the covariance matrix between classes, n_k is the number of observations in the k th class, the mean of the observed column vector y_i in the class k is denoted by $m_k = \frac{1}{n_k} \sum_{i \in C_k} y_i$, and $\bar{y} = \frac{1}{n} \sum_{k=1}^K n_k m_k$ is the mean column vector of the observations. The probabilistic framework of the FDA is

$$\mathbf{x} = \mathbf{m} + \mathbf{A}\mathbf{u}, \quad (67)$$

where $\mathbf{u} \sim \mathcal{N}(\mathbf{u}|\mathbf{v}, \mathbf{I})$, $\mathbf{v} \sim \mathcal{N}(\mathbf{v}|\mathbf{0}, \Psi)$, and $\mathbf{v}$ represents the class center. The corresponding graphical model is displayed in Figure 4.

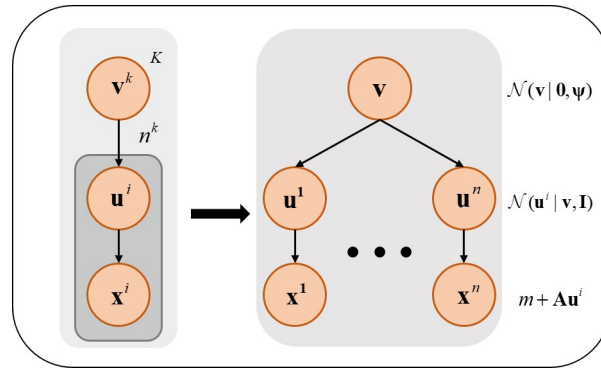


Figure 4. In the latent space, which is the space where the variables are independent, PFDA models the class center $\mathbf{v}$ and examples $\mathbf{u}$. The transformation $\mathbf{A}$ links the example $\mathbf{x}$ to its latent representation.

In the PFDA model, $\mathbf{m}$, Ψ , and $\mathbf{A}$ are unknown. The **log**-likelihood is given by

$$\mathcal{L}(\mathbf{x}^1, \dots, \mathbf{x}^n) = \sum_{k=1}^n \ln p(\mathbf{x}^i : i \in \mathcal{C}_n), \quad (68)$$

$$p(\mathbf{x}^1, \dots, \mathbf{x}^n) = \int \mathcal{N}(\mathbf{y} | \mathbf{0}, \Phi_b) \mathcal{N}(\mathbf{x}^1 | \mathbf{y}, \Phi_w) \cdots \mathcal{N}(\mathbf{x}^n | \mathbf{y}, \Phi_w) d\mathbf{y}, \quad (69)$$

where $p(\mathbf{x}^1 \cdots \mathbf{x}^n)$ is the joint distribution of a set of n patterns, provided they belong to the same class. By computing the integral, we have

$$\mathcal{L}(\mathbf{x}^1, \dots, \mathbf{x}^n) = -\frac{c}{2} (\ln |\Phi_b + \frac{1}{n} \Phi_w| + \text{tr}((\Phi_b + \frac{1}{n} \Phi_w)^{-1} \mathbf{S}_b) + (n-1) \ln |\Phi_w| + n \text{tr}(\Phi_w^{-1} \mathbf{S}_w)). \quad (70)$$

If Φ_w and Φ_b are both positive definite and Φ_w and Φ_b are both positive semi-definite, then we can maximize the value of $\mathcal{L}$. Without these limitations, basic matrix calculus provides

$$\Phi_w = \frac{n}{n-1} \mathbf{S}_w, \quad (71)$$

$$\Phi_b = \mathbf{S}_b - \frac{1}{n-1} \mathbf{S}_w. \quad (72)$$

The EM method is then used to update the parameters $\mathbf{m}$, $\mathbf{A}$, and Ψ to maximize the PFDA model's likelihood

$$\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^i, \quad (73)$$

$$\mathbf{A} = \mathbf{W}^{-T} \left(\frac{n}{n-1} \Lambda_w \right)^{\frac{1}{2}}, \quad (74)$$

$$\Psi = \max \left(0, \frac{n-1}{n} (\Lambda_b / \Lambda_w) - \frac{1}{n} \right). \quad (75)$$

4. Test Statistics

A fault diagnosis differs from a classification problem and should detect abnormalities from sampling data. Test statistics construct a threshold for judgment.

4.1. T^2 Test Statistic

The false alarm rate (FAR) is an elementary concept in fault detection, and it displays the probability of a false alert signal, which is given by

$$FAR = P\{J > J_{th} | f = 0\}, \quad (76)$$

where J denotes a test statistic, J_{th} denotes the threshold, and Equation (76) means the possibility that the decision logic may sound an alert for a malfunction even when one has not occurred. Then, the general formulation of the defect detection issue is provided by

$$y = g + \varepsilon \in \mathbb{R}^m, \varepsilon \sim \mathcal{N}(\mathcal{E}(\varepsilon), \Sigma), \quad (77)$$

where $\mathcal{E}(\varepsilon)$ and Σ are unknown, and assuming that sampling data $y_1, \dots, y_i, (i = 1, \dots, N)$ are available. Find a corresponding threshold J_{th} based on online measurement data $y_k, \dots, y_{k+j}$ with $FAR \geq \alpha$. Under the framework of model Equation (77), T^2 test statistic is defined by

$$\bar{y}_N = \frac{1}{N} \sum_{i=1}^N y_i, \quad (78)$$

$$\hat{\Sigma} = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y}_N)(y_i - \bar{y}_N)^T, \quad (79)$$

$$J = (\bar{y}_N - y_k)^T \hat{\Sigma}^{-1} (\bar{y}_N - y_k), \quad (80)$$

$$J_{th, T^2} = \frac{m(N^2 - 1)}{N(N - m)} \mathcal{F}_\alpha(m, N - m). \quad (81)$$

where the corresponding threshold is represented by J_{th, T^2} , and $\mathcal{F}_\alpha(m, N - m)$ denotes $\mathcal{F}$ -distribution with m and $(N - m)$ degrees of freedom. After obtaining each new measurement y_k , the test statistics will be checked. After calculating the test statistic, the alarm is triggered to indicate the fault by

$$\begin{cases} J \leq J_{th, Q} & \longrightarrow \text{fault-free} \\ J > J_{th, Q} & \longrightarrow \text{faulty and alarm} \end{cases} \quad (82)$$

4.2. SPE or Q Statistic

The inverse matrix of $\hat{\Sigma}$ is necessary for the T^2 statistic, while numerical trouble may incur in the computation by a high-dimensional or ill-conditioned $\hat{\Sigma}$. Q statistic can be alternatively chosen for detecting the fault

$$Q = y^T y. \quad (83)$$

The threshold $J_{th, Q}$ can be computed offline with the process data $y_1, \dots, y_N$ collected

$$\bar{y}_N = \frac{1}{N} \sum_{i=1}^N y_i, \quad (84)$$

$$g = \frac{\overline{Q^2} - \bar{Q}^2}{2\bar{Q}}, \quad (85)$$

$$h = \frac{2\bar{Q}^2}{\overline{Q^2} - \bar{Q}^2}, \quad (86)$$

$$\bar{Q} = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_N)^T (y_i - \bar{y}_N), \quad (87)$$

$$\overline{Q^2} = \frac{1}{N} \sum_{i=1}^N \left((y_i - \bar{y}_N)^T (y_i - \bar{y}_N) \right)^2. \quad (88)$$

Then, set $J_{th, Q}$ for a given significance level α

$$J_{th, Q} = g\chi_\alpha^2(h). \quad (89)$$

4.3. KL Divergence

Kullback–Leibler divergence (KLD) is well-known for measuring the divergence between two PDFs. The *KL* divergence of two continuous PDFs $p(x)$ and $q(x)$ is given by

$$KL(p(x), q(x)) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx. \quad (90)$$

The *KL* divergence is non-negative and zero if and only if $p(x)$ equals $q(x)$. For the PDFs $p(x)$ and $q(x)$, assume the random variable x satisfies Gaussian, $p(x) = \mathcal{N}(\mu_p, \Sigma_p)$, and $q(x) = \mathcal{N}(\mu_q, \Sigma_q)$, Equation (90) can be further written as

$$KL(p(x), q(x)) = \frac{1}{2} \left\{ (\mu_p - \mu_q)^T \Sigma_q^{-1} (\mu_p - \mu_q) + \log \left(\frac{|\Sigma_q|}{|\Sigma_p|} \right) + \text{tr}(\Sigma_q^{-1} \Sigma_p) \right\}. \quad (91)$$

Lei proposed a paper discussing the detection of an incipient fault condition in complex dynamic systems using the *KL* distance [93]. This paper proposes a methodology that can detect incipient anomalous behaviors based on *KL* divergence [94].

4.4. Hellinger Distance

The traditional statistical test cannot be effectively applied to detect the incipient fault. Hellinger distance was first proposed [95] to measure the similarity of two probability distributions. Assuming that $p(x)$ and $q(x)$ represent two continuous PDFs, the Hellinger distance (HD) can be defined as

$$H(p, q) = \sqrt{\frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx}. \quad (92)$$

HD is a symmetric bounded distance, and its possible values are between 0 and 1 as $0 \leq HD(p, q) = HD(q, p) \leq 1$. Based on the Lebesgue metric, the square of HD is expressed as

$$H^2(p, q) = 1 - \int \sqrt{p(x)q(x)} dx. \quad (93)$$

Given two PDFs that obey the normal distributions such that $p(x) \sim \mathcal{N}(\mu_p, \sigma_p^2)$ and $q(x) \sim \mathcal{N}(\mu_q, \sigma_q^2)$, $HD^2(p, q)$ of $p(x)$ with respect to $q(x)$ is given by

$$HD^2 = 1 - \left(\frac{2\sigma_p\sigma_q}{\sigma_p^2 + \sigma_q^2} \right)^{1/2} \exp \left(-\frac{(\mu_p - \mu_q)^2}{4(\sigma_p^2 + \sigma_q^2)} \right). \quad (94)$$

This work combined HD, Bayesian inference, and ICA to monitor a multiblock plant-wide process [96]. HD and KLD are combined to explore the isolation capability of an FDI test by Palmer [97]. Chen introduced HD into a multivariate statistical analysis framework to detect incipient faults for high-speed trains [98].

5. Recent Applications on Statistical Fault Diagnosis

Although statistical schemes for fault diagnosis have been widely applied, there are still many challenges in their practical applications. Abundant modifications are mounted on traditional statistical methods to ensure better performance in industrial applications with harsh working environments. Recent work and applications of statistical fault diagnosis are represented in several aspects in this section.

5.1. Approaches Targeted for Data with Outliers and Missing Values

The advanced sensor technique provides abundant information; however, the multi-rate sampling can lead to incomplete data entries, and the disturbances or measurement

errors shall induce outliers [99,100]. Although traditional statistical inference studies for process modeling and monitoring assume no missing data or outliers, industrial process data typically contain missing values, out-of-range values, and outliers. This greatly influences the statistical FD strategies for process modeling and monitoring.

Traditional statistical analysis would assume that the process data is clean since traditional statistics, such as mean and variance, are sensitive to outliers [101,102]. Additionally, the data gathered from industrial processes tends to not be distributed normally. The representative methods of multivariate statistical analysis usually sink into poor performance due to insufficient quality process data. As a result of the fact that the most frequently used test statistics work under the presumption that the sample data meets Gaussian distribution requirements, data quality also affects the determination of whether a fault exists [103]. The improved weighted k neighborhood standardization (WKNS)-PCA is applied to detect process outliers and its advantage lies in employing a single model for multi-mode industrial processes [104]. Multi-PCA models are trained and integrated with a modified exponentially weighted moving averages (EWMA) control chart to improve robustness to outliers and sensitivity to small and sudden abnormalities [105]. A novel dual robustness projection to latent structure method based on the L1 norm (L1-PLS) is proposed and illustrates its insensitivity to outliers [106]. The robust mixture PPCA model incorporates with a Bayesian soft decision fusion strategy for handling the missing data problem. The mixture of probabilistic principal component analysis (MPPCA) models was trained under multiple operational conditions, such as healthy conditions and anomalies with missing measurement data [107].

5.2. Modifications Designed for Non-Gaussian and Nonlinear Processes

The retrieved latent variables are assumed to be Gaussian in traditional statistical methods like PCA and PLS. Additionally, they require a linear correlation between the variables.

While the multiple manufacturing phases or operating conditions often lead to non-Gaussianity [108,109]. The control threshold and the boundary of normal operation may not be correct in non-Gaussianity-related situations. Therefore, the non-Gaussianity of industrial process data for traditional statistical analysis may cause false alarms [110]. ICA has been used to extract components that are non-Gaussian and statistically independent [111,112], while this method is cumbersome for some practice applications due to a number of drawbacks, including the unstable monitoring results and uncertain number selection of retained independent components [110,113,114].

The nonlinearity of industrial processes is mostly shown in two aspects. On the one hand, the nonlinearity is embodied in the relationship among time series $\mathbf{x}_{k-1} \rightarrow \mathbf{x}_k$. On the other hand, the nonlinear relationship is in different variables $\mathbf{x}_k \rightarrow \mathbf{y}_k$. With the ubiquity of nonlinearity in practical applications, these two aspects deserve attention. Compared with the technique for classifying a linear dataset [115], the feasibility of support vector machine [116] and proximal support vector machines [117] was illustrated. Probabilistic ICA serves as a statistical method for blind source separation and constructs a probabilistic model for uncertainty. It was extended to a variational Bayesian form to improve simplicity and robustness [118]. A semi-supervised learning framework was delivered for ICA [119]. Gaussian processes provided a probabilistic approach for dealing with nonlinearity, the model was derived by Lawrence [120], and the process monitoring was implemented by Ge and Song [121]. The support matrix machine, as an extension of SVM, was developed under a probabilistic framework [122]. Weighted difference principal component analysis (WDPCA) eliminates the multimodal and nonlinear characteristics of the original data by using the weighted difference method [123]. As an extension of CVA, the canonical variate analysis integrated with a dissimilarity-based index, for process incipient fault detection in nonlinear dynamic processes under varying operating conditions [124]. To overcome the poor prediction of PCA when inputs and outputs are nonlinear, a formalism integrating PCA and generalized regression neural networks (GRNNs) is in-

roduced, and it is a one-step procedure, which helps in faster development of nonlinear input–output models [125].

5.3. Approaches for Non-Stationary Processes

PCA and CCA are two types of multivariate analysis that are commonly utilized on processes with a single stationary condition. At the same time, their performance may degrade when confronted with highly dynamic systems. When these methods are applied in practice, the highly autocorrelated and time-dependent measurement data can influence the troubleshooting accuracy.

For online operating systems, one should consider moderate adjustments because of the sequential relationships in dynamic process data. Given that time-wise sample points are auto-correlated, one of the most important qualities of industrial systems should be their dynamic characteristics. As a result, extending statistical modeling from static to dynamic representations is preferred [100]. The state-space definition presents discrete-time dynamic systems to monitor dynamic processes [126]. The FIR-smoothing techniques obtain satisfactory performance regarding measurement with time-delay [127,128]. The methods [129–132] also improve immunity to disturbances using Bayesian inference. Dynamic independent component analysis (DICA) is applied to the augmenting matrix with time-lagged variables to deal with dynamic processes [133]. By combining the advantages of KPCA (Gaussian part) and KICA (non-Gaussian part), Zhang [134] developed a nonlinear dynamic approach to detect fault online compared to other nonlinear approaches.

5.4. Work on Robustness

Industrial process data are often mixed with disturbance, and measurement dimensions vary widely between scales. The data preprocessing involves data normalization to adjust ranges of values, while preprocessing methods need a large computation load, especially for voluminous industrial data. When processes are affected by disturbance, a robust strategy can tolerate unstable measurement quality. Different robust mechanisms of PCA have been researched as fundamental statistical tools for processing data and dimensionality reduction. By combining the projection pursuit (PP) with robust scatter matrix estimation, Hubert proposed a robust PCA [135]. Li and Chen created the robust PCA with PP first [136], and the generalized simulated algorithm was carried out by Xie [137]. Furthermore, the improvement was implemented to obtain stable numerical accuracy for the sake of high-dimensionality [138,139]. The methods mentioned above are deterministic, and the Bayesian methods can be flexible and alternative. The modification of PCA under a probabilistic framework can represent uncertainty and thus be a popular method [140]. To handle the heavy-tail distribution dataset, the Bayesian PCA employed Student's *t* distribution [141] and Laplace distribution [142]. Recently, successful industrial applications have been implemented [143–145].

5.5. Artificial Intelligence Approaches

The construction of system models and feature extraction can be replaced by network-based strategies [146]. When it comes to coping with nonlinearity and non-Gaussianity, neural networks (NN) excel. Besides being equipped with the ability to discover dynamic behaviors, NNs will be promising in FD systems. This promising trend for FD techniques is built firmly on NNs' enhanced computational power and explicability. By altering their weights based on input and output data, artificial NNs mimic the organization of the human brain. The use of counter propagation NNs and recurrent NNs may be seen in [147,148]. Convolutional NN combined with fractional Fourier transform and recurrence plot transform is applied under variable working conditions [149]. A membrane learnable residual spiking NN for autonomous vehicle sensor defects was proposed by Wang and Li [150].

6. Challenges and Open Problems

In practice, some challenges and problems like data preprocessing, real-time ability of FDD methods, and multichannel data from multiple sensors need further research. The existing methods also can be enhanced to deal with non-Gaussianity and non-linearity.

6.1. Preprocessing High-Dimensionality Data

The data sampled in a harsh working environment with disturbance, possibly of poor quality, will reduce the accuracy and effectiveness of FDD schemes. The preprocessing of sampling data to remove outliers will be beneficial for subsequent FDD steps. Modern industrial processes consist of various components, and each part can have abundant measured variables. This benefits real-time monitoring but is accompanied by problems with storing, managing, and preprocessing big data. The statistical methods generally use matrices to compute corresponding statistics, and the high-dimension of matrices increases with burgeoning size data. The dimensional explosion problem will induce a large computation load, demanding hardware facilities.

6.2. Statistical FD Schemes Developed without Real-Time Ability

Unlike a simple classification problem, the design of the FDD scheme needs to consider the dynamic behaviors of practical application. For any online system, the FDD schemes are ultimately designed to detect and diagnose faults in real time. The fast and effective decision-making process of judging whether a fault is reflected in the collected data is significant, especially for high-sample-frequency systems. Most methods reviewed in this paper only applied to static processes and have no ability for online implementation. The online implementation is a research gap for statistical FDD methods.

6.3. Enhancement on Existing Methods

Further enhancement on works presented in this review are suggested: (i) use different statistical methods to improve fault diagnosis; (ii) enhancement of statistical methods not explored by authors in their original works; and (iii) modification of methods to handle non-Gaussian and non-linear data. These methods should confirm their robustness when implemented in industrial processes (mainly chemical, mechanism, and bioengineering).

6.4. Development on Fault Diagnosis

Most methods reviewed in this paper focused on fault detection or isolation. That points to the research niche of the development of fault diagnosis methods.

6.5. Processes with Multichannel Data from Multiple Sensors

Research on the monitoring and diagnosis of general multichannel profiles is still limited in processing a single profile's data. In industrial applications, product quality is often characterized by profile data collected from multiple channels. By taking cross-correlations among multichannel profiles into consideration, profile monitoring is expected to become more sensitive to a variety of shifts.

7. Conclusions

Statistical methods have received increasing attention in recent years as an emerging and active research area in fault detection and diagnosis. This paper has sketched the probabilistic extensions of traditional statistical-based FDD schemes (such as PCA, ICA, CCA, and FDA) along with the test statistics. A brief review of challenges and perspectives on the statistical FDD strategy is presented. The review shows that each of the existing probabilistic approaches has its own strengths and limitations. In addition, several key open problems (such as non-Gaussian, non-linearity, non-stationary, and robustness) are discussed to demonstrate potential future research.

Author Contributions: Writing—original draft preparation, Y.Z. (Yanting Zhu); writing—review and editing, C.Z., S.Z. and J.W.; funding acquisition, S.Z.; supervision, Y.Z. (Yuxuan Zhang), S.Z.; resources, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Jiangsu Province (BK20211528), Fundamental Research Funds for the Central Universities (No. JUSRP123063), 111 Project (B23008).

Data Availability Statement: No new data were created.

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

References

1. Ma, C.Y.; Li, Y.B.; Wang, X.Z.; Cai, Z.Q. Early fault diagnosis of rotating machinery based on composite zoom permutation entropy. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108967. [\[CrossRef\]](#)
2. Tan, H.C.; Xie, S.C.; Ma, W.; Yang, C.X.; Zheng, S.W. Correlation feature distribution matching for fault diagnosis of machines. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 108981. [\[CrossRef\]](#)
3. Xia, P.C.; Huang, Y.X.; Tao, Z.Y.; Liu, C.L.; Liu, J. A digital twin-enhanced semi-supervised framework for motor fault diagnosis based on phase-contrastive current dot pattern. *Reliab. Eng. Syst. Saf.* **2023**, *235*, 109256. [\[CrossRef\]](#)
4. Liu, Z.H.; Chen, L.; Wei, H.L.; Wu, F.M.; Chen, L.; Chen, Y.N. A Tensor-based domain alignment method for intelligent fault diagnosis of rolling bearing in rotating machinery. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108968. [\[CrossRef\]](#)
5. Lu, L.; Han, X.; Li, J.; Hua, J.; Ouyang, M. A review on the key issues for lithium-ion battery management in electric vehicles. *J. Power Sources* **2013**, *226*, 272–288. [\[CrossRef\]](#)
6. Dong, Y.T.; Jiang, H.K.; Wu, Z.H.; Yang, Q.; Liu, Y.P. Digital twin-assisted multiscale residual-self-attention feature fusion network for hypersonic flight vehicle fault diagnosis. *Reliab. Eng. Syst. Saf.* **2023**, *235*, 109253. [\[CrossRef\]](#)
7. Yuan, Z.X.; Xiong, G.J.; Fu, X.F.; Mohamed, A.W. Improving fault tolerance in diagnosing power system failures with optimal hierarchical extreme learning machine. *Reliab. Eng. Syst. Saf.* **2023**, *236*, 109300. [\[CrossRef\]](#)
8. Choi, U.M.; Blaabjerg, F.; Lee, K.B. Study and handling methods of power IGBT module failures in power electronic converter systems. *IEEE Trans. Power Electron.* **2014**, *30*, 2517–2533. [\[CrossRef\]](#)
9. Song, Y.T.; Wang, B.S. Survey on reliability of power electronic systems. *IEEE Trans. Power Electron.* **2012**, *28*, 591–604. [\[CrossRef\]](#)
10. Riera Guasp, M.; Antonino Daviu, J.A.; Capolino, G.A. Advances in electrical machine, power electronic, and drive condition monitoring and fault detection: State of the art. *IEEE Trans. Ind. Electron.* **2014**, *62*, 1746–1759. [\[CrossRef\]](#)
11. Han, T.; Li, Y.F. Out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108648. [\[CrossRef\]](#)
12. Zhang, P.J.; Du, Y.; Habetler, T.G.; Lu, B. A survey of condition monitoring and protection methods for medium-voltage induction motors. *IEEE Trans. Ind. Appl.* **2010**, *47*, 34–46. [\[CrossRef\]](#)
13. Jung, J.H.; Lee, J.J.; Kwon, B.H. Online diagnosis of induction motors using MCSA. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1842–1852. [\[CrossRef\]](#)
14. CusidÓCusido, J.; Romeral, L.; Ortega, J.A.; Rosero, J.A.; Espinosa, A.G. Fault detection in induction machines using power spectral density in wavelet decomposition. *IEEE Trans. Ind. Electron.* **2008**, *55*, 633–643. [\[CrossRef\]](#)
15. Hameed, Z.; Hong, Y.S.; Cho, Y.M.; Ahn, S.H.; Song, C.K. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renew. Sustain. Energy Rev.* **2009**, *13*, 1–39. [\[CrossRef\]](#)
16. Garcia Marquez, F.P.; Mark Tobias, A.; Pinar Perez, J.M.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renew. Energy* **2012**, *46*, 169–178. [\[CrossRef\]](#)
17. Jiang, G.Q.; He, H.B.; Yan, J.; Xie, P. Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Trans. Ind. Electron.* **2018**, *66*, 3196–3207. [\[CrossRef\]](#)
18. Zhou, K.L.; Fu, C.; Yang, S.L. Big data driven smart energy management: From big data to big insights. *Renew. Sustain. Energy Rev.* **2016**, *56*, 215–225. [\[CrossRef\]](#)
19. Jardine, A.K.; Lin, D.M.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [\[CrossRef\]](#)
20. Oppenheimer, C.H.; Loparo, K.A. Physically based diagnosis and prognosis of cracked rotor shafts. In Proceedings of the Component and Systems Diagnostics, Prognostics, and Health Management II, Orlando, FL, USA, 1–5 April 2002; Volume 4733, pp. 122–132.
21. Zhu, Y.T.; Zhao, S.Y.; Zhang, C.X.; Liu, F. Tuning-free filtering for stochastic systems with unmodeled measurement dynamics. *J. Frankl. Inst.* **2023**, *361*, 933–943. [\[CrossRef\]](#)
22. Asadi, D. Actuator Fault detection, identification, and control of a multirotor air vehicle using residual generation and parameter estimation approaches. *Int. J. Aeronaut. Space Sci.* **2023**, *25*, 176–189. [\[CrossRef\]](#)
23. Zhao, S.Y.; Li, Y.Y.; Zhang, C.X.; Luan, X.L.; Liu, F.; Tan, R.M. Robustification of Finite Impulse Response Filter for Nonlinear Systems With Model Uncertainties. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 6506109. [\[CrossRef\]](#)
24. Zhao, S.Y.; Zhou, Z.; Zhang, C.X.; Wu, J.; Liu, F.; Shi, G.Y. Localization of underground pipe jacking machinery: A reliable, real-time and robust INS/OD solution. *Control Eng. Pract.* **2023**, *141*, 105711. [\[CrossRef\]](#)

25. Gao, Z.W.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part II: Fault Diagnosis with Knowledge-Based and Hybrid/Active Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3768–3774. [\[CrossRef\]](#)
26. Qin, S.J. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* **2012**, *36*, 220–234. [\[CrossRef\]](#)
27. Ding, S.X. Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results. *J. Process Control* **2014**, *24*, 431–449. [\[CrossRef\]](#)
28. Yu, X.L.; Zhao, Z.B.; Zhang, X.W.; Chen, X.F.; Cai, J.B. Statistical identification guided open-set domain adaptation in fault diagnosis. *Reliab. Eng. Syst. Saf.* **2023**, *232*, 109047. [\[CrossRef\]](#)
29. Joe Qin, S. Statistical process monitoring: Basics and beyond. *J. Chemom. J. Chemom. Soc.* **2003**, *17*, 480–502. [\[CrossRef\]](#)
30. Malhi, A.; Gao, R.X. PCA-based feature selection scheme for machine defect classification. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 1517–1525. [\[CrossRef\]](#)
31. Choqueuse, V.; Benbouzid, M.E.H.; Amirat, Y.; Turri, S. Diagnosis of three-phase electrical machines using multidimensional demodulation techniques. *IEEE Trans. Ind. Electron.* **2011**, *59*, 2014–2023. [\[CrossRef\]](#)
32. Misra, M.; Yue, H.H.; Qin, S.J.; Ling, C. Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Comput. Chem. Eng.* **2002**, *26*, 1281–1293. [\[CrossRef\]](#)
33. Vong, C.M.; Wong, P.K.; Ip, W.F. A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns. *IEEE Trans. Ind. Electron.* **2012**, *60*, 3372–3385. [\[CrossRef\]](#)
34. Li, G.; Qin, S.J.; Zhou, D.H. Geometric properties of partial least squares for process monitoring. *Automatica* **2010**, *46*, 204–210. [\[CrossRef\]](#)
35. Zhang, Y.W.; Zhou, H.; Qin, S.J.; Chai, T.Y. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares. *IEEE Trans. Ind. Inf.* **2009**, *6*, 3–10. [\[CrossRef\]](#)
36. Muradore, R.; Fiorini, P. A PLS-based statistical approach for fault detection and isolation of robotic manipulators. *IEEE Trans. Ind. Electron.* **2011**, *59*, 3167–3175. [\[CrossRef\]](#)
37. He, X.; Wang, Z.D.; Liu, Y.; Zhou, D.H. Least-squares fault detection and diagnosis for networked sensing systems using a direct state estimation approach. *IEEE Trans. Ind. Inf.* **2013**, *9*, 1670–1679. [\[CrossRef\]](#)
38. Kim, D.S.; Lee, I.B. Process monitoring based on probabilistic PCA. *Chemom. Intell. Lab. Syst.* **2003**, *67*, 109–123. [\[CrossRef\]](#)
39. Kim, D.S.; Yoo, C.K.; Kim, Y.I.; Jung, J.H.; Lee, I.B. Calibration, prediction and process monitoring model based on factor analysis for incomplete process data. *J. Chem. Eng. Jpn.* **2005**, *38*, 1025–1034. [\[CrossRef\]](#)
40. Choi, S.W.; Martin, E.B.; Morris, A.J.; Lee, I.B. Fault detection based on a maximum-likelihood principal component analysis (PCA) mixture. *Ind. Eng. Chem. Res.* **2005**, *44*, 2316–2327. [\[CrossRef\]](#)
41. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [\[CrossRef\]](#)
42. Chen, J.L.; Zhang, L.; Li, Y.F.; Shi, Y.F.; Gao, X.H.; Hu, Y.Q. A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. *Renew. Sustain. Energy Rev.* **2022**, *161*, 112395. [\[CrossRef\]](#)
43. Yu, J.B.; Zhang, Y. Challenges and opportunities of deep learning-based process fault detection and diagnosis: A review. *Neural Comput. Appl.* **2023**, *35*, 211–252. [\[CrossRef\]](#)
44. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4.
45. Si, X.S.; Wang, W.B.; Hu, C.H.; Zhou, D.H.; Pecht, M.G. Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Trans. Reliab.* **2012**, *61*, 50–67. [\[CrossRef\]](#)
46. Excoffier, L.; Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **1995**, *12*, 921–927. [\[PubMed\]](#)
47. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [\[CrossRef\]](#)
48. Zhao, S.Y.; Li, K.; Ahn, C.K.; Huang, B.; Liu, F. Tuning-Free Bayesian Estimation Algorithms for Faulty Sensor Signals in State-Space. *IEEE Trans. Ind. Electron.* **2022**, *70*, 921–929. [\[CrossRef\]](#)
49. Zhao, S.Y.; Huang, B.; Zhao, C.H. Online probabilistic estimation of sensor faulty signal in industrial processes and its applications. *IEEE Trans. Ind. Electron.* **2020**, *68*, 8853–8862. [\[CrossRef\]](#)
50. Zhao, S.Y.; Shmaliy, Y.S.; Ahn, C.K.; Zhao, C.H. Probabilistic monitoring of correlated sensors for nonlinear processes in state space. *IEEE Trans. Ind. Electron.* **2019**, *67*, 2294–2303. [\[CrossRef\]](#)
51. Lo, C.H.; Wong, Y.K.; Rad, A.B. Bond graph based Bayesian network for fault diagnosis. *Appl. Soft Comput.* **2011**, *11*, 1208–1212. [\[CrossRef\]](#)
52. Boudali, H.; Dugan, J.B. A discrete-time Bayesian network reliability modeling and analysis framework. *Reliab. Eng. Syst. Saf.* **2005**, *87*, 337–349. [\[CrossRef\]](#)
53. Cai, B.P.; Huang, L.; Xie, M. Bayesian networks in fault diagnosis. *IEEE Trans. Ind. Inf.* **2017**, *13*, 2227–2240. [\[CrossRef\]](#)
54. Cai, B.P.; Liu, H.L.; Xie, M. A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks. *Mech. Syst. Signal Process.* **2016**, *80*, 31–44. [\[CrossRef\]](#)
55. Cai, B.P.; Liu, Y.; Xie, M. A dynamic-Bayesian-network-based fault diagnosis methodology considering transient and intermittent faults. *IEEE Trans. Autom. Sci. Eng.* **2016**, *14*, 276–285. [\[CrossRef\]](#)
56. Trucco, P.; Cagno, E.; Ruggeri, F.; Grande, O. A Bayesian Belief Network modelling of organisational factors in risk analysis: A case study in maritime transportation. *Reliab. Eng. Syst. Saf.* **2008**, *93*, 845–856. [\[CrossRef\]](#)
57. Langseth, H.; Portinale, L. Bayesian networks in reliability. *Reliab. Eng. Syst. Saf.* **2007**, *92*, 92–108. [\[CrossRef\]](#)

58. Khakzad, N.; Khan, F.; Amyotte, P. Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 925–932. [\[CrossRef\]](#)
59. O'Hagan, A. Bayesian analysis of computer code outputs: A tutorial. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 1290–1300. [\[CrossRef\]](#)
60. Bobbio, A.; Portinale, L.; Minichino, M.; Ciancamerla, E. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliab. Eng. Syst. Saf.* **2001**, *71*, 249–260. [\[CrossRef\]](#)
61. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [\[CrossRef\]](#)
62. Jiang, Q.C.; Yan, X.F.; Huang, B. Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference. *IEEE Trans. Ind. Electron.* **2015**, *63*, 377–386. [\[CrossRef\]](#)
63. Da Silva, A.M.; Povinelli, R.J.; Demerdash, N.A. Induction machine broken bar and stator short-circuit fault diagnostics based on three-phase stator current envelopes. *IEEE Trans. Ind. Electron.* **2008**, *55*, 1310–1318. [\[CrossRef\]](#)
64. Zhou, T.T.; Zhang, L.B.; Han, T.; Droguett, E.L.; Mosleh, A.; Chan, F.T. An uncertainty-informed framework for trustworthy fault diagnosis in safety-critical applications. *Reliab. Eng. Syst. Saf.* **2023**, *229*, 108865. [\[CrossRef\]](#)
65. Zhou, T.T.; Han, T.; Droguett, E.L. Towards trustworthy machine fault diagnosis: A probabilistic Bayesian deep learning framework. *Reliab. Eng. Syst. Saf.* **2022**, *224*, 108525. [\[CrossRef\]](#)
66. Tenenbaum, J.B.; Silva, V.D.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [\[CrossRef\]](#)
67. Ahonen, T.; Hadid, A.; Pietikainen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2037–2041. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.Y. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 131–137. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1999**, *61*, 611–622. [\[CrossRef\]](#)
71. Kariwala, V.; Odiwei, P.E.; Cao, Y.; Chen, T. A branch and bound method for isolation of faulty variables through missing variable analysis. *J. Process Control* **2010**, *20*, 1198–1206. [\[CrossRef\]](#)
72. Yang, Y.W.; Ma, Y.X.; Song, B.; Shi, H.B. An aligned mixture probabilistic principal component analysis for fault detection of multimode chemical processes. *Chin. J. Chem. Eng.* **2015**, *23*, 1357–1363. [\[CrossRef\]](#)
73. He, B.; Yang, X.H.; Chen, T.; Zhang, J. Reconstruction-based multivariate contribution analysis for fault isolation: A branch and bound approach. *J. Process Control* **2012**, *22*, 1228–1236. [\[CrossRef\]](#)
74. Zhu, J.L.; Ge, Z.Q.; Song, Z.H. Robust modeling of mixture probabilistic principal component analysis and process monitoring application. *AIChE J.* **2014**, *60*, 2143–2157. [\[CrossRef\]](#)
75. Ge, Z.Q.; Song, Z.H. Robust monitoring and fault reconstruction based on variational inference component analysis. *J. Process Control* **2011**, *21*, 462–474. [\[CrossRef\]](#)
76. Zhu, J.L.; Ge, Z.Q.; Song, Z.H. Dynamic mixture probabilistic PCA classifier modeling and application for fault classification. *J. Chemom.* **2015**, *29*, 361–370. [\[CrossRef\]](#)
77. Zhu, J.L.; Ge, Z.Q.; Song, Z.H. HMM-driven robust probabilistic principal component analyzer for dynamic process fault classification. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3814–3821. [\[CrossRef\]](#)
78. de Andrade Melani, A.H.; de Carvalho Michalski, M.A.; da Silva, R.F.; de Souza, G.F.M. A framework to automate fault detection and diagnosis based on moving window principal component analysis and Bayesian network. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107837. [\[CrossRef\]](#)
79. Zheng, J.H.; Song, Z.H.; Ge, Z.Q. Probabilistic learning of partial least squares regression model: Theory and industrial applications. *Chemom. Intell. Lab. Syst.* **2016**, *158*, 80–90. [\[CrossRef\]](#)
80. Pérez, N.F.; Ferré, J.; Boqué, R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemom. Intell. Lab. Syst.* **2009**, *95*, 122–128. [\[CrossRef\]](#)
81. Zheng, J.H.; Song, Z.H. Semisupervised learning for probabilistic partial least squares regression model and soft sensor application. *J. Process Control* **2018**, *64*, 123–131. [\[CrossRef\]](#)
82. Xie, Y. Fault monitoring based on locally weighted probabilistic kernel partial least square for nonlinear time-varying processes. *J. Chemom.* **2019**, *33*, e3196. [\[CrossRef\]](#)
83. Botella, C.; Ferré, J.; Boqué, R. Classification from microarray data using probabilistic discriminant partial least squares with reject option. *Talanta* **2009**, *80*, 321–328. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Li, Q.H.; Pan, F.; Zhao, Z.G. Concurrent probabilistic PLS regression model and its applications in process monitoring. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 40–54. [\[CrossRef\]](#)
85. Beckmann, C.F.; Smith, S.M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **2004**, *23*, 137–152. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Bach, F.R.; Jordan, M.I. *A Probabilistic Interpretation of Canonical Correlation Analysis*; Technical Report; University of California Berkeley: Berkeley, CA, USA, 2005.

87. He, X.F.; Yan, S.C.; Hu, Y.X.; Niyogi, P.; Zhang, H.J. Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 328–340.
88. Yu, H.; Yang, J. A direct LDA algorithm for high-dimensional data—With application to face recognition. *Pattern Recognit.* **2001**, *34*, 2067–2070. [[CrossRef](#)]
89. Chen, L.F.; Liao, H.Y.M.; Ko, M.T.; Lin, J.C.; Yu, G.J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.* **2000**, *33*, 1713–1726. [[CrossRef](#)]
90. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
91. Liu, J.; Xu, H.T.; Peng, X.Y.; Wang, J.L.; He, C.M. Reliable composite fault diagnosis of hydraulic systems based on linear discriminant analysis and multi-output hybrid kernel extreme learning machine. *Reliab. Eng. Syst. Saf.* **2023**, *234*, 109178. [[CrossRef](#)]
92. Bouveyron, C.; Brunet, C. Probabilistic Fisher discriminant analysis: A robust and flexible alternative to Fisher discriminant analysis. *Neurocomputing* **2012**, *90*, 12–22. [[CrossRef](#)]
93. Xie, L.; Zeng, J.S.; Kruger, U.; Wang, X.; Geluk, J. Fault detection in dynamic systems using the Kullback–Leibler divergence. *Control Eng. Pract.* **2015**, *43*, 39–48. [[CrossRef](#)]
94. Chen, H.T.; Jiang, B.; Lu, N.Y. An improved incipient fault detection method based on Kullback-Leibler divergence. *ISA Trans.* **2018**, *79*, 127–136. [[CrossRef](#)] [[PubMed](#)]
95. Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Stat.* **1977**, 445–463. [[CrossRef](#)]
96. Jiang, Q.C.; Wang, B.; Yan, X.F. Multiblock independent component analysis integrated with Hellinger distance and Bayesian inference for non-Gaussian plant-wide process monitoring. *Ind. Eng. Chem. Res.* **2015**, *54*, 2497–2508. [[CrossRef](#)]
97. Palmer, K.A.; Bollas, G.M. Sensor selection embedded in active fault diagnosis algorithms. *IEEE Trans. Control Syst. Technol.* **2019**, *29*, 593–606. [[CrossRef](#)]
98. Chen, H.T.; Jiang, B.; Lu, N.Y. A newly robust fault detection and diagnosis method for high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2198–2208. [[CrossRef](#)]
99. Barnett, V.; Lewis, T. *Outliers in Statistical Data*; Wiley: New York, NY, USA, 1994; Volume 3.
100. Zhu, J.L.; Ge, Z.Q.; Song, Z.H.; Gao, F.F. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annu. Rev. Control* **2018**, *46*, 107–133. [[CrossRef](#)]
101. Yin, S.; Ding, S.X.; Xie, X.C.; Luo, H. A review on basic data-driven approaches for industrial process monitoring. *IEEE Trans. Ind. Electron.* **2014**, *61*, 6418–6428. [[CrossRef](#)]
102. Ding, S.X. *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*; Springer: Berlin/Heidelberg, Germany, 2014.
103. Chen, H.T.; Jiang, B.; Ding, S.X.; Huang, B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 1700–1716. [[CrossRef](#)]
104. Wang, G.Z.; Liu, J.C.; Zhang, Y.W.; Li, Y. A novel multi-mode data processing method and its application in industrial process monitoring. *J. Chemom.* **2015**, *29*, 126–138. [[CrossRef](#)]
105. Bakdi, A.; Kouadri, A.; Mekhilef, S. A data-driven algorithm for online detection of component and system faults in modern wind turbines at different operating zones. *Renew. Sustain. Energy Rev.* **2019**, *103*, 546–555. [[CrossRef](#)]
106. Zhou, J.L.; Zhang, S.L.; Wang, J. A dual robustness projection to latent structure method and its application. *IEEE Trans. Ind. Electron.* **2020**, *68*, 1604–1614. [[CrossRef](#)]
107. Ma, Z.; Luo, Y.Z.; Yun, C.B.; Wan, H.P.; Shen, Y.B. An MPPCA-based approach for anomaly detection of structures under multiple operational conditions and missing data. *Struct. Health Monit.* **2023**, *22*, 1069–1089. [[CrossRef](#)]
108. Jin, X.H.; Zhao, M.B.; Chow, T.W.; Pecht, M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Trans. Ind. Electron.* **2013**, *61*, 2441–2451. [[CrossRef](#)]
109. Chen, Z.W.; Ding, S.X.; Peng, T.; Yang, C.H.; Gui, W.H. Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms. *IEEE Trans. Ind. Electron.* **2017**, *65*, 1559–1567. [[CrossRef](#)]
110. Ge, Z.Q.; Song, Z.H.; Gao, F.R. Review of recent research on data-based process monitoring. *Ind. Eng. Chem. Res.* **2013**, *52*, 3543–3562. [[CrossRef](#)]
111. Li, R.F.; Wang, X.Z. Dimension reduction of process dynamic trends using independent component analysis. *Comput. Chem. Eng.* **2002**, *26*, 467–473. [[CrossRef](#)]
112. Kano, M.; Tanaka, S.; Hasebe, S.; Hashimoto, I.; Ohno, H. Monitoring independent components for fault detection. *AIChE J.* **2003**, *49*, 969–976. [[CrossRef](#)]
113. Chen, H.; Jiang, B. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 450–465. [[CrossRef](#)]
114. Chen, H.; Liu, Z.; Alippi, C.; Huang, B.; Liu, D. Explainable intelligent fault diagnosis for nonlinear dynamic systems: From unsupervised to supervised learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–14. [[CrossRef](#)]
115. Chiang, L.H.; Russell, E.L.; Braatz, R.D. *Fault Detection and Diagnosis in Industrial Systems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.
116. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.

117. Chiang, L.H.; Kotanchek, M.E.; Kordon, A.K. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Comput. Chem. Eng.* **2004**, *28*, 1389–1401. [\[CrossRef\]](#)
118. Chan, K.; Lee, T.W.; Sejnowski, T.J. Variational Bayesian learning of ICA with missing data. *Neural Comput.* **2003**, *15*, 1991–2011. [\[CrossRef\]](#)
119. Salazar, A.; Vergara, L.; Serrano, A.; Igual, J. A general procedure for learning mixtures of independent component analyzers. *Pattern Recognit.* **2010**, *43*, 69–85. [\[CrossRef\]](#)
120. Lawrence, N.; Hyvärinen, A. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *J. Mach. Learn. Res.* **2005**, *6*, 1783–1816.
121. Ge, Z.Q.; Song, Z.H. Nonlinear probabilistic monitoring based on the Gaussian process latent variable model. *Ind. Eng. Chem. Res.* **2010**, *49*, 4792–4799. [\[CrossRef\]](#)
122. Li, X.; Li, Y.; Yan, K.; Shao, H.D.; Lin, J.J. Intelligent fault diagnosis of bevel gearboxes using semi-supervised probability support matrix machine and infrared imaging. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108921. [\[CrossRef\]](#)
123. Guo, J.; Wang, X.; Li, Y.; Wang, G.Z. Fault detection based on weighted difference principal component analysis. *J. Chemom.* **2017**, *31*, e2926. [\[CrossRef\]](#)
124. Pilario, K.E.S.; Cao, Y. Canonical variate dissimilarity analysis for process incipient fault detection. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5308–5315. [\[CrossRef\]](#)
125. Kulkarni, S.G.; Chaudhary, A.K.; Nandi, S.; Tambe, S.S.; Kulkarni, B.D. Modeling and monitoring of batch processes using principal component analysis (PCA) assisted generalized regression neural networks (GRNN). *Biochem. Eng. J.* **2004**, *18*, 193–210. [\[CrossRef\]](#)
126. Roweis, S.; Ghahramani, Z. A unifying review of linear Gaussian models. *Neural Comput.* **1999**, *11*, 305–345. [\[CrossRef\]](#)
127. Zhao, S.Y.; Shmaliy, Y.S.; Liu, F. Batch Optimal FIR Smoothing: Increasing State Informativity in Nonwhite Measurement Noise Environments. *IEEE Trans. Ind. Inf.* **2022**. [\[CrossRef\]](#)
128. Zhao, S.Y.; Wang, J.F.; Shmaliy, Y.S.; Liu, F. Discrete Time q -Lag Maximum Likelihood FIR Smoothing and Iterative Recursive Algorithm. *IEEE Trans. Signal Process.* **2021**, *69*, 6342–6354. [\[CrossRef\]](#)
129. Zhang, T.Y.; Zhao, S.Y.; Luan, X.L.; Liu, F. Bayesian Inference for State-Space Models with Student- t Mixture Distributions. *IEEE Trans. Cybern.* **2022**, 4435–4445. [\[CrossRef\]](#) [\[PubMed\]](#)
130. Zhao, S.Y.; Huang, B. Trial-and-error or avoiding a guess? Initialization of the Kalman filter. *Automatica* **2020**, *121*, 109184. [\[CrossRef\]](#)
131. Zhao, S.Y.; Shmaliy, Y.S.; Andrade-Lucio, J.A.; Liu, F. Multipass optimal FIR filtering for processes with unknown initial states and temporary mismatches. *IEEE Trans. Ind. Inf.* **2020**, *17*, 5360–5368. [\[CrossRef\]](#)
132. Zhao, S.Y.; Shmaliy, Y.S.; Ahn, C.K.; Liu, F. Self-tuning unbiased finite impulse response filtering algorithm for processes with unknown measurement noise covariance. *IEEE Trans. Control Syst. Technol.* **2020**, *29*, 1372–1379. [\[CrossRef\]](#)
133. Lee, J.M.; Yoo, C.K.; Lee, I.B. Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chem. Eng. Sci.* **2004**, *59*, 2995–3006. [\[CrossRef\]](#)
134. Zhang, Y.W. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chem. Eng. Sci.* **2009**, *64*, 801–811. [\[CrossRef\]](#)
135. Hubert, M.; Rousseeuw, P.J.; Vanden Branden, K. ROBPCA: A new approach to robust principal component analysis. *Technometrics* **2005**, *47*, 64–79. [\[CrossRef\]](#)
136. Li, G.Y.; Chen, Z.L. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *J. Am. Stat. Assoc.* **1985**, *80*, 759–766. [\[CrossRef\]](#)
137. Xie, Y.L.; Wang, J.H.; Liang, Y.Z.; Sun, L.X.; Song, X.H.; Yu, R.Q. Robust principal component analysis by projection pursuit. *J. Chemom.* **1993**, *7*, 527–541. [\[CrossRef\]](#)
138. Hubert, M.; Rousseeuw, P.J.; Verboven, S. A fast method for robust principal components with applications to chemometrics. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 101–111. [\[CrossRef\]](#)
139. Croux, C.; Ruiz Gazen, A. High breakdown estimators for principal components: The projection-pursuit approach revisited. *J. Multivar. Anal.* **2005**, *95*, 206–226. [\[CrossRef\]](#)
140. Ding, X.H.; He, L.H.; Carin, L. Bayesian robust principal component analysis. *IEEE Trans. Image Process.* **2011**, *20*, 3419–3430. [\[CrossRef\]](#) [\[PubMed\]](#)
141. Luttinen, J.; Ilin, A.; Karhunen, J. Bayesian robust PCA of incomplete data. *Neural Process. Lett.* **2012**, *36*, 189–202. [\[CrossRef\]](#)
142. Gao, J.B. Robust L1 principal component analysis and its Bayesian variational inference. *Neural Comput.* **2008**, *20*, 555–572. [\[CrossRef\]](#) [\[PubMed\]](#)
143. Chen, T.; Martin, E.; Montague, G. Robust probabilistic PCA with missing data and contribution analysis for outlier detection. *Comput. Stat. Data Anal.* **2009**, *53*, 3706–3716. [\[CrossRef\]](#)
144. Chen, H.; Luo, H.; Huang, B.; Jiang, B.; Kaynak, O. Transfer Learning-motivated Intelligent Fault Diagnosis Designs: A Survey, Insights, and Perspectives. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 2969–2983. [\[CrossRef\]](#) [\[PubMed\]](#)
145. Chen, H.; Huang, B. Fault-tolerant soft sensors for dynamic systems. *IEEE Trans. Control. Syst. Technol.* **2022**, 2805–2818. [\[CrossRef\]](#)
146. Su, H.; Chong, K.T. Induction machine condition monitoring using neural network modeling. *IEEE Trans. Ind. Electron.* **2007**, *54*, 241–249. [\[CrossRef\]](#)

147. Li, C.J.; Huang, T.Y. Automatic structure and parameter training methods for modeling of mechanical systems by recurrent neural networks. *Appl. Math. Modell.* **1999**, *23*, 933–944.
148. Deuszkiewicz, P.; Radkowski, S. On-line condition monitoring of a power transmission unit of a rail vehicle. *Mech. Syst. Signal Process.* **2003**, *17*, 1321–1334. [[CrossRef](#)]
149. Bai, R.X.; Meng, Z.; Xu, Q.S.; Fan, F.J. Fractional Fourier and time domain recurrence plot fusion combining convolutional neural network for bearing fault diagnosis under variable working conditions. *Reliab. Eng. Syst. Saf.* **2023**, *232*, 109076. [[CrossRef](#)]
150. Wang, H.; Li, Y.F. Bioinspired membrane learnable spiking neural network for autonomous vehicle sensors fault diagnosis under open environments. *Reliab. Eng. Syst. Saf.* **2023**, *233*, 109102. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.