International Journal of Control, Automation, and Systems 20(9) (2022) 2892-2901
http://dx.doi.org/10.1007/s12555-021-0276-9
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Time-series Independent Component Analysis-aided Fault Detection for Running Gear Systems

Chao Cheng*, Sheng Yang, Yu Song, and Gang Liu

Abstract: By dealing with the non-Gaussian measurement and slow-change faults in running gear systems, this paper presents a fault detection (FD) scheme named time-series independent component analysis (TsICA), where the time-series characteristic is taken into account. Time-series algorithms can extract slow-change information in the data. The advantages of the proposed method are: 1) it can improve the FD power; 2) it considers the information in the data 3) it is suitable for non-Gaussian systems; 4) it is sensitive to slow-change faults; 5) it can effectively shorten the first time of fault detection. The feasibility of the proposed scheme is verified through a case study on running gear systems.

Keywords: Fault detection (FD), non-Gaussian, running gear systems, time-series independent component analysis (TsICA).

1. INTRODUCTION

Running gear systems are one of the crucial parts to ensure the safe operation of high-speed trains. It is a complex electromechanical system composed of gearboxes, traction motors, wheelsets, bearings, and other components, which plays a crucial role in supporting the car body and providing traction power [1–3]. If any slow-change or potential faults cannot be discovered timely and handled effectively, they may cause catastrophic accidents [4–6]. In order to avoid these traffic accidents, a large number of sensors are used to monitor the operation state of running gear systems, and massive amounts of data are collected [7–10]. Facing the close connection of data and non-Gaussian problems, it is necessary to design a new data-driven fault detection (FD) method for running gear systems.

In recent years, many experts and scholars have proposed a series of FD and diagnosis methods for high-speed trains, which are mainly divided into model-based methods [11], qualitative methods [12], and data-driven methods [1,13]. Among them, the model-based FD methods require accurate system models and operating mechanisms. However, the structure of running gear systems is complex, in which various components influence each other. Its nonlinearity is challenging to construct accurate mathematical models, so model-based FD methods have limitations in running gear systems. Besides, the primary task of FD methods based on qualitative experience is to obtain a complete fault diagnosis knowledge base. As running gear systems are affected by uncertain factors (such as temperature difference and variable working conditions), a complete fault knowledge base is difficult to establish. Compared with the previous two kinds, the data-driven FD method only needs analyze the operation status of running gear systems based on available data, which overcomes the complexity of the analytical model and the completeness of the fault knowledge base.

As an essential branch of data-driven FD methods [14–16], multivariate statistical process monitoring methods mainly include principal component analysis (PCA) [17,18], partial least squares (PLS) [19,20], independent component analysis (ICA) [21–24], correlated correlation analysis (CCA) [25,26] and so on [27–29]. Some of these methods have been successfully applied in running gear systems to improve the FD capability of slow-change faults. Specifically, the numerical fluctuation range of the three-coordinate vibration sensor is reduced in [17] via performing deep PCA on the initial data. In addition to in-depth analysis of data, parallel analysis of data from different angles can also effectively detect faults. For example, a two-dimensional CCA is developed to avoid poor detection results by detecting data from different dimensions separately [25]. Taking into account the correlation of the variables, a modified PLS is proposed in [19]. This algorithm divides the original variables into four subspaces to

Manuscript received April 8, 2021; revised July 30, 2021 and September 4, 2021; accepted November 2, 2021. Recommended by Associate Editor M. Chadli under the direction of Editor Kyoung Kwan Ahn. This work is supported by the National Natural Science Foundation of China (61903047,U20A20186), Jilin Science and Technology Department (20200401127GX).

Chao Cheng, Sheng Yang, Yu Song, and Gang Liu are with the School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China (e-mails: {chengx415, shengyang1997}@163.com, 2082439351@qq.com, lg@ccut.edu.cn).
* Corresponding author.

detect them separately. In summary, the improved PCA, CCA and PLS detect slow-change faults through linear or nonlinear data mapping. Although the computational efficiency is improved, the non-Gaussian problem of system data is ignored.

Compared with PCA, CCA and PLS methods, ICA improves the FD power by extracting valid information from non-Gaussian data without changing the data characteristic. An improved ICA based on model ideas is proposed to solve the non-Gaussian problem by analyzing massive data [22]. The non-Gaussian and nonlinear features of the data may exist simultaneously. For nonlinear and non-Gaussian problems, a weighted kernel ICA is adopted in [23]. It uses a Gaussian mixture model to assign weight values. Then FD is performed on some variables with high weight values. In addition to nonlinear and non-Gaussian problems, considering the instability of a single module, a hybrid dynamic ICA is adopted in [24] to improve the FD rate through multi-module parallel detection. Although the nonlinearity and non-Gaussian of the data are considered in the above methods, the influence caused by unknown noises is ignored. Besides, when looking for the optimization threshold, it is assumed that the data obeys Gaussian distribution, which causes the obtained threshold to be unreasonable and affects the FD result.

This paper takes running gear systems in high-speed trains as the research object. Aiming at the problem of non-Gaussian measurements and slow-change faults, a time-series independent component analysis (TsICA) is proposed. The main contributions of the method proposed in this article are listed as follows:

1) TsICA is highly feasible. It only relies on historic data and does not require prior knowledge.
2) TsICA is flexible. By extracting slow-change information from the data so that the residual is related to noise, it effectively solves the problem of separation of slow-change fault data and noise data.
3) TsICA effectively improves the detection efficiency of the slow-change fault in running gear systems.

The rest of this article is arranged explicitly as follows: In Section 2, running gear systems of high-speed trains and FD problems are briefly described. Section 3 of this article describes in detail how to detect faults. In Section 4, the reliability of the proposed scheme is verified by a case study on running gear systems. Finally, this paper ends up with a conclusion in Section 5.

2. PRELIMINARIES

This part briefly describes running gear systems of the high-speed train. Then, several problems of FD are introduced.

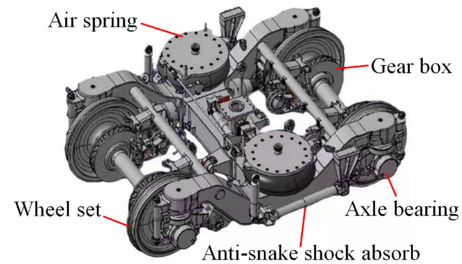

Fig. 1. The structure of running gears.

2.1. Running gear systems of high-speed train

Fig. 1 is a structural diagram of running gear systems, including running gears, gearboxes, and bearings. The data observed by sensors is the primary basis for the maintenance and repairing of these components. The data mainly includes the speed and acceleration of the bogie, the temperature and impact of the bearing, the temperature and vibration of the gearbox. From the internal and external aspects of the system, the data collected by sensors is non-Gaussian. From the internal perspective of the system, the fluctuation of data is related to the information control system of high-speed trains in [5,11]. For example, when a high-speed train decelerates or accelerates, the monitored axle temperature data has a decreasing or increasing trend. From the external perspective of the system, the varying train operating environment, uneven track treads, and the quality of component materials may cause random interference to the collected data in [30]. When the train is turning, the unbalanced height of the left and right wheelsets may cause significant fluctuations in the vibration data collected between the tread and the wheelsets.

In order to efficiently deal with non-Gaussian problems, it is necessary to establish a fully adapted model for high-speed train systems. In response to this problem, this article chooses to model running gear systems.

2.2. Problem description

According to the idea of data-driven fault detection, the following fault analysis model is given in [31]

$$\mathbf{X} = \mathbf{X}_f + \mathbf{N} = \mathbf{X}^* + \mathbf{F} + \mathbf{N}, \tag{1}$$

where $\mathbf{X}$ is initial data set, $\mathbf{X}^*$ is normal data set, $\mathbf{N}$ represents noise, and $\mathbf{F}$ is fault data. Equation (1) is a fault model under faulty conditions, where different signals can be separately described. It also indicates that, running gear systems are often accompanied by noise signals. Therefore, the initial data is closely related to the noise data. This increases the difficulty of completely separating different types of data. The fault analysis model under the actual conditions of running gear systems is given as follows:

$$\mathbf{X} = \mathbf{X}_f^* + \mathbf{N}^*, \tag{2}$$

where $\mathbf{X}_f^*$ includes separated normal data and fault data, and $\mathbf{N}^*$ represents the noise. In the non-Gaussian background, extracting slow-change information $\mathbf{X}_f^*$ from the initial data $\mathbf{X}$ to make it as close to $\mathbf{X}_f$ as possible is difficult.

3. METHODOLOGY

The TsICA algorithm is introduced in detail, mainly including slow-change data extraction and FD for running gear systems.

3.1. Time-series independent component analysis

ICA is a typical blind source separation algorithm, giving the following matrix

$$\mathbf{X} = \mathbf{FS} + \mathbf{E}, \tag{3}$$

$$\mathbf{S} = \mathbf{DUX}. \tag{4}$$

Equation (4) is transformed from (3), where $\mathbf{X}$ is the initial data matrix, $\mathbf{F}$ is mixing matrix, $\mathbf{S}$ is the independent component matrix, $\mathbf{E}$ is the residual matrix, $\mathbf{D}$ is the unmixing matrix, and $\mathbf{U}$ is the whitening matrix. The collection of initial data $\mathbf{X}$ is affected by noise data, which increases the difficulty of fault detection.

A TsICA algorithm is proposed to solve this problem. The algorithm extracts almost noiseless data from the initial data. The most fundamental principle of TsICA thinking comes from the autoregressive moving average (ARMA) model. Specifically, it can be described by

$$\mathbf{y}_k = \alpha_1 \mathbf{y}_{k-1} + \alpha_2 \mathbf{y}_{k-2} + \cdots + \alpha_s \mathbf{y}_{k-s}. \tag{5}$$

Latent variable $\mathbf{y}_k$ may be regarded as a linear combination of initial variable $\mathbf{x}_k$

$$\mathbf{y}_k = \mathbf{x}_k \boldsymbol{\beta},$$

$$\bar{\mathbf{y}}_k = \alpha_1 \mathbf{x}_{k-1}\boldsymbol{\beta} + \alpha_2 \mathbf{x}_{k-2}\boldsymbol{\beta} + \ldots + \alpha_s \mathbf{x}_{k-s}\boldsymbol{\beta}, \tag{6}$$

where $\bar{\mathbf{y}}_k$ is obtained from (5), k is the sampling moment, and s is the lag length. $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_s]$, $\boldsymbol{\beta}$ is weight vector of feature. The main idea of the TsICA model is to maximize the covariance between the latent variable and the predicted latent variable. It can be expressed as follows:

$$\frac{1}{N}\sum_{k=s+1}^{s+n} \bar{\mathbf{y}}_k^T \bar{\mathbf{y}}_k. \tag{7}$$

The initial data matrix

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n+s}], \tag{8}$$

the following matrix can be obtained through the initial data matrix $\mathbf{X}$

$$\mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+1}, \cdots, \mathbf{x}_{i+n-1}]^T,$$

$$\mathbf{A}_i = \dot{\mathbf{X}}_i^T \dot{\mathbf{X}}_i,$$

$$\mathbf{B}_i = \mathbf{X}_i^T \mathbf{X}_i,$$

$$\mathbf{C}_i = \mathbf{B}_i^{-1} \mathbf{A}_i, \tag{9}$$

where $i = 1, 2, ..., s+1$, $\dot{\mathbf{X}} = \mathbf{X}_i - \mathbf{X}_{i-1}$ is the first derivative. On the basis of (7), (10) can be obtained by combining the slow feature analysis.

$$\max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad \boldsymbol{\beta}^T \{\frac{1}{2}\sum_{i=1}^{s} (\mathbf{C}_{s+1}^T \mathbf{C}_i + \mathbf{C}_i^T \mathbf{C}_{s+1})\}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}),$$

$$\text{s.t.} \quad \|\boldsymbol{\alpha}\| = 1, \|\boldsymbol{\beta}\| = 1. \tag{10}$$

Remark 1: The main idea of slow feature analysis is the generalized eigenvalue decomposition of matrix $\boldsymbol{AW} = \boldsymbol{BW\Lambda}$. The matrix can be changed to $\mathbf{W}^T\mathbf{B}^{-1}\mathbf{AW} = \boldsymbol{\Lambda}$, where $\mathbf{W}$ is weight vector and $\boldsymbol{\Lambda}$ represent a diagonal matrix. The values inside represent the speed of change, arranged in descending order.

The maximum of (10) can be analyzed by Lagrangian multiplier method.

$$L = \boldsymbol{\beta}^T \overline{\mathbf{C}}(\boldsymbol{\alpha} \otimes \boldsymbol{\beta}) + \frac{1}{2} c_\beta (1 - \boldsymbol{\beta}^T \boldsymbol{\beta})$$

$$+ \frac{1}{2} c_\alpha (1 - \boldsymbol{\alpha}^T \boldsymbol{\alpha}), \tag{11}$$

where $\overline{\mathbf{C}} = \frac{1}{2}\sum_{i=1}^{s}(\mathbf{C}_{s+1}^T \mathbf{C}_i + \mathbf{C}_i^T \mathbf{C}_{s+1})$, $\boldsymbol{\alpha} \otimes \boldsymbol{\beta} = (\boldsymbol{\alpha} \otimes \mathbf{I})\boldsymbol{\beta} = (\mathbf{I} \otimes \boldsymbol{\beta})\boldsymbol{\alpha}$, $\otimes$ is Kronecker product.

Let L find the partial derivatives of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively, and set the partial derivatives to zero

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \overline{\mathbf{C}}(\boldsymbol{\alpha} \otimes \mathbf{I})\boldsymbol{\beta} + (\boldsymbol{\alpha} \otimes \mathbf{I})^T \overline{\mathbf{C}}^T \boldsymbol{\beta} - c_\beta \boldsymbol{\beta} = 0, \tag{12}$$

$$\frac{\partial L}{\partial \boldsymbol{\alpha}} = (\mathbf{I} \otimes \boldsymbol{\beta})^T \overline{\mathbf{C}}^T \boldsymbol{\beta} - c_\alpha \boldsymbol{\alpha} = 0. \tag{13}$$

Let $\mathbf{Z}_s = \overline{\mathbf{C}}(\boldsymbol{\alpha} \otimes \mathbf{I})$, the above formula can be simplified as

$$\frac{\partial J}{\partial \boldsymbol{\beta}} = \mathbf{Z}_s \boldsymbol{\beta} + \mathbf{Z}_s^T \boldsymbol{\beta} - c_\beta \boldsymbol{\beta} = 0, \tag{14}$$

$$\frac{\partial J}{\partial \boldsymbol{\alpha}} = (\mathbf{I} \otimes \boldsymbol{\beta})^T \overline{\mathbf{C}}^T \boldsymbol{\beta} - c_\alpha \boldsymbol{\alpha} = 0. \tag{15}$$

Multiplying $\boldsymbol{\beta}^T$ on both sides of (12) and multiplying $\boldsymbol{\alpha}^T$ on both sides of (13), we have

$$c_\beta = \boldsymbol{\beta}^T (\mathbf{Z}_s^T + \mathbf{Z}_s)\boldsymbol{\beta} = L,$$

$$c_\alpha = \boldsymbol{\alpha}^T (\mathbf{I} \otimes \boldsymbol{\beta})^T \overline{\mathbf{C}}^T \boldsymbol{\beta} = L. \tag{16}$$

It can be seen from (16) that the value of L is equal to c_β and c_α. Defining $\mathbf{t}$ as

$$\mathbf{t} = \mathbf{C}\boldsymbol{\beta}, \tag{17}$$

where $\mathbf{t}_i = \mathbf{C}_i \boldsymbol{\beta}$ for $i = 1, 2, ..., s+1$. The following formula can be obtained

Algorithm 1: Time-series Algorithm

1: Normalize the initial data $\mathbf{X}$;
2: Randomly give the unit vector $\boldsymbol{\beta}$;
3: The expression of $\boldsymbol{\alpha}$ can be obtained;

$$\boldsymbol{\alpha} = \frac{1}{2}\{[(\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_s)^T \mathbf{t}_{s+1}] + [\mathbf{t}_{s+1}^T(\mathbf{t}_s, \cdots, \mathbf{t}_2, \mathbf{t}_1)]\}.$$

4: After $\boldsymbol{\alpha}$ is determined, $\boldsymbol{\beta}$ can be obtained;

$$\boldsymbol{\beta} = \sum_{i=1}^{s} \boldsymbol{\alpha}_i(\mathbf{C}_{s+1}^T \mathbf{t}_i + \mathbf{C}_i^T \mathbf{t}_{s+1}).$$

5: Unitize $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

$$\boldsymbol{\beta} = \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|}, \ \boldsymbol{\alpha} = \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}.$$

6: Update $\mathbf{C}$ as

$$\mathbf{q} = \mathbf{C}^T\mathbf{t}/\mathbf{t}^T\mathbf{t}; \ \mathbf{C} := \mathbf{C} - \mathbf{t}\mathbf{q}^T; \tag{19}$$

7: Go back to the third step and continue to extract variables until l variables are extracted.

$$c_\alpha \boldsymbol{\alpha} = \frac{1}{2}\{[(\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_s)^T \mathbf{t}_{s+1}] + [\mathbf{t}_{s+1}^T(\mathbf{t}_s, \cdots, \mathbf{t}_2, \mathbf{t}_1)]\},$$
$$c_\beta \boldsymbol{\beta} = \sum_{i=1}^{s} \boldsymbol{\alpha}_i(\mathbf{C}_{s+1}^T \mathbf{t}_i + \mathbf{C}_i^T \mathbf{t}_{s+1}). \tag{18}$$

It can be seen from (21) that when $\boldsymbol{\beta}$ is a certain value, $\boldsymbol{\alpha}$ can also be determined. The unit vector $\boldsymbol{\beta}$ can be randomly given, and the non-Gaussian fault information is extracted from the initial data. The implementation procedures are is given in Algorithm 1.

Remark 2: The number of latent variable l and the time delay variable s should be determined. If s has been determined, there is a particular functional relationship between l and s, expressed as $l = l(s)$. The selection principle of l is that the last calculated L accounts for 5% of the total L. In this paper, the value s and l are selected as 3 and 13, respectively.

Algorithm 1 can be understood as extracting a slow-change variable from the initial data in each loop. The slow-change data obtained through the loop can be considered unbiased, and the remaining residual data is related to noise. The slow-change score is $\mathbf{T} = [\mathbf{t}_1, \cdots, \mathbf{t}_l]$. The following formula is obtained by $\mathbf{T}$

$$\mathbf{T}_i = [\mathbf{t}_i, \ ..., \ \mathbf{t}_{l-s+i-1}] \ \text{ for } i = 1, \ 2, \ \cdots, \ s+1. \tag{20}$$

The $\mathbf{T}_1$, $\mathbf{T}_2$, and $\mathbf{T}_s$ can be used to predict the value of $\mathbf{T}_{s+1}$ through the ARMA model

$$\begin{aligned}\mathbf{T}_{s+1} &= \mathbf{T}_1\boldsymbol{\Xi}_s + \mathbf{T}_2\boldsymbol{\Xi}_{s-1} + \cdots + \mathbf{T}_s\boldsymbol{\Xi}_1 + \mathbf{V} \\ &= \overline{\mathbf{T}}_s\boldsymbol{\Xi} + \mathbf{V},\end{aligned} \tag{21}$$

where $\overline{\mathbf{T}}_s = [\mathbf{T}_1, \mathbf{T}_2, \cdots, \mathbf{T}_s]$ and $\boldsymbol{\Xi} = [\boldsymbol{\Xi}_s, \boldsymbol{\Xi}_{s-1}, \cdots, \boldsymbol{\Xi}_1]$.

To make the predicted value $\overline{\boldsymbol{\Xi}}$ unbiased, the residual should be minimized. We have

$$\boldsymbol{\Upsilon}_i = \mathbf{T}_{s+1} - \overline{\mathbf{T}}_s\overline{\boldsymbol{\Xi}}. \tag{22}$$

The $\hat{\mathbf{T}}_{s+1}$ is predicted

$$\hat{\mathbf{T}}_{s+1} = \overline{\mathbf{T}}_s\overline{\boldsymbol{\Xi}}, \tag{23}$$

where $\overline{\boldsymbol{\Xi}}$ can be expressed as

$$\overline{\boldsymbol{\Xi}} = (\overline{\mathbf{T}}_s^T\overline{\mathbf{T}}_s)^{-1}\overline{\mathbf{T}}_s^T\mathbf{T}_{s+1}. \tag{24}$$

The $\hat{\mathbf{X}}$ is further calculated as follows:

$$\hat{\mathbf{X}} = \hat{\mathbf{T}}_{s+1}\mathbf{Q}^T\mathbf{X}_{s+1}, \tag{25}$$
$$\hat{\mathbf{S}} = \mathbf{D}\mathbf{U}\hat{\mathbf{X}}, \tag{26}$$

where $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_l]$ is loading matrix in each $\mathbf{q}_i$ defined in (19), and $\hat{\mathbf{X}}$ is the non-Gaussian data estimated from the initial data $\mathbf{X}$. Equations (23) and (24) are explained in detail in Appendix A.

Remark 3: In this paper, the ARMA model is selected with an autoregressive parameter of 1, called the autoregressive model. A series of derivations are carried out on this basis.

3.2. Fault detection

SPE statistic is also called Q statistic, which is a particularly popular statistic. According to (26), it can be defined as

$$I^2 = \hat{\mathbf{S}}^T\hat{\mathbf{S}}, \tag{27}$$
$$Q = \mathbf{e}_k^T\mathbf{e}_k = (\hat{\mathbf{S}} - \mathbf{U}^{-1}(\mathbf{U}\mathbf{A})^T\hat{\mathbf{S}}). \tag{28}$$

Assuming $\mathbf{e}_k \sim \mathcal{N}(0, \Sigma_{\mathbf{e}_k})$, the distribution of Q can be approximately expressed by the following formula

$$Q \sim g\chi_h^2, \tag{29}$$

where χ_h^2 is the distribution of χ^2 with h degrees of freedom, and

$$g = \frac{\overline{Q^2} - \overline{Q}^2}{2\overline{Q}}, \ h = \frac{2\overline{Q}^2}{\overline{Q^2} - \overline{Q}^2}, \ \bar{\hat{\mathbf{S}}} = \frac{1}{N}\sum_{j=1}^{N}\hat{\mathbf{S}}(j),$$
$$\overline{Q^2} = \frac{1}{N}\sum_{j=1}^{N}\{(\hat{\mathbf{S}}(j) - \bar{\hat{\mathbf{S}}})^T(\hat{\mathbf{S}}(j) - \bar{\hat{\mathbf{S}}})\}^2,$$
$$\overline{Q} = \frac{1}{N}\sum_{j=1}^{N}(\hat{\mathbf{S}}(j) - \bar{\hat{\mathbf{S}}})^T(\hat{\mathbf{S}}(j) - \bar{\hat{\mathbf{S}}}). \tag{30}$$

Calculate the data of SPE statistics under normal conditions, then set the threshold J_{th} for a significance level μ

$$J_{th,Q} = g\chi_\mu^2(h). \tag{31}$$

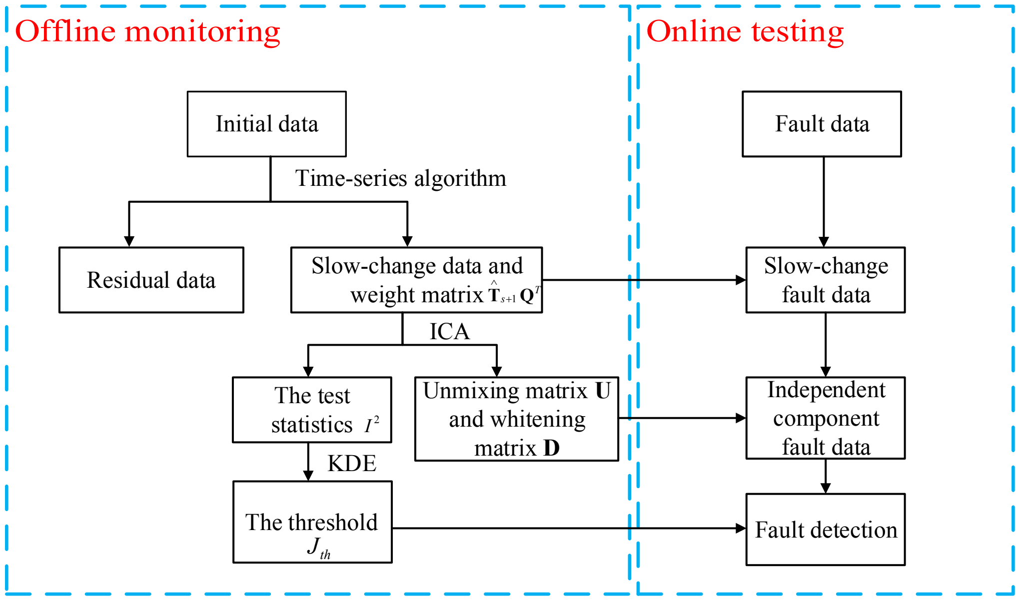

Fig. 2. Flow chart of TsICA.

The I^2 statistic and the Q statistic use the same method to set the threshold. By introducing risk assessment, the FD problem of TsICA is transformed into a test statistics problem. The specific test results are as follows:

$$\begin{aligned} &H_0: I^2 \le J_{th,I^2} \text{ and } Q \le J_{th,Q} \Longrightarrow \text{fault free}, \\ &H_1: I^2 > J_{th,I^2} \text{ or } Q > J_{th,Q} \Longrightarrow \text{faulty}, \end{aligned} \tag{32}$$

where J_{th,I^2} and $J_{th,Q}$ are the threshold for setting a boundary between the null hypothesis H_0 and alternative hypothesis H_1. The main idea of the TsICA algorithm is shown in Fig. 2. The specific detection process is as follows:

Offline detection:

Step 1: Collect normal operating data $\mathbf{X}$ in offline state;

Step 2: Normalize the initial data $\mathbf{X}$;

Step 3: Extract the slow-changing data $\hat{\mathbf{X}}$ from the normalized data $\mathbf{X}$ by Algorithm 1;

Step 4: The matrix $\hat{\mathbf{X}}$ is unmixed to obtain the matrix $\mathbf{S}$;

Step 5: The offline detection statistic I^2 and Q are obtained ;

Step 6: Determine the threshold J_{th} by kernel density estimation.

Online detection:

Step 1: Collect online data $\mathbf{X}_{on-line}$;

Step 2: Standardize data $\mathbf{X}_{on-line}$;

Step 3: Find the slow-changing fault data $\hat{\mathbf{X}}_{on-line}$;

Step 4: Find the slow-changing independent component matrix $\hat{\mathbf{S}}_{on-line}$;

Step 5: Obtain online detection statistics $I^2_{on-line}$ and $Q_{on-line}$;

Step 6: The fault is detected by comparing the detection statistics with the threshold.

4. VERIFICATIONS

4.1. Experimental vertification

To improve the reliability of the method proposed in this paper, a testing bench is introduced for simulating the running gears of high-speed trains. The parameter test-bed is developed by Changchun Railway Rolling Stock Co., Ltd, shown in Fig. 3.

As everyone knows, the running speed of high-speed train will directly affect the temperature and vibration data collected by sensors. In order to collect accurate data, the speed of the traction motor of the running part should be above 950 r/min. At this time, the train runs stably, and

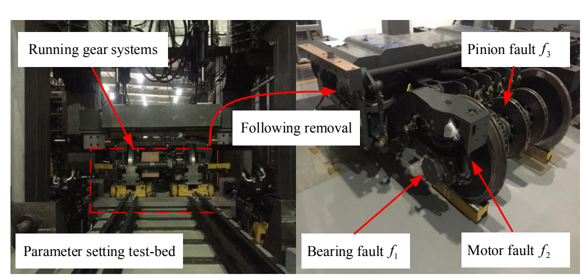

Fig. 3. Operation platform and fault position of running gear in high-speed trains.

Table 1. The physical meaning of variables and the range of values in different states.

Variable	Position	Normal range/unit	Abnormal range/symbol/unit
x_1	Motor non-drive side	35-49/°C	50-65/ f_1 /°C
x_2	Bearing	60-90/°C	91-105 / f_2 /°C
x_3	Pinion box	14-16/Hz	17-19 / f_3 /Hz
x_4	Side of bogie	25-40/°C	- -

Table 2. Test results of running gear system.

Method	f_1		f_2		f_3	
	MAR	FAR	MAR	FAR	MAR	FAR
DPCA	6.43%	3.13%	5.00%	3.75%	1.43%	3.44%
DICA	1.44%	1.88%	36.44%	**1.25%**	1.44%	1.88%
The proposed method based on TsICA	**0.72%**	**0.94%**	**0%**	1.88%	**0%**	**1.56%**

the collected data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\} \in \Re^{600\times 4}$ has stable characteristics. The physical meaning and value range of the variables are listed in Table 1. It should be noted that the number of sampling samples is 600 (where 300 is the training data and 300 is the test data), and the sampling interval is $t = 1$ min.

4.1.1 Fault injection

Motor faults f_1, bearing faults f_2 and pinion box faults f_3 are respectively injected into the running gear systems in Fig. 3. The time of fault injection is 321 min in the test data. The three types of faults fluctuate in the fault range corresponding to Table 1.

4.1.2 Fault detection

As shown in Fig. 4, the detection results of TsICA is satisfactory when the fault is not injected. Figs. 5-7 are the detection results of TsICA after fault injection. Both I^2 and SPE statistics effectively detect the fault. Where the red dashed line and the solid blue line are thresholds and detection statistics. The detected missing alarm rate (MAR) and fault alarm rate (FAR) values are listed in Table 2. It needs to be pointed out that the MAR detected by

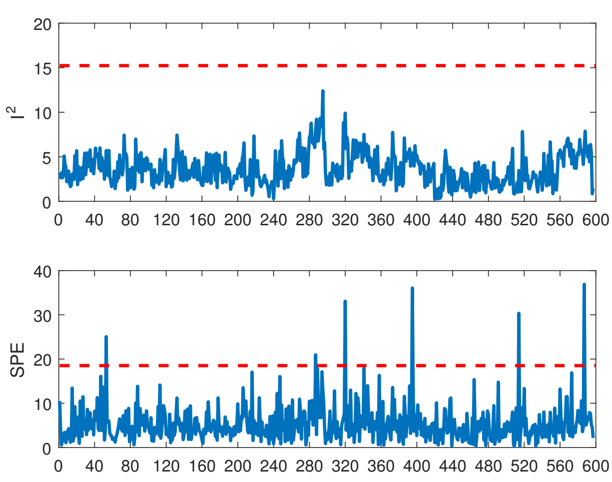

Fig. 4. Detection result under fault-free conditions.

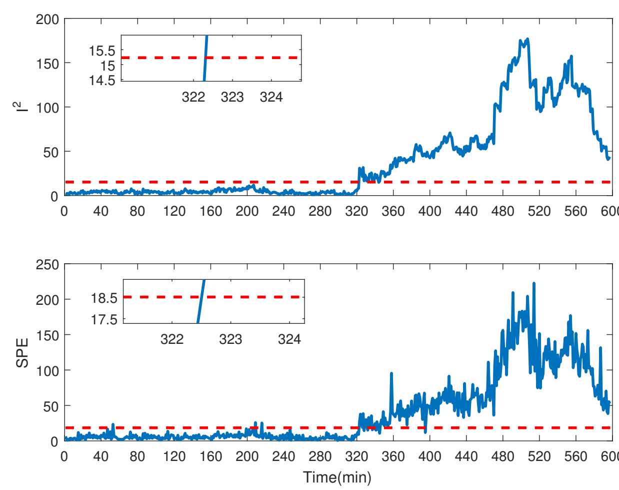

Fig. 5. Detection results of f_1 by TsICA.

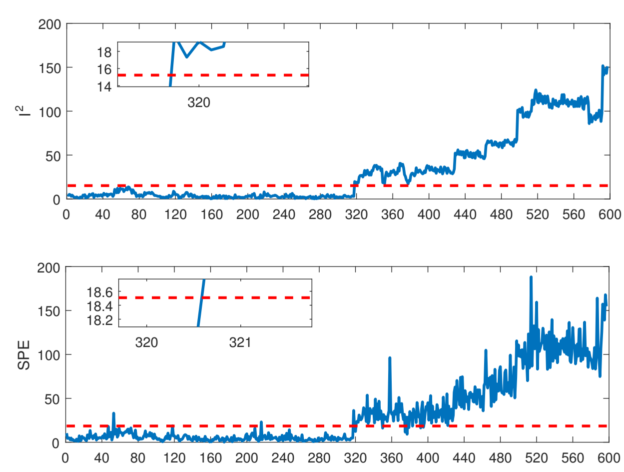

Fig. 6. Detection results of f_2 by TsICA.

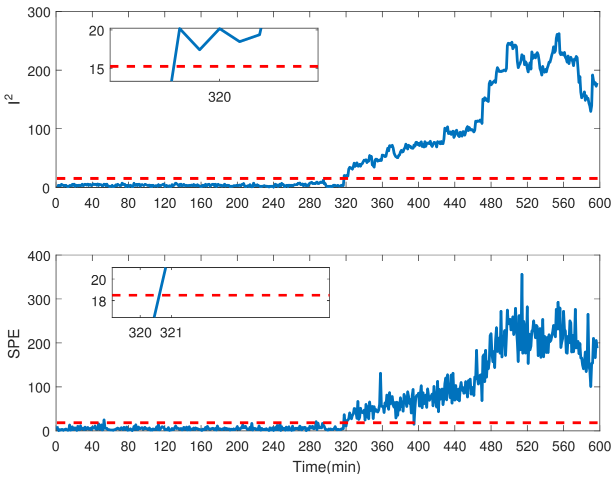

Fig. 7. Detection results of f_3 by TsICA.

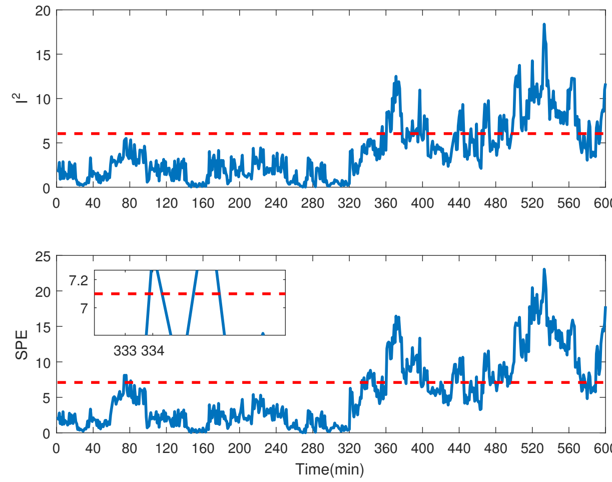

Fig. 9. Detection results of f_2 by DICA.

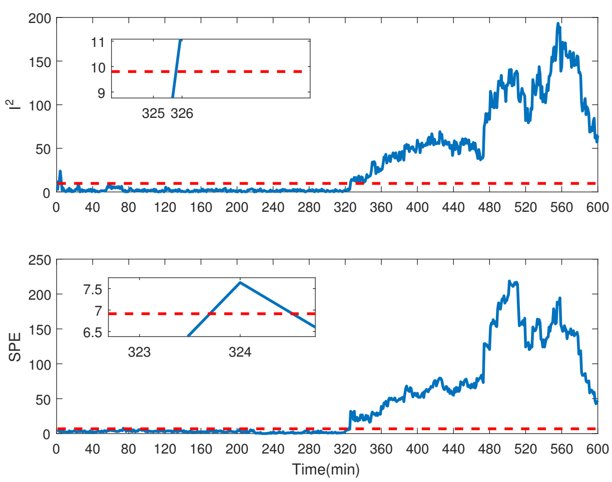

Fig. 8. Detection results of f_1 by DICA.

Table 3. Time when the fault is first detected

Method	Detection time (min)		
	f_1	f_2	f_3
DPCA	327	**321**	326
DICA	324	334	325
The proposed based on TsICA	**323**	**321**	**321**

TsICA is less than 1%, and the MAR is in the range of (0%, 2%). In addition, For the three types of failures, the first detection time is showed in Table 3. Next, analyze the three types of faults in detail:

- Fault f_1 is a motor fault caused by the eccentricity of the motor rotor. As shown in Fig. 5, after 500 min, the detection statistics showed a downward trend. The main reason is that due to the inherent fault tolerance of running gear systems, the impact of the fault on the system is constantly being weakened.
- Fault f_2 is a bearing fault caused by corrosion of parts in the bearing. According to the test results shown in Fig. 6, TsICA has achieved very satisfactory test results. Specifically, when a fault occurs, it is immediately detected by I^2 and SPE statistics.
- Fault f_3 is a pinion box fault caused by insufficient lubrication between parts. As shown in Fig. 7, after the fault is injected, I^2 and SPE statistics fluctuate above the threshold over time.

4.1.3 Comparative analysis

This paper compares the two data-driven FD methods, DPCA and DICA, with the TsICA method to prove the superiority of this method. A simple comparative analysis of the generated result graphs is carried out to illustrate the superiority of TsICA better.

- The detection results of the three faults by DICA are shown in Figs. 8-10. For f_2, I^2 and SPE do not detect the fault in time and showed a fluctuating state. Therefore, ICA is less sensitive to slow-change faults than TsICA.
- Figs. 11-13 are the detection results of DPCA for three faults. The MAR value is as high as 6.43% in f_1, which reduces the credibility of the method and cannot meet the actual demand.

4.2. Discussions

The discussion is divided into two main parts: 1) Comparative analysis based on MAR and FAR; 2) Whether FAR value can be reduced to 0.

4.2.1 Detection efficiency

For the three types of faults, the first detection time of the method proposed in this paper is 323 min, 321 min, and 321 min, respectively, which are better than DPCA and ICA. This result is achieved by extracting slow-change information from the data by TsICA.

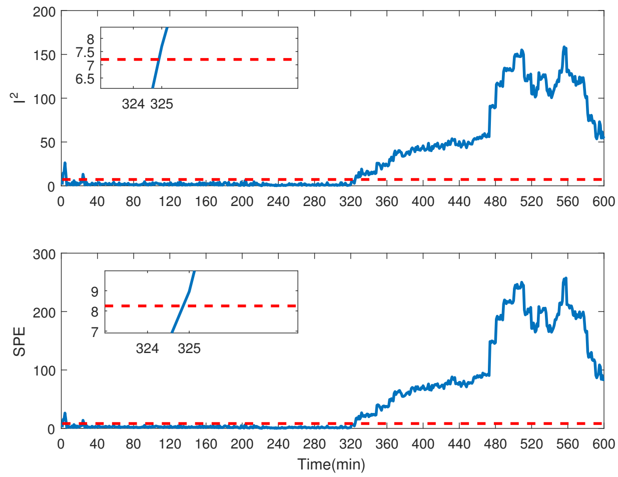

Fig. 10. Detection results of f_3 by DICA.

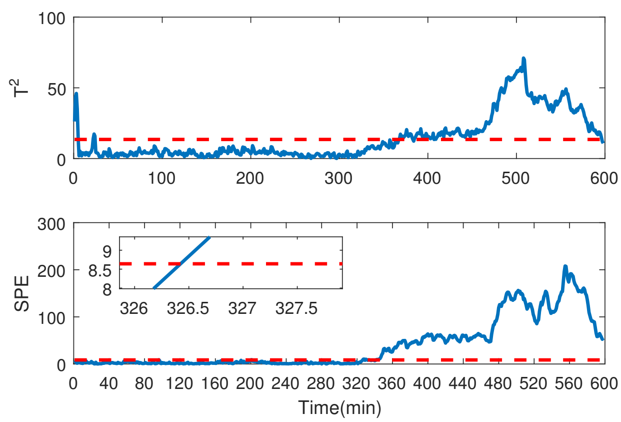

Fig. 11. Detection results of f_1 by DPCA.

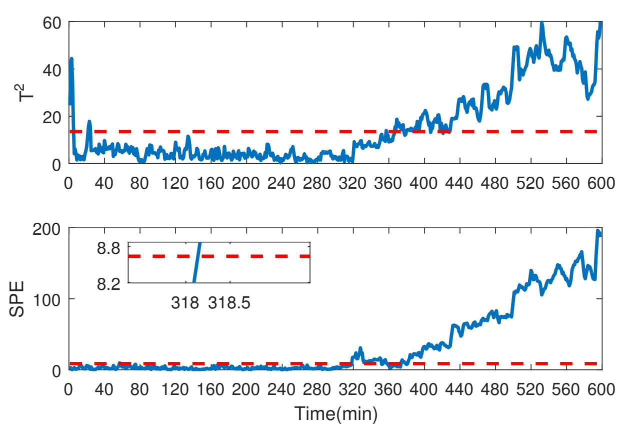

Fig. 12. Detection results of f_2 by DPCA.

4.2.2 The value of false alarm rate

Under ideal conditions, FAR will be infinitely close to 0. In this article, FAR (FAR $\in (0\%, 2\%)$) may be caused by the following two situations: 1) The system is in a non-stationary state; 2) The time-series algorithm covers a small amount of noisy data during multiple iterations of extraction.

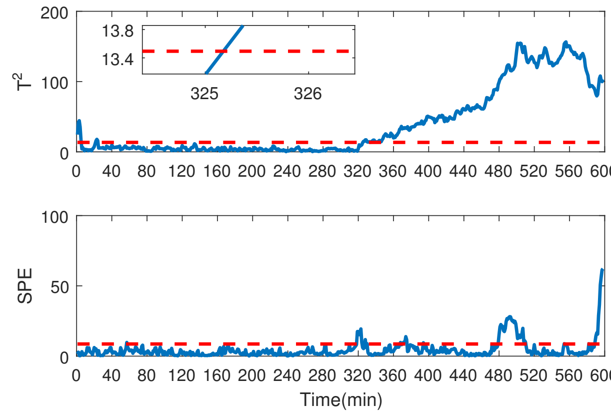

Fig. 13. Detection results of f_3 by DPCA.

5. CONCLUSIONS

This paper proposes a TsICA algorithm to solve the FD problem of running gear systems on a high-speed train. TsICA comprehensively analyzes the slow-change data of running gear systems and overcomes the problem that the fault data and the noise data are difficult to separate. The scientificity of TSICA is proved by rigorous formula derivation. The effectiveness of TsICA is fully verified on the parameter test bench of running gear systems. It needs to be pointed out that 1) Compared with traditional methods, TsICA is more sensitive to slow-change faults; 2)The MAR of TsICA is better than DICA and DPCA for the three types of failures, and the FAR value is less than 2%; 3) The first fault detection time is better than traditional methods, which makes TsICA more suitable for running gear systems; 4) The FD method proposed in this paper is based on data-driven, highly feasible and flexible so that it can be extended to other industrial processes.

APPENDIX A: MATHEMATICAL VERIFICATIONS OF (23) AND (24)

Proof: The sun of errors is $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}_1^2 + \boldsymbol{\Upsilon}_2^2 + \ldots + \boldsymbol{\Upsilon}_s^2, \Upsilon$ can be written as

$$\boldsymbol{\Upsilon} = (\mathbf{T}_{s+1} - \overline{\mathbf{T}}_s \overline{\boldsymbol{\Xi}}^T)(\mathbf{T}_{s+1} - \overline{\mathbf{T}}_s \overline{\boldsymbol{\Xi}}).$$

Let the first-order deviation of $\boldsymbol{\Upsilon}$ to $\overline{\boldsymbol{\Xi}}$ be 0

$$\frac{\partial \boldsymbol{\Upsilon}}{\partial \overline{\boldsymbol{\Xi}}} = -\mathbf{T}_{s+1}^T \overline{\mathbf{T}}_s - \mathbf{T}_{s+1}^T \overline{\mathbf{T}}_s + 2\overline{\boldsymbol{\Xi}}^T \overline{\mathbf{T}}_s^T \overline{\mathbf{T}}_s = 0.$$

The estimated value by least squares is

$$\overline{\boldsymbol{\Xi}} = (\overline{\mathbf{T}}_s^T \overline{\mathbf{T}}_s)^{-1} \overline{\mathbf{T}}_s^T \mathbf{T}_{s+1},$$
$$\hat{\mathbf{T}}_{s+1} = \overline{\mathbf{T}}_s \overline{\boldsymbol{\Xi}}.$$

Hence, (23) and (24) can be obtained.

REFERENCES

[1] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1700-1716, 2022.

[2] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 11, pp. 6418-6428, 2014.

[3] D. Zhou, Y. Zhao, Z. Wang, X. He, and M. Gao, "Review on diagnosis techniques for intermittent faults in dynamic systems," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2337-2347, 2020.

[4] L. Yin, P. Zhu, and T. Li, "Fault detection and diagnosis for delay-range-dependent stochastic systems using output PDFs," *International Journal of Control, Automation, and Systems*, vol. 15, no. 4, pp. 1701-1709, 2017.

[5] C. Cheng, X. Qiao, H. Luo, G. Wang, W. Teng, and B. Zhang, "Data-driven incipient fault detection and diagnosis for the running gear in high-speed trains," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9566-9576, 2020.

[6] J. M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 59, no. 1, pp. 223-234, 2004.

[7] L. Rong, X. Peng, and B. Zhang, "A reduced-order fault detection filter design for polytopic uncertain continuous-time Markovian jump systems with time-varying delays," *International Journal of Control, Automation, and Systems*, vol. 16, no. 8, pp. 2021-2032, 2018.

[8] C. Yang, C. Yang, T. Peng, X. Yang, and W. Gui, "A fault-injection strategy for traction drive control systems," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5719-5727, 2017.

[9] H. Henao, S. H. Kia, and G.-A. Capolino, "Torsional-vibration assessment and gear-fault diagnosis in railway traction system," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1707-1717, 2011.

[10] B. Zhou and X. Gu, "Multi-block statistics local kernel principal component analysis algorithm and its application in nonlinear process fault detection," *Neurocomputing*, vol. 376, no. 1, pp. 222-231, February 2020.

[11] C. Cheng, W. Wang, H. Chen, B. Zhang, J. Shao, and W. Teng, "Enhanced fault diagnosis using broad learning for traction systems in high-speed trains," *IEEE Transactions on Power Electronics*, vol. 36, no. 7, pp. 7461-7469, 2020.

[12] Q. Zhu, L. Li, C. Chen, C. Liu, and G. Hu, "A low-cost lateral active suspension system of the high-speed train for ride quality based on the resonant control method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4187-4196, 2018.

[13] X. Lu, Z. Liu, J. Zhang, H. Wang, Y. Song, and F. Duan, "Prior-information-based finite-frequency H_∞ control for active double pantograph in high-speed railway," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 8723-8733, 2017.

[14] H. Chen, Z. Chen, Z. Chai, B. Jiang, and B. Huang, "A single-side neural network-aided canonical correlation analysis with applications to fault diagnosis," *IEEE Transactions on Cybernetics*, pp. 1-13, 2021. DOI: 10.1109/TCYB.2021.3060766

[15] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 450-465, 2020.

[16] E. Tian and C. Peng, "Memory-based event-triggering H_∞ load frequency control for power systems under deception attacks," *IEEE Transactions on Cybernetics*, vol. 50, no. 11, pp. 4610-4618, 2020.

[17] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 6, pp. 2526-2540, 2019.

[18] P. Nomikos and J. F. MacGregor, "Monitoring batch processes using multiway principal component analysis," *AIChE Journal*, vol. 40, no. 8, pp. 1361-1375, 1994.

[19] S. Yin, G. Wang, and H. Gao, "Data-driven process monitoring based on modified orthogonal projections to latent structures," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1480-1487, 2016.

[20] G. Lee, S.-O. Song, and E. S. Yoon, "Multiple-fault diagnosis based on system decomposition and dynamic PLS," *Industrial & Engineering Chemistry Research*, vol. 42, no. 24, pp. 6145-6154, 2003.

[21] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, June 2000.

[22] J. Yu, J. Yoo, J. Jang, J. H. Park, and S. Kim, "A novel hybrid of auto-associative kernel regression and dynamic independent component analysis for fault detection in nonlinear multimode processes," *Journal of Process Control*, vol. 68, no. 4, pp. 129-144, 2018.

[23] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 1, pp. 122-135, 2017.

[24] S. Zhang and C. Zhao, "Hybrid independent component analysis (H-ICA) with simultaneous analysis of high-order and second-order statistics for industrial process monitoring," *Chemometrics and Intelligent Laboratory Systems*, vol. 185, pp. 47-58, 2019.

[25] Q. Jiang, F. Gao, X. Yan, and H. Yi, "Multiobjective two-dimensional CCA-based monitoring for successive batch processes with industrial injection molding application," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3825-3834, 2019.

[26] Z. Chen, C. Yang, T. Peng, H. Dan, C. Li, and W. Gui, "A cumulative canonical correlation analysis-based sensor precision degradation detection method," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 8, pp. 6321-6330, 2019.

[27] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang, and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3721-3730, 2020.

[28] H. Luo, S. Yin, T. Liu, and A. Q. Khan, "A data-driven realization of the control-performance-oriented process monitoring system," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 521-530, 2020.

[29] Y. Dong and S. J. Qin, "Dynamic latent variable analytics for process operations and control," *Computers & Chemical Engineering*, vol. 114, pp. 69-80, 2018.

[30] C. Cheng, J. Wang, Z. Zhou, W. Teng, Z. Sun, and B. Zhang, "A BRB-based effective fault diagnosis model for high-speed trains running gear systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 110-121, 2022.

[31] H. Chen, B. Jiang, W. Chen, and H. Yi, "Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4716-4725, 2019.

Chao Cheng received his Ph.D. degree from Jilin University, Changchun, China, in 2014. He is currently a Teacher with the Changchun University of Technology, Changchun. He has been a Post-Doctoral Fellow in process control engineering with the Department of Automation, Tsinghua University, Beijing, China, since 2018. He has also been a Post-Doctoral Fellow with the National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., China, since 2018. His research interests include dynamic system fault diagnosis and predictive maintenance, wireless sensor network, artificial intelligence, and data-driven method.

Sheng Yang received his B.Eng. degree from Wuhan East Lake University, Wuhan, China, in 2014. He is currently working toward an M.Eng. degree in computer science and engineering with the Changchun University of Technology. His research interests include complex system fault diagnosis, health status estimation, and data-driven fault detection and diagnosis.

Yu Song received his B.S. degree from Changchun University of technology, and received an M.S. degree from Shenyang University of technology. He currently works in Changchun University of technology. He is mainly engaged in fault detection, embedded systems, and UAVs.

Gang Liu is a teacher at Changchun University of Technology. He mainly teaches courses such as Introduction to Computer Subjects; Software Requirements; Software Design Patterns and Software Architecture. He is mainly engaged in data mining, distributed systems, and software engineering.