Article

Disturbance Detection of a Power Transmission System Based on the Enhanced Canonical Variate Analysis Method

Shubin Wang [1], Yukun Tian [1,2], Xiaogang Deng [1,*] , Qianlei Cao [2], Lei Wang [2] and Pengxiang Sun [2]

1 College of Control Science and Engineering, China University of Petroleum, Qingdao 266580, China; shubinw@126.com (S.W.); tianyk42085@hundsun.com (Y.T.)
2 Qingdao Topscomm Communication Co., Ltd., Qingdao 266109, China; caoqianlei@topscomm.com (Q.C.); wanglei5@topscomm.com (L.W.); sunpengxiang@topscomm.com (P.S.)
* Correspondence: dengxiaogang@upc.edu.cn

Abstract: Aiming at the characteristics of dynamic correlation, periodic oscillation, and weak disturbance symptom of power transmission system data, this paper proposes an enhanced canonical variate analysis (CVA) method, called SLCVAkNN, for monitoring the disturbances of power transmission systems. In the proposed method, CVA is first used to extract the dynamic features by analyzing the data correlation and establish a statistical model with two monitoring statistics T^2 and Q. Then, in order to handling the periodic oscillation of power data, the two statistics are reconstructed in phase space, and the k-nearest neighbor (kNN) technique is applied to design the statistics nearest neighbor distance DT^2 and DQ as the enhanced monitoring indices. Further considering the detection difficulty of weak disturbances with the insignificant symptoms, statistical local analysis (SLA) is integrated to construct the primary and improved residual vectors of the CVA dynamic features, which are capable to prompt the disturbance detection sensitivity. The verification results on the real industrial data show that the SLCVAkNN method can detect the occurrence of power system disturbance more effectively than the traditional data-driven monitoring methods.

Keywords: canonical variate analysis; disturbance detection; power transmission system; k-nearest neighbor analysis; statistical local analysis

1. Introduction

With the increasing demand on the power energy in the modern industry, power transmission systems are becoming more and more large-scale and complicated [1,2]. Due to the system complexity, anomalies and disturbances are often unavoidable in real power systems. If these unexpected events are not handled timely, they may cause huge accident risks and even the widespread power outages, which are companied by the huge economic loss and severe life inconvenience. Therefore, it is of great value to detect the abnormal events quickly and maintain the safe running of power systems [3]. In recent years, the wide area measurement system (WAMS) based on synchronous phaser technology has been successfully applied in the power industry. The phasor measurement units in WAMS provide the basic data support for the real-time dynamic monitoring of the power system [4]. Accordingly, safety monitoring and disturbance detection of power systems based on the measurement data analysis has been a hot topic in academic and engineering fields [5–7].

Aiming at the power system disturbance detection task, researchers have conducted a lot of studies, which can be roughly divided into two categories: time/frequency domain analysis and multivariate statistical analysis. The time/frequency domain analysis investigates the power system changes from the perspective of the signal processing, which involves the time domain, frequency domain, or time-frequency domain. In consideration of the good time-frequency localization property, Huang et al. [8] discussed the application of the Morelet wavelets method in power system disturbance detection. The Hilbert Huang

Transform is another time-frequency signal analysis tool. Manglik et al. [9] applied it to the disturbance detection for the electric power system. Ghaderi et al. [10] proposed the time-frequency analysis method assisted by current waveform energy and normalized joint time-frequency moment and demonstrated its performance in the high-impedance ground fault detection. Salehi et al. [11] designed a morphological edge detection filter to obtain the transient features of fault signals. Liu et al. [12] used the wavelet packet Tsallis singularity entropy algorithm for disturbance detection. In general, the time/frequency domain analysis methods mainly analyze the single signal and fail to fully consider the correlation between different parameters. In response to this shortcoming, some scholars started their work by applying multivariate statistical analysis. Multivariate statistical analysis (MSA) methods can realize the simultaneous detection of multiple parameter changes and have outstanding advantages in the complex industrial systems. However, most of the present MSA studies focus on the system modeling and disturbance detection in the chemical process, steel industry, and high-train system [13–17], but MSA's application to power system monitoring is very rare. Barocio et al. [18] first introduced the principal component analysis (PCA) method into the field of power system monitoring and discussed the detection and visualization of power system disturbances based on PCA. Guo et al. [19] built a transmission line fault detection method by combining PCA and support vector machine. Considering the masking influence caused by the oscillation trend and strong noise of power system data, Cai et al. [20] further proposed a PCAkNN method, which is superior to the basic PCA method in the numerical model testing and New England power system model data. These research articles point out that the multivariate statistical analysis has great application potential in the field of power system monitoring.

Although PCA and PCAkNN methods have achieved significant success in the power system monitoring field, they have some shortcomings deserving further studies. On the one hand, these methods do not take into account the dynamic characteristics of power system data, which easily leads to a high missing detection rate. Different from the other industrial process data with the steady operation mode, the power system data, such as the voltage and current, are with obvious dynamic trends. On the other hand, the present methods do not consider how to enhance the detection of weak disturbances. In real applications, some disturbances may be with small amplitudes, slow changes, unclear disturbance characteristics, and are easy to be covered by noises [21,22]. How to enhance the detection capability on these weak disturbances is one challenging task.

Aiming at the aforementioned problems, this paper proposes a SLCVAkNN-based disturbance detection method for power transmission system monitoring by combining canonical variate analysis (CVA), kNN, and statistical local analysis (SLA). Compared with the traditional PCA-based power system monitoring methods, CVA has a stronger dynamic feature extraction ability [23–25], which provides a new and powerful tool for power system data analysis. Referring to the present PCAkNN method, the CVAkNN statistical model is developed to deal with the dynamic periodic oscillation signals. Furthermore, in order to enhance the detection of weak faults, SLA is integrated for SLCVAkNN modeling, which mines the local statistical information for better weak disturbance monitoring.

The rest of the paper's content is arranged as follows. The principle of the proposed SLCVAkNN methodology is given in the Section 2, while the corresponding disturbance detection procedure is detailed in Section 3. One case study on the actual industrial data is used to verify the effectiveness of the proposed method.

2. The Proposed Methodology

This section clarifies the proposed SLCVAkNN-based power system disturbance detection method. The whole methodology involves three parts: dynamic system modeling using canonical variate analysis, monitoring index construction based on kNN, and weak disturbance detection by statistical local analysis.

2.1. CVA Monitoring Model

A power system is a classical dynamic process [26,27], where the measurement data demonstrate the clear trend along the sampling time. The measured three-phase electric field and current waveform change with time, and the current data point has a certain correlation with the historical samples. Therefore, it is more reasonable to apply the dynamic data analysis tool to extract the process features.

Canonical variate analysis (CVA) is an effective dynamic data analysis tool, which has been applied to the model identification and control in the multivariate dynamic system [28,29]. This paper introduces it to deal with power system data. For a certain power transmission line, the data points under normal system operation have a fixed correlation along the time dimension. When a disturbance occurs, this correlation may be destroyed. By monitoring the correlation among the time series data, CVA can find the system disturbance effectively. When CVA is applied to data modeling, the training data are firstly divided into the historical data set and the future data set, and the CVA optimization problem is designed to find the maximum correlation between these two data sets for describing the data dynamic features. The algorithm details are clarified as follows.

For the power system measurement data vector $x_h \in R^m$ at the h-th sampling instant, its corresponding historical data vector p_h and future data vector f_h are constructed as

$$p_h = [x_h^T, x_{h-1}^T, \cdots, x_{h-l+1}^T]^T \in R^M \tag{1}$$

$$f_h = [x_{h+1}^T, x_{h+2}^T, \cdots, x_{h+l}^T]^T \in R^M \tag{2}$$

where $M = m \times l$, and l represents the time lag order.

Given the projection vectors a and b, they are used to transform the historical and future vectors into their respective projections $d = a^T p_h$ and $v = b^T f_h$. CVA is to optimize the vector pair a and b so that the correlation between d and v is maximized, which are also called canonical variates. This can be described by the mathematical expression as

$$\begin{cases} \max_{a,b} \rho(d,v) = a^T \Sigma_{pf} b \\ \text{s.t. } var(d) = a^T \Sigma_{pp} a = 1 \\ \quad var(v) = b^T \Sigma_{ff} b = 1 \end{cases} \tag{3}$$

where Σ_{pf} represents the cross-covariance matrix of the historical and future data vectors, and Σ_{pp}, Σ_{ff} denote the covariance matrix of the historical and future data vectors, respectively.

Suppose that the training data set includes n samples as $X = [x_1^T, x_2^T, \cdots, x_n^T]^T \in R^{n \times m}$, then the historical and future data matrix can be expressed by

$$P = [p_l^T, p_{l+1}^T, \cdots, p_{n-l}^T]^T \in R^{N \times M} \tag{4}$$

$$F = [f_l^T, f_{l+1}^T, \cdots, f_{n-l}^T]^T \in R^{N \times M} \tag{5}$$

where $N = n - 2l + 1$ is the sample number of the historical and future data matrix. Then the covariance matrices defined in Equation (3) can be calculated by

$$\Sigma_{pf} = \frac{1}{N-1} P^T F \tag{6}$$

$$\Sigma_{pp} = \frac{1}{N-1} P^T P \tag{7}$$

$$\Sigma_{ff} = \frac{1}{N-1} F^T F \tag{8}$$

Solving the optimization problem described by Equation (3) leads to a singular value decomposition on the matrix $\Xi = \Sigma_{pp}^{-1/2}\Sigma_{pf}\Sigma_{ff}^{-1/2}$, which is expressed by

$$\Xi = U\Lambda V^T \tag{9}$$

The solution of Equation (9) is further used to build the a series of the projection vectors a_i and $b_i (1 \le i \le M)$, which are computed by

$$a_i = \Sigma_{pp}^{-1/2}U(:,i), \ b_i = \Sigma_{ff}^{-1/2}V(:,i), \tag{10}$$

where $(:,i)$ represents the i-th column of the matrix. The vectors a_i and b_i are ordered by the corresponding correlation degree, which is given in the diagonal elements of matrix Λ, also meaning the correlation coefficients. The first s pairs of projection vectors $\{a_i, b_i, 1 \le i \le s\}$ describe the stronger correlation and indicate the close relationship between the historical data and the future data. Therefore, a projection matrix $A_s = [a_1 a_2 \cdots a_s]$ is defined to extract the canonical variate vector d_h as

$$d_h = A_s^T p_h. \tag{11}$$

which describes the main dynamic features of process data. Here, s is determined so that the corresponding canonical variate vectors describe a cumulative percentage of 90% of correlation coefficients.

As A_s only involves the first s projection directions, it cannot cover all the data information. The rest information in the CVA model can be described by the CVA residual vector e_h as

$$e_h = (I - A_s A_s^T)p_h \tag{12}$$

Based on the canonical variate vector and CVA residual vector, two monitoring statistics T^2 and Q are often used to judge the process state. The T^2 statistic describes the changes of principal dynamic states, while the Q statistic monitors the changes of the residual information. For the h-th sample, the statistics are written by

$$T_h^2 = d_h^T d_h \tag{13}$$

$$Q_h = e_h^T e_h \tag{14}$$

In the normal operation, these two statistics should satisfy $T_h^2 \le T_{h,lim}^2$ and $Q_h \le Q_{h,lim}$, where $T_{h,lim}^2$ and $Q_{h,lim}$ are the corresponding confidence limits. In some literature, the confidence limits of these two statistics can be obtained by assuming the prior distribution [30]. However, these distribution assumptions are often difficult to satisfy. Therefore, this paper applies the data-driven kernel density estimation to determine the confidence limit [31,32].

2.2. CVAkNN Model Based on kNN Monitoring Index

As the measurement data of power transmission systems have the periodic fluctuation characteristic, the traditional CVA monitoring statistics T^2 and Q behave unsteadily with the periodic changes. In this case, disturbance detection by directly monitoring the amplitudes of monitoring statistics cannot discover the disturbance signals effectively and may lead to a high disturbance missing rate.

In order to overcome this defect, this paper introduces the k-nearest neighbor analysis (kNN) to enhance the basic monitoring statistics. kNN is one effective multimodal data analysis tool and does not depend on the amplitude changes before and after the disturbance. In the literature [33,34], kNN was introduced and adapted for real-time detection of system disturbances. By combining the CVA model and the kNN-based monitoring statistics, the improved method, which is called CVAkNN, has a stronger capability of dealing with the periodic oscillation data property. The main idea of CVAkNN is to first

reconstruct the monitoring statistic in the phase space and then build the monitoring index based on the distance between the reconstructed statistic vector and its k-nearest neighbor.

Phase space reconstruction is a good method to deal with time series analysis. This method regards one-dimensional time series as the result of nonlinear dynamic system motion and constructs the phase vectors by re-arranging the time series. This theory has been successfully applied in the fields of chaotic time series prediction and equipment failure data analysis [35,36]. Here it is introduced to deal with the CVA monitoring statistics for the further kNN analysis.

For the training data set with n samples $x_1, x_2, \ldots, x_n$, the corresponding statistics vectors are obtained by the CVA modeling as

$$\boldsymbol{T}^2 = [T_l^2 \ T_{l+1}^2 \ \cdots \ T_n^2] \tag{15}$$

$$\boldsymbol{Q} = [Q_l \ Q_{l+1} \ \cdots \ Q_n] \tag{16}$$

Further, the phase reconstruction statistics matrix can be formulated as follows:

$$\boldsymbol{MT}^2 = \begin{bmatrix} T_l^2 & \cdots & T_{l+L-2}^2 & T_{l+L-1}^2 \\ T_{l+1}^2 & \cdots & T_{l+L-1}^2 & T_{l+L}^2 \\ \vdots & \vdots & \vdots & \vdots \\ T_{n-L+1}^2 & \cdots & T_{n-1}^2 & T_n^2 \end{bmatrix} \tag{17}$$

$$\boldsymbol{MQ} = \begin{bmatrix} Q_l & \cdots & Q_{l+L-2} & Q_{l+L-1} \\ Q_{l+1} & \cdots & Q_{l+L-1} & Q_{l+L} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{n-L+1} & \cdots & Q_{n-1} & Q_n \end{bmatrix} \tag{18}$$

where L is the embedding dimension defining the length of the reconstructed phase vector. Based on the results of the phase space reconstruction, the dynamic behavior of the statistics can be better described, which is conducive to the detection of power system disturbances.

In the online monitoring stage, a new testing sample x_t is collected at the t-th sampling instant. Then the monitoring statistics can be computed by applying Equations (13) and (14), and the reconstructed phase vectors are described as

$$\boldsymbol{NT}_t^2 = \left[T_{t-L+1}^2 \ \cdots \ T_{t-1}^2 \ T_t^2 \right] \tag{19}$$

$$\boldsymbol{NQ}_t = [Q_{t-L+1} \ \cdots \ Q_{t-1} \ Q_t] \tag{20}$$

To determine whether the test data x_t is normal, it is necessary to compare the similarity between $\boldsymbol{NT}_t^2, \boldsymbol{NQ}_t^2$, and the reconstructed statistics matrix in Equations (17) and (18). If the reconstructed statistics $\boldsymbol{NT}_t^2, \boldsymbol{NQ}_t^2$ are strongly similar to one column of the training statistics vectors in Equations (17) and (18), then the test data x_t describe the normal working condition. Otherwise, it means that some faults occur in the power transmission system. Therefore, the key is how to perform this similarity comparison. This paper introduces the k-nearest neighbor (kNN) analysis to construct a kNN-based distance measurement indicator: statistical nearest neighbor distance (SNND).

The idea of SNND is to find the first k-th nearest neighbors of the test vector in the given matrix data and compute the distance between the test vector and the k-th nearest neighbors as a disturbance detection criterion. The SNND index for $\boldsymbol{NT}_t^2$ is defined as

$$DT_t^2 = \left\| \boldsymbol{NT}_t^2 - \boldsymbol{MT}^2(j_k, :) \right\|, \tag{21}$$

where $\boldsymbol{MT}^2(j_k, :)$ represents the j_k-th row in the $\boldsymbol{MT}^2$ matrix, which corresponds to the k-th nearest neighbor of $\boldsymbol{NT}_t^2$, and $||.||$ represents the L2 norm calculation. By analogy, the SNND indicator of NQ_t can be established as

$$DQ_t = ||\boldsymbol{NQ}_t - \boldsymbol{MQ}(j_k, :)||. \tag{22}$$

Under normal operating conditions, the above two indicators should fluctuate within a relatively small range. That means $DT_t^2 \leq DT_{lim}^2$ and $DQ_t \leq DQ_{lim}$ for the normal running status. Once the threshold is exceeded, it means that there is a system disturbance. The threshold can be obtained by the kernel density estimation method.

2.3. SLCVAkNN Model Assisted by Statistical Local Analysis

In the power transmission system, some weak disturbances are often difficult to detect, such as the high-impendence single-phase ground fault. When this kind of disturbance occurs, the changes reflected by the measure voltage and current variables are very small. Further, considering the influence of modeling error and process noise, this kind of disturbance may be concealed and viewed as the normal process changes. Therefore, enhacning the weak disturbance detection is of great value to ensure the safety of power transmission systems. In this paper, we integrate the statistical local analysis (SLA) with CVAkNN and propose an improved SLCVAkNN monitoring model for better weak disturbance monitoring performance.

SLA was originally proposed by Basseville [37] for inspecting the process parameter changes. In recent years, some researchers have introduced it into the chemical process fault detection and demonstrated its effectiveness [38–40]. In this paper, we will perform the statistical local analysis on the CVA model. To look back into the CVA monitoring statistics in Equations (13) and (14), it is found that the monitoring statistics used to indicate the process status are composed of the canonical variate vector $\boldsymbol{d}_h$ and the CVA residual vector $\boldsymbol{e}_h$. Therefore, if we attempt to improve the weak disturbance monitoring of CVA statistics, the vectors $\boldsymbol{d}_h$ and $\boldsymbol{e}_h$ must be improved with stronger disturbance sensitivity.

According to the statistical local analysis theory, given the system observation $\boldsymbol{z}_j$ and the system parameter $\boldsymbol{\vartheta}$, a primary residual vector $\varphi(\boldsymbol{z}_j, \boldsymbol{\vartheta})$ can be defined for disturbance detection if it meets the following conditions: [37,38]

- $E\{\varphi(\boldsymbol{z}_j, \boldsymbol{\vartheta})\} = 0$, if $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$;
- $E\{\varphi(\boldsymbol{z}_j, \boldsymbol{\vartheta})\} \neq 0$, if $\boldsymbol{\vartheta}$ is in the neighborhood of ϑ_0, but $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$;
- $\varphi(\boldsymbol{z}_j, \boldsymbol{\vartheta})$ is differentiable with $\boldsymbol{\vartheta}$;
- $\varphi(\boldsymbol{z}_j, \boldsymbol{\vartheta})$ exists for $\boldsymbol{\vartheta}$ in the neighborhood of $\boldsymbol{\vartheta}_0$.

Here $\boldsymbol{\vartheta}_0$ represents the parameters under the normal condition.

By investigating the i-th element in the vector $\boldsymbol{d}_h$, which is denoted as $d_{h,i}$, it is easily derived by Equation (11) that $d_{h,i} = \boldsymbol{a}_i^T \boldsymbol{p}_h$. Naturally, the variance of $d_{h,i}$ can be computed as

$$E\{d_{h,i}^2\} = \boldsymbol{a}_i^T E\{\boldsymbol{p}_h \boldsymbol{p}_h^T\} \boldsymbol{a}_i \tag{23}$$

For the statistical samples, $E\{\boldsymbol{p}_h \boldsymbol{p}_h^T\}$ is factually equal to the covariance matrix $\boldsymbol{\Sigma}_{pp}$. Further combining the first constraint on the vector $\boldsymbol{a}$ in Equation (3), it is known that $\boldsymbol{a}_i^T E\{\boldsymbol{p}_h \boldsymbol{p}_h^T\} \boldsymbol{a}_i = 1$. Therefore, we build the SLA primary residual of the canonical variate as

$$\varphi_{d_{h,i}} = d_{h,i}^2 - 1. \tag{24}$$

which meets the condition $E\{\varphi_{d_{h,i}}\} = 0$ in the normal condition.

Similarly, we analyze the variance of $e_{h,i}$ to obtain

$$E\{e_{h,i}^2\} = \boldsymbol{A}_r(i, :)E\{\boldsymbol{p}_h \boldsymbol{p}_h^T\}\boldsymbol{A}_r(i, :)^T \tag{25}$$

As A_r can be obtained in the model training procedure, the above expression must be equal to a fixed value, which is denoted as $\sigma_i = A_r(i,:)E\{p_h p_h^T\}A_r(i,:)^T$. Therefore, the SLA primary residual of the CVA residual can be built as

$$\varphi_{e_{h,i}} = e_{h,i}^2 - \sigma_i. \tag{26}$$

which meets the condition $E\{\varphi_{e_{h,i}}\} = 0$ for the normal data.

For a more sensitive disturbance detection, the SLA improved residual is applied in a moving window with the width of w, which is expressed by

$$\psi_{d_{h,i}} = \begin{cases} \frac{1}{\sqrt{h}}\sum\limits_{j=1}^{h}\varphi_{d-j,i}, & h < w \\ \frac{1}{\sqrt{w}}\sum\limits_{j=h-w+1}^{h}\varphi_{d_{j,i}}, & h \geq w \end{cases} \tag{27}$$

$$\psi_{e_{h,i}} = \begin{cases} \frac{1}{\sqrt{h}}\sum\limits_{j=1}^{h}\varphi_{e_{j,i}}, & h < w \\ \frac{1}{\sqrt{w}}\sum\limits_{j=h-w+1}^{h}\varphi_{e_{j,i}}, & h \geq w \end{cases} \tag{28}$$

Up to now, we can obtain the SLA improved residual vectors $\boldsymbol{\psi}_{d,h} = [\psi_{d_{h,1}}\ \psi_{d_{h,2}}\ \cdots\ \psi_{d_{h,s}}]^T$ and $\boldsymbol{\psi}_{e,h} = [\psi_{e_{h,1}}\ \psi_{e_{h,2}}\ \cdots\ \psi_{e_{h,M}}]^T$. These residual vectors are used to replace the original CVA features d_h and e_h so that the monitoring model is modified to the SLCVAkNN model.

With the SLA improved residual vectors, the monitoring statistics are constructed as follows:

$$T_h^2 = \boldsymbol{\psi}_{d,h}^T \boldsymbol{\psi}_{d,h} \tag{29}$$

$$Q_h = \boldsymbol{\psi}_{e,h}^T \boldsymbol{\psi}_{e,h} \tag{30}$$

3. Disturbance Detection Procedure Based on SLCVA*k*NN

Power transmission system disturbance detection based on SLCVA*k*NN method is divided into two stages: offline modeling stage and online detection stage. The corresponding flowchart is shown in Figure 1.

Stage 1: offline modeling stage

1. Acquire the normal condition data to constitute the training data set $X = [x_1^T, x_2^T, \cdots, x_n^T]^T \in R^{n \times m}$ and perform data normalization processing. Here, the mentioned normal condition data mean the data from a section of transmission line between two adjacent nodes. For different lines, the corresponding modelings are needed separately.
2. Construct historical data sets P and future data sets F according to Equations (4) and (5), calculate the covariance matrices by Equations (6)–(8), and solve the CVA optimization by the SVD as Equation (9).
3. Extract the canonical variate vector d_h and the CVA residual vector e_h, as shown in Equations (11) and (12).
4. Perform Equations (24) and (26) to obtain the SLA primary residual vectors and further calculate the SLA improved residual vectors by Equations (27) and (28).
5. Compute the monitoring statistics T_h^2 and Q_h for all the training samples according to Equations (29) and (30).
6. Construct the statistics matrix in the phase space according to Equations (17) and (18).
7. Calculate the SNND monitoring indices DT^2 and DQ for all the training samples and determine the 95% confidence limits DT_{lim}^2 and DQ_{lim} by kernel density estimation.

Stage 2: online detection stage

1. Obtain online new data x_t and normalize it with the training data.
2. Construct the corresponding historical vector p_t and project the p_t to the CVA model and obtain the canonical variate vector d_t and e_t according to Equations (11) and (12).
3. Apply Equations (24), (26)–(28) to compute the SLA primary residual vector and the improved residual vector orderly.
4. Compute the monitoring statistics T_t^2 and Q_t for the online new sample x_t according to Equations (29) and (30).
5. Construct the phase space statistics vector NT_t^2 and NQ_t^2, and calculate the SNND index DT_t^2 and DQ_t^2 by Equations (21) and (22).
6. Compare the SNND indices with the corresponding confidence limits DT_{lim}^2 and DQ_{lim}. If any one exceeds the confidence limit, a disturbance sample is indicated.

Here, it is pointed out that the local neighborhood standardization (LNS) [41] may be used to enhance the traditional z-score standardization. Compared with the traditional z-score method, LNS has better capability to deal with the non-steady data with the periodic oscillations.
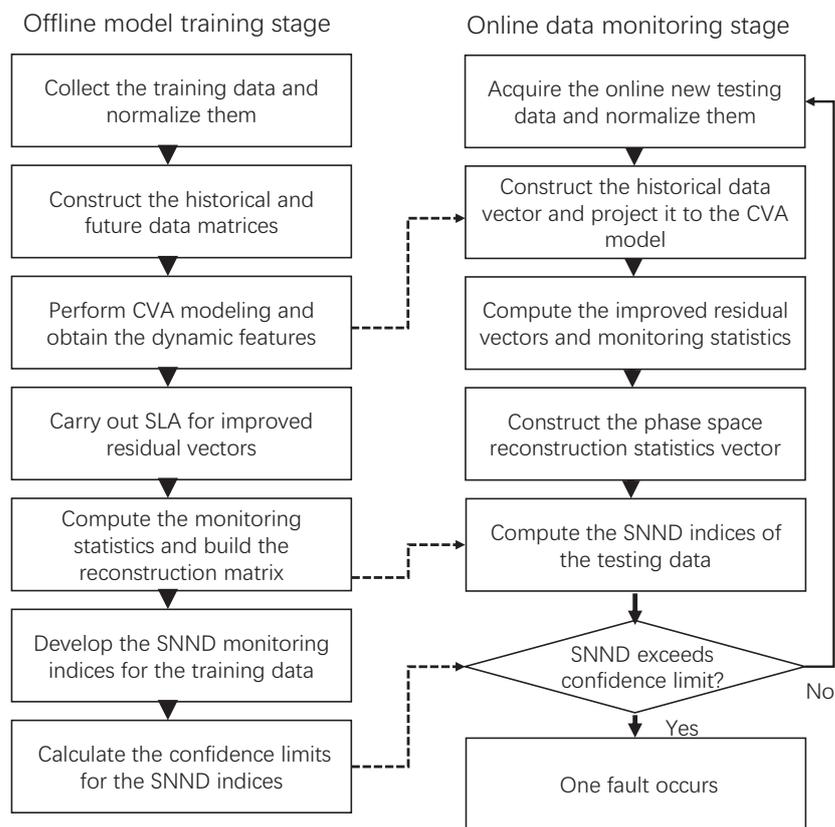
Figure 1. Flow chart of disturbance detection by SLCVA*k*NN.

4. Case Analysis

In order to verify the advantages of the SLCVA*k*NN method in the power transmission system disturbance detection, this section gives the case study on the real industrial data collected from the actual power transmission system. For method comparison, four methods, including the proposed SLCVA*k*NN method and three other methods of PCA, PCA*k*NN, and CVA*k*NN, are all applied to build the monitoring models for disturbance detection. The PCA method has two monitoring statistics T^2 and Q, while the other three methods are with the *k*NN-based statistics DT^2 and DQ. When these methods are used,

they indicate the system status by the monitoring charts, where the monitoring indices of normal and faulty samples are given by black and blue solid lines, respectively, while the detection threshold, that is the 95% confidence limit of the monitoring index, is plotted by the red dashed line. One evaluation index, called the disturbance detection rate (DDR), is used to evaluate the different monitoring methods. DDR is the percentage of the abnormal samples exceeding the detection threshold over all the abnormal samples.

The used real industrial data are collected from the seven transmission lines in a power supply station in August 2018. These lines are radially connected. Their data are collected because all of them involve the ground fault. The data acquisition units, designed by Qingdao Topscomm Communication CO. LTD, are used to collect the electric field intensity and current. Here, the real line voltage is up to 110 KV so that the existing equipment can not directly measure it. Therefore, the electric field intensity is applied to reflect the voltage trend. For each transmission line, one corresponding data set is recorded that involves the normal state and the abnormal state. The data set has the length of about 1300 samples, where the disturbance starting time (DST) is different in different transmission lines. The detailed information about the acquired data sets are listed in Table 1, where DST data record the sample number corresponding to the disturbance starting time. A demonstration of the collected data for the DATA-A is given in Figure 2, where six measured variables, including the electric field intensities of phase A, B, and C, and the currents of phase A, B, and C, are involved. Due to the existence of the harmonic load, the current sine wave distortion can be seen in these curves.

Table 1. The collected industrial data sets.

No.	Description	DST
DATA-1	Data set collected from line 904 exit	446
DATA-2	Data set collected from line 906, pole 116-3	456
DATA-3	Data set collected from line 906, pole 90-2	445
DATA-4	Data set collected from line 906, pole 151-5	458
DATA-5	Data set collected from line 906, pole 97-1	452
DATA-6	Data set collected from line 906 exit	493
DATA-7	Data set collected from line 907 exit	420
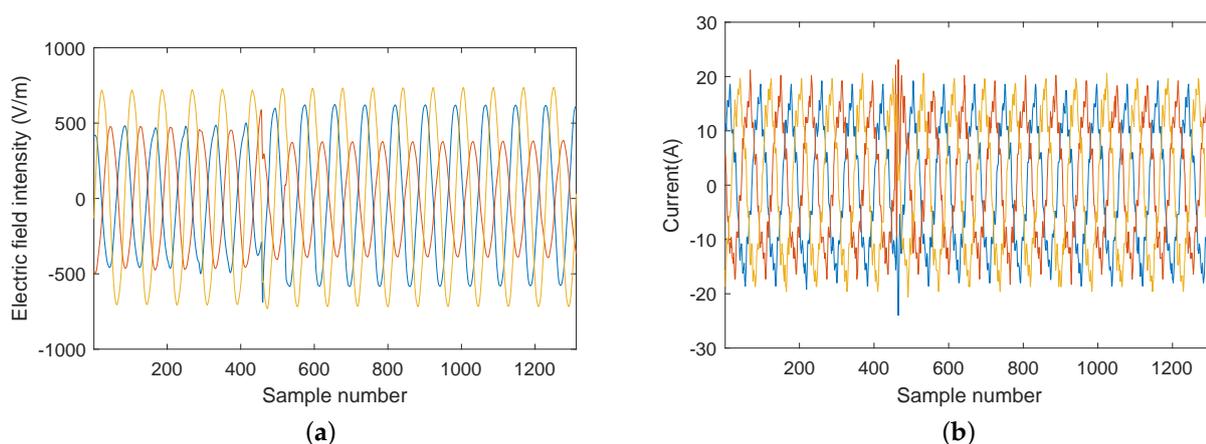
Figure 2. Data waveform collected from 904 line exit. (**a**) Three-phase electric field intensity; (**b**) Three-phase current.

Taking the data set DATA-2 as one example, it is collected from the pole 116-3 of the line 906. This data set includes 1312 samples. To investigate it with the help of on-site engineers, it is known that the disturbance occurs from the 456th sample. Although engineers can find this disturbance by careful analysis, this manual way is very time-consuming and inefficient, so it is difficult to implement in large-scale transmission system monitoring. Therefore, building an automatic multivariate data analysis tool is very necessary. In

this section, we apply four MSA methods, which are PCA, PCA*k*NN, CVA*k*NN, and SLCVA*k*NN, to perform the automatic fault detection. When the statistical models are developed, the model parameters are set as follows: $k = 3, L = 10, l = 2, w = 20$. For the data set DATA-2, the first 320 sampling point are considered to be in a normal operating state, they can be utilized as the training data set for model development, while monitoring charts of PCA, PCA*k*NN, CVA*k*NN, and SLCVA*k*NN are demonstrated in the Figures 3–6, respectively. By the PCA monitoring results shown in Figure 3, it can be seen that the disturbance cannot be detected very effectively. The DDR of PCA T^2 is 4.43%, while the Q is a little better with the DDR of 29.52%. When PCA*k*NN is used, the DT^2 has a similarly poor detection rate, but the DQ statistic achieves clear improvement with the DDR of 57.76%. These results demonstrate that the PCA*k*NN method proposed by Cai et al. [20] can deal with the power system data with oscillation characteristic effectively. However, from these figures, the monitoring statistics do not exceed the confidence limits significantly. This may lead to the uncertain judgement on the occurrence of disturbance. When the CVA*k*NN is applied in Figure 5, the DQ statistic performs a little better with the DDR of 49.71%. However, its DT^2 indicator clearly improves the DDR to 92.51%, which means a significant detection rate improvement of about 70% in contrast with the PCA*k*NN's DQ index. The best monitoring results on this data set is provided by SLCVA*k*NN, which are shown in Figure 6. By this figure, it is observed that the disturbance is detected very clearly with the DDRs of 97.25% and 96.80% for DT^2 and DQ, respectively. This case gives a comprehensive comparison on the four methods of PCA, PCA*k*NN, CVA*k*NN, and SLCVA*k*NN. The applications show that PCA*k*NN does better than PCA due to the use of *k*NN, while SLCVA*k*NN further prompts the disturbance detection performance with the integration of CVA and SLA.
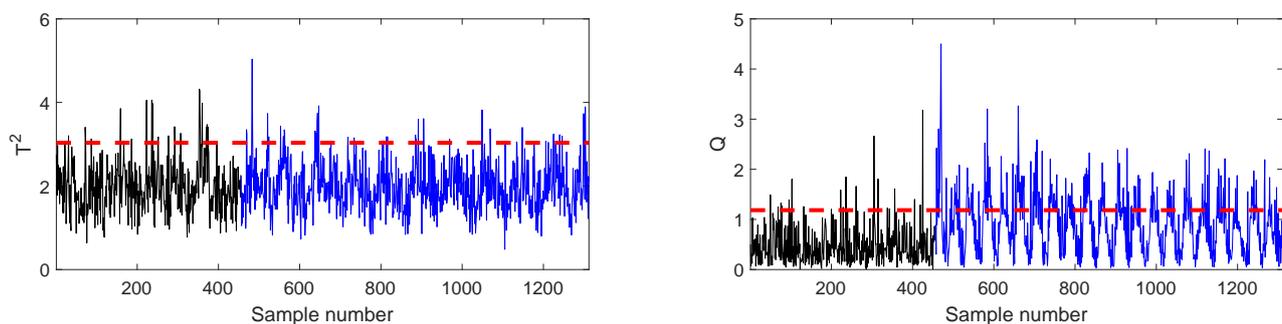
Figure 3. PCA monitoring results on the DATA-2 case.
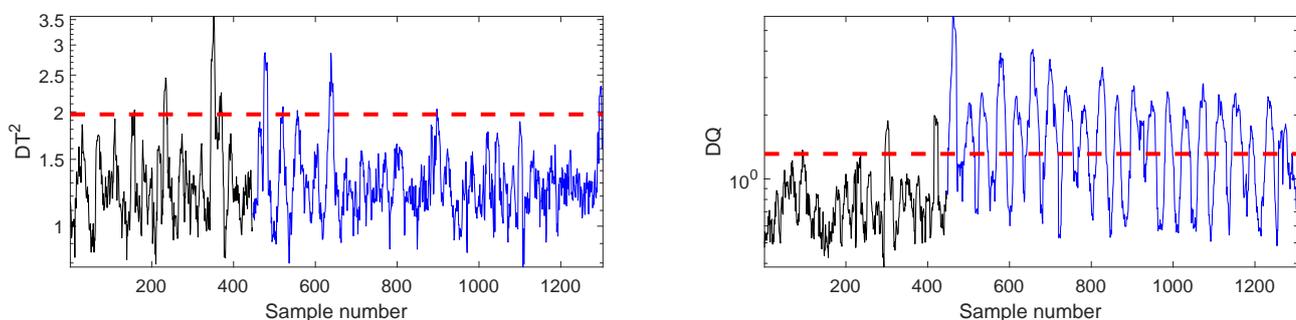
Figure 4. PCA*k*NN monitoring results on the DATA-2 case.
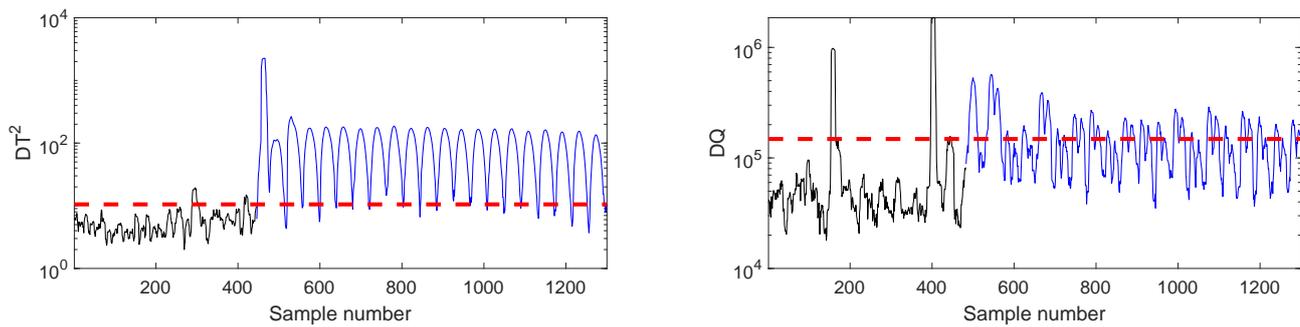
Figure 5. CVA*k*NN monitoring results on the DATA-2 case.
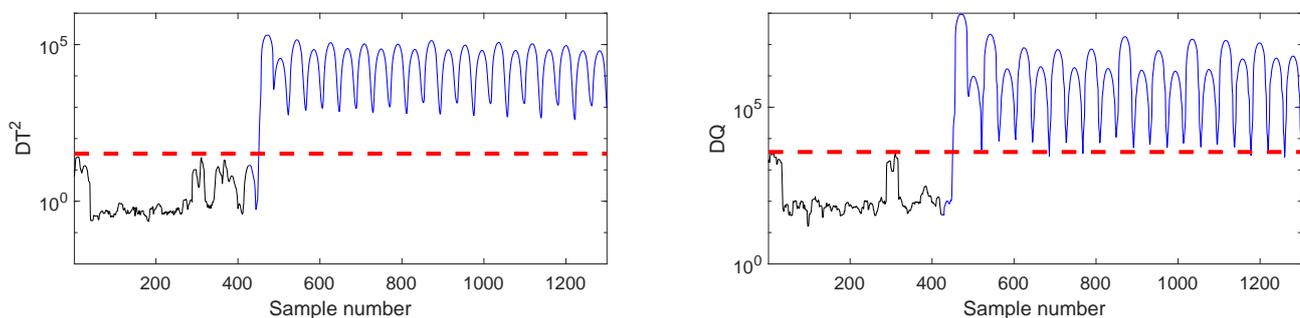
Figure 6. SLCVAkNN monitoring results on the DATA-2 case.

Another example on the data set DATA-6 is illustrated, which corresponds to the line 906 exit. The modeling procedure is similar to the above case. Here we only give the monitoring charts of CVA*k*NN and SLCVA*k*NN, as shown in the Figures 7 and 8. With the consideration of system dynamics, the CVA*k*NN DT^2 monitoring chart gives a higher DDR of 88.51%. Compared with the CVA*k*NN method, which has only one effective monitoring statistic, SLCVA*k*NN has two well-behaved monitoring statistics. The DT^2 and DQ have the DDRs of 97.37% and 97.25%, respectively. The testing results on DATA-6 further verify the advantage of the proposed method over the CVA*k*NN method.
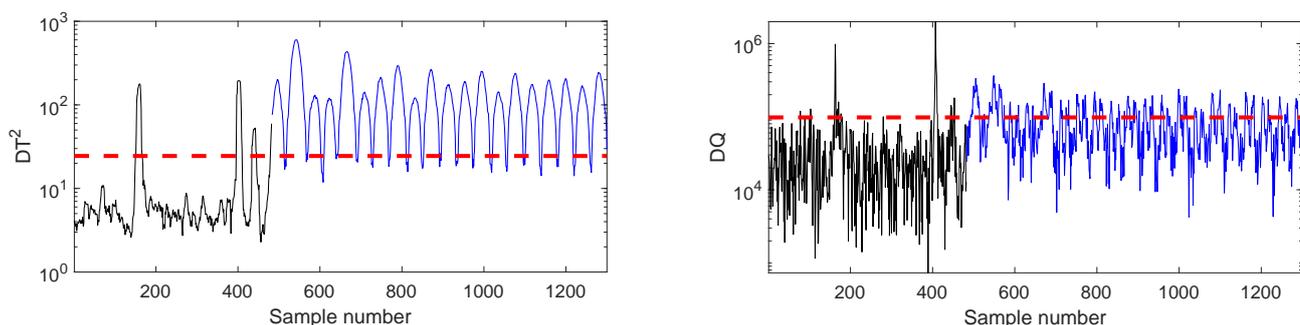
Figure 7. CVA*k*NN monitoring results on the DATA-6 case.

The summary of disturbance detection rates for all seven data sets are shown in Table 2. From this table, it is shown that the faults in DATA-2 and DATA-4 are difficult to detect by PCA, whose DDRs are all lower than 30%. By the use of PCA*k*NN, these two faults are detected with higher DDRs, which are 57.76% and 26.78%, respectively. By contrast, CVA*k*NN does better on the two faults. In particular, its DT^2 statistic gives the DDR higher than 90%. When SLCVAkNN is used, its two monitoring statistics have the higher DDRs than 95%. For the sets of DATA-1, DATA-3, DATA-6, and DATA-7, PCA can detect these faults with about 70-80% DDR on one statistic. That means PCA can alarm these faults, but

the alarm degree is not very sufficient. The PCA*k*NN and CVA*k*NN improve the DDR to about 90%. Further combining the SLA technique, SLCVA*k*NN achieves higher DDR than CVA*k*NN on these four sets. As to DATA-5, all these four methods give a similarly good performance with the DDRs higher than 95%. Considering all seven of these data sets, we observe that the average detection rates of CVA*k*NN outperforms the PCA and PCA*k*NN method, while the ones of SLCVA*k*NN statistics can reach 97.46% and 96.29%, which are the highest among these four methods.
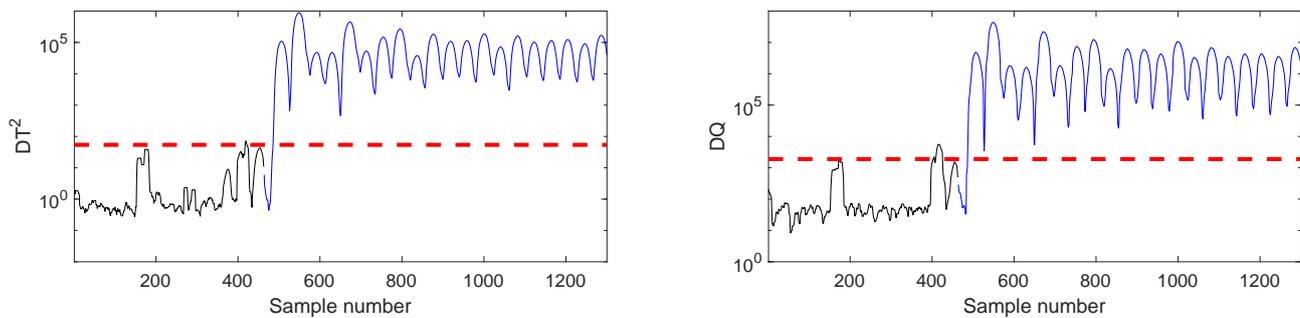
Figure 8. SLCVA*k*NN monitoring results on the DATA-6 case.

Table 2. The disturbance detection rate of PCA, PCA*k*NN, CVA*k*NN, and SLCVA*k*NN for the tested data sets.

NO.	PCA		PCA*k*NN		CVA*k*NN		SLCVA*k*NN	
	T^2	Q	DT^2	DQ	DT^2	DQ	DT^2	DQ
DATA-1	7.50%	70.47%	26.18%	90.89%	96.76%	64.39%	96.83%	89.48%
DATA-2	4.43%	29.52%	3.85%	57.76%	92.51%	49.71%	97.25%	96.80%
DATA-3	5.99%	83.06%	19.93%	97.81%	100.00%	100.00%	97.85%	97.97%
DATA-4	5.15%	10.18%	8.54%	26.78%	96.60%	51.93%	97.36%	96.79%
DATA-5	48.78%	96.05%	82.81%	99.54%	100.00%	100.00%	98.06%	98.29%
DATA-6	4.39%	77.44%	5.49%	92.20%	88.51%	41.20%	97.37%	97.25%
DATA-7	7.95%	84.99%	11.09%	95.97%	93.49%	55.33%	97.47%	97.47%
Average	12.03%	64.53%	22.56%	80.13%	95.41%	66.08%	97.46%	96.29%

To sum up, the applications on real industrial data verify the effectiveness of the proposed SLCVA*k*NN in the power transmission system monitoring. All the tested faults are about the ground faults. Although this paper does not provide the results on the other disturbances such as 1,3-phase short circuits, overvoltages, the presented algorithm is also suitable for these cases because they similarly lead to the changes of voltage and current. However, one related issue should be noted. In this article, this method detects all the occurred disturbances, including normal disturbances such as load power variations. To judge whether the disturbance is a fault or a normal disturbance is a further job. In fact, as to this issue, one solution is to enrich the modeling data with different normal changes. As the *k*NN used in this method can deal with the multimodal data case, the trained model can distinguish the faults and normal disturbances effectively when the normal changing data are considered in the model training procedure.

5. Conclusions

This paper proposes a power transmission system disturbance detection method based on SLCVA*k*NN. The real industrial data collected from the field transformer station are applied to verify the proposed method. By investigating the application results, we can draw the following conclusions.

- **CVA-based monitoring method can provide better dynamic information mining**. The dynamic data analysis tool CVA is introduced to deal with the power transmission system data. By observing the application results, we find that CVA*k*NN has a higher detection rate than PCA*k*NN.
- **The statistical local analysis can further enhance the disturbance monitoring**. Considering that many high-impendence ground faults in the real power systems are with insignificant symptoms, the weak disturbance detection methods are very important in improving the disturbance detection sensitivity. By focusing on the statistical local information of CVA features, the proposed SLCVA*k*NN method outperforms the CVA*k*NN method.

Nomenclature

CVA	canonical variate analysis
MSA	multivariate statistical analysis
SLA	statistical local analysis
WAMS	wide area measurement system
A_s	projection matrix
d_h	canonical variate vector
DQ_t	SNND monitoring index
f_h	future data vector at the h-th sample instant
MT^2	phase reconstruction statistics matrix
NT_t^2	reconstructed statistics vector
p_h	historical data vector at the h-th sample instant
Q_h	monitoring statistic at the h-th sample instant
x_h	data vector at the h-th sample instant
$\phi_{d_{h,i}}$	SLA primary residual of canonical variate
$\psi_{d_{h,i}}$	SLA improved residual of canonical variate
Σ	Covariance matrix
kNN	k-nearest neighbor
PCA	principal component analysis
SNND	statistical nearest neighbor distance
a	projection vector
b	projection vector
DT_t^2	SNND monitoring index
e_h	residual vector at the h-th sample instant
F	future data matrix

MQ	phase reconstruction statistics matrix
NQ_t	reconstructed statistics vector
P	historical data matrix
T_h^2	monitoring statistic at the h-th sample instant
X	data matrix
$\phi_{e_{h,i}}$	SLA primary residual of CVA residual
$\psi_{e_{h,i}}$	SLA improved residual of CVA residual

References

1. Samuelsson, O.; Hemmingsson, M.; Nielsen, A.H.; Pedersen, K.O.H.; Rasmussen, J. Monitoring of power system events at transmission and distribution level. *IEEE Trans. Power Syst.* **2006**, *21*, 1007–1008. [CrossRef]
2. Patel, B.; Bera, P. Fast fault detection during power swing on a hybrid transmission line using WPT. *IET Gener. Transm. Distrib.* **2019**, *13*, 1811–1820. [CrossRef]
3. Musa, M.H.H.; He, Z.; Fu, L.; Deng, Y. Linear regression index-based method for fault detection and classification in power transmission line. *IEEJ Trans. Electr. Electron. Eng.* **2018**, *13*, 979–987. [CrossRef]
4. Chang, H.-H.; Linh, N.V.; Lee, W.-J. A novel nonintrusive fault identification for power transmission networks using power-spectrum-based hyperbolic s-transform-part i: Fault classification. *IEEE Trans. Ind. Appl.* **2018**, *54*, 5700–5710. [CrossRef]
5. Costa, F.B. Fault-induced transient detection based on real-time analysis of the wavelet coefficient energy. *IEEE Trans. Power Deliv.* **2013**, *29*, 140–153. [CrossRef]
6. Math, H.J.B.; Das, R.; Djokic, S.; Ciufo, P.; Meyer, J.; Ronnberg, S.K.; Zavoda, F. Power quality concerns in implementing smart distribution-grid applications. *IEEE Trans. Smart Grid* **2016**, *8*, 391–399.
7. Cheng, C.; Wang, W.; Chen, H.; Zhang, B.; Shao, J.; Teng, W. Enhanced fault diagnosis usign broad learning for traction systems in high-speed trains. *IEEE Trans. Power Electron.* **2021**, *36*, 7461–7469. [CrossRef]
8. Huang, S.; Hsieh, C.; Huang, C. Application of morlet wavelets to supervise power system disturbances. *IEEE Trans. Power Deliv.* **1999**, *14*, 235–243. [CrossRef]
9. Manglik, A.; Li, W.; Ahmad, S.U. Fault Detection in power system using the Hilbert-Huang transform. In Proceedings of the 2016 IEEE Canadian Conference On Electrical And Computer Engineering (CCECE), Vancouver, BC, Canada, 14–18 May 2016.
10. Ghaderi, A.; Mohammadpour, H.A.; Ginn, H.L.; Shin, Y.-J. High-impedance fault detection in the distribution network using the time-frequency-based algorithm. *IEEE Trans. Power Deliv.* **2015**, *30*, 1260–1268. [CrossRef]
11. Salehi, M.; Namdari, F. Fault classification and faulted phase selection for transmission line using morphological edge detection filter. *IET Gener. Transm. Distrib.* **2018**, *12*, 1595–1605. [CrossRef]
12. Liu, Z.; Hu, Q.; Cui, Y.; Zhang, Q. A new detection approach of transient disturbances combining wavelet packet and tsallis entropy. *Neurocomputing* **2014**, *142*, 393–407. [CrossRef]
13. Zhang, X.; Kano, M.; Song, Z. Optimal weighting distance-based similarity for locally weighted PLS modeling. *Ind. Eng. Chem. Res.* **2020**, *59*, 11552–11558. [CrossRef]
14. Deng, X.; Du, K. Efficient batch process monitoring based on random nonlinear feature analysis. *Canadian J. Chem. Eng.* **2021**, *in press*, 1–12. [CrossRef]
15. Zhang, X.; Kano, M.; Matsuzaki, S. A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking. *Comput. Chem. Eng.* **2019**, *130*, 106575. [CrossRef]
16. Chen, H.; 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**, *in press*, 1–17. [CrossRef]
17. Chen, H.; Jiang, B. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 450–465. [CrossRef]
18. Barocio, E.; Pal, B.C.; Fabozzi, D.; Thornhill, N.F. Detection and visualization of power system disturbances using principal component analysis. In Proceedings of the 2013 IREP Symposium Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid, Rethymnon, Greece, 25–30 August 2013.
19. Guo, Y.; Li, K.; Liu, X. Fault diagnosis for power system transmission line based on PCA and SVMs. In Proceedings of the 2nd International Conference on Intelligent Computing for Sustainable Energy and Environment (ICSEE), Shanghai, China, 12–13 September 2012.
20. Cai, L.; Thornhill, N.F.; Kuenzel, S.; Pal, B.C. Wide-area monitoring of power systems using principal component analysis and k-nearest neighbor analysis. *IEEE Trans. Power Syst.* **2018**, *33*, 4913–4923. [CrossRef]
21. Chen, H.; Wu, J.; Jiang, B.; Chen, W. A modified neighborhood preserving embedding-based incipient fault detection with applications to small-scale cyber-physical systems. *ISA Trans.* **2020**, *104*, 175–183. [CrossRef] [PubMed]
22. Chen, H.; Jiang, B.; Zhang, T.; Lu, N. Data-driven and deep learning-based detection and diagnosis of incipient faults with application to electrical traction systems. *Neurocomputing* **2020**, *396*, 429–437. [CrossRef]
23. Jiang, B.; Braatz, R.D. Fault detection of process correlation structure using canonical variate analysis-based correlation features. *J. Process. Control* **2017**, *58*, 131–138. [CrossRef]
24. Li, X.; Yang, Y.; Bennett, I.; Mba, D. Condition monitoring of rotating machines under time-varying conditions based on adaptive canonical variate analysis. *Mech. Syst. Signal Process.* **2019**, *131*, 348–363. [CrossRef]

25. Chen, Z.; Yang, C.; Peng, T.; Dan, H.; Li, C.; Gui, W. A cumulative canonical correlation analysis-based sensor precision degradation detection method. *IEEE Trans. Ind. Electron.* **2019**, *66*, 6321–6330. [CrossRef]

26. Han, S.; Xu, Z.; Sun, B.; He, L. Dynamic characteristic analysis of power system interarea oscillations using HHT. *Int. J. Electr. Power Energy Syst.* **2010**, *32*, 1085–1090. [CrossRef]

27. Hu, Z. Method considering the dynamic coupling characteristic in power system for stability assessment. *IET Gener. Transm. Distrib.* **2017**, *11*, 2534–2539. [CrossRef]

28. Zhang, S.; Zhao, C.; Huang, B. Simultaneous static and dynamic analysis for fine-scale identification of process operation statuses. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5320–5329. [CrossRef]

29. Li, X.; Duan, F.; Bennett, I.; Mba, D. Canonical variate analysis, probability approach and support vector regression for fault identification and failure time prediction. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3771–3783. [CrossRef]

30. Chiang, L.H.; Russell, E.L.; Braatz, R.D. *Fault Detection and Diagnosis in Industrial Systems*; Springer: London, UK, 2001.

31. Odiowei, P.E.P.; Cao, Y. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Trans. Ind. Inform.* **2010**, *6*, 36–45. [CrossRef]

32. Zhang, Z.; Deng, X. Anomaly detection using improved deep SVDD model with data structure preservation. *Pattern Recognit. Lett.* **2021**, *148*, 1–6. [CrossRef]

33. Zhang, X.; Li, Y. Multiway principal polynomial analysis for semiconductor manufacturing process fault detection. *Chemom. Intell. Lab. Syst.* **2018**, *181*, 29–35. [CrossRef]

34. Cai, L.; Thornhill, N.F.; Kuenzel, S.; Pal, B.C. Real-time detection of power system disturbances based on k-nearest neighbor analysis. *IEEE Access* **2017**, *5*, 5631–5639. [CrossRef]

35. Zhang, A.; Xu, Z. Chaotic time series prediction using phase space reconstruction based conceptor network. *Cogn. Neurodynamics* **2020**, *14*, 849–857. [CrossRef]

36. Yu, P.; Yan, X. Stock price prediction based on deep neural networks. *Neural Comput. Appl.* **2020**, *32*, 1609–1628. [CrossRef]

37. Basseville, M. On-board component fault detection and isolation using the statistical local approach. *Automatica* **1998**, *34*, 1391–1415. [CrossRef]

38. Kruger, U.; Kumar, S.; Littler, T. Improved principal component monitoring using the local approach. *Automatica* **2007**, *43*, 1532–1542. [CrossRef]

39. Ge, Z.; Yang, C.; Song, Z. Improved kernel PCA-based monitoring approach for nonlinear processes. *Chem. Eng. Sci.* **2009**, *64*, 2245–2255. [CrossRef]

40. Deng, X.; Cai, P.; Deng, J.; Cao, Y.; Song, Z. Primary-auxiliary statistical local kernel principal component analysis and its application to incipient fault detection of nonlinear industrial processes. *IEEE Access* **2019**, *7*, 122192–122204. [CrossRef]

41. Ma, H.; Hu, Y.; Shi, H. A novel local neighborhood standardization strategy and its application in fault detection of multimode processes. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 287–300. [CrossRef]