


Article

# Clustering Analysis of Voltage Sag Events Based on Waveform Matching

Chenyan Hao and Jun Jin \* 

College of Mathematical Sciences, Yangzhou University, Yangzhou 225002, China; hcy5826@163.com

\* Correspondence: 118020208006@smail.swufe.edu.cn

**Abstract:** Voltage sags are a serious problem within power supplies, which pose threats to both residential electricity and industrial manufacturing. Since any one sag may be recorded by multiple monitoring devices from different substations, the issue of redundant information in data arises. In this regard, a novel method for voltage sag events based on projection technology, shape dynamic time warping (shapeDTW), and spectral clustering is proposed. The main contributions of this paper may be summarized as follows: (1) We present a new method for extracting the voltage anomaly waveform, which is a fast projection segmentation algorithm (FPSA). The voltage sag waveform is only a part of the voltage anomaly waveform, so the voltage anomaly waveform contains more information. (2) ShapeDTW and spectral clustering are used to match and cluster voltage anomaly waveforms, so as to achieve the normalization of voltage sag events. (3) In practical engineering, the proposed method in the paper can be used to obtain the impact of voltage sags, reduce computational complexity, and ease the workload of the operation and maintenance engineers. Experiments were conducted using voltage sag data from voltage sag events recorded by the 10 kV monitoring points in Beijing, China. The results showed the effectiveness and reliability of our proposed methods.

**Keywords:** fast projection segmentation algorithm (FPSA); voltage sag event; shapedtw; spectral clustering



**Citation:** Hao, C.; Jin, J. Clustering Analysis of Voltage Sag Events Based on Waveform Matching. *Processes* **2022**, *10*, 1337. <https://doi.org/10.3390/pr10071337>

Academic Editor: Jean-Pierre Corriou

Received: 20 January 2022

Accepted: 25 May 2022

Published: 8 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Currently, voltage sags have become an increasingly relevant issue for both utility companies and their consumers. In certain industries, a sag event only a few milliseconds in length can interrupt critical processes, which may then take several hours to be restarted. This results in large financial losses due to downtime [1]. Moreover, in industrial applications, voltage sags are the most frequently occurring disturbances among all the possible disturbances and account for 92% to 98% of interruptions. They may also be responsible for other disturbances in the power network [2]. Therefore, the treatment and research of voltage sags are important in improving the reliability of power systems and ensuring a safe power supply in industrial production and daily life [3]. The Institute of Electrical and Electronics Engineers' (IEEE) definition for voltage sag is as follows: the root mean square (RMS) voltage drops to 10–90% of the rated value, and the duration is 10 ms–60 s [4]. At present, the data collection, detection, classification, and identification of voltage sags are the focus of research; see [5–12].

Many studies have been performed regarding the pattern recognition of voltage sags [3]. In general, there are three methods: data mining, signal processing, and physical modeling. Distinct methods for identifying the source of voltage sags were analyzed in [5], which are based on the disturbance energy, voltage current characteristic, or active current component. Reference [13] proposed a Kullback–Leibler divergence measure and standard deviation for voltage sag and harmonics identification. A support vector machine and support vector regression were used to identify fault types and estimate fault resistance. The Euclidean distance approach was used to identify the fault distance in [14]. K-means clustering was used to identify voltage sags in [15]. Reference [16]

detected and classified different power quality disturbances using the half- and one-period windowing technique based on a continuous S-transform and neural networks. Several signal-processing techniques have been proposed for power quality (PQ) disturbance detection and segmentation. In [17], sags and interruptions were detected by the RMS value. Its main drawback is the poor results achieved with non-stationary signals due to the dependency of the RMS on both periodicity and wave shape. A sag/swell detection algorithm based on wavelet transform, operating even in the presence of flicker and harmonics in the voltage source, was presented in [18]. In [19], the Kalman filter and an expert system were used to segment and identify different types of voltage dips (fault-induced, transformer saturation, induction motor starting) and interruptions (non-fault and fault-induced). The main drawbacks of this method are related to either the failure in detecting very small changes in the voltage magnitude or the time resolution problems. In [20], a method for voltage sag source identification that combined wavelet analysis and modified dynamic time warping (DTW) distance was proposed. In [21], a statistical analysis of variance (MANOVA) was directly proposed to extract the attributes of voltage events from the voltage and current waveforms. In [22], Thevenin's equivalent circuit was used to replace any power network. By determining the sign of the internal resistance in Thevenin's equivalent circuit, the origin of a voltage sag disturbance could be easily identified. In [7], the authors proposed a method of detecting sags for applications in dynamic voltage restorers. The method is based on the fundamental amplitude calculation using the d-q components of each phase of the voltage signal in half-cycle windows. The method was efficient for different levels of sags, phase-jump, harmonics, and variation in fundamental frequency. In [23], the authors used the delayed Legendre wavelet, ant lion optimization algorithm, and a classifier ensemble to detect and classify eight types of sags.

The monitoring and analysis of voltage sag events can provide an effective scientific basis for power system operations management, accident investigation, fault location, and sag management [3]. With the advancement of science and technology, a large amount of original sampling data may be collected by monitoring systems. These data completely encapsulate the transient waveforms of each detected sag event. However, the data of the same voltage sag event would be recorded by multiple monitoring points. These data not only consume computer memory unnecessarily, but also leave higher workloads for engineers. The analysis of these data is then extremely inefficient due to a lack of effective data-processing procedures. Therefore, it is very important to identify whether the data recorded by different monitoring points are the voltage sag caused by the same event. As far as we know, there is no relevant literature on this aspect. In order to meet the needs of practical engineering applications, a novel clustering analysis method for voltage sag events is proposed in this paper. At present, all classification methods are mainly based on the characteristics of the voltage sag waveform. However, this waveform is only a part of the voltage anomaly waveform. The voltage anomaly waveform contains more information than the voltage sag waveform. Therefore, we propose a new method for extracting the voltage anomaly waveform instead of the voltage sag waveform, which is an FPSA. Then, we use shapeDTW and spectral clustering to match and cluster voltage anomaly waveforms, so as to achieve the normalization of voltage sag events.

The rest of this paper is organized as follows. In Section 2, our novel clustering analysis method for voltage sag events based on projection technology, shapeDTW, and spectral clustering is introduced. In Section 3, real data analyses are used to verify the effectiveness of the FPSA and clustering analysis method. The conclusions are summarized in Section 4.

## 2. Methodology

In this section, the FPSA is presented as an effective solution to correctly identify non-stationary and quasi-stationary states in the measured voltage waveforms. We further investigate the classification of voltage sag events by combining shapeDTW and spectral clustering. A total of twenty recording files of voltage sag events are available to this end, which are labeled as id 01, 02,  $\dots$ , 20. These signals are events recorded by PQ instruments

and contain both pre-trigger and post-trigger information. Figure 1 shows voltage data (id 01). Next, we mainly use id01 data to illustrate our proposed FPSA.

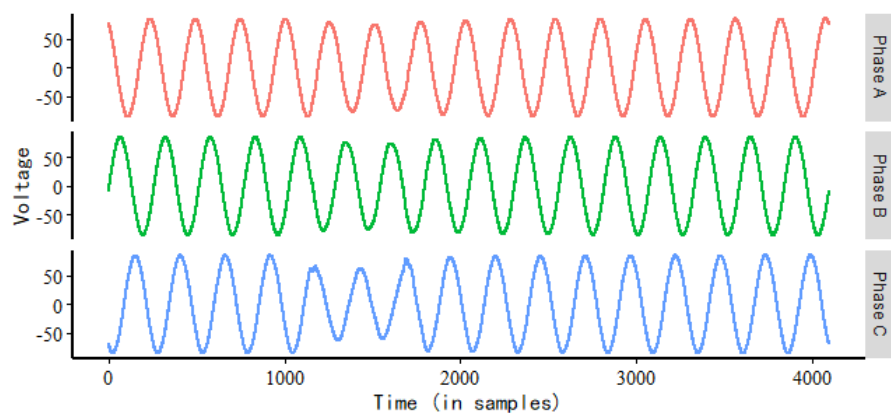


Figure 1. Voltage data (id 01).

### 2.1. Fast Projection Segmentation Algorithm

The model used to detect a signal transition can be divided into three steps: signal modeling, acquisition of the detection parameter (DP), and extraction of the voltage anomaly segment. The basic principle of signal modeling is to generate residuals by comparing the actual responses with the expected responses of the system using mathematical models. These residuals are expected to be zero (or zero mean) under no-fault conditions. In practical situations, the residuals are corrupted by the presence of noise, unknown disturbances, and uncertainties in the system model. Hence, the aim of the method is to generate robust residuals that are insensitive to these noises and uncertainties, while remaining sensitive to faults. To this end, a simple and effective filtering method is proposed in this paper.

#### 2.1.1. Signal Modeling Based on Filtering

Since the voltage data are usually periodic, it was assumed that the current voltage sag monitoring terminal records the data as  $K$  points/period ( $K = 256$  in this paper). The voltage data of length  $N$  are expressed as  $\{x_i, i = 1, 2, \dots, N\}$ , and then, the filtering method can be divided into the following three steps:

*Step 1* : Given  $K$ , the original voltage data  $\{x_i, i = 1, 2, \dots, N\}$  can be divided into  $L$  segments:  $\{x_1, x_2, \dots, x_{K_1}\}, \{x_{K_1+1}, \dots, x_{K_2}\} \dots, \{x_{K_{L-1}+1}, \dots, x_{K_L}\}$ , where  $L = \lfloor N/K \rfloor$ ,  $K_1 = K, K_2 = 2 \times K, K_L = L \times K$  and  $\lfloor \cdot \rfloor$  denotes rounding down. Voltage data (id 01) segmentation results are shown in Figure 2, where the black dotted lines are the dividing lines.

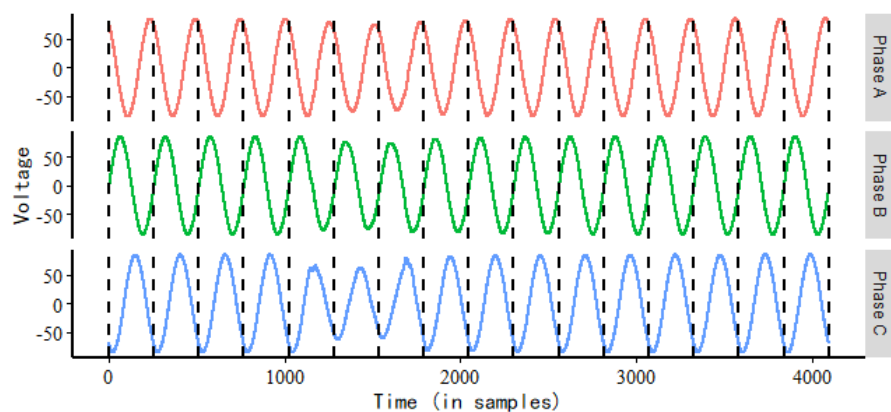


Figure 2. Voltage data (id 01) are divided into  $L$  segments according to  $K$ .

*Step 2* : According to IEEE Standard 1159–1995 [2], the RMS is used to initially determine whether  $\{x_{K_{j-1}+1}, \dots, x_{K_j}\}$  has a sag. If  $\{x_{K_{j-1}+1}, \dots, x_{K_j}\}$  is a sag, then  $\{x_{K_{j-1}+1}, \dots, x_{K_j}\}$  is removed, where  $j = 1, 2, \dots, L$ . The remaining data  $X$  after removal are obtained as follows.

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_l & \cdots & X_{K_1} \\ X_{K_1+1} & X_{K_1+2} & \cdots & X_{K_1+l} & \cdots & X_{K_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{K_{j-1}+1} & X_{K_{j-1}+2} & \cdots & X_{K_{j-1}+l} & \cdots & X_{K_j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{K_{L-1}+1} & X_{K_{L-1}+2} & \cdots & X_{K_{L-1}+l} & \cdots & X_{K_L} \end{bmatrix}, l = 1, 2, \dots, K, j = 1, 2, \dots, L \quad (1).$$

Since the median can characterize the center position of the distribution and has the advantage of being robust and unaffected by outliers, we used the median of the data  $\{x_l, x_{K_1+l}, \dots, x_{K_{L-1}+l}\}$  as the estimate of the  $l$ th column of the matrix  $X$ . Let  $\hat{x}_l = \text{median}\{x_l, x_{K_1+l}, \dots, x_{K_{L-1}+l}\}$ , where  $l = 1, 2, \dots, K$ . The process of Step 2 on the voltage sag data (id 01) is shown in Figure 3.

*Step 3* : Sequence 1 can be obtained by Step 2 as follows:

$$\text{Sequence 1: } \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_K\},$$

Repeat Sequence 1 on the length of the original data ( $N$ ) to obtain Sequence 2:

$$\text{Sequence 2: } \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_K, \hat{x}_1, \hat{x}_2, \dots, \hat{x}_K \dots\}.$$

Sequence 2 is our estimate of the original data. Then, we can obtain the filter sequence (residuals' sequence) as follows:

$$\text{Filter sequence: } \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N\} \quad \varepsilon_i = \hat{x}_i - x_i, \quad i = 1, 2, \dots, N.$$

The filtering process for voltage sag data (id 01) is shown in Figure 4.

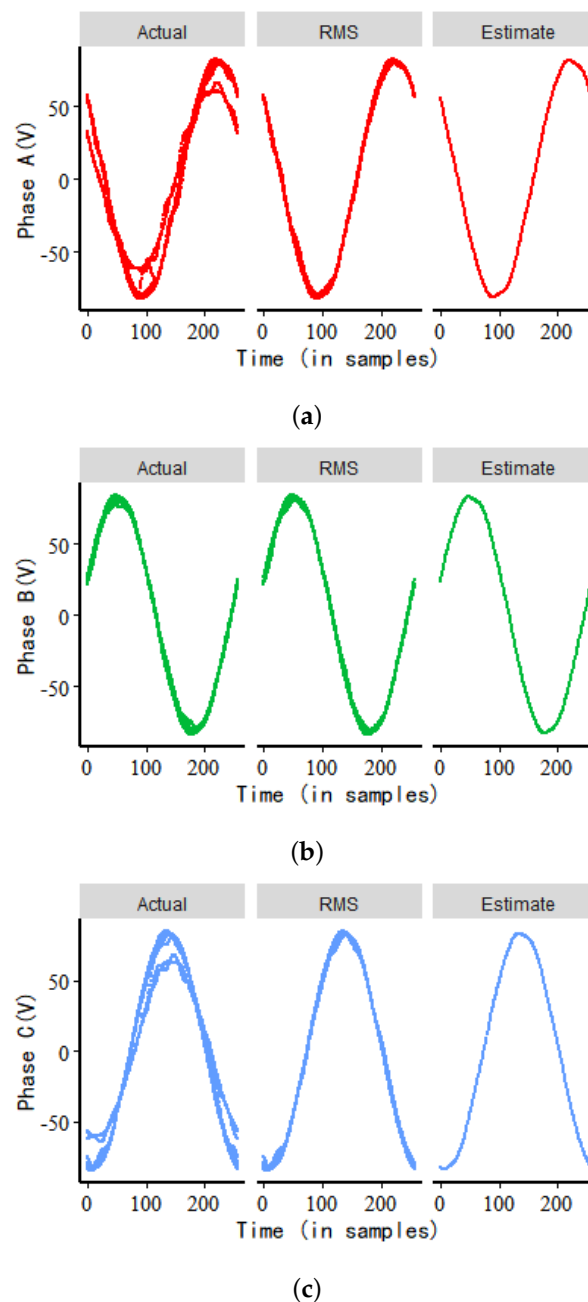
### 2.1.2. Detection Parameter Based on Sharp Drop Point

The voltage sag data are projected onto the vertical axis, and the frequency histogram can be obtained after projection (as shown in the red part of Figure 5b). The kernel density curve  $f(x)$  of the frequency histogram can be obtained by kernel density estimation (the red curve shown in Figure 5b). In the real data, the amount of voltage anomaly (voltage sag) data is always significantly smaller than the total amount of data. Based on this fact, we can clearly find that the anomaly data are mainly distributed in the tail region of the projection distribution, while the non-anomaly data are mainly concentrated in the central region of the projection distribution. The projected distribution of voltage sag data (id 01) is shown in Figure 5.

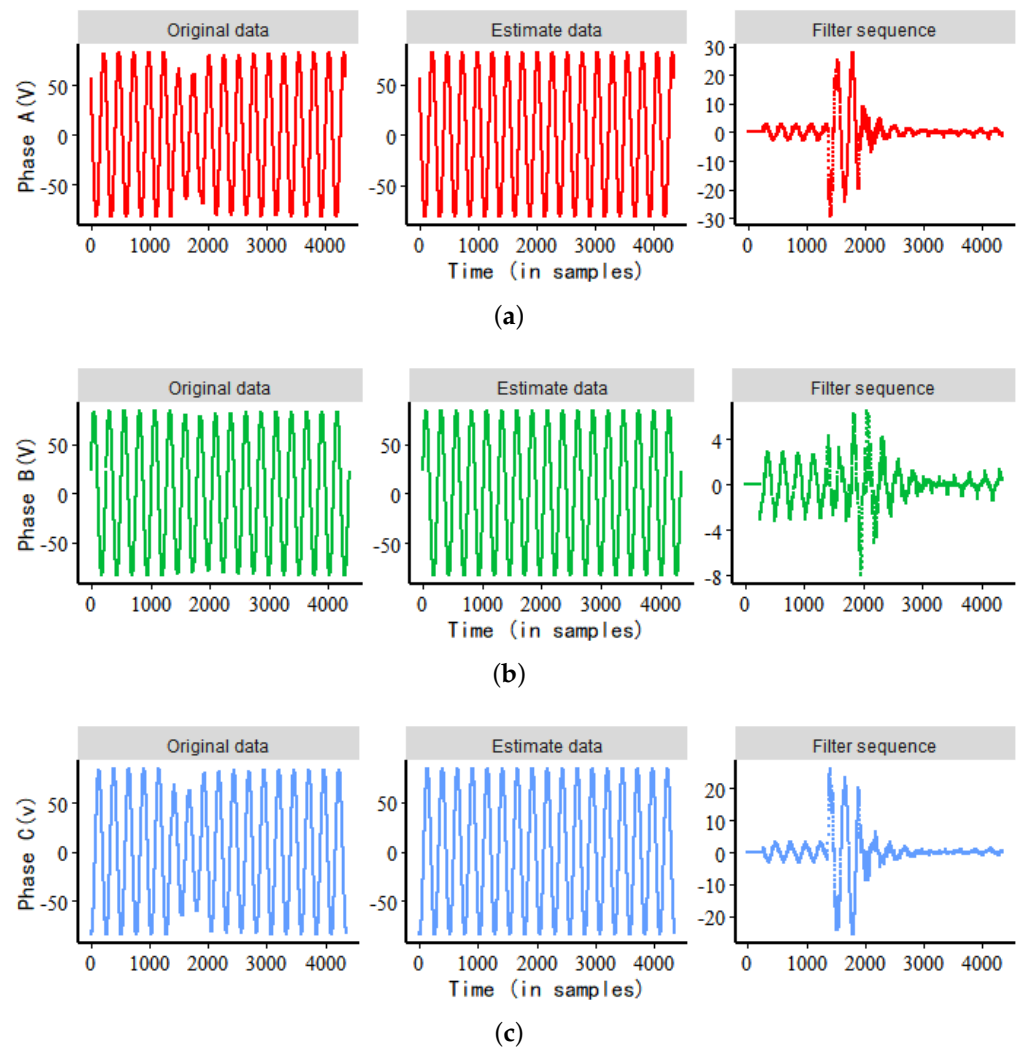
In this paper, we propose to use the local sudden-drop point (sdp) of the kernel density function as the basis for dividing the tail region and the central region of the kernel density curve, where the tail region of the kernel density curve is the anomaly data. In this way, the sdp is also called the DP. The local sudden-drop point of  $f(x)$  is defined as the location at which the density curve falls rapidly from high to low. A formal definition is given as follows.

**Definition 1.** (Local sudden-drop point) For any point  $x$ , the interval  $(x - d, x + d)$  is considered as the  $d$ -neighborhood of  $x$ . Then,  $x_0$  is a local sudden-drop point if  $x_0 = \operatorname{argmax}_x \frac{d^2 f(x)}{dx^2}$  for all  $x \in (x_0 - d, x_0 + d)$ , when  $\frac{df(x)}{dx} < 0$ ; see [24].

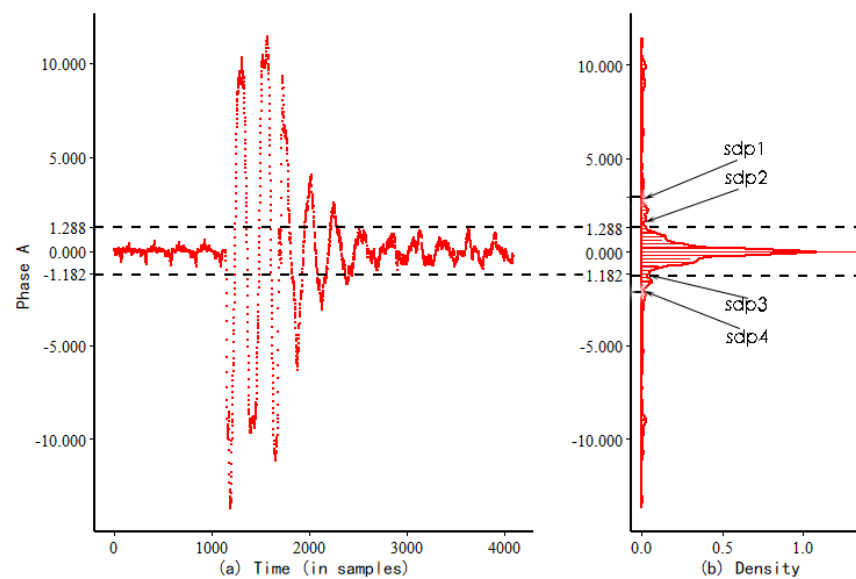
**Remark 1.** Through the definition of the sdp, we can conclude that if the voltage sags, the sdp of the kernel density function of the voltage data must exist, but there may be multiple sharp-drop points, such as sdp 1, sdp 2, sdp 3, and sdp 4 in Figure 6. In power data analysis, we can choose sdp 2 and sdp 3 as the division of the tail region and the central region of the kernel density function, where the region between sdp 2 and sdp 3 is the central region of the kernel density function.



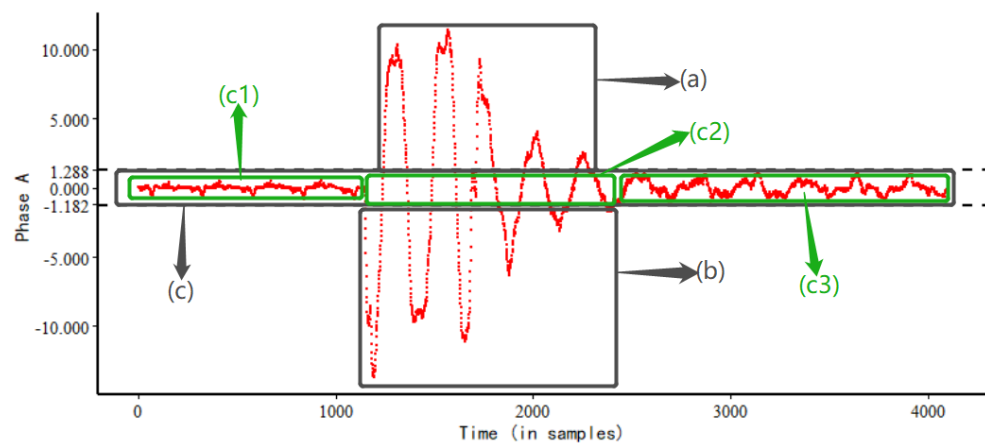
**Figure 3.** The processing of Step 2 on voltage sag data (id 01), where “Actual” represents the actual response, “RMS” represents the real data after passing the RMS detection, “Estimate” represents  $\hat{x}_l, l = 1, 2, \dots, K$ , and the subfigures (a–c) represent the three-phase voltage sag data respectively.



**Figure 4.** The processing of Step 3 on voltage sag data (id 01), where subfigures (a–c) represent the three-phase voltage sag data respectively.



**Figure 5.** Projection fast segmentation algorithm of A-phase voltage sag data (id 01), where (a) is the filter sequence of the A-phase (id 01) and (b) is the corresponding kernel density curve.



**Figure 6.** Region segmentation of the filter sequence (id 01).

### 2.1.3. Extracting Voltage Anomaly Segment

The region segmentation of the filtered sequence is shown in Figure 6. From Figure 6, we can see that the filtering sequence can be divided into three regions (a), (b), and (c) by *sdp 2* and *sdp 3*. It is obvious that the voltage of the (a) and (b) regions has voltage anomalies. In fact, the voltage anomalies also occurred in the (c2) region of the (c) region. To this end, we propose a method for identifying the (c2) region in the (c) region as an anomaly region.

Through the above, a discontinuous voltage sequence can be obtained:

$$\{\varepsilon_{l_1}, \varepsilon_{l_1+1}, \dots, \varepsilon_{l_1+p_1}\}, \dots, \{\varepsilon_{l_s}, \varepsilon_{l_s+1}, \dots, \varepsilon_{l_s+p_s}\}, \dots, \{\varepsilon_{l_n}, \varepsilon_{l_n+1}, \dots, \varepsilon_{l_n+p_n}\},$$

where the total number of sequences is  $n$ ,  $s = 1, 2, \dots, n$ . If the discontinuity distance  $h = (l_s + p_s) - (l_{s-1} + p_{s-1})$  between adjacent sequences is less than one period ( $K$ ), we consider that the sequence point between the discontinuities has a sag, and based on the above method, the (c2) region in the (c) region can be identified as an anomaly.

Figure 7 shows the results of anomaly detection using the FPSA for voltage sag data (id 01), where the dashed lines indicate the respective start and end positions of the anomaly and the shaded portion is the anomaly segment. According to Figure 7, it can be concluded that the start and end positions of the A-phase are 1406 and 2694, respectively (those of the B-phase are 1405 and 2746 and the C-phase are 1404 and 2718).

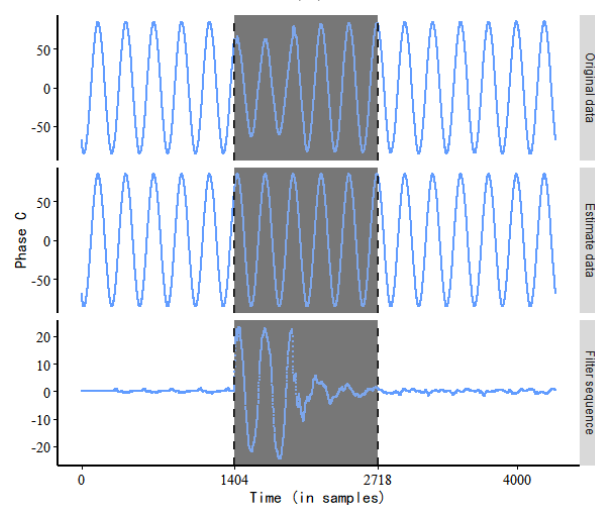
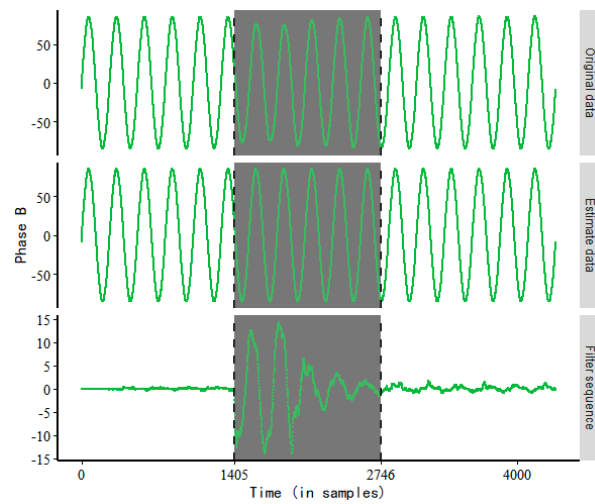
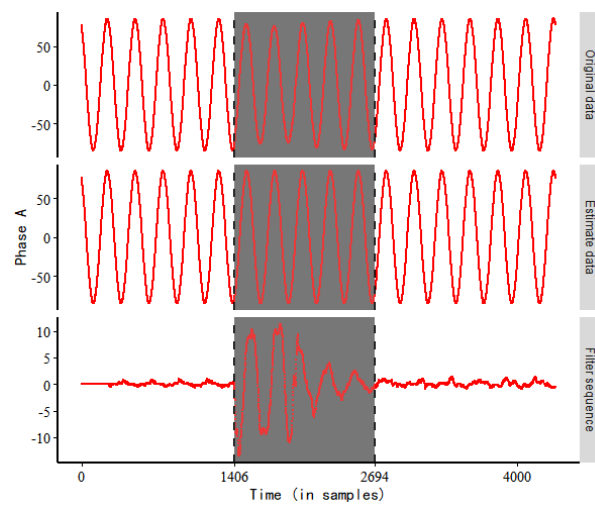
## 2.2. Clustering Analysis of Voltage Sag Events

In this section, shapeDTW and spectral clustering are used to match and cluster voltage anomaly waveforms, so as to achieve the normalization of voltage sag events. The clustering analysis method is mainly divided into two steps. The first step uses shapeDTW to calculate the shape similarity between the anomaly curves extracted by the FPSA. The second step uses the obtained similarity values from the first step to construct the similarity matrix of spectral clustering, and then, the voltage anomaly curve is classified by spectral clustering.

### 2.2.1. Waveform Matching Based on Shapedtw

The DTW was first introduced in the 1970's for audio analysis by [25,26] before being used for general time series analysis. By minimizing the cumulative distance between the signals  $(x(i))_{1 \leq i \leq M}$  and  $(y(j))_{1 \leq j \leq Q}$  (time series), DTW provides a non-linear alignment optimal path between these two time series. It is a relatively straightforward process that first estimates the distance map  $d$  between the signals with elements  $d_{i,j}$  given by:

$$d_{i,j} = (x(i) - y(j))^2 \quad (2)$$



**Figure 7.** FPSA for voltage sag data (id 01), where subfigures (a–c) represent the three-phase voltage sag data respectively.



A cumulative distance map  $D$  with elements  $D_{i,j}$  is then computed from the distance map  $d$ .  $D_{i,j}$  represents the minimal accumulated distance to reach the point  $(i, j)$  starting from the origin  $(1, 1)$ . It is given by:

$$D_{i,j} = d_{i,j} + \min \begin{cases} D_{i,j-1} \\ D_{i-1,j} \\ D_{i-1,j-1} \end{cases} \quad i = 2, \dots, M \quad j = 2, \dots, Q, \quad (3)$$

where we have the following initial conditions:

$$D_{1,1} = d_{1,1}, \quad (4)$$

$$D_{1,j} = \sum_{p=1}^j d_{1,p}, \quad j = 1, \dots, Q, \quad (5)$$

$$D_{i,1} = \sum_{q=1}^i d_{q,1}, \quad i = 1, \dots, M. \quad (6)$$

This cumulative distance is used to define a warping path between signals  $(x(i))_{1 \leq i \leq M}$  and  $(y(j))_{1 \leq j \leq Q}$ , denoted  $\phi_{xy}$ :

$$\phi_{xy} : \begin{cases} \llbracket 1; S \rrbracket & \longrightarrow \llbracket 1; M \rrbracket \times \llbracket 1; N \rrbracket \\ s & \longmapsto \phi_{xy}(s) = (\phi_{xy}^x(s), \phi_{xy}^y(s)) \end{cases} \quad (7)$$

$\phi_{xy}$  verifies several constraints:

1. The monotonicity constraint guarantees the time ordering.
2. The boundary constraints:  $\phi_{xy}(1) = (1, 1)$  and  $\phi_{xy}(S) = (M, Q)$ .
3. The step size conditions:  $0 \leq \phi_{xy}^x(s) - \phi_{xy}^x(s-1) \leq 1$  and  $0 \leq \phi_{xy}^y(s) - \phi_{xy}^y(s-1) \leq 1 \quad \forall s \in 2 \dots S$ ; see Equation (3).

This path minimizes the final cumulative distance. Its length  $S$  depends on the signals to be aligned and is determined during the DTW process.

Figure 8 presents two examples of DTW alignments between two time series, with Figure 8a,b showing the alignments between series and Figure 8c,d showing the cumulative distance maps  $D$  and the warping paths in green for these two examples.

Although DTW obtains a global optimal solution, it does not necessarily achieve locally sensible matchings. Concretely, two temporal points with entirely dissimilar local structures may be matched by DTW. To address this problem, [27] proposed an improved alignment algorithm, named shape dynamic time warping (shapeDTW), which enhances DTW by taking pointwise local structural information into consideration.

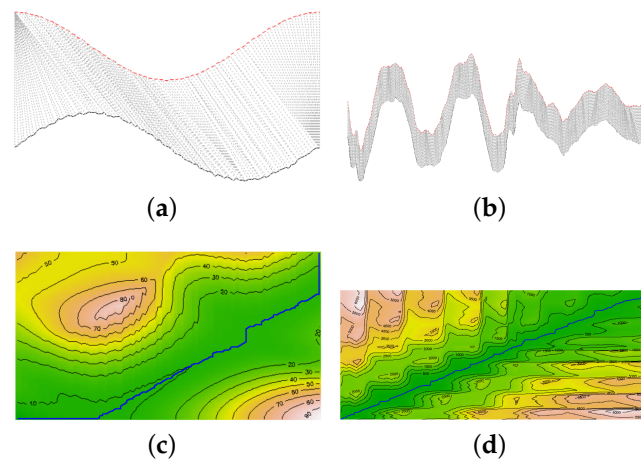
ShapeDTW is a temporal alignment algorithm, which consists of two sequential steps: Firstly, represent each temporal point by some shape descriptor, which encodes structural information of local subsequences around that point; in this way, the original time series is converted into a sequence of descriptors. Secondly, use DTW to align two sequences of descriptors. The shapeDTW time series alignment procedure is shown in Figure 9 [27]. This section mainly uses the shapeDTW algorithm to calculate the shape similarity between voltage sag sequences extracted by the FPSA.

**Remark 2.** Reference [27] gave four shape descriptors: piecewise aggregate approximation (PAA), discrete wavelet transform (DWT), slope, and derivative. Derivative is used in this paper.

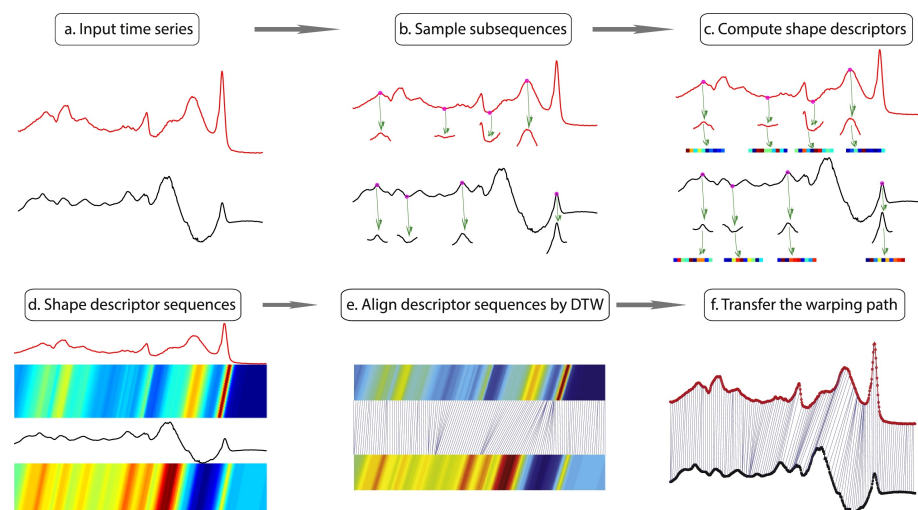
### 2.2.2. Spectral Clustering Based on shapeDTW

The spectral clustering algorithm is based on the spectral theory of graph theory and is widely used in image segmentation, computer vision, and pattern recognition.

Spectral clustering is a class of methods based on eigen-decompositions of graph affinity matrices. Given a set of data points  $S = \{x_1, \dots, x_n\}$ , a weighted graph  $G = (V, E)$  is first constructed for which every vertex corresponds to a point in  $S$  and each edge is weighted by the similarity between the connected points. The Laplacian graph  $L$  [28] is then derived from the adjacent matrix  $W$  of  $G$ , and the eigenvectors of  $L$  are computed. Finally, the traditional  $K$ -means method is applied to the low-dimensional representations of the original data. There are many spectral clustering algorithms that are based on the above procedures [29–31].



**Figure 8.** Illustration of a DTW process for aligning two pairs of signals (one per column). The first row: the matching between points of two pairs of signals (in red and black) is symbolized by grey lines. The second row: superimposition of the warping path on the cumulative distance matrix  $D$ . Yellow areas of  $D$  correspond to larger cumulative distances, whereas green ones correspond to smaller ones. The first case of alignment (a,c) is an example of speech data. The second case of alignment (b,d) is an example of voltage sag data. The two pairs of signals are the phase A of voltage sag data id 01 and 02, respectively.



**Figure 9.** Pipeline of shapeDTW. shapeDTW consists of two major steps: encode local structures by shape descriptors, and align descriptor sequences by DTW. Concretely, we sample a subsequence from each temporal point and further encode it by some shape descriptor. As a result, the original time series is converted into a descriptor sequence of the same length. Then, we align two descriptor sequences by DTW and transfer the found warping path to the original time series.

In this paper, our proposed shapeDTW spectral clustering algorithm is developed in the Jordan–Weiss (NJW) framework [31]. Therefore, the NJW algorithm is briefly reviewed

as Algorithm 1, for the sake of completeness. Algorithm 1 mainly involves the determination of the number of clusters  $k$ , the calculation of affinity matrix  $W$ , and the calculation of Laplacian matrix  $L$ .

---

**Algorithm 1** NJW spectral clustering algorithm.

---

**Input:** Dataset  $S = \{x_1, \dots, x_n\}$  in  $\mathfrak{R}^1$  and the number of clusters  $k$

**Output:**  $k$ -way partition of the input data

Construct the affinity matrix  $W$ :

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix};$$

2: The degree matrix  $D$  is defined as the diagonal matrix with the degrees  $d_1, \dots, d_n$  on the diagonal,

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}, \text{ where } d_i = \sum_{j=1}^n w_{ij};$$

Compute Laplacian matrix  $L$ ;

4: Compute the  $k$  eigenvectors of  $L$ ,  $e_1, e_2, \dots, e_k$ , which are associated with the  $k$  largest eigenvalues, and form the matrix  $\mathbf{X} = [e_1, e_2, \dots, e_k]$ ;

Renormalize each row to form a new matrix  $\mathbf{Y} \in \mathfrak{R}^{n \times k}$  with  $\mathbf{Y}_{ij} = \mathbf{X}_{ij} / \left( \sum_j \mathbf{X}_{ij}^2 \right)^{1/2}$ , so that each row of  $\mathbf{Y}$  has a unit magnitude;

6: Treat each row of  $\mathbf{Y}$  as a point in  $\mathfrak{R}^k$ , and partition the  $n$  points ( $n$  rows) into  $k$  clusters via a general cluster algorithm, such as the  $K$ -means algorithm;

Assign the original point  $x_i$  to the cluster  $c$  if and only if the corresponding row  $i$  of the matrix  $\mathbf{Y}$  is assigned to the cluster  $c$ .

---

(1) There are mainly three methods for determining the number of clusters  $k$ : elbow method, average silhouette method, and gap statistic method. In this paper, we adopted the average silhouette method for determining the number of clusters  $k$ .

(2) The methods for calculating the affinity matrix  $W$  are mainly the  $\varepsilon$ -neighborhood graph,  $k$ -nearest neighbor graph, and fully connected graph. In this paper, the fully connected graph is used to calculate  $W$ .

$$w_{ij} = w_{ji} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), & i \neq j \\ 0 & , i = j \end{cases} \quad (8)$$

where  $\|x_i - x_j\|^2$  represents the Euclidean distance between  $x_i$  and  $x_j$ . In this paper,  $x_i$  and  $x_j$  represent the voltage filtering sequence. Since the lengths of  $x_i$  and  $x_j$  are different, the Euclidean distance between  $x_i$  and  $x_j$  cannot be calculated. For this, we used the shapeDTW similarity between  $x_i$  and  $x_j$  instead of the Euclidean distance between  $x_i$  and  $x_j$ . Then,

$$w_{ij} = w_{ji} = \begin{cases} \exp\left(\frac{-\text{shapeDTW}(x_i, x_j)}{25\sigma^2}\right), & i \neq j \\ 0, & i = j \end{cases} \quad (9)$$

where  $\text{shapeDTW}(x_i, x_j)$  represents the shapeDTW similarity between  $x_i$  and  $x_j$ . We aimed to improve spectral clustering by shapeDTW, so that it can cluster datasets of voltage filtering sequences.

(3) There are mainly three methods for the calculation of Laplacian matrices  $\mathbf{L}$ : unnormalized Laplacian matrices, random walk Laplacian matrices, and symmetric Laplacian matrices [29]. In this paper, we adopted symmetric Laplacian matrices, so:

$$\mathbf{L} = \mathbf{D}^{-1/2}(\mathbf{D} - \mathbf{W})\mathbf{D}^{-1/2}. \quad (10)$$

### 2.3. Procedure Flow Chart of the Methodology

A procedure flow chart of the proposed methodology is presented in Figure 10.

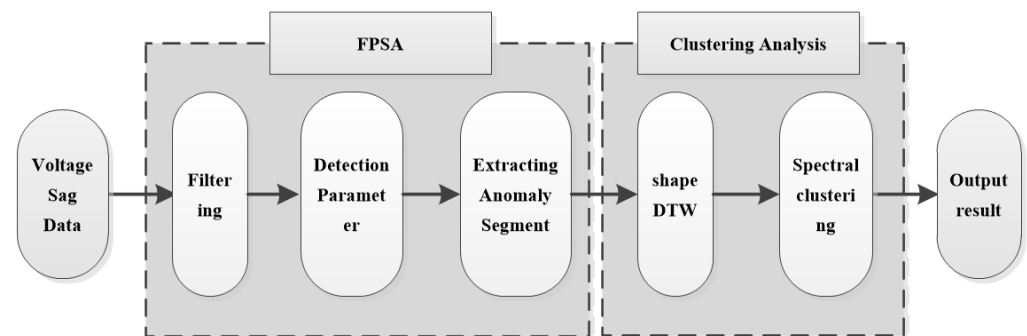


Figure 10. Procedure flow chart of the methodology.

## 3. Empirical Analysis

### 3.1. Data

To verify the effectiveness of the method proposed in this paper, real data samples from a voltage sag monitoring system were selected. These data recorded the voltage sag events that occurred in Beijing in May 2018. In total, twenty valid samples were retained after excluding invalid samples. These samples belonged to three types of sag events. The sag types and number of data samples are shown in Table 1. According to the process in Figure 10, firstly, the start position and end position of the anomaly of the voltage sag data are detected. Secondly, the anomaly waveform is extracted. Thirdly, the anomaly waveform is classified to achieve the classification purpose of the voltage sag events.

Table 1. The sag types and number of data samples.

Type	Number	id
1	10	01, 02, 03, 04, 05, 06, 07, 08, 09, 10
2	6	11, 12, 13, 14, 15, 16
3	4	17, 18, 19, 20

### 3.2. Empirical Results

To illustrate the accuracy of the FPSA detection results, we compared the FPSA with the RMS method [16]. The results are shown in Table 2, where “RMS” represents the root mean square of the half-period refresh voltage detection method and “FPSA” represents the FPSA. The formula for calculating the RMS is as follows:

$$U_{\text{rms}}(k) = \sqrt{\frac{1}{N} \sum_{i=1+(k-1)\frac{N}{2}}^{(k+1)\frac{N}{2}} u^2(i)} \quad (11)$$

where  $N$  is the number of sampling points per period ( $K = 256$ ),  $u(i)$  is the instantaneous value of the voltage waveform acquired for the  $i$ th time, and  $k$  is the calculated window ( $k = 1, 2, \dots, K$ ). The threshold of the detection voltage was set to 0.9 pu according to the definition of a voltage sag.

**Table 2.** Detection results of voltage (*start position, stop position*).

id	A-Phase		B-Phase		C-Phase	
	RMS	FPSA	RMS	FPSA	RMS	FPSA
01	[1600,1888]	[1406,2694]	[1602,1893]	[1405,2746]	[1605,1886]	[1404,2718]
02	[1569,1950]	[1461,2848]	[1604,1912]	[1460,2797]	[1610,1945]	[1460,2627]
03	[1593,1862]	[1379,2768]	[1556,1885]	[1379,2718]	[1560,1886]	[1377,2680]
04	[1627,1818]	[1420,2663]	[1612,1896]	[1419,2747]	[1607,1903]	[1417,2704]
05	[1591,1860]	[1381,2793]	[1586,1879]	[1376,3233]	[1580,1888]	[1376,2944]
06	[1611,1895]	[1409,2802]	[1601,1902]	[1407,3050]	[1609,1899]	[1407,2714]
07	[1563,1847]	[1364,2782]	[1567,1866]	[1364,2706]	[1568,1857]	[1363,2667]
08	[1598,1777]	[1376,2985]	[1603,1786]	[1373,2932]	[1607,1791]	[1371,2934]
09	[1591,1804]	[1394,2870]	[1586,1810]	[1393,2735]	[1593,1805]	[1392,2694]
10	[1530,1910]	[1427,2868]	[1528,1905]	[1426,2755]	[1550,1912]	[1425,2677]
11	[1431,1885]	[1368,2263]	[1422,1887]	[1408,2600]	[1431,1879]	[1369,2330]
12	[1471,1927]	[1409,2306]	[1476,1931]	[1417,2640]	[1481,1933]	[1409,2370]
13	[1471,1926]	[1408,2306]	[1478,1935]	[1429,2639]	[1476,1931]	[1409,2370]
14	[1468,1916]	[1402,2304]	[1470,1918]	[1399,2796]	[1471,1821]	[1402,2451]
15	[1404,1953]	[1439,2341]	[1408,1962]	[1436,2707]	[1402,1956]	[1439,2487]
16	[1485,1928]	[1406,2834]	[1486,1931]	[1416,2666]	[1479,1929]	[1406,2629]
17	[1247,2130]	[971,3247]	[1251,2133]	[975,3275]	[1248,2110]	[876,3121]
18	[1161,1793]	[1032,2943]	[1170,1812]	[923,3068]	[1173,1821]	[1036,3070]
19	[1249,2380]	[1213,3389]	[1253,2376]	[1118,3515]	[1257,2374]	[1128,3654]
20	[1394,1613]	[314,3120]	[1402,1619]	[314,3123]	[1407,1609]]	[314,3121]

From Table 2, we can conclude that the sag segments detected by the RMS are all inside the abnormal segment detected by the FPSA, so the abnormal segment contains more information than the sag segment. Therefore, it is reasonable to use the abnormal segment for clustering analysis of the voltage sag events.

We used the method proposed in this paper to cluster the voltage sag events for A-phase voltage data, B-phase voltage data, and C-phase voltage data in the voltage sag data, respectively. The results of the clustering analysis of A-phase voltage data, B-phase voltage data, and C-phase voltage data are shown in Tables 3–5 and Figures 11–13, respectively. Tables 3–5 is an affinity matrix  $\mathbf{W}_A$  ( $\mathbf{W}_B$ ,  $\mathbf{W}_C$ ) calculated using the abnormal waveform segment of the A-phase (B-phase, C-phase) voltage data. The area of the same

color (gray, green, and yellow) in  $W_A$  ( $W_B$ ,  $W_C$ ) represents the voltage sag caused by the same event. The calculation method of  $W_A$  ( $W_B$ ,  $W_C$ ) is in Equation (7). Figures 11–13 show the clustering results of the abnormal voltage segments in the A-phase (B-phase, C-phase) voltage data.

**Table 3.** The affinity matrix  $W_A$ .

id	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	0.000	0.981	0.963	0.984	0.939	0.969	0.974	0.944	0.977	0.987	0.804	0.794	0.795	0.800	0.803	0.886	0.777	0.705	0.740	0.736
02	0.981	0.000	0.970	0.974	0.953	0.976	0.980	0.951	0.983	0.980	0.812	0.802	0.805	0.810	0.812	0.901	0.788	0.711	0.759	0.744
03	0.963	0.970	0.000	0.965	0.977	0.987	0.982	0.965	0.985	0.956	0.765	0.758	0.759	0.764	0.767	0.858	0.778	0.798	0.745	0.731
04	0.984	0.974	0.965	0.000	0.944	0.973	0.973	0.945	0.977	0.984	0.801	0.789	0.791	0.801	0.805	0.881	0.774	0.703	0.737	0.730
05	0.939	0.953	0.977	0.944	0.000	0.973	0.967	0.970	0.972	0.931	0.747	0.740	0.742	0.754	0.757	0.837	0.787	0.788	0.754	0.726
06	0.969	0.976	0.987	0.973	0.973	0.000	0.991	0.965	0.990	0.962	0.773	0.764	0.765	0.770	0.773	0.864	0.775	0.798	0.740	0.729
07	0.974	0.980	0.982	0.973	0.967	0.991	0.000	0.960	0.990	0.966	0.783	0.774	0.775	0.781	0.784	0.872	0.777	0.701	0.742	0.733
08	0.944	0.951	0.965	0.945	0.970	0.965	0.960	0.000	0.963	0.937	0.757	0.750	0.751	0.752	0.756	0.839	0.808	0.709	0.774	0.748
09	0.977	0.983	0.985	0.977	0.972	0.990	0.990	0.963	0.000	0.971	0.784	0.774	0.775	0.781	0.785	0.874	0.781	0.704	0.747	0.735
10	0.987	0.980	0.956	0.984	0.931	0.962	0.966	0.937	0.971	0.000	0.813	0.802	0.804	0.810	0.815	0.892	0.777	0.707	0.741	0.736
11	0.804	0.812	0.765	0.801	0.747	0.773	0.783	0.757	0.784	0.813	0.000	0.985	0.985	0.967	0.963	0.920	0.686	0.776	0.699	0.657
12	0.794	0.802	0.758	0.789	0.740	0.764	0.774	0.750	0.774	0.802	0.985	0.000	0.988	0.967	0.965	0.916	0.669	0.762	0.690	0.629
13	0.795	0.805	0.759	0.791	0.742	0.765	0.775	0.751	0.775	0.804	0.985	0.988	0.000	0.966	0.964	0.917	0.680	0.770	0.695	0.652
14	0.800	0.810	0.764	0.801	0.754	0.770	0.781	0.752	0.781	0.810	0.967	0.967	0.966	0.000	0.982	0.916	0.686	0.767	0.702	0.665
15	0.803	0.812	0.767	0.805	0.757	0.773	0.784	0.756	0.785	0.815	0.963	0.965	0.964	0.982	0.000	0.921	0.682	0.770	0.697	0.656
16	0.886	0.901	0.858	0.881	0.837	0.864	0.872	0.839	0.874	0.892	0.920	0.916	0.917	0.916	0.921	0.000	0.763	0.864	0.770	0.727
17	0.777	0.788	0.778	0.774	0.787	0.775	0.777	0.808	0.781	0.777	0.686	0.669	0.680	0.686	0.682	0.763	0.000	0.867	0.939	0.885
18	0.705	0.711	0.798	0.703	0.788	0.798	0.701	0.709	0.704	0.707	0.776	0.762	0.770	0.767	0.770	0.864	0.867	0.000	0.838	0.811
19	0.740	0.759	0.745	0.737	0.754	0.740	0.742	0.774	0.747	0.741	0.699	0.690	0.695	0.702	0.697	0.770	0.939	0.838	0.000	0.843
20	0.736	0.744	0.731	0.730	0.726	0.729	0.733	0.748	0.735	0.736	0.657	0.629	0.652	0.665	0.656	0.727	0.885	0.811	0.843	0.000

**Table 4.** The affinity matrix  $W_B$  of B-phase voltage sag data.

id	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	0.000	0.986	0.977	0.986	0.942	0.972	0.985	0.952	0.985	0.986	0.777	0.777	0.777	0.858	0.854	0.874	0.800	0.877	0.720	0.734
02	0.986	0.000	0.973	0.985	0.935	0.969	0.984	0.948	0.985	0.988	0.766	0.765	0.766	0.845	0.842	0.863	0.797	0.876	0.722	0.732
03	0.977	0.973	0.000	0.978	0.955	0.978	0.986	0.966	0.986	0.967	0.786	0.786	0.787	0.859	0.856	0.884	0.805	0.879	0.713	0.730
04	0.986	0.985	0.978	0.000	0.947	0.975	0.983	0.954	0.985	0.984	0.776	0.776	0.777	0.852	0.850	0.872	0.802	0.880	0.727	0.735
05	0.942	0.935	0.955	0.947	0.000	0.967	0.951	0.975	0.949	0.930	0.779	0.777	0.776	0.839	0.834	0.855	0.833	0.907	0.747	0.754
06	0.972	0.969	0.978	0.975	0.967	0.000	0.979	0.976	0.979	0.963	0.779	0.778	0.778	0.858	0.852	0.873	0.817	0.893	0.724	0.744
07	0.985	0.984	0.986	0.983	0.951	0.979	0.000	0.960	0.991	0.979	0.783	0.783	0.783	0.862	0.858	0.882	0.804	0.879	0.714	0.732
08	0.952	0.948	0.966	0.954	0.975	0.976	0.960	0.000	0.960	0.942	0.776	0.776	0.776	0.840	0.840	0.865	0.833	0.910	0.750	0.759
09	0.985	0.985	0.986	0.985	0.949	0.979	0.991	0.960	0.000	0.981	0.782	0.783	0.784	0.862	0.858	0.882	0.802	0.879	0.715	0.731
10	0.986	0.988	0.967	0.984	0.930	0.963	0.979	0.942	0.981	0.000	0.763	0.764	0.764	0.842	0.840	0.859	0.792	0.870	0.722	0.730
11	0.777	0.766	0.786	0.776	0.779	0.779	0.783	0.776	0.782	0.763	0.000	0.990	0.994	0.911	0.914	0.891	0.745	0.758	0.654	0.738
12	0.777	0.765	0.786	0.776	0.777	0.778	0.783	0.776	0.783	0.764	0.990	0.000	0.991	0.909	0.913	0.890	0.740	0.753	0.652	0.730
13	0.777	0.766	0.787	0.777	0.776	0.778	0.783	0.776	0.784	0.764	0.994	0.991	0.000	0.910	0.914	0.891	0.739	0.753	0.651	0.730
14	0.858	0.845	0.859	0.852	0.839	0.858	0.862	0.840	0.862	0.842	0.911	0.909	0.910	0.000	0.985	0.972	0.713	0.769	0.612	0.659
15	0.854	0.842	0.856	0.850	0.834	0.852	0.858	0.840	0.858	0.840	0.914	0.913	0.914	0.985	0.000	0.981	0.713	0.773	0.614	0.658
16	0.874	0.863	0.884	0.872	0.855	0.873	0.882	0.865	0.882	0.859	0.891	0.890	0.891	0.972	0.981	0.000	0.716	0.784	0.615	0.652
17	0.800	0.797	0.805	0.802	0.833	0.817	0.804	0.833	0.802	0.792	0.745	0.740	0.739	0.713	0.713	0.716	0.000	0.902	0.902	0.885
18	0.877	0.876	0.879	0.880	0.907	0.893	0.879	0.910	0.879	0.870	0.758	0.753	0.753	0.769	0.773	0.784	0.902	0.000	0.841	0.819
19	0.720	0.722	0.713	0.727	0.747	0.724	0.714	0.750	0.715	0.722	0.654	0.652	0.651	0.612	0.614	0.615	0.902	0.841	0.000	0.858
20	0.734	0.732	0.730	0.735	0.754	0.744	0.732	0.759	0.731	0.730	0.738	0.730	0.730	0.659	0.658	0.652	0.885	0.819	0.858	0.000

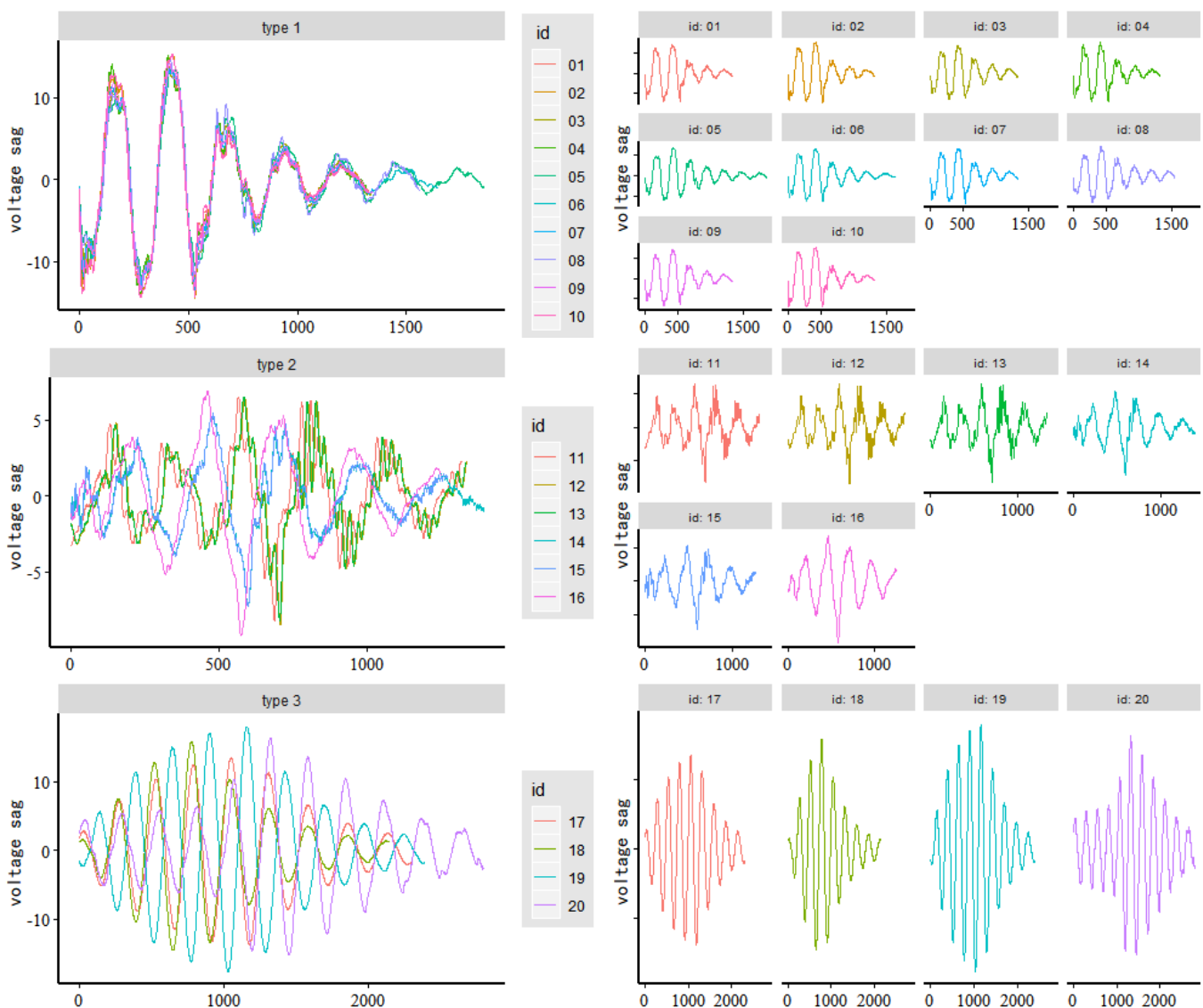


Figure 11. Clustering result based on A-phase voltage sag data.

From Tables 3–5, we can see that the value of the same color area is obviously larger than the value of the blank area, indicating that the abnormal waveform segment of the voltage sag caused by the same event is more similar. It can be obtained from Figures 11–13 that the clustering accuracy of the clustering analysis method proposed in this paper is 100% for the voltage sag data (id 01–20).

Table 6 shows the clustering results of the affinity matrix of different abnormal voltage segments using the average of the affinity matrix of the voltage data of phase A, phase B, and phase C, where  $W_{ABC} = \frac{W_A + W_B + W_C}{3}$ .  $W_{ABC}$  has the same clustering results as  $W_A$ ,  $W_B$ , and  $W_C$ .

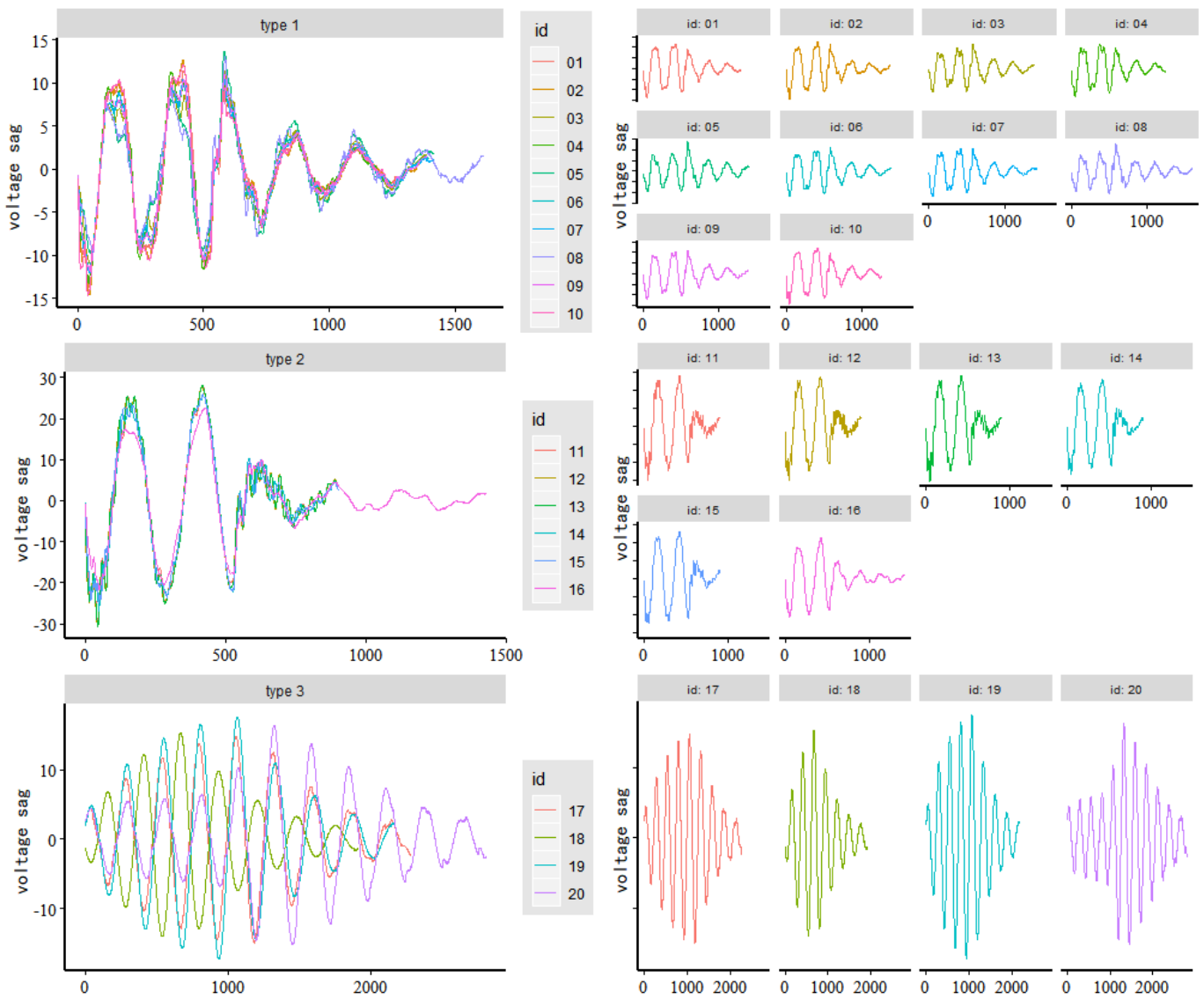


Figure 12. Clustering result based on B-phase voltage sag data.

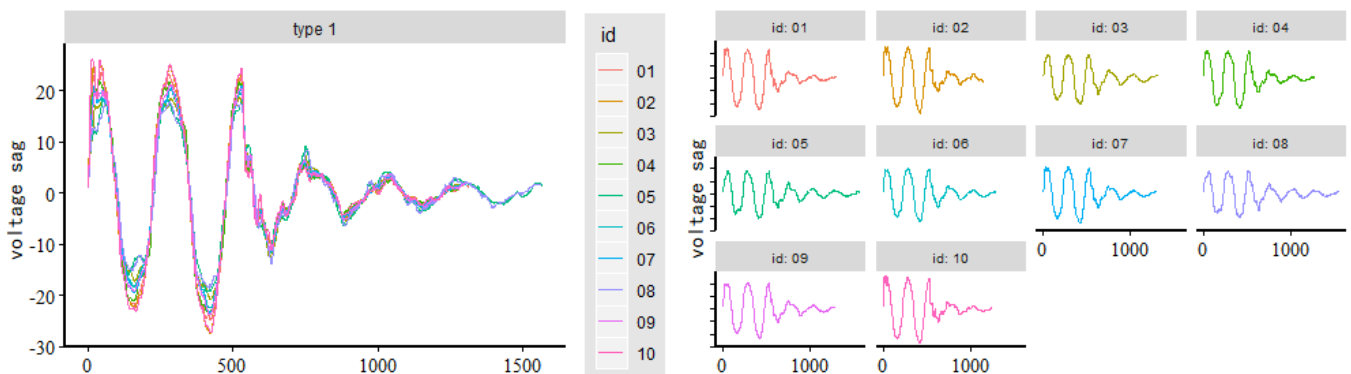


Figure 13. Cont.



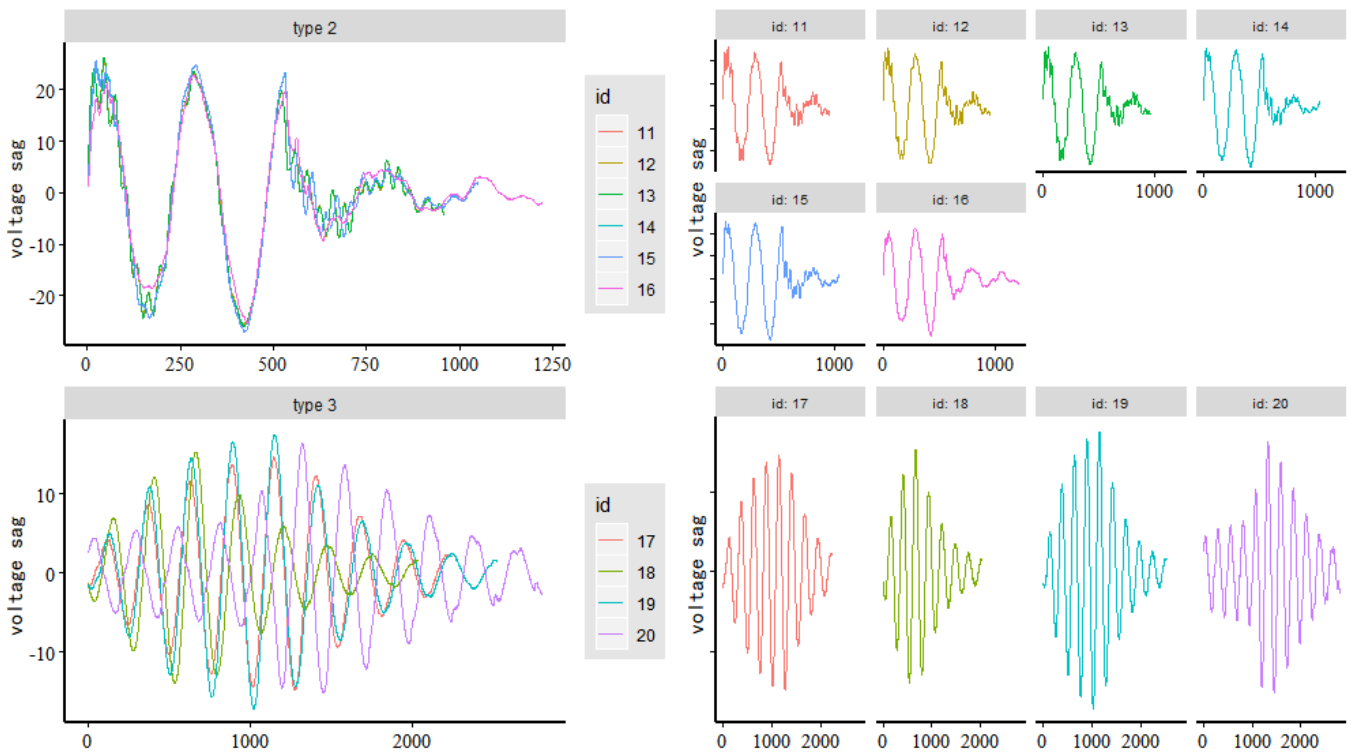


Figure 13. Clustering result based on C-phase voltage sag data.

Table 5. The affinity matrix  $W_C$  of C-phase voltage sag data.

id	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
01	0.000	0.967	0.947	0.980	0.917	0.966	0.974	0.909	0.972	0.978	0.932	0.935	0.932	0.958	0.957	0.955	0.702	0.786	0.714	0.678
02	0.967	0.000	0.978	0.946	0.949	0.946	0.954	0.954	0.954	0.924	0.961	0.894	0.944	0.954	0.886	0.962	0.679	0.757	0.691	0.659
03	0.947	0.978	0.000	0.960	0.964	0.984	0.975	0.961	0.976	0.925	0.898	0.901	0.897	0.911	0.912	0.949	0.764	0.833	0.768	0.728
04	0.980	0.946	0.960	0.000	0.932	0.975	0.981	0.924	0.979	0.955	0.928	0.932	0.928	0.946	0.947	0.949	0.711	0.795	0.726	0.683
05	0.917	0.949	0.964	0.932	0.000	0.954	0.946	0.982	0.946	0.941	0.877	0.879	0.876	0.891	0.891	0.939	0.808	0.882	0.805	0.768
06	0.966	0.946	0.984	0.975	0.954	0.000	0.987	0.950	0.987	0.944	0.912	0.915	0.911	0.929	0.929	0.945	0.746	0.824	0.752	0.713
07	0.974	0.954	0.975	0.981	0.946	0.987	0.000	0.940	0.989	0.951	0.921	0.925	0.921	0.938	0.938	0.939	0.731	0.814	0.740	0.701
08	0.909	0.954	0.961	0.924	0.982	0.950	0.940	0.000	0.940	0.948	0.871	0.875	0.870	0.884	0.885	0.941	0.814	0.883	0.812	0.772
09	0.972	0.954	0.976	0.979	0.946	0.987	0.989	0.940	0.000	0.964	0.895	0.846	0.858	0.889	0.838	0.941	0.734	0.815	0.743	0.704
10	0.978	0.924	0.925	0.955	0.941	0.944	0.951	0.948	0.964	0.000	0.939	0.940	0.939	0.960	0.960	0.962	0.687	0.764	0.695	0.663
11	0.932	0.961	0.898	0.928	0.877	0.912	0.921	0.871	0.895	0.939	0.000	0.987	0.996	0.956	0.955	0.949	0.677	0.751	0.684	0.662
12	0.935	0.894	0.901	0.932	0.879	0.915	0.925	0.875	0.846	0.940	0.987	0.000	0.987	0.957	0.957	0.962	0.683	0.755	0.690	0.671
13	0.932	0.944	0.897	0.928	0.876	0.911	0.921	0.870	0.858	0.939	0.996	0.987	0.000	0.956	0.955	0.923	0.676	0.749	0.683	0.660
14	0.958	0.954	0.911	0.946	0.891	0.929	0.938	0.884	0.889	0.960	0.956	0.957	0.956	0.000	0.986	0.962	0.682	0.760	0.691	0.658
15	0.957	0.886	0.912	0.947	0.891	0.929	0.938	0.885	0.838	0.960	0.955	0.957	0.955	0.986	0.000	0.965	0.684	0.762	0.693	0.661
16	0.955	0.962	0.949	0.949	0.939	0.945	0.939	0.941	0.941	0.962	0.949	0.962	0.923	0.962	0.965	0.000	0.714	0.805	0.715	0.689
17	0.702	0.679	0.764	0.711	0.808	0.746	0.731	0.814	0.734	0.687	0.677	0.683	0.676	0.682	0.684	0.714	0.000	0.868	0.934	0.872
18	0.786	0.757	0.833	0.795	0.882	0.824	0.814	0.883	0.815	0.764	0.751	0.755	0.749	0.760	0.762	0.805	0.868	0.000	0.833	0.804
19	0.714	0.691	0.768	0.726	0.805	0.752	0.740	0.812	0.743	0.695	0.684	0.690	0.683	0.691	0.693	0.715	0.934	0.833	0.000	0.854
20	0.678	0.659	0.728	0.683	0.768	0.713	0.701	0.772	0.704	0.663	0.662	0.671	0.660	0.658	0.661	0.689	0.872	0.804	0.854	0.000

**Table 6.** The affinity matrix  $W_{ABC}$  of voltage sag data.

id	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
1	0.000	0.978	0.962	0.983	0.933	0.969	0.978	0.935	0.978	0.984	0.838	0.835	0.835	0.872	0.871	0.905	0.760	0.856	0.725	0.716
2	0.978	0.000	0.974	0.968	0.946	0.964	0.973	0.951	0.974	0.964	0.846	0.820	0.838	0.870	0.847	0.909	0.755	0.848	0.724	0.712
3	0.962	0.974	0.000	0.968	0.965	0.983	0.981	0.964	0.982	0.949	0.816	0.815	0.814	0.845	0.845	0.897	0.782	0.870	0.742	0.730
4	0.983	0.968	0.968	0.000	0.941	0.974	0.979	0.941	0.980	0.974	0.835	0.832	0.832	0.866	0.867	0.901	0.762	0.859	0.730	0.716
5	0.933	0.946	0.965	0.941	0.000	0.965	0.955	0.976	0.956	0.934	0.801	0.799	0.798	0.828	0.827	0.877	0.809	0.892	0.769	0.749
6	0.969	0.964	0.983	0.974	0.965	0.000	0.986	0.964	0.985	0.956	0.821	0.819	0.818	0.852	0.851	0.894	0.779	0.872	0.739	0.729
7	0.978	0.973	0.981	0.979	0.955	0.986	0.000	0.953	0.990	0.965	0.829	0.827	0.826	0.860	0.860	0.898	0.771	0.865	0.732	0.722
8	0.935	0.951	0.964	0.941	0.976	0.964	0.953	0.000	0.954	0.942	0.801	0.800	0.799	0.825	0.827	0.882	0.818	0.901	0.779	0.760
9	0.978	0.974	0.982	0.980	0.956	0.985	0.990	0.954	0.000	0.972	0.820	0.801	0.806	0.844	0.827	0.899	0.772	0.866	0.735	0.723
10	0.984	0.964	0.949	0.974	0.934	0.956	0.965	0.942	0.972	0.000	0.838	0.835	0.836	0.871	0.872	0.904	0.752	0.847	0.719	0.710
11	0.838	0.846	0.816	0.835	0.801	0.821	0.829	0.801	0.820	0.838	0.000	0.987	0.992	0.945	0.944	0.920	0.703	0.762	0.679	0.686
12	0.835	0.820	0.815	0.832	0.799	0.819	0.827	0.800	0.801	0.835	0.987	0.000	0.989	0.944	0.945	0.923	0.697	0.757	0.677	0.677
13	0.835	0.838	0.814	0.832	0.798	0.818	0.826	0.799	0.806	0.836	0.992	0.989	0.000	0.944	0.944	0.910	0.698	0.757	0.676	0.681
14	0.872	0.870	0.845	0.866	0.828	0.852	0.860	0.825	0.844	0.871	0.945	0.944	0.944	0.000	0.984	0.950	0.694	0.765	0.668	0.661
15	0.871	0.847	0.845	0.867	0.827	0.851	0.860	0.827	0.827	0.872	0.944	0.945	0.944	0.984	0.000	0.956	0.693	0.768	0.668	0.658
16	0.905	0.909	0.897	0.901	0.877	0.894	0.898	0.882	0.899	0.904	0.920	0.923	0.910	0.950	0.956	0.000	0.731	0.818	0.700	0.689
17	0.760	0.755	0.782	0.762	0.809	0.779	0.771	0.818	0.772	0.752	0.703	0.697	0.698	0.694	0.693	0.731	0.000	0.879	0.925	0.881
18	0.856	0.848	0.870	0.859	0.892	0.872	0.865	0.901	0.866	0.847	0.762	0.757	0.757	0.765	0.768	0.818	0.879	0.000	0.837	0.811
19	0.725	0.724	0.742	0.730	0.769	0.739	0.732	0.779	0.735	0.719	0.679	0.677	0.676	0.668	0.668	0.700	0.925	0.837	0.000	0.852
20	0.716	0.712	0.730	0.716	0.749	0.729	0.722	0.760	0.723	0.710	0.686	0.677	0.681	0.661	0.658	0.689	0.881	0.811	0.852	0.000

The results of the above empirical analysis show that the proposed method can effectively achieve the clustering of voltage sag events.

#### 4. Concluding Remarks

With the development of electrical power technology and expansions of the monitoring scope, the amount of voltage sag data has also dramatically increased over time. The increase in the amount of voltage sag data generates a huge amount of work for power engineers due to the need for the analysis of the data. In addition, accurate clock synchronization between voltage sag events is difficult to achieve due to the costs of monitoring such power systems. Therefore, it is impossible to know whether the voltage sag events detected by different monitoring stations correspond to the same event. Therefore, valid information cannot be processed in time, and valuable information is discarded. The clustering analysis method proposed in this paper is effective in solving this problem, and the test results obtained based on real data reflect the accuracy of the approach. The voltage anomaly waveform of voltage sags caused by the same event has a high degree of similarity, and the voltage anomaly waveform of the voltage sags caused by different events is significantly different. Our method can effectively judge whether voltage sags detected by different monitoring points were caused by the same event. The range of influence of the voltage sag event can be obtained simultaneously.

**Author Contributions:** Conceptualization, J.J. and C.H.; methodology, C.H.; software, J.J.; writing—original draft preparation, J.J. writing—review and editing, C.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank Nanjing Shining Electric Automation Co., Ltd. for providing the field measurement waveforms. Nanjing Shining Electric Automation supported the development of the proposed method for power quality monitoring systems.

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

## References

1. Vegunta, S.C.; Milanovic, J.V. Estimation of cost of downtime of industrial process due to voltage sags. *IEEE Trans. Power Deliv.* **2011**, *26*, 576–587. [[CrossRef](#)]
2. Arias-Guzman, S. Analysis of voltage sag severity case study in an industrial circuit. *IEEE Trans. Ind. Appl.* **2017**, *53*, 15–21. [[CrossRef](#)]
3. Mei, F.; Ren, Y.; Wu, Q.; Zhang, C.; Pan, Y.; Sha, H.; Zheng, J. Online recognition method for voltage sags based on a deep belief network. *Energies* **2018**, *12*, 43. [[CrossRef](#)]
4. Xiao, X.Y.; Chen, Y.Z.; Wang, Y.; Ma, Y.Q. Multi-attribute analysis on voltage sag insurance mechanisms and their feasibility for sensitive customers. *Int. Gener. Transm. Dis.* **2018**, *12*, 3892–3899. [[CrossRef](#)]
5. Polajzer, B.; Tumberger, G.S.; Seme, S.; Dolinar, D. Detection of voltage sag sources based on instantaneous voltage and current vectors and orthogonal Clarke's transformation. *IET Gener. Transm. Distrib.* **2008**, *2*, 219–226. [[CrossRef](#)]
6. Kezunovic, M.; Liao, Y. A new method for classification and characterization of voltage sags. *Electr. Pow. Syst. Res.* **2001**, *58*, 27–35. [[CrossRef](#)]
7. Sadigh, A.K.; Smedley, K.M. Fast and precise voltage sag detection method for dynamic voltage restorer (DVR) application. *Electr. Pow. Syst. Res.* **2016**, *130*, 192–207. [[CrossRef](#)]
8. Xi, Y.; Li, Z.; Zeng, X.; Tang, X.; Liu, Q.; Xiao, H. Detection of power quality disturbances using an adaptive process noise covariance Kalman filter. *Digit. Signal. Process.* **2018**, *76*, 34–49. [[CrossRef](#)]
9. Saini, M.K.; Beniwal, R.K. Detection and classification of power quality disturbances in wind-grid integrated system using fast time-time transform and small residual-extreme learning machine. *Int. Trans. Electr. Energy Syst.* **2018**, *28*, e2519. [[CrossRef](#)]
10. Jeevitha, S.R.S.; Mabel, M.C. Novel optimization parameters of power quality disturbances using novel bio-inspired algorithms: A comparative approach. *Biomed. Signal Process. Control.* **2018**, *42*, 253–266. [[CrossRef](#)]
11. Branco, H.M.; Oleskovicz, M.; Coury, D.V.; Delbem, A.C. Multiobjective optimization for power quality monitoring allocation considering voltage sags in distribution systems. *Int. J. Electr. Power Energy Syst.* **2018**, *97*, 1–10. [[CrossRef](#)]
12. Bagheri, A.; Gu, I.; Bollen, M.; Balouji, E. A Robust Transform-Domain Deep Convolutional Network for Voltage Dip Classification. *IEEE Trans. Power Deliv.* **2018**, *33*, 2794–2802. [[CrossRef](#)]
13. Kapoor, R.; Gupta, R.; Jha, S.; Kumar, R. Boosting performance of power quality event identification with KL Divergence measure and standard deviation. *Measurement* **2018**, *126*, 134–142. [[CrossRef](#)]
14. Gururajapathy, S.S.; Mokhlis, H.; Illias, H.A.; Awal, L.J. Support vector classification and regression for fault location in distribution system using voltage sag profile. *IEEE J. Trans. Electr. Electron. Eng.* **2017**, *12*, 519–526. [[CrossRef](#)]
15. Garcia-Sánchez, T.; Gómez-Lázaro, E.; Muljadi, E.; Kessler, M.; Muñoz-Benavente, I.; Molina-García, A. Identification of linearised RMS-voltage dip patterns based on clustering in renewable plants. *IET Gener. Transm. Distrib.* **2018**, *12*, 1256–1262.
16. Daud, K.; Abidin, A.F.; Ismail, A.P. Voltage Sags and Transient Detection and Classification Using Half/One-Cycle Windowing Techniques Based on Continuous S-Transform with Neural Network. In Proceedings of the 2nd International Conference on Applied Physics and Engineering (ICAPE), Penang, Malaysia, 2–3 November 2016.
17. Meena, P.; Rao, K.U.; Ravishankar, D. A modified simple algorithm for detection of voltage sags and swells in practical loads. *IEEE Int. Conf. Power Syst.* **2009**, *12*, 1–6.
18. Latran, M.B.; Teke, A. A novel wavelet transform based voltage sag/swell detection algorithm. *Int. J. Electr. Power Energy Syst.* **2015**, *71*, 131–139. [[CrossRef](#)]
19. Styvaktakis, E.; Bollen, M.H.; Gu, I.Y. Expert system for classification and analysis of power system events. *IEEE Trans. Power Deliv.* **2002**, *17*, 423–428. [[CrossRef](#)]
20. Chu, J.W.; Yuan, X.D.; Chen, B.; Wang, X.C.; Qiu, H.F.; Gu, W. A Method for Distribution Network Voltage Sag Source Identification Combining Wavelet Analysis and Modified DTW Distance. *Power Syst. Technol.* **2018**, *42*, 637–643.
21. Nunez, V.B.; Velandia, R.; Hernandez, F.; Melendez, J.; Vargas, H. Relevant Attributes for Voltage Event Diagnosis in Power Distribution Networks. *Rev. Iberoam. Autom. Inform. Ind.* **2013**, *10*, 73–84.
22. Tang, Y.; Wei, R.; Chen, K.; Fang, Y. Voltage sag source identification based on the sign of internal resistance in a "Thevenin's equivalent circuit". *Int. Trans. Electr. Energy Syst.* **2017**, *27*, e2461. [[CrossRef](#)]
23. Saini, M.K.; Aggarwal, A. Fractionally delayed Legendre wavelet transform based detection and optimal features based classification of voltage sag causes. *J. Renew. Sustain. Energy* **2019**, *11*, 25–36. [[CrossRef](#)]
24. Zhuang, D.; Liu, Y.; Liu, S.; Ma, T.; Ong, S.H. A shape-based cutting and clustering algorithm for multiple change-point detection. *J. Comput. Appl. Math.* **2020**, *6*, 112623. [[CrossRef](#)]
25. Rabiner, L.R. Considerations in dynamic time warping algorithms for discrete word recognition. *J. Acoust. Soc. Am.* **1978**, *63*, 575–582. [[CrossRef](#)]

26. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust.* **1978**, *26*, 43–49. [[CrossRef](#)]
27. Zhao, J.P.; Laurent, I. Shapedtw: shape dynamic time warping. *Pattern Recogn.* **2018**, *74*, 171–184. [[CrossRef](#)]
28. Chung, F.R.K. Spectral Graph Theory, Regional Conference Series in Mathematics. *AMS* **1997**, *92*, 142–162.
29. Luxburg, U. A tutorial on spectral clustering. *Stat Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
30. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
31. Ng, A.; Jordan, M.; Weiss, Y. On spectral clustering: analysis and an algorithm. In *Advances in Neural Information Processing Systems (NIPS)*; MIT Press: Cambridge, MA, USA, 2002; Volume 17.