


## Article

# A Novel Fault Location Method for Power Cables Based on an Unsupervised Learning Algorithm <sup>†</sup>

Mingzhen Li <sup>1,\*</sup>, Jialong Bu <sup>1</sup>, Yupeng Song <sup>1</sup>, Zhongyi Pu <sup>1</sup>, Yuli Wang <sup>2</sup> and Cheng Xie <sup>3</sup>

<sup>1</sup> School of Electrical Engineering, Nantong University, No. 9, Seyuan Road, Nantong 226019, China; 1712021069@stmail.edu.ntu.cn (J.B.); 1708031027@stmail.ntu.edu.cn (Y.S.); 1712021010@stmail.edu.ntu.cn (Z.P.)

<sup>2</sup> China Electric Power Research Institute, Wuhan 430074, China; wangyuli@epri.sgcc.com.cn

<sup>3</sup> Zhejiang Electric Power Research Institute, Hangzhou 310014, China; chengxie@zj.sgcc.com.cn

\* Correspondence: mzli@ntu.edu.cn

<sup>†</sup> This work was extensions of our paper published in Xie, C.; Shao, X.; Jin, Y.; Wang, Y.; Li, M. A Novel Fault Location Method for Power Cables Based on Unsupervised Learning. In Proceedings of 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Beijing, China, 6–10 September 2020.

**Abstract:** In order to locate the short-circuit fault in power cable systems accurately and in a timely manner, a novel fault location method based on traveling waves is proposed, which has been improved by unsupervised learning algorithms. There are three main steps of the method: (1) build a matrix of the traveling waves associated with the sheath currents of the cables; (2) cluster the data in the matrix according to its density level and the stability, using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN); (3) search for the characteristic cluster point(s) of the two branch clusters with the smallest density level to identify the arrival time of the traveling wave. The main improvement is that high-dimensional data can be directly used for the clustering, making the method more effective and accurate. A Power System Computer Aided Design (PSCAD) simulation has been carried out for typical power cable circuits. The results indicate that the hierarchical structure of the condensed cluster tree corresponds exactly to the location relationship between the fault point and the monitoring point. The proposed method can be used for the identification of the arrival time of the traveling wave.

**Keywords:** power cable; fault location; sheath current; traveling wave; unsupervised learning



**Citation:** Li, M.; Bu, J.; Song, Y.; Pu, Z.; Wang, Y.; Xie, C. A Novel Fault Location Method for Power Cables Based on an Unsupervised Learning Algorithm. *Energies* **2021**, *14*, 1164. <https://doi.org/10.3390/en14041164>

Academic Editor: Issouf Fofana

Received: 25 January 2021

Accepted: 17 February 2021

Published: 22 February 2021

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



**Copyright:** © 2021 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

Due to the limitations of urban land resources and the requirements for power reliability, power cables are widely used in urban power transmission and distribution systems [1–3]. With the rapid increase in the scale of power cable applications, the number of short-circuit faults has also increased [4,5]. Although the overall short-circuit fault rate of the power cables is low, the origin of the short-circuit faults has many possible factors, which are difficult to predict, and there will be wide social impacts and large economic losses caused by the faults. In order to shorten the repair time of the faulty power cables, efforts are needed to develop reliable and accurate fault location methods.

According to the fault location principles, currently, online fault location methods used in power systems are mainly based on the impedance and the traveling wave. Both of the impedance-based and the traveling-wave-based methods originated from the fault location for overhead lines. The impedance-based method [6–8] mainly uses the relay protection device to collect fault voltage/current data, and locates the fault based on the parameter identification of the system impedance. Since the impedance monitored by this method may have a nonlinear relationship with the distance between the fault point, it is difficult to directly locate the fault accurately through the relationship between the impedance and

the distance between the fault point. The traveling-wave-based methods [9–11] locate the fault through the identification of the propagation time of the fault traveling wave. When the line distribution parameters are large or non-uniform, it is difficult to accurately extract the head and/or the arrival time of the traveling wave.

Recently, efforts have been made to improve the accuracy and applicability of impedance-based methods. In [12], mutual inductance effects have been taken into account to make the method suitable for the medium voltage distribution network in the double-circuit lines. In [13], the fault location criteria for different fault sections of the distributed network have been proposed, taking into account the electrical characteristics of nodes in the distributed network under fault and non-fault conditions. In [14], a fault line selection method of distributed network has been proposed based on the direction vector characteristics of the fault currents. In [15], a Fibonacci search algorithm has been used to analyze a specific network structure under the premise of low fault resistance, the results indicate that this method improves the robustness of the impedance-based method with respect to the fault resistance. In [16], the influence of the load uncertainty on the accuracy of fault location has been studied, and a method to reduce the influence has been proposed. The research above can be summarized in Table 1.

**Table 1.** The summary of the improvements of the impedance-based method.

References	[12]	[13]	[14]	[15]	[16]
Line model	$\pi$ -type	Distributed	$\pi$ -type	Distributed	Distributed
Double or multiple loops	$\checkmark^1$	$\checkmark$	$\checkmark$	- <sup>2</sup>	-
Fault section location	-	$\checkmark$	-	$\checkmark$	-
Non-uniform transmission line	$\checkmark$	-	$\checkmark$	-	-
Branches	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Robustness to fault resistance	-	-	-	$\checkmark$	$\checkmark$
Distributed Power	-	-	$\checkmark$	-	$\checkmark$
Load uncertainty	-	-	$\checkmark$	-	$\checkmark$

(<sup>1</sup>  $\checkmark$  means applicable; <sup>2</sup> - means not applicable).

As shown in Table 1, the improvements of the impedance-based method mainly focus on the applicability of different application scenarios. With the different focuses, the fault location criterion cannot be uniform or mutually applicable.

Additionally, efforts have been made to improve the accuracy and the applicability of the traveling-wave-based method in recent years. In [17], the regular pattern of the reflection of voltage and current traveling waves at the end of the line has been analyzed, then a fault searching algorithm has been proposed based on mid-point time, which can be used for the fault location of mixed line with discontinuous wave velocity. In [18], time-frequency analysis of fault traveling waves has been carried out using continuous wavelet transform method, so that the arrival time of the traveling wave could be identified according to the characteristic frequency of the fault traveling wave. In [19], the identification accuracy of the high frequency part of the fault traveling wave has been improved using a multilayer neural network. In [20], the problem of multi-channel signal synchronization has been solved by using the time difference between the first arrival wave and the first reflected wave. In [21], a normalized fault location criterion has been proposed, which does not require external time reference and accurate traveling wave velocity. However, the time difference may be aliased in multiple discharges of the fault, which is common in cable breakdown discharges. The above studies have improved the traveling-wave-based method in terms of traveling wave propagation, time-frequency analysis, and feature recognition, which can be summarized in Table 2.

**Table 2.** The summary of the improvements of the traveling-wave-based method.

References	[17]	[18]	[19]	[20]	[21]
Frequency range	10~100 kHz	2.5 kHz~1 GHz	- <sup>2</sup>	Within 100 kHz	Within 1 MHz
Mixed lines	√ <sup>1</sup>	√	√	-	-
Non-uniform transmission line	√	-	√	√	-
Branches	√	√	√	-	-
No signal synchronization required	-	-	√	√	√
No traveling wave speed required	-	-	-	√	√
No line parameters required	-	-	-	-	√
Clock skew	-	-	-	-	√

(<sup>1</sup> √ means applicable; <sup>2</sup> - means not applicable).

It can be seen from the above-mentioned literature analysis that, as time goes by, there are continuous researches to improve the traveling-wave-based method, which also makes the method supplemented and improved. With the improvement of computing performance and signal synchronization accuracy, many applications of intelligent algorithms have emerged in recent years, but these improvements and applications are limited to a single type of line or specific application scenarios, and it is difficult to form a unified fault location criterion. The improvements have not fundamentally solved the inherent problems of traveling wave head attenuation and noises.

The authors of this work previously proposed an unsupervised learning method for fault location of power cables and to improve on the traditionally applied traveling-wave-based method [22,23]. The present paper further expands on their previously published work, and makes modifications of the fault location process. There are four steps for the fault location in the previous contribution [23], i.e., (1) build a matrix of the traveling waves associated with the sheath currents of the cables; (2) apply t-distributed Stochastic Neighbor Embedding (t-SNE) to reconstruct the matrix into a low dimension; (3) cluster the data in the matrix according to its closeness, using Density-Based Spatial Clustering of Applications with Noise (DBSCAN); (4) search for the maximum slope point of the non-noise cluster with the fewest samples to identify the arrival time of the traveling wave. The improvement mainly lies in the high-dimensional clustering for the identification of the arrival time of the traveling waves of multiple signals. Thus, Steps 2 and 3 can be merged as one, the information of the original signals can be reserved without the dimensionality reduction, and the waveforms of the traveling waves can be identified directly via an unsupervised learning algorithm. In addition, the presented method can output a condensed cluster tree which is believed to correspond with the traveling waves. The specific relation between the condensed cluster tree and the traveling wave will be studied in further research, which could help to make the process of fault occurrence and traveling wave propagation clearer.

## 2. Analysis of the Traveling Wave Signals and the Fault Location

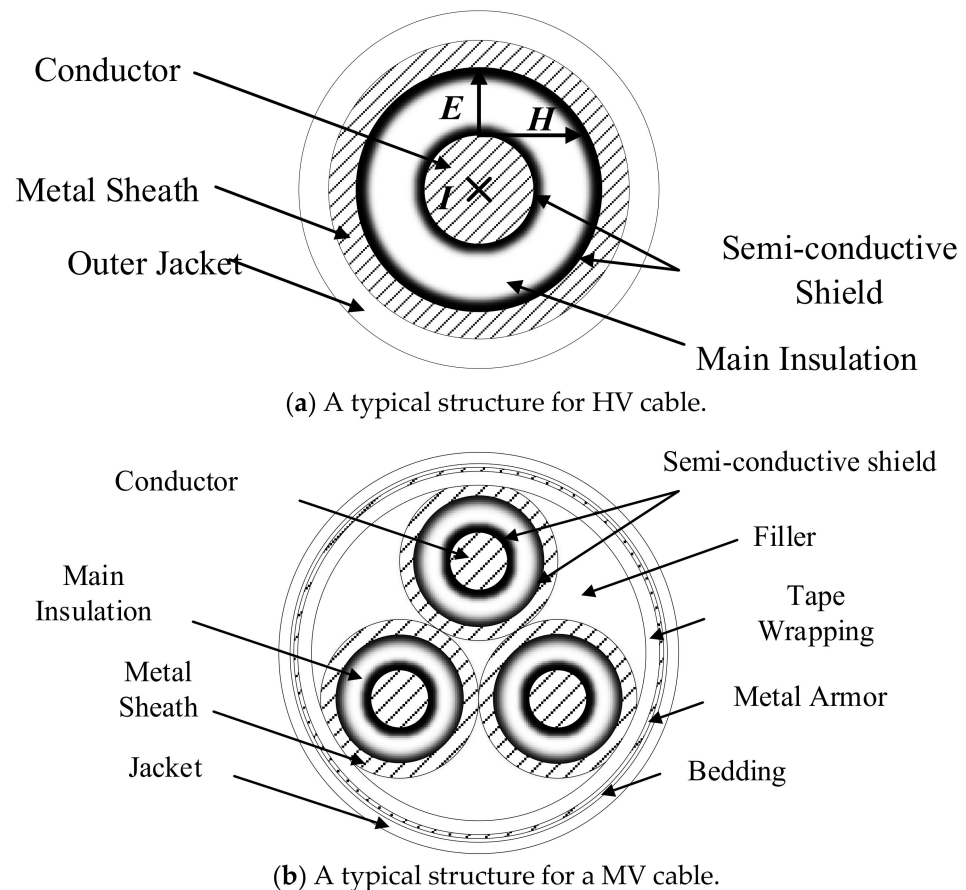
### 2.1. Typical Cable Structures and the Monitoring of The Traveling Wave Signals

#### 2.1.1. Typical Cable Structures

A typical power cable for high voltage (HV: 110 kV or more) has a single-core structure, and a typical power cable for medium voltage (MV: 10 to 35 kV) has a three-core structure, as shown in Figure 1a,b, respectively.

Since the metal sheath of the HV cable has been reliably grounded at least one end, the sheath voltage is close to the ground level during normal operation. In addition, when the load current flows through the core conductor, the direction of the electric field on the cable cross section is radial, shown as  $E$  in Figure 1a, and the direction of the magnetic field is shown as  $H$ . Therefore, the direction of the Poynting vector of the energy transmission is the  $E \times H$  direction, which indicates that the energy propagates along the line axis between the main insulation between the two layers of metal. When a short-circuit fault occurs at any position of the cable, there will be fault signals on the core conductor and the

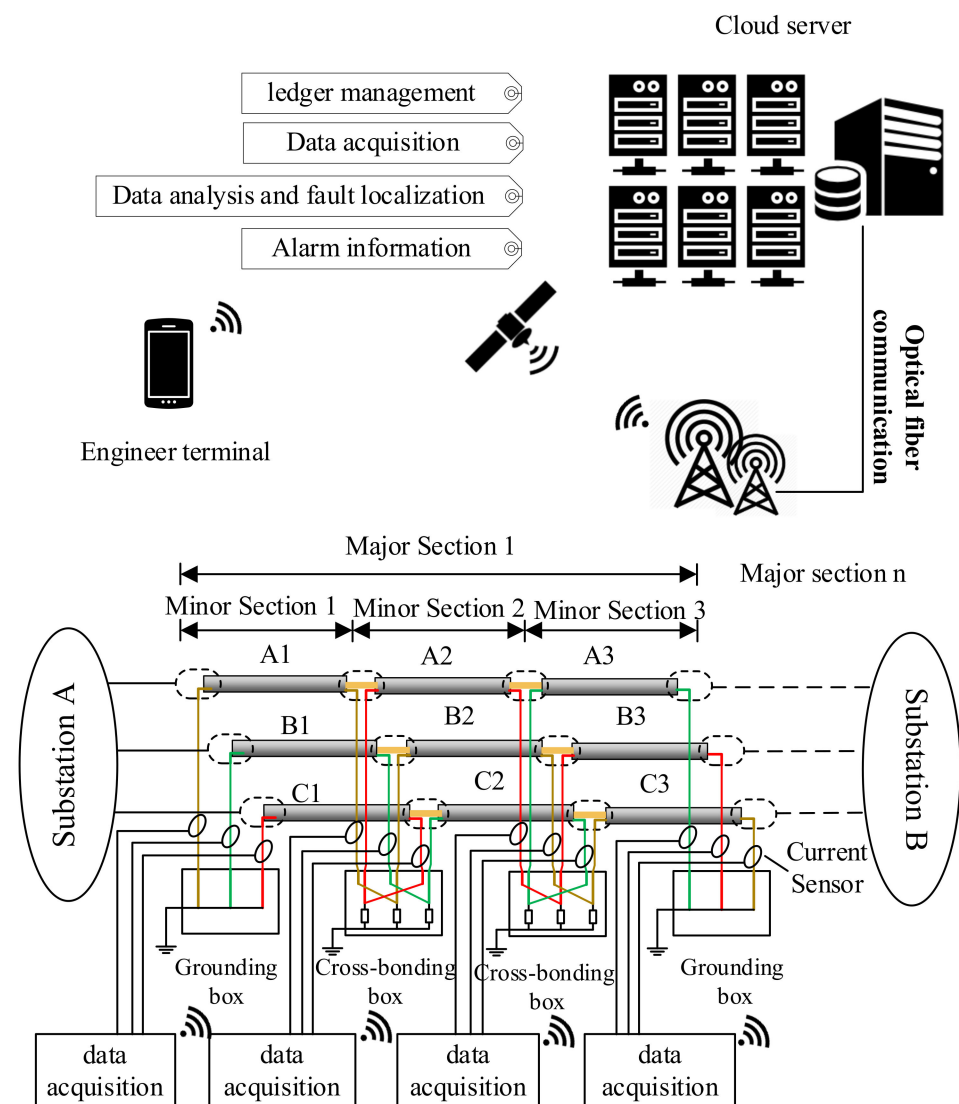
metal sheath. Therefore, regardless of whether the core conductor or the metal sheath is monitored, the arrival time of the traveling wave is the same. For safety and convenience, the metal sheath has been chosen to monitor for the traveling wave signals in the paper. For the MV cable shown in Figure 1b, both the metal sheath and the metal armor are grounded at the same grounding position. Therefore, the method of obtaining traveling waves by monitoring the sheath current is also suitable for MV cables.



**Figure 1.** Cross-sectional structure diagram of typical power cables.

### 2.1.2. The Monitoring of the Traveling Wave Signals

Figure 2 shows the sheath current monitoring system, where a typical HV cable circuit is taken as an example. The system consists of four parts, namely, the data acquisition module, the communication module, the location analysis software installed in the cloud server and the interface for cable maintenance engineers. Among them, the data acquisition module includes sheath current sensors, and it is installed in the grounding boxes and the cross-bonding boxes. When a short-circuit fault occurs in the cable, the sheath current in the loop will rise to the level of the fault current. At the same time, the data acquisition module will be awakened and will start to upload the data. The communication module of the data acquisition system can transmit the recorded data to a designated cloud server, where the data analysis software will perform data analysis, after that, the location results will be sent to the maintenance engineer. The recorded data can also be downloaded from the server for further analysis.



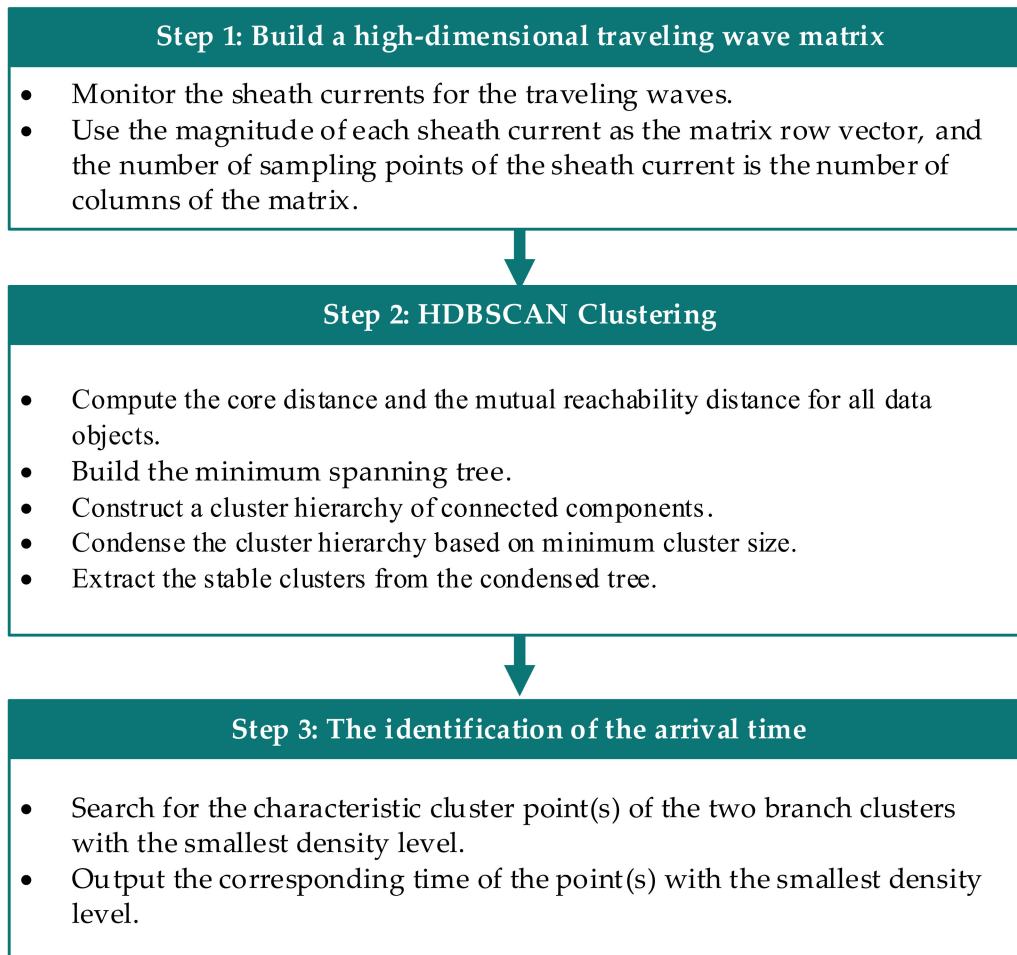
**Figure 2.** The designed sheath currents monitoring system (The yellow line represents phase A, the green line represents phase B, the red line represents phase C. A1–C3 are the labels of different cable sections).

## 2.2. The Fault Location

An autonomous learning mechanism based on the two-terminal traveling wave method is proposed. Thus, the arrival time of the fault traveling wave can be identified accurately with multiple monitored data. Since the error of the two-terminal traveling wave method mainly comes from the identification of the arrival times at the two terminals of the cable, moreover, the main source of identification errors is randomness, such as synchronization time delay, noise interference, etc., and more monitoring points can help to reduce random errors, thereby improving accuracy. In order to reduce random errors, the sheath currents of the three-phase cables and adjacent cables in the same cable passage are designed to be monitored. Thus, a multi-dimensional data matrix can be built. Through the analysis of the multi-dimensional matrix, the arrival times can be identified more accurately. The process of the identification of the arrival times at the two terminals of the cable is shown in Figure 3.

There are three main steps of the proposed method, as presented in Figure 3. Firstly, a high-dimensional data matrix needs to be built using the monitored sheath currents. The amplitude of each sheath current is taken as the row vector of the matrix, and the number of sampling points of the sheath current is the number of columns of the matrix.

Secondly, the high-dimensional matrix can be directly clustered using HDBSCAN, which is a density-based hierarchical clustering algorithm. The specific process of HDBSCAN is relatively complicated, which will be demonstrated in Sections 2.3 and 2.4. Finally, the arrival time of the traveling wave can be identified by searching for the characteristic cluster point(s) of the two branch clusters with the smallest density level.



**Figure 3.** The process of the identification of the arrival time at the two terminals of the cable.

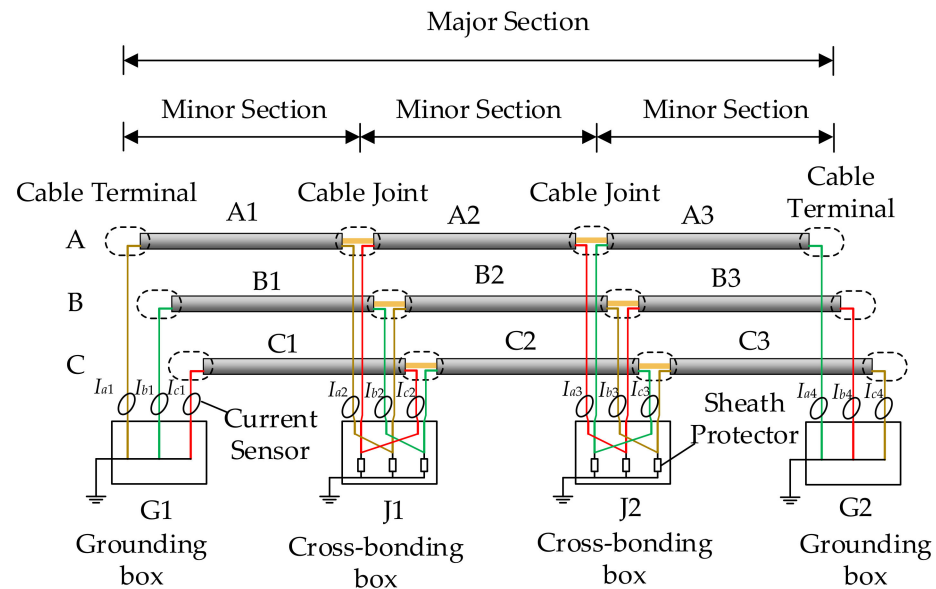
It is to be noticed that, in [23], unsupervised learning (t-SNE + DBSCAN) has been used to improve the identification of traveling wave arrival time. The results show that even if there is noise and/or random time errors in the recorded data, the arrival time of the traveling wave can be identified. However, when the matrix dimension is greater than 3, the dimensionality reduction process will sacrifice the integrity of the data information. After the dimensionality reduction, the data will lose the original information to a certain extent, for example, the physical meaning. The paper omits the step of the dimensionality reduction and uses the recorded data directly for identification. This allows the original data information to be retained to the greatest extent, and in principle makes the error as small as possible.

### 2.3. The Construction of the Traveling Wave Matrix

As shown in Figure 4, the sheath currents can be monitored at the grounding boxes and the cross-linking boxes in actual HV cross-bonded cable systems. Usually, there are multiple cables sharing the same passage/duct. If a short-circuit fault occurs in a cable, an inductive coupling signal can be detected on the sheath of the adjacent cable. Thus, multiple sheath currents can be analyzed together, which is believed to make the



identification accuracy of the arrival time to be improved. Hence, a high-dimensional data matrix can be built using the amplitude of each monitored sheath current as the row vector, and the number of sampling points of the sheath current is the number of columns of the matrix.



**Figure 4.** The schematic diagram of a typical high voltage (HV) cross-bonded cable circuit (The yellow line represents phase A, the green line represents phase B, the red line represents phase C. A1–C3 are the labels of different cable sections).

A typical HV cross-bonded cable circuit model, as shown in Figure 4, has been established using PSCAD. The cross-sectional structure parameters of the cable are shown in Table 3. As shown in Figure 4, 12 current sensors have been used in a major section to monitor the sheath current. The sheath currents measured at the 12 sensors are, respectively,  $I_{a1}, I_{b1}, I_{c1}$ , the sheath currents recorded by the current sensors of phase A, B, C at G1;  $I_{a2}, I_{b2}, I_{c2}$ , the sheath currents recorded in phase A, B, C at J1;  $I_{a3}, I_{b3}, I_{c3}$  the sheath currents recorded in phase A, B, C at J2;  $I_{a4}, I_{b4}, I_{c4}$ , the sheath currents recorded in phase A, B, C at G2. A short-circuit fault has been set in cable section A1, 300 m away from G1. The 12 sheath currents generated in the simulation are presented in Figure 5. In the case, the sheath current matrix  $I$  can be presented in (1).

$$I = [ I_{a1} \quad I_{b1} \quad I_{c1} \quad \dots \quad I_{c3} ]^T \quad (1)$$

$$v_0 = \frac{1}{\sqrt{\mu\epsilon}} \quad (2)$$

**Table 3.** The cross-sectional structure parameters of the HV cable.

	Structure	Outer Radius/mm
1	Core conductor (Copper)	17.0
2	Inner semi-conductor (Nylon belt)	18.4
3	Main insulation (Ultra-clean XLPE)	34.4
4	Outer semi-conductor (Super-smooth semi-conductive shielding material)	35.4
5	Water-blocking layer (Semi-conductor)	39.4
6	Metal sheath (aluminum)	43.9
7	Jacket (PVC)	48.6

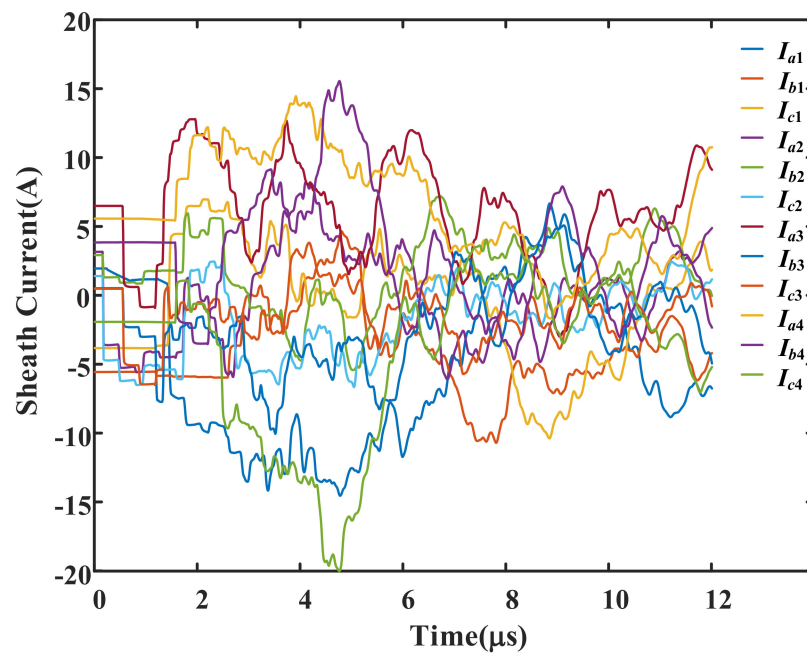


Figure 5. The simulated 12 sheath currents of the HV cable.

Since the velocity of the traveling wave depends on the material parameters of the cable, it can be calculated using (2), where  $\mu$  is the magnetic permeability and  $\mu\epsilon$  is the insulation permittivity. In the case,  $\mu = 4\pi \times 10^{-7}$  and  $\epsilon = 3.63 \times 10^{-11}$ , so  $v_0 = 1.48 \times 10^8$  m/s. Hence, the arrival time of the traveling wave is expected to be  $2.03 \mu\text{s}$  at G1 and  $1.35 \mu\text{s}$  at J1.

#### 2.4. The Unsupervised Learning Algorithm

DBSCAN [24] is a typical density-based clustering algorithm, which is able to identify arbitrarily shaped clusters and noise (outliers) in data. However, the main problem of such algorithms in application is the ability to process high-dimensional data. HDBSCAN [25] follows Hartigan's classic density contour clustering/tree model, and improves the existing density-based clustering algorithms for different aspects [26]. Moreover, it provides paths of minimum spanning tree for high-dimensional data [27].

Instead of using parameters  $Eps$  and  $MinPts$  ( $Eps$  is the threshold of the neighborhood distance of a certain sample;  $MinPts$  is the threshold of the number of samples near the sample whose distance is  $Eps$ .) to describe the closeness of the sample distribution in the neighborhood as in DBSCAN, HDBSCAN uses parameters  $MinPts$  and  $MinClustSize$  ( $MinClustSize$  indicates the minimize cluster size) to limit the possible clusters, and makes the high-dimensional clustering problem an optimization problem.

##### 2.4.1. The Measurement of High-Dimensional Data

The first problem introduced by high-dimensional data is the measurement method. It has been proven that the Euclidean distances of all data objects in the high-dimensional space tend to be equal [28]. Therefore, the high-dimensional space needs to be measured in a more suitable way.

Let  $X = \{x_1, \dots, x_n\}$  be a dataset containing  $n$  data objects, each of which is described by an attribute vector  $x$ . The Euclidean distance between  $x_p$  and  $x_q$  denotes as  $d(x_p, x_q)$ . An object  $x_p$  is called a *core object* if its  $\epsilon$ -neighborhood ( $\epsilon$  means the value of the minimum radius) contains at least  $MinPts$  objects. The *core distance* of an object  $x_p$ ,  $d_{core}(x_p)$ , is the Euclidean distance from  $x_p$  to its  $MinPts$ -nearest neighbor (including  $x_p$ ). The *mutual reachability distance* between two objects  $x_p$  and  $x_q$  is defined as (3).

$$d_{mreach}(x_p, x_q) = \max\{d_{core}(x_p), d_{core}(x_q), d(x_p, x_q)\} \quad (3)$$



Under the measurement of *mutual reachability distance*, dense objects, whose core distance are small, remain the same distance from each other, but sparser objects have been dispersed to at least the core distance from any other object. It is to be noticed that the effectiveness depends on the value of *MinPts*. The larger *MinPts* values the more data objects are affected by the measurement. While the value of *MinPts* cannot be too large, or the number of the final clusters would be very small.

#### 2.4.2. The Hierarchical Clusters of High-Dimensional Data

With the measurement of *mutual reachability distance*, the density discrimination of the data objects is more significant. As the data objects are high-dimensional, the density cannot be directly expressed on a plane, in other words, it is not “flat”. Thus, the data objects can be regarded as a weighted graph, where the data object is a vertex, and the weight of the edge between any two data objects is equal to the *mutual reachability distance*.

Now the clustering process can be regarded as the process of merging vertices and deleting edges. Similar to DBSCAN, a threshold parameter *Eps* is used in the process. This time, *Eps* is not a constant value. The threshold starts from a relatively large value, and is steadily being lowered. Simultaneously, the edges whose weights are above the threshold value are deleted. Eventually, the connected graph is turned into a completely disconnected graph which is composed by a hierarchy of connected components. This process can be achieved by building the minimum spanning tree [29].

#### 2.4.3. The Construction of the Clustering Hierarchy

Since the minimum spanning tree of the high-dimensional data objects is still high dimensionally, it needs to be converted into the hierarchy of connected components to make the clustering hierarchy “flat”. This process is completed by the following steps:

- (1) Sort the edges of the minimum spanning tree by the weights (in ascending order);
- (2) Traverse all the branches;
- (3) Create a new merged cluster for each edge.

It is to be noticed that Step 3 can be completed via a union-find data structure [30]. Thus, each data object has a clear corresponding start and end in the hierarchy (with the ordinate of *mutual reachability distance*), which can be used for arbitrary clustering through cutting branches with a single horizontal line. The next step is to find the reasonable horizontal line(s) for the clustering, not just by setting a fixed parameter of the *mutual reachability distance*. Essentially, the clustering process here of DBSCAN is achieved by setting such a fixed parameter, which is very unintuitive. Therefore, a better solution is needed to select the clusters.

#### 2.4.4. Condense the Cluster Hierarchy

The first step in the cluster extraction is to condense the huge and complex cluster hierarchy into a smaller tree, and attach a little more data to each node. Rather than setting a fixed parameter of the *mutual reachability distance*, HDBSCAN uses *MinClustSize* (minimum cluster size) limit the possible clusters. Thus, when the hierarchy are traversed, each split is judged that whether the number of points in one of the new clusters created by each split is less than *MinClustSize*. If the number of the data objects is less than *MinClustSize*, the data object falls outside the cluster, and the larger cluster retains the cluster identity of the parent node. Then, the data objects with its *mutual reachability distance* which fall outside the cluster are marked. On the other hand, if the split is into two clusters each at least as large as *MinClustSize*, the split persists in the tree. Thus, the clusters are easier to view and process, especially for clustering problems with simple data structures. However, the clusters are still needed to be selected as ‘flat’ clusters.

#### 2.4.5. Extract the Clusters

In fact, the ‘flat’ clusters have the highest stability. To illustrate the notion of the stability, several important mathematical measurement parameters are introduced. First,

a different measurement is needed to denote the persistence of clusters, as shown in (4), representing the density level.

$$\lambda = \frac{1}{\text{distance}} \quad (4)$$

For a cluster  $C_i$ , the *excess of mass* of  $C_i$  can be defined by (5), where  $f(\mathbf{x})$  is a density function, representing the density of each cluster.

$$E(C_i) = \int_{\mathbf{x} \in C_i} (f(\mathbf{x}) - \lambda_{\min}(C_i)) d\mathbf{x} \quad (5)$$

The *relative excess of mass* of a cluster  $C_i$ , which appears at level  $\lambda_{\min}(C_i)$  can be defined by (6), where  $\lambda_{\max}(\mathbf{x}, C_i) = \min\{f(\mathbf{x}), \lambda_{\max}(C_i)\}$  and  $\lambda_{\max}(C_i)$  is the density level at which  $C_i$  is split or disappears.

$$E_R(C_i) = \int_{\mathbf{x} \in C_i} (\lambda_{\max}(\mathbf{x}, C_i) - \lambda_{\min}(C_i)) d\mathbf{x} \quad (6)$$

The *stability* of a cluster  $C_i$  can be defined by (7), where  $\lambda_{\min}(C_i)$  is the minimum density level at which  $C_i$  exists,  $\lambda_{\max}(\mathbf{x}_j, C_i)$  is the density level beyond which object  $\mathbf{x}_j$  no longer belongs to cluster  $C_i$ .

$$S(C_i) = \sum_{\mathbf{x}_j \in C_i} (\lambda_{\max}(\mathbf{x}_j, C_i) - \lambda_{\min}(C_i)) \quad (7)$$

Therefore, the step can be regarded as an optimization problem, and the objective is to maximize the overall stability of the extracted clusters. After optimization, the most prominent non-overlapping cluster will be the output.

### 3. Results

#### 3.1. The Arrival Time Identification

There are big differences between the 12 sheath current amplitudes recorded at different locations, as shown in Figure 5, which also constitute the elements of the high-dimensional matrix  $I$ , the input. The next step is the clustering. In order to make the distinction of the clustering more remarkable, a relatively large *MinClustSize* and a relatively small *MinPts* are required. In this case, *MinClustSize* and *MinPts* are set to 12 and 7, respectively. By applying HDBSCAN to the matrix  $I$  (the original data to be input into the algorithm), the condensed cluster tree can be obtained, as plotted in Figure 6.

The red nodes plotted in Figure 6 represent the final stable clusters, and the black nodes represent the clusters which are not extracted after stability optimization. The serial numbers of the red nodes are same as those of the clusters of the tree, and they can be renumbered as 1–9 for convenience. In Figure 6, the abscissa has no physical meaning; the ordinate represents the relative magnitude of the  $\lambda$  value of each cluster, which increases from top to bottom.

The cluster scatter plots of sheath currents  $I_{a2}$ ,  $I_{b2}$  and  $I_{c2}$  are presented in Figure 7, where Cluster 1–9 represent the final stable clusters, and Cluster 0 represents “noise”. It should be noted that Cluster 0 does not have to be the real noise in practice. As mentioned above, whether the object  $\mathbf{x}_p$  is labeled as “noise” only depends on the hierarchical structure of the data, which is related to the density level and stability of the high-dimensional matrix. In the case, the arrival time point of the traveling wave that needs to be extracted has similar characteristics to the “noise”. Therefore, Cluster 0 should be analyzed together with the ordinary clusters. For further analysis, Figure 8 shows the number of samples in each of the 10 clusters.

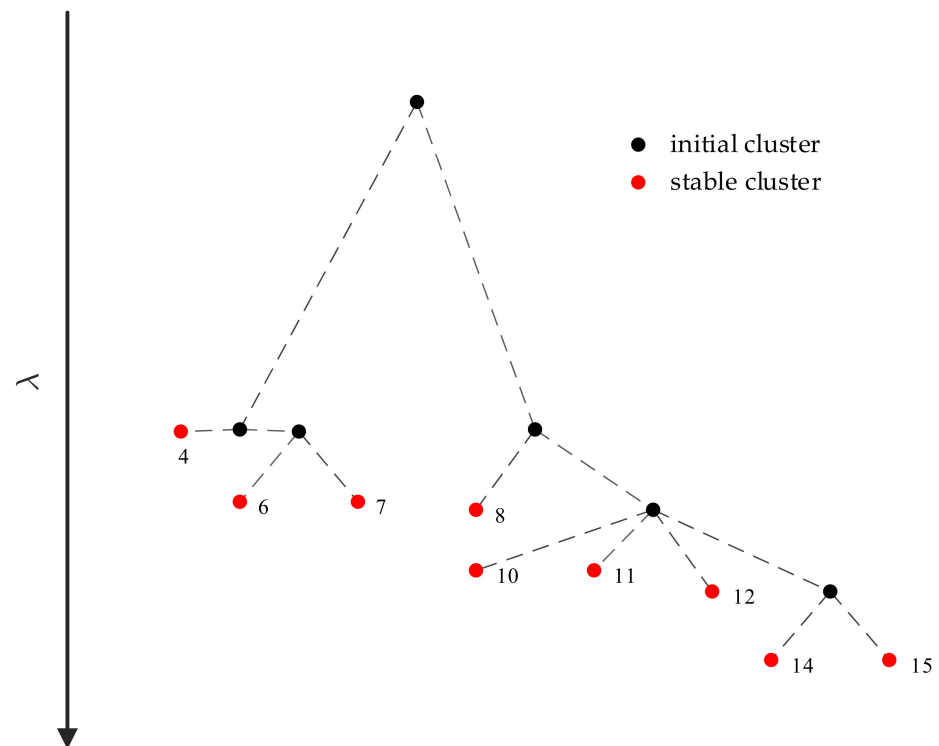


Figure 6. The schematic diagram of the condensed cluster tree for the matrix  $I$ .

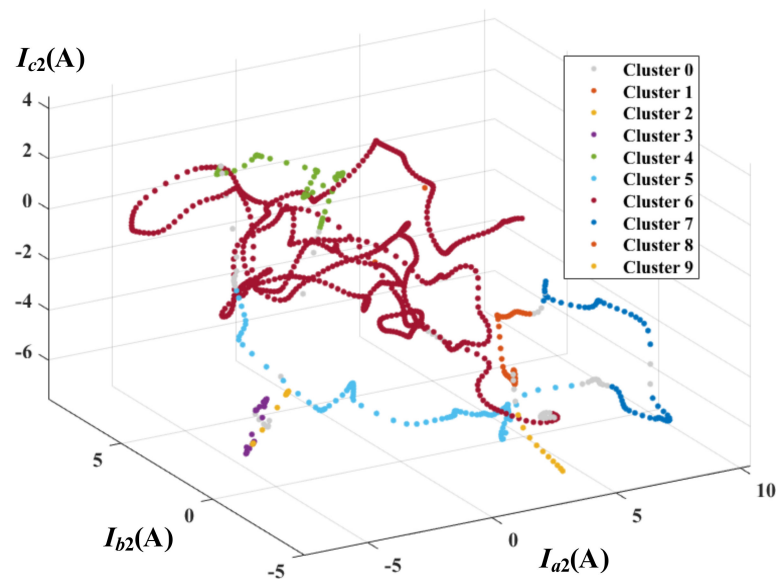
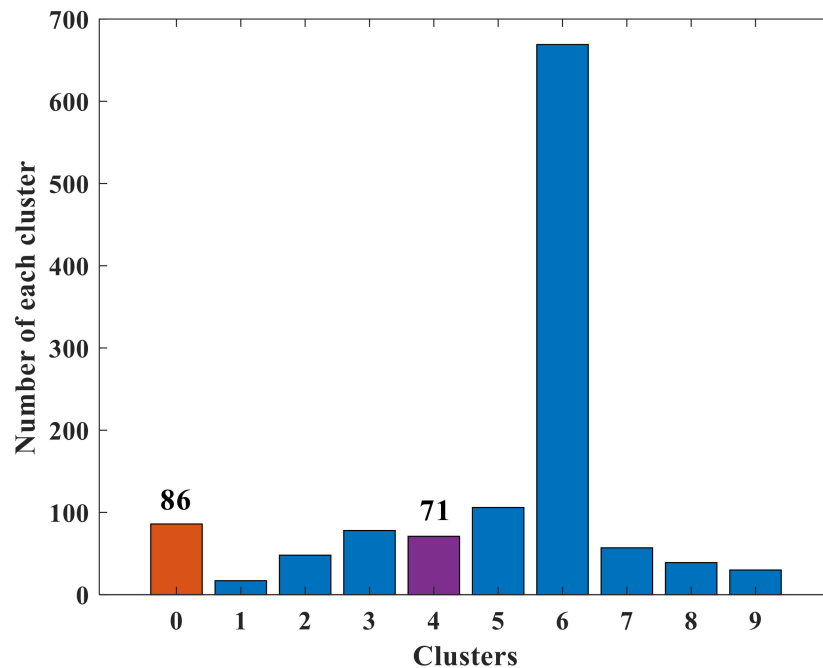


Figure 7. The cluster scatter scheme of sheath currents  $I_{a2}$ ,  $I_{b2}$  and  $I_{c2}$ .

Since the traveling waves propagate to the two ends of the cable in two opposite directions, and they are roughly mirrored on the two sides of the fault, the monitoring position closest to the fault will record traveling waves earliest. In addition, there are signal steps in the sheath current when a fault occurs, and they correspond exactly to the arrival time of the traveling wave. For the density-based hierarchical clustering method, they are the two branch nodes with the smallest  $\lambda$  value. In the case, they correspond to node 4 and node 8 in Figure 6. It is noted that the fault is closer to J1, corresponding to node 8 in Figure 6 or Cluster 4 in Figure 7. Therefore, the samples of Cluster 4 and Cluster 0 can be analyzed together. As shown in Table 4, the time stamp of the only sample labeled as

Cluster 0 in Cluster 4 samples is  $1.35 \mu\text{s}$ . It is exactly the expected time of the first arriving traveling wave recorded at J1, corresponding to the sheath current  $I_{a2}$ .



**Figure 8.** The number of the sheath current samples in each cluster of the HV cable circuit.

**Table 4.** The clustering results of some data labels.

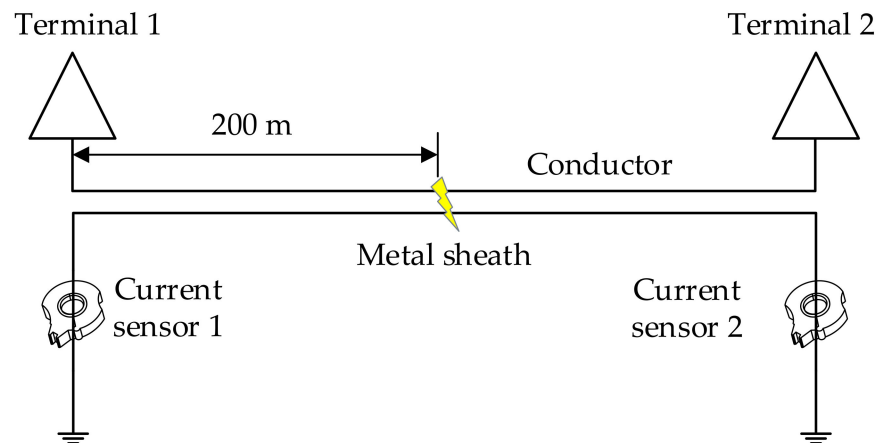
Serial Number *	Cluster Label	Time Stamp/ $\mu\text{s}$
131	4	1.31
132	4	1.32
133	4	1.33
134	4	1.34
<b>135</b>	<b>0</b>	<b>1.35</b>
136	4	1.36
137	4	1.37

(\* The sampling rate is 100 MHz).

### 3.2. The Application for a MV Cable Circuit

The method also can be applied for MV cable circuits, whose radial structure was shown in Figure 1b. The metal sheath and the metal armor are connected and grounded together at the same position of the two terminals, as presented in Figure 9. In this case, the MV cable circuit is 400 m and the fault was set to 200 m from terminal 1. Table 5 shows the parameters of cross-sectional structure of the MV cable. It is to be noticed that the material parameters of the main insulation (XLPE, crosslinked polyethylene) are the same as those in the HV cables. The sheath currents are monitored by the two current sensors, as shown in Figure 9, where  $I_{m1}$  is the sheath current recorded by current sensor 1 and  $I_{m2}$  is the sheath current recorded by current sensor 2.

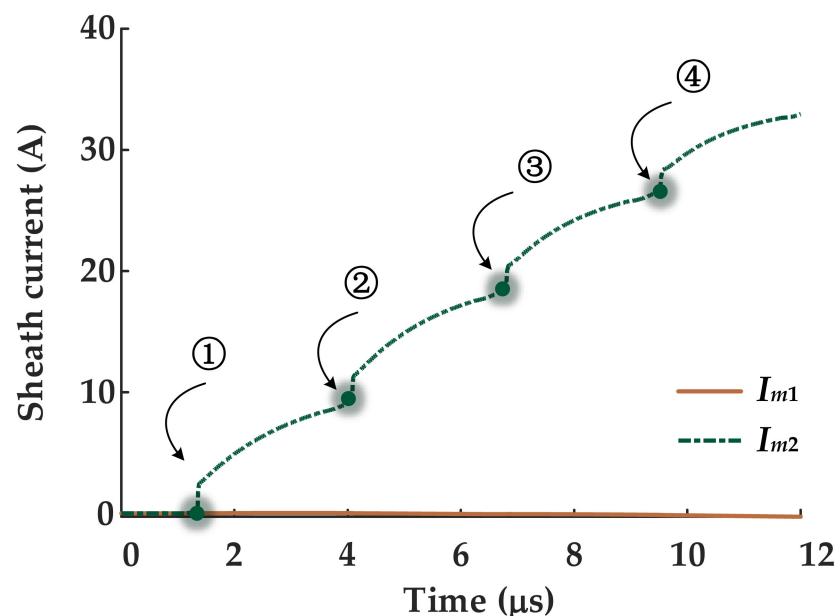
In this case, the initial matrix dimension is greatly reduced, and the process is basically the same. As shown in Figure 10, the initial data is relatively monotonous. In the case, *MinClustSize* and *MinPts* are set to 7 and 7, respectively. By applying HDBSCAN to the sheath current matrix of the MV cable circuit, the condensed cluster tree can be presented in Figure 11.



**Figure 9.** The schematic diagram of a double-end bonded medium voltage (MV) cable system. Note: ‘conductor’ has three cores, one for each phase.

**Table 5.** The parameters of cross-sectional structure of the MV cable.

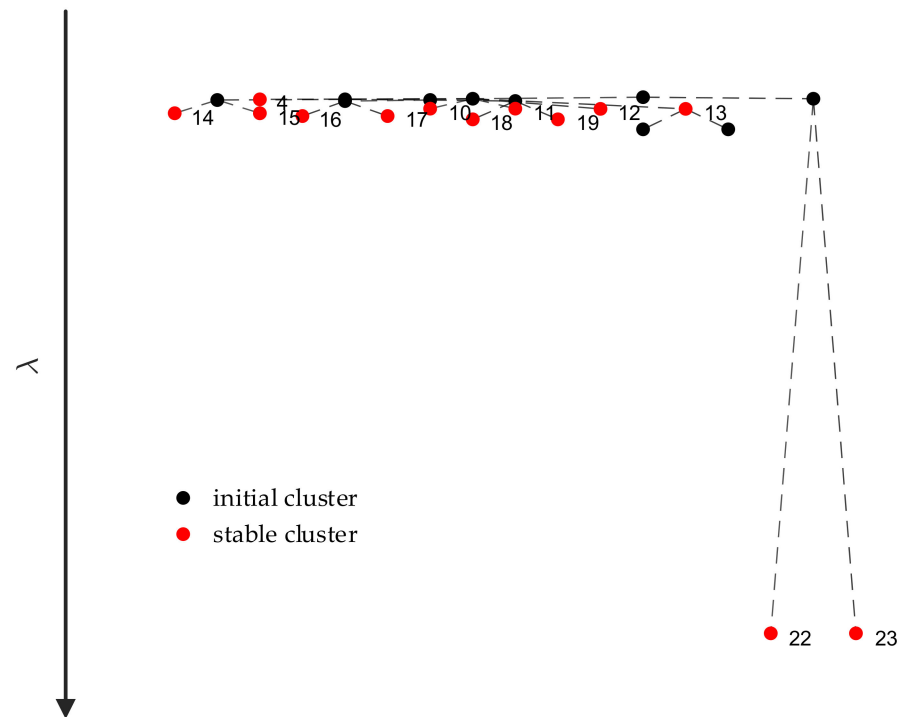
	Structure	Outer Radius/mm
1	Core conductor (Copper)	10.30
2	Main insulation (including inner and outer semi-conductor)	16.60
3	Metal sheath (Copper tape)	16.75
4	Filler and tape wrapping (semi-conductor)	36.50
5	Metal armor (Steel)	38.50
6	Bedding (Water-blocking PVC)	40.10
7	Jacket (Fire-resistant PVC)	44.90



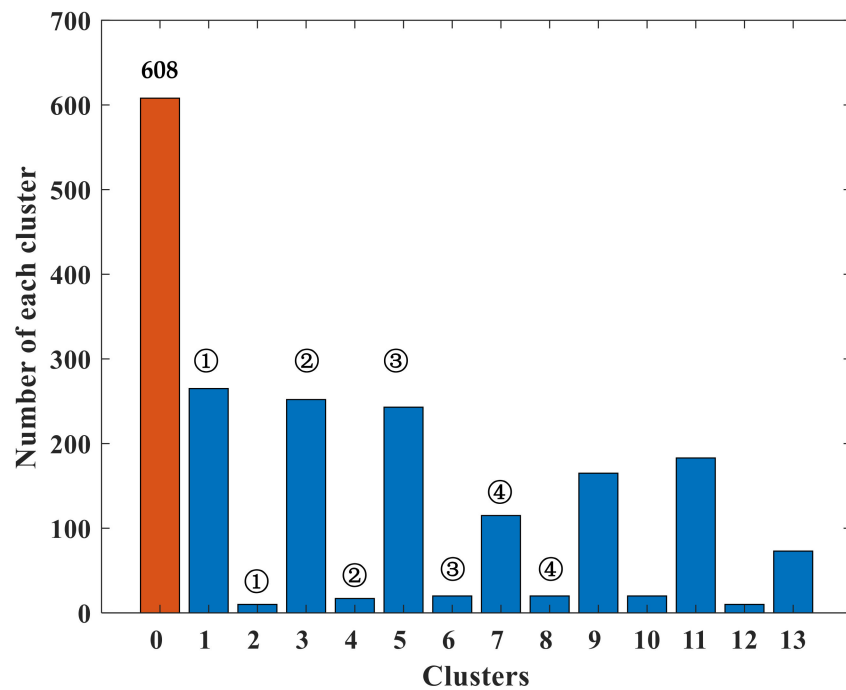
**Figure 10.** The sheath currents monitored by the current sensors at the two ends. (①~④) represent the traveling wave steps no.1 to no.4).

The meaning of the icons in Figure 11 are the same as Figure 6. There are more clusters in this case, and there are two clusters, namely Cluster 22 and Cluster 23, with much bigger  $\lambda$  value than other clusters. They are the two ‘flat’ clusters with the highest density. Therefore, they are not the clusters containing the point of the arrival time, as a result of

the cluster should be the one with the fewest samples. For further analysis, the number of samples in each of the clusters is shown in Figure 12.



**Figure 11.** The schematic diagram of the condensed cluster tree for sheath current matrix of the MV cable circuit.



**Figure 12.** The number of the sheath current samples in each cluster of the MV cable circuit (①~④ represent the traveling wave steps no.1 to no.4).

Similar to the results of the HV cable circuit shown in Figure 6 and Table 4, Cluster 0 still has a large number of the sheath current samples, even with the largest number in this case. As stated, Cluster 0 does not have to be the real noise in practice. Instead, the



point of the arrival time can be just in Cluster 0. The two branches of the traveling wave just correspond to the branch tree of Figure 11 and Cluster 1–13 in Figure 12. The traveling wave steps ①~④ in Figure 10 just correspond to Cluster 1–8 in Figure 12.

#### 4. The Superiority of the Presented Method

##### 4.1. The Comparisons for the Arrival Time Identification

In [23], five popular algorithms for the arrival time identification have been compared. They are, respectively, (A) the threshold method; (B) the wavelet-based multi-scale time-frequency analysis method; (C) the cross-correlation algorithm; (D) the cumulative energy method; (E) t-SNE + DBSCAN. Now add the presented method to the comparison, numbered as method F.

Simulations have been carried out for the MV and HV cable circuits, and the arrival time identification has been performed using the 6 algorithms.

Firstly, simulations have been carried out for the MV cable circuit shown in Figure 9, where two signals have been assessed. The fault position has been set from current sensor 1 to 50 m, 100 m, 150 m, 200 m, 250 m, 300 m and 350 m in sequence. Thus, the corresponding time differences are 2.03  $\mu$ s, 1.35  $\mu$ s, 0.68  $\mu$ s, 0  $\mu$ s,  $-0.68$   $\mu$ s,  $-1.35$   $\mu$ s and  $-2.03$   $\mu$ s, respectively. The time difference results obtained by using the six algorithms under ideal conditions are presented in Table 6.

**Table 6.** The time difference results obtained by using the six algorithms.

Method	TD * ( $\mu$ s)						
	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
A	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
B	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
C	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
D	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
E	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$
F	2.03	1.35	0.68	0	$-0.68$	$-1.35$	$-2.03$

(\*: TD is short for the time difference).

As shown in Table 6, under ideal conditions, all the six algorithms can identify the time difference accurately for MV cables.

Secondly, simulations have been carried out for the HV cable circuit shown in Figure 4, where 12 signals have been assessed. The fault position has been set from the left monitoring position (G1) to 300 m. In order to make the simulations more realistic, white noise, between 0 and 50 dB, has been added to the waveforms. The time difference between the traveling waves arriving at the detection points G1 and J1 should be  $-0.68$   $\mu$ s. The time difference results from the six algorithms are presented in Table 7.

**Table 7.** The time difference results from the six algorithms with white noise.

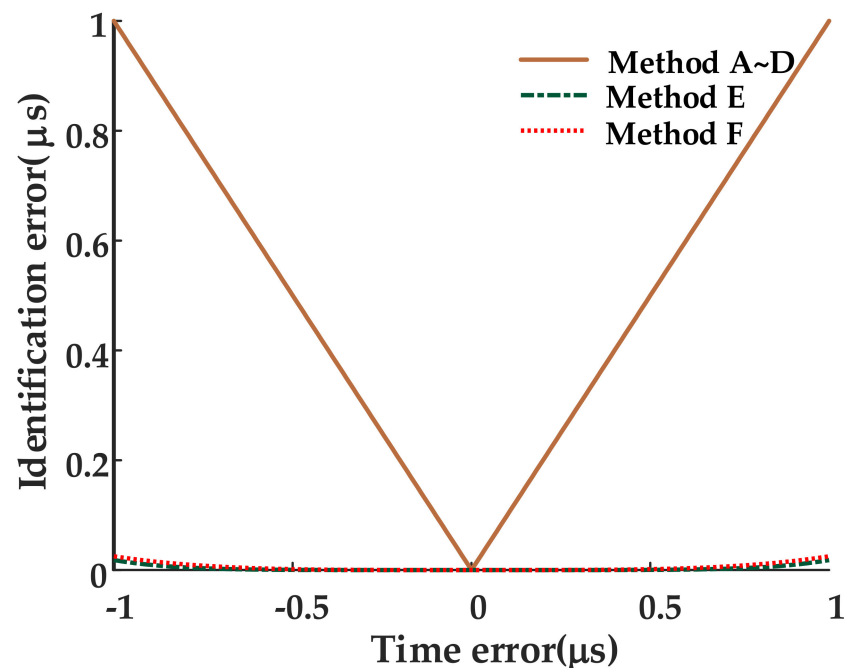
Method	SNR *					
	50 dB	40 dB	30 dB	20 dB	10 dB	0 dB
A	$-0.68$ $\mu$ s	$-0.59$ $\mu$ s	$-0.76$ $\mu$ s	$-0.54$ $\mu$ s	$-0.47$ $\mu$ s	$-0.25$ $\mu$ s
B	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.63$ $\mu$ s	$-0.59$ $\mu$ s	$-0.73$ $\mu$ s	$-0.37$ $\mu$ s
C	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.66$ $\mu$ s	$-0.75$ $\mu$ s	$-0.75$ $\mu$ s	$-0.35$ $\mu$ s
D	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.67$ $\mu$ s	$-0.55$ $\mu$ s	$-0.42$ $\mu$ s
E	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.67$ $\mu$ s	$-0.53$ $\mu$ s
F	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.68$ $\mu$ s	$-0.69$ $\mu$ s	$-0.49$ $\mu$ s

(\*: SNR is short for signal-to-noise ratio).

As shown in Table 7, the six algorithms have different sensitivities to noise. Methods A to E/F sequentially have better noise suppression capabilities; while there is no essential difference between method E and F. When the noise level becomes too high, e.g., 0 dB,

all six algorithms lose the accuracy to varying degrees. In practice, there are many cables which equip with a minimal number of sensors. The noise recorded by these sensors is too large and the synchronization error is too high for the accurate fault location. In order to reduce the influence of randomly generated errors, multiple recorded data from the power cables are recommended to be analyzed together using method(s) E and/or F.

Still with the simulations for the HV cable circuit, all simulation conditions and parameters are the same. This time, a random error has been added to the recording time of the current sensor  $I_{a1}$ , which is, simulating the loss of synchronization time stamp(s). The tolerance range of the random error applied is  $[-1 \mu\text{s}, +1 \mu\text{s}]$ . Under this condition, the performances of the six algorithms are shown in Figure 13.



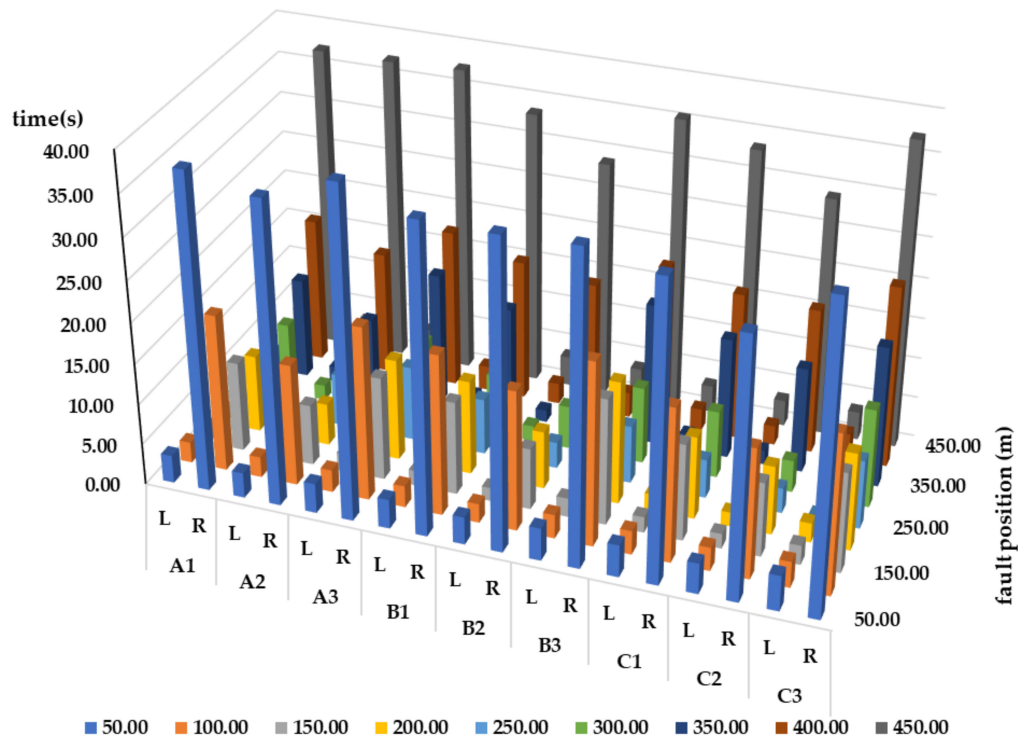
**Figure 13.** The performances of methods A–F with a recording time error added to  $I_{a1}$ . (The ordinate axis represents the absolute value of the identification error).

As shown in Figure 13, when there is a time deviation within the recorded data from  $I_{a1}$ , methods A–D have almost no ability to reduce errors, while the errors of method E and F are almost negligible. Therefore, the performances of method E and F are much better than those of methods A–D. Although the issue has a low probability of occurring in practice, when it does occur, it is catastrophic for the fault location using other algorithms. Even if the error is only microseconds, the fault location error is hundreds of meters. The results indicate that method E and F are suitable for use in fault location with a few random time errors in the recorded data.

#### 4.2. The Improvement from *t*-SNE + DBSCAN

The superiority of the fault location method of the paper mainly lies in the processing of multiple signals, even faster than the former method (*t*-SNE + DBSCAN) [23]. Simulations have been carried out for the comparison. The typical HV cross-bonded cable circuit shown in Figure 2 has been used for the simulation, the fault has been set at cable section A1–C3, respectively, and the fault position has been set, in turn, at 50 m, 100 m, 150 m, 200 m, 250 m, 300 m, 350 m, 400 m and 450 m from the left end of the cable. All simulations have been carried out on an AMD Ryzen Threadripper 3970X 32-Core, 64 GB RAM computer. The running time of each simulation can be presented in Figure 14.

In Figure 14, the running time of R is 2.23~20.94 times than L. Although the time saved is just dozens of seconds for the typical HV cross-bonded cable circuit, it can be hours for a much more complicated circuit or days for practical cable circuits.



**Figure 14.** The running time of each simulation. (L represent the method presented in this paper; R represent the algorithm of t-SNE + DBSCAN; A1~C3 represent the cable sections where the fault occurs).

Another improvement mainly lies in the tree structure. The condensed cluster tree is basically the two-branch-tree structure, which is believed to be corresponded to the traveling waves, generated at the fault position, headed to the two opposite directions to the ends of the cable. Thus, the condensed cluster tree makes the method a better visualization, which is more convenient for the analysis of the state at the beginning of the fault.

## 5. Discussions and Conclusions

### 5.1. Discussions

HDBSCAN is not only facilitates high-dimensional clustering, but also outputs a condensed cluster tree which is believed to be corresponded to the traveling waves. As the condensed cluster tree is basically the two-branch-tree structure, and the hierarchical structure is related to the density level (or the distance), on the other hand, the fault traveling waves generated at the fault position and headed to the two opposite directions to the ends of the cable, therefore, the reason for the appearance of the condensed cluster tree could be related to the propagation process (catadioptric reflection) of the traveling waves. Although the propagation process of the traveling waves is not the direct or the only influencing factor for the appearance of the condensed cluster tree, it could be one of the root factors. The specific relation between the condensed cluster tree and the traveling wave needs further study in the future work. The specific structure of the condensed cluster tree will be the key research object. The distribution of the  $\lambda$  value of Cluster 0 node needs to be figured out. In turn, the proposed method will be more usable in practice. In addition, the true-type test based on high-dimensional data will be carried out in future research.

## 5.2. Conclusions

This paper proposed a novel traveling-wave-based method for fault location of power cables. It improved the former method (t-SNE + DBSCAN) by reducing the data processing steps to reduce data loss. The proposed approach consisted of three main steps: (1) build a matrix of the traveling waves associated with the sheath currents of the cables; (2) cluster the data in the matrix according to its density level and the stability, using HDBSCAN; (3) search for the characteristic cluster point(s) of the two branch clusters with the smallest density level to identify the arrival time of the traveling wave. The simulation comparison of the six methods validated the superiority of the proposed method. The following conclusions can be drawn from the results:

- (1) HDBSCAN runs faster than the former method (t-SNE + DBSCAN). The running time of the former method is 2.23–20.94 times than the proposed method by the simulation test.
- (2) The hierarchical structure of the condensed cluster tree corresponds exactly to the location relationship between the fault point and the monitoring point.
- (3) The proposed method can be used for the identification of the arrival time of the traveling wave.

**Author Contributions:** Conceptualization, M.L.; methodology, M.L. and J.B.; software, Y.S. and Z.P.; validation, Y.W. and C.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Key Scientific and Technical Funds of Zhejiang Electric Power Corporation, grant number 5211DS17000Y.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Dong, X.; Yang, Y.; Zhou, C.; Hepburn, D.M. Online monitoring and diagnosis of HV cable faults by sheath system currents. *IEEE Trans. Power Del.* **2017**, *32*, 2281–2290. [[CrossRef](#)]
2. Li, M.; Zhou, C.; Zhou, W.; Zhang, J.; Zhang, L.; Yao, L. Feasibility Study on Lengthening the High Voltage Cable Section and Reducing the Number of Cable Joints via Alternative Bonding Methods. *High Volt.* **2019**, *4*, 292–299. [[CrossRef](#)]
3. Razzaghi, R.; Lugrin, G.; Rachidi, F.; Paolone, M. Assessment of the Influence of Losses on the Performance of the Electromagnetic Time Reversal Fault Location Method. *IEEE Trans. Power Del.* **2017**, *32*, 2303–2312. [[CrossRef](#)]
4. Zhou, C.; Yang, Y.; Li, M.; Zhou, W. An integrated cable condition diagnosis and fault localization system via sheath current monitoring. In Proceedings of the 2016 International Conference on Condition Monitoring and Diagnosis (CMD), Xi'an, China, 25–28 September 2016. [[CrossRef](#)]
5. Wang, K.; Sun, D.; Zhang, J.; Xu, Y.; Luo, K.; Zhang, N.; Zou, J.; Qiu, L. An acoustically matched traveling-wave thermoacoustic generator achieving 750 W electric power. *Energy* **2016**, *103*, 313–321. [[CrossRef](#)]
6. Xu, Z.Y.; Jiang, S.J.; Yang, Q.X.; Bi, T.S. Ground distance relaying algorithm for high resistance fault. *IET Gener. Transm. Distrib.* **2009**, *4*, 27–35. [[CrossRef](#)]
7. Eissa, M.M. Ground distance relay compensation based on fault resistance calculation. *IEEE Trans. Power Del.* **2006**, *21*, 1830–1835. [[CrossRef](#)]
8. Li, M.; Zhou, C.; Zhou, W. A Revised Model for Calculating HV Cable Sheath Current Under Short-Circuit Fault Condition and Its Application for Fault Location—Part 1: The Revised Model. *IEEE Trans. Power Del.* **2019**, *34*, 1674–1683. [[CrossRef](#)]
9. Manesh, H.M.; Lugrin, G.; Razzaghi, R.; Romero, C.; Paolone, M.; Rachidi, F. A new method to locate faults in power networks based on Electromagnetic Time Reversal. In Proceedings of the IEEE, International Workshop on Signal Processing Advances in Wireless Communications, Cesme, Turkey, 17–20 June 2012. [[CrossRef](#)]
10. Lugrin, G.; Razzaghi, R.; Rachidi, F.; Paolone, M. Electromagnetic time reversal applied to fault detection: The issue of losses. In Proceedings of the IEEE International Symposium on Electromagnetic Compatibility, Dresden, Germany, 16–22 August 2015. [[CrossRef](#)]
11. Bawart, M.; Marzinotto, M.; Mazzanti, G. Diagnosis and location of faults in submarine power cables. *IEEE Elect. Insul. Mag.* **2016**, *32*, 24–37. [[CrossRef](#)]

12. Dashti, R.; Salehizadeh, S.; Shaker, H.; Tahavori, M. Fault Location in Double Circuit Medium Power Distribution Networks Using an Impedance-Based Method. *Appl. Sci.* **2018**, *8*, 1034. [[CrossRef](#)]
13. Bahmanyar, A.; Jamali, S. Fault location in active distribution networks using non-synchronized measurements. *Int. J. Electr. Power Energy Syst.* **2017**, *93*, 451–458. [[CrossRef](#)]
14. Dashti, R.; Ghasemi, M.; Daisy, M. Fault location in power distribution network with presence of distributed generation resources using impedance based method and applying  $\pi$  line model. *Energy* **2018**, *159*, 344–360. [[CrossRef](#)]
15. Orozco-Henao, C.; Bretas, A.S.; Chouhy-Leborgne, R.; Herrera-Orozco, A.R.; Marín-Quintero, J. Active distribution network fault location methodology: A minimum fault reactance and Fibonacci search approach. *Int. J. Electr. Power Energy Syst.* **2017**, *84*, 232–241. [[CrossRef](#)]
16. Dashti, R.; Sadeh, J. Applying Dynamic Load Estimation and Distributed-parameter Line Model to Enhance the Accuracy of Impedance-based Fault-location Methods for Power Distribution Networks. *Electr. Power Compon. Syst.* **2013**, *41*, 1334–1362. [[CrossRef](#)]
17. Ji, T. Study on Fault Location of Distribution Feeders Based on Transient Traveling Waves. Ph.D. Thesis, Shandong University, Jinan, China, March 2006. [[CrossRef](#)]
18. Borghetti, A.; Bosetti, M.; Nucci, C.A.; Paolone, M.; Abur, A. Integrated Use of Time-Frequency Wavelet Decompositions for Fault Location in Distribution Networks: Theory and Experimental Validation. *IEEE Trans. Power Deliv.* **2010**, *25*, 3139–3146. [[CrossRef](#)]
19. Da Silva, A.P.A.; Lima, A.C.S.; Souza, S.M. Fault location on transmission lines using complex-domain neural networks. *Electr. Power Energy Syst.* **2012**, *43*, 720–727. [[CrossRef](#)]
20. Lopes, F.V. Settings-free traveling-wave-based earth fault location using unsynchronized two-terminal data. *IEEE Trans. Power Del.* **2016**, *31*, 2296–2298. [[CrossRef](#)]
21. Lopes, F.V.; Dantas, K.M.; Silva, K.M.; Flávio, B.C. Accurate two-terminal transmission line fault location using traveling waves. *IEEE Trans. Power Deliv.* **2018**, *33*, 873–880. [[CrossRef](#)]
22. Xie, C.; Shao, X.; Jin, Y.; Wang, Y.; Li, M. A Novel Fault Location Method for Power Cables Based on Unsupervised Learning. In Proceedings of the 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Beijing, China, 6–10 September 2020. [[CrossRef](#)]
23. Li, M.; Liu, J.; Zhu, T.; Zhou, W.; Zhou, C. A Novel Traveling-Wave-Based Method Improved by Unsupervised Learning for Fault Location of Power Cables via Sheath Current Monitoring. *Sensors* **2019**, *19*, 2083. [[CrossRef](#)] [[PubMed](#)]
24. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the 2nd ACM SIGKDD, Portland, OR, USA, 4–8 August 1996; pp. 226–231.
25. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*; PAKDD 2013. Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7819. [[CrossRef](#)]
26. Hartigan, J.A. Estimation of a convex density contour in two dimensions. *J. Am. Statist. Assoc.* **1987**, *82*, 267–270. [[CrossRef](#)]
27. Campello, R.J.G.B.; Moulavi, D.; Zimek, A.; Sander, J. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans. Knowl. Discov. Data* **2015**, *10*, 1–51. [[CrossRef](#)]
28. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, Miami, FL, USA, 20–25 June 2009. [[CrossRef](#)]
29. Carrabs, F.; Gaudio, M. A Lagrangian approach for the minimum spanning tree problem with conflicting edge pairs. *Networks* **2020**, 1–14. [[CrossRef](#)]
30. Todo, S.; Matsuo, H.; Shitara, H. Parallel loop cluster quantum Monte Carlo simulation of quantum magnets based on global union-find graph algorithm. *Comput. Phys. Commun.* **2019**, *239*, 84–93. [[CrossRef](#)]