A dimension reduction method used in detecting errors of distribution transformer connectivity

Tang, Zeyang; Shen, Yu; Zhou, Kunpeng; Cao, Kan; Yang, Fan; Cai, Defu; Zhou, Chengke

A Dimension Reduction Method Used in Detecting Errors of Distribution Transformer Connectivity

ZEYANG TANG [ID][1], (Member, IEEE), YU SHEN [ID][1], KUNPENG ZHOU [ID][1], KAN CAO [ID][1], FAN YANG [ID][1], DEFU CAI [ID][1], AND CHENGKE ZHOU [ID][2], (Senior Member, IEEE)

[1]Electric Power Research Institute, State Grid Hubei Electric Power Company Ltd., Wuhan 430077, China
[2]School of Engineering and Built Environment, Glasgow Caledonian University, Glasgow G4 0BA, U.K.

Corresponding author: Zeyang Tang (whu_tzy@126.com)

ABSTRACT Many power utilities may have problems with the quality of data in records about which feeder a distribution transformer is connected to. This affects the operation and maintenance of smart grid infrastructure, outage management, line loss management, and workforce safety. The traditional manual way of verifying and updating the distribution transformer connectivity (DTC) is time-consuming and labor-intensive. Researchers have proposed the method which makes use of secondary side single-phase voltage of distribution transformers to verify DTC. However, when the windings of distribution transformers are yyn0 connected, the performance of the single-phase voltage method (SPVM) is unsatisfactory due to the unbalanced three-phase voltages. This paper proposed a dimension reduction method (DRM) that can convert the unbalanced three-phase voltage to a balanced voltage. The data which include 29 days' secondary side voltages of 3866 distribution transformers have been collected. The performances of the DRM and the SPVM have been compared. Results show that the DRM always has higher accuracy compared with SPVM. The influences of the maximum voltage difference and the reference distribution transformer have also been discussed. The DRM's results are more stable compared with SPVM.

INDEX TERMS Distribution transformer connectivity, data-driven, dimension reduction method, similarity, unbalanced.

I. INTRODUCTION

Distribution network topology (DNT), as an important part of basic network data, has recently received increasing attention. It is known that the correctness of DNT will significantly affect the management of distribution network losses and maintenance of the distribution network. In order to reduce active network losses [1], to balance load among the three phases [2], [3] and to improve power system reliability [4], it is necessary to adjust the DNT. However, many power utilities in China are facing the problem of verifying and updating DNT information in their records. In general, DNT data includes secondary side and primary side topology. Fig. 1 shows the types of DNT information. Secondary side topology includes two types of information, i.e. Type 1: to which distribution transformer a customer is connected; Type 2: to which phase a customer is connected.

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Yuan Chen [ID].
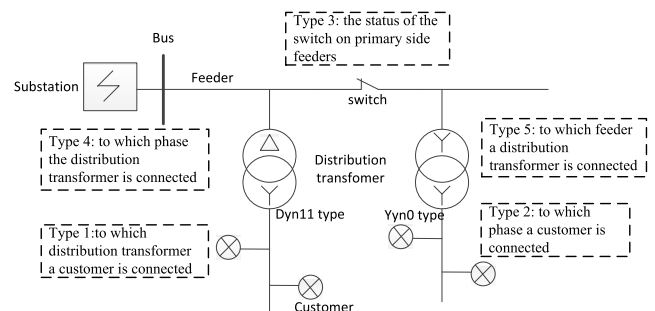

FIGURE 1. Types of DNT information.

Primary side topology includes three types of information, i.e. Type 3: the status of the switch on primary side feeders; Type 4: to which phase the distribution transformer is connected; Type 5: to which feeder a distribution transformer is connected (distribution transformer connectivity, DTC). This paper focuses on Type 5 DNT information.

The traditional manual way of verifying DNT is time-consuming and labor-intensive. Therefore, researchers have proposed device-driven and data-driven methods to solve the problem. The device-driven methods rely on additional high-precision phasor measurement units in distribution networks, such as micro-Phasor Measurement Units (μPMUs) or synchro-phasors [5]. Although these devices provide accurate results, they are very costly and have not been widely used.

Large amounts of operational data, such as the secondary side voltages of distribution transformers, etc. These data not only directly reflect the operating status of the distribution network but also indirectly reflect the topology of the distribution network.

To deal with Type 1 DNT data, [6] proposed a distribution network connectivity (DNC) verification method based on the analysis of smart meter data. The method identified the neighboring meters by voltage profile correlation analysis and predicted customers' upstream and downstream connectivity by voltage magnitude comparison. Reference [7] put forward a method of topology verification for use in low-voltage transformer areas based on discrete Fréchet distance. The methods proposed in [6] and [7] have one thing in common: the similarities of the customer voltage curves have been calculated. Similarity calculation is the basic method to verify DNC.

In respect of Type 2 DNT information, analysis methods can be categorized into two types based on the data type. The first type, based on voltage data [8]–[10], assumed that customers within the same phase shared similar voltage patterns. The second group is based on load data analysis [11]–[13], i.e. finding the optimal combination of households to provide similar load aggregation to that of the known phase load. To deal with Type 3 DNT information, [14] analyzed the data from high-precision phasor measurement units to detect the topology changes in distribution networks, using the IEEE 33-bus model for validation.

To deal with Type 4 DNT data, [15] presents a method for verification and estimation of phase connectivity of the transformer. An alternative approach, in [16], identifies the phase on single-phase taps (laterals and transformers) using voltage measurements of smart meters. The similarities of voltage curves have been calculated and used to identify to which phase a transformer is connected.

Although little literatures have paid attention to type 5 DNT, the literatures [6], [16] about type 1 and type 4 DNT have provided an essential idea that can also be used in type 5 DNT. The essential idea is that we can make use of the similarity of the voltage curves of customers or distribution transformers when verifying DNT. The voltage curves of customers that are connected to the same transformers have high similarity and the voltage curves of distribution transformers that are connected to the same feeders also have high similarity.

As for customers or distribution transformers, the voltage curves may be single-phase or three-phase. The method proposed in [6] and [16] only considered the single-phase

voltage curves. The method used in [6] and [16] can be called single phase voltage method (SPVM). How to deal with three-phase voltage curves has not been mentioned by any literatures. When SPVM is used to deal with three-phase voltage curves, the performance is unsatisfactory and confusing results will be obtained. Especially when the three-phase voltage curves are unbalanced, how to deal with this problem is still worth studying.

In order to solve this problem, the dimension reduction method (DRM) has been originally proposed in this paper. The DRM can convert the unbalanced three-phase voltage curves to the balanced voltage curves. After that, the SPVM can be used verify DTC. Otherwise, confusing results may be obtained when unbalanced three-phase voltage curves are under considered. The details of why SPVM can not be directly used to deal with unbalanced three-phase voltage curves can be found in Section II. The DRM can be recognized as an improvement of SPVM.

The paper is organized as follows. Section II introduces the methods which mainly include how to calculate the similarity between voltage curves and how the dimension reduction method (DRM) works. Case studies have been carried out in section III to compare the performances of DRM and the single-phase voltage method (SPVM). The impacts of maximum voltage difference and reference distribution transformer have been discussed in section IV. Conclusions are made in section V.

II. METHODS FOR DETECTING ERRORS OF DTC
The operation experience shows that the distribution transformers which are connected to the same feeder have high similarity in voltage curves. On the contrary, the distribution transformers which are connected to different feeders have low similarity in voltage curves. This is the basic rule that will be used to detect the errors of DTC. Before talking about the DRM, the overview about how to detect error of DTC will be introduced. The authors have proposed a method to detect the errors of DTC. Fig. 2 shows the flow chart of detecting errors of DTC. There are four steps.

a) Data collection and preprocessing.
b) Methods used to calculate similarity.
c) Determination of the similarity threshold.
d) Determination of the correctness of DTC.

Due to the limitation of page length, in this paper we mainly focus on step a) and b). The SPVM and DRM are used to calculate similarity, the performances of the two methods will be compared. The similarity threshold is set casually. The validity of the threshold will not be discussed in this paper. If the similarity of two distribution transformers is bigger than similarity threshold, it can be concluded that the two distribution transformers are connected to the same feeder. Otherwise, the two distribution transformers are connected to the different feeder.

A. DATA COLLECTION AND PREPROCESSING
To detect the errors of DTC, account data and secondary side voltage data of distribution transformers should be collected.
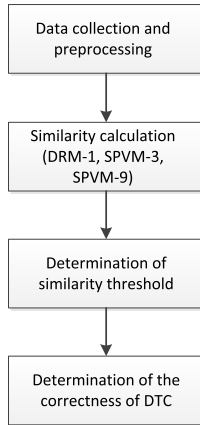
FIGURE 2. Flow chart of detecting errors of DTC.

(a) Distribution transformer YS#1

(b) Distribution transformer YS#2

FIGURE 3. Secondary side voltage curves of distribution transformers (January 27th, 2017).

Account data under consideration includes the names of substations, the names of feeders, and the names of distribution transformers. Secondary side voltage data of distribution transformers refers to the three-phase secondary side voltages data of distribution transformers. In order to improve the accuracy of the procedure, it is necessary to preprocess the collected data, including deleting records containing missing data. In this work, only the distribution transformers with complete data will be considered in the analysis. For example, the secondary side voltage curves of a distribution transformer contain three phases and every phase has 96 points. If one or more data points are missing in any phase voltage curve, the data of the distribution transformers will not be considered in the analysis.

B. SIMILARITY CALCULATION OF VOLTAGE CURVES

1) CORRELATION COEFFICIENT

In reference [18], the value of correlation coefficient is used to reflect the similarity of two variables. In this paper, the correlation coefficient of the voltage curves will be calculated. The greater the absolute value of the correlation coefficient, the greater the probability of the two distribution transformers connected to the same feeder.

The method of calculating the correlation coefficient is shown in (1) :

$$P = \frac{N \sum_{i=1}^{N} (U_{\mathrm{T}1i} U_{\mathrm{T}2i}) - \sum_{i=1}^{N} U_{\mathrm{T}1i} \sum_{i=1}^{N} U_{\mathrm{T}2i}}{\sqrt{N \sum_{i=1}^{N} U_{\mathrm{T}1i}^2 - \left(\sum_{i=1}^{N} U_{\mathrm{T}1i}\right)^2} \sqrt{N \sum_{i=1}^{N} U_{\mathrm{T}2i}^2 - \left(\sum_{i=1}^{N} U_{\mathrm{T}2i}\right)^2}} \quad (1)$$

where $U_{\mathrm{T}1i}$ is the transformer T1's voltage of point i, $U_{\mathrm{T}2i}$ is the transformer T2's voltage of point i, $i = 1 \cdots N$, N is equal to 96.

In this paper, single phase voltage method(SPVM) and dimension reduction method(DRM) will be used to calculate the correlation coefficients. The details of the two methods will be illustrated in the following.
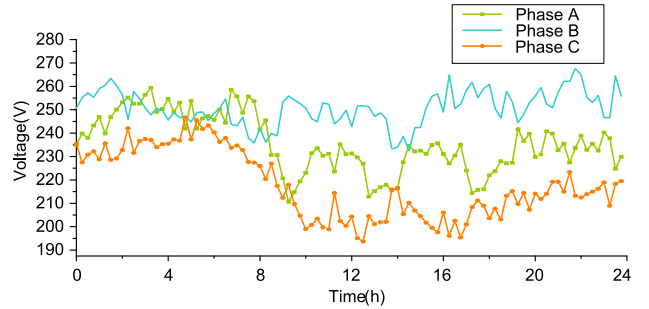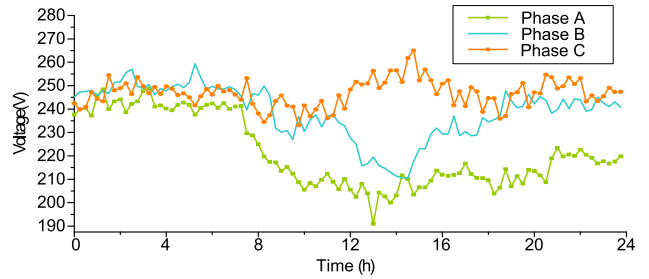
2) SINGLE PHASE VOLTAGE METHOD

For better understanding of SPVM, two distribution transformers will be used as an example. Fig. 3 shows the secondary side voltage curves of two distribution transformers (whose windings are yyn0 connected), (a) YS#1 and (b) YS#2, on January 27th, 2017. It is known that transformers YS#1 and YS#2 are connected to the same feeder. As presented in Fig. 3, the secondary side voltage curves of the two distribution transformers are obviously different. The power consumption information acquisition system collects the secondary side voltages every 15 minutes. Thus, there are 96 points for every phase.

If the phases A of the two distribution transformers are connected to the phase A of the feeder, it can be recorded as A-A. Otherwise, if the phase A of a distribution transformer is connected to the phase A of the feeder, while the phase A of the other distribution transformer is connected to the phase B of the feeder, it can be recorded as A-B.

The SPVM has two subtypes (SPVM-3, SPVM-9). When the phase of the two distribution transformers is consistent (e.g. A-A, B-B, C-C), SPVM-3 will be used to calculate correlation coefficient. When the phase of the two distribution transformers is inconsistent (e.g. A-B, B-C, C-A), SPVM-9 will be used to calculate correlation coefficient.

SPVM-3: The SPVM-3 includes the following steps.

a) Calculate the correlation coefficients (CC1, CC5, CC9 as shown in TABLE 1) using the equation (1).

When SPVM-3 is used, it is assumed that the phase of the two distribution transformers is consistent (e.g. A-A, B-B, C-C). Thus if the two distribution transformers (YS#1 and YS#2) are connected to the same feeder, the three phases

TABLE 1. The correlation coefficient between the Distribution transformers YS#1 and YS#2 (January 27th).

Distribution transformers	Correlation coefficients		
	YS#2 Phase A	YS#2 Phase B	YS#2 Phase C
YS#1 Phase A	0.8149(CC1)	0.7305(CC2)	0.0671(CC3)
YS#1 Phase B	0.1182(CC4)	0.2507(CC5)	-0.0606(CC6)
YS#1 Phase C	0.9009(CC7)	0.7478(CC8)	-0.0057(CC9)

FIGURE 4. Diagram of the shift of high voltage side neutral point of a yyn0 transformer when load imbalance occurs.

voltages curves of the two distribution transformers are expected to have high similarity. However, the results of TABLE 1 show that the similarity of phase A of the two distribution transformers is 0.8149. While the results of phase B and phase C are 0.2507 and -0.0057 respectively. Although the phase A of the two distribution transformers have high similarity, the phase B can C have low similarity. Inconsistent and confusing results have been obtained by SPVM. In order to overcome the inconsistent problem, we will use the maximum of the three correlation coefficients (CC1, CC5, CC9) to represent the similarity of the voltage curves of the two distribution transformer.

b) Calculate the maximum of the three correlation coefficients (CC1, CC5, CC9 as shown in TABLE 1).

c) The maximum correlation coefficient will be used to represent the similarity of the two distribution transformers. If the maximum correlation coefficient is less than the similarity threshold, it shows that DTC has an error.

SPVM-9: The SPVM-9 includes the following steps.

a) Calculate the correlation coefficients (CC1, CC2, CC3, CC4, CC5, CC6, CC7, CC8, CC9 as shown in TABLE 1) using the equation (1).

b) Calculate the maximum of the nine correlation coefficients.

c) The maximum correlation coefficient will be used to represent the similarity of the two distribution transformers. If the maximum correlation coefficient is less than the similarity threshold, it shows that DTC has an error.

3) DIMENSION REDUCTION METHOD

The following section will illustrate why the dimension reduction method (DRM) is needed.

The distribution transformers are three-phase transformers whose windings are yyn0 connected or Dyn11 connected. The operation experience shows that the three-phase secondary side voltages of Dyn11 transformers are balanced, whereas the three-phase secondary side voltages of yyn0 transformers are unbalanced.

The main reason is that when a transformer whose windings are yyn0 connected [10], the high-voltage side is star-connected and ungrounded. If the three-phase loads are unbalanced, the zero-sequence current can not flow in the high voltage side winding. The zero sequence current becomes an excitation current and a 'zero' potential is generated on the high voltage side. This results in the offset of
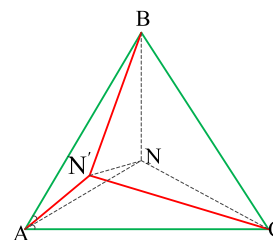
the neutral of the high voltage side and imbalance among the three-phase voltages on the high-voltage side. Correspondingly, the low-voltage side of the three-phase voltages will also be unbalanced. Operation experience shows that many non-energy-saving yyn0 distribution transformers have been installed in the rural distribution network.

As stated above, the unbalance of three-phase loads results in the unbalance of the secondary side voltages of yyn0 transformer (as shown in Fig. 4). The DRM can convert the unbalanced three-phase voltages to a balanced voltage. The balanced voltage will be used to calculate correlation coefficient instead of the unbalanced three-phase voltages. The detailed description of the proposed method is shown as follows.

A neutral point offset diagram of the high-voltage side of a yyn0 transformer is shown in Fig. 4. As previously stated, when a transformer winding is yyn0 connected, the high-voltage side has a star-type connection and is ungrounded. If the three-phase loads are balanced, the secondary side voltage is balanced and the three high-voltage values are equal.

As shown in Fig. 4, AN, BN, CN represent the three-phase voltages when three-phase loads are balanced. If the three-phase loads are unbalanced, the secondary side voltage is also unbalanced, causing a shift on the high voltage side and the neutral point N shifts to N'. In Fig. 4, AN', BN', CN' represent the three-phase voltages when three-phase loads are unbalanced. AB, AC, BC represent line voltages.

When the three-phase loads are balanced, assume $AN = BN = CN = x$, then $AB = \sqrt{3}x$. If the three-phase loads are unbalanced, according to the cosine theorem, the following relationships exist:

$$\cos \angle BAN' = \frac{AN'^2 + AB^2 - BN'^2}{2AN'AB} \quad (2)$$

$$\cos \angle CAN' = \frac{AN'^2 + AC^2 - CN'^2}{2AN'AC} \quad (3)$$

At the same time,

$$\angle BAN' + \angle CAN' = 60° \quad (4)$$
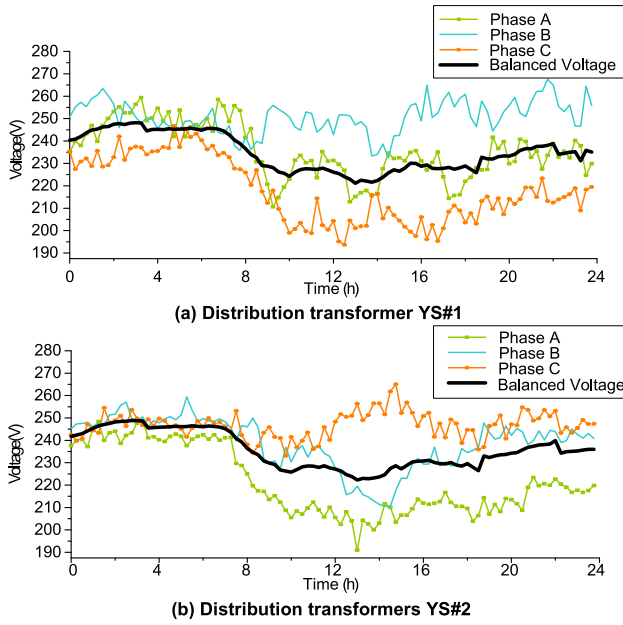
$$AB = AC \quad (5)$$

FIGURE 5. Secondary side voltage curves and balanced voltage curves of distribution transformers (January 27th, 2017).

From (2)-(5), (6) can be obtained.

$$\arccos \frac{AN'^2 + AB^2 - BN'^2}{2AN'AB} + \arccos \frac{AN'^2 + AB^2 - CN'^2}{2AN'AB} = 60° \quad (6)$$

Therefore, if the unbalanced voltages AN', BN', and CN' are known, the balanced voltages AN, BN, and CN can be calculated from (6). Equation (6) is the core of DRM.

As it has been mentioned in section II, every phase of the distribution transformer has 96 points voltages. In Fig. 5, the three-phase voltages of distribution transformer YS#1 are unbalanced. The first points of the three-phase voltages of distribution transformer YS#1 are used as an example. The first points of phase A, phase B, and phase C are 235V, 250.7V, and 235.4V respectively. So $AN' = 235$, $BN' = 250.7$, $CN' = 235.4$. If the three variables are put into equation (6), then we can obtain equation (7).

$$\arccos \frac{235^2 + AB^2 - 250.7^2}{2 \times 235 \times AB}$$
$$+ \arccos \frac{235^2 + AB^2 - 235.4^2}{2 \times 235 \times AB} = 60° \quad (7)$$

The numerical solution of equation (7) is 426.1. Thus $AB = 426.1$. Because $AB = \sqrt{3}AN = \sqrt{3}BN = \sqrt{3}CN$, we can calculate the balanced voltages ($AN = BN = CN = 240.2$). When all the 96 points of unbalanced three-phase voltages are calculated by equation (6), we can obtain the 96 points of balanced voltages. The details can be found in Appendix.

In Fig. 5, the unbalanced three-phase voltages have been converted to a balanced voltage by using the DRM. The black thick lines represent the balanced voltage of distribution transformer YS#1 and YS#2.

TABLE 2. The results of SPVM-3, SPVM-9, and DRM-1when the Distribution transformers YS#1 and YS#2 are analyzed.

Distribution transformers	Correlation coefficients
SPVM-3	0.8149
SPVM-9	0.9009
DRM-1	0.9969

DRM-1: The DRM-1 includes the following steps.

a) For every distribution transformer, if its' windings are yyn0 connected, DRM will be used to calculate the balanced voltage. Otherwise, if a distribution transformer's windings are Dyn11 connected, one of the three-phase voltages will be used to calculate similarity.

b) Calculate the correlation coefficient between the two distribution transformers (the balanced voltage curves are used)

c) The correlation coefficient will be used to represent the similarity of the two distribution transformers. If the maximum correlation coefficient is less than the similarity threshold, it shows that DTC has an error.

III. CASE STUDY
A. CASE STUDY OF THE TYPICAL DISTRIBUTION TRANSFORMERS
TABLE 2 shows the results of SPVM-3, SPVM-9, and DRM-1 when the distribution transformers YS#1 and YS#2 are analyzed. It can found that the result of DRM-1 is 0.9969 which is the biggest.

If the similarity threshold is set as 0.92, the results of SPVM-3 and SPVM-9 are less than similarity threshold. Thus, we will conclude that the distribution transformers YS#1 and YS#2 are connected to different feeders. Actually, the two distribution transformers are connected to the same feeder. In this situation, SPVM-3 and SPVM-9 will make a mistake, whereas the judgement of DRM-1 is correct.

If the similarity threshold is set as 0.85, only the result of SPVM-3 is less than similarity threshold. In this situation, SPVM-3 will make a mistake, whereas the judgements of DRM-1 and SPVM-9 are correct.

B. CASE STUDY OF A RURAL POWER SUPPLY COMPANY
In order to compare the performances of DRM-1, SPVM-3, and SPVM-9 furtherly, 29 days' data (from 2019 April 1th to April 29th) have been collected from a rural power supply company of China. The data include 189 feeders and 3866 distribution transformers. The connectivity of the 3866 distribution transformers has been verified to be correct.

A reference distribution transformer should be set for every feeder. For example, if a feeder has 25 distribution transformers which can be recorded as $DT_1, DT_2, \cdots DT_{25}$, according to ascending order DT_1 can be set as the reference distribution transformer. The correlation coefficients between the other distribution transformers and the reference

distribution transformer can be calculated and recorded as $CC_{1-2}, CC_{1-3}, \cdots CC_{1-25}$.

1) ONE DAY'S RESULTS

In order to compare the performances of the three methods, the correlation coefficients of 3866 distribution transformers on April 1th have been calculated by using the three methods (DRM-1, SPVM-3, SPVM-9). For every feeder, a reference distribution transformer has been determined according to ascending order. It can be found in Fig. 6 that the average correlation coefficients of DRM-1, SPVM-3 and SPVM-9 are 0.89, 0.75 and 0.78. As has been mentioned, the connectivity of all the 3866 distribution transformers has been verified to be correct. The greater the absolute value of the correlation coefficient, the greater the probability of the two distribution transformers connected to the same feeder. Thus, it can be concluded that the performance of DRM-1 is better than the other two methods because its average correlation coefficient is the greatest.

In order to further analyze and compare the performances of the three methods, 5 threshold values (0.7, 0.75, 0.8, 0.85 and 0.9) have been set. For a feeder, if the correlation coefficients between all the other distribution transformers and the reference distribution transformers are greater than the threshold value, then it shows that the distribution transformer connectivity of the feeder is correct. On the contrary, if the correlation coefficient between a distribution transformer and the reference distribution transformer is less than the threshold value, then it shows that the distribution transformer connectivity of the feeder has an error. In this situation, the method makes mistake. Fig. 7 shows the accuracy of the three methods under different threshold values. When the threshold value is 0.9, the accuracies of the DRM-1, SPVM-3 and SPVM-9 are 56.8%, 22.0%, and 24.5% respectively. With the decrease of the threshold value, the accuracy of the three method increase. Especially when the threshold value is 0.75, the accuracies of the DRM-1, SPVM-3 and SPVM-9 are 92.4%, 57.4%, and 60.9% respectively. In this situation, the performance of DRM-1 is acceptable because the probability of making mistakes is only 7.6%. However, the performances of SPVM-3 and SPVM-9 are still unsatisfactory. No matter what the threshold values are, the performances of DRM-1 are always the best among the three methods. Besides, the results of SPVM-3 and SPVM-9 are quite close. The results of SPVM-9 are a little better than SPVM-3 obviously.

2) 29 DAYS' RESULTS

The similarity calculation of voltage curves may be affected by the quality of the data. The quality of the data may be different every day. The 29 days' data have been used and the average correlation coefficients of the 3866 distribution transformers have been calculated. The results in Fig. 8 show that the average correlation coefficients of DRM-1 are always the biggest among the three methods.
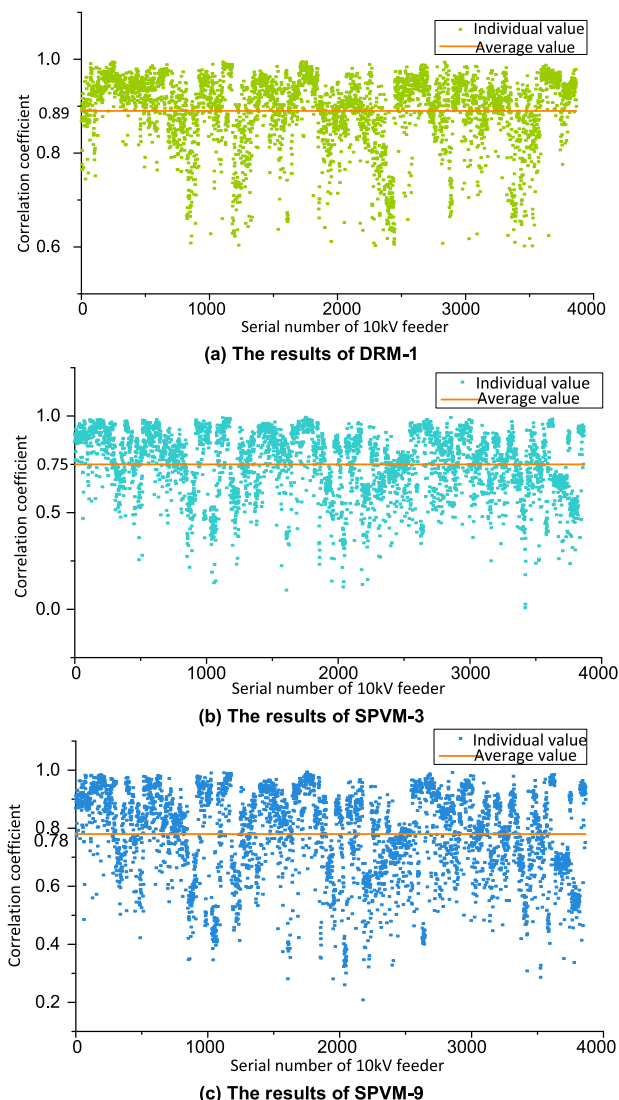

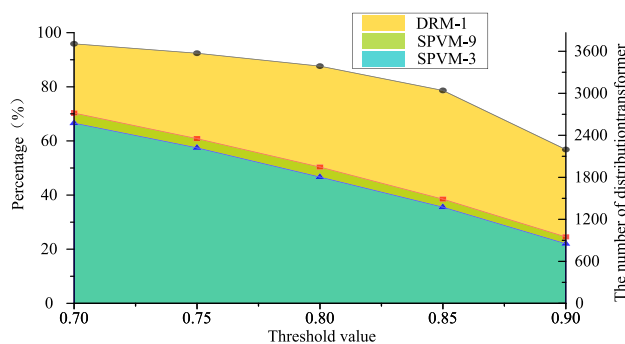**FIGURE 6.** One day's results of the three methods.

FIGURE 7. Accuracy of the three methods under different threshold values.

Fig. 9 shows the accuracy of the three methods in 29 days. It can be found that the accuracy of DRM-1 is always the highest among the three methods and the accuracy of SPVM-3 and SPVM-9 is quite close. Especially on
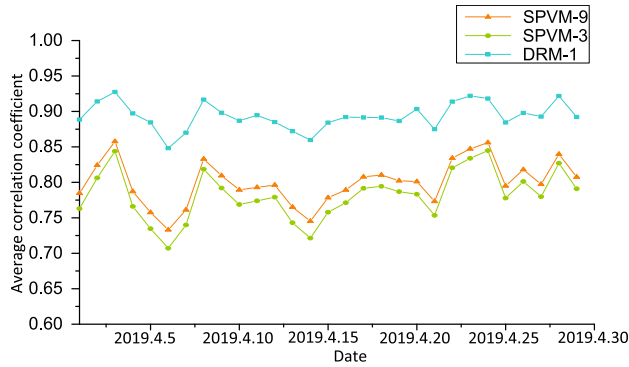
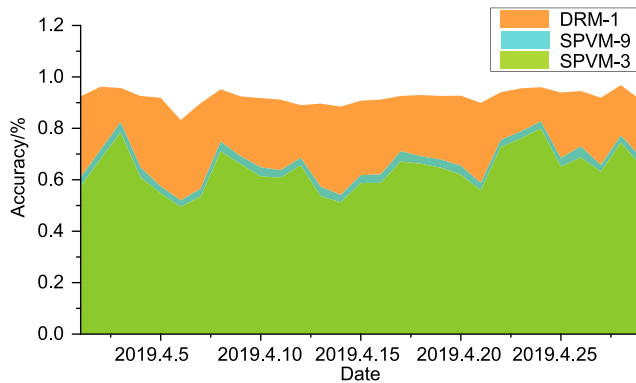**FIGURE 8.** 29 days' results of the three methods.

FIGURE 9. 29 days' accuracy of the three methods when the threshold value is 0.75.

April 28th, the accuracy of DRM-1, SPVM-3 and SPVM-9 are 96.8%, 74.4%, and 77.2% respectively.

From the case study of a rural power supply company, it can be found that the performance of DRM is always better compared with SPVM. No matter which threshold is chosen, the accuracy of DRM is always higher compared with SPVM.

IV. DISCUSSIONS

A. THE IMPACT OF MAXIMUM VOLTAGE DIFFERENCE

As it has been mentioned before, when a distribution transformer is yyn0 connected, the unbalanced of three-phase load will result in the unbalanced secondary side voltage.

For the sake of analyzing the impact of unbalanced secondary side voltage on similarity calculation, an index that can reflect the extent of unbalanced voltage should be defined first.

Assume the three-phase secondary side voltages of a distribution transformer are $U_A = \{U_{A1}, U_{Ai}, \cdots, U_{An}\}$, $U_B = \{U_{B1}, U_{Bi}, \cdots, U_{Bn}\}$, $U_C = \{U_{C1}, U_{Ci}, \cdots, U_{Cn}\}$. Where $n = 96$. In China, the power consumption information acquisition system only acquires 96 points data every day.

For point i, the voltage differences between phase A and phase B is $\Delta U_{ABi} = |U_{Ai} - U_{Bi}|$, the voltage difference between phase A and phase C is $\Delta U_{ACi} = |U_{Ai} - U_{Ci}|$, the voltage difference between phase B and phase C is $\Delta U_{BCi} = |U_{Bi} - U_{Ci}|$. The maximum voltage difference
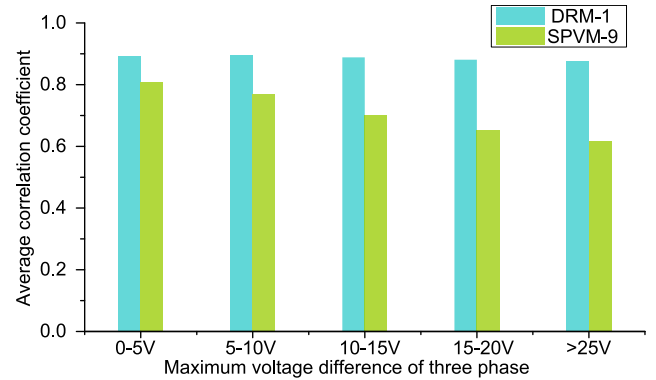

FIGURE 10. The relationship between maximum voltage difference of three-phase and average correlation coefficient.

of a distribution transformer at point i is $\Delta U_{\max i} = \max(\Delta U_{ABi}, \Delta U_{BCi}, \Delta U_{ACi})$. Thus, the maximum voltage difference (MVD) of a distribution transformer for one day is $\Delta U_{\max} = \max(\Delta U_{\max 1}, \Delta U_{\max 2}, \cdots, \Delta U_{\max 96})$.

Fig. 10 shows the relationship between MVD and the average correlation coefficient. In order to compare the performances of DRM-1 and SPVM-9, April 1th data are used. It can be concluded from the results that the impact of MVD on DRM-1 is small, whereas the influence of MVD on SPVM-9 is significant. When MVD is between 0 and 5V, the average correlation coefficient of DRM-1 and SPVM-9 are 0.89 and 0.81 respectively. With the increase of MVD, the differences between DRM-1 and SPVM-9 also increase. Especially when MVD is bigger than 25V, the average correlation coefficient of DRM-1 and SPVM-9 are 0.88 and 0.62 respectively.

When the maximum voltage difference increases, the average correlation coefficient of DRM-1 decrease from 0.89 to 0.88 and the average correlation coefficient of SPVM-9 decrease from 0.81 to 0.62.

From the above analysis, it can be concluded that maximum voltage difference has little influence on DRM and the performance of DRM-1 is more stable compared with SPVM-9.

B. THE IMPACT OF REFERENCE DISTRIBUTION TRANSFORMER

As mentioned above that reference transformer is set according to ascending order. A feeder may have many distribution transformers. If the reference distribution transformer is changed, what will happen? As it has been mentioned in section III, the data which include 189 feeders and 3866 distribution transformers has been collected. Feeder L15 is one of the 189 feeders. The feeder L15 will be used as an example to study the impact of reference distribution transformer on similarity calculation. 12 distribution transformers are connected to Feeder L15. The 12 distribution transformers will be used as the reference distribution transformers in turn. In every turn, the average correlation coefficient will be calculated. Fig. 11 shows that the average correlation
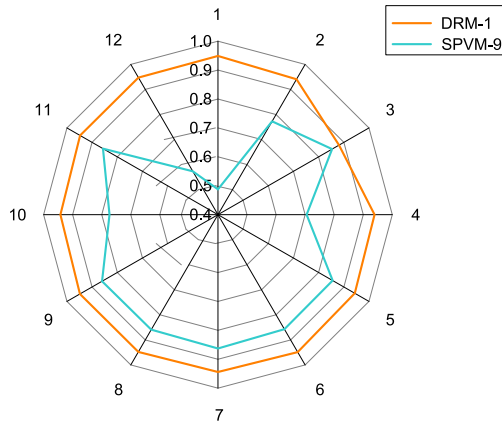
FIGURE 11. The results of feeder L15 when reference distribution transformer is changed.
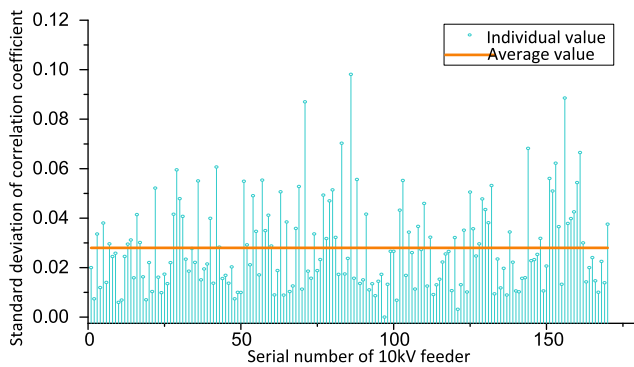

FIGURE 12. The standard deviation of correlation coefficients of 189 feeders when DRM-1 is used.

coefficients of DRM-1 are quite stable no matter which distribution transformer is used as a reference. However, the average correlation coefficients of SPVM-9 have significant differences. The smallest value of DRM-1 is 0.88, whereas the biggest value of DRM-1 is 0.95. The smallest value of SPVM-9 is 0.49, whereas the biggest value of SPVM-9 is 0.86. The results show that the volatility of DRM-1 is smaller compared with SPVM-9. Thus, DRM-1 is more stable compared with SPVM-9.

For the purpose of illustrating the performance of DRM-1 and SPVM-9 further, all the distribution transformers on the 189 feeders have been studied. The standard deviations of correlation coefficients of the 189 feeders have been calculated. Fig. 12 shows that the maximum and average standard deviations of the 189 feeders are 0.098 and 0.028 respectively when DRM-1 is used. However, the results in Fig. 13 show that the maximum and average standard deviations of the 189 feeders are 0.218 and 0.075 respectively when SPVM-9 is used. The maximum and average standard deviations of DRM-1 are smaller compared with SPVM-9. The volatility of the results of DRM-1 is smaller and the performance of DRM-1 is more stable compared with SPVM-9.
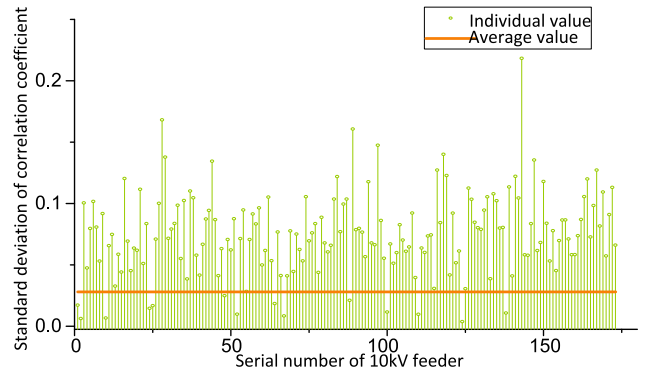

FIGURE 13. The standard deviation of correlation coefficients of 189 feeders when SPVM-9 is used.

V. CONCLUSION

When SPVM is used to verify DTC, the accuracy is unsatisfactory especially when the secondary three-phase voltages are unbalanced. In order to improve the accuracy, a dimension reduction method (DRM) has been proposed. The DRM can convert the unbalanced three-phase voltage to a balanced voltage.

The data which include 29 days' secondary side voltages of 3866 distribution transformers have been collected. Case studies have been carried out. The performances of the DRM and the SPVM have been compared. Results show that the DRM always has higher accuracy compared with SPVM. Especially when the threshold value is 0.75, the accuracies of the DRM-1, SPVM-3 and SPVM-9 are 92.4%, 57.4%, and 60.9% respectively. The influences of the maximum voltage difference and the reference distribution transformer on similarity calculation have also been discussed. The DRM's results are more stable compared with SPVM.

Due to the limitation of page length, the validity of the similarity threshold has not been discussed in this paper. In the next paper, we will discuss the performance of the single threshold and multi threshold that based on voltage fluctuation rate.

APPENDIX
ILLUSTRATION OF HOW DRM CONVERTS UNBALANCED THREE-PHASE VOLTAGES TO THE BALANCED VOLTAGE

As it has been mentioned in section II, every phase of the distribution transformer has 96 points voltages. In Fig. 5, the three-phase voltages of distribution transformer YS#1 are unbalanced. The first points of the three-phase voltages of the distribution transformer YS#1 are used as an example. The first points of phase A, phase B, and phase C are 235V, 250.7V, and 235.4V respectively. So $AN' = 235, BN' = 250.7, CN' = 235.4$. If the three variables are put into equation (6), then we can obtain equation (7).

$$\arccos \frac{235^2 + AB^2 - 250.7^2}{2 \times 235 \times AB}$$
$$+ \arccos \frac{235^2 + AB^2 - 235.4^2}{2 \times 235 \times AB} = 60° \quad (7)$$

TABLE 3. Example of converting unbalanced three-phase voltages to the balanced voltage

Points	The unbalanced three-phase voltages of YS#1			The balanced voltages	Points	The unbalanced three-phase voltages of YS#1			The balanced Voltages
	Phase A	Phase B	Phase C			Phase A	Phase B	Phase C	
1	235	250.7	235.4	240.2	49	231.3	242.8	204.3	225.6
2	239.8	255.2	227.5	240.6	50	229.6	251.7	195.1	224.3
3	238	257.3	230.8	241.7	51	226.9	251.7	193.7	222.8
4	243.2	255.4	232.2	243.4	52	212.9	251.3	204.5	221.8
5	246.9	259	228.8	244.6	53	215.2	247.2	201.1	220.2
6	239.9	260.5	235.6	245.1	54	216.9	248.7	201.9	221.5
7	246.9	263.4	228.6	245.9	55	217.9	246.2	202.1	221.2
8	250.1	260.2	229.2	246.2	56	215.5	233.3	215.9	221.4
9	253.2	256.5	232.8	247.3	57	216.6	234	216.3	222.1
10	255.2	246	242	247.7	58	227.6	238.6	205.4	223.4
11	252.6	257.8	231.6	247.1	59	234.6	232.5	210.2	225.5
12	252.6	254.8	236.6	247.9	60	232.2	242.3	206.9	226.6
13	256.4	250.7	237.5	248.1	61	232.5	242.5	204.5	225.9
14	259.4	247.9	237.1	247.9	62	231.2	250.8	201.7	227.0
15	249.1	250.5	234	244.4	63	234.9	255.8	199.5	228.9
16	250.3	249.8	235.3	245.0	64	235.7	256.8	197.6	228.8
17	254.6	245.6	235.5	245.1	65	231.1	249	206	228.0
18	249.1	248.8	237.3	245.0	66	227.1	264.9	196.1	227.5
19	253	246.6	236.8	245.3	67	230.4	250.5	202.5	226.9
20	242	247.1	246.7	245.2	68	235.1	252.4	195.4	226.4
21	253.8	244.8	237.3	245.2	69	224	258.4	201	226.5
22	241.9	248.8	245.5	245.4	70	214.4	261.8	208.4	226.7
23	246.3	249.1	241.8	245.7	71	215.7	255.2	211.2	226.4
24	247.3	245.4	243.3	245.3	72	216	259.1	209	226.8
25	245.7	248.4	240.3	244.8	73	222	260.9	203.6	227.4
26	250	250.7	236.3	245.5	74	223.8	250.7	207.7	226.6
27	244.4	254.6	237.9	245.5	75	228	246.4	203.1	225.1
28	258.5	243.8	233.7	245.1	76	227.1	257.9	213.2	231.9
29	255.7	243.1	234.7	244.3	77	227.3	253.5	215.2	231.4
30	248.7	246.9	232.8	242.7	78	241.7	244.5	209.7	231.5
31	255.7	237.8	227.7	240.1	79	236.7	248.1	214.4	232.6
32	253.6	235.8	227.4	238.6	80	239.7	253.1	207.3	232.6
33	241.7	241.9	225.9	236.4	81	229.9	255.9	214.1	232.6
34	245.5	236.2	220.4	233.8	82	230.9	259.5	211.9	233.2
35	230.7	239.9	227	232.4	83	240.8	252.5	214.1	235.3
36	230.6	238.9	217.5	228.8	84	239.8	250.9	219.2	236.3
37	220.7	253.3	212.4	228.0	85	232.7	257.6	219.2	235.9
38	210.6	255.9	217.9	227.1	86	235.5	262.2	215	236.7
39	214.8	254.2	209.7	225.2	87	227.5	262.2	223.3	236.9
40	219.4	252.7	204.7	224.6	88	233.6	267.6	213.2	237.0
41	223	250.7	199	223.2	89	238.9	265.1	212.5	237.8
42	231.4	246.2	200.7	225.3	90	232.9	255.4	213.9	233.4
43	233.5	244.8	203.4	226.6	91	235.4	253.2	215	234.0
44	230.3	253	199.7	226.6	92	232.5	256.2	216.2	234.3
45	231.2	252.3	198.9	226.4	93	240.2	246.6	218.9	234.9
46	223.6	243.9	214.4	226.9	94	237.8	246.6	209	230.6
47	235.2	246.1	202.4	227.2	95	224.8	264.5	218.3	234.8
48	231	249.8	200.4	226.1	96	229.9	255.8	219.5	234.5

The numerical solution of equation (7) is 426.1. Thus $AB = 426.1$. Because $AB = \sqrt{3}AN = \sqrt{3}BN = \sqrt{3}CN$, we can calculate the balanced voltages ($AN = BN = CN = 240.2$). When all the 96 points of unbalanced three-phase voltages are calculated by equation (6), we can obtain the 96 points of balanced voltages.

As shown in Table 3, all the 96 points of balanced voltages have been calculated by using the equation (6). Equation (6) is the core of DRM. By using DRM, the unbalanced three-phase voltages of YS#1 have been converted to balanced voltages.

REFERENCES

[1] D. Q. Hung, N. Mithulananthan, and R. C. Bansal, "Integration of PV and BES units in commercial distribution systems considering energy loss and voltage stability," *Appl. Energy*, vol. 113, pp. 1162–1170, Jan. 2014.
[2] C.-S. Chen, T.-T. Ku, and C.-H. Lin, "Design of phase identification system to support three-phase loading balance of distribution feeders," *IEEE Trans. Ind. Appl.*, vol. 48, no. 1, pp. 191–198, Jan. 2012.
[3] C. Xiaoqi, J. Xiaolong, and J. Hongjie, "Reconfiguration strategy for active distribution network considering spatial temporal electrical vehicle load model," *High Voltage Eng.*, vol. 43, no. 3, pp. 1049–1056, 2017.
[4] Y. Jianlin, C. Long, and G. Mingxing, "Reliability evaluation method of distribution network in consideration of prearranged power cut," *High Voltage Eng.*, vol. 43, no. 4, pp. 1248–1255, 2017.
[5] V. Nougain, M. K. Jena, and B. K. Panigrahi, "Synchro-phasors assisted back-up protection of transmission line," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 14, pp. 3414–3420, Aug. 2018.
[6] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1964–1971, Jul. 2015.
[7] G. Juncheng, Z. Xiaofei, and G. Zhimin, "Topology verification of low-voltage transformer areas based on discrete Fréchet distance and editing nearest-neighbors method," *Elect. Meas. Instrum.*, vol. 54, no. 5, pp. 50–55, 2017.
[8] H. Pezeshki and P. J. Wolfs, "Consumer phase identification in a three phase unbalanced LV distribution network," in *Proc. 3rd IEEE PES Innov. Smart Grid Technol. Eur. (ISGT Europe)*, Oct. 2012, pp. 1–7.
[9] H. Pezeshki and P. Wolfs, "Correlation based method for phase identification in a three phase LV distribution network," in *Proc. 22nd Australas. Universities Power Eng. Conf. (AUPEC)*, Sep. 2012, pp. 1–7.
[10] B. K. Seal and M. F. McGranaghan, "Automatic identification of service phase for electric utility customers," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–3.
[11] V. Arya, D. Seetharam, S. Kalyanaraman, K. Dontas, C. Pavlovski, S. Hoy, and J. R. Kalagnanam, "Phase identification in smart grids," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 25–30.
[12] S. J. Pappu, N. Bhatt, R. Pasumarthy, and A. Rajeswaran, "Identifying topology of low voltage distribution networks based on smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5113–5122, Sep. 2018.
[13] S. P. Jayadev, A. Rajeswaran, N. P. Bhatt, and R. Pasumarthy, "A novel approach for phase identification in smart grids using graph theory and principal component analysis," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 5026–5031.
[14] G. Cavraro, R. Arghandeh, G. Barchi, and A. von Meier, "Distribution network topology detection with time-series measurements," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2015, pp. 1–5.
[15] V. D. Krsman and A. T. Sarić, "Verification and estimation of phase connectivity and power injections in distribution network," *Electr. Power Syst. Res.*, vol. 143, pp. 281–291, Feb. 2017.
[16] T. A. Short, "Advanced metering for phase identification, transformer identification, and secondary modeling," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 651–658, Jun. 2013.
[17] Z. Tang, K. Cao, D. Cai, L. Wan, X. Yu, J. Xin, and K. Zhou, "Reduction method of outlet voltage of transformer with unbalanced three-phase load," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Nov. 2017, pp. 1–6.
[18] A. Dejamkhooy, A. Dastfan, and A. Ahmadifard, "K-means clustering and correlation coefficient based methods for detection of flicker sources in non-radial power system," *Russian Electr. Eng.*, vol. 85, no. 4, pp. 251–259, Apr. 2014.

ZEYANG TANG (Member, IEEE) was born in Jingzhou, Hubei, China, in 1987. He received the B.Sc. degree in electrical engineering from Tianjin University, Tianjin, China, in 2009, and the Ph.D. degree in electrical engineering from Wuhan University, Wuhan, China, in 2015. He is currently with the Electric Power Research Institute, State Grid Hubei Electric Power Company Ltd. His main research interests include data analysis and distribution networks.

YU SHEN was born in Xiaogan, China, in 1983. He received the master's degree in high voltage and insulation technology from Wuhan University, China, in 2005. He is currently the Director of the Electric Power Research Institute, State Grid Hubei Electric Power Company Ltd, China. His research interests include energy internet and smart distribution network key technology.

KUNPENG ZHOU received the Ph.D. degree in electrical engineering from Wuhan University, China, in 2004. From 2004 to 2007, he was a Research and Development Engineer with Nanjing Nari Group Company, Nanjing, China. In 2007, he joined the Power Research Institute, State Grid Hubei Electric Power Company Ltd, Wuhan, China, where he is currently a Professorate Senior engineer with the Department of Power System Technologies. His research interests include power system analysis, operation and control, and emerging focus mainly on integration of renewable generation and/or controllable loads, data mining, and power assets management.

KAN CAO received the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 2011. His research interests include power system reliability, and power system operation and control.

FAN YANG was born in 1982. He received the Ph.D. degree from Wuhan University, in 2011. He is currently a Senior Engineer with the Power Research Institute, State Grid Hubei Electric Power Company Ltd. His research interests include distribution automation, electrical equipment fault detection, and diagnosis technology. He is a Senior Member of CSEE.

DEFU CAI received the B.Eng. and Ph.D. degrees in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2014, respectively.

He is currently with the Power Research Institute, State Grid Hubei Electric Power Company Ltd. His research interests include renewable energy and power system analysis.

CHENGKE ZHOU (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in hydropower automation from the Huazhong University of Science and Technology, Wuhan, China, in 1983 and 1986, respectively, and the Ph.D. degree in electrical science from The University of Manchester, Manchester, U.K., in 1994. He joined the School of Engineering and Computing, Glasgow Caledonian University (GCU), Glasgow, U.K., in 1994. He has been a Postdoctoral Research Fellow, a Lecturer, and a Senior Lecturer, since August 2006. He joined Heriot-Watt University, Edinburgh, Scotland, U.K., as a Reader. In 2007, he returned to GCU as a Professor. He has more than 20 years of research experience in power systems and partial discharge-based high-voltage plant condition monitoring. He has acted as a Consultant with EDF Energy, Scottish Power plc, and British Energy. He has published more than 100 articles. He is a Fellow of IET and a Chartered Engineer.

• • •