Received July 9, 2019, accepted July 10, 2019, date of publication July 19, 2019, date of current version August 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2930094

A Novel Method for Screening the PMU Phase Angle Difference Data Based on Hyperplane Clustering

ANCHENG XUE[1], (Member, IEEE), SHUANG LENG[1], YECHENG LI[1], FEIYANG XU[1], (Student Member, IEEE), KENNETH E. MARTIN[2], (Fellow, IEEE), AND JINGSONG XU[1,3]

[1]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Source, North China Electric Power University, Beijing 102206, China
[2]Electric Power Group (EPG), Pasadena, CA 91101, USA
[3]Yinchuan Electric Power Supply Company, State Grid Ningxia Electric Power Company, Ltd., Yinchuan 750001, China

Corresponding author: Ancheng Xue (acxue@ncepu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0902900, in part by the National Natural Science Foundation of China under Grant 51627811 and Grant 51725702, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018ZD01.

ABSTRACT Compared with the traditional supervisory control and data acquisition (SCADA) data, phasor measurement unit (PMU) data is characterized by phase angle measurement and high reporting speed (perhaps 100 Hz). The high reporting speed provides dynamic characteristics of the power system frequency, voltage, and current measurement. PMUs have become one of the important data sources for smart grid monitoring. PMU/WAMS (wide area measurement system) based advanced applications have been widely used in the dispatch centers. Some of the applications, such as line parameter identification and state estimation, depend not only on phase angle data but also on phase angle difference between different locations. Field data can suffer from errors, such as time synchronization error, transducer error, PMU algorithm error, hardware error or malicious attacks, etc. A time synchronization error can directly cause an error in the phase angle difference calculated between the two ends of a transmission line that could degrade a PMU based application. In this paper, a novel method to cluster the phase angle difference data, assess the data quality and screen out the bad PMU phase angle difference data is proposed. First, we develop the hyperplane cluster method to cluster the phase angle difference data. Second, in order to screen out the right data type, this paper compares the virtual reactance parameters of each data type obtained by voltage mean to the line reactance parameter given by the system model. Finally, the performance of the proposed methods has been verified by a simulation. The efficiency of the proposed method has been analyzed. The application of the proposed method using field measured PMU data shows the engineering practicability of the proposed method. In addition, the comparison of the proposed method with other clustering methods is discussed.

INDEX TERMS Data screening, hyperplane clustering, measured PMU data, phasor angle.

I. INTRODUCTION

The synchronized phasor data measured by PMUs is an important source for power grid monitoring [1]. Currently, a series of advanced applications have been developed based on PMU/WAMS [1], [2], such as parameter identification [2]–[4], state estimation [5], [6], low frequency oscillation monitoring traceability and wide area damping control, wide area protection, out-of-step separation [7], [8] and so on.

The reliability of PMU data is essential for its applications. However, experience shows that the measured PMU phase angle has had some quality problems [9]. Studies have shown that time synchronization, power transformers, PMU algorithms, malicious attacks and other factors may cause deviations in measured PMU angle data [10]–[17]. Time

The associate editor coordinating the review of this manuscript and approving it for publication was Lin Zhang.

synchronization problems have been a major contributor to PMU phase angle data error. These problems may due to problems existing in the substation's own time synchronization system, or the GPS signal receiver. For example, the bad weather may influence the GPS receiver and result in biases in the time synchronization [10]. An attack or GPS signal interference by a device may also generate a timing deviation [12], [13]. If a GPS receiver loses the satellite timing signal, it resorts to a local oscillator that will drift in time and cause the PMU phase angle measurement to drift as well [14]. This drift in time can be described as a steady-state deviation Δt between the two substation's time scale and UTC time causing a deviation of $2\pi f_0 \Delta t$ in the PMU phase angle measurement in the substation [15]. In practice, large phase-angle mismatch has been observed in real PMU measurements [16], [17].

Other aspects of data problems have been observed including ferromagnetic resonance and transient response of capacitive voltage transformers, saturation of current transformers [18], the error of spectrum leakage and frequency fluctuation in PMU algorithm. These all have an impact on the accuracy of the PMU data.

To ensure the reliability of PMU applications, bad data detection and screening of the PMU is required. At present, the PMU data detection methods can be divided into two categories: (1) traditional model-based statistical methods; (2) artificial intelligence/data driven based methods which may base on power system model.

The first approach is used by traditional model-based statistical methods realizing the PMU data detection with power system applications such as state estimation, etc. Reference [19] uses the expectation maximization algorithm to screen out the PMU data which may have been attacked. Reference [20] constructs different objective functions and sensitivity matrices by means of hypothesis testing in statistics to achieve bad data detection. Reference [21] detects the bad data by a statistics mutation change.

The second approach adopts artificial intelligence/data-driven based methods for detecting bad data. References [22], [23] use support vector regression estimation to identify abnormal data by comparing the residual between predicted and measured values. Reference [24] uses a density-based spatial clustering method to screen and calibrate PMU data online at both ends of the line. The method does not need accurate model parameters, can detect and calibrate the deviation caused by the PMU and its measurement channel, and has good anti-noise performance. Reference [25] realizes data detection by constructing multiple unsupervised learning models, automatic weight assignment and online recognition.

Furthermore, some data correction methods have been proposed to recover abnormal data [26]–[30]. Reference [26] presents a new approach for synchronized phasor measurement-based state estimation, which can perform phasor angle bias correction with given measurement redundancy. Reference [27] proposes a correction method of PMU phase angle data under GPS spoofing attack, based on synchrophasor-based power system state estimation. For detection and correction of errors in PMU measurements, Ref. [29] utilizes several mathematical techniques based on least-squares optimization to integrate synchronized measurements and power system operation planning data, allowing the method to be used for real-time synchronized measurement calibration. Reference [30] presents a novel data mining-based approach which detects and fixes data manipulation attacks to the wide area monitoring system.

Various methods for detecting and calibrating bad data from PMU have been proposed. Some of them apply clustering methods, such as density-based spatial clustering for the detection of bad data in Ref. [25] and realizing online calibration of PMU in Ref. [24] and Ref. [30]. There is no doubt that applying clustering methods is a way to improve data quality of PMU. In Ref. [31], a review of clustering methods has been presented and the traditional clustering algorithms and modern clustering algorithms are compared.

Among those clustering methods, the K-means clustering algorithm has the nature of relatively low time complexity and high computing efficiency in general. In addition, centroid points of different clusters can be obtained by a K-means clustering algorithm. That makes it possible for applying K-means clustering to online screening for abnormal PMU data. However, the weakness of K-means clustering algorithm is that the K-means results can be influenced by initial points and easily drawn into local a optimum which may lead to an incorrect result. Therefore, in recognizing with the above problem and the problem raised by field data, we propose a PMU bad phase angle data detection and screening method based on hyperplane clustering which is modified from K-means clustering using line electrical connection parameters applied to the measured phase angle difference data of two substations. The innovations of this work are as follows:

(1) Aiming at the features of PMU phase angle difference data error, the clustering method has been improved.

(2) A scheme for detecting and screening the abnormal data of phase angle difference is proposed with high computational efficiency, and strong online application capability.

The remainder of this paper is organized as follows: Section II presents the data and initial analysis for the problem. Section III introduces hyperplane-based data classification method. Section IV introduces the normal data screening method with virtual reactance and the online application strategy. Section V verifies the feasibility of the method through simulation. Section VI applies the proposed method to the actual data and shows the practicality of the method Finally, the conclusions are presented in Section VI.

II. INITIAL ANALYSIS TO MEASURED PMU DATA

A. RELATIONSHIP BETWEEN ACTIVE POWER AND VOLTAGE PHASE ANGLE

A transmission line with PMUs measurement at both ends m and n is shown in Fig. 1. Bus m is designated the sending end and bus n is the receiving end.

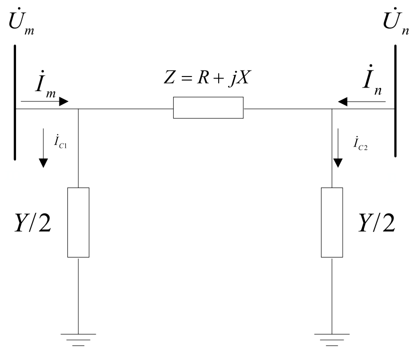


FIGURE 1. Transmission line model.

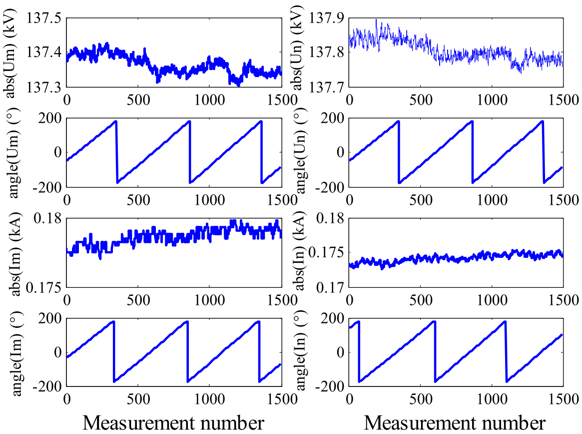


FIGURE 2. Amplitude and phase angle voltage and current at both ends of the line.

By ignoring the line-to-ground admittance branch, the active power on the transmission line can be expressed as:

$$P_{mn} = \left(U_m^2 - U_m U_n \cos \Delta\theta_{mn}\right) G - U_m U_n \sin \Delta\theta_{mn} B \tag{1}$$

where $G + jB = 1/Z = 1/(R + jX)$ is the branch series conductance and series susceptance respectively.

For the high-voltage transmission line, the resistance is much smaller than reactance ($R \ll X$), that is, $B \gg G$ and $B \approx$-1/X. Meanwhile, the phase angle difference between the two ends of the line is small, thus, $\sin\Delta\theta_{mn} \approx \Delta\theta_{mn}$, then

$$P_{mn} = -U_m U_n \Delta\theta_{mn} B = \frac{U_m U_n \Delta\theta_{mn}}{X} \tag{2}$$

Equation (2) shows that the active power P_{mn} is approximately proportional to the voltage phase angle difference $\Delta\theta_{mn}$.

B. PHASE ANGLE "JUMPING" EXAMPLE IN RECORDED DATA

For a 220kV line that is 40.358km long, the measured voltage & current phase angle and amplitude data and the active/reactive power data at both ends of the line are shown in Figures 2 and 3.

Figures 2 and 3 show that, during this period of time, the voltage and current amplitudes at both ends of the line only have small fluctuation. If the phase measurement in a PMU makes a change, it affects both the voltage and current phase angle together, so the power calculations from a single PMU are not affected. This shows the active/reactive power fluctuations are small.

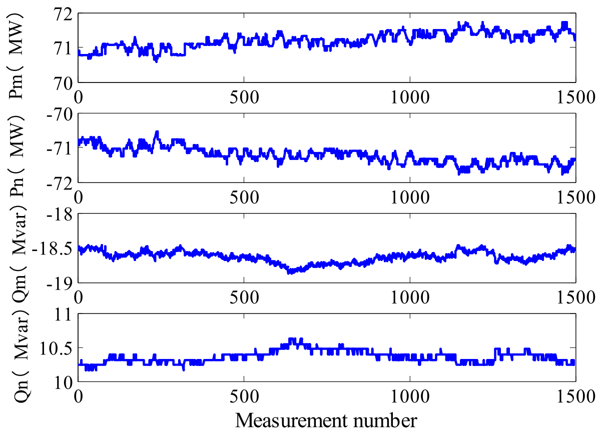


FIGURE 3. Active and reactive power at both ends of the line.

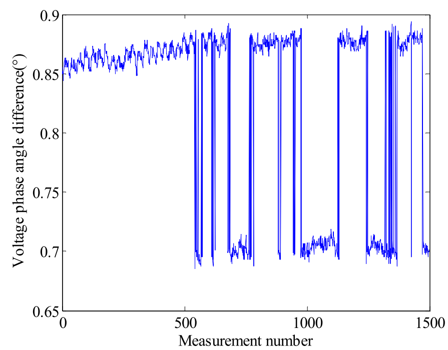


FIGURE 4. Voltage phase angle difference across the line.

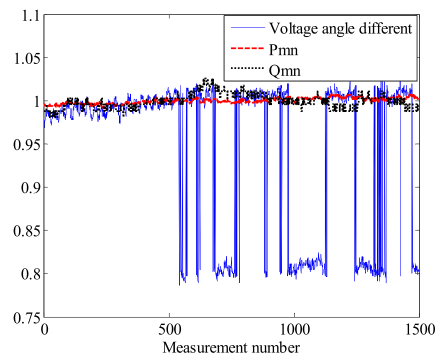


FIGURE 5. Comparison with active and reactive power after normalization.

The voltage phase angle difference between the two ends of the line is shown in Figure 4 and 5. It shows an obvious step jump in the value of measured phase angle difference, but it does not exist in the active power data. Thus, the power calculation remains constant even though the phase angle changes. (2) above shows the active power is proportional to the voltage phase angle difference between the two ends, so here where the power measured at each end does not change when the angle changes indicates the phase angle data may be bad.

Furthermore, the phase angle difference can be roughly classified into two groups, thus, which group is the right one?

The mathematical formulation of the above problem could be stated as follows: for a transmission line, if there is an angle difference step problem in PMU data, then use a clustering method to divide the data followed by an indicator to determine which data is normal.

This paper presents the usage of clustering method to realize the detection and screening of bad data in phase angle difference in the following sections.

Remark: The primary mechanism creating phase angle jumps are jumps in clock states caused by synchronization problems. While the clock may transition between many points, more than two stable states have not been observed. Therefore, the proposed method is only for the two clusters case.

III. CLUSTERING THE DATA

A. THE K-MEANS CLUSTERING METHOD

The K-means clustering method is a widely used unsupervised machine learning method. Given the number of data clusters, the K-means clustering method calculates the distance from all the data to the K cluster centers, and divides it into clusters accordingly.

The specific process of K-means clustering method could be as follows:

(1) Determine the required number of data clusters K;

(2) Randomly select K initial center points;

(3) Calculate the distance from each data point to all the centers to find the nearest cluster center. The data point is assigned to the cluster corresponding to the nearest center. For example, calculate the distances between the i-th data and K centers. If the distance from i-th data to k-th center is the smallest one, then i-th data belongs to k-th cluster;

(4) Calculate the median of the distances from the data to the center in each cluster, and replace the cluster center with the median;

(5) Repeat step (3) and (4) until the cluster's centers do not change.

The K-means algorithm can divide the data into the desired number of categories according to actual needs, but the classification of the data by the K-means algorithm is locally optimal and susceptible to the initial centroid. Furthermore, it is difficult to ensure that the clustering result is the desired one. With the above consideration, the hyperplane clustering method is proposed based on the K-means method combined with electrical connection parameters.

B. HYPERPLANE CLUSTERING FOR PHASE ANGLES

This subsection presents the phase angle difference classification based on hyperplane clustering.

The equation (2) indicates that the active power P_{mn} and the phase angle difference $\Delta\theta_{mn}$ are approximately proportional

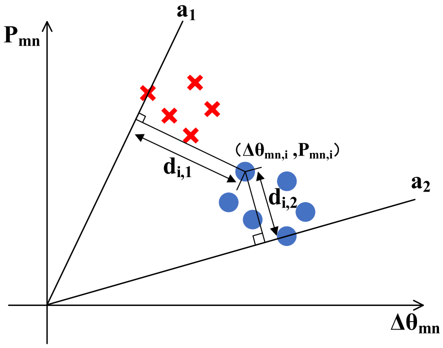


FIGURE 6. The distance from the i-th data point to the center line of the k-th data cluster.

which is a linear relationship. Thus, if there is a step change in phase angle difference data, the P_{mn} will also have a step change, but the ratio between them will not change unless there is an error in the data. The difference of ratio between different data sets can be used to realize hyperplane clustering in screening out abnormal data.

The PMU data contained following data taken from both ends of the line at different times: the amplitude of voltage, the angle of voltage, the amplitude of current, the angle of current, the active power and the reactive power. It has dimension 6 (measurements) $*$ 2 (locations) $= 12$. However, only the phase angle difference $\Delta\theta_{mn}$ and the active power P_{mn} can be selected as features to implement hyperplane clustering for data classification. Therefore, the proposed hyperplane clustering method only needs two-dimensions. It uses hyperplane as data center which is different from K-means clustering. In detail, the hyperplane used in this method is the linear line which relates the linear relationship of the active power P_{mn} and the phase angle difference $\Delta\theta_{mn}$. Therefore, the physical constraints are heuristically included in the proposed clustering algorithm.

Set $\Delta\theta_{mn}$ and P_{mn} as the x-axis and y-axis respectively. The specific hyperplane clustering method is as follows:

(1) Calculate $P_{mn}/\Delta\theta_{mn}$ of each data;

(2) Calculate the maximum and minimum of the $P_{mn}/\Delta\theta_{mn}$ called a_1 and a_2, make them the initial slope values of center line. Initial cluster center line $P_{mn} = a_1\Delta\theta_{mn}, P_{mn} = a_2\Delta\theta_{mn}$ are formed;

(3) Calculate the distance between all data points and the straight line $P_{mn} = a_1\Delta\theta_{mn}$ and $P_{mn} = a_2\ \Delta\theta_{mn}$. Classify each data point into a cluster associated with the nearest straight line.

The distance from the i-th data point to the center line of the k-th data cluster is shown in figure 6 and defined as follow:

$$d_{i,k} = \left|P_{mn,i} - a_k * \Delta\theta_{mn,i}\right| \Big/ \sqrt{1 + a_k^2} (k = 1, 2) \quad (3)$$

(4) In the k-th cluster ($k = 1, 2$), calculate the slopes of all the data points ($P_{mn,i}/\Delta\theta_{mn,i}$, denoted as $a_{k,i}$ for i-th data point) in the cluster, then calculate their median (denoted as $a_{k,median}$), and replace the slope of the center line (denoted

as a_k) obtained by the (n-1)-th iteration with $a_{k,median}$. That is, $a_k^n = a_{k,median}^n$.

(5) Repeat steps (3) and (4) until a_k^n $(k = 1, 2)$ do not change. Then the slope value of each cluster center line $a_k^\infty = \lim_{n\to\infty} a_k^n$ $(k = 1, 2)$ could be obtained.

Remark: Theoretically, the proposed hyperplane clustering method has two advantages over the K-means clustering method to the problem raised in the paper.

i) The proposed hyperplane clustering method adopts the physical feature of the data heuristically, i.e., it uses the feature of the active power P is approximately proportional to the voltage phasor angle difference $\Delta\theta$.

ii) The initialization method in the proposed hyperplane clustering method can avoid random parameter selection, which may degrade the result in K-means clustering.

IV. DATA SCREENING METHOD AND APPLICATION

A. DATA SCREENING METHOD

This section provides a method way to screen out the correct data in the case that line reactance is known, after the hyperplane clustering.

For each point in different clusters, the virtual line reactance of the i-th point in the cluster can be calculated as follows:

$$X_{mn,\mathrm{i}} = U_m^i U_n^i \Delta\theta_{mn}^i / P_{mn}^i \tag{4}$$

If the line reactance X_{ref} is known, the comparison of virtual reactance and actual line reactance X_{ref} could be used to screen for correct phase angle difference. (X_{ref} is the accepted line reactance used in the control room)

In this paper, comparisons of the virtual impedance of different clusters with the actual impedance is used. For the k-th $(k = 1, 2)$ cluster, the central virtual impedance is calculated as follows:

Step 1: Calculate the voltage mean ($\overline{U_m^k}$, $\overline{U_n^k}$) of the k-th clusters.

Step 2: Calculate the virtual impedance X_{vk} of cluster k by using a_k^∞ and $\overline{U_m^k}$, $\overline{U_n^k}$ from (5).

$$X_{vk} = \overline{U_m^k U_n^k} / a_k^\infty \tag{5}$$

Step 3: the indicators of different cluster can be calculated as follows:

$$\gamma_{Xk} = (1 - |X_{ref} - X_{vk}| \big/ X_{ref}) \times 100\% \tag{6}$$

Step 4: the data cluster whose index is closest to 100%, will be determined as the data cluster corresponding to the normal data.

B. DETAILS IN APPLICATION

The proposed method can be applied in the local center, as shown in Figure 7, to where the data of two ends of a line measured by the PMUs would be uploaded. In order to detect the step jumping in the phase angle difference data, a start-up criterion can be set to initiate the proposed method.

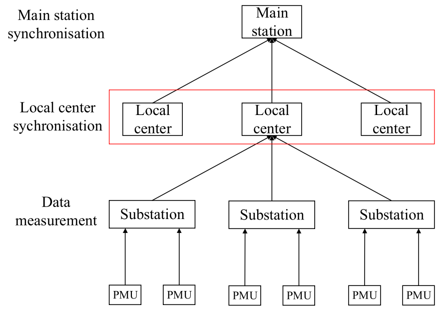


FIGURE 7. Structure of PMU data Collection.

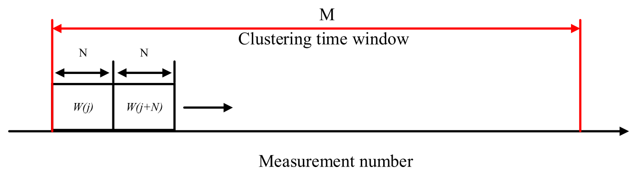


FIGURE 8. Model of detection.

Supposing the amount of data in a detection time window is N, the amount of data in a clustering time window is M. The start-up index is as follows.

$$\sigma(j) = \max\{std(W(j))/std(W(j+N)), \\ std(W(j+N))/std(W(j))\} \tag{7}$$

where j, $j+N$, are the number label for the data. $W(j)$ is the set of angle differences between 2 ends of the line from number label j to $j+N$-1. *Std(W(j))* is the standard deviation of the angles in set $W(j)$.

When there is no step jump and large fluctuation in two detection time windows, the ratio of standard deviation of the data, that is, $\sigma(j)$ is close to 1. When a step jump occurs in $W(j+N)$, σ will become a number larger than 1. In general, the threshold σ_{th} is influenced by the length of the time window N.

Thus, with the above start-up index $\sigma(j)$ calculated in the local center, which receive the data at the both ends of the line, the proposed method could be applied. Once $\sigma(j) > \sigma_{th}$ a timing problem is detected. The data window $C(j)$ used in the proposed method is set from j to $(j+M)$.

In detail, the steps in the process of applying the proposed method could be stated as follows.

Step 1: Calculates the start-up criterion $\sigma(j)$.

Step 2: Is $\sigma(j)$ bigger than σ_{th}? Yes: go to Step 3; No: go to Step 4.

Step 3: Run the proposed method in the time window C(j). Then slide to $W(j+M)$ and $W(j+M+N)$. Repeat step 1 when data is uploaded. Go to step 1.

Step 4: Slide to next detection time window $W(j+1)$ and $W(j+N+1)$. Repeat step (1) and (2) when data is uploading.

In addition, the flowchart of the proposed method is shown in Figure 9.

Remark: In the application of the proposed method, the time delay of data transmission is up to 100–1000ms,

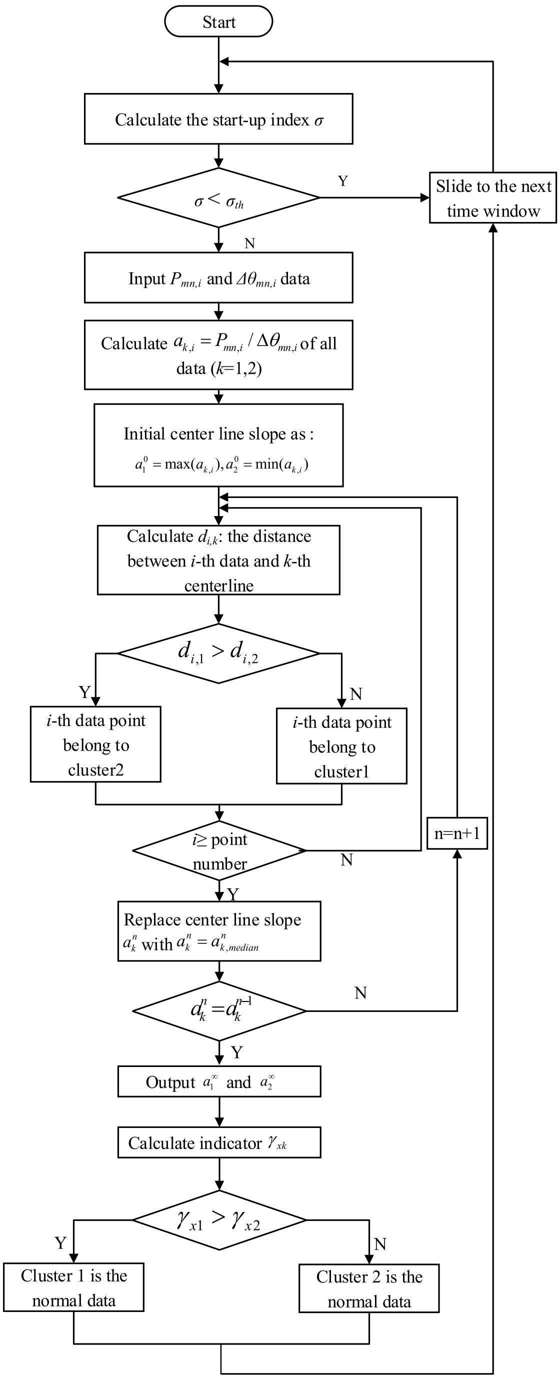
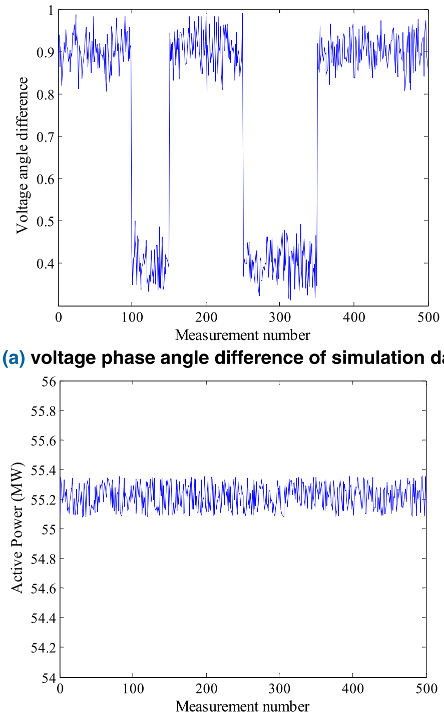


FIGURE 9. Method flow chart.



(a) voltage phase angle difference of simulation data

(b) Power flow of simulation data

FIGURE 10. Voltage angle difference and power flow of simulation data.

or even more, and the time cost of collecting data depends on the length of clustering time window M, which will be several seconds. Then, the application of the proposed method could be "on-line", but it would be delayed with some seconds (may be limited in 10 seconds).

V. CASE STUDY WITH PSCAD

A 500kV 110km single-circuit transmission line is built in PSCAD, and the line is with the typical parameters, i.e., $R = 1.661\Omega$, $X = 28.71\Omega$, and $B = 5.676e\text{-}4S$. The PMUs measure the data of the steady-state positive sequence voltage, current phasor, active and reactive power at both ends of the line is obtained, with the sampling frequency 50 Hz. A total of 4 sets of 500 data points each (10s long) are used in this paper.

To simulate the phase angle step, a phase angle step deviation $-0.5°$ is added in each data set in two places, one from data number 100 to 150 and the other from 250 to 350. One data set has no added noise and the other 3 have zero mean added Gaussian noise. In detail, noise with a standard deviation of 0.1° is added to voltage phase angle and with a standard deviation of 0.5% is added to active power in all 3 cases. In addition, noise with standard deviation intensities of 0.1%, 0.2%, and 0.3% are added to the amplitude of voltage, these are shown in fig 10. The data windows $C(j)$ for simulation and measured data are set as the data window which include all data.

In the simulations, the parameters are set as follows, the threshold of start-up $\sigma_{th} = 2$ and the length of detection time window $N = 50$. With the above data, the start-up index σ at different time can be obtained, as shown in Fig. 11. Fig.11 shows that start-up index $\sigma(j)$ are bigger than threshold σ_{th} at some data points. Therefore, hyperplane clustering data should be initiated.

With the proposed hyperplane cluster method, the data shown in Fig. 4 could be classified into two sets, as shown in Fig. 12 and Fig. 13, for the case of with and without noise, respectively.

Furthermore, the proposed indicator according to different group can obtained, as shown in Table 1.

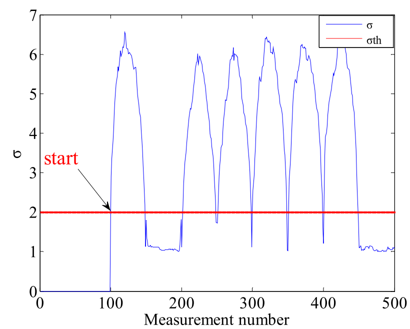


FIGURE 11. The results of start-up index for simulation data.

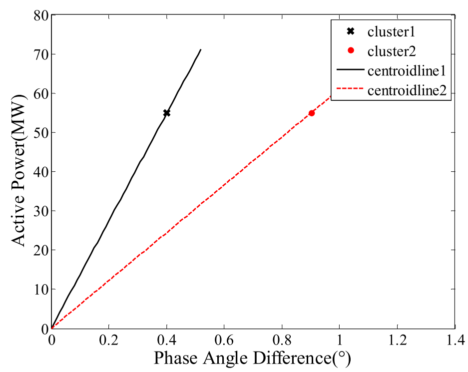


FIGURE 12. Clustering result without noise.

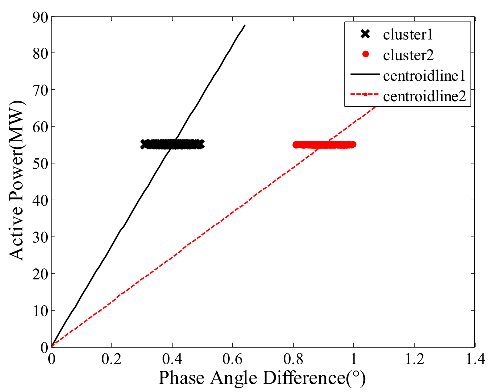


FIGURE 13. Clustering result under 0.2% superimposed noise.

Table 1 shows that, in the case without/with noise, the indicator of the cluster set 1 is much smaller than the cluster set 2, which is consistent with the simulation setting.

Therefore, the proposed method can accurately identify the correct data with/without noise.

Furthermore, Table 1 shows that, although intensities of the Gaussian noise are changed but the indicator γ_{Xk} of the two clusters of data changed little (within 1%).

In addition, the simulations are performed on the computer with Intel Core i5-8400 2.8Ghz, running in the environment of MATLAB®. The time cost and its average for the program containing 500 sets of data is shown in Table 2.

Table 2 shows that the proposed method is fast, and can provide a data check in <0.1s.

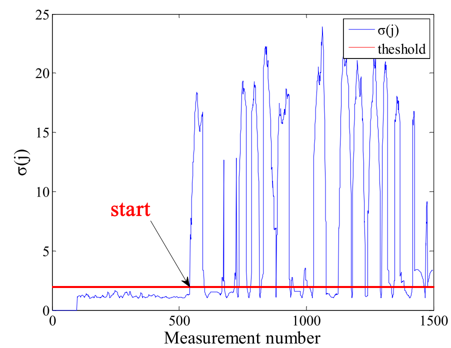


FIGURE 14. The results of start-up index for the measured data A.

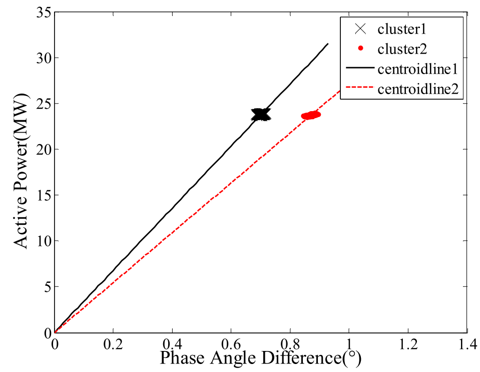


FIGURE 15. Cluster Results for the measured data A.

VI. CASE STUDY WITH MEASURED PMU DATA

A. RESULT OF HYPERPLANE CLUSTERING

This section presents the application of the proposed method to the measured data A mentioned in Section II.

For the 220kV line in section 2, its length is 40.358km, and its line offline parameter is $R = 0.04\Omega$/km, $X = 0.304\Omega$/km, $B = 1.954$e-6S/km. The PMU upload frequency is 25Hz. For the 1500 sets of data (1 minute) with phase angle difference step, the result of start-up index is shown in Fig. 14. The Fig.14 shows that start-up index $\sigma(j)$ are bigger than threshold σ_{th} and the hyperplane clustering data should be initiated.

The result of the proposed hyperplane clustering can be obtained as shown in Fig. 15.

In addition, the proposed indicator according to different group can obtained, as shown in Table 3. Table 3 shows that for the measured data, the indicator of data cluster 2 is bigger than data cluster 1. According to simulation result, data in cluster 2 are the normal data.

B. COMPARISON WITH DIFFERENT CLUSTERING METHOD

For the simulation data and the measured data A, the K-means and Fuzzy C-means could give out the same result as the proposed hyperplane clustering method. However, in different cases, there may be different results. This subsection presents the related examples.

In this subsection, another measured data B of the same transmission line is used, as shown in Fig. 16.

TABLE 1. Proposed indicator for different clusters with/ without noise.

	Noise intensity	Data deviation	Data cluster number	γ_{Xk}	Normal/ Error
No noise	0	-0.5°	1	46.15%	Error
		0°	2	96.41%*	Normal
Noise	0.1%	-0.5°	1	46.23%	Error
		0°	2	96.98%	Normal
	0.2%	-0.5°	1	45.29%	Error
		0°	2	97.58%	Normal
	0.3%	-0.5°	1	46.28%	Error
		0°	2	96.48%	Normal

* Since the power flow equation (2) is an approximation, the simulation result without noise is not exactly 100%

TABLE 2. Running time data multiple times.

	1	2	3	4	5	Mean
Time(s)	0.0532	0.0535	0.0536	0.0547	0.0545	0.0539

TABLE 3. Proposed indicator for the measured data.

Data cluster number	Data cluster 1	Data cluster 2
γ_{Xi}	79.31%	98.73%

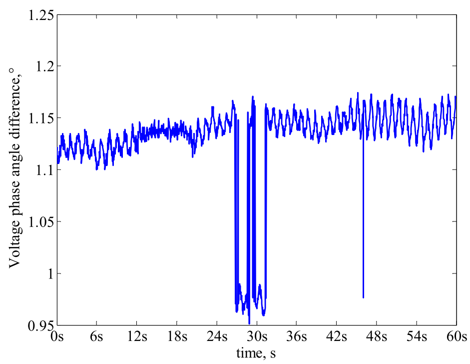


FIGURE 16. Measured data of phase angle difference, measured data B.

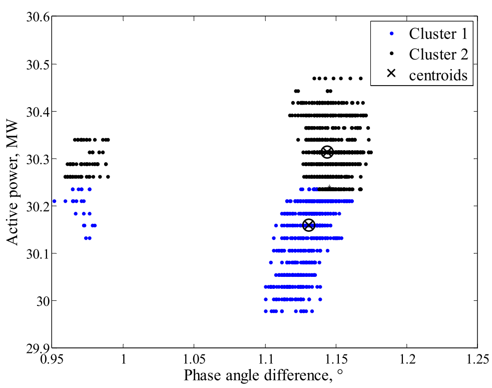


FIGURE 17. The results of K-means with measured data B.

The result of K-means and Fuzzy C-means could be obtained, as shown in Fig. 17, and Fig 18, respectively. The result of proposed hyperplane clustering method is shown in Fig. 19.

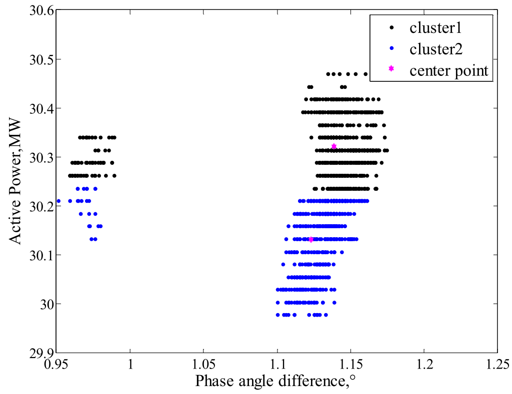


FIGURE 18. The results of Fuzzy C-means with measured data B.

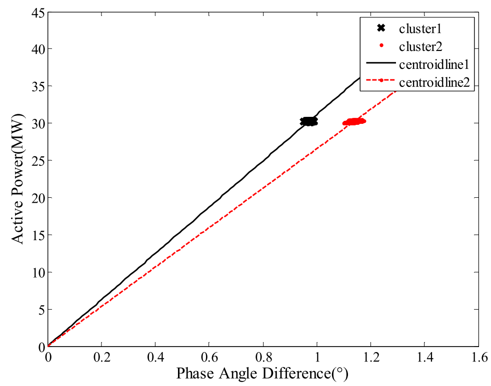


FIGURE 19. The result of hyperplane clustering.

Comparing the result of Fig 19 to the Fig. 17 and Fig. 18, it indicates that the K-means and Fuzzy C-means clustering methods could not give the desired result. Thus, the proposed hyperplane clustering method is recommended.

The reason may be that the proposed hyperplane clustering method heuristically adopts the physical feature of the data, i.e., the active power is approximately proportional to the voltage phase angle difference.

VII. CONCLUSION

A novel method for screening the PMU phase angle difference data based on hyperplane clustering is proposed in this paper. The method is suitable for detecting phase angle difference bias error data, which may be caused by time synchronization error, malicious attack, etc. The proposed paper can classify the PMU phase angle difference data set with and without bias error efficiently based on hyperplane clustering. Through comparing the virtual reactance obtained by the DC power flow equation with the reactance in control room, the abnormal PMU data set can be screened out. If the phase angle error is very small, it may not be possible to distinguish between the right data and bad data. However, for most of the cases, the proposed method works well and it is practical to implement.

REFERENCES

[1] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York, NY, USA: Springer, 2008.

[2] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, Jun. 2010.

[3] S. Gajare, A. K. Pradhan, and V. Terzija, "A method for accurate parameter estimation of series compensated transmission lines using synchronized data," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4843–4850, Nov. 2017.

[4] H. Zhou, X. Zhao, D. Shi, H. Zhao, and C. Jing, "Calculating sequence impedances of transmission line using PMU measurements," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Denver, CO, USA, Jul. 2015, pp. 1–5.

[5] C. Wang, Z. Qin, Y. Hou, and J. Yan, "Multi-area dynamic state estimation with PMU measurements by an equality constrained extended Kalman filter," *IEEE Trans. Smart Grid.*, vol. 9, no. 2, pp. 900–910, May 2018.

[6] R. Su, J. Zhao, G. Zhang, and H. Sun, "A hybrid power system state estimation method considering measurement correlations," (in Chinese), *Power Syst. Technol.*, vol. 42, no. 8, pp. 2651–2658, Aug. 2018.

[7] C. Lu, B. Shi, X. Wu, and H. Sun, "Advancing China's smart grid: Phasor measurement units in a wide-area management system," *IEEE Power Energy Mag.*, vol. 13, no. 5, pp. 60–71, Sep./Oct. 2015.

[8] Y. T. Wang, Y. Tang, L. J. Ding, C. Zhang, Z. Chen, and Y. Huang, "Research and development of new wide area out-of-step control technology for power systems," (in Chinese), *Power Syst. Technol.*, vol. 37, no. 7, pp. 1827–1833, Jul. 2013.

[9] W. Qi, X. Xiao, D. Li, X. Xu, and G. Shao, "The difference between SCADA and WAMS real-time data in dispatching center," (in Chinese), *South Power Syst. Technol.*, vol. 7, no. 5, pp. 87–91, 2013e.

[10] Y. Wang, X. Ren, and Q. Fan, "Synchronous phasor measurement unit and error analysis," (in Chinese), *Electr. Power Autom. Equip.*, vol. 24, no. 1, pp. 66–71, Jan. 2004.

[11] T. Gregorius and G. Blewitt, "The effect of weather fronts on GPS measurements," *GPS World*, vol. 9, no. 5, pp. 52–60, 1998.

[12] Z. Zhang, S. Gong, A. D. Dimitrovski, and H. Li, "Time synchronization attack in smart grid: Impact and analysis," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 87–98, Mar. 2013.

[13] S. Gong, Z. Zhang, M. Trinkle, A. D. Dimitrovski, and H. Li, "GPS spoofing based time stamp attack on real time wide area monitoring in smart grid," in *Proc. IEEE 3rd Int. Conf. Smart Grid Commun. (SmartGridComm)*, Tainan, Taiwan, Nov. 2012, pp. 300–305.

[14] Y. Liu, Y. Jia, Z. Lin, Y. Zhang, L. Wang, K. Tomsovic, and Y. Liu, "Impact of GPS signal quality on the performance of phasor measurements," in *Proc. 16th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, Hersonissos, Greece, Sep. 2011, pp. 1–6.

[15] L. Vanfretti, J. H. Chow, S. Sarawgi, and B. Fardanesh, "A phasor-data-based state estimator incorporating phase bias correction," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 111–119, Feb. 2011.

[16] Q. Zhang, X. Luo, D. Bertagnolli, S. Maslennikov, and B. Nubile, "PMU data validation at ISO New England," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Vancouver, BC, Canada, Jul. 2013, pp. 1–5.

[17] W. Yao, Y. Liu, D. Zhou, Z. Pan, M. J. Till, J. Zhao, L. Zhu, L. Zhan, Q. Tang, and Y. Liu, "Impact of GPS signal loss and its mitigation in power system synchronized measurement devices," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1141–1149, Mar. 2018.

[18] T. Bi, H. Liu, X. Zhou, and Q. Yang, "Impact of transient response of instrument transformers on phasor measurements," in *Proc. IEEE PES Gen. Meeting*, Providence, RI, USA, Jul. 2010, pp. 1–6.

[19] D. Lee and D. Kundur, "Cyber attack detection in PMU measurements via the expectation-maximization algorithm," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Atlanta, GA, USA, Dec. 2014, pp. 223–227.

[20] R. Baldick, K. A. Clements, Z. Pinjo-Dzigal, and P. W. Davis, "Implementing nonquadratic objective functions for state estimation and bad data rejection," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 376–382, Feb. 1997.

[21] S. Pal, B. Sikdar, and J. Chow, "Classification and detection of PMU data manipulation attacks using transmission line parameters," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5057–5066, Sep. 2018.

[22] D. Zhang, W. Li, and Y. Liu, "Reconstruction method of active power historical operating date for wind farm," (in Chinese), *Autom. Electr. Power Syst.*, vol. 38, no. 5, pp. 14–18, Mar. 2014.

[23] L. Wang, R.-Q. Zhang, W. Sheng, and Z.-G. Xu, "Regression forecast and abnormal data detection based on support vector regression," (in Chinese), *Proc. CSEE*, vol. 29, no. 8, pp. 92–96, Mar. 2009.

[24] X. Wang, D. Shi, Z. Wang, C. Xu, Q. Zhang, X. Zhang, and Z. Yu, "Online calibration of phasor measurement unit using density-based spatial clustering," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1081–1090, Jun. 2018.

[25] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, "Ensemble-based algorithm for synchrophasor data anomaly detection," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2979–2988, Mar. 2018. doi: 10.1109/TSG.2018.2816027.

[26] L. Vanfretti, J. Chow, S. Sarawgi, and B. Fardanesh, "A phasor-data-based state estimator incorporating phase bias correction," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 111–119, Feb. 2011.

[27] X. Fan, L. Du, and D. Duan, "Synchrophasor data correction under GPS spoofing attack: A state estimation-based approach," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4538–4546, Sep. 2018.

[28] D. Shi, D. J. Tylavsky, and N. Logic, "An adaptive method for detection and correction of errors in PMU measurements," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1575–1583, Dec. 2012.

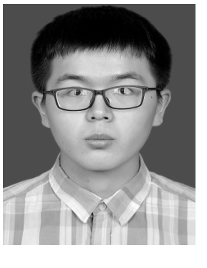
[29] Q. Zhang, V. Vittal, G. T. Heydt, N. Logic, and S. Sturgill, "The integrated calibration of synchronized phasor measurement data in power transmission systems," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2573–2581, Oct. 2011.

[30] X. Wang, D. Shi, J. Wang, Z. Yu, and Z. Wang, "Online identification and data recovery for PMU data manipulation attack," *IEEE Trans. Smart Grid*, to be published. doi: 10.1109/TSG.2019.2892423.

[31] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, 2015.

ANCHENG XUE (M'08) was born in Jiangsu, China, in 1979. He received the B.Sc. degree in applied mathematics and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2001 and 2006, respectively. He was a Postdoctoral with the Institute of System Science, Chinese Academy of Sciences. He is currently a Professor with North China Electric Power University, where he joined in March 2008. His research interests include the model and data-driven power system stability analysis, control and protection, and especially the applications of the PMU.

SHUANG LENG was born in Hubei, China, in 1996. He received the B.S degree in electrical engineering from North China Electric Power University, where he is currently pursuing the master's degree. His research interests include PMU data quality and the application of PMU.



YECHENG LI was born in Shandong, China, in 1998. He is currently a Junior with North China Electric Power University. His research interests include PMU data quality and PMU application.

FEIYANG XU (S'18) was born in China, in 1995. He received the B.Sc. degree in electrical engineering from North China Electric Power University, where he is currently pursuing the master's degree. His research interest includes the application of PMU in power system.

KENNETH E. MARTIN (F'08) started working with synchrophasor measurement with the original PMUs, in 1987, and conducted the first PMU tests. He has over 40 years' experience in the electric utility industry, first at the Bonneville Power Administration (BPA) in communication, precise timing, instrumentation, and testing. He developed the phasor measurement system at BPA, including building the first phasor data concentrator and supported similar developments at many utilities. He was a lead for the IEC 61850, part 90-5, and is the convener for 60255-118-1 developing the joint IEC-IEEE measurement standard. He is currently a Principal Engineer with the Electric Power Group (EPG). He is also a registered Professional Engineer. He has authored or coauthored more than 60 articles and technical papers. He chaired the development of the IEEE C37.118 Synchrophasor Standards from the 2005 original, through the 2014 Amendment.

JINGSONG XU was born in China, in 1992. He is currently pursuing the master's degree with North China Electric Power University. He will join the State Grid Ningxia Yinchuan Electric Power Company, in August 2019. His research interest includes parameter identification in power system.