IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 17, 2020, accepted January 29, 2020, date of publication February 3, 2020, date of current version February 11, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2971030*

# A Data-Driven Vulnerability Evaluation Method in Grid Edge Based on Random Matrix Theory Indicators

**KAI DING[1], YIMIN QIAN[1], YI WANG[1], PAN HU[1], AND BO WANG[2]**

[1]Hubei Electric Power Research Institute, Wuhan 430070, China
[2]School of Electrical and Automation Engineering, Wuhan University, Wuhan 430070, China

Corresponding author: Bo Wang (whwdwb@whu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 51477121.

**ABSTRACT** In order to solve the problem of vulnerability assessment of complex power systems facing complex structures and large sizes, a novel data driven method based on random matrix theory is proposed in this paper. Firstly, with the use of phasor measurement units (PMUs) big data, evaluation matrices are constructed to extract statistical characteristics of power systems operation. Then, with the combination of random matrix theory and entropy theory, vulnerability evaluation index are constructed considering the degree of influence of some faults in power systems. With full use of big data, the model-free method is more accurate and comprehensive. Simulation results in IEEE 39-bus test system and a real-world power grid in China verify the effectiveness of the method.

**INDEX TERMS** PMU measurement, voltage stability, power system vulnerability, random matrix theory, big data.

## I. INTRODUCTION

Due to the investment of new power electronic devices and the implementation of power market policy, a power system is more and more prone to disturbance, resulting in economic losses and serious social impact [1]–[4]. Research shows that the vulnerability identification of the power grid plays a very important role in the occurrence and expansion of the fault [5], it has important theoretical and applied research value, which can provide valuable data reference for the safe and stable operation of the power network and early warning, and then put forward reasonable and effective improvement measures in time. It can also improve the reliability of the power grid, reduce the occurrence of power outages, and maintain the safe and stable operation of the power grid [6]–[11].

There are mainly two types of methods for power system vulnerability identification [12]–[19]. The first one is based on complex network theory. The evaluation indices have been constructed based on the power grid topological structure and then used to explore the reason that caused the failure and evaluate the system vulnerability [12], [20]–[24]. Although this method can build a cascading failure model accurately, it relies too much on the topological structure, which makes it hard to obtain a general model. The second identification method is based on the power system operational conditions by analyzing power flow. Several identification indices regarding voltage stability problem and various aspects have been proposed in the references [15], [16], [25], [26]. However, the complex calculation of power flow and the proposed indices makes it hard to be applied to any given system. In addition, operational condition-based methods failed to capture the system dynamic response after a disturbance and therefore cannot be used for the on-line identification process. Besides, along with the expansion of the power grid scale, the access of new types of equipment, and the implementation of the electricity market, the establishment of the actual model that can reflect all physical characteristics of the grid, is getting more and more difficult.

The development of the Wide Area Measurement System (WAMS) transformed the traditional power system vulnerability problem into a problem that can be solved in a data-driven way [27]–[30]. However, due to the high dimension of measured data sets and a fast sampling rate, it is hard to process and detect the synchronized data samples. Therefore, finding effective big data modeling methods and

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

extracting key features through data mining methods are inevitable requirements for the application of WAMS systems in on-line stability and vulnerability analysis of large-scale power system.

Random Matrix Theory (RMT) originated from the investigation of quantum physics. Many laws were discovered by numerical study in mathematical physics and have been widely used in the fields of finance and communications [31]–[38]. The high-dimensional random matrix was first introduced into the power system field as a useful mathematical tool for analyzing power system big data. The RMT has been widely used for identifying the operational status of power systems and power equipment. Using RMT to identify the critical buses in the power system does not need the system topological structure and can also considerate the dynamic response. As a result, the identification error can be reduced and a general model can be obtained as well. It is easy to be calculated and further applied to the online identification process. A vulnerability identification model has been proposed in paper [39]–[43], which can recognize system critical buses well. Paper [44]–[49] argues calculating the vulnerability of each bus based on random matrix theory and entropy theory, but the index proposed is not comprehensive and may cause misjudgment under some circumstances; Our previous work [50]–[53] also proposes a data-driven method to evaluate vulnerability of power network, however, the index is not comprehensive and accurate enough. In fact, in our previous work, both $s_1$ and $s_2$ calculate the degree of change of the system caused by fault from the perspective of augmented matrix, that is, from the perspective of a single bus, only different calculation methods are used. Therefore, in order to consider the reaction of both single bus and overall system, $s_1$ has been improved in this paper, thus a higher identification accuracy is achieved and complex modeling of the system is avoided.Section II presents the mathematical formula of random matrix theory and its corresponding laws; Section III shows the methodology of identifying critical buses using random matrix theory; Simulation results and analysis in IEEE 39-bus test system and real-world case are given in Section IV;and Section V is the conclusions and future work.

The contributions of this paper are highlighted as follows:

- Apply the random matrix model into the critical bus identification problem;
- Improve the index by considering more comprehensively and more accurately;
- Find the relationship between the characteristics of the M-P law, ring law and entropy and the power system stability status;

## II. MATHEMATICAL FORMULA

Power systems are developing very fast nowadays both in size and complexity. Traditional data analysis tools, by establishing hypotheses and simplified models, are difficult to meet the requirement for dynamic and universal, requiring a new non-model approach. Furthermore, the widespread use of WAMS and smart meters have generated typical space-time data for power systems, thus a new analytical method for big data is required since traditional methods will fail when dealing with big data.

Random matrix theory(RMT) is a statistical-based model-free approach that can be applied to big data analysis, which can handle the challenges above. Without knowing much about the model, the use of raw data can help to easily implement vulnerability optimization in power systems.

### A. RANDOM MATRIX THEORY

Random matrix theory is an effective mathematical tool to analyze complex systems. The elements in the random matrix can be deterministic data or random numbers that follow certain distributions. Although the random matrix theory requires that the dimension of the matrix tends to be infinite, a fairly accurate result can also be observed in a relatively modest matrix (from tens to hundreds of dimensions). This is the fundamental of applying RMT to solve the practical engineering problem.

Let $A$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$. If all $\lambda_j s$ are real, then we can construct a 1-dimensional empirical distribution function

$$F^A(x) = \frac{1}{n}\sum_{j=1}^{n} I(\lambda_j \leq x) \tag{1}$$

Otherwise, we may construct a 2-dimensional empirical distribution function by the real and imaginary parts of $\lambda_j$,

$$F^A(x, y) = \frac{1}{n}\sum_{j=1}^{n} I(\Re(\lambda_j) \leq x; \quad \Im(\lambda_j) \leq y) \tag{2}$$

Then, $F^A$ is called the Empirical Spectral Distribution (ESD) of $A$. The main task of RMT is to investigate limiting properties of $F^A$ in the case where $A$ is random and the order $n$ tends to infinity. Eigenvalues can reflect some characteristics of the matrix, in this way, empirical distribution function of eigenvalues can help to extract characteristics.

### B. MARCHENKO-PASTUR LAW (M-P LAW)

Let $X = \{x_{ij}\}_{1\leq i\leq N, 1\leq j\leq T}$ be a random $N \times T$ matrix whose entries with the mean $\mu(x) = 0$ and the variance $\sigma^2(x) < \infty$, are independently identically distributed (i.i.d.). $N$ is an integer such that $N/T = c \in (0, 1]$. Then the empirical eigenvalue distribution of the corresponding sample covariance matrix $S = 1/N(XX^H)$ converges to M-P law with distribution density function:

$$f_{MP}(x) = \begin{cases} \dfrac{\sqrt{(b-x)(x-a)}}{2\pi x c\sigma^2}, & a \leq x \leq b. \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

where $a = \sigma^2(1 - \sqrt{c})^2$, $b = \sigma^2(1 + \sqrt{c})^2$. M-P law gives the asymptotic behavior of singular values of large rectangular random matrices. Figure 1 shows the relationship

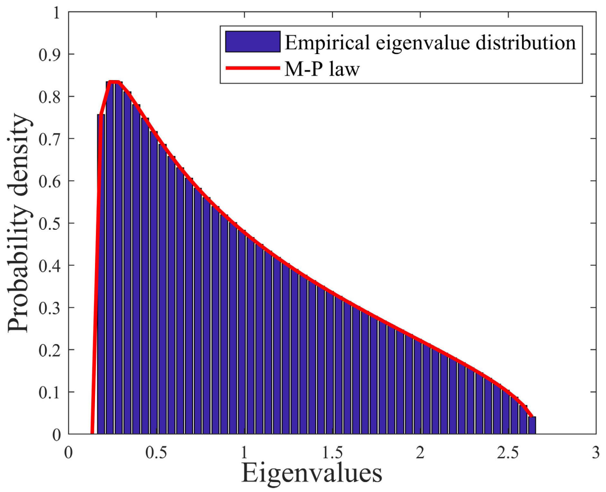



**FIGURE 1.** M-P law representation.

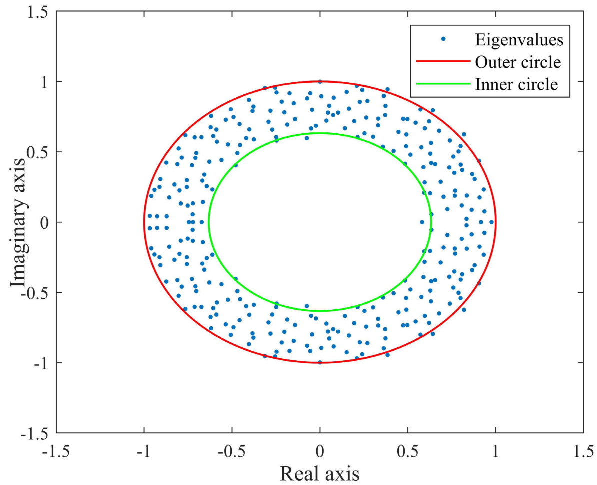



**FIGURE 2.** Ring law representation.

of empirical eigenvalue distribution and the M-P law. The probability density distribution is limited by the M-P law and has a quite uniform distribution shape. It means that if 1) the size of the matrix tends to be infinite 2) the elements of the matrix are independently identically distributed 3) $c \in (0, 1]$, then the empirical distribution of eigenvalues will converge to M-P law, which is determined by $c$ and $\sigma^2$.

### C. THE RING LAW

Considering the matrix product $Z = \sum_{i=1}^{\alpha} \hat{X}_i$, where $\hat{X}_i \in C_{N\times N}$ is the singular value equivalent of rectangular $N \times T$ non-Hermitian random matrix $X_i$, whose entries are i.i.d., variables are with the mean $\mu(x) = 0$ and the variance $\sigma^2(x) = 1$. The empirical eigenvalue distribution of $Z$ converges almost surely to limit given by:

$$f_Z(z) = \begin{cases} \dfrac{1}{\pi x \alpha}|z|^{\frac{2}{\alpha-2}}, & (1-c)^{\frac{\alpha}{2}} \le |z| \le 1. \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

As $N, T \to \infty$ with the ratio $N/T = c \in (0, 1]$. On the complex plane of the eigenvalues, the inner circle radius is $(1-c)^{\alpha/2}$ and outer circle radius is unity. And the singular value equivalent matrix $\hat{X}$ is calculated by:

$$\hat{X} = \sqrt{XX^H}U \tag{5}$$

where $U \in C^{N\times N}$ is a Haar Unitary matrix. The Ring Law extends the RMT to large non-Hermitian random matrices and is one of the most remarkable developments in the modern probability. Figure 2 is the ring law representation. The eigenvalues are distributed between the inner circle and outer circle.

### D. LINEAR EIGENVALUE STATISTICS

The linear eigenvalue statistics describe the distribution of the eigenvalues of the random matrix and can reflect the trace of the random matrix. The Mean Spectral Radius (MSR) is a commonly used linear eigenvalue statistic for random matrices. It is the distance from the origin of the matrix eigenvalues in the complex plane. MSR is defined as follows:

$$r_{MSR} = \frac{1}{N}\sum_{i=1}^{N} |\lambda_i|, \ i = 1, 2, \cdots, N \tag{6}$$

where $r_{MSR}$ is the MSR and $\lambda_i$ is the $i^{th}$ eigenvalue.

## III. IDENTIFYING CRITICAL BUSES USING RANDOM MATRIX THEORY

Modeling the traditional power system vulnerability problem into a data-driven problem has two main advantages. The physical model is no longer needed and the influence of fault types can be ignored.

According to RMT, when the dimensions of a random matrix are sufficiently large, and the elements of which are independent identically distributed, then the empirical spectral distribution converges to M-P law and ring law.

In power systems, the voltage of each bus remains constant when in normal operation, only with some fluctuations caused by measuring errors, noises or small disturbance, and the voltages of all buses are independent and in normal distribution, the data shows a statistical random characteristic, thus the empirical spectral distribution converges to M-P law and ring law, and MSR is larger than the inner circle. While in abnormal operation, there exists correlation and the random characteristic is broken, so the empirical spectral distribution will no longer conform with the M-P law and the ring law, furthermore, MSR is less than the inner circle.

The proposed method uses bus voltage data as the input for the analysis. The system has different behaviors under normal and abnormal operating conditions. Thus the status of power systems can be observed by analyzing the M-P law, ring law and MSR of the random matrix. Figure 3 shows the data sample distribution under two operational conditions, the red line represents the inner circle in the last two figures.

Analyzing the problem from the M-P law point of view, when the abnormal condition occurs, the M-P law cannot

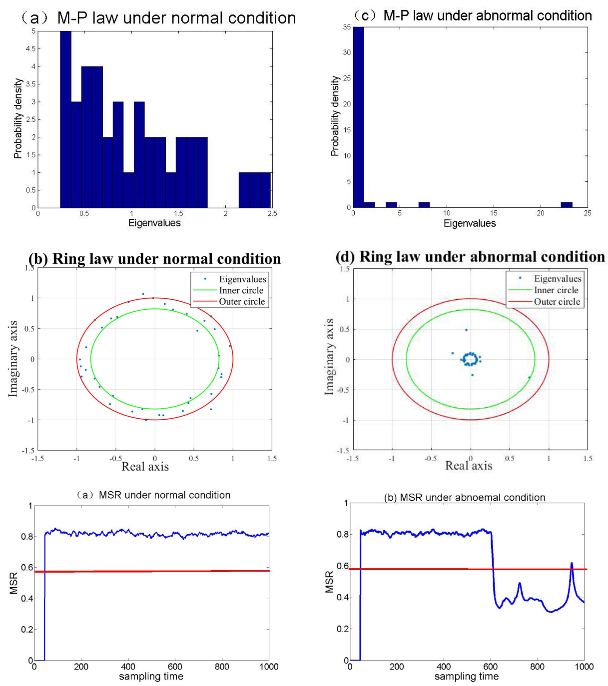



**FIGURE 3.** Data sample distribution under different conditions.

be met. The spectrum distribution histogram becomes narrower and denser and the probability distribution becomes more and more uneven. Furthermore, an abnormally large eigenvalue is obtained as shown in Figure 3. This change can be used to assess whether the system has transitioned from a normal state to an abnormal one.

The trace of the random matrix changes when system operational condition changes. Eigenvalue distribution will against the law when the system is in abnormal conditions. Figure 3 shows that when a system is in the normal condition, all the eigenvalues are distributed between the outer circle and the inner circle. About 2.5% data samples are distributed around the edge of the outer circle because of the error and would not affect the identification result. When the system starts to be abnormal, a large number of abnormal data points within the inner ring begin to appear. After the system is completely destabilized, data samples only exist in the inner circle.

MSR is the mean value of eigenvalues, it can be seen that when the power system is in normal operation, it fluctuates at a value larger than the inner circle; while in abnormal operation, it drops to an inner circle. In conclusion, system operation status can be determined by comparing the MSR and the inner circle radius.

### *A. CONSTRUCTING RANDOM MATRIX USING WAMS DATA*

In a given system, Phasor Measurement Units (PMUs) are implemented at $n$ buses. $k$ key variables are being monitored through each PMU. In total, $N = n \times k$ state variables are obtained. Events in power system often lead to voltage change, so this paper uses nodal voltage to evaluate the critical buses, thus $k = 1$. At one sampling moment $t_i$, the column vector is:

$$\mathbf{x}(t_i) = [x_1(t_i), x_2(t_i), \cdots, x_N(t_i)]^T \tag{7}$$

As the sampling time increases, the $N$ column vectors appear as a matrix in time series:

$$\mathbf{X} = [\mathbf{x}(t_1), \mathbf{x}(t_2), \cdots, \mathbf{x}(t_i), \cdots] \tag{8}$$

For the on-line calculation purpose, a sliding time window is used to collect the PMU data. The window width is $T$ and the random matrix is therefore constructed as:

$$\mathbf{X}_{N\times T} = [\mathbf{x}(t_{i-T+1}), \mathbf{x}(t_{i-T+2}), \cdots, \mathbf{x}(t_i)] \tag{9}$$

This random matrix includes both temporal and spatial characteristics as the sliding time window moves forward to another time point. Note that it is hard to have both $N, T \to \infty$ and $N/T = c \in (0, 1]$ in reality. The sliding time window is actually a way to obtain the matrix. Using the sliding time window to obtain the measurement data of the current time and its historical sampling time from the data source, the width of the time window is actually the number of columns of the matrix. For IEEE39-bus test system, the window width needs to be greater than 39. On this basis, if the window width is too large, the calculation speed is slow, while the window width is too small, the amount of data is less and the accuracy is reduced. In this paper, different values of window width are given in the simulation, including 50, 80, 100, 120, 150, 180 and so on, it is found that good results can be obtained when the window width is about 100, so T is set to be 120 in IEEE-39 test system. As for real-world model, we let $N = 27$ and $T = 80$. In addition, all the data samples should be pre-processed and normalized before the calculation stage. According to [34], when analyzing the correlation between a particular bus and the status of the whole power system, the voltage of this bus is duplicated m times to increase its weight in the matrix, thus the random matrix can better reflect the characteristics of the bus, and can better reflect the contribution of this bus to the status of the power system. The augmented matrix $\hat{\mathbf{X}}'$ is:

$$\hat{\mathbf{X}}' = [\hat{\mathbf{X}}, \hat{x}_j^1, \hat{x}_j^2, \cdots, \hat{x}_j^m] \tag{10}$$

where $\hat{\mathbf{X}}$ is the original random matrix, $\hat{x}_j$ is the voltage related to the particular buses, and $m$ indicates the number of duplications.

When identifying critical buses, each bus is duplicated for an equal number of times to form an augmented matrix that can reflect the characteristics of changes in each bus. The critical buses of the system can be obtained by studying the statistical rules of each augmented matrix using the M-P law and ring law.

### *B. VULNERABILITY IDENTIFICATION INDEX*

The overall system vulnerability identification index is constructed using three proposed indices: M-P law based index, MSR based index and system energy based index.

#### 1) M-P LAW BASED INDEX

M-P law describes the spectrum distribution characteristic of a high-dimensional random matrix. When the system suffers from a change in its status or structure, the spectrum distribution changes accordingly and obviously. The M-P law based index is showing the dynamics of the system before and after the disturbance occurs, which can be defined as:

$$s_1 = \frac{\sum_{i=1}^{\tau}\sum_{j=1}^{n}|\lambda_{ij} - E(\lambda_i)|}{n*\tau} \tag{11}$$

where $\tau$ is the number of sampling points from the start of the fault to the end of the fault, that is, each moment during the fault is considered to increase the accuracy. $\lambda_{ij}$ is the $j^{th}$ eigenvalue obtained by the $i^{th}$ sampling point. $E(\lambda_i)$ is the mean value of all the eigenvalues of the matrix corresponding to this moment. Instead of using the augmented matrix, we use $\hat{\mathbf{X}}$ to calculate the reaction of the whole system.

$s_1$ represents the impact of a fault on the system dynamic performance and it can capture the transition from a normal operating condition to an abnormal operating condition. The system is suffering a huge change of status when $s_1$ is smaller.

#### 2) MSR BASED INDEX

Analyzing the MSR curve associated with the voltage dynamic response, it is found that the results match with each other. When MSR is more than the inner circle radius, the system is more stable. The MSR decreased when a fault occurs in the system and it goes up again when the fault is been cleared. Therefore, the MSR based index is defined as:

$$s_2 = \sum_{j=1}^{n}\frac{E(MSR_{0j}) - E(MSR_j)}{n} \tag{12}$$

where $E(MSR_{0j})$ is the mean MSR of system matrix under steady state and $E(MSR_j)$ is the mean MSR of $j^{th}$ system matrix under the disturbance.

MSR based index $s_2$ is showing the impact of a fault on the other buses. A larger value of $s_2$ indicates a stronger impact on the system.

#### 3) SYSTEM ENTROPY BASED INDEX

When a fault occurs, if the energy changes can be absorbed by all the other buses, the instability problem is less likely to happen. The impact of a fault on a bus can be interpreted by the rate of change in MSR. Entropy theory can be applied to calculate the capability of a system absorbing the energy changes. It is defined as:

$$H = -C\sum_{i=1}^{l}p(\omega_i)lnp(\omega_i) \tag{13}$$

where $C$ is a constant, $l$ is the total number of states, and $p(\omega_i)$ is the probability of each state. Generally, entropy refers to the disorder or uncertainty of a system. The system is under a more disordered condition when the entropy is higher and the system is more stable when the entropy is smaller. In this problem, when a fault a occurs at a bus, the change of MSR in all other buses are calculated and formed the state variables.

$$\omega_j = E(MSR_{0j}) - E(MSR_j) \tag{14}$$

where $E(MSR_{0j})$ is the mean MSR of system matrix under steady state and $E(MSR_j)$ is the mean MSR of $j^{th}$ system matrix under the disturbance. All the $\omega_j$ are normalized as:

$$\mu_j = \frac{\omega_j}{\sum_{j=1}^{n}\omega_j} \tag{15}$$

where $n$ is the total number of buses. Therefore the system entropy-based index is calculated as:

$$s_3 = -C\sum_{j=1}^{n}\mu_j ln\mu_j \tag{16}$$

where $C$ is a constant.

System entropy-based index $s_3$ shows the capability of a system to absorb the transient power changes. The bigger $s_3$ is, the more disorder the system is and the impact of the fault is more severe.

Considering all the above proposed indices, a general vulnerability identification index $S$ can be calculated by:

$$S = \frac{s_2 s_3}{s_1} \tag{17}$$

This equation considers the degree of response of the whole system ($s_1$), the degree of response of single bus ($s_2$), and the degree of chaos in the system ($s_3$). Thus the general index $S$ showed the overall importance of a bus in the system. If $S$ is bigger, the bus is more critical, and the system will suffer from a severer stability issue if the fault occurs at this bus.

### C. OVERALL IDENTIFICATION FLOW CHART

Figure 4 shows the overall flowchart of the proposed critical bus identification method.

## IV. SIMULATION RESULTS AND ANALYSIS

Simulation results are obtained from IEEE 39-bus test system. First of all, the identification results of a detailed example is presented to show the advantage of the proposed method over the previous method. Then the critical buses are identified and the results are compared with all the other existing methods. The simulations are performed in Power System Analysis Software Package (PSASP) and all the calculations are done by MATLAB R2018a.

### A. EXAMPLE OF A DETAILED CASE

The IEEE 39-bus test system is well known as 10-machine New-England Power System. Generator 1 represents the aggregation of a large number of generators. The system has 10 generators and 46 lines. Figure 5 is showing the structure of the test system.

Because various external factors generate random noise in the grid, making the real-time data volatile, the random noise is introduced to reduce the correlation of data in the simulation. The disturbance is set as follows: the system has

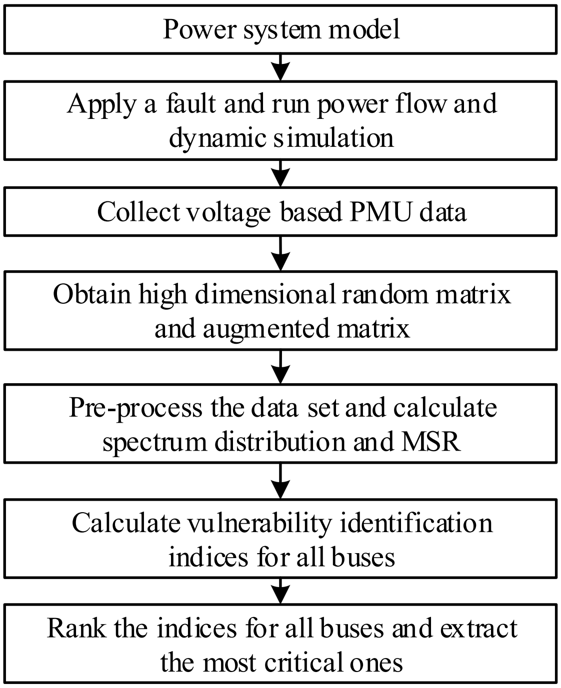



**FIGURE 4.** Flowchart of proposed method.

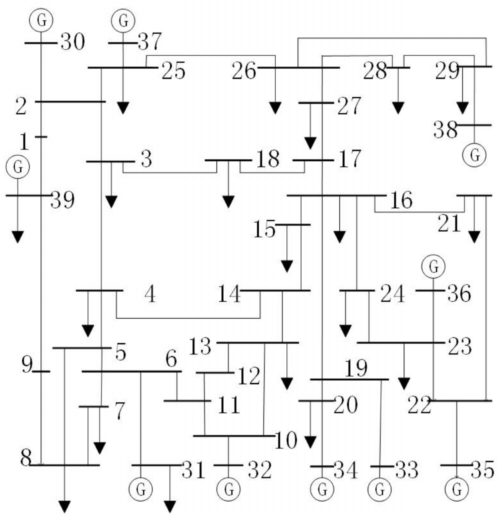



**FIGURE 5.** IEEE 39-bus test system structure.

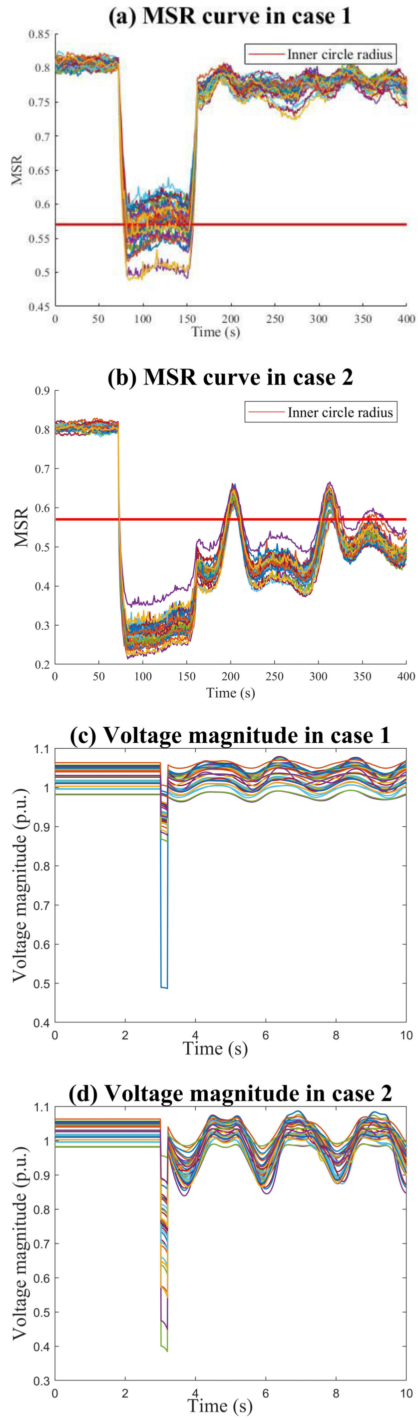



**FIGURE 6.** Voltage and MSR in case 1 and 2.

a two-phase short-circuit fault occurred at $3s$ and cleared at $3.2s$. The fault location is close to the starting point of the line, for example, $4-5$ indicates that the fault is close to bus 4. Apply the same fault near bus 1 and bus 2, the calculated MSR and voltage curves at all the other buses are shown in Figure 6. With the exception of the fault bus, the MSR and voltage curve of all the other 38 nodes in the period from the fault occurrence to the fault removal in the two kinds of faults are shown in figure 6, where case1 and case2 respectively correspond to the fault on bus1 and bus2.

From the change of MSR and the voltage response, it is easy to observe that when the fault is applied at bus 2 the system suffers more than at bus 1. After the disturbance occurred at bus 2, we can clearly see that the changes in all the remained buses are greater and even some instability issue may occur. Bus 2 should be more critical than bus 1.

Using the method proposed in paper [44], which gives each bus different weights considering the degree of impact on the grid after the fault, the vulnerability index of bus 1 and bus 2 are respectively calculated at the time period when the fault occurs:

$$p_{11} = 0.0012$$
$$p_{21} = 0.0716 \tag{18}$$

where $p_{11}$ is the vulnerability evaluation index for bus 1 and $p_{21}$ is the vulnerability evaluation index for bus 2. Applying the criteria of paper [44], the bus is more critical if the vulnerability coefficient is smaller, bus 1 is more critical compared with bus 2. The identification result does not match with the real-time system dynamic response and it caused a misjudgment.

If these two coefficients are calculated during the system recovery time period, the results are obtained:

$$p_{11} = 0.0221$$
$$p_{21} = 0.1116 \tag{19}$$

The identification result is still the same that bus 1 is more critical than bus 2, which again against the real response.

Calculating performance evaluation index $s_1$ in the same scenario.

$$s_{11} = 0.3942$$
$$s_{21} = 0.2320 \tag{20}$$

According to the identification rule proposed in section III, bus 2 is more critical than bus 1 because the value of its first index is smaller. It matches well with reality, and the voltage drop is higher when the fault occurs at bus 2 than bus 1.

The misjudgment of the previous method mainly comes from the following two aspects:1) the situation in which the absolute value is involved in the original judgment index, resulting in the misjudgment. 2) the index is constructed only according to the MSR of the corresponding matrix, which is too single. If we get the wrong results because of the statistical error, it will directly affect the final conclusion. The proposed method does not need to rely on the absolute value and the combined evaluation index also can reduce the error brought by the system chaos conditions.

In conclusion, the method proposed in this paper have two improvements compared with that in [23]: (1) The method proposed in [23] can cause misjudgment in some cases, as shown in part A of section IV, while the method proposed in this paper can avoid this mistake. (2) Index of vulnerability in [23] does not consider the degree of influence of the fault on the whole system, while this paper considers the degree of response of the whole system ($s_1$), the degree of response of single bus ($s_2$), and the degree of chaos in the system ($s_3$), thus increasing the accuracy of the calculation.

### B. CRITICAL BUSES IDENTIFICATION RESULTS AND COMPARISON

In this part, a fault is applied at each bus in the IEEE 39-bus test system. The real-time voltage data is collected every 0.02$s$ and a white noise of 0.001 times is added to the voltage data. The size of the white noise is determined by the size of the data given to simulation. Through simulations, the proper size of white noise can be gotten. When the augmented matrix is constructed, $m$ is taken as 40.

**TABLE 1.** Critical buses identification results under three fault conditions.

| Fault 1 | | Fault 2 | | Fault 3 | | Mean value | |
|---|---|---|---|---|---|---|---|
| Bus | S | Bus | S | Bus | S | Bus | E(S) |
| 16 | 19.13 | 29 | 13.37 | 17 | 18.94 | 16 | 17.00 |
| 17 | 15.31 | 16 | 13.21 | 16 | 18.68 | 17 | 15.69 |
| 19 | 13.31 | 17 | 12.83 | 19 | 18.33 | 19 | 13.65 |
| 15 | 13.09 | 3 | 11.99 | 15 | 17.48 | 3 | 13.44 |
| 2 | 12.60 | 10 | 10.74 | 20 | 17.67 | 15 | 13.33 |
| 29 | 12.57 | 2 | 10.52 | 21 | 17.00 | 20 | 13.11 |
| 6 | 11.63 | 15 | 9.43 | 3 | 16.77 | 2 | 13.00 |
| 3 | 11.56 | 19 | 9.32 | 14 | 16.14 | 29 | 12.94 |
| 10 | 10.75 | 4 | 8.45 | 10 | 15.87 | 10 | 12.45 |
| 19 | 9.63 | 20 | 9.57 | 11 | 15.01 | 14 | 12.10 |

In order to further reduce the error, three types of faults are applied: 1) two-phase short-circuit fault occurs at the end of each bus at 3$s$ and is cleared at 3.2$s$; 2) three-phase short-circuit fault occurs at the end of each bus at 2$s$ and is cleared at 2.12$s$; 3) two-phase short-circuit fault occurs at the end of each bus at 3$s$ and is cleared at 3.4$s$. The vulnerability identification index $S$ for each bus is calculated and the mean value of the index under three fault conditions is obtained and ranked. Table 1 shows the top ten critical buses under three fault conditions.

As is concluded in the previous section, If $S$ is bigger, the bus is more critical, and the system will suffer from a severer stability issue if the fault occurs at this bus. Take fault 1 as an example, bus 16 has the biggest identification index, which means it is the most critical bus under fault 1. The corresponding voltage response and calculated results are shown in Figure 7.

From the real-time voltage response, M-P law, ring law and the change of MSR, it is observed that when a fault is applied at bus 16, a voltage instability problem showed up. What's more, the system lost its voltage stability only when the fault occurred at bus 16. This is proof that the bus 16 is the most critical bus in the whole system and the proposed method is effective. Similar conclusions can also be obtained in the case of fault 2 and fault 3. Considering all three types of fault, the most critical buses in the IEEE 39-bus test system are bus 16, 17 and 19.

To further compare the results obtained from the proposed method with the existing methods, the identification results from other papers are obtained. Table 2 shows the comparison results. The ranking results go from the most critical one to the less critical ones. $M_1$ is the method proposed in this paper, and $M_2$-$M_6$ stand for identification methods based on Analytic hierarchy process, comprehensive analysis, power flow analysis, complex network, and structural evaluation.

This paper identifies bus 16, 17 and 19 as the most critical buses in the system. Taking a look back at the system structure, it is known that bus 16 and 19 are located on the important transmission channels for the power transmission between machine 33 and 34. A failure occurs at these buses would cause an imbalance of power and therefore cause

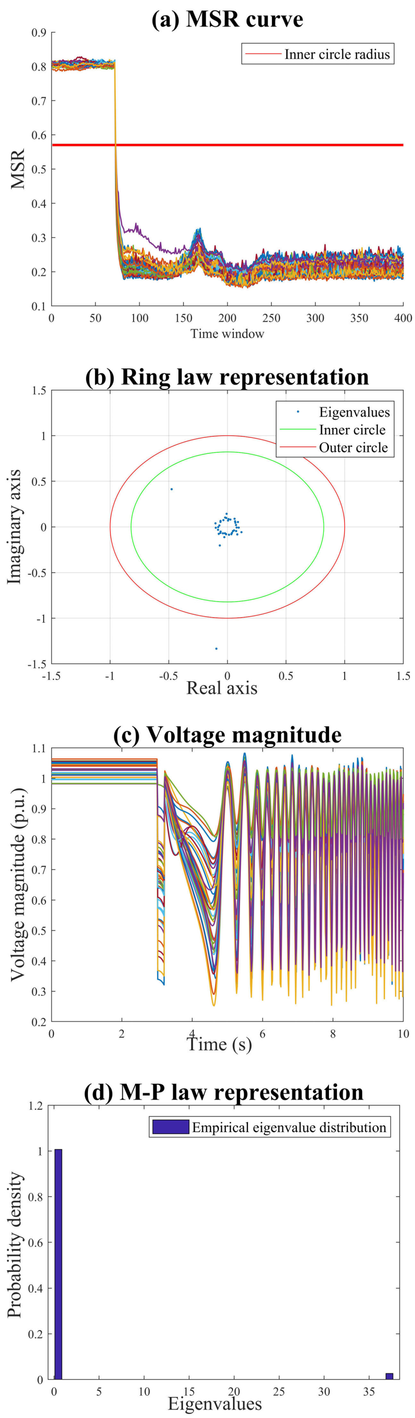



**FIGURE 7.** Corresponding results of perturbation at bus 16.

a large power shortage problem. For example, the disconnection of the branch 16-17 will cause the power of the bus 18 and bus 27 to be unbalanced, which may cause a series of problems such as transient stability problem. The disconnection of the branch 16-19 will trigger the system decoupling. From the theoretical studies, these buses also have an important position from the stability point of view, which proves that the identification results in this paper are in line with the actual operation of the system. Compared with

**TABLE 2.** Ranking results from different identification methods.

| Ranking | $M_1$ | $M_2$ | $M_3$ | $M_4$ | $M_5$ | $M_6$ |
|---|---|---|---|---|---|---|
| 1 | 16 | 16 | 29 | 8 | 16 | 12 |
| 2 | 17 | 6 | 22 | 7 | 6 | 7 |
| 3 | 19 | 17 | 21 | 5 | 5 | 8 |
| 4 | 15 | 5 | 10 | 6 | 9 | 14 |
| 5 | 3 | 2 | 6 | 13 | 17 | 13 |
| 6 | 10 | 4 | 16 | 11 | 26 | 4 |
| 7 | 20 | 14 | 3 | 14 | 19 | 5 |
| 8 | 29 | 29 | 20 | 9 | | 11 |
| 9 | 2 | 19 | 19 | 12 | | 6 |
| 10 | 14 | 10 | 2 | 10 | | 10 |

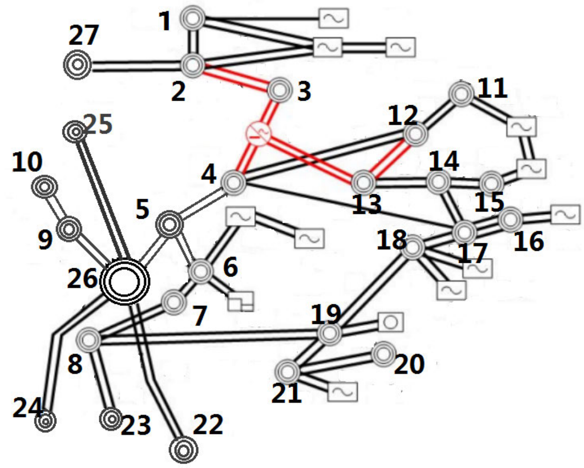



**FIGURE 8.** Real-world test system structure in China.

other methods, the obtained results are generally the same, but they are not exactly the same. This is because the focus of each method is different, so the identification results will be different. In summary, the proposed method is effective and can guarantee a reduced identification error.

### C. REAL-WORLD CASE STUDIES

We use a 500kV AC/DC hybrid power grid in a certain area of China to identify the vulnerability of each bus. The bus number is shown in Figure 8, where there are 27 buses, each of which is a substation. Both the capacity of bus 22 and 26 are 2 $\times$ 3000MVA, and there are 4 and 6 500kV lines, respectively. We use PSASP to build this model and all the calculations are done by MATLAB R2018a.

Since there are 27 buses in this model, we take $m = 25$ in this section. The same faults as part $B$ are set near each bus to identify the vulnerability. Table 3 shows the top seven critical buses under three fault conditions.

From table 3, we can detect signals based on the following analysis:

I. bus 26, 22 and 25 are at the top the table, which means that these three buses are most critical.

II. Bus 26 is the hub bus of the power grid and has six outgoing lines. From the simulation, when the fault occurs near this bus, voltage instability occurs, the DC commutation fails continuously, and the power is greatly reduced, which causes the local power angle instability. Thus bus 26 is the most critical bus.

**TABLE 3. Critical buses identification results for real-world model structure.**

| Fault 1 | | Fault 2 | | Fault 3 | | Mean value | |
|---|---|---|---|---|---|---|---|
| Bus | S | Bus | S | Bus | S | Bus | E(S) |
| 26 | 18.83 | 22 | 17.86 | 26 | 19.66 | 26 | 18.60 |
| 25 | 17.65 | 26 | 17.32 | 22 | 18.06 | 22 | 17.67 |
| 22 | 17.10 | 25 | 15.98 | 25 | 17.76 | 25 | 17.13 |
| 2 | 14.26 | 3 | 15.08 | 3 | 14.43 | 2 | 13.79 |
| 3 | 11.72 | 2 | 14.12 | 10 | 13.21 | 3 | 13.74 |
| 9 | 11.45 | 10 | 11.43 | 2 | 13.00 | 10 | 11.8 |
| 10 | 10.76 | 9 | 10.78 | 9 | 12.63 | 9 | 11.62 |

III. From the perspective of a power-on level, bus 22 receives about 40% of power, which has less power on, and weaker system voltage support; (2) From the perspective of DC response, it takes a lot of reactive power from the system during DC active power recovery after fault clearing. So that the system voltage cannot be recovered, and finally, the voltage is unstable. Thus bus 22 is critical.

In general, the simulation results are consistent with the actual situation, which verifies the validity and practicability of the proposed method.

## V. CONCLUSION AND FUTURE WORK

In this paper, big data from PMUs is used to construct a big data matrix that characterizes the grid state. Based on random matrix theory, the relationship between state deterioration and data fluctuation is analyzed. With the combination of entropy theory, a model-free vulnerability evaluation method in complex power systems is proposed.The main conclusions are as follows:

1) The data characteristics of power systems can characterize the response degree of the buses to the faults;
2) From different perspectives, M-P law based index, MSR based index and system entropy based index are constructed to identify the critical buses in power systems;
3) The data-driven method proposed can solve the problems of vulnerability assessment of complex power systems without modeling and has better accuracy than previous methods.
4) The future work may include more types of faults and increase the accuracy of the vulnerability index.

## REFERENCES

[1] X. Wang, L. T. Yang, Y. Wang, X. Liu, Q. Zhang, and M. J. Deen, "A distributed tensor-train decomposition method for cyber-physical-social services," *ACM Trans. Cyber-Phys. Syst.*, vol. 3, no. 4, pp. 1–15, Oct. 2019.

[2] X. Yin, Q. Chen, and B. Wang, "Vulnerability assessment system of complicated power system based on four-level set model," *Trans. China Electrotechn. Soc.*, vol. 28, no. 1, pp. 225–233, 2013.

[3] L. Qi, X. Zhang, W. Dou, and Q. Ni, "A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2616–2624, Nov. 2017.

[4] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, "Power system transient stability assessment based on big data and the core vector machine," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016.

[5] T. Wang, X. Yue, and X. Gu, "Comprehensive evaluation of power grid vulnerability based on hesitant fuzzy decision making method," *Power Syst. Technol.*, vol. 41, no. 7, pp. 2272–2281, 2017.

[6] Y. Liu, J. Liu, and M. Wang, "Fast assessment method for transient vulnerability of transmission lines based on kinetic energy injection betweenness," *Proc. CSEE*, vol. 31, no. 13, pp. 40–47, 2011.

[7] L. Qi, P. Dai, J. Yu, Z. Zhou, and Y. Xu, "'Time–location–frequency'–aware Internet of Things service selection based on historical records," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 1, 2017, Art. no. 1550147716688696.

[8] H. Song, G. A. Fink, and S. Jeschke, *Security and Privacy in Cyber-Physical Systems: Foundations, Principles, and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[9] M. A. Rahman, M. Y. Mukta, A. Yousuf, A. T. Asyhari, M. Z. A. Bhuiyan, and C. Y. Yaakub, "IoT based hybrid green energy driven highway lighting system," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervas. Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Aug. 2019, pp. 587–594.

[10] X. Liu, R. Zhu, B. Jalaian, and Y. Sun, "Dynamic spectrum access algorithm based on game theory in cognitive radio networks," *Mobile Netw. Appl.*, vol. 20, no. 6, pp. 817–827, Dec. 2015.

[11] Z. Guan, X. Lu, N. Wang, J. Wu, X. Du, and M. Guizani, "Towards secure and efficient energy trading in IIoT-enabled energy Internet: A blockchain approach," *Future Gener. Comput. Syst.*, to be published.

[12] J. Lin, Y. Su, Y. Cheng, C. Lu, L. Zhu, H. Huang, and Y. Liu, "A robust complex-domain state estimator using synchrophasor measurements," *Int. J. Electr. Power Energy Syst.*, vol. 96, pp. 412–421, Mar. 2018.

[13] H. Liu, H. Kou, C. Yan, and L. Qi, "Link prediction in paper citation network to construct paper correlation graph," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–12, 2019.

[14] X. Wang, L. T. Yang, H. Li, M. Lin, J. Han, and B. O. Apduhan, "NQA: A nested anti-collision algorithm for RFID systems," *ACM Trans. Embed. Comput. Syst.*, vol. 18, no. 4, pp. 1–21, Jul. 2019.

[15] E. Bompard, E. Pons, and D. Wu, "Extended topological metrics for the analysis of power grid vulnerability," *IEEE Syst. J.*, vol. 6, no. 3, pp. 481–487, Sep. 2012.

[16] B. Liu, Z. Li, X. Chen, Y. Huang, and X. Liu, "Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 65, no. 3, pp. 346–350, Mar. 2018.

[17] A. Dwivedi and X. Yu, "A maximum-flow-based complex network approach for power system vulnerability analysis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 81–88, Feb. 2013.

[18] I. Kassabalidis, M. El-Sharkawi, R. Marks, L. Moulin, and A. Alves Da Silva, "Dynamic security border identification using enhanced particle swarm optimization," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 723–729, Aug. 2002.

[19] W. Ju and Y. Li, "Identification of critical lines and nodes in power grid based on maximum flow transmission contribution degree," *Autom. Electr. Power Syst.*, vol. 36, no. 9, pp. 6–12, 2012.

[20] L. Qi, X. Zhang, S. Li, S. Wan, Y. Wen, and W. Gong, "Spatial-temporal data-driven service recommendation with privacy-preservation," *Inf. Sci.*, vol. 515, pp. 91–102, Apr. 2020.

[21] Z. Guan, X. Liu, L. Wu, J. Wu, R. Xu, J. Zhang, and Y. Li, "Cross-lingual multi-keyword rank search with semantic extension over encrypted data," *Inf. Sci.*, vol. 514, pp. 523–540, Apr. 2020.

[22] H. Song, R. Srinivasan, T. Sookoor, and S. Jeschke, *Smart Cities: Foundations, Principles, and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[23] M. A. Rahman, M. M. Hasan, A. T. Asyhari, and M. Z. A. Bhuiyan, "A 3D-collaborative wireless network: towards resilient communication for rescuing flood victims," in *Proc. IEEE 15th Int. Conf. Dependable, Autonomic Secure Comput., 15th Int. Conf. Pervas. Intell. Comput., 3rd Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Nov. 2017, pp. 385–390.

[24] S. Wan, Z. Gu, and Q. Ni, "Cognitive computing and wireless communications on the edge for healthcare service robots," *Comput. Commun.*, vol. 149, pp. 99–106, Jan. 2020.

[25] W. Gong, L. Qi, and Y. Xu, "Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–8, Apr. 2018.

[26] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Netw. Appl.*, pp. 1–13, 2019.

[27] W. Yao, J. Zhao, M. J. Till, S. You, Y. Liu, Y. Cui, and Y. Liu, "Source Location Identification of Distribution-Level Electric Network Frequency Signals at Multiple Geographic Scales," *IEEE Access*, vol. 5, pp. 11166–11175, 2017.

[28] Y. Jiang, H. Song, Y. Yang, H. Liu, M. Gu, Y. Guan, J. Sun, and L. Sha, ''Dependable model-driven development of CPS: From stateflow simulation to verified implementation,'' *ACM Trans. Cyber-Phys. Syst.*, vol. 3, no. 1, pp. 1–31, Aug. 2018.

[29] W. Wei, M. A. Rahman, I. F. Kurniawan, A. T. Asyhari, S. M. N. Sadat, and L. Yao, ''Immune genetic algorithm optimization and integration of logistics network terminal resources,'' in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Feb. 2019, pp. 435–436.

[30] R. Zhu, X. Zhang, X. Liu, W. Shu, T. Mao, and B. Jalaian, ''ERDT: Energy-efficient reliable decision transmission for intelligent cooperative spectrum sensing in industrial IoT,'' *IEEE Access*, vol. 3, pp. 2366–2378, 2015.

[31] S. Wan, M. Li, G. Liu, and C. Wang, ''Recent advances in consensus protocols for blockchain: A survey,'' in *Wireless Networks*. 2020, doi: 10.1007/s11276-019-02195-0.

[32] S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K.-K.-R. Choo, ''Multi-dimensional data indexing and range Query processing via Voronoi diagram for Internet of Things,'' *Future Gener. Comput. Syst.*, vol. 91, pp. 382–391, Feb. 2019.

[33] B. Han, L. Luo, G. Sheng, G. Li, and X. Jiang, ''Framework of random matrix theory for power system data mining in a non-Gaussian environment,'' *IEEE Access*, vol. 4, pp. 9969–9977, 2016.

[34] X. Xu, X. He, Q. Ai, and R. C. Qiu, ''A correlation analysis method for power systems based on random matrix theory,'' *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1811–1820, Jul. 2017.

[35] X. He, L. Chu, R. C. Qiu, Q. Ai, and Z. Ling, ''A novel data-driven situation awareness approach for future grids–using large random matrices for big data modeling,'' *IEEE Access*, vol. 6, pp. 13855–13865, 2018.

[36] L. Luo, B. Han, J. Chen, G. Sheng, and X. Jiang, ''Partial discharge detection and recognition in random matrix theory paradigm,'' *IEEE Access*, vol. 5, pp. 8205–8213, 2017.

[37] M. Hassan and A. Bermak, ''Robust Bayesian inference for gas identification in electronic nose applications by using random matrix theory,'' *IEEE Sensors J.*, vol. 16, no. 7, pp. 2036–2045, Apr. 2016.

[38] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao, and H. Liu, ''A big data architecture design for smart grids based on random matrix theory,'' *IEEE Trans. smart Grid*, vol. 8, no. 2, pp. 674–686, 2017.

[39] X. Wang, L. T. Yang, X. Xie, J. Jin, and M. J. Deen, ''A cloud-edge computing framework for cyber-physical-social services,'' *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 80–85, Nov. 2017.

[40] L. Qi, W. Dou, W. Wang, G. Li, H. Yu, and S. Wan, ''Dynamic mobile crowdsourcing selection for electricity load forecasting,'' *IEEE Access*, vol. 6, pp. 46926–46937, 2018.

[41] F. Ud Din, A. Ahmad, H. Ullah, A. Khan, T. Umer, and S. Wan, ''Efficient sizing and placement of distributed generators in cyber-physical power systems,'' *J. Syst. Archit.*, vol. 97, pp. 197–207, Aug. 2019.

[42] X. Wang, W. Wang, L. T. Yang, S. Liao, D. Yin, and M. J. Deen, ''A distributed HOSVD method with its incremental computation for big data in cyber-physical-social systems,'' *IEEE Trans. Comput. Soc. Syst.*, vol. 5, no. 2, pp. 481–492, Jun. 2018.

[43] B. Wang, S. Wan, X. Zhang, and K.-K. R. Choo, ''A novel index for assessing the robustness of integrated electrical network and a natural gas network,'' *IEEE Access*, vol. 6, pp. 40400–40410, 2018.

[44] W. Liu, D. Zhang, Y. Ding, Q. Wu, C. Deng, and D. Liu, ''Power grid vulnerability identification methods based on random matrix theory and entropy theory,'' *Proc. CSEE*, vol. 37, no. 20, pp. 5893–5901, 2017.

[45] L. Qi, X. Zhang, W. Dou, C. Hu, C. Yang, and J. Chen, ''A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment,'' *Future Gener. Comput. Syst.*, vol. 88, pp. 636–643, Nov. 2018.

[46] S. Wan and S. Goudos, ''Faster R-CNN for multi-class fruit detection using a robotic vision system,'' *Comput. Netw.*, vol. 168, Feb. 2020, Art. no. 107036.

[47] B. Jalaeian, R. Zhu, H. Samani, and M. Motani, ''An optimal cross-layer framework for cognitive radio network under interference temperature model,'' *IEEE Syst. J.*, vol. 10, no. 1, pp. 293–301, Mar. 2016.

[48] L. Qi, Z. Zhou, J. Yu, and Q. Liu, ''Data-sparsity tolerant Web service recommendation approach based on improved collaborative filtering,'' *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 9, pp. 2092–2099, 2017.

[49] S. Wan, X. Li, Y. Xue, W. Lin, and X. Xu, ''Efficient computation offloading for Internet of vehicles in edge computing-assisted 5g networks,'' *J. Supercomput.*, pp. 1–30, Oct. 2019.

[50] B. Wang, J. Wang, D. Liu, and S. Chen. (2019). *Research on Evaluating Vulnerability of Power Network Based on High-Dimensional Random Matrix Theory*. [Online]. Available: http://kns.cnki.net/kcms/detail/11.2107.tm.20190125.1151.008.html

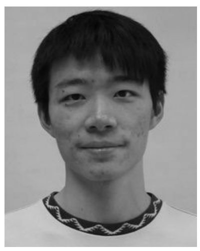
[51] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, ''Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction,'' *Complexity*, vol. 2019, pp. 1–13, Nov. 2019.

[52] J. Wang, J. Luo, X. Liu, Y. Li, S. Liu, R. Zhu, and A. Anjum, ''Improved Kalman filter based differentially private streaming data release in cognitive computing,'' *Future Gener. Comput. Syst.*, vol. 98, pp. 541–549, Sep. 2019.

[53] B. Wang, H. Ma, X. Wang, G. Deng, Y. Yang, and S. Wan, ''Vulnerability assessment method for cyber-physical system considering node heterogeneity,'' *J. Supercomput*, pp. 1–21, Oct. 2019.

**KAI DING** received the B.Eng. degree in electrical and electronic engineering and the Ph.D. degree from Wuhan University, China, in 2003 and 2007, respectively. He is currently a Senior Engineer and the President of the Smart Grid Department at Hubei and CEPRI, where he was mainly responsible for power quality and power system stability analysis and control research. His research interests include integration of renewable energy, power quality, and integrated energy systems.

**YIMIN QIAN** received the Engineering Master degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2013. He is currently an Intermediate Engineer and a Researcher at Hubei, CEPRI, and State Grid. His interest involves with generator-grid coordination, electric vehicle, renewable generation, and integrated energy systems.



**YI WANG** received the B.Eng. degree in electrical power engineering from the University of Bath, in 2014, and the Master of Science degree in power system engineering from the University College London, in 2015. He is currently an Engineer and a Researcher with the Hubei Electric Power Research Institute. His professional fields include wind and solar power generation, and power quality.

**PAN HU** received the B.Eng. degree in electrical and automation engineering and the Ph.D. degree from Wuhan University, China, in 2015 and 2018, respectively. In 2017, he became a Visiting Ph.D. Student with the University of Alberta, Canada. He is currently a Senior Engineer and a Researcher at Hubei, CEPRI, and State Grid. His interest involves with power quality issues as well as power system stability analysis, renewable generation, and integrated energy system.

**BO WANG** received the Ph.D. degree in computer science from Wuhan University, Wuhan, China, in 2006. He was a Postdoctoral Researcher with the School of Electrical Engineering, Wuhan University, from 2007 to 2009. He is currently an Associate Professor with the School of Electrical Engineering, Wuhan University. His research interests include power system online assessment, big data, integrated energy systems, and smart city.