Received November 29, 2016, accepted January 4, 2017, date of publication January 9, 2017, date of current version January 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2649841

Framework of Random Matrix Theory for Power System Data Mining in a Non-Gaussian Environment

BEI HAN, (Member, IEEE), LINGEN LUO, GEHAO SHENG, (Member, IEEE), GUOJIE Li, (Senior Member, IEEE), AND XIUCHEN JIANG

Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: L. Luo (llg523@sjtu.edu.cn)

This work was supported in part by the National High Technology Research and Development Program of China (863 Program) under Grant SS2015AA050204, in part by the National Natural Science Foundation of China under Grant 514070116, and in part by the State Grid Science and Technology under Project GWKJ-2016-180.

ABSTRACT A novel empirical data analysis methodology based on the random matrix theory (RMT) and time series analysis is proposed for the power systems. Among the ongoing research studies of big data in the power system applications, there is a strong necessity for new mathematical tools that describe and analyze big data. This paper used RMT to model the empirical data which also treated as a time series. The proposed method extends traditional RMT for applications in a non-Gaussian distribution environment. Three case studies, i.e., power equipment condition monitoring, voltage stability analysis and low-frequency oscillation detection, illustrate the potential application value of our proposed method for multi-source heterogeneous data analysis, sensitive spot awareness and fast signal detection under an unknown noise pattern. The results showed that the empirical data from a power system modeled following RMT and in a time series have high sensitivity to dynamically characterized system states as well as observability and efficiency in system analysis compared with conventional equation-based methods.

INDEX TERMS Random matrix theory, data mining, time series analysis, non-Gaussian, condition monitoring, static voltage stability, low frequency oscillation.

I. INTRODUCTION

BIG data is considered one of the most promising techniques to adapt conventional fields to the internet era [1]. The growing intelligence of power system appliances and operations has resulted in dramatically increased amounts of information and computations, arousing interest in the use of big data techniques in Smart Grid applications [2], [3]. In other words, Smart Grid analysis would be characterized by data-based models developed from system raw data rather than traditional system models that are built with hypotheses and simplifications.

Among the ongoing big data studies, researchers primarily concentrate on the 4 Vs data (data with volume, velocity, variety and veracity) collection, distributed storage and computation, visualization, etc. [4]. Regarding Smart Grid applications, there is a necessity for new data-based modeling and analysis tools for the efficient description of big data structures as well as for accurate detection of big data correlations and distributions to gain insights into power system characteristics and stability.

Random Matrix Theory (RMT) has emerged as a useful framework in wireless communication, neural networks, network science, and cognitive radio technology [5]–[8] for processing multivariate data. Traditional RMT aims to solve problems with infinite dimensions, which usually can be found with asymptotic convergence; however, recent developments in RMT concerning non-asymptotic RMT [9], [10] have perfectly resolved finite-dimensional problems. This may extend RMT to finite- dimensional engineering applications, as has been illustrated and proven by results from multiple applications in this paper. Apart from the dimensional limit of the research topics, traditionally, RMT mainly focuses on a data environment with Gaussian distribution, which would also significantly limit its applications in real power systems, where empirical data is better characterized as a non-Gaussian distribution. This paper has adopted the method of analyzing these non-Gaussian data under time series [11]–[14]. A framework of modelling power system data which treated as a time series by RMT to data mining in a non-Gaussian environment is proposed for the first time.

The potentials of RMT as well as RMT in time series analysis as a significant extension to the existing applications of RMT are studied and validated with three different applications from power systems in this paper.

The first application solved by the proposed method focuses on power line condition monitoring. Power line condition monitoring for both transmission and distribution systems provides an important measure for ensuring power system security. A significant obstacle in applying the technique of power line condition monitoring is to extract valuable information and identify condition features of the system by processing a large quantity of multi-source heterogeneous data obtained from numerous pieces of monitoring equipment in real time. The proposed random matrix and eigenvalue spectrum analysis in this paper provides a more convenient and intuitive method of assessing and determining system conditions through big data from field measurements, thereby ensuring real-time processing and instant fault judgments, even under increasing volume, velocity, and variety of data.

The second case study involves application of the new method for analyzing voltage stability, which has been considered a major factor leading to power failures. On one hand, it is increasingly challenging to compute the voltage stability margin of large power networks in real-time while maintaining an adequate accuracy with traditional stability metrics. On the other hand, taking into consideration the randomness (because of the newly adopted distributed generations, electric vehicles, combined head and power techniques, etc., that constitute more flexible power demands) in the computation of power grid voltage stability in a real-time manner can even aggravate the situation. The proposed method eliminates the procedure of constructing, reducing and solving system equations that are required for the existing methods, such as continuation power flow [15] and modal analysis [16], and provides an approach for real-time determination of voltage stability conditions,

The third application concerns low-frequency oscillation analysis with WAMS data. Because of growing system scale and operation complexity, the eigenanalysis of system model methods [17], where mathematical equations must be simplified and solved, have low accuracy and are not suitable for practical applications. Moreover, the widespread installations of WAMS provide a data platform for online recognition of the low-frequency oscillation patterns [18]. The Prony algorithm based on WAMS data is a typical method for oscillation pattern recognition [19], [20]. However, the computation complexity by higher system order for better fitting results limits its practical on-line application [21]. The third application of this paper is based on the proposed use of RMT for disturbance signal detection under an unknown noise pattern; this approach involves solving the Prony algorithm by switching problems for better field application.

The remainder of this paper is organized as follows: section 2 introduces the mathematical concepts and models used in this paper, including the M-P law, the ring law and corresponding extensions in time series. The method of problem formation through real-time empirical data is also developed at the end of the section. Three case studies indicating the potential applications of RMT in power systems are given in sections 3, 4 and 5 concerning power line condition monitoring, voltage stability analysis and low frequency oscillation detection, respectively. Section 6 highlights the key contributions and section 7 draws the conclusions.

II. RMT AS A BIG DATA MODELING TOOL

A. BASICS OF RMT

When the dimensions of a random matrix are sufficiently large, the empirical spectral distribution (ESD) of its eigenvalues always converge to their theoretical limits [9]. Note that although the asymptotic convergence in RMT is considered under infinite dimensions, the asymptotic results are remarkably accurate for relatively moderate matrix sizes. This is the main intention of using RMT as a practical solution to real world engineering problems. Firstly, some known results, i.e., the M-P law and ring law from RMT are summarized as flowing which also presented in our previous work [22]:

1) MARCHENKO-PASTUR LAW (M-P LAW)

Let $X = \{x_{ij}\}_{1\le i\le N,1\le j\le T}$ be a random $N \times T$ matrix whose entries with the mean $\mu(x) = 0$ and the variance $\sigma^2(x) < \infty$ are independent identically distributed (i.i.d.). N is an integer such that $N/T = c \in (0, 1]$. Thus, the ESD of the corresponding sample covariance matrix $S = 1/N(XX^H)$ converges to the M-P law with the following distribution density function [9]:

$$f_{MP}(x) = \begin{cases} \dfrac{\sqrt{(b-x)(x-a)}}{2\pi x c\sigma^2}, & a \le x \le b \\ 0, & otherwise \end{cases} \quad (1)$$

where $a = \sigma^2(1-\sqrt{c})^2$, $b = \sigma^2(1+\sqrt{c})^2$.

The M-P law proves that there exists an asymptotic behavior of the eigenvalues distribution in large random matrices.

2) THE RING LAW

Consider the matrix product $Z = \sum_{i=1}^{\alpha} \tilde{X}_i$, where $\tilde{X}_i \in \mathbb{C}^{N\times N}$ is the singular value equivalent [10] of rectangular $N \times T$ non-Hermitian random matrixX_i, whose entries are i.i.d. variables with the mean $\mu(x) = 0$ and the variance $\sigma^2(x) = 1$. Then the empirical eigenvalue distribution of $\mathbf{Z}$ converges almost surely to the limit given by:

$$f_Z(z) = \begin{cases} \dfrac{1}{\pi x\alpha} |z|^{\frac{2}{\alpha}-2}, & (1-c)^{\frac{\alpha}{2}} \le |z| \le 1 \\ 0 & , elsewhere \end{cases} \quad (2)$$

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 the outer circle radius is unity. In addition, the singular value equivalent matrix $\tilde{X}$ is calculated by:

$$\widetilde{X} = \sqrt{XX^H}U \quad (3)$$

where $U \in \mathbb{C}^{N\times N}$ is a Haar Unitary matrix.

The ring law extends RMT to large non-Hermitian random matrices. As shown in [23], [24], the assemble method of $\boldsymbol{Z}$ by product operation allows us to study the streaming datasets both space and time.

B. RMT IN TIME SERIES

When applying RMT to power system data mining, a significant difficulty is that the empirical data may not follow a Gaussian distribution. Furthermore, the measured filed data may be correlated in space or time. Therefore, in this paper we would model power system data as a time series first. Then, based on the theoretical results for extending RMT in time series instead of pure Gaussian constraints [25], [26], the algorithm of power system data modelling as a random matrix in time series aspect is proposed.

Considering a stationary ARMA(p,q) equation of a time series first [27]:

$$\phi(B)y_t = \theta(B)\varepsilon_t \tag{4}$$

where $\{y_t : t = 0, \pm 1, \ldots\}$ is a real variables sequence, $\{\varepsilon_t : t = 0, \pm 1, \ldots\}$ denotes a white noise vector with zero mean and σ^2 variance, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B + \ldots + \theta_p B^p$ are real polynomials in B, which is a backshift operator $B^j y_t = y_{t-j}, j = 0, 1, \ldots$.

Let $X_1 = (X_{11}, X_{21}, \ldots X_{T1}), \ldots X_N = (X_{1N}, X_{2N}, \ldots X_{TN})$ are N independent copies of $y = (y_1, y_1, \ldots y_T)$, we have:

$$X = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{T1} \\ X_{12} & X_{22} & \cdots & X_{T2} \\ \vdots & \vdots & \ldots & \vdots \\ X_{1N} & X_{2N} & \cdots & X_{TN} \end{pmatrix}_{N\times T} = (X_1, X_2, \ldots, X_{\mathrm{N}})' \tag{5}$$

Keeping $c = N/T \in (0, 1]$, then the corresponding ESD of $S = 1/N(XX^H)$ tends to a probability distribution F whose Stieltjes transform satisfies the equation [25]:

$$z = -\frac{1}{s} + \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cs + \{2\pi f(\lambda)\}^{-1}} d\lambda \tag{6}$$

where $s_F(z) = \int \frac{1}{x-z} F(dx), z \in \mathbb{C}^+$, and $f(\lambda)$ is the spectral density of the ARMA(p,q) model:

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left|\frac{\theta(e^{-i\lambda})}{\phi(e^{-i\lambda})}\right|^2, \quad \lambda \in [0, 2\pi) \tag{7}$$

Equation (6) gives an implicit solution of F which could be solved easily in some special cases, i.e., AR(1) and MA(1) [25]. For more complicated cases, a numerical solution is proposed in [28] to compute F which is depicted in the following:

We can write (6) as:

$$s = \frac{1}{-z + A(s(z))} \tag{8}$$

where $A(s(z)) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{cs+\{2\pi f(\lambda)\}^{-1}} d\lambda$.

For a given real x, let ε be a small enough positive value, and set $z = x + i\varepsilon$. Choose an initial value $s_0(z) = u + i\varepsilon$ and iterate for $k \geq 0$ according to the iterative equation:

$$s_{k+1}(z) = \{-z + A(s_k(z))\}^{-1}, \tag{9}$$

until convergence, and then let $s_k(z)$ be the final value.

Thus, we have the density function $f_T(x)$:

$$f_T(x) = \frac{1}{\pi}\Im s_k(z) \tag{10}$$

As addressed above, the power system empirical data is assembled by RMT from a time series point of view which extends RMT from pure Gaussian to a non-Gaussian context.

C. PROBLEM FORMING FROM EMPIRICAL DATA

Most importantly, a feasible matrix structure for data description and analysis in a power system must be constructed considering a big data environment. This paper presents an improvement of the method introduced in [23] and [24], fitting RMT for heterogeneous data both from a multi-source and in a time series.

For multi-source data, assuming n types of measurable variables (which may be heterogeneous) are sampled at time t_i, the collected data can be built as a column vector $x(t_i) = (x_1, x_2, \ldots, x_n)^H$. For single source data, we could also form the matrix following the assembly method introduced by (5).

In addition, a sliding split time window is used for the raw data intercept to meet the requirements for real-time data processing in power system applications. Specifically, a sliding window is used to truncates the measured data into vectors continuously whose length is $N \times T$. Each vector then will be assembled as a matrix $X \in \mathbb{C}^{N\times T}$ for further analysis. Therefore, T is denoted as the sliding offset and can be used to adjust the matrix size ($c = N/T \in (0, 1]$) which is important for real-time processing requirements.

To fulfill the preconditions of RMT, a normalization operation should be performed to get the normalized non-Hermitian matrix $\overline{X} \in \mathbb{C}^{N\times T}$:

$$\overline{X}_{i,j} = \frac{X_{i,j} - MEAN(X_i)}{stdDEV(X_i)} \tag{11}$$

where $MEAN(X_i)$ is the mean value of vector X_i, $stdDEV(X_i)$ is the standard deviation of vector X_i.

Therefore, a general procedure for empirical data processing can be summarized as follows:

Step 1: Obtain raw data matrix $\boldsymbol{X}$ according to the multi-source or single-source scenario.

Step 2: Calculate $\overline{X}$ by (11) to convert original matrix $\boldsymbol{X}$ to a standard non-Hermitian matrix.

Step 3: Obtain the singular value equivalent matrix $\widetilde{X}$ through $\overline{X}$ by (3).

Step 4: Calculate the sample covariance matrix $\boldsymbol{S}$ and the matrix product $\boldsymbol{Z}$.

Step 5: Calculate $f_{MP}(x)$ of $\boldsymbol{S}$ according to (1).

Step 6: Calculate the eigenvalue of $\boldsymbol{Z}$ and the inner circle radius according to (2).

Step 7: For the time series, obtain the parameters $\phi(B)$ and $\theta(B)$ of ARMA(p, q) using the MATLAB toolbox.

Step 8: Calculate $f_T(x)$ according to the proposed numerical method in section II.B.

In this paper, we set α in (2) as 1 for simplicity. All the data in the three case studies introduced later are analyzed according to this procedure.

III. RMT APPLICATION IN CONDITION MONITORING

In power system applications of condition monitoring and diagnosis targeting for power equipment, data from multiple monitoring devices are always analyzed separately [29]. With the extensive installation of monitoring devices, such as PMUs and AMI, a large volume and variety of data concerning system conditions are now available quite easily. Thus, the required feature identification and analysis from these multi-source heterogeneous data starts to play a more important role in the whole process of system diagnosis.

We propose a method based on RMT for analyzing fundamental voltage and harmonics from PMUs. This method is first implemented with eight PMUs sampling bus voltages every 15 minutes alone for each of the three monitored power transmission lines. The mean and standard deviation of the eight PMUs are used to evaluate the operation state for each line.

For example, the fundamental voltage and the third harmonic waveforms monitored from phase A of the three lines are shown in Fig. 1, where the curves denote the mean values and the shadowed areas denote the standard deviations of the corresponding curves. These time-domain waveforms can display the empirical data of each power line intuitively. Even some immediate comparisons and judgments can be made through observations. However, it is difficult to analyze and extract more useful hidden information of the data, such as power quality of the line from time-domain observations.

In Fig. 1 (a), the fundamental voltage average value of power line 1 is much smoother and closer to the nominal voltage than those of lines 2 and 3, which can always lead to an immediate judgment that line 1 has a more stable voltage waveform (better power quality). In addition, in Fig. 1 (a), the voltage variance of line 3 is larger than that of line 2 in the first 500 minutes; however, the situation is reversed afterwards, which leads to an implicit obstacle to determine which line has better power quality, even with more detailed reference to the third harmonics from Fig. 1 (b).

A traditional monitoring and analysis method in the time domain is unable to perform the task of power quality evaluation given intricate situations with large volume multi-source heterogeneous data.

Following the method introduced in section II.C, we combine the fundamental voltage and harmonic data from above as a random matrix $\boldsymbol{X}$, and its corresponding matrix $\boldsymbol{Z}$ is also obtained. The eigenvalues of matrix $\boldsymbol{Z}$ are shown in Fig. 2 (upper part of each line), along with a defined theoretical ring between the inner circle (green) and the outer circle (blue).

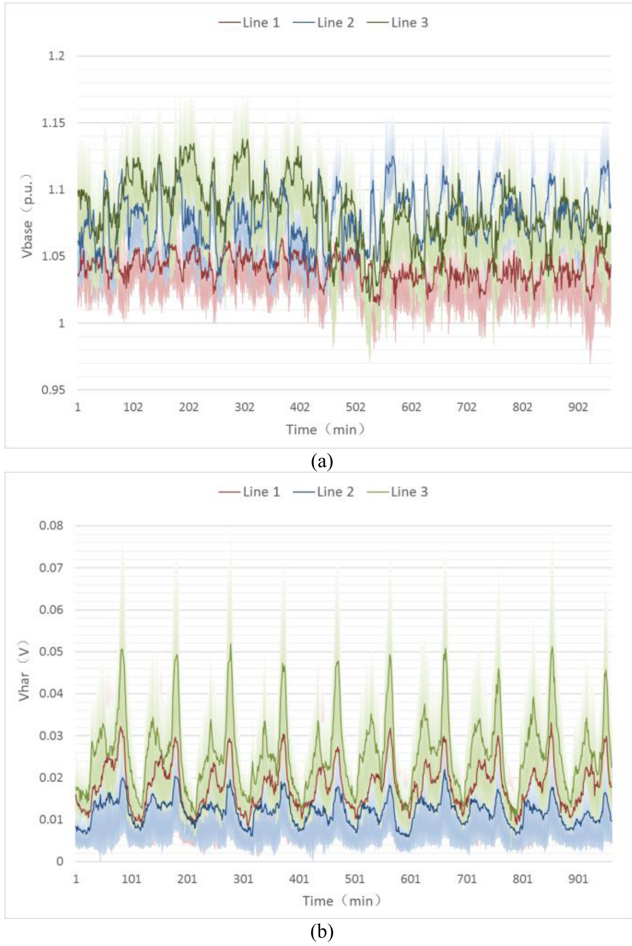

FIGURE 1. (a) The fundamental voltage waveforms of phase A and (b) the third harmonic waveforms of phase A for three power transmission lines.

It can be observed that the eigenvalues of the assembled random matrix $\boldsymbol{Z}$ are generally outside of the theoretical ring, and most of them are within the inner circle. According to the ring law, we can infer that the corresponding data forming matrix $\boldsymbol{X}$ do not follow i.i.d.; in other words, the data indicate valid signals of power quality rather than Gaussian noise [8]. This situation coincides with the fact that the monitored power line fundamental voltage and harmonic data forming $\boldsymbol{X}$ are time or/and space correlated, which show different features in RMT based data analysis with Gaussian noise.

Correspondingly, it can be found in Fig. 2 (bottom part of each line) that the eigenvalue distribution of the sample covariance matrix of $\boldsymbol{X}$ (denoted as $\boldsymbol{S}$) is not consistent with the theoretical M-P law (blue curves) for each line. Some eigenvalues are outside of the region of M-P law, which also denote valid signals present in the measured data rather than Gaussian noise [8].

Furthermore, it is obvious that the eigenvalues are more sparsely distributed for line 3 than for the other two lines. Note that the eigenvalues tend to be concentrated (especially for lines 1 and 2) in a minimal circle with a varying radius, which is much larger for line 3 than for the other two lines.

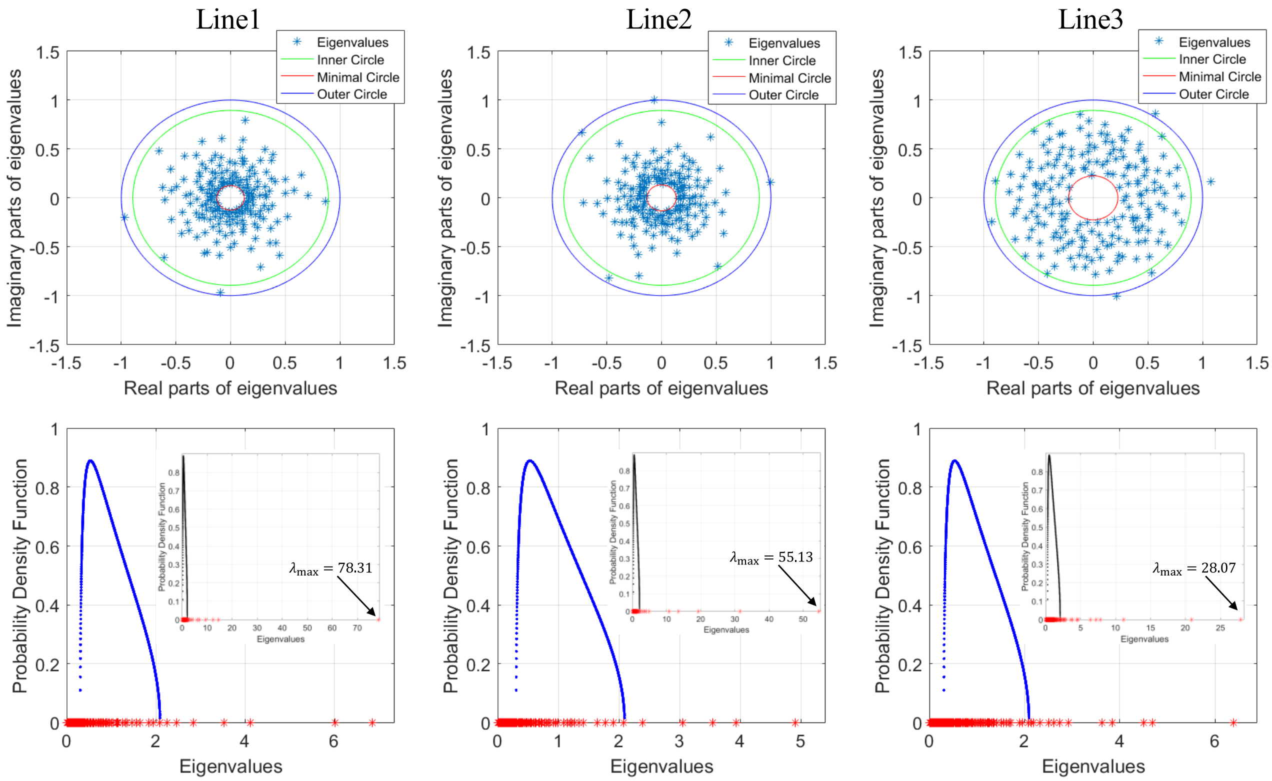

FIGURE 2. **Eigenvalue distribution of matrix *Z* assembled by fundamental voltage and third harmonics data (top), and the eigenvalue distribution of matrix *S* and corresponding M-P law in the three lines (bottom).**

In addition, the mean spectral radius (MSR) [23] metrics could be used in quantifying the sparseness and the radius of the eigenvalue distribution in our analysis, which would obviously show the same phenomena. From these features indicated by the RMT analysis method, we can distinguish line 3 from lines 1 and 2, which is in agreement with the fact that line 3 is an industrial power line with a more complicated demand environment, whereas lines 1 and 2 are ordinary residential power lines. This information is impossible to obtain through traditional time-domain data analysis.

To conclude, valid information implicated by a large volume of heterogeneous data can be easily extracted from Gaussian noise using the RMT method, and the varying densities of the eigenvalue distributions and the radii of the minimal circles can be used in analyzing and displaying different features hidden in the multi-source heterogeneous data.

IV. RMT APPLICATION IN STATIC VOLTAGE STABILITY AWARENESS

In this section, we develop RMT for extracting more intrinsic power system features besides data expression and observations, with the application of critical point detection of static voltage stability.

The continuation power flow is one of the most important methods to find a complete PV curve starting at some base load and leading to the steady state voltage stability limit (critical point) [15]. Power flow equations considered with uniformly increasing loads are illustrated as follows:

$$\begin{cases} P_{is} - V_i \sum_{j=1}^{n} V_j (G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}) - \lambda P_{id} = 0 \\ Q_{is} - V_i \sum_{j=1}^{n} V_j (G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}) - \lambda Q_{id} = 0 \end{cases} \tag{12}$$

where V and θ are the voltage value and angle, respectively, P_{is}, Q_{is} and P_{id}, Q_{id} indicate initial and increasing active/reactive power vectors of bus i, respectively, and $\lambda(0 \leq \lambda \leq \lambda_m)$ is used as the load growth factor [15].

In this case study, the IEEE 9-bus system, shown in appendix A, is used to compare the awareness of the critical point between the PV curve at bus 9 and the eigenvalue distributions evolution of bus 9 voltage with increasing λ. The simulation is based on the continuation method and performed by the MATPOWER toolbox [30]. For observing the eigenvalue distribution of bus 9 voltage at each load factor λ, we change λ with a normal distribution of zero mean and 5% variance and perform one million times power flow to obtain bus 9 voltage datasets.

Following the procedure presented in section II.C, the simulated bus 9 voltages at each λ are assembled as a 500×2000 dimensional matrix $\boldsymbol{X}$. The eigenvalue distribution of matrix $\boldsymbol{Z}$ is reported in the left part of Figs. 3 (a), (b) and (c).

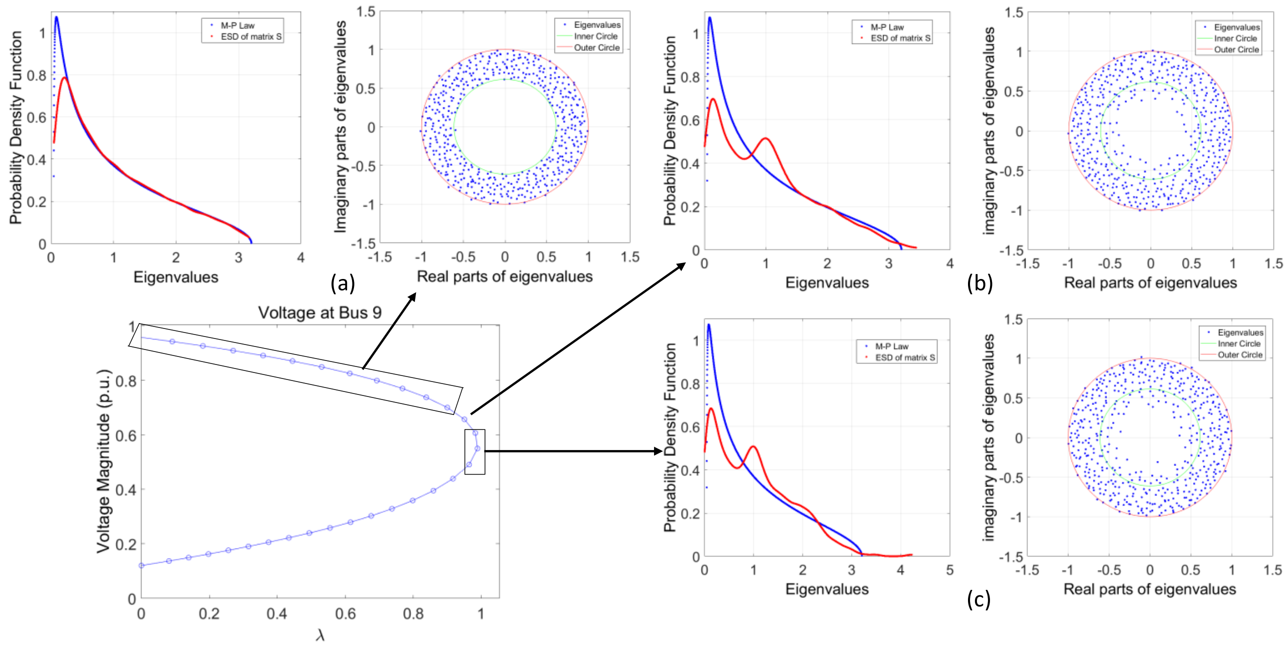

FIGURE 3. PV curve of bus 9 in IEEE 9-bus system and its evolution of eigenvalue distribution.

The ESD of the sample covariance matrix $S = 1/500(XX^H)$, and accordingly, $f_{MP}(x)$ is calculated and reported in the right part of Figs. 3 (a), (b) and (c).

From observing and comparing the eigenvalue distributions during normal operation in Fig. 3 (a) and at/near the critical point in Figs. 3 (b) and (c) by RTM, we can see that, during normal operations, the eigenvalue distributions of matrix $\boldsymbol{S}$ coincide with the M-P law, and the eigenvalues of matrix $\boldsymbol{Z}$ are basically distributed in the theoretical ring with a relatively large radius; in contrast, at/near the critical point, eigenvalue distributions of matrix $\boldsymbol{S}$ fluctuate apart from the M-P law, and the eigenvalues of matrix $\boldsymbol{Z}$ escape from the theoretical ring and into the inner circle.

In other words, the consistencies of the eigenvalue distributions with both the M-P law and the ring law can reveal the status of voltage stability. The distributions of eigenvalues following RMT are highly sensitive to the system voltage bifurcation point, which provides strong evidence of the effectiveness of the use of RMT-based big data modeling for the detection of intrinsic system characteristics.

Compared with traditional complete system modeling and continuation power flow computations for achieving the stability limits, RTM-based big data modeling can fulfill the task of voltage stability critical point detection more intuitively and immediately using the measured voltage data. During real-time operation, each bus of the system can be characterized with a PDF and theoretical ring plots of eigenvalues from the corresponding matrices, which can be directly observed for making judgments on emergencies. Furthermore, with the latter method, only measured voltage data at the related bus are required as inputs for the RMT models, instead of detailed structures and parameters of system components (lines, buses, loads, etc.) for complete system modeling and power flow calculations. The RMT based data analysis in power system can thus provide a new method involving the flexible and efficient use of massive information collected through PMUs.

V. RMT APPLICATION IN LOW-FREQUENCY OSCILLATION DETECTION UNDER UNKNOWN NOISE PATTERN

Some adverse effects, e.g., noise and system order, greatly restrain the practical applications of the Prony method. On the one hand, WAMS data measured throughout the grid are always subject to random load changes or system operations at various locations of the network known as ambient data, acting like colored noise signals [31]. On the other hand, research studies have been performed to reduce the system order to accelerate computation speed. Moreover, as addressed in [21], a sliding data window analysis method has been proposed to achieve an on-line application. This study found that the efficiency of WAMS data processing could be improved further if the starting time of disturbance could be detected rapidly because no Prony algorithm is required during the data window without disturbance. Therefore, there is a requirement for fast detection of disturbances apart from the ambient data.

To tackle the problem of smooth switching of algorithms, we propose a fast disturbance detection method based on RMT under ambient data characterized as colored noise with an unknown statistical pattern. According to the RMT, if there is no (disturbance) signal, then the ESD of the sample covariance matrix ($\boldsymbol{S}$) follows the M-P law [8]. Correspondingly, whenever the ESD of matrix $\boldsymbol{S}$ departs from the M-P law, (disturbance) signal is implied by the data. In our case, the noise is modeled by ARMA, and the M-P law need to be extended in time series as introduced in section II.B.

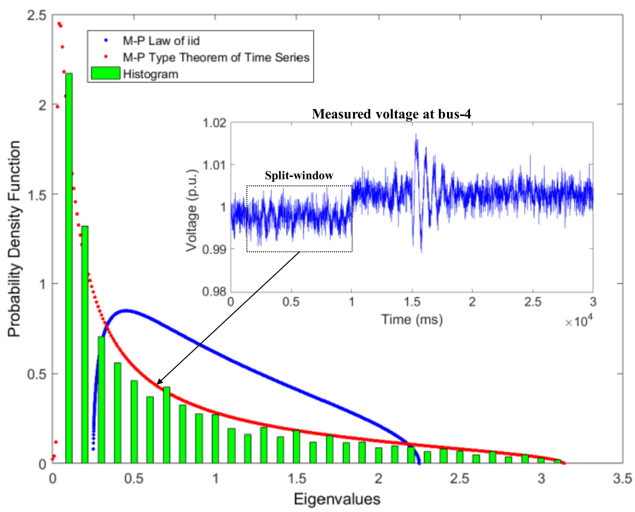

FIGURE 4. Comparative M-P law and M-P law in time series for the ESD of measured voltage at bus 4.

In this case study, the PMU measured data are obtained by time domain simulations of a 14-bus system in PSAT [32], as presented in Appendix B. The whole simulation time is set as 30 seconds, and the integration time step is set as 0.001 second. The disturbance is designed as a three-phase ground fault on bus 9 starting at 15 seconds that is removed 0.1 seconds later. We choose bus 4 voltage described in the time domain as the ambient data which is illustrated in the subgraph of Fig. 4 together with the disturbance signal for analysis.

For forming the time series data as matrix $\boldsymbol{X}$, the sliding offset unit T is set as 200 ms and the split-window length is set to be 10000 ms. For example, the first 10000 data is assembled to a 50 × 200 matrix by (5) which denoted as $\boldsymbol{X}$. The ESD of its covariance matrix $S = 1/50(XX^H)$ could be obtained and reported in Fig. 4 as the green bars. We can observe that the ESD of matrix $\boldsymbol{S}$ is obviously different with the M-P law (blue curve), whereas the RMT in time series by (6) (red curve) could generally match the ESD with some deviations due to the finite sampling data size (theoretically infinite data would provide a perfect match).

Assuming in the sliding data window based Prony method, the split window gradually moves to the right with the 10000 ms window length and 200 ms offset unit T. Along with the window moving, a disturbance can be found in the original voltage data as shown in the subgraph of Fig.4. By our proposed method, the disturbance in non-Gaussian noise (modeled by ARMA) can be detected instantly with the presence of the maximum eigenvalue of matrix $\boldsymbol{S}$. Fig. 5 illustrates the linkage between the maximum eigenvalue (red dots in black curve) and disturbance (blue curve indicated by the arrow): the disturbance appears at 15 seconds

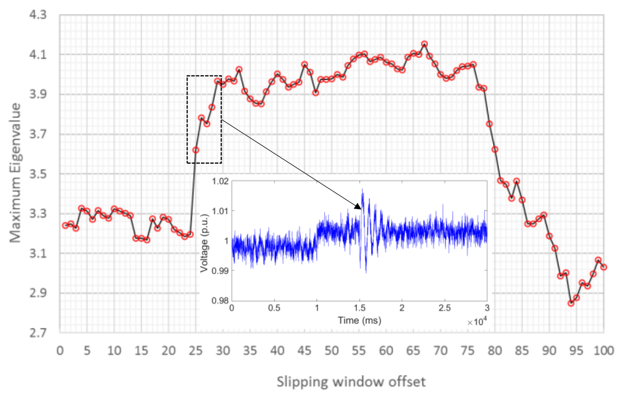

FIGURE 5. The maximum eigenvalue based disturbance detection according to the sliding split window.

(5 s after sliding the window) in the time-domain voltage data. In other words, for the starting time of the disturbance to be detected rapidly, we just need perform an eigenvalue calculation of a 50 × 200 matrix which is an easy task for existing electronic equipment.

With this application, the RMT-based data modeling approach has shown the potential to address the case of non-Gaussian noise. This would extend the applicable range of RMT in power systems with data obtained from field tests.

VI. SUMMARY OF THE EVALUATIONS

In this paper, we found that RMT-based data modeling together with the related theoretical deductions and conclusions can provide an effective analysis method for power systems with large volume multi-source heterogeneous data. With the results of the three case studies, the following contributions and potentials of the RMT-based method are highlighted.

First, with its specialty of modeling infinite-dimensional matrices, the RMT-based method can obtain a simple structure for representing multi-source heterogeneous data and thus provide a novel representation and analysis scheme through eigenvalue distributions of matrices.

Second, by considering eigenvalues of data models under M-P law and ring law and analyzing the correlations between data models and system states, the method can provide in-depth system analyses, such as power system voltage stability analysis, without solving equations.

Third, the RMT-based data models are able to dynamically characterize real-time system states with high sensitivity. Because the data models can be updated and comparatively analyzed under RMT without high-order equations, real-time analysis of large systems is possible. Thus, RMT can provide efficient processing of continuously measured data in power systems and be aware of emergent faults or sudden changes of the system states.

Fourth, with the extensions of RMT in time series, the empirical data models can also be analyzed under unknown

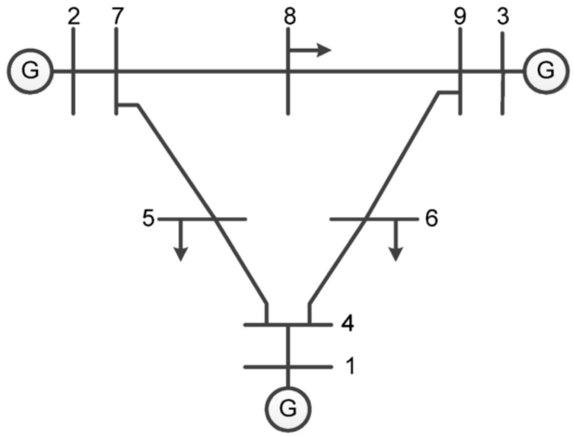

FIGURE 6. IEEE 9-bus system used in section IV.

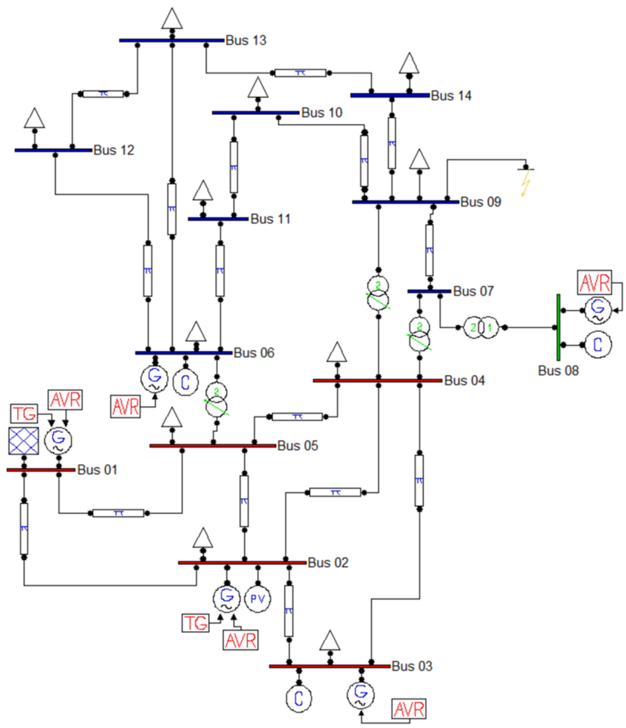

FIGURE 7. A 14-bus system in PSAT used in section V.

noise patterns and thus be more applicable for power system analyses, which are containing not only environmental noise but also small fluctuations from non-local system operations.

Finally, the representing and analyzing structures of RMT-based models can be implemented with parallel computing, providing high potentials for use of RMT in engineering applications for large power system analysis in a big data environment.

VII. CONCLUSIONS

To conclude, the RMT based data mining method could be efficiently applied to complicated fields, including the power system, by making sufficient use of the historical and multi-variate data, instead of or in parallel with the traditional analytical methods. In this paper, we conducted initial attempts in three power system applications, concerning condition monitoring, voltage stability awareness, and low-frequency oscillation detection. And with the three cases, an improved RMT model to be applied in time series and non-Gaussian noise environment is validated, revealing the flexibility of our proposed method.

APPENDIX A

See Fig. 6.

APPENDIX B

See Fig. 7.

REFERENCES

[1] Science. (Feb. 2011). *Special Online Collection: Dealing With Data*. [Online]. Available: http://www.sciencemag.org/site/special/data/

[2] IBM. (Jul. 1, 2015). *The Four V's of Big Data*. [Online]. Available: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

[3] H. Jiang, K. Wang, Y. Wang, M. Gao, and Y. Zhang, "Energy big data: A survey," *IEEE Access*, vol. 4, pp. 3844–3861, 2016.

[4] P. Zhang, X. Wu, and J. He, "Review on big data technology applied in active distribution network," *Electr. Power Construction*, vol. 36, no. 1, pp. 52–59, 2015.

[5] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, Jun. 2009.

[6] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Trans. Emerg. Telecommun. Technol.*, vol. 10, no. 6, pp. 585–595, 1999.

[7] N. Asendorf and R. R. Nadakuditi, "The performance of a matched subspace detector that uses subspaces estimated from finite, noisy, training data," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1972–1985, Apr. 2013.

[8] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[9] T. Tao, *Topics in Random Matrix Theory*. Providence, RI, USA: AMS, 2012.

[10] R. C. Qiu, Z. Hu, H. Li, and M. C. Wicks, *Cognitive Radio Communications and Networking: Principles and Practice*. Hoboken, NJ, USA: Wiley, 2012.

[11] J. Y. Fan, T. H. Ortmeyer, and R. Mukundan, "Power system stability improvement with multivariable self-tuning control," *IEEE Trans. Power Syst.*, vol. 5, no. 1, pp. 227–234, Feb. 1990.

[12] H. V. Haghi, M. T. Bina, and M. A. Golkar, "Nonlinear modeling of temporal wind power variations," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 838–848, Oct. 2013.

[13] V. S. Perić and L. Vanfretti, "Power-system ambient-mode estimation considering spectral load properties," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1133–1143, May 2014.

[14] V. Zvaritch and E. Glazkova, "Application of linear AR and ARMA processes for simulation of power equipment diagnostic systems information signals," in *Proc. Comput. Problems Elect. Eng. (CPEE)*, Sep. 2015, pp. 259–261.

[15] V. Ajjarapu and C. Christy, "The continuation power flow: A tool for steady state voltage stability analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 416–423, Feb. 1992.

[16] B. Gao, G. K. Morison, and P. Kundur, "Voltage stability evaluation using modal analysis," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1529–1542, Nov. 1992.

[17] D. J. Trudnowski and J. E. Dagle, "Effects of generator and static-load nonlinearities on electromechanical oscillations," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1283–1289, Aug. 1997.

[18] J. F. Hauer, D. J. Trudnowski, and J. G. DeSteese, "A perspective on WAMS analysis tools for tracking of oscillatory dynamics," in *Proc. IEEE Power Eng. Soc. General Meeting*, Jun. 2007, pp. 1–10.

[19] D. P. Wadduwage, U. D. Annakkage, and K. Narendra, ''Identification of dominant low-frequency modes in ring-down oscillations using multiple Prony models,'' *IET Generat., Transmiss. Distrib.*, vol. 9, no. 15, pp. 2206–2214, 2015.

[20] S. Rai, D. Lalani, S. K. Nayak, T. Jacob, and P. Tripathy, ''Estimation of low-frequency modes in power system using robust modified Prony,'' *IET Generat., Transmiss. Distrib.*, vol. 10, no. 6, pp. 1401–1409, 2016.

[21] J. Xiao, X. Xie, Y. Han, and J. Wu, ''Dynamic tracking of low-frequency oscillations with improved Prony method in wide-area measurement system,'' in *Proc. IEEE Power Eng. Soc. General Meeting*, Jun. 2004, pp. 1104–1109.

[22] L. Luo, B. Han, J. Chen, G. Sheng and X. Jiang, ''Partial discharge detection and recognition in random matrix theory paradigm,'' *IEEE Access*, Dec. 2016. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7763871

[23] 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*, Jan. 2015. [Online]. Available: http://arxiv.org/pdf/1501.07329.pdf

[24] X. Xu, X. He, Q. Ai, and C. Qiu, ''A correlation analysis method for power systems based on random matrix theory,'' *IEEE Trans. Smart Grid*, Jun. 2015. [Online]. Available: http://arxiv.org/pdf/1506.04854.pdf

[25] C. Wang, B. Jin, and B. Miao, ''On limiting spectral distribution of large sample covariance matrices by VARMA(p,q),'' *J. Time Ser. Anal.*, vol. 32, no. 5, pp. 539–546, 2011.

[26] M. Banna and F. Merlevède, ''Limiting spectral distribution of large sample covariance matrices associated with a class of stationary processes,'' *J. Theor. Probab.*, vol. 28, no. 2, pp. 745–783, 2013.

[27] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.

[28] J. Yao, ''A note on a Marčenko–Pastur type theorem for time series,'' *Statist. Probab. Lett.*, vol. 82, no. 1, pp. 22–28, 2012.

[29] Y. Han and Y. H. Song, ''Condition monitoring techniques for electrical equipment: A literature survey,'' *IEEE Power Eng. Rev.*, vol. 22, no. 9, p. 59, Sep. 2002.

[30] R. D. Zimmerman and C. Murillo-Sánchez. *MATPOWER User's Manual, Version 4.1*. [Online]. Available: http://www.pserc.cornell.edu/matpower/docs

[31] G. K. Pai and J. W. Pierre, ''A real-time scheme for validation of an auto-regressive time series model for power system ambient inter-area mode estimation,'' in *Proc. IEEE Hawaii Int. Conf. Syst. Sci.*, Jan. 2014, pp. 2436–2443.

[32] F. Milano, ''An open source power system analysis toolbox,'' *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug. 2005.

BEI HAN (M'13) received the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and the Ph.D. degree in electrical engineering from the Politecnico di Torino, Italy, in 2013.

She is currently an Assistant Professor with Shanghai Jiao Tong University. Her research interests are complex power system modeling, data mining, and microgrid.

LINGEN LUO received the M.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008, and the Ph.D. degree in electrical engineering from the Politecnico di Torino, Italy, in 2014.

He is currently a Staff Researcher with Shanghai Jiao Tong University. His research interests include the condition monitoring of power apparatus, vulnerability analysis of power systems.

GEHAO SHENG (M'08) was born in Hunan, China, in 1974. He received the B.E., M.S., and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1996, 1999, and 2003, respectively, all in electric power system and automation engineering.

From 2003 to 2005, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, where he is currently an Associate Professor with the Department of Electrical Engineering. His research interests include the condition monitoring of power apparatus.

GUOJIE LI (M'09-SM'12) received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1989 and 1993, respectively, and the Ph.D. degree from the School of EEE, Nanyang Technological University, Singapore, in 1999.

He was an Associate Professor with the Department of Electrical Engineering, Tsinghua University, Beijing. He is currently a Professor with the Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include power system analysis and control, wind and PV power control and integration.

XIUCHEN JIANG was born in Shandong, China, in 1965. He received the B.E. degree in high voltage and insulation technology from Shanghai Jiaotong University, Shanghai, China, in 1987, the M.S. degree in high voltage and insulation technology from Tsinghua University, Beijing, China, in 1992, and the Ph.D. degree in electric power system and automation from Shanghai Jiao Tong University in 2001.

He is currently a Professor with the Department of Electrical Engineering, Shanghai Jiao Tong University. His research interests include electrical equipment online monitoring, condition-based maintenance, and automation.