
Early Anomaly Detection for Power Systems Based on Kullback-Leibler Divergence Using Factor Model Analysis

Qing Feng, Ghadir Radman, Xuebin Li

Department of Electrical and Computer Engineering, Tennessee Technological University, Cookeville, USA

Email address:

qfeng@tntech.edu (Qing Feng), gradman@tntech.edu (G. Radman), xli44@tntech.edu (Xuebin Li)

To cite this article:

Qing Feng, Ghadir Radman, Xuebin Li. Early Anomaly Detection for Power Systems Based on Kullback-Leibler Divergence Using Factor Model Analysis. *American Journal of Electrical Power and Energy Systems*. Vol. 10, No. 4, 2021, pp. 60-73.

doi: 10.11648/j.epes.20211004.12

Received: August 3, 2021; **Accepted:** August 21, 2021; **Published:** August 30, 2021

Abstract: Real-time anomaly detection is a critical monitoring task for power systems. Most studies of power network detection fail to identify small fault signals or disturbances that might lead to damages or system-wide blackout. This work presents a methodology for analyzing high-dimensional PMU data and detecting early events for large-scale power systems in a non-Gaussian noise environment. Also, spatio-temporal correlations of PMU data are explored and determined by the factor model for anomaly detection. Based on random matrix theory, the factor model monitors the variation of spatio-temporal correlations in PMU data and estimates the number of dynamic factors. Kullback-Leibler Divergence is employed to measure the deviation between two spectral distributions: the empirical spectral distribution of the covariance matrix of residuals from online monitoring data and its theoretical spectral distribution determined by the factor model. Using IEEE 57-bus, IEEE 118-bus, and Polish 2383-bus systems, three different case studies demonstrate that the proposed method is more effective in identifying early-stage anomalies in high-dimensional PMU data collected from large-scale power networks. Performance evaluations validate that this method is sensitive and robust to small fault signals compared with other statistical approaches. The proposed method is a data-driven approach that doesn't require any prior knowledge of the topology of power networks.

Keywords: Anomaly Detection, Spatio-Temporal Correlation, Kullback-Leibler Divergence (KLD), Factor Model

1. Introduction

Anomalies in power systems include deviations from expected measurements resulting from grid faults, such as load fluctuations and system oscillations. Anomaly detection is an important data analysis task that detects outliers from a given dataset. Also, real-time anomaly monitoring and detection are essential functions of the smart grid. Therefore, anomalies should gain more attention because they might lead to significant but rare events which might result in widespread blackouts. Discovering these faulty behaviors promptly and taking countermeasures will ensure the quality of the power grid and reduce economic losses [1].

Comprehensive studies of anomaly detection for large-scale power systems have been presented in recent years. Signal processing solution is a traditional method for event detection. For instance, the wavelet energy function is employed to examine the non-stationary occurrence of significant changes in signals

[2]. Also, principal component analysis (PCA)[3, 4] and singular value decomposition (SVD) [5] techniques are applied to event detection by reducing the system model dimension and extracting energy coherent structures. Recently, a large number of research topics focus on fault detection with machine learning. According to Wang [6], the online power system transient stability assessment problem is characterized as a two-class classification problem using a core vector machine. Other machine learning methods have been used in fault detection, including artificial neural networks (ANN) [7], KMeans [8], and isolated forest [9].

With the increasing scale of the power grid, such as Polish 2383-bus system, how to monitor such large power networks has been challenging for online monitoring analysis. Recently, a deep learning scheme has been successfully applied to anomaly detection for large-scale power systems, including long short-term memory (LSTM) [10, 11], graph neural networks (GNN) [12], deep neural network (DNN) [13].

However, calculations of these methods are growing exponentially with the depth and complexity of networks. Also, there are not enough labeled anomalies corresponding to the normal samples for training and testing.

Random matrix theory [14] as a statistic solution to big data has been mentioned to analyze massive data and successfully applied to the analysis of large-scale power networks. Using random matrix theory such as Qiu [15], state estimation and fault detection for power grid are investigated with PMU data. Ma [16] applies a graph partitioning algorithm to determine fault location by dividing the complex networks into several groups. Then the largest eigenvalue is calculated to detect the fault and locate the event position. Moreover, large size matrices of capacitor voltages are analyzed and identified the faulty submodules using single ring law analysis [17]. The linear eigenvalue statistics (LES) method is utilized to study situation awareness of power systems based on random matrix theory [18].

These methods mentioned above are helpful to identify faults. However, real-world power systems have strong interactions, mutual coupling effects, which exhibit spatio-temporal correlations in the power grid [15, 19]. Correlations existing in power systems lead to degraded accuracy and performance of traditional methods under Gaussian assumptions. Therefore, spatio-temporal correlations in power networks should be significantly considered. Investigation of the correlation structure of data is vital towards improving event detection performance [20, 21]. However, only a few studies focus on the above issue, which motivates us to explore an efficient correlation analysis method for power systems.

In recent years, the factor model for high-dimensional data analysis has attracted more attention. The main task of the factor model aims to use a small number of factors to characterize a large number of variables and extract correlation information from high-dimensional data. The factor model has been widely used in the analysis of big data recently. Yeo [22] applies high-dimensional factor models to capture economic market dynamics. Kapetanios [23] estimates the number of factors needed in high-dimensional factor models by studying the behavior of the eigenvalues. Pelger [24] explores factor models in high-frequency data and proposes an estimation approach to separate continuous and rare jump risk factors. Furthermore, Lee [25] generates a dynamic factor model to estimate power correlations between load and wind. In this paper, the factor model is used to analyze high-dimensional PMU data and study the spatio-temporal correlation of power systems.

Early anomaly detection aims to discover faults in a low signal-to-noise ratio (SNR) environment. Some faulty behaviors usually exhibit dynamic characteristics in their early stage. Due to noise interference, fault signals are submerged in the noise, which causes many algorithms inefficiency in early fault recognition. Recently, the event detection approach based on Kullback-Leibler divergence (KLD) has been successfully applied to early fault detection in some fields and approved its high sensitiveness and

robustness. For example, Bakdi [26] involves the KLD method to power quality monitoring in RMPV systems. Gupta [27] uses KLD based approach to monitor early warning signals of a blackout. Additionally, KLD with PCA technique proves high performance on incipient fault detection [28-30]. Moreover, a fault detection method based on Jensen-Shannon divergence is proposed to detect the surface material cracks using experimental ECT data [31].

Under the random matrix theory framework, this paper proposes a factor model-based KLD method (FKLD) to detect early fault events in large-scale power systems. The main contributions of this paper are as follows:

- 1) Propose an early anomaly detection method for large-scale power systems based on the KLD technique. The proposed method is sensitive to incipient faulty behaviors in power networks, which efficiently identifies significant failures in their early stages.
- 2) The method in this paper captures the correlation properties of high-dimensional power grid data by using factor model analysis. Therefore, the factor model-based FKLD detection method has more accuracy and higher performance than conventional statistic methods in non-Gaussian noise environment systems.
- 3) Present a statistical method based on random matrix theory, and it is robust to interference noise and bad data, including data errors.
- 4) This paper introduces a data-driven approach to detect small anomaly behaviors without prior knowledge of the complex topology of power systems.

The rest of the paper is organized as follows: In Section 2, data modeling and random matrix theory are described. Then correlation analysis for power systems is introduced in Section 3. Furthermore, factor model analysis for power systems is presented in Section 4. The proposed method is then introduced in Section 5. Next, cases are studied, and the results are exhibited in Section 6. Following that, performance evaluations are made by comparing with other algorithms in Section 7. Finally, conclusions are drawn in Section 8.

2. Data Modeling and Random Matrix Theory

Large power systems are modeled as high-dimensional random matrices, which transfers a complex anomaly detection problem of large-scale power networks into the analysis of matrices [14]. In this section, data processing is introduced firstly. High dimensional phasor measurement unit (PMU) [32, 33] data is employed to be status data in this paper, such as voltage magnitude and active power. Then Marchenko-Pasture Law, a fundamental theory of random matrix theory, is described.

2.1. PMU Data Processing

Let X be status data, $X \in \mathbb{C}^{N \times T}$ is a matrix of N variables for T sampling times collected from PMUs shown in Figure 1. X is split into the fixed-length moving windows of K

sampling times, $K < T$. A measurement matrix $E_{T(i)} \in \mathbb{C}^{N \times K}$ at current time $T(i)$ is generated as:

$$E_{T(i)} = (X_{T(i)-K+1}, X_{T(i)-K+2}, \dots, X_{T(i)}) \quad (1)$$

Where $X_{T(i)}$ is an N -row vector of X at current time $T(i)$, $X_{T(i)} = (X_{1,T(i)}, X_{2,T(i)}, \dots, X_{N,T(i)})^T$. Then the measurement matrix $E_{T(i)}$ can be normalized as:

$$\tilde{E}_{j,h} = (E_{j,h} - u(E_j)) \times \frac{\sigma(\tilde{E}_j)}{\sigma(E_j)} + u(\tilde{E}_j) \quad (2)$$

Where $u(\tilde{E}_j) = 0$ and $\sigma(\tilde{E}_j) = 1$, $u(E_j)$ and $\sigma(E_j)$ are the mean and standard variance of the j^{th} row vector in the matrix $E_{T(i)}$, $j \in (1, 2, \dots, N)$, $h \in (1, 2, \dots, K)$.

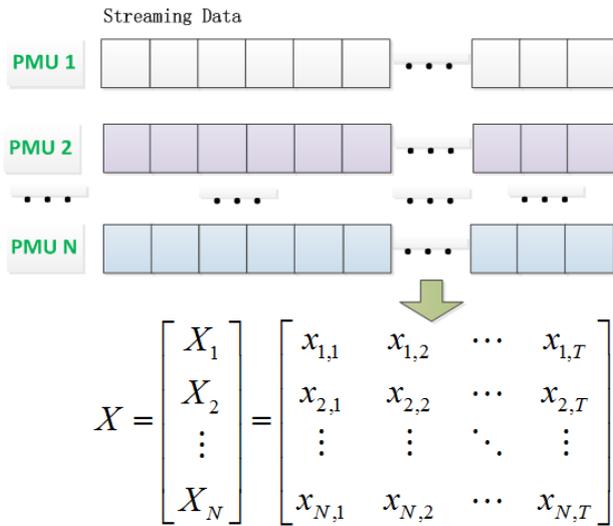


Figure 1. Streaming data collected from PMUs is constructed as a random matrix.

2.2. Marchenko-Pasture Law

Marchenko-Pasture Law (MP-law) [34] is described the asymptotic behavior of empirical spectral density (ESD). Let $X = [x_1, \dots, x_T]$ be $N \times T$ random matrix whose elements are independent identically distributed with mean $u(X) = 0$ and standard variance $\sigma(X) < \infty$. The sample covariance matrix of X is defined as:

$$S = \frac{1}{n} X X^H \quad (3)$$

Assume that $\lambda_1, \lambda_2, \dots, \lambda_N$ are the eigenvalues of the sample covariance matrix of X . ESD of $N \times N$ sample covariance matrix of X is generated by its eigenvalues:

$$\rho_s(X) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(S)} \quad (4)$$

Where δ_i denotes the dirac point measure at i . As $N, T \rightarrow \infty$ with $N/T = c \in (0, 1]$, spectral distribution converges to the distribution of MP-Law with density function:

$$f(x) = \begin{cases} \frac{1}{2\pi x c \sigma^2} \sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

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

3. Correlation Analysis for Power Systems

With the growing scale of power systems, such as distribution generations and flexible loads, the structure of power networks becomes much larger and more complex than it used to be. All electrical components are integrated into one system, which leads to strong interactions, mutual coupling effects, and high randomness when large fluctuations occur in power systems [19]. These behaviors mentioned above exhibit spatio-temporal correlations [15, 35] in power networks. Some studies have been discovered and proved their existence in power systems. Such as Xu [19] finds that correlations exist between the influential factors and system status utilizing augmented matrix and demonstrates that certain factors influence the operating states of power networks.

Additionally, Chakhchoukh [21] shows that cross-correlations occur when several PMUs are placed close to each other by analyzing Akaike information criterion, which reveals that different PMU measurements at different locations can be correlated as well. Furthermore, Ghanavati [36] finds that autocorrelation increases with the power system approaching instability and estimates autocorrelation. Here, autocorrelation (temporal correlation) is the correlation between two measurements $u_{n,t}$ and $v_{n,t-i}$ at the same bus n , while the correlation between two measurements $u_{n,t}$ and $v_{m,t+\tau}$ from different buses (bus n and bus m) is cross-correlation (spatio-correlation). In this paper, spatio-temporal correlation is modeled and estimated using factor model analysis with high dimensional PMU data collected in a non-Gaussian environment.

4. Factor Model Analysis for Power Systems

Traditional studies theoretically use Gaussian assumptions [18]. However, in the real world, large-scale power systems exhibit behaviors of spatio-temporal correlation based on the discussion above. In this section, the factor model [22] is utilized to capture the correlation structure of PMU data and predict the power state mode. The main task of this section is to build factor model based on random matrix theory. In this section, the subspace of residuals is construed, firstly. Then estimated eigenvalue distribution and the modeled eigenvalue distribution of residuals are obtained. Finally, by minimizing the distance between estimated and modeled spectral distributions, the spatio-temporal correlations of power systems can be determined.

4.1. Estimated Eigenvalue Distribution Analysis

The estimated eigenvalue distribution is produced by the ESD of the covariance matrix of residuals after eliminating p factors from PMU data.

Let X be a $N \times T$ matrix. X is decomposed into p principal components (factors) and residuals as $X = R + \sum_{p=1}^n L_p F_p$. Thus residual matrix is generated as below:

$$R = X - \sum_{p=1}^n L_p F_p \quad (6)$$

Where R is $N \times T$ residual matrix obtained after p principal components are removed from X . F_p is a matrix of factors, and L_p is a matrix of the factor loading, p is the number of factors to be subtracted from PMU data. Furthermore, the residuals are modeled by the covariance matrix shown as:

$$C_p = \frac{1}{T} R R^H \quad (7)$$

Finally, the ESD of the covariance matrix C_p , represented by $F^{real}(p)$, is obtained by equation (4).

4.2. Modeled Eigenvalue Distribution Analysis

The next step is to investigate the modeled eigenvalue distribution of the covariance matrix of residuals. A doubly correlated form is constructed to build the spatio-temporal correlation structure as below:

$$R = A_N^{1/2} \varepsilon B_T^{1/2} \quad (8)$$

where ε is an $N \times T$ matrix with independent identical distribution entries. A_N and B_T representing cross-covariance and autocovariance, respectively, are symmetric and non-negative definite matrices.

$$a^4 c^2 M^4 + 2a^2 c(-1 + b^2)z + a^2 c M^3 + ((1 - b^2)^2 z^2 - 2a^2 c(1 + b^2)z + (c^2 - 1)a^4)M^2 - 2a^4 M - a^4 = 0 \quad (9)$$

Where $|b| < 1$, $a = \sqrt{1 - b^2}$ and $c = N/T$. Basic derivations of the polynomial equation (9) are provided in the Appendix.

2) Next, the Stieltjes Transform can be generated from moment generating function:

$$G^{model}(z) = \frac{M^{model}(z)+1}{z} \quad (10)$$

3) Finally, the modeled eigenvalue distribution $F^{model}(b)$ is estimated from the imaginary part of $G^{model}(z)$ given by:

$$F^{model}(b) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0^+} \Im m G(\lambda + i\varepsilon) \quad (11)$$

4.3. Spectral Distribution Distance

The estimated and modeled eigenvalue distributions of the covariance matrix are constructed in residual subspace from the discussion above. The former is controlled by p , the number of factors to be removed. The latter depends on parameter b . Now the objective of estimation is to predict two parameters of (p, b) by finding the best fit of $F^{real}(p)$ and $F^{model}(b)$. Jensen-Shannon divergence, which is the symmetric form of KLD, measures the spectral distance between $F^{real}(p)$ and $F^{model}(b)$ as:

In this paper, the structures of A_N and B_T are restricted so that they are expressed by simple parameter sets η and α , for example, each residual has the same cross-correlation, η , to other residuals, and each residual has an exponentially decaying temporal autocorrelations with a parameter α , so A_N and B_T are determined by two parameters η and α , since $A_N = \{(A_N)_{ii} = 1, (A_N)_{ij, i \neq j} = \eta, i, j = 1, \dots, N\}$, and $B_T = \{(B_T)_{st} = \exp(-\frac{|s-t|}{\alpha}), s, t = 1, \dots, T\}$. Thus, the modeled eigenvalue distribution can be represented by $F^{model}(\eta, \alpha)$.

Two assumptions are made to simplify the estimation in the factor model:

Assumption 1:

Total cross-correlations are efficiently eliminated from p principal components so that cross-correlations can be negligible. Thus $A_N \approx I_{N \times N}$ or $\eta = 0$.

Assumption 2:

The autocorrelations are exponentially decreasing, which can be expressed by $\{B_T\} = b^{|i-j|}$, with $|b| < 1$.

Based on the above two assumptions, $F^{model}(\eta, \alpha)$ can be reduced to $F^{model}(b)$ with only parameter b . Since it is difficult to obtain the limiting spectral distribution from Stieltjes Transform, free random variable techniques are provided to calculate the result for $F^{model}(b)$, which can be generated by the following three steps:

1) By using the moments generating function $M \equiv M^{model}(z)$ and its inverse relation to N transform, we can derive the following polynomial equation:

$$D_{KL}(F^{real} || F^{model}) = \frac{1}{2} \sum_i F_i^{real} \log \frac{F_i^{real}}{M_i} + \frac{1}{2} \sum_i F_i^{model} \log \frac{F_i^{model}}{M_i} \quad (12)$$

Where $M_i = \frac{F_i^{real} + F_i^{model}}{2}$. More details of factor model estimation are shown in Algorithm I. To quickly search for b in the minimizing process, we use the Newton method to solve equation (12). p and b approximately approach their actual values when the distance between estimated and theoretical distribution is small. Here p and b stand for cross-correlation and autocorrelation, respectively. p and b exhibit dynamic characters of data and indicate the different operating state modes. Therefore, they can be used as indices for anomaly detection.

Algorithm I. Estimation for spatio-temporal correlations p_i, b_i , and $F_{T(i)}^{model}(b_i)$:

- 1: Construct PMU data matrix $X \in \mathbb{C}^{N \times T}$;
- 2: At each sampling point $T(i)$;
- 3: Separate X into fixed-length windows. Each window forms a matrix $E_{T(i)} \in \mathbb{C}^{N \times K}$ following equation (1);
- 4: Obtain residuals $R_{T(i)}$ after p_i factors removed from $E_{T(i)}$ by equation (6), $p_i \in \mathbb{Z}^+$;
- 5: Normalize residuals $R_{T(i)}$ through equation (2);
- 6: Obtain covariance matrix of $R_{T(i)}$ using equation (3);

- 7: Calculate $F_{T(i)}^{real}(p_i)$ by equation (4);
- 8: Use the Newton method to find b_i by solving the function $D_{KL}^{T(i)}(b_i)$ in equation (12);
- 9: Find a solution to $F_{T(i)}^{model}(b_i)$ using equation (11).

5. The Proposed Method for Early Anomaly Detection

In this section, the proposed method for anomaly detection is mainly introduced. Firstly, KLD anomaly detector for power systems is presented. Then the framework of the proposed method is illustrated. Eventually, a fault-to-noise ratio (FNR) is given to evaluate the performance of our approach.

5.1. KLD Anomaly Detector

KLD anomaly detector [37, 38] is based on Kullback-Leibler divergence, which is a metric to measure the distance between two probability distribution $P(x)$ and $Q(x)$. KLD is defined as:

$$D_{KL}(P||Q) = \sum_i P_i \log \frac{P_i}{Q_i} \quad (13)$$

Jensen-Shannon divergence [39], the symmetric version of KLD, is used in this paper. Jensen-Shannon divergence is defined as follows:

$$J(P||Q) = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M) \quad (14)$$

Where $M = \frac{1}{2}(P + Q)$. Equations (15) and (16) are given to

solve equation (14):

$$D_{KL}(P||M) = \sum_i P_i \log \frac{P_i}{M_i} \quad (15)$$

$$D_{KL}(Q||M) = \sum_i Q_i \log \frac{Q_i}{M_i} \quad (16)$$

$D_{KL}(P||Q)$ is always positive, and it is equal to zero when two probability distributions are identical. Otherwise, it becomes higher when one probability distribution deviates from another, which is described as follows:

$$\begin{cases} D_{KL}(P||Q) = 0, & \text{if } P = Q \\ D_{KL}(P||Q) > 0, & \text{if } P \neq Q \end{cases} \quad (17)$$

In practice, there is always a non-zero deviation caused by interference noise. Therefore, a moderate threshold of KLD detector must be considered, which is described as:

$$D_{KL}(P||Q) \leq H_{th} \quad (18)$$

Conventionally, a higher-level noise may lead to a bigger threshold. There are two schemes to mitigate this situation: One approach is increasing the data dimension to reduce the impact of statistic noise on the KLD detector. It is known that central limit theory requires $N, T \rightarrow \infty$. However, T can be very large in the real world to satisfy the theory limit, but N (dimension) is often moderate [14]. Therefore, the low dimension of data degrades the performance of many statistical approaches. By increasing the data dimension, spectral distribution is more approximate to theoretical measures, which leads to statistical noise deduction. Another method is finding an appropriate model for raw data to reduce the error of detection.

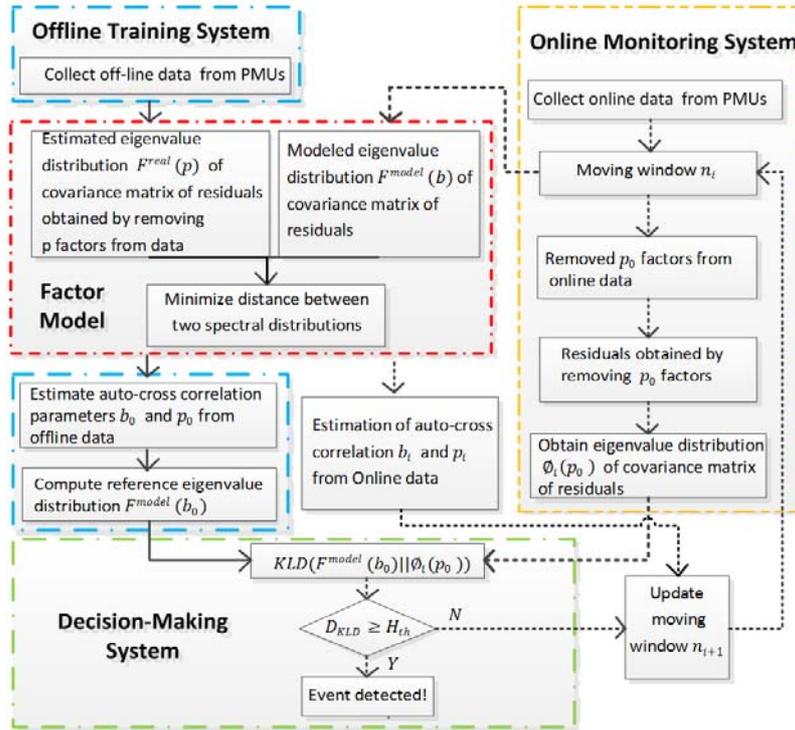


Figure 2. The architecture of the FKLD detection method.

5.2. The Framework of the Proposed Method

Concluding techniques mentioned above, the procedure of the proposed method is outlined to make an entire point of view for this detection process, which is shown in Figure 2.

Offline Training System:

KLD detector compares two spectral distributions: the eigenvalue distribution of online data and the reference spectral distribution. The main task of the offline training process is to estimate an appropriate reference spectral distribution for KLD detector. Initially, the factor model is established by two spectral distributions $F^{real}(p_0)$ and $F^{model}(b_0)$. $F^{real}(p_0)$ is the estimated eigenvalue distribution of covariance matrix of residuals obtained by removing p_0 factors from data. And $F^{model}(b_0)$ stands for the reference eigenvalue distribution. Once the distance of two distributions is minimized, the auto-cross correlations parameter p_0 and b_0 can be determined. It is noted that the offline data is collected from a fault-free environment of power networks. Considering a non-Gaussian environment, p_0 and b_0 are reference parameters that represent spatio-temporal correlations in a normal operating state, which are determined according to Algorithm I. Thus, a reference spectral distribution $F^{model}(b_0)$, also named factor model, is obtained using Algorithm I and ready for the following KLD detector.

Online Monitoring System:

Online monitoring is the process of generating the eigenvalue distribution of online data. Firstly, to set up the process, online PMU data is segmented into a fixed-length window of n sampling points. The detection algorithm is initialized by the first n sample points. And then the moving window is updated every n sampling points.

At the beginning of the process, residuals are obtained after p_0 factors subtracted from online data, p_0 was a pre-estimated parameter from the offline system. Next, the eigenvalue distribution of the covariance matrix of residuals from online data, which is represented as $\phi_i(p_0)$, is determined by equation (4). When fault signals or disturbances occur, $\phi_i(p_0)$ deviates from its reference spectral distribution $F^{model}(b_0)$ since spatial-temporal correlations of status data (such as bus voltage magnitude, active power) changed. If the index of KLD detector exceeds its threshold, then an anomaly event will be issued. Following that, b_i and p_i at the i^{th} segmented windows are estimated by factor model using online data.

Decision-Making System:

The main task of the decision-making process is to compare the eigenvalue distribution of online data with its reference spectral distribution. In this process, a KLD anomaly detector measures the deviation between two different distributions $F^{model}(b_0)$ and $\phi_i(p_0)$. Then a threshold is given to determine whether an event or non-event exists. We formulate this decision-making process as a hypothesis test as below:

$$\begin{cases} H_0: D_{KL}(F^{model}(b_0)||\phi_i(p_0)) < H_{th} \\ H_1: D_{KL}(F^{model}(b_0)||\phi_i(p_0)) \geq H_{th} \end{cases} \quad (19)$$

Where H_{th} is a threshold of KLD detector. The threshold H_{th} equals zero on the condition that the estimated distribution is equal to its reference. However, due to noises are existing in real situations, there are always slight deviations that reflect as H_{th} .

5.3. Fault-to-Noise Ratio Evaluation

Fault-to-noise ratio (FNR) [30, 31] is employed to evaluate FKLD performance, which measures the detector sensitivity to faults in the presence of noise by comparing the level of noises and fault signals. FNR is defined as:

$$FNR = 10 \times \log_{10} \frac{\sigma_e^2}{\sigma_0^2} \quad (20)$$

where σ_e^2 and σ_0^2 are estimated variances. σ_e^2 is related to changes due to fault occurrence, σ_0^2 is estimated by PMU data collected from a fault-free environment.

6. Case Studies

In this section, the proposed approach is simulated with PMU data from IEEE 57-bus, IEEE 118-bus, and Polish 2383-bus systems. Details about the three test systems above can be found in Matpower Package [40]. Three cases in different scenarios are simulated on the test systems, and the effectiveness of the proposed method is validated. In the first case, the proposed method detects fault events from single line tripping. Then voltage collapse event is analyzed in the second case. In the third case, multiple-fault detection is presented by comparing the linear eigenvalue statistics (LES) algorithm with the FKLD method. The synthetic data is constructed as $X = U + r\zeta$, where U is generated data. In this paper, voltage magnitude is utilized to be generated data in all three cases. ζ is non-Gaussian noise which is expressed: $\zeta_{i,t} = 0.5 \times \zeta_{i,t-1} + \varepsilon_{i,t}$. $\varepsilon_{i,t}$ is white noise, $\varepsilon_{i,t} \in \mathcal{N}(0, 1 - 0.5^2)$ so that variance of $\zeta_{i,t}$ is 1. r is the magnitude of noise added to synthetic data. Meanwhile, faults or significant disturbances are regarded as fault signals represented by ω_{fault} . Thus, the signal-to-noise ratio (SNR) is calculated by: $SNR(dB) = 20 \log|\omega_{fault}/r|$.

Case 1: Anomaly Detection for Line Tripping Event

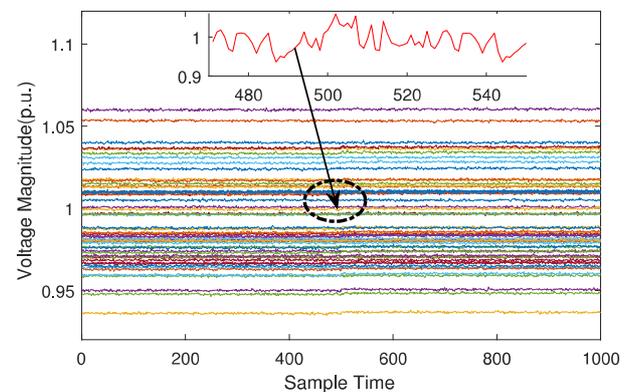


Figure 3. Synthetic data is generated from the IEEE 57-bus system. An event is set at $t = 501s$ in case 1.

In this case, the generated data shown in Figure 3 is created from the IEEE 57-bus system (see Figure 4), containing 57 voltage magnitude variables with 1000 sampling points. Assume that a transmission line connecting bus 9 to bus 10 is tripped at $t_s = 501s$. Details of this event are shown in Table 1. Furthermore, the moving window size is set to be 57×80 , and SNR is 30 dB in a non-Gaussian environment. Then fault detection performance is verified by the proposed method. Figure 5 presents the simulation result detected by the proposed method over continuously moving windows. Due to moving window initializing, the first testing time starts from 81. During $t_s = 0 \sim 500s$, values of FKLD are almost constant, which implies that the power system operates in a

normal state. Starting from $t_s = 501s$, FKLD values increase rapidly, which indicate that an event is occurring at $501^{th}s$. After several sample points, FKLD values decrease to normal, with the system returning to a new steady state. In addition, the changes of spectral distribution in different operating states are investigated by comparing the eigenvalue distribution, factor model, and MP-law illustrated in Figure 6. It demonstrates that the factor model can successfully fit the eigenvalue distribution of covariance matrix of residuals, which is an efficient model to capture the spatio-temporal correlation structure of grid data. In contrast, MP-Law is not able to characterize the spectral distribution of data with spatio-temporal correlations.

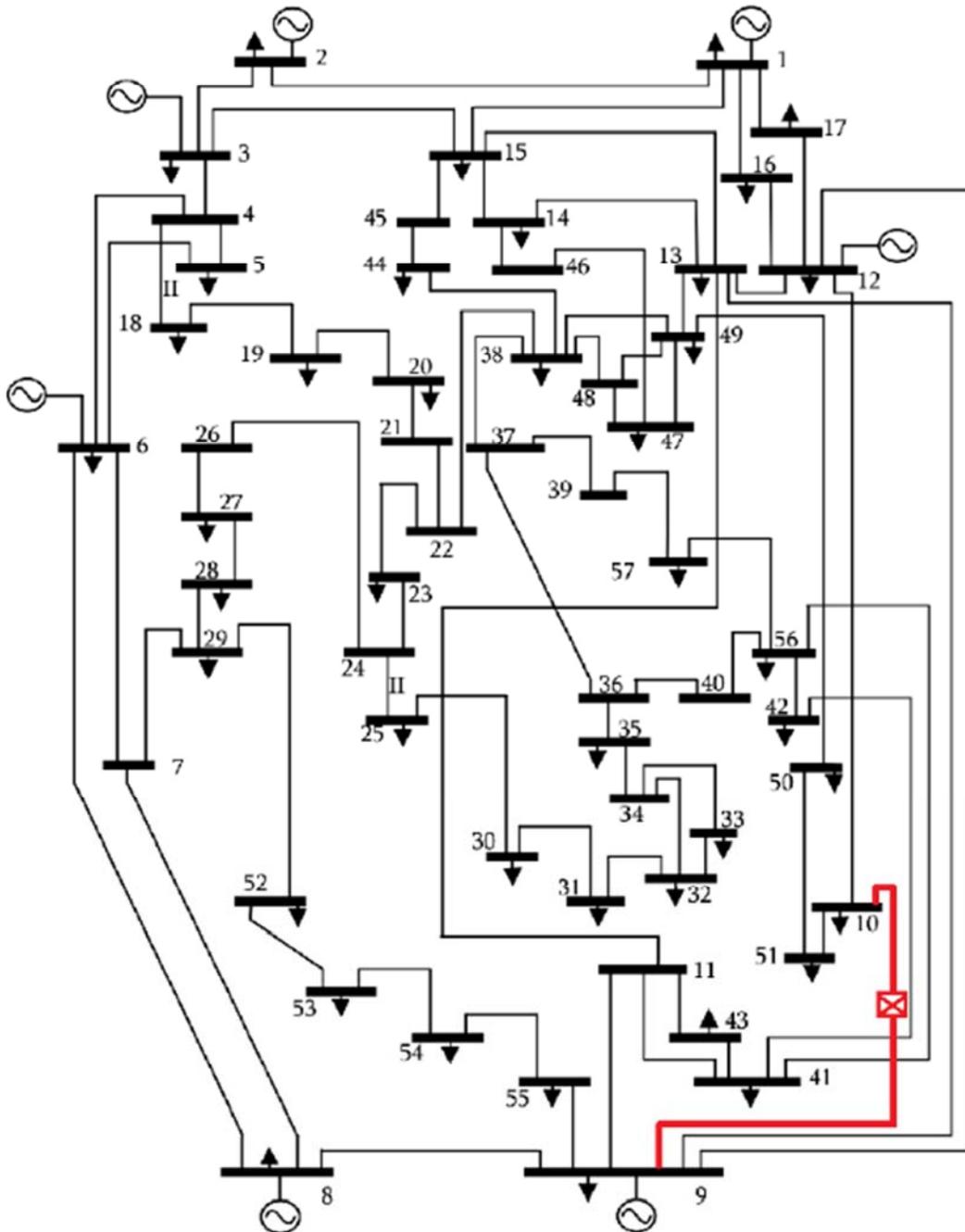


Figure 4. In the IEEE 57-bus system [41], a line connecting bus 9 to bus 10 is tripped at $t_s = 501s$ in case 1.

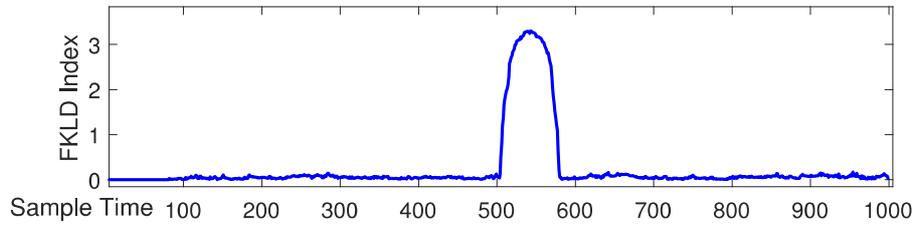


Figure 5. FKL D anomaly detection results in case 1.

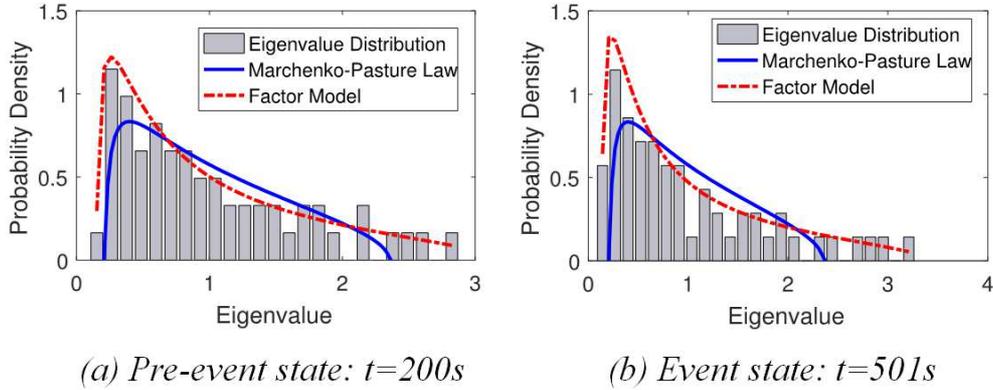


Figure 6. Comparison of eigenvalue distribution, factor model, and MP-law in pre-event state and event state in case 1, $c=0.71$.

Table 1. Line tripping event in case 1.

Bus _{from}	Bus _{to}	Sampling Time(s)	Impedance (p.u.)
9	10	$t_s = 0\sim 500s$	0.02
9	10	$t_s = 501\sim 1000s$	10
Others		$t_s = 1\sim 1000s$	Unchanged

Case 2: Anomaly Detection for Voltage Collapse Event

In this case, the generated data is created from IEEE 118-bus system. A single overload fault is assumed at bus 52, with the details shown in table 2. The voltage magnitude data is collected containing 118 variables with 1000 sampling points exhibited in Figure 7. Furthermore, we set the size of moving window to be 118×240 . And SNR is 27 dB in a non-Gaussian environment. Next, we test the performance of the proposed method. As illustrated in Figure 8, during $t_s = 0\sim 500s$, FKL D values remain almost constant, which indicates the power system operates in a normal state. From $t_s = 501s$, FKL D values gradually increase, which coincides with the variation of voltages caused by gradually increasing load at bus 52. From $t_s = 901s$, FKL D values are growing with further increasing voltage arising from continuing variation of load. Starting from $t_s = 1301s$, FKL D values increase rapidly, which coincides with voltage collapse.

In this case, we further investigate correlation variations in different operating states. As details are seen from Figure 9, b and p are estimated in the various operating states. b is increasing from 0.5 in the pre-event state to 0.68 in the state of voltage collapse. This result reveals that b increases as the increment of faults signal in power networks. In other words, b is related to the level of fluctuation on each bus. When

overload happens, the variations of power flow on the faulty buses affect the neighbor buses. Therefore, cross-correlations between buses near the event location vary rapidly at the moment of the faults occurring. Here, P reveals the level of cross-correlation between buses. Furthermore, we compare spectral distribution in different operating states. As seen in Figure 9, spectral distributions are changing with the variation of b and p , which indicates that the factor model can capture specific correlation structures of dynamic data.

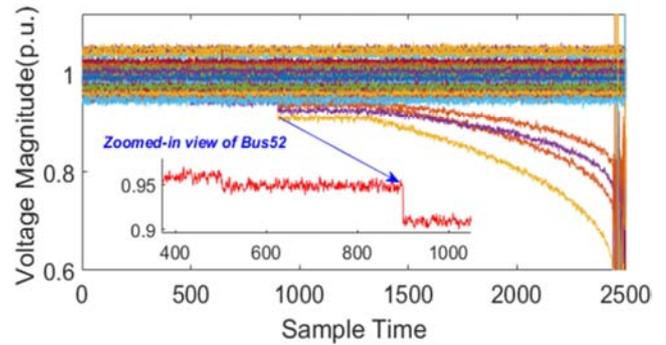


Figure 7. The synthetic data is generated from IEEE 118-bus test system. An event occurs at $t = 501s$. With the evolution of fault, voltage collapse takes place in the system in case 2.

Table 2. Voltage collapse event in case 2.

Time(s)	1~500s	501~900s	901~1300s	1301~2500s
$P_{Bus52}(MW)$	10	40	130	$t/5 - 130$

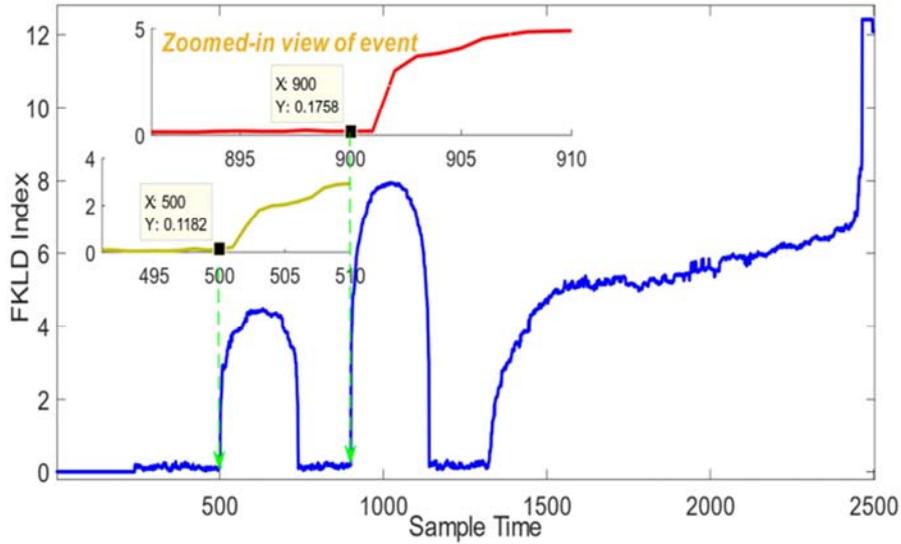


Figure 8. The detection result of voltage collapse event and partial zoomed-in view of the results in case 2.

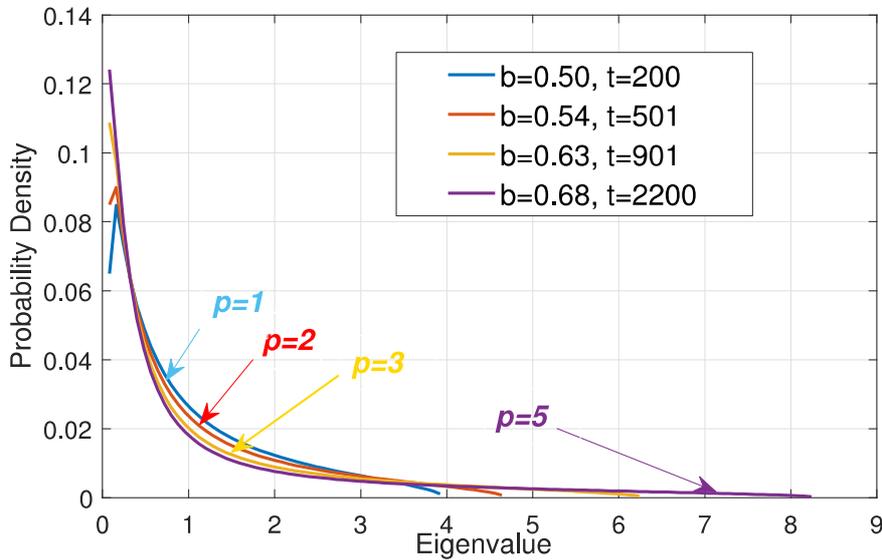


Figure 9. The spectral distributions in different operating states of the power system in case 2, $c=0.49$.

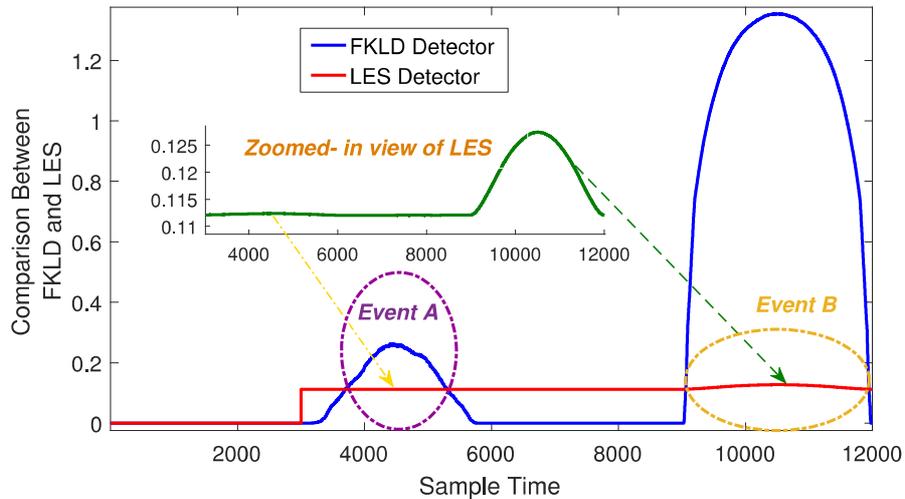


Figure 10. Comparison of Detection results between FKLD and LES algorithms on Polish 2383-bus system and partial zoomed-in view of the results in case 3, where LES fails to identify Event A.

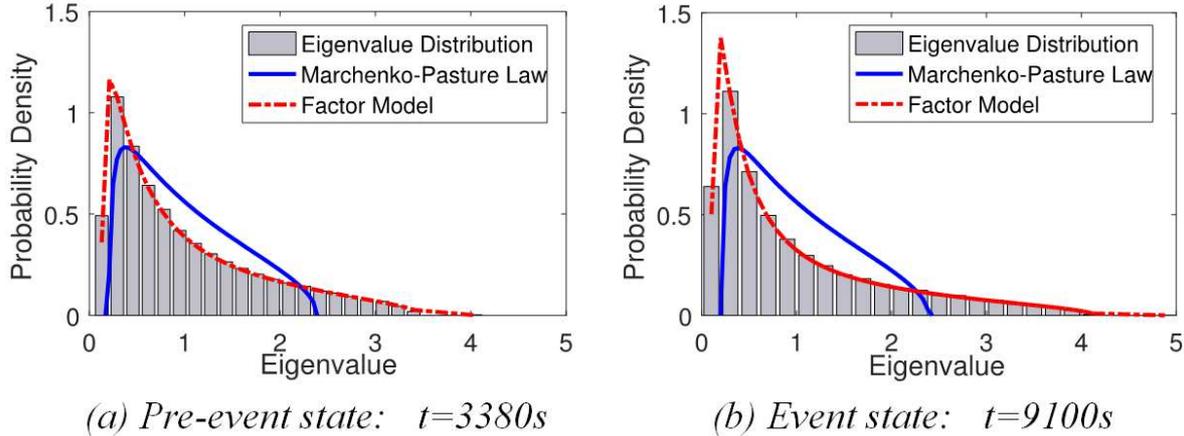


Figure 11. Comparison of eigenvalue distribution, factor model, and MP-law at different times in case 3, $c=0.79$.

Case 3: Multiple Event Detection

In this case, the main task is to verify the performance of the proposed method on multiple events in large-scale power networks. Voltage magnitude data is created from the polish 2383-bus test system, containing 2383 variables with 12000 sampling points in a non-Gaussian environment with 25 dB SNR.

Multiple faults are set at 59th bus, 61th bus, and 126th bus, respectively. More details of events are presented in Table 3. Additionally, the size of the moving window is set to be 2383×3000. In this case, the detection performance of the proposed method in high dimensionality is showed by comparing FKLD with the LES approach [18][20]. Moreover, LES algorithm can be constructed by using various test functions. The commonly used test functions of LES include Chebyshev Polynomial ($\varphi(\lambda) = 2\lambda^2 - 1$), Information Entropy ($\varphi(\lambda) = -\lambda \ln \lambda$), Likelihood Ratio Function ($\varphi(\lambda) = \lambda - \ln \lambda - 1$), and Wasserstein Distance ($\varphi(\lambda) = \lambda - 2\sqrt{\lambda} + 1$), respectively. In this simulation, Chebyshev Polynomial Function is used to be a test function. The detection result and its partial zoomed-in view are presented in Figure 10. The result exhibits that the LES method (red) is not able to detect Event A occurring at $t_s = 3001s$. In contrast, the proposed method (blue) can easily recognize it, indicating that the FKLD method has higher sensitivity to small fluctuations than the LES method in high dimensionality. The result demonstrates that FKLD is an efficient method for small fault recognition in large-scale power networks. Meanwhile, comparisons between the eigenvalue distribution of online data and the factor model in different operating states of the power network are shown in Figure 11. It reveals that the factor model fits the eigenvalue distribution of covariance matrix of residuals from high dimensional PMU data. By way of contrast, MP-law is plotted at the same time.

Furthermore, p and b are investigated in multiple event detection. p and b represent the different state modes. For example, as seen from table 4, one factor is estimated in a normal state. Due to the number of events arising, p increases from two to three, implying that p is related to the increasing number of faults. The variation of p is caused by mutual

effects between the buses when multiple events occur. Meanwhile, following the varies of faults signals, b is constantly increasing as the fluctuations of power systems.

Table 3. Multiple events in case 3.

Event ID	Bus	Sampling Time(s)	Active Load (MW)
Event A	59	$t_s = 1\sim 3000s$	0
		$t_s = 3001\sim 12000s$	100
Event B	61	$t_s = 1\sim 9000s$	0
		$t_s = 9001\sim 12000s$	100
	126	$t_s = 1\sim 9000s$	25
Others		$t_s = 9001\sim 12000s$	500
		$t_s = 1\sim 12000s$	Unchanged

Table 4. Comparison of FKLD values, p , and b in different states in case 3.

System State	Event Time(s)	p	b	FKLD
Normal State	2000s	1	0.5	5×10^{-5}
Event A	3380s	2	0.581	0.0204
Event B	9100s	3	0.673	0.3613

7. Performance Comparison with Other Algorithms

In the above section, the performance of the proposed method is verified in different scenarios. The FKLD algorithm is evaluated for further analysis by comparing it with other test functions of the LES method. All the simulations in this section are tested on an IEEE 118-bus test system with 5000 sample points.

The performance evaluation of the FKLD method is validated with probabilities of false alarm under non-hypothesis as:

$$P_{fa} = P(D_{KL} > H_{th}|H_0) \tag{21}$$

For a fixed P_{fa} , the detection threshold is determined by:

$$H_{th} = m_0 + \gamma\sigma_0 \tag{22}$$

Where m_0 and σ_0 are the mean and standard variance of data which is collected from fault-free environment of power systems. γ is the threshold adjustment. For a fixed P_{fa} , γ is determined by Monte Carlo studies at a specified SNR.

In this evaluation, detection probability is analyzed to prove the performance of the proposed method by comparing it with other algorithms. Considering the worst and the best cases of SNR, the range of SNR is set to be [10, 30] dB, and the probability of false alarm is fixed with 1%. As shown in Figure 12, the proposed method exhibits good performance above 15 dB (87% Detection Probability). However, the rest of the algorithms require higher SNR to achieve the same detection probability as the FKLD method. In a word, the proposed method outperforms the LES algorithm in early anomaly detection.

The FNR performance of the FKLD method is studied by comparing it with different algorithms to evaluate sensitivity to fault signals in the presence of noise. The range of SNR is set to be [10, 30] dB. As exhibited in Figure 13, the FNR values of the FKLD method are the highest among all the other algorithms. The result demonstrates that the FKLD method has a high sensitivity to small fault signals, which explains why the proposed method has good performance in a relatively low SNR environment.

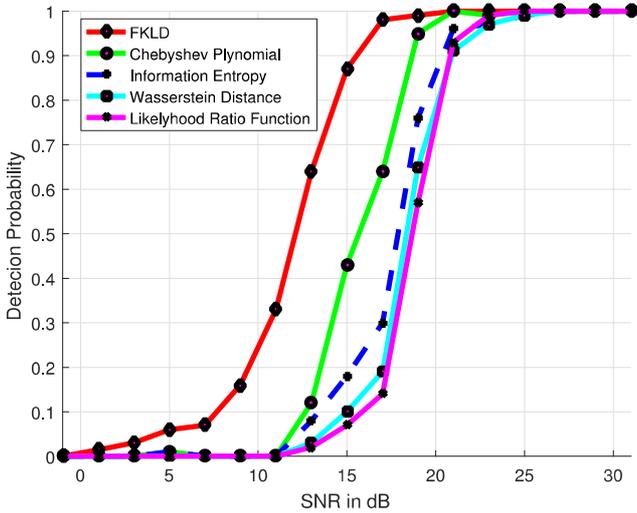


Figure 12. Detection Probability of FKLD and different algorithms.

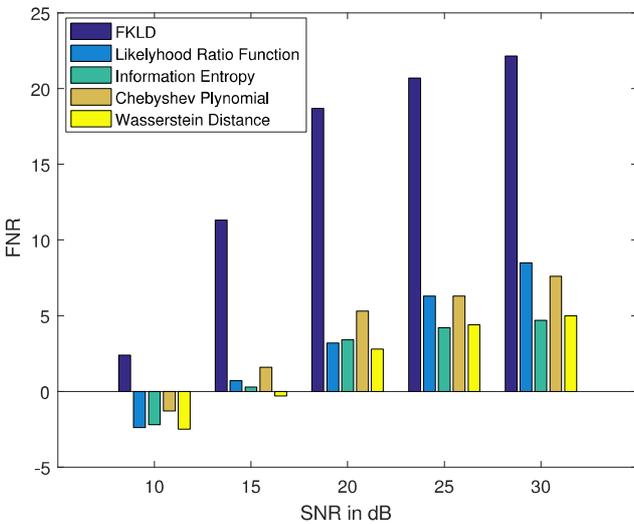


Figure 13. FNR versus SNR in the comparison of different algorithms.

8. Conclusion

This paper proposes the FKLD method to analyze high-dimensional PMU data and detect early fault events in a non-Gaussian noise environment. Spatial-temporal correlations are determined by factor model analysis considering existing correlation structures in the power grid, which captures dynamic information of high-dimensional PMU data. Furthermore, performances of the proposed method are verified by using standard, large, and synthetic systems, including IEEE 57-bus, IEEE 118-bus, and Polish 2383-bus systems in different scenarios. Results demonstrate that the proposed method is effective in discovering early-stage events in power networks.

Performance evaluations of the proposed method reveal that the FKLD method is more sensitive and robust to high-dimensional data anomalies than other statistical methods in a relatively low SNR environment. This methodology applied to online monitoring and fast diagnosis for power systems is data-driven without any topology knowledge.

Appendix

Key Concepts and Derivation of Polynomial Equation

Definition 1: The Green's Function or Stieltjes Transform is defined as:

$$G_X(z) = \int \frac{\rho_X(\lambda)}{z-\lambda} d\lambda \quad (23)$$

Where $\rho_X(\lambda)$ is ESD of the covariance matrix of X . z is a complex variable. Also, the spectral density $\rho_X(\lambda)$ can be obtained from the imaginary part of Green's Function:

$$\rho_X(\lambda) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \Im m G_X(\lambda + i\varepsilon) \quad (24)$$

Definition 2:

The n -th moment of ESD is defined as:

$$m_n = \frac{1}{N} \langle \text{Tr} X^n \rangle = \int \rho_X(\lambda) \lambda^n d\lambda \quad (25)$$

Definition 3:

Moment Generating Function:

$$G_X(z) = \sum_{n \geq 0} \frac{m_n}{z^{n+1}} \quad (26)$$

$$M_X(z) = \sum_{n \geq 1} \frac{m_n}{z^{n+1}} \quad (27)$$

The relationship of $G_X(z)$ and $M_X(z)$ as:

$$M_X(z) = zG_X(z) - 1 \quad (28)$$

Definition 4:

N-Transform is a function that is the inverse transform of moment generating function $M_X(z)$:

$$M_X(N_X(z)) = N_X(M_X(z)) = z \quad (29)$$

Definition 5:

S-Transform is a basic transform of free probability theory

which is given by:

$$S(z) = \frac{z+1}{z} \gamma^{-1}(z) \tag{30}$$

where $\gamma(s) = \int \frac{sx}{1-sx} dp(x) = -s^{-1}G(s^{-1}) - 1$.

For two free matrices X_1 and X_2 , random variable multiplication law is obtained:

$$S_{X_1 X_2}(z) = S_{X_1}(z) S_{X_2}(z) \tag{31}$$

The relationship between S Transform and N-Transform is found as below:

$$S_X(z) = \frac{1+z}{zN_X(z)} \tag{32}$$

We can then use the free random variable (FRV) multiplication law to obtain the N-Transform of multiplication of X_1 and X_2 ,

$$\begin{aligned} N_{X_1 X_2}(z) &= \frac{1+z}{zS_{X_1 X_2}(z)} \\ &= \frac{1+z}{zS_{X_1}(z)S_{X_2}(z)} \\ &= \frac{1+z}{z \frac{1+z}{zN_{X_1}(z)} \frac{1+z}{zN_{X_2}(z)}} \\ &= \frac{z}{1+z} N_{X_1}(z) N_{X_2}(z) \end{aligned} \tag{33}$$

Based on the basic concepts and properties mentioned above, N-Transform for empirical covariance matrix can be derived in the following steps:

Initially, define a general form of empirical covariance: $R = A^{1/2} \epsilon B^{1/2}$ Where ϵ is $N \times T$ matrix with identical independent distribution entries. A and B are $N \times N$ and $T \times T$ symmetric non-negative definite matrices. Then Covariance matrix of R is obtained as:

$$C = \frac{1}{T} R R^T = \frac{1}{T} A^{1/2} \epsilon B^{1/2} (A^{1/2} \epsilon B^{1/2})^T = \frac{1}{T} A^{1/2} \epsilon B^{1/2} (B^{1/2})^T \epsilon^T (A^{1/2})^T = \frac{1}{T} A^{1/2} \epsilon B \epsilon^T A^{1/2} \tag{34}$$

Following that, N-Transform of the covariance matrix of C is derived by using the cyclic property of trace and FRV multiplication law as:

Finally, we obtain a fourth-order polynomial equation (9) with $a^2 = 1 - b^2$.

$$\begin{aligned} N_{\frac{1}{T} A^{1/2} \epsilon B \epsilon^T A^{1/2}}(z) &= N_{\frac{1}{T} A^{1/2} \epsilon B \epsilon^T}(z) \\ &= N_{\frac{1}{T} A \epsilon B \epsilon^T}(z) \text{ (cyclic property of trace)} \\ &= \frac{z}{1+z} N_A(z) N_{\frac{1}{T} \epsilon B \epsilon^T}(z) \text{ (FRV multiplication law)} \\ &= \frac{z}{1+z} N_A(z) N_{\frac{1}{T} \epsilon^T \epsilon B}(z) \text{ (cyclic property of trace)} \\ &= \frac{z}{1+z} N_A(z) \frac{rz}{1+rz} N_B(rz) N_{\frac{1}{T} \epsilon^T \epsilon}(z) \text{ (FRV multiplication law)} \\ &= \frac{z}{1+z} N_A(z) \frac{rz}{1+rz} N_B(rz) \frac{(1+z)(1+rz)}{z} \\ &= rz N_A(z) N_B(rz) \end{aligned} \tag{35}$$

Where $N_{\frac{1}{T} \epsilon^T \epsilon}(z) = \frac{(1+z)(1+rz)}{z}$. Considering $M \equiv M(z)$ and $z = M = M(z)$ which are defined in equation (29), the above equation (35) can be rewritten:

$$N_{\frac{1}{T} A^{1/2} \epsilon B \epsilon^T A^{1/2}}(M(z)) = r M N_A(M) N_B(rM) = z \tag{36}$$

Assume that cross-correlation of residual is completely removed, the cross-correlation matrix is satisfied $A = I_N$. Then we have $N_A(z) = N_I(z) = 1 + 1/z$. Therefore, the equation (36) is conveniently rewritten as:

$$rM = M_B\left(\frac{z}{r(1+M)}\right) \tag{37}$$

Now we need to find M_B for equation (37). The auto-covariance matrix B from (34) is AR (1) process, which can be considered as a simple form $B = b^{|i-t|}$. M transform of the matrix B can be generated using a Fourier transform as:

$$M_B(z) = -\frac{1}{\sqrt{1-z} \sqrt{1-\frac{(1+b^2)^2}{1-b^2} z}} \tag{38}$$

References

- [1] Anderson, D., Zhao, C., Hauser, C., Venkatasubramanian, V., Bakken, D., & Bose, A. (2012). Real-time simulation for smart grid control and communications design. *IEEE Power Energy Mag.*, 10 (1), 49-57.
- [2] Kim, D. I., Chun, T. Y., Yoon, S. H., Lee, G., & Shin, Y. J. (2015). Wavelet-based event detection method using PMU data. *IEEE Transactions on Smart grid*, 8 (3), 1154-1162.
- [3] Rafferty, M., Liu, X., Laverty, D. M., & McLoone, S. (2016). Real-time multiple event detection and classification using moving window PCA. *IEEE Transactions on Smart Grid*, 7 (5), 2537-2548.
- [4] Ge, Y., Flueck, A. J., Kim, D. K., Ahn, J. B., Lee, J. D., & Kwon, D. Y. (2015). Power system real-time event detection and associated data archival reduction based on synchrophasors. *IEEE Transactions on Smart Grid*, 6 (4), 2088-2097.
- [5] Lim, J. M., & DeMarco, C. L. (2015). SVD-based voltage stability assessment from phasor measurement unit data. *IEEE Transactions on Power Systems*, 31 (4), 2557-2565.
- [6] Wang, B., Fang, B., Wang, Y., Liu, H., & Liu, Y. (2016). Power system transient stability assessment based on big data and the core vector machine. *IEEE Transactions on Smart Grid*, 7 (5), 2561-2570.
- [7] Zhou, D. Q., Annakkage, U. D., & Rajapakse, A. D. (2010). Online monitoring of voltage stability margin using an artificial neural network. *IEEE Transactions on Power Systems*, 25 (3), 1566-1574.
- [8] Khaledian, E., Pandey, S., Kundu, P., & Srivastava, A. K. (2020). Real-Time Synchrophasor Data Anomaly Detection and Classification Using Isolation Forest, KMeans, and LoOP. *IEEE Transactions on Smart Grid*, 12 (3), 2378-2388.

- [9] Shen, L., Du, H., Liu, S., Chen, S., Qiao, L., Liu, S.,... & Li, J. (2020). Real time outlier monitoring for power transformer fault diagnosis based on isolated forest. In IOP Conference Series: Materials Science and Engineering (Vol. 715, No. 1, p. 012033). IOP Publishing.
- [10] Tan, Y., Hu, C., Zhang, K., Zheng, K., Davis, E. A., & Park, J. S. (2020). LSTM-Based Anomaly Detection for Non-Linear Dynamical System. *IEEE Access*, 8, 103301-103308.
- [11] Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015, April). Long short term memory networks for anomaly detection in time series. In *Proceedings* (Vol. 89, pp. 89-94).
- [12] Shi, X., Qiu, R., Mi, T., He, X., & Zhu, Y. (2019). Adversarial feature learning of online monitoring data for operational risk assessment in distribution networks. *IEEE Transactions on Power Systems*, 35 (2), 975-985.
- [13] Zhang, L., Wang, G., & Giannakis, G. B. (2019). Real-time power system state estimation and forecasting via deep unrolled neural networks. *IEEE Transactions on Signal Processing*, 67 (15), 4069-4077.
- [14] Qiu, R. C., & Antonik, P. (2017). Smart grid using big data analytics: a random matrix theory approach. John Wiley & Sons.
- [15] Qiu, R., Chu, L., He, X., Ling, Z., & Liu, H. (2018). Spatiotemporal Big Data Analysis for Smart Grids Based on Random Matrix Theory. *Transportation and Power Grid in Smart Cities: Communication Networks and Services*, 591-633.
- [16] Ma, D., Hu, X., Zhang, H., Sun, Q., & Xie, X. (2019). A hierarchical event detection method based on spectral theory of multidimensional matrix for power system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.
- [17] Zhou, W., Sheng, J., Luo, H., Li, W., & He, X. (2018). Detection and localization of submodule open-circuit failures for modular multilevel converters with single ring theorem. *IEEE Transactions on Power Electronics*, 34 (4), 3729-3739.
- [18] He, X., Qiu, R. C., Ai, Q., Chu, L., Xu, X., & Ling, Z. (2016). Designing for situation awareness of future power grids: An indicator system based on linear eigenvalue statistics of large random matrices. *IEEE Access*, 4, 3557-3568.
- [19] Xu, X., He, X., Ai, Q., & Qiu, R. C. (2015). A correlation analysis method for power systems based on random matrix theory. *IEEE Transactions on smart grid*, 8 (4), 1811-1820.
- [20] Shi, X., Qiu, R., Ling, Z., Yang, F., Yang, H., & He, X. (2019). Spatio-temporal correlation analysis of online monitoring data for anomaly detection and location in distribution networks. *IEEE Transactions on Smart Grid*, 11 (2), 995-1006.
- [21] Chakhchoukh, Y., Vittal, V., & Heydt, G. T. (2013). PMU based state estimation by integrating correlation. *IEEE Transactions on Power Systems*, 29 (2), 617-626.
- [22] Yeo, J., & Papanicolaou, G. (2016). Random matrix approach to estimation of high-dimensional factor models. *arXiv preprint arXiv: 1611.05571*.
- [23] Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics*, 28 (3), 397-409.
- [24] Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208 (1), 23-42.
- [25] Lee, D., & Baldick, R. (2016). Load and wind power scenario generation through the generalized dynamic factor model. *IEEE Transactions on power Systems*, 32 (1), 400-410.
- [26] Bakdi, A., Bounoua, W., Mekhilef, S., & Halabi, L. M. (2019). Nonparametric Kullback-divergence-PCA for intelligent mismatch detection and power quality monitoring in grid-connected rooftop PV. *Energy*, 189, 116366.
- [27] Gupta, S., Waghmare, S., Kazi, F., Wagh, S., & Singh, N. (2016, March). Blackout risk analysis in smart grid WAMPAC system using KL divergence approach. In *2016 IEEE 6th International Conference on Power Systems (ICPS)* (pp. 1-6). IEEE.
- [28] Chen, H., Jiang, B., & Lu, N. (2018). An improved incipient fault detection method based on Kullback-Leibler divergence. *ISA transactions*, 79, 127-136.
- [29] Harmouche, J., Delpha, C., & Diallo, D. (2014). Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part I. *Signal processing*, 94, 278-287.
- [30] Harmouche, J., Delpha, C., & Diallo, D. (2015). Incipient fault detection and diagnosis based on Kullback-Leibler divergence using principal component analysis: Part II. *Signal Processing*, 109, 334-344.
- [31] Zhang, X., Delpha, C., & Diallo, D. (2020). Incipient fault detection and estimation based on Jensen-Shannon divergence in a data-driven approach. *Signal Processing*, 169, 107410.
- [32] De La Ree, J., Centeno, V., Thorp, J. S., & Phadke, A. G. (2010). Synchronized phasor measurement applications in power systems. *IEEE Transactions on smart grid*, 1 (1), 20-27.
- [33] Sarri, S., Zanni, L., Popovic, M., Le Boudec, J. Y., & Paolone, M. (2016). Performance assessment of linear state estimators using synchrophasor measurements. *IEEE Transactions on Instrumentation and Measurement*, 65 (3), 535-548.
- [34] Marchenko, V. A., & Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114 (4), 507-536.
- [35] Burda, Z., Jarosz, A., Nowak, M. A., & Snarska, M. (2010). A random matrix approach to VARMA processes. *New Journal of Physics*, 12 (7), 075036.
- [36] Ghanavati, G., Hines, P. D., & Lakoba, T. I. (2015). Identifying useful statistical indicators of proximity to instability in stochastic power systems. *IEEE Transactions on Power Systems*, 31 (2), 1360-1368.
- [37] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22 (1), 79-86.
- [38] Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.
- [39] Fuglede, B., & Topsoe, F. (2004, June). Jensen-Shannon divergence and Hilbert space embedding. In *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings*. (p. 31). IEEE.

- [40] Zimmerman, R. D., Murillo-Sánchez, C. E., & Thomas, R. J. (2010). MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26 (1), 12-19.
- [41] Marcelino, C. G., Almeida, P. E., Wanner, E. F., Baumann, M., Weil, M., Carvalho, L. M., & Miranda, V. (2018). Solving security constrained optimal power flow problems: a hybrid evolutionary approach. *Applied Intelligence*, 48 (10), 3672-3690.