Spatio-Temporal Correlation Analysis of Online Monitoring Data for Anomaly Detection in Distribution Networks

Xin Shi\textsuperscript{1}, Robert Qiu\textsuperscript{1,2}, Fellow, IEEE, Zenan Ling\textsuperscript{1}, Fan Yang\textsuperscript{1}, Xing He\textsuperscript{1}

Abstract—The online monitoring data in distribution networks contain rich information on the running states of the system. By leveraging the data, this paper proposes a spatio-temporal correlation analysis approach for anomaly detection in distribution networks. First, spatio-temporal matrix for each feeder in the distribution network is formulated and the spectrum of its covariance matrix is analyzed. The spectrum is complex and exhibits two aspects: 1) bulk, which arises from random noise or fluctuations and 2) spikes, which represents factors caused by anomaly signals or fault disturbances. Then by connecting the estimation of the number of factors to the limiting empirical spectral density of the covariance matrix of the modeled residual, the anomaly detection problem in distribution networks is formulated as the estimation of spatio-temporal parameters (i.e., the reducing factors $p$ and the autoregressive rate $b$), during which free random variable techniques are used. Furthermore, as for the estimated factors, we define and calculate a statistical magnitude for them as the spatial indicator to indicate the system state. Simultaneously, we use the autoregressive rate to measure the varieties of the temporal correlations of the data for tracking the system movement. Our approach is purely data driven and it is capable of discovering the anomalies in an early phase by exploring the variations of the spatio-temporal correlations of the data, which makes it practical for real applications. Case studies on the synthetic data verify the effectiveness of our approach and analyze the implications of the spatio-temporal parameters. Through the real-world online monitoring data, we further validate our approach and compare it with another spectrum analysis approach using the Marchenko-Pastur law. The results show that our approach is more accurate and it can serve as a primitive for analyzing the spatio-temporal data in a distribution network.

Index Terms—anomaly detection, distribution network, online monitoring data, spatio-temporal correlation analysis, free random variable

I. INTRODUCTION

This paper is driven by the need of anomaly detection using online monitoring data in a distribution network. The anomalies caused by some fault disturbances may present intermittent, asymmetric, and sporadic spikes, which are random in magnitude and could involve sporadic bursts as well, and exhibit complex, nonlinear, and dynamic characteristics \cite{1}. What’s more, with numerous branch lines and changeable network topology, it is questionable that traditional model-based approaches are capable of fully and accurately detecting the anomalies in the distribution network, because they are usually based on certain assumptions and simplifications.

With significant deployment of online monitoring devices in a distribution network, a large amount of data is collected through these devices. The massive amount of data contains rich information on the operating states of the distribution network. In recent years, there has been abundant research on utilizing the online monitoring data to realize anomaly detection. For example, in \cite{2}, a real-time fault detection and faulted line identification approach is proposed by using phasor measurement unit (PMU) data collected from a PMU-based monitoring system. In \cite{3}, dimensionality of synchrophasor data is studied and an algorithm based on dimension reduction analysis is developed for early event detection. In \cite{4}, time-series voltage data from online monitoring system is used to compute Lyapunov component to estimate voltage stability. In \cite{5}, by modeling streaming PMU data as random matrix flow, an algorithm based on multiple high dimensional covariance matrix tests is developed for system state estimation.

For a system with multiple measurement devices installed, the multi-dimensional data collected through them contains rich information on the system states. In terms of the data structure, spatio- (cross-) and temporal correlation (autocorrelation) should be considered when analyzing the system states. Then several open questions are raised, for example: 1) What is the spatio-temporal correlation of the data? 2) How to characterize or measure the spatio-temporal correlation of the data? 3) What is the relationship between the spatio-temporal correlation of the data and the state of the system? It is questionable for those conventional model-based methods to model the complex system, let alone addressing those new questions.

Factor models are important tools for reducing the dimensionality and extracting the relevant information in analyzing high-dimensional data. The estimation of high-dimensional factor models have been well studied in recent years \cite{6,7,8,9}. In \cite{10}, based on random matrix theory (RMT), a new approach to estimate high-dimensional factor models was proposed. It explores the spatial correlation of the data by estimating the factors of its covariance matrices. Simultaneously, it measures the temporal correlation of the data by analyzing its residuals obtained by subtracting...
principal components. The proposed approach can effectively capture the structural information of the data and outperforms other known methods. Considering the spectrum from the real data is complex and cannot be trivially dissected by simple techniques, it is meaningful to apply factor models for the real data analysis in the distribution network.

In this paper, based on exploring the spatio-temporal correlation of data amongst multiple-time-instant monitoring devices in a distribution network, a new algorithm for anomaly detection is designed. It leverages the spatio-temporal similarities amongst online monitoring data, and realizes anomaly detection by detecting the variations of the spatio-temporal correlation of the data. The main advantages of our approach can be summarized as follows: 1) It is a purely data-driven approach without requiring too much prior knowledge on the complex topology of the distribution network or model parameters, which eliminates potential detection errors caused by inaccurate system information or model parameters. 2) Our approach is sensitive to the variation of the spatio-temporal correlation of the online monitoring data and it is capable of detecting the anomalies in an early phase, because the correlation of the data will change immediately once an anomaly signal occurs. 3) It is theoretically justified that our approach is sensitive to the weak factors of the spectrum, which is good for improving the anomaly detection accuracy. 5) The approach can be conducted under both normal and abnormal operating conditions, which makes it practical for both online and offline analysis.

The rest of this paper is organized as follows. Section II analyzes the empirical spectrum distribution of the online monitoring data and the anomaly detection problem is formulated as the estimation of spatio-temporal parameters. In Section III, our anomaly detection approach based on spatio-temporal correlation analysis is proposed and discussed. Both synthetic data from IEEE-118 bus system and real-world online monitoring data from a grid are used to validate the effectiveness of our approach in Section IV. Conclusions are presented in Section V.

II. PROBLEM FORMULATION

In this section, the empirical spectrum distribution (ESD) of the covariance matrix of the online monitoring data in a distribution network under both normal and abnormal operating conditions is first analyzed. Then, the corresponding residual obtained by subtracting principal components is formulated and discussed. The anomaly detection problem in the distribution network is modeled as the problem of estimating the number of factors and the autoregressive rate of the residual.

A. Empirical Spectrum Distribution of the Online Monitoring Data

Marchenko-Pastur law (MP-law): Let $X = \{x_{i,j}\}$ be a $N \times T$ random matrix, whose entries are independent identically distributed (i.i.d.) variables with the mean $\mu(x) = 0$ and the variance $\sigma^2(x) < \infty$. The corresponding covariance matrix is defined as $\Sigma = \frac{1}{N} XX^H$. As $N, T \to \infty$ but $c = \frac{N}{T} \in (0, 1]$, according to the MP-law [1], the ESD of $\Sigma$ converges to the limit with probability density function (PDF)

$$f_{MP}(x) = \begin{cases} \frac{1}{2\pi\sigma^2} \sqrt{(b - x)(x - a)}, & a \leq x \leq b \\ 0, & \text{others} \end{cases}$$

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

We apply the MP-law for the online monitoring data from a distribution network. Assume matrix $R$ is the three-phase voltage measurements from 77 monitoring devices installed on the low voltage side of distribution transformers within one feeder. The data is sampled every 15 minutes and the sampling time is 7 days. We convert $R$ into the standard form $\hat{R}$ through

$$\hat{r}_{ij} = (r_{ij} - \mu(r_i)) \times \frac{\sigma(\hat{r}_i)}{\sigma(r_i)} + \mu(\hat{r}_i),$$

where $r_i = (r_{i1}, r_{i2}, \ldots)$, $\mu(\hat{r}_i) = 0$, and $\sigma(\hat{r}_i) = 1$. The covariance matrix of $\hat{R}$ is calculated and the corresponding ESD under both normal and abnormal operating states of the feeder is shown in Figure 1.

Fig. 1: The ESD of the covariance matrix of $\hat{R}$ and its comparison with the theoretical MP-law under both normal and abnormal operating conditions of the feeder.

From Figure 1, we can see that the spectrum of the covariance matrix of $\hat{R}$ (i.e., the blue bars) typically exhibits two aspects: bulk and spikes (i.e., the deviating eigenvalues). The bulk arises from random noise or fluctuations and the spikes are mainly caused by large disturbances or anomaly signals. No matter when the feeder operates in normal or abnormal state, the spectrum can not be fitted by the MP-law. To be noted is that the region of the bulk and the size of the spikes are different when the feeder operates in different states. Therefore, the spectrum can not be trivially dissected by using the MP-law and we must consider a new approach instead to analyze the spectrum for assessing the operating states of the feeder more accurately.

B. Residual Formulation and Discussion

From subsection II-A, the spectrum of the covariance matrix of $\hat{R}$ inspires us to decompose the real-world online monitoring data into systematic components (factors) and idiosyncratic noise (residuals). Assume matrix $R$ is of $N$ measurements and $T$ observations, thus a factor model regarding $R$ can be written as

$$R = LF + U,$$
where $L$ is an $N \times p$ matrix of factor loadings, $F$ is a $p \times T$ matrix of factors, $p$ is the number of factors, and $U$ is an $N \times T$ matrix of residuals. For the real-world online monitoring data, the ESD of the covariance matrix of the residuals does not fit to the MP-law, no matter how many factors are removed, as is shown in Figure 2.

In order to estimate the spectrum of the real residuals, we connect the estimation of the number of factors to the limiting ESD of the covariance matrix of the residuals from the real-world online monitoring data. The latter is controlled by the number of the covariance matrix of residuals from the real-world online monitoring data. The former is approximated by the number of the residuals constructed by removing $p$ factors from the real data, according to work [13], principal components can approximately mimic all true factors. Considering the factor data, the cross-covariance matrix characterized by parameter $b$ data, the cross-covariance matrix $A^T_1A_1$ approximately mimic all true factors. Considering the factor data, the cross-covariance matrix characterized by parameter $b$ and the autoregressive rate $\rho$.

### III. Anomaly Detection in Distribution Networks

Based on the discussions above, by using the online monitoring data, we propose a spatio-temporal correlation analysis approach for anomaly detection in distribution networks. In this section, the estimation method of factor models built in Section II is described in detail, in which, free random variable techniques proposed in Burda’s work [12] are used to calculate the modeled spectral density. Then, specific procedures of our proposed anomaly detection approach are given and characteristics regarding them are further analyzed. Finally, more discussions about our approach are carried out.

#### A. Factor Model Estimation

From Section II we can estimate $p$ and $b$ by minimizing the distance between the ESD of the covariance matrix of the residuals from the real-world online monitoring data and the limiting eigenvalue density of $\Sigma$, which is stated as

$$\{\hat{p}, \hat{b}\} = \arg \min_{p,b} D(\rho_{real}(p), \rho_{model}(b)),$$

where $\rho_{real}(p)$ represents the ESD of the covariance matrix of the residuals constructed by removing $p$ factors from the real data, $\rho_{model}(b)$ is the limiting spectral density of the modeled covariance matrix characterized by parameter $b$, and $D$ is the spectral distance measure.

In order to obtain $\rho_{real}(p)$, we firstly calculate the residuals by removing $p$ largest principal components from the real-world online monitoring data. Because for high dimensional data, according to work [13], principal components can approximately mimic all true factors. Considering the factor model in Equation (5), the $p$-level residual $\hat{U}^{(p)}$ is calculated by

$$\hat{U}^{(p)} = R - \hat{L}^{(p)} \hat{F}^{(p)},$$

where $\hat{F}^{(p)}$ is an $p \times T$ matrix composed of $p$ principal components from correlation matrix of $R$, $\hat{L}^{(p)}$ is an $N \times p$ matrix estimated by multivariate least squares regression, namely

$$\hat{L}^{(p)} = R * inv(\hat{F}^{(p)}),$$

where $inv$ denotes the pseudo-inverse method. We can calculate the covariance matrix of $\hat{U}^{(p)}$ by

$$\Sigma^{(p)} = \frac{1}{T} \hat{U}^{(p)} \hat{U}^{(p)^T},$$

and $\hat{\rho}_{real}(p)$ is the bulk spectrum of $\Sigma^{(p)}$.

Then we calculate $\rho_{model}(b)$ by using free random variable techniques proposed in [12]. For the autoregressive model $U$, by using the moments’ generating function $M \equiv M_{\Sigma_{model}}(z)$
and its inverse relation to N-transform \[^{10}\], we can derive the following polynomial equation
\[
\begin{align*}
\dot{a}^4 c^2 M^4 + 2a^2 c(-1 + b^2)z + a^2 c M^3 + (1 - b^2)^2 z^2 \\
-2a^2 c(1 + b^2)z + (c^2 - 1)a^4 M^2 - 2a^4 M - a^4 = 0
\end{align*}
\]
where \( a = \sqrt{1 - b^2} \) and \( c = \frac{\zeta}{\eta} \). See Appendix for the derivation details. Thus, we can obtain \( M_{\text{model}}(z) \) by solving Equation (8). Then we can calculate the Stieltjes transform \( G_{\text{model}}(z) \) through
\[
G_{\text{model}}(z) = \frac{M_{\text{model}}(z) + 1}{z} \quad \text{for } |z| \neq 0.
\]
The mean spectral density \( \rho_{\text{model}}(b) \) can be reconstructed from the imaginary part of \( G_{\text{model}}(z) \) as
\[
\rho_{\text{model}}(b) = -\frac{1}{\pi} \lim_{\epsilon \to 0^+} \Im G_{\text{model}}(\lambda + i\epsilon).
\]
The spectral distance measure \( D \) must be sensitive to the information disparity in \( \rho_{\text{real}}(p) \) and \( \rho_{\text{model}}(b) \). Here, we use Jensen-Shannon divergence, which is defined as
\[
D(\rho_{\text{real}} || \rho_{\text{model}}) = \frac{1}{2} \sum_i \log \frac{\rho_{\text{real}}(i)}{\rho_{\text{model}}(i)} \\
+ \frac{1}{2} \sum_i \log \frac{\rho_{\text{model}}(i)}{\rho(i)},
\]
where \( \rho = \rho_{\text{real}} + \rho_{\text{model}} \). We can see that \( D(\rho_{\text{real}} || \rho_{\text{model}}) \) becomes smaller as \( \rho_{\text{real}} \) approaches \( \rho_{\text{model}} \), and vice versa. Therefore, we can match \( \rho_{\text{model}}(b) \) to \( \rho_{\text{real}}(p) \) by minimizing \( D \), through which the optimal parameter set \((\hat{p}, \hat{b})\) is obtained.

B. spatio-temporal Correlation Analysis Approach for Anomaly Detection

From Section \[^{11}\] we know that the number of removed factors \( p \) and the autoregressive rate \( b \) can be used to indicate the variations of spatial and temporal correlation of the original data. Based on the estimated parameter \( \hat{p} \), we design a statistical indicator for the top \( \hat{p} \) eigenvalues of the covariance matrix of the online monitoring data to measure the spatial correlation, which is defined as
\[
N_{\phi} = \sum_{i=1}^{\hat{p}} \phi(\lambda_i),
\]
where \( \lambda_1 > \lambda_2 > \cdots > \lambda_{\hat{p}} \), and \( \phi(\cdot) \) is a test function that makes a linear or nonlinear mapping for the eigenvalues \( \lambda_i \). Details about the test function can be found in our previous work \[^{12}\]. As an indicator to measure the spatial correlation of the data, \( N_{\phi} \) is more accurate and robust than the estimated number of removed factors \( \hat{p} \), because the latter is susceptible to the weak factors caused by some normal fluctuations or disturbances.

Meanwhile, the estimated parameter \( \hat{b} \) is directly used to measure the temporal correlation of the real data. It can effectively emulate the variation of the temporal correlation of the data, and provide an insight into system dynamics. To be mentioned is that, if the residual processes of the real online monitoring data are not auto-correlated, \( \hat{b} \) will be far different from the true value.

In real applications, we can move a certain length window on the collected data set \( D \) at continuous sampling times and the last sampling time is the current time, which enables us to track the variations of spatio-temporal correlations of the online monitoring data in real-time. For example, at the sampling time \( t_j \), the obtained raw data matrix \( R(t_j) \in \mathbb{R}^{N \times T} \) is formulated by
\[
R(t_j) = (d(t_j-T+1), d(t_j-T+2), \cdots, d(t_j)),
\]
where \( d(t_k) = (d_1, d_2, \cdots, d_N)^T \) for \( t_j-T+1 \leq t_k \leq t_j \) is the sampling data at time \( t_k \). Thus, \( N_{\phi} \) and \( \hat{b} \) are produced for each sampling time.

Based on the works above, an anomaly detection approach based on spatio-temporal correlation analysis is designed. The fundamental steps are given as follows. Steps 4 \~ 8 are conducted for calculating the ESD of the covariance matrix of the real residuals and steps 9 \~ 10 are for calculating the limiting spectral density of the built autoregressive model, and the spectral distance of them are calculated and saved in each iteration shown in step 11. We can obtain the optimal parameter set by getting those corresponding the minimum spectral distance for each sampling time in step 12. During the steps above, \( N_{\phi} \) and \( \hat{b} \) are calculated as indicators to assess the system states.

### Steps of spatio-temporal Correlation Analysis for Anomaly Detection in Distribution Networks

1. For each feeder, construct a spatio-temporal data set \( D \) by arranging three-phase voltage measurements of all public transformers within the feeder in a series of time.
2. At each sampling time \( t_j \);
3. Obtain the original data matrix \( R(t_j) \in \mathbb{R}^{N \times T} \) by moving a \( N \times T \) window on \( D \);
4. For the number of removing factors \( p = 1, 2, \cdots \);
5. Get the real residual \( U^{(p)}(t_j) \) through Equation \[^{3}\];
6. Normalize \( U^{(p)}(t_j) \) into the standard form through Equation \[^{5}\];
7. Calculate the covariance matrix of the real residual, i.e., \( X_{\text{real}}^{(p)}(t_j) \);
8. Obtain \( \rho_{\text{real}}(t_j) \) through getting the bulk spectrum of \( X_{\text{real}}^{(p)}(t_j) \);
9. For the autoregressive rate \( b \sim U[0, 1] \);
10. Obtain \( \rho_{\text{model}}(t_j) \) through Equation \[^{8}\], \[^{9}\] and \[^{10}\];
11. Calculate the spectral distance \( D(\rho_{\text{real}}(t_j) || \rho_{\text{model}}(t_j)) \) through Equation \[^{11}\] and save them;
12. Obtain the optimal parameter set \((\hat{p}(t_j), \hat{b}(t_j))\) through Equation \[^{12}\];
13. Calculate the spatial indicator \( N_{\phi}(t_j) \) through Equation \[^{13}\];
14. Draw the \( N_{\phi} - t \) curve and \( \hat{b} - t \) curve for each feeder in a series of time \( t \) to realize anomaly detection.

The anomaly detection approach proposed is driven by the online monitoring data in distribution networks, and based on high-dimensional statistical theories. It reveals the variations of spatio-temporal correlations of the input data when anomalies happen and can detect the weak anomalies occurring in the system through controlling both the number of factors \( p \) and the autoregressive rate \( b \). Compared with traditional model-based methods, our approach is purely driven by data and does not require too much knowledge about the complex topology of the complex distribution network. It is theoretically robust against small random fluctuations and measuring errors in the system, which can help to improve anomaly detection accuracy.
and reduce the potential false detection probability. What’s more, our proposed approach is practical for real-time anomaly detection by moving a certain length window method.

C. More Discussions About the Proposed Approach

In Section II-B we assume that the cross-correlations of the real residuals can be effectively eliminated by removing \( p \) factors. However, for the real-world online monitoring data in a distribution network, whether this assumption holds is questionable. Meanwhile, the factor model estimation method in Section III-A is suitable for large-dimensional data matrices in theory. However, in practice, the dimensions of the online monitoring data for some feeders in a distribution network are often moderate, such as hundreds or less. Here, we will check how well the built autoregressive model can fit the real residuals, the result of which are shown in Figure 3.

Fig. 3: Fit of the built autoregressive model to the real data. The data was sampled through the online monitoring devices installed on the low voltage side of the public transformers within one feeder. It was sampled when the feeder operated under both normal and abnormal states, respectively sampling 672 times. The dimension of the data is 231. The built model with estimated \( \hat{p} \) and \( \hat{b} \) fits the residuals of the real data very well. For comparison, MP-law for the real residual matrix is plotted.

Figure 3(a) and 3(b) respectively show the fitted result of our built autoregressive model to the real residuals under both normal and abnormal operating states. We can see that, with optimal estimated parameter set \( \hat{p}, \hat{b} \), our built model can fit the real residuals well no matter whether the feeder runs in normal or abnormal states. In contrast, the MP-law does not fit the real residuals well no matter whether the feeder runs in normal or abnormal operating states. We can see that, with estimated \( \hat{p} \) and \( \hat{b} \) fits the residuals of the real data very well. For comparison, MP-law for the real residual matrix is plotted.

IV. CASE STUDIES

In this section, our proposed anomaly detection approach is validated through using both the synthetic data from the IEEE-118 bus system and real-world online monitoring data in a distribution network. Four cases under different scenarios are designed: 1) The first three cases, leveraging the synthetic data, test the effectiveness of our approach for anomaly detection and analyze the implications of parameter \( p \) and \( b \). 2) The last case, using the real-world online monitoring data, validate our approach and compare it with the spectrum analysis approach using MP-law in the work [14][15].

A. Case Study with Synthetic Data

The synthetic data was sampled from the simulation results of the standard IEEE-118 bus system [16], with a sampling rate of 50 Hz. During the simulations, a sudden change of the active load at one bus was considered an anomaly signal and little white noise was introduced to represent random fluctuations and system errors.

1) Case Study on the Effectiveness of Our Approach for Anomaly Detection: In this case, the synthetic data set contains 118 voltage measurement variables with sampling 1000 times. In order to test the effectiveness of the proposed anomaly detection approach, an assumed anomaly signal was set for bus 20 and others stayed unchanged, which is shown in Table I. In the experiment, the size of the split-window was \( 118 \times 200 \) and a little random autoregressive noise with a decaying rate \( b = 0.5 \) was introduced into each split-window to represent the auto-correlations of the residuals. The experiment was repeated for 20 times and the results were averaged. Here, we chose the likely-hood radio function (LRF) as the test function.

TABLE I: Assumed Signals for Load of Bus 20 in Case 1.

| Bus | Sampling Time | Active Power (MW) |
|-----|---------------|------------------|
| 20  | \( t_s = 1 \sim 500 \) | 20 |
| 20  | \( t_s = 501 \sim 1000 \) | 80 |
| Others | \( t_s = 1 \sim 1000 \) | Unchanged |

Fig. 4: Effectiveness of our approach for anomaly detection in case 1.

(a) \( \mathcal{N}_\phi - t \) curve  
(b) \( \hat{b} - t \) curve

Figure 4(a) and Figure 4(b) show the anomaly detection result respectively from the perspective of space and time. It is noted that the \( \mathcal{N}_\phi - t \) curve and \( \hat{b} - t \) curve begin at \( t_s = 200 \), because the initial split-window includes 199 times of historical sampling and the present sampling data. It is also important to note that the index number starts from 0 in Python. The detection processes are shown as follows:

I. During \( t_s = 200 \sim 500 \), \( \mathcal{N}_\phi \) and \( \hat{b} \) remains almost constant, which means the system operates in normal state.
and the spatio-temporal correlations of the data stay almost unchanged. As is shown in Figure 5(a), the ESD of the covariance matrix of the residuals can be fitted well by the built model with $\hat{p} = 0, \hat{b} = 0.4752$.

II. From $t_s = 501 \sim 700$, $N_{\phi}$ and $\hat{b}$ change dramatically, which denotes an anomaly signal which occurs at $t_s = 501$, where the spatio-temporal correlations of the data begin to change. The curves are almost inverted U-shape and the delay lag of the anomaly signal to the spatio-temporal indicators is equal to the split-window’s width. Figure 5(b) shows the ESD of the covariance matrix of the residuals fits the probability density function (PDF) of our built model with $\hat{p} = 1, \hat{b} = 0.6435$.

III. At $t_s = 701$, $N_{\phi}$ and $\hat{b}$ recover to their initial values and remain constant afterwards, which indicates the system has returned to normal.

From the analyzes above, we can conclude that, by exploring the variations of the spatio-temporal correlations of the data, our proposed approach is able to detect anomaly signal effectively. It is noted that the estimated $\hat{p}$ and $\hat{b}$ are different when the system operates in different states. In the following cases, we will discuss their implications.

2) Case Study on the Implication of $p$: In our experiments, we found the estimated optimal parameter $\hat{p}$ is related to the system states. In this case, we will further discuss what drives it. In order to validate the relations of $p$ and the number of anomaly events occurred in the system, different number of assumed anomaly signals were set, which is shown in Table II. All the other parameters were set the same as in case 1) and the experiment was also repeated for 20 times with results being averaged.

TABLE II: Assumed Signals for Active Load of Bus 20, 30 and 60 in Case 2.

| Bus | Sampling Time | Active Power(MW) |
|-----|---------------|------------------|
| 20  | $t_s = 1 \sim 500$ | 20 |
|     | $t_s = 501 \sim 1000$ | 80 |
| 30  | $t_s = 1 \sim 550$ | 20 |
|     | $t_s = 551 \sim 1000$ | 80 |
| 60  | $t_s = 1 \sim 600$ | 20 |
|     | $t_s = 601 \sim 1000$ | 80 |
| Others | $t_s = 1 \sim 1000$ | Unchanged |

The generated $\hat{p} - t$ curve by continuously moving windows is shown in Figure 6. Interpretations of $\hat{p}$ are stated as follows:

I. From $t_s = 200$ to $t_s = 500$, The estimated $\hat{p}$ remains almost at 7. During our experiment, we observed that no strong factors appeared during this period. The most likely explanation is that our approach is sensitive to weak factors caused by some normal fluctuations or disturbances and can identify them effectively.

II. From $t_s = 501$ to $t_s = 550$, one strong factor is observed during our experiment and the estimated $\hat{p}$ is between 1 and 2. This indicates that one anomaly event occurred, which is consistent with our assumed result in Table II. Then, from $t_s = 551$ to $t_s = 600$, the number of estimated factors is between 2 and 3, which is also consistent with the number of assumed anomaly signals contained in the window. A similar analysis result can be concluded during $t_s = 601 \sim 650$.

III. From $t_s = 651$ to $t_s = 800$, the value of $\hat{p}$ decreases 1 every other 50 sampling times. This is because the width of our moving window is 200 and the number of anomaly events contained in the window decreases 1 every 50 sampling times with its movement.

IV. At $t_s = 801$, the value of $\hat{p}$ returns to nearly 7 and remains afterwards, which indicates the system has returned to the normal state.

From the analyzes above, we can conclude that $p$ is related to the number of anomaly events occurring in the system. Theoretically, the estimated $\hat{p}$ is equal to the number of anomaly events. However, It is noted that our approach is able to identify several weak factors even if the system operates in normal state. Therefore, in our approach, we designed a statistical magnitude regarding the estimated $\hat{p}$ factors instead of simply using $\hat{p}$ as the spatial indicator to detect the anomalies.

3) Case Study on the Implication of $b$: From Case 1), we see that the parameter $b$ can be used to reflect the system states from the perspective of time. In this case, we will further discuss the implication of $b$ and what drives it. In order to test the relations of $b$ and the degree of abnormality in the system, assumed anomaly signals with different scale degrees were set for bus 20, which is shown in Table III. All the other parameters were set the same as in Case 1). The experiment was also repeated for 20 times with results being averaged.

The $\hat{b} - t$ curve generated by continuously moving windows is shown in Figure 7. Interpretations of $\hat{b}$ are described as follows:
TABLE III: Assumed Signals With Different Scale Degrees
Set for Bus 20 in Case 3.

| Bus | Sampling Time | Active Power(MW) |
|-----|---------------|-----------------|
| 20  | $t_s = 1 \sim 500$ | 20 |
|     | $t_s = 501 \sim 1000$ | 60/90/120 |
| Others | $t_s = 1 \sim 1000$ | Unchanged |

I. From $t_s = 200$ to $t_s = 500$, the estimated $\hat{b}$ remains almost constant, which indicates the system operates in a normal state.

II. From $t_s = 501$ to $t_s = 700$, the $\hat{b} - t$ curves are almost inverted U-shaped and reach their global maximum at $t_s = 600$. It is noted that the estimated $\hat{b}$ corresponding to the assumed signal of active power from 20 to 120 has the largest maximum value and that from 20 to 60 has the smallest value. It shows that the higher the abnormality degree is, the larger the estimated value of $\hat{b}$ is.

III. At $t_s = 701$, the estimated $\hat{b}$ recover to their initial values and remain constant afterwards, which indicate that the system has returned to the normal state.

From the analyzes above, it can be concluded that $b$ is related to the degree of abnormality in the system. The higher the abnormality degree of the system is, the greater the deviation of the estimated $\hat{b}$ from its normal standard is. It reflects essential information on the system movement, which explains why it can be used as a temporal indicator to measure the system state in our approach.

B. Case Study with Real-World Online Monitoring Data

In this subsection, the online monitoring data obtained from a real-world distribution network is used to test our proposed approach. The feeder in the distribution network consists of different levels of branch lines and substations with 77 distribution transformers in total. On the low voltage side of each distribution transformer, one online monitoring device is installed, through which we can obtain many types of measurement variables such as three-phase voltages, three-phase currents, active load, etc. The data was sampled every 15 minutes and the sampling time was from 2017/3/15 00:00:00 to 2017/4/6 23:45:00. Here, we chose three-phase voltages as the elements to formulate the data set of size $231 \times 2208$. Anomaly time and fault type for the feeder were recorded, which are shown in Figure 8. It can be seen that the anomaly begins at 2017/3/24 12:05:00 and the anomaly type is three-phase voltage imbalance.

In the experiment, we used a $231 \times 672$ split-window, and moved it at each sampling time, which enabled us to track the variation of the spatio-temporal correlation of the data. With moving windows, the generated spatio-temporal indicators for each sampling time is shown in Figure 9. In the figure, the red dashed line marks the beginning time of the anomalies. It is noted that the indicator curves begin at 2017/3/7 23:45:00, because our initial split-window includes 7 days’ sampling time.

From the spatial/temporal indicator $- t$ curves, we can realize anomaly detection for the feeder as follows:

I. From 2017/3/15 00:00:00 to 2017/3/24 08:45:00, the values of the spatio-temporal indicators remain almost constant, which indicates the feeder operates in a normal state. As is shown in Figure 10(a), the ESD of covariance matrix of residuals can be fitted well by the built model with $\hat{p} = 18, \hat{b} = 0.4158$ when the feeder runs in normal state, but it does not fit the MP-law.

II. At 2017/3/24 09:00:00, the values of the spatio-temporal indicators begin to change dramatically, which indicate anomaly signals occurring and the operating state of the feeder beginning to change. Comparing with the recorded anomaly time, we can see that the anomaly can be detected earlier than the recorded time. Furthermore, from 2017/3/24 09:00:00 to 2017/4/2 22:30:00, the spatial/temporal indicator $- t$ curves are almost inverted U-shaped, which is consistent with our simulation result in Section IV-B. Figure 10(b) shows, in abnormal state, the ESD of covariance matrix of residuals can be fitted well by our built model with $\hat{p} = 37, \hat{b} = 0.5544$.

III. From 2017/4/2 22:45:00, the calculated spatio-temporal indicators recover their initial values and remain afterwards,
which indicates the feeder has returned to the normal state.

Fig. 10: The ESD of the covariance matrix of the residuals from the real-world online monitoring data can be fitted very well by our built model with the estimated $\hat{p}$ and $\hat{b}$, while can not be fitted by the MP-law. The optimal parameters estimated are different when the feeder runs in different states.

The analyzes above show that our approach is capable of detecting anomalies in an early phase in distribution networks. In real-time analysis, we may not observe the completely inverted U-shape curve and we can use the variance ratio of the spatial/temporal indicator $-t$ curves as the basis to judge whether an anomaly is occurring. It is noted that the delay lag of the anomaly signals to the spatio-temporal indicators is not equal to the split-window’s width, which is not in accord with our simulation results. The reason is that the anomaly events in the real data usually last for a period of time and we can estimate that through subtracting the split-window’s width from the width of the inverted ‘U’.

Furthermore, we compare our approach with the spectrum analysis (SPA) method using the MP-law. SPA has been well studied in the work [14][15]. The difference of SPA to our approach is that it uses the MP-law for analyzing the spectrum of the covariance matrix of the data and calculates a statistical indicator for it. In order to compare the anomaly detection performances of the two approaches, we normalized the calculated indicators into $[0,1]$, which is shown in Figure [11]

Fig. 11: Comparison of anomaly detection performance between our approach and SPA using the MP-law.

From the figure, we can see that both our approach (spatial indicator (SI)/temporal indicator (TI) $-t$ curves) and SPA are capable of detecting the anomalies effectively. However, the SI/TI $-t$ curves has higher variance ratio than SPA using the MP-law, which indicates our approach is more sensitive to the anomalies in the distribution network.

The reason is that the spectrum from the real data is complex and it is not accurate to trivially dissect it by using the MP-law adopted in SPA. In contrast, our approach uses factor model for the spectrum analysis of the real data, and it estimates the factor model by controlling both the number of factors and the autoregressive rate of the residuals, making it capable of tracking the data behavior more accurately.

V. CONCLUSION

By analyzing the structure information of the online monitoring data in distribution networks, this paper proposes a spatio-temporal correlation analysis approach to realize anomaly detection. It is capable of detecting the anomalies in an early phase by exploring the variations of the spatio-temporal correlations of the data. The spatial and temporal indicators we designed are able to indicate the data behaviour accurately. Our approach is purely data-driven and it does not require prior knowledge on the complex topology or modeling of the distribution network. It is robust to random fluctuations in the data and sensitive to the weak factors in the spectrum, which help improve the anomaly detection accuracy. Furthermore, our approach can be conducted both online and offline and has fast computational speed, which makes it suitable for real-time applications. The cases studies with synthetic data verify the effectiveness of our approach and offer explanations on the estimated spatio-temporal parameters. Through the real-world online monitoring data from a grid, we validate the effectiveness of our approach and compare it with the spectrum analysis approach by using the MP-law. The results show that our approach is more sensitive to anomaly detection and is more practical for analyzing the spatio-temporal data in distribution networks.

APPENDIX

Definition 1 The Green’s Function (or Stieltjes Transform).

$$G_H(z) = \int \frac{\rho_H(\lambda)}{z - \lambda} d\lambda, \quad (14)$$

where $\rho_H(\lambda)$ is the spectral density of the random matrix $H$, which can be reconstructed from the Green’s Function by calculating its imaginary part

$$\rho_H(\lambda) = -\frac{1}{\pi} \lim_{\varepsilon \to 0^+} \Im G_{\Sigma, \text{model}}(\lambda + i\varepsilon). \quad (15)$$

Definition 2 Moment.
The $n$-th moment of $\rho_H(\lambda)$ is defined as

$$m_{H,n} = \int \rho_H(\lambda)\lambda^n d\lambda. \quad (16)$$

Definition 3 Moment generating function.

$$G_H(z) = \sum_{n=0}^{\infty} \frac{m_{H,n}}{z^{n+1}}. \quad (17)$$

$$M_H(z) = \sum_{n=1}^{\infty} \frac{m_{H,n}}{z^{n+1}}. \quad (18)$$
Thus, the relation between \( M_H(z) \) and \( G_H(z) \) can be derived through Equation (17) and (18)
\[
M_H(z) = zG_H(z) - 1. \tag{19}
\]

**Definition 4 N-transform.**

\( N_H(z) \) is the inverse transform of \( M_H(z) \)
\[
M_H(N_H(z)) = N_H(M_H(z)) = z. \tag{20}
\]

For the empirical covariance matrix \( \Sigma = \frac{1}{T} U U^T = \frac{1}{T} A_N^{-1/2} S B_T S^T A_N^{1/2} \), the N-transform of \( \Sigma \) can be derived as
\[
N_\Sigma(z) = \frac{1}{T} A_N^{-1/2} S B_T S^T A_N^{1/2} = \frac{1}{T} \frac{1}{z} \frac{N_\Sigma(z)}{N_\Sigma(z) + 1} N_A(z).
\]

Considering \( M = M_\Sigma(z) \) and its inverse relation to N-transform, we can obtain
\[
z = rM N_B(rM) N_A(rM). \tag{22}
\]

In Section II-B we assume the residual \( U_{it} = U_{it} - \frac{1}{T} \sum_{i=1}^T U_{it} \) follows the autoregressive model, namely \( U_{it} = b_i \delta(t) \) \(|b| < 1\). For the autoregressive process, the cross-correlation matrix \( A_N = I_N \). Then \( N_A(z) = N_I(z) = 1 + 1/z \). Combining Equation (22), we can obtain
\[
z = r(1 + M) N_B(rM) \\frac{z}{r(1 + M)}.
\]

For the auto-covariance matrix \( B_T \) of vector autoregressive process, \( B_{T,t} = b_i^{t-l} \). By using Fourier-transform, the moment generation function of \( B_T \) is obtained by
\[
M_{B_T}(z) = \frac{1}{\sqrt{1 - z^2}} \sqrt{1 - (1 + b_i^2)z}. \tag{24}
\]

Thus, by solving Equation (23), we can obtain the object polynomial in Equation (5).

**REFERENCES**

[1] M. R. Jaafari Mousavi, “Underground distribution cable incipient fault diagnosis system,” Ph.D. dissertation, Texas A&M University, 2007.

[2] M. Pignati, L. Zanni, P. Romano, R. Cherkauoi, and M. Paolone, “Fault detection and faulted line identification in active distribution networks using synchrophasors-based real-time state estimation,” IEEE Transactions on Power Delivery, vol. 32, no. 1, pp. 381–392, 2017.

[3] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” IEEE Transactions on Power Systems, vol. 29, no. 6, pp. 2784–2794, 2014.

[4] S. Dasgupta, M. Paramasivam, U. Vaidya, and V. Aijjarapu, “Real-time monitoring of short-term voltage stability using pmu data,” IEEE Transactions on Power Systems, vol. 28, no. 4, pp. 3702–3711, 2013.

[5] L. Chu, R. C. Qiu, X. He, Z. Ling, and Y. Liu, “Massive streaming pmu data modeling and analytics in smart grid state evaluation based on multiple high-dimensional covariance tests,” IEEE Transactions on Big Data, vol. PP, no. 99, pp. 1–1, 2016.

[6] S. C. Ahn and A. R. Horenstein, “Eigenvalue ratio test for the number of factors,” Econometrica, vol. 81, no. 3, pp. 1203–1227, 2013.

[7] J. Bai and S. Ng, “Principal components estimation and identification of static factors,” Journal of Econometrics, vol. 176, no. 1, pp. 18–29, 2013.

[8] M. Harding, “Estimating the number of factors in large dimensional factor models,” 2013.

[9] M. Pelger, “Large-dimensional factor modeling based on high-frequency observations,” Social Science Electronic Publishing, 2015.

[10] J. Yeo and G. Papanicolaou, “Random matrix approach to estimation of high-dimensional factor models,” Papers, 2016.

[11] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” Mathematics of the USSR-Sbornik, vol. 1, no. 4, p. 457, 1967.

[12] Z. Burda, A. Jarosz, M. A. Nowak, and M. Snarska, “A random matrix approach to varma processes,” vol. 12, no. 1002.0934, pp. 1653–1655, 2010.

[13] J. H. Stock and M. W. Watson, “Forecasting using principal components from a large number of predictors,” Publications of the American Statistical Association, vol. 97, no. 460, pp. 1167–1179, 2002.

[14] X. Shi, R. Qiu, X. He, L. Chu, and Z. Ling, “Incipient fault detection and location in distribution networks: A data-driven approach,” arXiv preprint arXiv:1801.01669, 2018.

[15] 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 transactions on smart Grid, vol. 8, no. 2, pp. 674–686, 2017.

[16] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “Matpower: Steady-state operations, planning, and analysis tools for power systems research and education,” IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 12–19, Feb 2011.