International Journal of Engineering and Technology Innovation, vol. 15, no. 3, 2025, pp. 286-299

A Novel Method for Detecting Voltage Anomaly in Distribution Networks Based on Improved Standard Deviation Filters

Chenggong Chen[1], Zhiqiang Xiang[2,*], Aiyuan Li[1], Ling Luo[1], Tao Tang[2], Yulong Chen[2]

[1]State Grid Zhuzhou Electric Power Co., Ltd., Zhuzhou, China

[2]School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha, China

Received 03 December 2024; received in revised form 29 March 2025; accepted 07 April 2025

DOI: https://doi.org/10.46604/ijeti.2024.14580

Abstract

Accurate voltage anomaly detection is the prerequisite for the reliable operation of the distribution network. However, the traditional detection methods are prone to missed and false alarms. In practice, the distribution of phase voltage difference satisfies a normal distribution during normal operation and deviates from the distribution during faults. This paper proposes a novel method for voltage anomaly detection in distribution networks based on an improved standard deviation filter. The proposed method identifies an anomaly by evaluating the dispersion degree based on the mean and standard deviation of the phase voltage difference dataset. The short-term cycle, long-term cycle, and weighting coefficient are adopted rationally, thus large data storage requirements and repeated calculations can be avoided. Compared with the clustering and the isolation forest algorithms, the proposed method can identify voltage anomalies more accurately. The reliability of the proposed method is verified by on-site data.

Keywords: distribution network, voltage anomaly detection, standard deviation filter, mean

1. Introduction

The distribution network is essential in the power system to connect the generation side and the consumer side, and its operational reliability is closely related to the national economy [1-3]. Voltage anomaly is one of the main factors affecting the power system's reliability. If the voltage anomaly causes an accident tripping, it will affect the normal operation of the power system and may even result in irreversible damage to the electrical equipment [4-5]. To improve the safety and reliability of power system operation, accurate monitoring of voltage anomalies is particularly critical [6-7]. Currently, the methods for detecting voltage anomaly in distribution networks can generally be categorized into the voltage over-limit setting method [8-13] and the abnormal data identification method based on deep learning and clustering algorithms [14-22].

The principle of the voltage over-limit value setting method is relatively simple. When the measured voltage value exceeds the set voltage over-limit value, an alarm signal is sent. However, setting the voltage over-limit value reasonably remains challenging. Firstly, the detection accuracy and sensitivity can hardly be simultaneously combined. A lower alarm threshold ensures detection sensitivity, whereas voltage fluctuations in normal operation may result in false alarms [8-9]. A higher alarm threshold improves detection accuracy but significantly reduces detection sensitivity [10-11]. Secondly, the unified voltage over limit value cannot be applied to certain environments. For example, a three-phase imbalance exists in

* Corresponding author. E-mail address: 18473270892@163.com

certain bus voltages during normal operation, and a unified fixed value for the over-limit alarm will lead to false alarms or missed detections. Personalized settings and fixed value settings are cumbersome and cannot adapt to voltage changes caused by changes in the mode of operation. [12-13].

The abnormal data identification method based on deep learning and clustering algorithms can detect voltage anomaly by building a time series prediction model and comparing the error between the real and predicted values [14]. However, the method has difficulty in accurately distinguishing between "abnormal" and "normal" values statistically. To prevent frequent false alarms of voltage anomaly, the detection sensitivity is usually compromised. Such a situation makes it impossible to accurately alarm when voltage anomaly characteristics are not apparent, resulting in faults often being undiscovered and untreated in time [15].

Liu et al. [16] proposed a method for voltage anomaly detection in distribution networks based on an improved K-Means clustering algorithm (K-Means). This method utilizes the improved elbow method and silhouette coefficient algorithm to reduce computation time and improve the accuracy and efficiency of online monitoring. It builds upon the foundation of K-Means. Zhang [17] proposed a method of extracting voltage anomaly features based on the K-Means. This method performs well in voltage signal acquisition, yet the accuracy of voltage anomaly monitoring is lower than expected. Although these methods can detect abnormal voltage data to a certain degree, sometimes the distribution network data is irregular, which may lead to missed abnormal data [18].

Kuang et al. [19] and Ke et al. [20] proposed a neural network-based voltage anomaly detection method in distribution networks. This method constructs a detection model using historical data and detects anomalies by comparing the difference between predicted values and actual values. However, the method fails when the voltage anomaly is not characterized. Guo et al. [21] is based on support vector machines for data anomaly detection, which has a narrow scope of application and a high rate of missed detection. Similarly, this method may fail when the fault characteristics are elusive. A dynamic filling algorithm for missing data in the distribution network was proposed by Mei et al. [22], enabling accurate and automatic selection of anomaly identification thresholds and effectively detecting abnormal data in distribution networks.

To improve the detection accuracy, the standard deviation filter based on the mathematical statistics method has gained considerable attention. This method can measure the dispersion degree of a data set by calculating the mean and standard deviation of a set of data, thus realizing the detection of abnormal data [23-24]. Shi et al. [25] cleanses anomalous data by calculating the standard deviation of the data and setting a threshold. However, due to the lack of preprocessing during the handling of raw data, there may be instances of missed detection during the anomaly data cleansing process. Li et al. [26] and Yan et al. [27] detected anomalous data by calculating the standard deviation of the sampled data. However, this method does not consider the computational challenges posed by the storage and processing of large-scale data. The more buses to be monitored, the more severe these problems become [28].

To facilitate algorithm implementation and reduce storage and computation demands, this paper proposes an improved standard deviation calculation method. First, the mean (E_i) and standard deviation (σ_i) of each short-term data cycle are calculated, where E_i and σ_i represent the mean and standard deviation of the i-th short-term cycle, respectively. Then, based on the calculated E_i, σ_i, and sample sizes (n_i) of each short-term cycle, the overall mean (E) and standard deviation (σ) of the long-term cycle are updated through a rolling calculation, where E and σ denote aggregated statistical parameters. During this process, only the mean, standard deviation, and sample size of each short-term data set need to be stored, significantly reducing data storage requirements and avoiding repeated computations on large historical datasets. The proposed method supports an intelligent fault detection and analysis system for distribution networks, enabling sensitive detection of voltage anomalies such as high-resistance grounding and line breaks.

2. Principle of Improved Standard Deviation Filter

The implementation of standard deviation filters for data anomaly detection has attracted considerable attention from the research community. This chapter introduces the principle of the improved standard deviation filter. The approach achieves a reduction in both data storage and computational requirements through short-period rolling computation. By introducing the time-weight coefficient, the system demonstrates heightened sensitivity to recent variations. Voltage anomaly identification is accomplished by analyzing distribution discrepancies between normal and abnormal data.

2.1. Principal definition

The neutral point of 10 to 35 kV distribution networks in China adopts a non-direct grounding system. When faults such as grounding, line break, and ferromagnetic resonance occur in the distribution network, the neutral point potential of the system will shift, resulting in abnormal bus voltage [29]. Based on the consistency characteristics of the three-phase voltage changes on the bus, the phase voltage difference is selected as the characteristic value.

A sliding time window standard deviation filter can be constructed using historical eigenvalues. First, the phase voltage data sampled from the distribution network is divided into three sets of long-period phase voltage difference datasets, and the long-period datasets are further grouped into individual short-period datasets. There will be multiple points in the retained dataset that are consistently unchanged due to communication and other reasons at the time of sampling. At this point, it is necessary to preprocess this data to eliminate it. Subsequently, E_i and σ_i for the current cycle are calculated using the short-cycle data. Based on E_i, σ_i, and n_i obtained from each short-term cycle, a time-weighting coefficient is introduced to compute E and σ of the long-period datasets through a rolling process. The short-period mean and standard deviation are corrected according to the temporal distance to ensure that the improved standard deviation filter is more sensitive to recent data changes and adapts to the new characteristics of the bus voltage more quickly.

Moreover, since the phase voltage difference dataset, composed of historical data, contains anomalous data when faults occur in the distribution network, it is necessary to eliminate the data in the long-period dataset that deviates from the normal distribution range. After the elimination, the grouping calculation is performed again according to the initial steps to obtain the mean and standard deviation of the three data sets.

2.2. Short cycle selection

Firstly, the system obtains the sampled long-period phase voltage data. If the standard deviation filter uses historical data from the past M days, the long period of the sliding time window is M days. The first calculation needs to save the three-phase bus voltages for the past M days. Then, the phase voltage difference is calculated from the sampled phase voltages using the following formula:

$$U_{ab} = U_a - U_b \tag{1}$$

$$U_{bc} = U_b - U_c \tag{2}$$

$$U_{ca} = U_c - U_a \tag{3}$$

In this case, the following three sets of long-period phase voltage difference data can be obtained:

$$D_{Uab} = \{U_{ab1}, U_{ab2}, \ldots, U_{abm}\} \tag{4}$$

$$D_{Ubc} = \{U_{bc1}, U_{bc2}, \ldots, U_{bcm}\} \tag{5}$$

$$D_{Uca} = \{U_{ca1}, U_{ca2}, \ldots, U_{cam}\} \tag{6}$$

D_{Uij} represents the long-period phase voltage difference data set, i = a, b, or c, and j =b, c, or a for the phase sequence. m represents the number of samples in each data set. If M = 30 is selected, and historical data of three-phase voltage is recorded every minute, then m = 30 × 24 × 60 = 43,200. The improved method of calculating the standard deviation in this paper is to initially divide the long-period dataset into short-period datasets. The short period is chosen to be N days, where M is an integer multiple of N. The long-period phase voltage difference data set D_{Uij} is divided into L = M / N groups by period:

$$D_{Uij} = \{U_{Uij1}, U_{Uij2}, \ldots, U_{UijL}\} \tag{7}$$

$$D_{Uijs} = \{U_{Uijs1}, U_{Uijs2}, \ldots, U_{Uijsn}\} \tag{8}$$

D_{Uijs} represents the short-period phase voltage difference data set, where s = 1, 2, ..., L denotes the short cycle number, and n is the number of samples in each short period. The value of n is calculated as $n = m / L$. If N = 1 is chosen, then L = 30 and n = 1,440.

The quantile-quantile plot (QQ plot) of the phase voltage difference U_{ab} at a substation over a 10 months period is shown in Fig. 1. The dataset contains abnormal voltage data points. The points in the QQ plot are near the straight line, indicating that a normal distribution is satisfied. The presence of points away from the straight line on the left and right sides of the graph is due to the presence of data anomalies in the 10-month data. It can be seen that the data satisfy the normal distribution when the distribution network operates normally. After the exclusion of abnormal voltage data, the resultant histogram of the phase voltage difference U_{ab} during normal operation of the distribution network is shown in Fig. 2. Since the system samples are discrete, the data obtained is jumpy rather than continuous. The resulting interval segments of the phase voltage difference data set are not continuous.

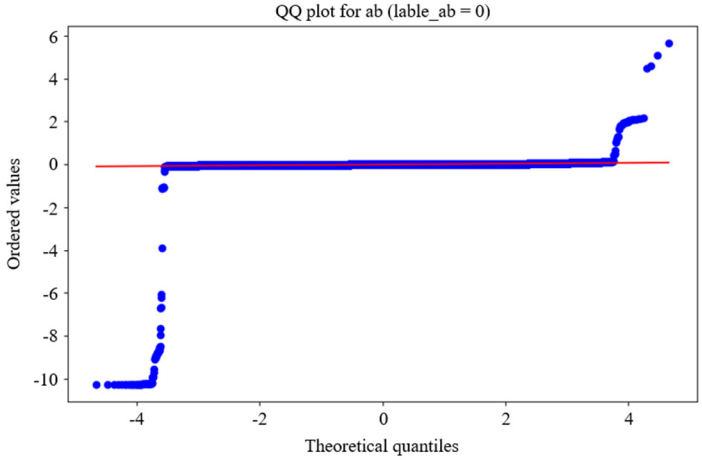


Fig. 1 QQ plot of U_{ab} at the 10 kV bus terminal (10-month data window)

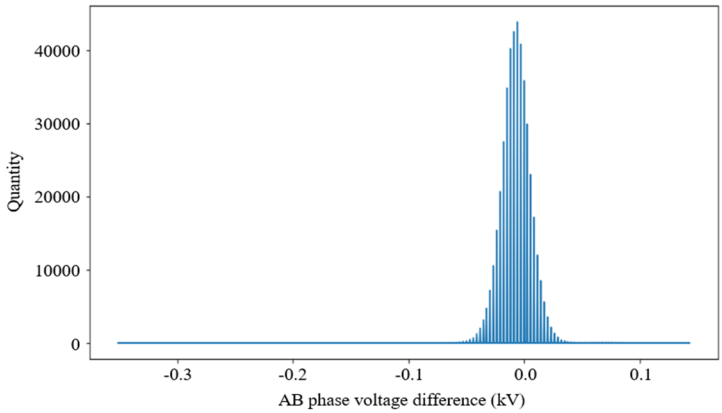


Fig. 2 Probability density plot of U_{ab} at the 10 kV bus terminal (10-month data window)

2.3. Data preprocessing

During normal operation of the distribution network, the phase voltage difference satisfies the normal distribution. Due to problems in sampling, transmission, and storage, saved historical data may have multiple persistent no-change points. As shown in Fig. 3, persistent non-refreshment of historical data results in localized spikes in the probability density figure.

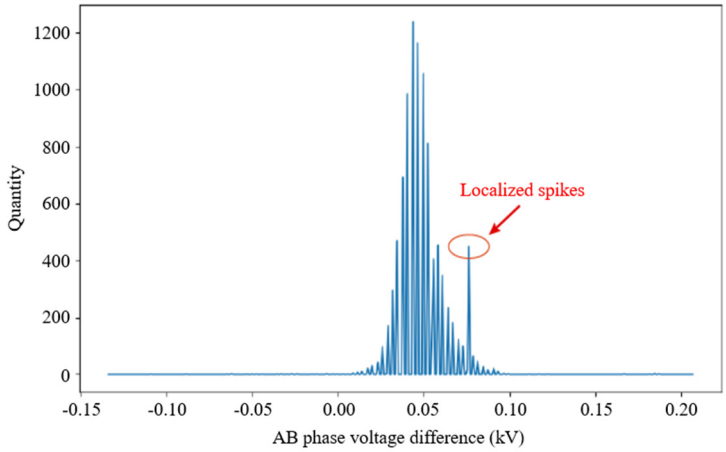


Fig. 3 Probability density plot of U_{ab} at the 10 kV bus terminal (1-week data window)

Due to the presence of localized spikes, the computed mean and standard deviation values become artificially inflated, thereby compromising voltage anomaly detection accuracy. In this case, localized spikes should be removed through data preprocessing. Since the bus voltage telemetry is accurate to two decimal places, there are fluctuations in the bus voltage during normal operation. According to the statistics, it is rare for the bus voltage to remain unchanged for five minutes, which can be ascribed to the communication terminal or data not being refreshed. Meanwhile, the data points that remain unchanged for five or more consecutive minutes throughout the day need to be eliminated. Since the system picks one point a minute, it is considered that in the case of five or more consecutive identical data points, the identical data other than the first data points will be excluded.

2.4. Mean and standard deviation

The sample number of each short-period phase voltage difference data set D_{Uijs} after preprocessing is denoted as n'_{ijs}, whereas the number of samples before preprocessing is n. The E'_{ijs} and σ'_{ijs} are calculated separately for each short-period data set D_{Uijs} as follows. E'_{ijs} and σ'_{ijs} represent the mean and standard deviation of each short period after data preprocessing, respectively.

$$E'_{ijs} = \frac{U_{ijs1} + U_{ijs2} + \cdots + U_{ijsn}}{n'_{ijs}} \tag{9}$$

$$\sigma'^{2}_{ijs} = \frac{\left(U_{ijs1} - E'_{ijs}\right)^2 + \left(U_{ijs2} - E'_{ijs}\right)^2 + \cdots + \left(U_{ijsn} - E'_{ijs}\right)^2}{n'_{ijs}} \tag{10}$$

Based on the mean and standard deviation of each short-period dataset, the long-period mean E'_{ij} and σ'_{ij} of the phase voltage difference data set D_{Uij} are calculated based on the previously obtained mean and standard deviation of each short-period dataset, as shown in the formulas below. E'_{ij} and σ'_{ij} represent the mean and standard deviation of each long period after data preprocessing, respectively.

$$E'_{ijs} = \frac{n'_{ij1}E'_{ij1} + n'_{ij2}E'_{ij2} + \cdots + n'_{ijL}E'_{ijL}}{n'_{ij1} + n'_{ij2} + \cdots + n'_{ijL}} \tag{11}$$

$$\sigma_{ij}^{'2}=\frac{n_{ij1}^{'}\sigma_{ij1}^{'2}+n_{ij2}^{'}\sigma_{ij2}^{'2}+\cdots+n_{ijL}^{'}\sigma_{ijL}^{'2}}{n_{ij1}^{'}+n_{ij2}^{'}+\cdots+n_{ijL}^{'}}+\frac{n_{ij1}^{'}\left(E_{ij1}^{'}-E_{ij}^{'}\right)^{2}+n_{ij2}^{'}\left(E_{ij2}^{'}-E_{ij}^{'}\right)^{2}+\cdots+n_{ijL}^{'}\left(E_{ijL}^{'}-E_{ij}^{'}\right)^{2}}{n_{ij1}^{'}+n_{ij2}^{'}+\cdots+n_{ijL}^{'}} \tag{12}$$

From the above equations, it can be seen that the long-period mean and standard deviation can be derived from the saved short-period mean and standard deviation without repeated calculation of the mean and standard deviation.

When the mode of operation changes, such as the parallel operation of the bus, the main transformer is out of power, or the bus is supplied by the opposite side of the joint circuit, the voltage characteristics of the bus will change, which can easily cause a false judgment of voltage anomaly. Regarding sudden changes in bus voltage, as the selected time window lengthens, the adaptability of the historical data-based standard deviation filter decreases.

Meanwhile, the time window is too short for the standard deviation based on a small amount of sample data to accurately characterize the data. Therefore, a time-dependent weighting coefficient is introduced into the conventional standard deviation calculation equation. The short-period mean and standard deviation are corrected according to the length of time. This ensures that the improved standard deviation filter is more sensitive to recent data changes and adapts more quickly to the new characteristics of the bus voltage. With the introduction of the time-dependent weighting coefficient K_{ijs}, the above equation becomes:

$$E_{ijs}^{'}=\frac{K_{ij1}n_{ij1}^{'}E_{ij1}^{'}+K_{ij2}n_{ij2}^{'}E_{ij2}^{'}+\cdots+K_{ijL}n_{ijL}^{'}E_{ijL}^{'}}{K_{ij1}n_{ij1}^{'}+K_{ij2}n_{ij2}^{'}+\cdots+K_{ijL}n_{ijL}^{'}} \tag{13}$$

$$\begin{aligned}\sigma_{ij}^{'2}=&\frac{K_{ij1}n_{ij1}^{'}\sigma_{ij1}^{'2}+K_{ij2}n_{ij2}^{'}\sigma_{ij2}^{'2}+\cdots+K_{ijL}n_{ijL}^{'}\sigma_{ijL}^{'2}}{K_{ij1}n_{ij1}^{'}+K_{ij2}n_{ij2}^{'}+\cdots+K_{ijL}n_{ijL}^{'}}\\&+\frac{K_{ij1}n_{ij1}^{'}\left(E_{ij1}^{'}-E_{ij}^{'}\right)^{2}+K_{ij2}n_{ij2}^{'}\left(E_{ij2}^{'}-E_{ij}^{'}\right)^{2}+\cdots+K_{ijL}n_{ijL}^{'}\left(E_{ijL}^{'}-E_{ij}^{'}\right)^{2}}{K_{ij1}n_{ij1}^{'}+K_{ij2}n_{ij2}^{'}+\cdots+K_{ijL}n_{ijL}^{'}}\end{aligned} \tag{14}$$

The setting of the weighting coefficient K_{ijs} is set based on a linear or exponential function of the dataset storage time. As the dataset storage time shortens, the weighting coefficient increases accordingly; conversely, it decreases as the storage time lengthens. The time weighting coefficients take values in the range of 0 to 1. The longer the data are retained, the closer the weighting coefficient approaches 0; conversely, it approaches 1 as the retention time decreases. In particular, when K_{ijs} = 1, the mean and standard deviation calculations become time-independent.

With the introduction of the weighting coefficient, the phase voltage difference dataset in the historical data will include abnormality data generated upon faults in the distribution network. In this case, data other than $E'_{ij} \pm 3\sigma'_{ij}$ need to be excluded from each long-period phase voltage difference dataset. The number of samples in each short-period phase voltage difference data set D_{Uijs} after exclusion is denoted as n_{ijs}, while the number of samples before exclusion is n'_{ijs}. At this point, recalculate E_{ijs} and σ_{ijs} for each dataset D_{Uijs} after elimination. After the calculation, the means E_{ab}, E_{bc}, E_{ca}, and standard deviations σ_{ab}, σ_{bc}, σ_{ca} of the three datasets are obtained.

3. Proposed Method

This section presents the principle and implementation of the proposed voltage anomaly detection method for distribution networks. Based on the concept of the improved standard deviation filter, anomalies can be identified by analyzing the distribution patterns of normal and abnormal data. In practical operations, however, certain evident fault conditions can also be directly detected through voltage analysis. Building on this, this study integrates actual operational data with the improved standard deviation filter to enhance voltage anomaly detection in distribution networks.

3.1. Voltage anomaly detection

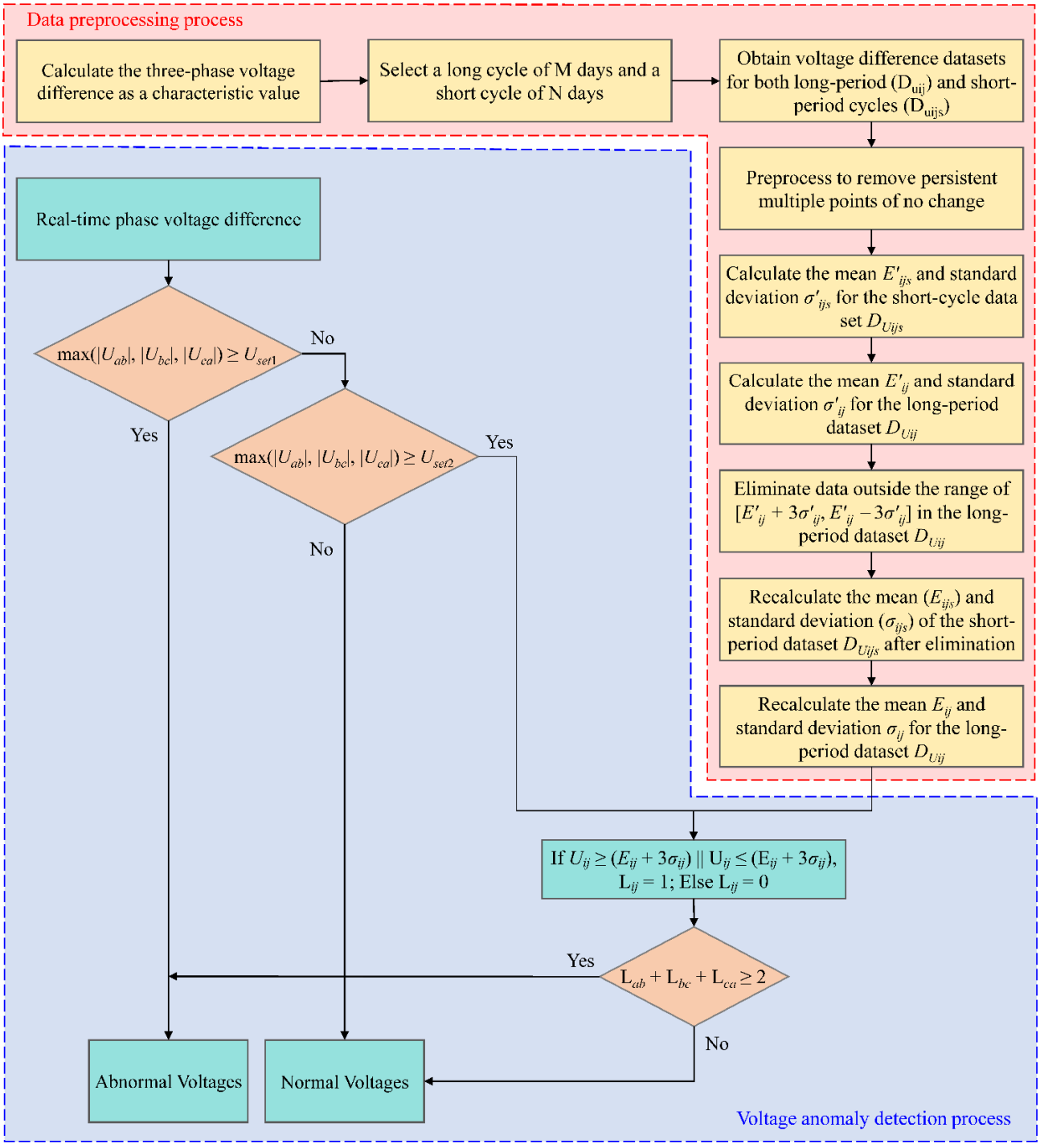


Fig. 4 Flowchart for voltage anomaly detection in distribution network

The absolute value of the real-time phase voltage difference in the (L+1)th short period (after M days) is taken. If the absolute value of the voltage difference in any phase is greater than or equal to U_{set1} (i.e., $\max(|U_{ab}|, |U_{bc}|, |U_{ca}|) \geq U_{set1}$), it is directly judged as a voltage anomaly. Based on actual operating experience, U_{set1} can be set to 0.8 kV. If the maximum value of the absolute phase voltage difference is less than U_{set1} but greater than or equal to U_{set2} (i.e., $U_{set2} \leq \max(|U_{ab}|, |U_{bc}|, |U_{ca}| \leq U_{set1})$, then it is necessary to determine whether the absolute value of each phase voltage difference lies within the range of $E_{ij} \pm 3\sigma_{ij}$. Taking U_{ab} as an example, this requires verifying whether $E_{ab} - 3\sigma_{ab} \leq U_{ab} \leq E_{ab} + 3\sigma_{ab}$ is satisfied. On the other hand, if the absolute value of the real-time phase voltage difference fails to satisfy two or more of the above conditions simultaneously, it is also judged as a voltage anomaly. To ensure sufficient sensitivity to bus voltage anomalies, U_{set2} can be set to 0.4 kV.

The voltage thresholds (U_{set1}, U_{set2}) mentioned here are derived from statistical data on three-phase unbalance of busbar voltages in hundreds of operational substations. During normal operation, the phase voltage of the distribution network does not exceed U_{set1}, whereas during fault conditions, the phase voltage does not fall below U_{set2}. A fault is considered to have possibly occurred only when the phase voltage lies between these two thresholds. The flowchart for detecting anomalies in the bus voltage of the distribution network is shown in Fig. 4.

3.2. Real-time data processing

At the end of the (L+1)th short period, the data preprocessing described in Section 2.3 is repeated to exclude both persistent and discrete abnormal data. After excluding the data, mean $E_{ij(L+1)}$, standard deviation $\sigma_{ij(L+1)}$, and number of samples $n_{ij(L+1)}$ of the phase voltage difference dataset for the (L+1)th short period are calculated and saved. The overall E_{ij} and σ_{ij} from the 2nd to the (L+1)th short period are then computed using a rolling approach to support real-time anomaly detection in the (L+2)th short period. Due to the large capacity of the system, the impact of distributed generation (DG) integration is relatively minor. Moreover, the connection or disconnection of distributed power sources does not lead to an imbalance in the bus load that would cause a three-phase voltage imbalance at the bus. Therefore, this method is also applicable to the integration of DG.

4. Case Study

To verify the accuracy of the method proposed in this paper for detecting abnormal voltage data in distribution networks, this section first validates the effect of the improved standard deviation filter based on distribution network data. By comparing with different anomaly detection methods, the method proposed in this paper is more accurate in detecting abnormal data. Secondly, the effectiveness of the method is confirmed through on-site real-time monitoring verification.

4.1. Analysis of the impact of sliding time windows on data anomaly identification

The case parameters described below are designed uniformly as follows: the long-period M is set as 30 days, the short-period N is set as 1 day, U_{set1} is set as 0.8 kV, and U_{set2} is set as 0.4 kV. First, the historical voltage data of the distribution network for the past 30 days was obtained. The historical data contains abnormal voltage data. After data elimination and rolling calculation, the mean and standard deviation of the data set during normal operation are obtained. At this point, voltage anomaly detection can be carried out. Fig. 5 and Fig. 6 display voltage anomaly identification results at the same bus with one-week and one-month time windows, respectively. Circular markers denote normal data, while triangular markers indicate anomaly dates. The method shows high sensitivity to deviations from normal ranges. Shorter time windows may distort the results due to limited statistical samples, potentially causing normal data to be misclassified as anomalies.

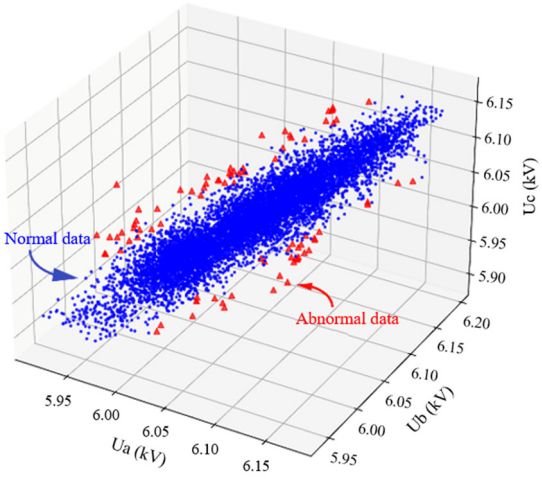


Fig. 5 Voltage anomaly identification (1-week window)

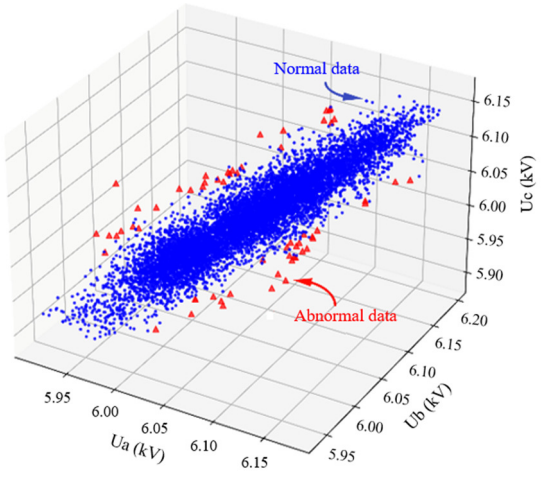


Fig. 6 Voltage anomaly identification (1-month window)

4.2. Case study of improved standard deviation filter

Figs. 7 to 9 show the comparisons of the probability density plots of the 10 kV bus terminal U_{ab} of a certain substation, with data windows of 1-week, 1-month, and 1-month with the introduction of a time-weighting coefficient, respectively. The left is the probability density distribution diagram smoothed by the Gaussian kernel, and the right is the original probability density diagram. As seen in Fig. 7 and Fig. 8, the dataset for long-time windows is more statistically regular compared to short-time windows. It can be seen from Fig. 8 and Fig. 9 that the introduction of the time weight coefficient does not decrease the normal distribution of the data set, whereas it will be more sensitive to recent data.

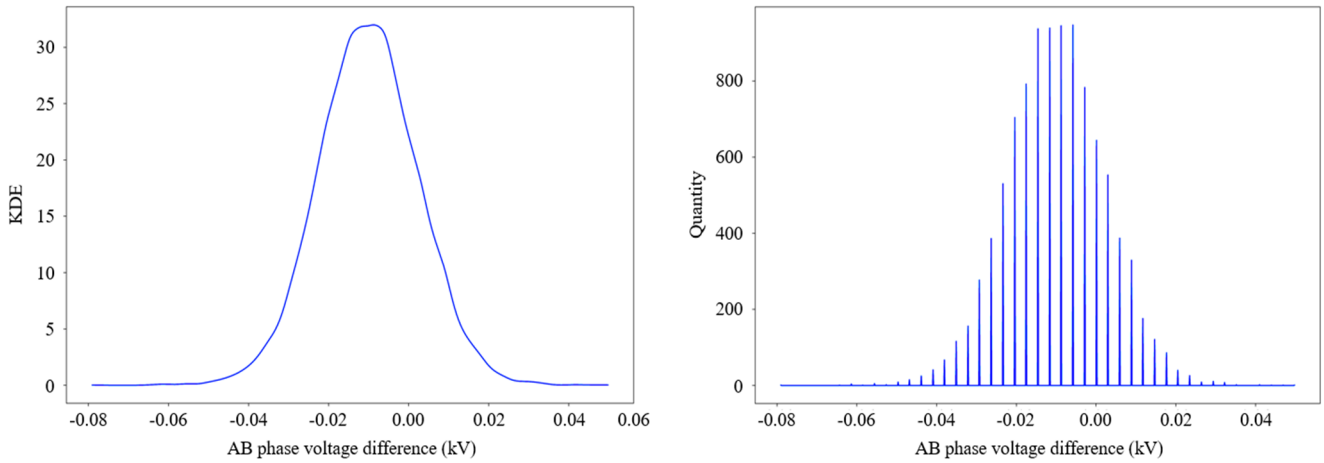


Fig. 7 Probability density of U_{ab} at 10 kV bus (1-week window)

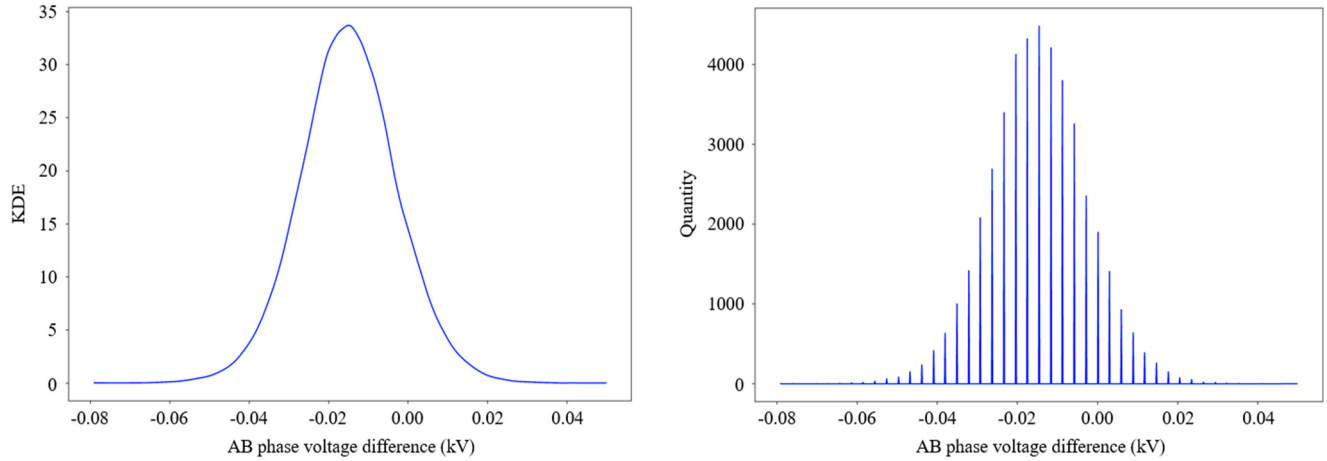


Fig. 8 Probability density of U_{ab} at 10 kV bus (1-month window)

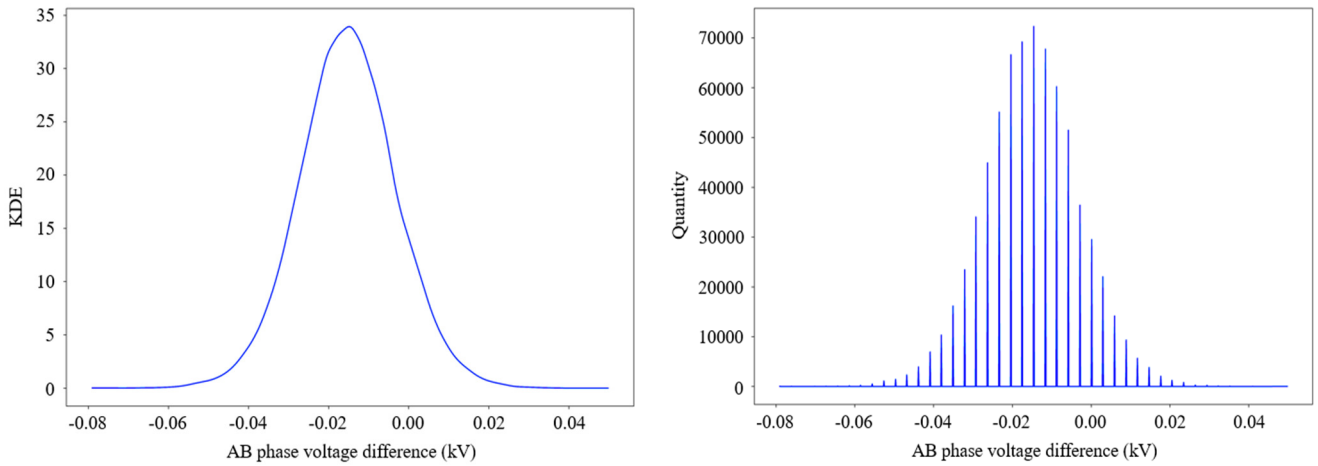


Fig. 9 Probability density of U_{ab} at 10 kV bus (1-month window with time-weighting coefficient)

4.3. Case study of voltage anomaly

The normal transient voltage of the 10 kV bus of a certain substation is shown in Table 1. As indicated in the table, during normal operating conditions of the distribution network, phase voltages remain within specified ranges (typically ±5% of the nominal value). Therefore, the phase voltage difference is close to 0, and the three-phase voltage differences are similar to each other.

Table 1 Normal three-phase voltage of the 10 kV bus of a substation

Time	U_a (kV)	U_b (kV)	U_c (kV)	$\|U_{ab}\|$ (kV)	$\|U_{bc}\|$ (kV)	$\|U_{ca}\|$ (kV)
03:07:00	5.7949	6.0352	5.8184	0.2403	0.2168	0.0235
03:08:00	5.7949	6.0322	5.8125	0.2373	0.2197	0.0176
08:40:00	5.7832	6.0059	5.8008	0.2227	0.2051	0.0176

When solid grounding, high-resistance grounding, and line break faults occur in a substation, three methods of improved standard deviation filter, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering, and isolation forest are used to identify the bus voltage. The identification results are shown in Table 2. A value of "0" indicates that the bus voltage is recognized as normal, and a value of "1" indicates that the bus voltage is recognized as abnormal. "Data.2" represents low-resistance grounding, "Data.3" and "Data.4" represent single-phase line break faults, and "Data.5", "Data.1", and "Data.6" represent high-resistance grounding faults.

Table 2 Comparison of three different methods

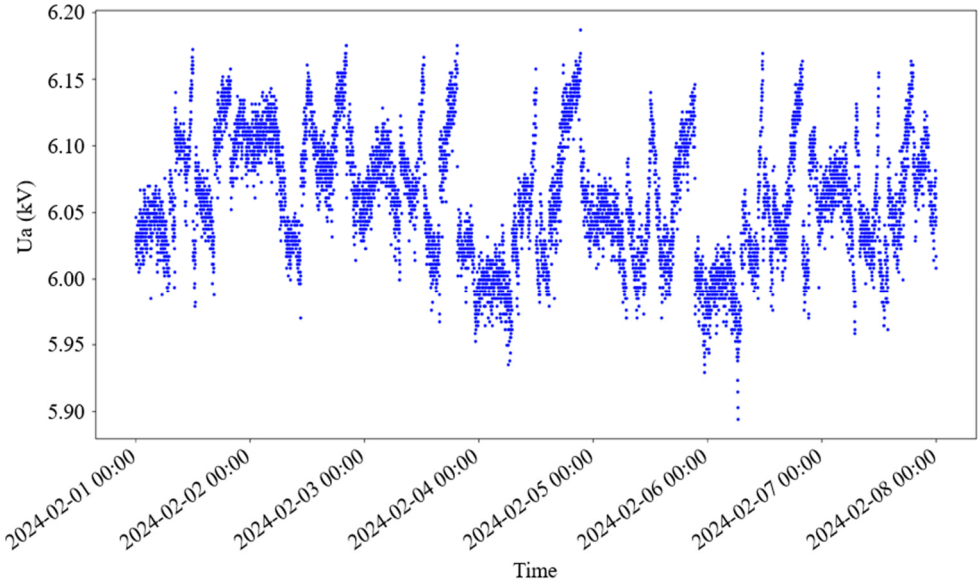
Data	U_a (kV)	U_b (kV)	U_c (kV)	U_{ab} (kV)	U_{bc} (kV)	U_{ca} (kV)	The improved standard deviation filter	DBSCAN clustering	Isolation forest
1	5.50	5.86	6.10	-0.36	-0.24	0.60	1	0	0
2	0.16	10.20	10.51	-10.04	-0.31	10.35	1	1	1
3	5.6396	6.2578	5.5957	-0.6182	0.6621	-0.0439	1	1	1
4	5.7041	6.1143	5.6543	-0.4102	0.46	-0.0498	1	0	1
5	5.9004	6.0732	5.5166	-0.1728	0.5566	-0.3838	1	0	0
6	5.8975	6.082	5.5342	-0.1845	0.5478	-0.3633	1	0	0

The recognition results indicate that the improved standard deviation filter achieves the highest accuracy in detecting voltage anomalies. In contrast, both the clustering algorithm and the isolation forest are less suitable for identifying abnormal voltages close to normal values. In the "Data.4" line break fault, the bus voltage is slightly abnormal, and the clustering algorithm cannot accurately identify it. For "Data.5" high-resistance grounding fault, neither the clustering algorithm nor the isolation forest can correctly identify the voltage anomaly.

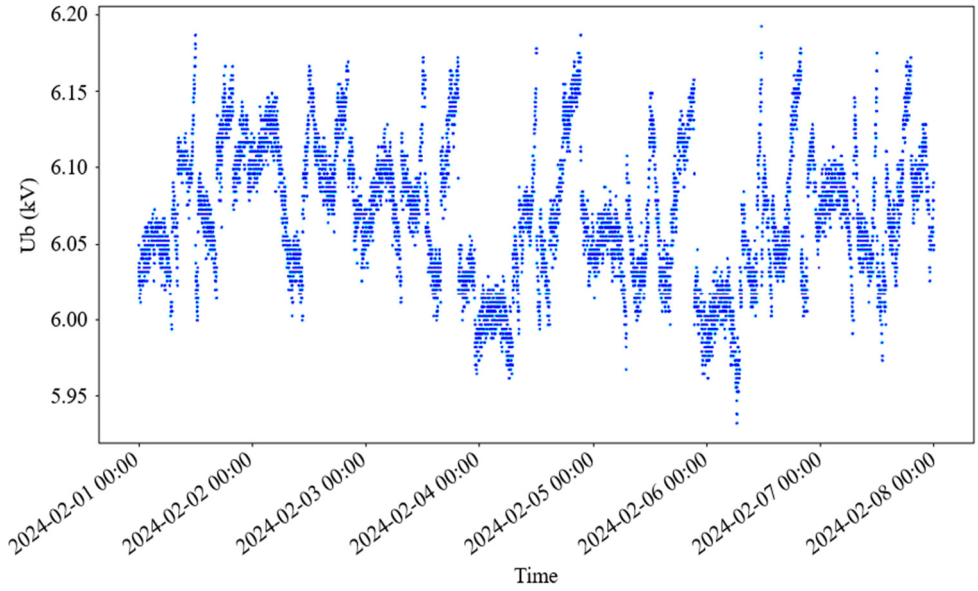
4.4. On-site case

The intelligent detection and analysis system for distribution network faults, developed based on the proposed method, has been successfully implemented. This system is capable of sensitively detecting voltage anomalies such as high-resistance grounding and line break, thereby providing essential prerequisites for subsequent fault type identification, timely fault detection, and rapid resolution.

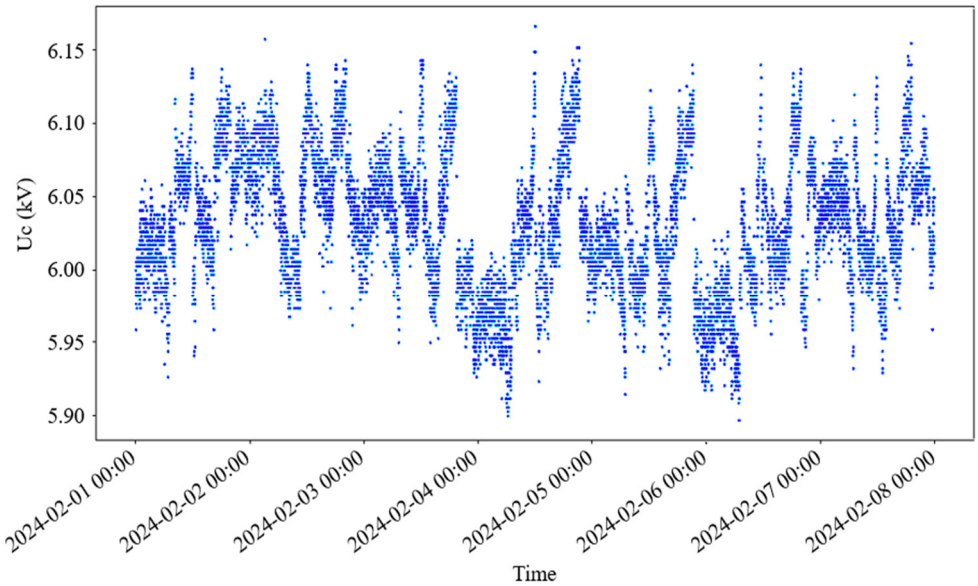
Fig. 10 shows a short-period three-phase voltage dataset for a certain distribution network. The short period N is selected as 7. Thus, the number of sample points is $60 \times 24 \times 7 = 10{,}080$. It can also be seen from Fig. 10 that under normal operating conditions of the distribution network, the phase voltages are maintained within the specified range (usually ±5% of the rated value).



(a) Short-period phase voltage data set of phase A



(b) Short-period phase voltage data set of phase B



(c) Short-period phase voltage data set of phase C

Fig. 10 Short-period three-phase voltage dataset for a certain distribution network

After obtaining the short-period dataset, it needs to be preprocessed to exclude data that remains constant at multiple continuous points. Subsequently, the mean and standard deviation of the short-period dataset are calculated. The mean and standard deviation of the long-period dataset are derived from a rolling calculation using the mean and standard deviation of the short-period dataset. A time-dependent weighting coefficient is introduced during the rolling calculation process. After data exclusion and recalculation, the distribution network voltage anomaly detection table is shown in Table 3.

An alarm signal is issued when the proposed method detects a voltage anomaly in the distribution network. The voltage anomaly detection method based on the improved standard deviation filter accurately identifies the voltage anomaly and issues an alarm signal regardless of solid grounding, high-resistance grounding, line break faults, and ferromagnetic resonance faults.

Table 3 Distribution network voltage anomaly detection table of the proposed method

Substation	Bus	Fault type	Instantaneous voltage (kV)	
			Phase voltage	Line voltage
Zhuzhou.Guanzhuang	10 kV.II	High resistance grounding	U_a: 5.73 U_b: 6.24 U_c: 6.12	U_{ab}: 10.44
Zhuzhou.Guanzhuang	10 kV.II	High resistance grounding	U_a: 5.65 U_b: 6.08 U_c: 5.93	U_{ab}: 10.21
Zhuzhou.Xiangshiling	10 kV.II	High resistance grounding	U_a: 5.90 U_b: 5.59 U_c: 6.07	U_{ab}: 10.14
Zhuzhou.Guanzhuang	10 kV.II	High resistance grounding	U_a: 5.78 U_b: 6.34 U_c: 6.21	U_{ab}: 10.58
Zhuzhou.Shizitan	10 kV.II	Disconnection fault	U_a: 5.77 U_b: 6.12 U_c: 5.72	U_{ab}: 10.18
Zhuzhou.Xiaojiang	10 kV.II	High resistance grounding	U_a: 5.63 U_b: 6.12 U_c: 5.80	U_{ab}: 10.13
Zhuzhou.Fugang	10 kV.I	Disconnection fault	U_a: 6.41 U_b: 5.67 U_c: 5.59	U_{ab}: 10.17
Zhuzhou.Junchu	10 kV.I	Ferromagnetic resonance	U_a: 3.38 U_b: 7.44 U_c: 7.35	U_{ab}: 10.14

5. Conclusion

To address the shortcomings of fixed over-limit alarms in detecting voltage anomalies in distribution networks, such as omissions and false alarms. This paper proposed a detection method based on an improved standard deviation filter. The method identifies voltage anomalies by evaluating the dispersion degree of the phase voltage difference dataset using the mean and standard deviation. As a result, the reliability and timeliness of voltage anomaly detection were improved. The effectiveness of the proposed method was verified through theoretical analysis and on-site data. The main conclusions are summarized as follows:

(1) The short-period mean and standard deviation were utilized to compute the long-period statistics. During this process, only the mean, standard deviation, and sample size of each short-period dataset needed to be stored, which significantly reduced data storage requirements and eliminated repeated calculations on large volumes of historical data.

(2) Experimental results demonstrated that deep-learning and isolation forest algorithms had limitations in detecting voltage anomalies, leading to low identification accuracy. In contrast, the proposed method accurately identified voltage anomalies across various fault types.

(3) The proposed method was unaffected by operational changes, such as parallel bus operation, main transformer outages, or supply switching from the opposite side of the tie-line.

Acknowledgments

The authors acknowledge the support of State Grid Zhuzhou Electric Power Co., Ltd. This work was funded by the Natural Science Foundation of Hunan Province (2023JJ30036).

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] G. Pathirage, R. Razzaghi, and L. Andrew, "Fault Location in Power Distribution Networks Using Frequency Analysis of Traveling Waves," Electric Power Systems Research, vol. 249, article no. 111971, 2025.

[2] B. Tang, J. Bao, N. Pan, M. Liu, J. Li, and Z. Xu, "Grid Operation and Inspection Resource Scheduling Based on an Adaptive Genetic Algorithm," International Journal of Engineering and Technology Innovation, vol. 14, no. 2, pp. 152-164, 2024.

[3] J. Y. Ren, J. W. Zhao, N. Pan, N. B. Zhang, and J. W. Yang, "Prediction of Distribution Network Line Loss Rate Based on Ensemble Learning," International Journal of Engineering and Technology Innovation, vol. 14, no. 1, pp. 103-114, 2024.

[4] Q. Ren, W. Kang, X. Yang, Q. Wang, and Q. Huang, "Multi-dimensional Deep Learning-Based Anomaly Detection and Adaptive Security Strategy for Sustainable Power Grid Operation," Measurement, vol. 252, article no. 117313, 2025.

[5] Z. Lu, G. Zhao, X. Kong, J. Chen, X. Guo, and J. Zhang, "A Detection Based on Particle Filtering and Multivariate Time-Series Anomaly Detection via Graph Attention Network for Automatic Voltage Control Attack in Smart Grid," Sustainable Energy, Grids and Networks, vol. 40, article no. 101494, 2024.

[6] J. Fuentes-Velazquez, E. Beltran, E. Barocio, and C. Angeles-Camacho, "A Fast Automatic Detection and Classification of Voltage Magnitude Anomalies in Distribution Network Systems Using PMU Data," Measurement, vol. 192, article no. 110816, 2022.

[7] L. Klein, J. Dvorský, D. Seidl, and L. Prokop, "Novel Lossy Compression Method of Noisy Time Series Data with Anomalies: Application to Partial Discharge Monitoring in Overhead Power Lines," Engineering Applications of Artificial Intelligence, vol. 133, part C, article no. 108267, 2024.

[8] V. Kachesov, A. Lebedev, and E. Kitova, "Monitoring in 6–35 kV Power Networks, Location of Single-Phase Ground Fault and Detection of Fault Feeder," International Journal of Electrical Power & Energy Systems, vol. 152, article no. 109271, 2023.

[9] L. L. Zhang, Y. Zhang, Y. D. Xue, C. Wang, and Z. M. Shao, "Fault Phase Identification of Non-solidly Grounding System Considering System Asymmetry," Electric Power Automation Equipment, vol. 39, no. 4, pp. 24-29, 2019. (In Chinese)

[10] H. Zeng and C. Yang, "Fault Location Matrix Algorithm for Distribution Network Based on Voltage Amplitude Correction," Electric Power Science and Engineering, vol. 36, no. 7, pp. 41-46, 2020. (In Chinese)

[11] Q. Li, J. Xue, J. Wang, J. Xiao, Y. Fan, M. Zhou, et al., "Protection Performance Assessments of Two Types of Sheath Voltage Limiters in 220kV Single Point Bonded Cable System," Electric Power Systems Research, vol. 212, article no. 108637, 2022.

[12] Y. Chen, C. Liu, J. Li, R. Yang, C. Hou, and P. Yao, "Self-Adaptive Identification Method for 110kV Line Disconnection Fault Based on Characteristics of Voltage Changes," Proceedings of the CSU-EPSA, vol. 34, no. 8, pp. 93-101, 2022. (In Chinese)

[13] J. Lin, J. You, M. Guo and Q. Hong, "Amplitude-Phase-Characteristics-Based Fault Phase Detection Method for Grounding Fault in Distribution Network," IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-12, 2024.

[14] E. B. Espejo, F. R. S. Sevilla, and P. Korba, Monitoring and Control of Electrical Power Systems Using Machine Learning Techniques, Amsterdam, Netherlands: Elsevier, pp. 137-161, 2023.
[15] A. P. Shafei, J. F. A. Silva, and J. Monteiro, "Convolutional Neural Network Approach for Fault Detection and Characterization in Medium Voltage Distribution Networks," e-Prime - Advances in Electrical Engineering, Electronics and Energy, vol. 10, article no. 100820, 2024.
[16] M. Liu, X. He, R. Qin, H. Jiang, and X. Meng, "Anomaly Detection of Distribution Network Voltage Data Based on Improved K-means Clustering K-value Selection Algorithm," Journal of Electric Power Science and Technology, vol. 37, no. 6, article no. 10, 2022.
[17] C. Zhang, "Application of K-Means Clustering Algorithm in Substation Voltage Anomaly Monitoring," Application of IC, vol. 40, no. 02, pp. 288-290, 2023. (In Chinese)
[18] F. Liu, R. Liu, and Z. Dong, " Anomalous Dynamic Data Detection Method for Smart Meters Based on K-Means Clustering," Electronic Design Engineering, vol. 31, no. 05, pp. 84-88, 2023. (In Chinese)
[19] H. Kuang, X. He, M. He, R. S. Qin, and H. Jiang, "Abnormal Voltage Data Detection of Distribution Network Based on Bidirectional Long Short-Term Memory Neural Network," Science Technology and Engineering, vol. 21, no. 24, pp. 10291-10297, 2021.
[20] Z. Ke, C. Luo, and S. Li, "Detection Method of Abnormal Voltage Data in Distribution Network Based on Recurrent Neural Network," Electronic Design Engineering, vol. 32, no. 01, pp. 106-110, 2024. (In Chinese)
[21] J. Guo, Y. Hou, L. Ding, and Z. Jin, "Multi-type Data Anomaly Detection in Power System State Estimation Using Support Vector Machine," Foreign Electronic Measurement Technology, vol. 43, no. 04, pp. 152-161, 2024. (In Chinese)
[22] Y. Mei, Y. Li, W. Zhou, Y. Guo, W. Deng, and X. Qiao, "Dynamic Data Cleaning Method of Abnormal and Missing Data in a Distribution Network Based on Machine Learning," Power System Protection and Control, vol. 51, no. 7, pp. 158-169, 2023. (In Chinese)
[23] K. Y. Liu, W. X. Sheng, and L. J. Hu, "Power System Data Adjustment Method Based on Robust Unscented Kalman Filter," Science Technology and Engineering, vol. 20, no. 31, pp. 12857-12862, 2020. (In Chinese)
[24] C. T. Yen and U. H. Chen, "Design of Deep Learning Acoustic Sonar Receiver with Temporal/ Spatial Underwater Channel Feature Extraction Capability," International Journal of Engineering and Technology Innovation, vol. 14, no. 2, pp. 115-136, 2024.
[25] M. Shi, R. Yin, A. Hu, and J. Wu, "A Novel Photovoltaic Array Outlier Cleaning Algorithm Based on Moving Standard Deviation," Power System Protection and Control, vol. 48, no. 6, pp. 108-114, 2020. (In Chinese)
[26] J. Li, R. Liu, T. Lu, L. Pan, C. Ma, Q. Li, et al., "Early Fault Diagnosis of Lithium-Ion Battery Packs Based on Improved Local Outlier Detection and Standard Deviation Method," Energy Storage Science and Technology, vol. 12, no. 9, pp. 2917-2926, 2023.
[27] Y. Yan, G. Sheng, Y. Liu, X. Du, H. Wang, and X. Jiang, "Anomalous State Detection of Power Transformer Based on Algorithm SlidingWindows and Clustering," High Voltage Engineering, vol. 42, no. 12, pp. 4020-4025, 2016. (In Chinese)
[28] Q. Yang, W. Hao, L. Ge, W. Ruan, and F. Chi, "FARIMA Model-Based Communication Traffic Anomaly Detection in Intelligent Electric Power Substations," IET Cyber-Physical Systems: Theory & Applications, vol. 4, no. 1, pp. 22-29, 2019.
[29] L. Zhang, B. Xu, Y. Xue, and H. Gao, "Transient Fault Locating Method Based on Line Voltage and Zero-Mode Current in Non-Solidly Earthed Network," Proceedings of the CSEE, vol. 32, no. 13, pp. 110-115+198, 2012. (In Chinese)