Electrical Power and Energy Systems 172 (2025) 111266

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Statistical filtering-based least square method for short-circuit capacity identification of traction power supply system

Shaobing Yang [a], Liang Hu [a,b], Wanqi Zhang [a,*], Tingting He [a], Dylan Dah-Chuan Lu [c], Mingli Wu [a]

[a] School of Electrical Engineering, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China
[b] Power Supply Branch of Guoneng Xinshuo Railway Co., Ltd, Ordos, Inner Mongolia 010300, China
[c] School of Electrical and Data Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia

ARTICLE INFO

Keywords:
Thevenin equivalent model
Short circuit capacity
Data-driven
Statistical filtering
Traction power supply system

ABSTRACT

The short-circuit capacity (SCC) of electrified railways extends beyond power supply and voltage stability, encompassing traffic organization, relay protection, and fault diagnosis. However, due to the dual fluctuations of source and loads, it is difficult to establish Thevenin equivalent model (TEM) and calculate the SCC at feeders of traction power supply system (TPSS). This paper proposes a statistical filtering-based least square method to identify SCC. Two statistical filtering rules are described to improve the consistence between voltage and power variations. The data that satisfy two rules are obtained by power threshold and box plots. It can reduce the error of TEM caused by source-side fluctuations and low voltage-power variation consistency. Preliminary results are obtained by the least square regression and evaluated by the fitting quality. The result with poor fitting quality is modified by probability distribution of historical results. Considering the loose thresholds, Kalman filter is introduced to reduce the error caused by obvious outliers. Finally, the proposed algorithm is validated by simulations and field test evaluated by Jensen-Shannon divergence. The SCC is roughly same as the actual value with an improvement of 15.76% compared to conventional methods. Besides, the sensitivity analysis shows the great generalization performance of the proposed methods.

1. Introduction

The development of power system with the aim of transitioning towards low-carbon energy sources is promoted by the government in recent years. Microgrids in power system, especially renewable energy sources, have garnered widespread attention [1–4]. The operation of power systems featuring a substantial integration of renewable energy encounters notable challenges arising from factors, such as diminished impedance, reduced inertia, and limited support capacity. Furthermore, the presence of electrified railways, distinguished by high power demand, pronounced impact, and fluctuations, significantly affects power system in regions abundant in renewable energy yet possessing relatively weak support capacity. To address fluctuations occurring on both the source and load sides of electrified railways, it is imperative to conduct analyses of grid robustness at feeder levels and to ensure secure operation within the designated power supply capacity.

Grid strength, serving as a metric for assessing the interaction between the power system and its connecting infrastructure, is defined as the ratio of short-circuit capacity (SCC) at the bus to the equipment capacity [5–7]. Due to the exceedance of loads, several cases, such as, voltage drop, traction blockage of trains, and transformer overload, show negative impacts on train operation. This is mainly attributed to the low SCC, which makes it difficult to supply the load. Quantitative descriptions of the load support capacity through SCC are essential for design, maintenance, and renovation of electrified railways [8,9].

Thevenin equivalent model (TEM) is a typical method to identify SCC. It has been widely applied in scenarios, such as, wind power integration [10], relay protection, and voltage stability analysis [11,12]. Current theories can be classified into two groups: network analytical and local measurement-based theories. The former one requires information of the grid architecture or the power flow data at each node. TEM is built by solving the entire network or inverting the admittance matrix [13,14]. These methods are mostly used by grid dispatch centers. However, it is difficult to obtain the detailed network information for railway departments. This makes it impossible to obtain SCC via the network analytical theory in traction power supply system (TPSS).

* Corresponding author.
E-mail address: 20117014@bjtu.edu.cn (W. Zhang).

https://doi.org/10.1016/j.ijepes.2025.111266
Received 17 January 2025; Received in revised form 27 September 2025; Accepted 13 October 2025
Available online 21 October 2025

To solve the problem of SCC identification, the local measurement-based theory only requires information from local nodes, including voltage, current, power, etc., which is more suitable for TPSS. Based on this, there are three general methods: two-section method (TSM), multiple-section method (MSM) and single-section method (SSM). TSM is a fundamental method for identifying SCC, but the accuracy is low due to dual fluctuations of source and loads. Further, MSM is introduced to mitigate parameter drift and reduce measurement and process errors. Most scholars have studied TEM based on the least square method (LSM) [15–18]. To reduce the error of phase changes, an improved approach for online voltage stability analysis is proposed [19]. Besides, the graphical method (GM) is applied to reduce the impact caused by fluctuations in source-side, but it has poor robustness [20–22]. In general, MSM requires the assumption of minimal fluctuations on the source-side and appropriate fluctuations on the load-side. It is difficult to apply them in electrified railways. Although some existing schemes can capture voltage fluctuations using supervised or unsupervised learning approaches [23–25], such algorithms demand high computational power, which current electrified railway systems are unable to meet. To address the issue, SSM is proposed using the sensitivity analysis or complete differential equations [26,27]. However, it relies on initial values heavily, which are hard to obtain. Moreover, it has significant time cost and computational complexity.

Nowadays, statistic-based approaches are explored to solve the aforementioned problems by selecting data [28,29]. Traditional methods typically assume "source-side stability." However, TPSSs, characterized by intense fluctuations on both the "source" (upstream grid) and "load" (train loads) sides, represent an extreme scenario where traditional methods are prone to large identification errors and insufficient usable data. To solve problems of all above methods, a statistical filtering-based least square (SFLS) algorithm is proposed to obtain the TEM and SCC. Main contributions are summarized as follows:

(1) A statistical filtering mechanism oriented towards "dual-side fluctuations" is proposed. Utilizing two statistical criteria and the box plot tool, it quantitatively screens data exhibiting consistent voltage-power changes, effectively distinguishing benign fluctuations dominated by load changes from adverse fluctuations caused by source-side disturbances. This lays the foundation for adapting traditional methods to special operating conditions.

(2) A result correction strategy for poor data conditions is constructed. By introducing a real-time correction mechanism based on historical probability distribution, a lightweight solution enhances algorithm robustness, ensures continuous result output, and improves engineering practicality.

(3) An integrated framework of "statistical filtering + least squares identification + dynamic probability correction" is formed. This not only improves identification accuracy and stability but also provides a new approach for parameter identification in similar volatile systems.

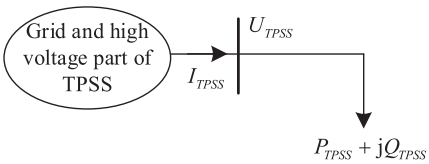
The rest of this paper is organized as follows. Section 2 studies electrical characteristics of TPSS and the framework of proposed algorithm. Section 3 illustrates the details of SFLS for identifying SCC. On this basis, the applicability of proposed method is validated by simulated cases and field test in Section 4. Conclusions are drawn in Section 5.

2. Characteristics analysis and TEM framework

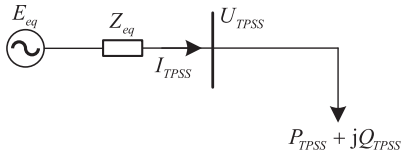
2.1. Thevenin equivalence model for TPSS

The schematic of TPSS is shown in Fig. 1(a). It can be regarded as a two-node system with an equivalent source E_{eq} and its corresponding impedance Z_{eq}. U_{TPSS}, I_{TPSS}, P_{TPSS} and Q_{TPSS} are the voltage, current, active power and reactive power at the feeder, respectively. The equivalent circuit is depicted in Fig. 1(b) with respect to the feeder. The SCC, denoted as S_{eq}, can be calculated by

$$S_{eq} = \|E_{eq}\|_2^2 / \|Z_{eq}\|_2 \tag{1}$$



(a) The schematic of actual TPSS



(b) The equivalent circuit of TEM

Fig. 1. Schematic and TEM circuit of TPSS.

where $\|\cdot\|_2$ is 2-norm. The system model can be written as

$$E_{eq} = U_{TPSS} + Z_{eq} I_{TPSS} \tag{2}$$

Assumption I: For TSM, the parameters in TEM are constant in two adjacent sampling periods, written as

$$\begin{cases} E_{eq}^{(k)} = E_{eq}^{(k+1)} \\ Z_{eq}^{(k)} = Z_{eq}^{(k+1)} \end{cases} \tag{3}$$

where k and $k+1$ represent two adjacent sampling times. TEM can be written as

$$\begin{cases} Z_{eq}^{(k)} = -\dfrac{U_{TPSS}^{(k+1)} - U_{TPSS}^{(k)}}{I_{TPSS}^{(k+1)} - I_{TPSS}^{(k)}} \\ E_{eq}^{(k)} = \dfrac{U_{TPSS}^{(k)} I_{TPSS}^{(k+1)} - U_{TPSS}^{(k+1)} I_{TPSS}^{(k)}}{I_{TPSS}^{(k+1)} - I_{TPSS}^{(k)}} \end{cases} \tag{4}$$

Since *Assumption I* cannot be satisfied in most scenarios, the real characteristic cannot be identified accurately by TSM. Samples with appropriate disturbance need to be selected to reduce the error in practical application. Otherwise, there will be following scenarios where TEM of source cannot be identified accurately.

Scenario I: If there is large disturbance on the load power, *Assumption I* is not valid due to the change of TEM parameters.

Scenario II: If there is almost no disturbance on the load power, (4) will approach the form of 0/0, resulting in the parameter drift issue.

Scenario III: If there is no disturbance on the load side and only disturbance occurs in the source side, the value of Z_{eq} equals to the load impedance.

2.2. Characteristic analysis and algorithm framework

To solve the above problem, characteristics of TPSS are analyzed and the SFLS algorithm is proposed for TEM and SCC identification. Electrical characteristics of a Substation in XS Railway under on-load and no-load conditions are shown in Figs. 2 and 3, respectively. Under on-load conditions, power fluctuates frequently within (1.82, 13.95) MVA, and voltage fluctuates within (27.12, 29.28) kV. The correlation coefficient of −0.8845 between their variations indicates that it is suitable to identify TEM parameters. However, the load power of no-load conditions is too small to induce voltage fluctuations with a correlation of −0.0161. They are greatly affected by external disturbances. At around 6:24 in Fig. 3(a), although the traction load power was relatively low, a significant voltage variation occurred, primarily due to substantial fluctuations on the power supply side. To improve the robustness of algorithm in such cases, a statistical filtering method is adopted.

To improve the accuracy and scenario compatibility of LSM for TEM, the SFLS algorithm framework is proposed by integrating statistical

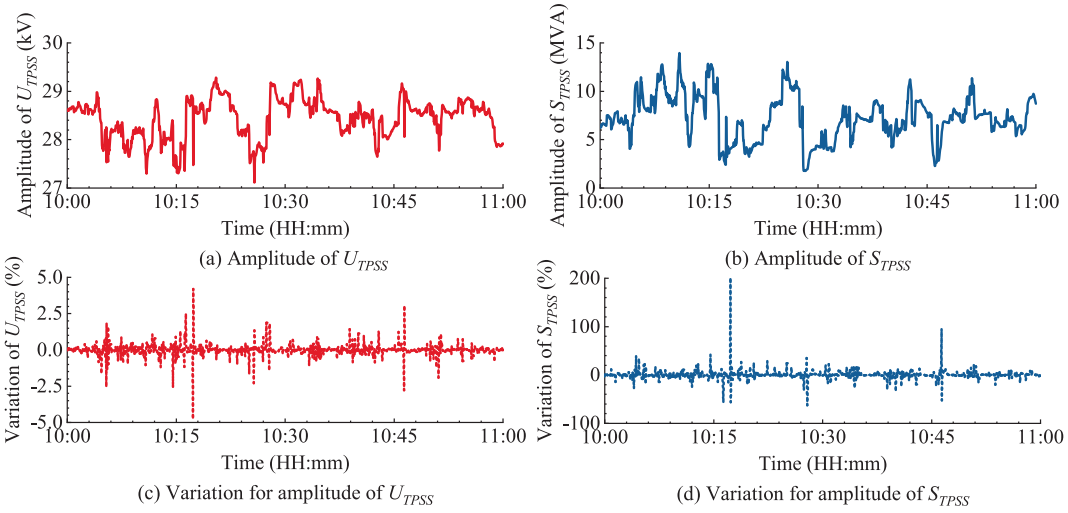


Fig. 2. Electrical characteristics at feeders under on-load conditions.

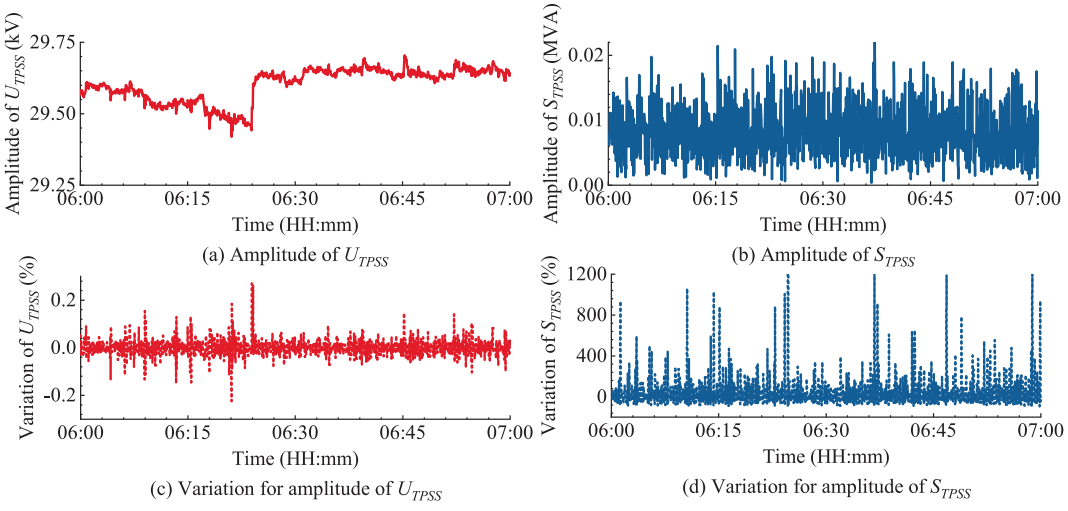


Fig. 3. Electrical characteristics at feeders under no-load conditions.

filtering processes and dynamic correction mechanisms, depicted in Fig. 4. To mitigate rounding errors, the field data is normalized in the first stage. Then the critical challenge in TEM identification, the parameter drift issue caused by inconsistent power-voltage variations, is addressed through two statistical rules. The load power threshold and the box plot are adopted for eliminating outliers with poor correlation between power and voltage variations. For certain time intervals with absence or limited availability of data, the obtained TEM through the LSM is assessed using the fitting quality (R^2). Results with low R^2 are corrected by probability distribution of historical results. Due to the threshold setting and actual scenarios where identified parameters fluctuate frequently, Kalman Filter (KF) is incorporated to smooth transient fluctuations in identified TEM parameters. The similarity between results and reference values is evaluated by Jensen-Shannon divergence (JSD).

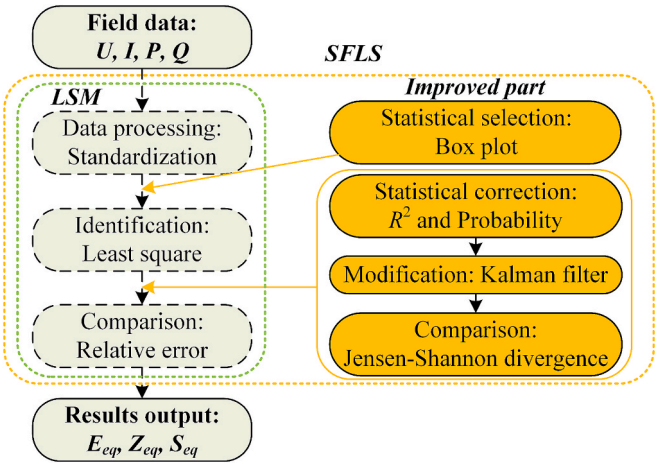
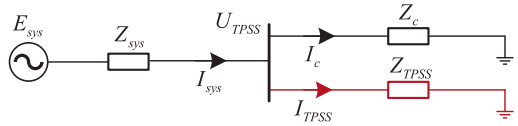


Fig. 4. Algorithm framework of proposed SFLS.

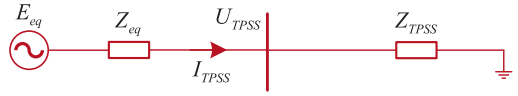
3. Statistical filtering-based least square algorithm

3.1. Improved LSM based on box plots

The equivalent circuit of TPSS is shown in Fig. 5(a). E_{sys} and Z_{sys} are the equivalent voltage and system impedance of remote power system respectively. I_{sys} is the total load current. Z_c denotes the equivalent load impedance of the upstream and adjacent nodes that affect U_{TPSS}. I_c is the corresponding load current. Z_{TPSS} is the local load impedance. The TEM is shown in Fig. 5(b). Since E_{sys} and Z_{sys} are usually adjusted by grid



(a) The equivalent circuit of TPSS



(b) The equivalent circuit of TEM

Fig. 5. Equivalent circuits of TPSS and TEM.

dispatch centers, they are fixed within several adjacent samples. E_{eq} and Z_{eq} are mainly influenced by Z_c, shown as

$$\begin{cases} Z_{eq} = (Z_{sys}Z_c)/(Z_{sys}+Z_c) \\ E_{eq} = (E_{sys}Z_c)/(Z_{sys}+Z_c) \end{cases} \tag{5}$$

Due to the randomness of Z_c, the solving equation considering fluctuations of source side is shown as

$$\begin{cases} E_{eq}^{(k)} = Z_{eq}^{(k)} I_{TPSS}^{(k)} + U_{TPSS}^{(k)} \\ E_{eq}^{(k+1)} = Z_{eq}^{(k+1)} I_{TPSS}^{(k+1)} + U_{TPSS}^{(k+1)} \end{cases} \tag{6}$$

where

$$\begin{cases} E_{eq}^{(k+1)} = E_{eq}^{(k)} + \Delta E_{eq}^{(k)} \\ Z_{eq}^{(k+1)} = Z_{eq}^{(k)} + \Delta Z_{eq}^{(k)} \end{cases}$$

where Δx represents the variation of x. x can be E_{eq}, Z_{eq}, U_{TPSS} or I_{TPSS}.

According to (6), $Z_{eq}^{(k+1)}$ is obtained as

$$Z_{eq}^{(k+1)} = -\frac{U_{TPSS}^{(k+1)} - U_{TPSS}^{(k)}}{I_{TPSS}^{(k+1)} - I_{TPSS}^{(k)}} + \frac{\Delta E_{eq}^{(k)} - I_{TPSS}^{(k)} \Delta Z_{eq}^{(k)}}{I_{TPSS}^{(k+1)} - I_{TPSS}^{(k)}} \tag{7}$$

Since relative variation ratios of E_{eq} and Z_{eq} are almost same, the product of $I_{TPSS}^{(k)}$ and $\Delta Z_{eq}^{(k)}$ is

$$I_{TPSS}^{(k)} \Delta Z_{eq}^{(k)} = \frac{Z_{eq}^{(k)}}{Z_{TPSS}^{(k)} + Z_{eq}^{(k)}} \Delta E_{eq}^{(k)} \tag{8}$$

$Z_{TPSS}^{(k)}$ is much larger than $Z_{eq}^{(k)}$ in actual TPSS, thus (8) equals to zero. $Z_{eq}^{(k+1)}$ can be rewritten as

$$Z_{eq}^{(k+1)} \approx \frac{\Delta E_{eq}^{(k)} - \left(U_{TPSS}^{(k+1)} - U_{TPSS}^{(k)}\right)}{I_{TPSS}^{(k+1)} - I_{TPSS}^{(k)}} \tag{9}$$

Two rules of statistical filtering need to be satisfied for minimizing influence of $\Delta E_{eq}^{(k)}$.

Rule I: The change in power is large to avoid a 0/0 situation.

Rule II: The consistency between voltage and power variations is great, e.g., small power fluctuations with small voltage fluctuations and vice versa.

According to the analysis in Section 2.2, a power threshold λ is defined to select samples which satisfies

$$S_{TPSS} > \lambda \tag{10}$$

This guarantees the selected data can reduce errors caused by other disturbance. Furthermore, the box plot is adopted to eliminate data for satisfying *Rule II*. A ratio of variations δ_k is defined as

$$\delta_k = \left\| \frac{S_{TPSS}^{(k)} \Delta U_{TPSS}^{(k)}}{U_{TPSS}^{(k)} \Delta S_{TPSS}^{(k)}} \right\|_2 \tag{11}$$

As shown in Fig. A1 of Appendix A, the box plot measures the data dispersion and variability by constructing quartiles. Samples are divided into four intervals. The interquartile spacing *IQR* is used to distinguish outliers, defined as

$$IQR = Q_3 - Q_1 \tag{12}$$

where Q_1 is the upper quartile, Q_3 the lower quartile, σ the standard deviation. The available data is defined as

$$Q_1 - 1.5IQR \leqslant \delta_k \leqslant Q_3 + 1.5IQR \tag{13}$$

After the statistical filtering, samples that satisfy *Rule I* and *Rule II* are obtained. The objective function of TEM is constructed as

$$\min f(Z_{eq}, E_{eq}) = \sum_{k=1}^{m} \left(E_{eq} - Z_{eq} I_{TPSS}^{(k)} - U_{TPSS}^{(k)}\right)^2 \tag{14}$$

where m is the total number of post-filtered samples. The solution is

$$\begin{bmatrix} E_{eq} & -Z_{eq} \end{bmatrix}^{\mathrm{T}} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y} \tag{15}$$

where

$$\begin{cases} \boldsymbol{X}_{m\times 2} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ I_{TPSS}^{1} & I_{TPSS}^{2} & \cdots & I_{TPSS}^{m} \end{bmatrix}^{\mathrm{T}} \\ \boldsymbol{Y}_{m\times 1} = \begin{bmatrix} U_{TPSS}^{1} & U_{TPSS}^{2} & \cdots & U_{TPSS}^{m} \end{bmatrix}^{\mathrm{T}} \end{cases}$$

The robustness is poor if there is a significant change of E_{eq} and Z_{eq} caused by the dispatch center in a time window. To avoid this condition, R^2 is used to evaluate the credibility of regression results, defined as

$$R^2 = 1 - \frac{\sum_{k=1}^{m} \left(U_{TPSS}^{(k)} - \widehat{U}_{TPSS}^{(k)}\right)^2}{\sum_{k=1}^{m} \left(U_{TPSS}^{(k)} - \overline{U}_{TPSS}^{(k)}\right)^2} \tag{16}$$

where "^" and "-" stand for estimated and average values. If $R^2 < \varepsilon$, pre-modified results need to be corrected using the probability distribution of historical parameters. Real-time data correction has been widely applied in multiple problem-solving scenarios, which is vital for improving the performance of algorithms [30].

3.2. SFLS with Kalman filter

To improve the accuracy further, KF is applied in this paper to smooth the obtained data from the above subsection. The state variable x represents E_{eq}, Z_{eq}, and S_{eq}. KF consists of two steps, "*Prediction*" and "*Correction*" [31,32].

Prediction: The current state $\widetilde{x}_k$ and the priori error covariance $\widetilde{p}_k$ are obtained as

$$\begin{cases} \widetilde{x}_k = a x_{k-1} + e_k \\ \widetilde{p}_k = a p_{k-1} a + q \end{cases} \tag{17}$$

where a is the state transfer coefficient, e_k the prediction error, and q is the covariance of process errors.

Correction: The Kalman gain g_k, the optimal estimation value x_k, and the posteriori estimation covariance p_k are obtained as

$$\begin{cases} g_k = \widetilde{p}_k h / \left(h \widetilde{p}_k h + r\right) \\ x_k = \widetilde{x}_k + g_k \left(z_k - h\widetilde{x}_k\right) \\ p_k = (1 - g_k h)\widetilde{p}_k \end{cases} \tag{18}$$

where z_k is the observation value, r the average observation error, and h is the elastic coefficient.

3.3. The proposed algorithm procedure

The procedure of the proposed SFLS (shown in Fig. 6) can be summarized as:

Step 1: Apply standardization to reduce the rounding error.

Step 2: Select data satisfy *Rule I* and *Rule II* to reduce the error caused by parameter drift and low consistency.

Step 3: Apply the least square regression to identify TEM parameters and SCC using the post-filtered data.

Step 4: Calculate R^2 and compare it with ε to reduce the error caused by the parameter mutation in a time window. If $R^2 > \varepsilon$, repeat *Step 2* for the next time window until all time windows are finished. If $R^2 < \varepsilon$, modify TEM and SCC by the probability distribution.

Step 5: Adopt KF to smooth identified TEM parameters and SCC.

This paper employs second-level data granularity for SCC identification, a choice balanced among the algorithm's principle, data analysis needs, and engineering feasibility. The Thevenin equivalent-based algorithm focuses on steady-state voltage fluctuations, requiring no millisecond-level transient data. Second-level data sufficiently captures sustained power changes (e.g., train traction) while filtering high-frequency noise; minute-level data lacks the resolution for detailed fluctuation analysis, whereas millisecond-level data imposes excessive storage and computational burdens. Furthermore, second-level data is a standard output of existing railway monitoring systems, facilitating practical engineering application and ensuring a balance between accuracy, efficiency, and feasibility.

3.4. Performance evaluation method

Considering the random nature of SCC in real TPSS scenarios, it is typically treated as a statistical value over a period. To assess the performance of various approaches, an index is introduced to quantify the similarity between random variables. Generally, JSD is used to measure the distance of two probability distributions [33,34]. For two probability distributions (α, β) of a discrete random variable ξ, the *JSD* of α and β is defined as

$$\begin{cases} JSD = 0.5\left(\sum_k\left(\alpha(k)\log\left(\frac{\alpha(k)}{\tau(k)}\right) + \beta(k)\log\left(\frac{\beta(k)}{\tau(k)}\right)\right)\right) \\ \tau(k) = 0.5(\alpha(k) + \beta(k)) \end{cases} \tag{19}$$

The closer its value is to 0, the more similar the two probability distributions are.

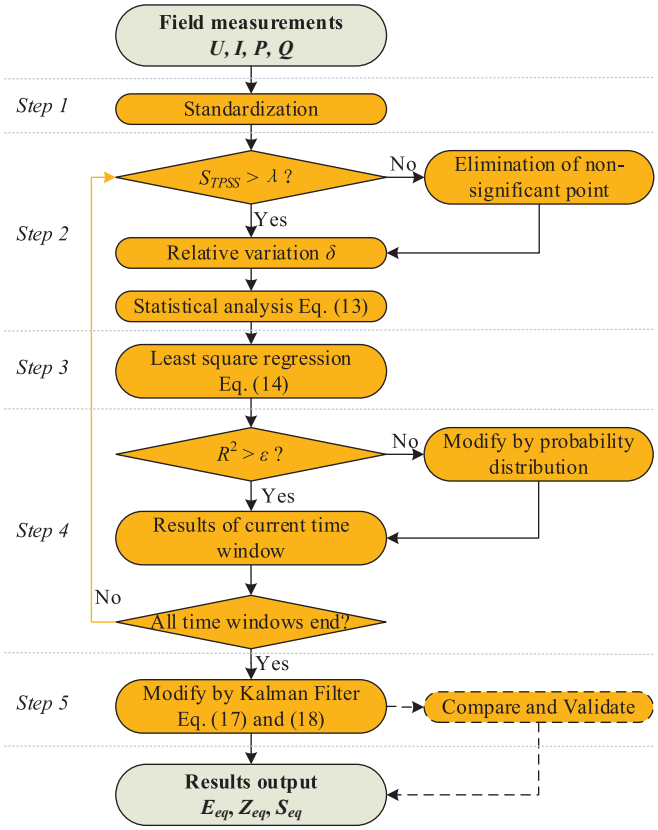


Fig. 6. The algorithm procedure diagram of SFLS.

4. Case study

Simulation and experimental data are used to verify the effectiveness of the proposed SFLS. Experimental data is obtained from a substation in Railway XS. To strengthen the availability and reduce the preliminary error caused by the asynchronous data, the data analysis equipment adopts EMAP electrical comprehensive test software and HS4 data acquisition device with 4 sampling channels. The sampling rate is 20 kSps, the sampling accuracy is 0.0015 %, and the signal–noise ratio is 96 dB. The equipment installed in the Railway XS is shown in Fig. 7. Parameter settings of the proposed algorithm are shown in Table 1. The variations in magnitude and phase angle represent stochastic changes or fluctuations in the upstream grid of actual TPSS. Simulated parameters are set within following constraints: $E_{eq}\in(25.00, 30.00)$ kV, $Z_{eq}\in(0.02 + j1.95, 0.20 + j6.80)$ Ω. The thresholds are set as: $\lambda = 0.02$ per unit (p.u.), $\varepsilon = 0.8$.

4.1. Simulated case

In this section, two cases considering the dual fluctuations of source-load side are established. Base values of voltage, impedance and power are 27.5 kV, 27.5 Ω and 27.5 MVA, respectively. The per unit values of E_{eq} and Z_{eq} are shown in Table 1. The power applied in simulations is actual load power of TPSS.

Numerical mutation is applied to simulate changes of TEM. Noise generated by random distribution is utilized to simulate external interference. To demonstrate the robustness, *Case I* represents the scenario with adverse noise. The amplitude noise of E_{eq} and Z_{eq} is $(-0.01, 0.01)$ p.u., and the phase angle noise of them is $(-1, 1)$ degree. *Case II* is a common noise scenario in TPSS. The noise is $(-0.001, 0.001)$ p.u. for amplitude. Since the identification error is proportional to the

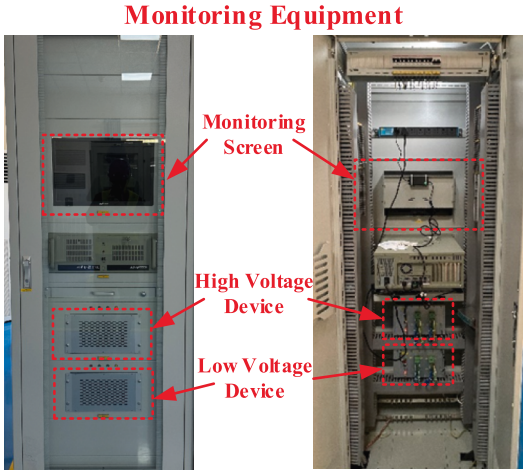


Fig. 7. Equipment installed in the Railway XS.

Table 1
Parameters setting of simulation cases.

Interval	E_{eq} Amplitude (p.u.)	Angle (Deg)	Z_{eq} Amplitude (p.u.)	Angle (Deg)
17:00 ~ 19:20	1.000	0.000	0.182	85.944
19:20 ~ 21:40	0.964	1.000	0.218	83.079
21:40 ~ 00:00	0.927	1.000	0.255	83.079

interference, the subsequent analysis is presented based on *Case I* to show the performance of proposed SFLS in the scenario with adverse noise.

The simulated data of a time window is shown in Fig. 8(a). The time window is set as 10 min. The relationships of ΔU_{TPSS} and ΔS_{TPSS} with the ratio concentrated on the range of (0.26, 10.78) are depicted in Figs. 8 (b) and 8(c), respectively. Anomalies occur at 17:04:21 with a value of 117.65 due to significant changes in S_{TPSS} and reverse changes in E_{eq}. After filtering by box plot, remained data are shown in Fig. 8(d) where it accounts for 90.5 %. Further, with the least square regression, we can obtain: E_{eq} = 0.9960p.u., Z_{eq} = 0.2393p.u., S_{eq} = 4.1455p.u., R^2 = 0.9787. From the results, it shows *Step 3* can identify the real characteristics of power system. However, R^2 is low in the scenario contains mutation of TEM parameters. *Step 4* is used to modify these results. Fig. 9 (a) illustrates the variation of R^2 throughout the entire duration 17:00 ~ 00:00. It can be seen that in some periods, such as around 19:20, R^2 reaches a lower value. When t = 19:13:16, R^2 is approximately 0.7991. Probability distributions of E_{eq}, Z_{eq}, S_{eq} are shown in Fig. 9(b). Then, identified parameters are modified as: E_{eq} = 0.9999p.u., Z_{eq} = 0.1883p.u., S_{eq} = 5.2101p.u.

Fig. 10 illustrates the results with and without KF. The mutation of identified parameters is unreasonable obviously. The magnitude of mutation is reduced with KF, which is appliable for improving the robustness of SFLS. Taking the transition period from 19:10 to 19:20 as an example, the extreme deviation of S_{eq} is 2.75 and the coefficient of variation is 0.21 after adopting KF. The stability is improved compared to the original result (3.20, 0.33) without KF. Fig. 11 shows the TEM parameters obtained from SFLS. JSD and average relative error of several algorithms are depicted in Tables 2 and 3, respectively. Due to the assumption of fixed TEM and parameter drift, traditional TSM and GM have larger errors. LSM is also inaccurate because it contains points with low consistence of ΔU_{TPSS} and ΔS_{TPSS} and mutations of E_{eq} and Z_{eq}. The SFLS achieves a 9.43 % improvement in accuracy compared with LSM. The improvement is more significant for *Case II*, which is around 28.18 %.

4.2. Field test

Fig. 12(a) depicts the field data of a substation in Railway XS. After the statistical filtering, R^2 is shown in Fig. 12(b) and is poor during intervals 19:28 ~ 19:30 and 19:45 ~ 19:54. The former is mainly caused by the low load power, while the latter is caused by the random mutation affected by adjacent loads. Despite all this, results are more stable after the statistical correction and filtering process, as illustrated in Fig. 12(c). Fig. 12(d) and Table 4 present the SCC identification results of different algorithms. To objectively evaluate accuracy, the results of each method were compared with the measured benchmark value (6.91p.u.) obtained through short-circuit testing, and their relative errors were calculated. Additionally, Table 4 lists the standard deviations of the results from each method to assess their volatility.

5. Discussion and future works

5.1. Sensitivity analysis

According to the case study, the data causing large errors is eliminated by the thresholds, so the accuracy of identification is great. The sensitivity analysis needs to be studied for the robustness of the SFLS. Results are shown in Table 5. The relative error is smaller with the larger λ, but the standard deviation is worse. Besides, λ is a threshold to eliminate the data with small load power. Thus, it cannot be set as a large value. ε is set for selecting the identified TEM parameters, so the relative error and standard deviation get smaller when ε increases. However, the accuracy is worse due to the strict value, e.g., ε = 0.9. To ensure the great applicability and accuracy of proposed SFLS, thresholds can be set within following constraints: $\lambda \in (0.01, 0.03)$ p.u., $\varepsilon \in (0.8, 0.9)$.

Although the boundary coefficient for outliers in box plots is empirically determined based on statistical experience, various modified versions of box plots with different coefficient settings have been developed in several applications. Given that this coefficient may influence algorithm performance, an analysis of identification results under different coefficient settings was conducted based on empirical data. As shown in Table 6, the error of identification results remains stable at approximately −8.00 % across different threshold settings, with a standard deviation around 0.75. This demonstrates the robustness of the proposed algorithm.

5.2. Supplementary works

The accuracy of proposed algorithm has been verified from the perspective of system-side SCC through both simulation and field test cases. In practice, SCC represents only one of the performance indicators for TPSSs. Other critical metrics include voltage unbalance and short-circuit ratio. A comprehensive analysis incorporating multiple

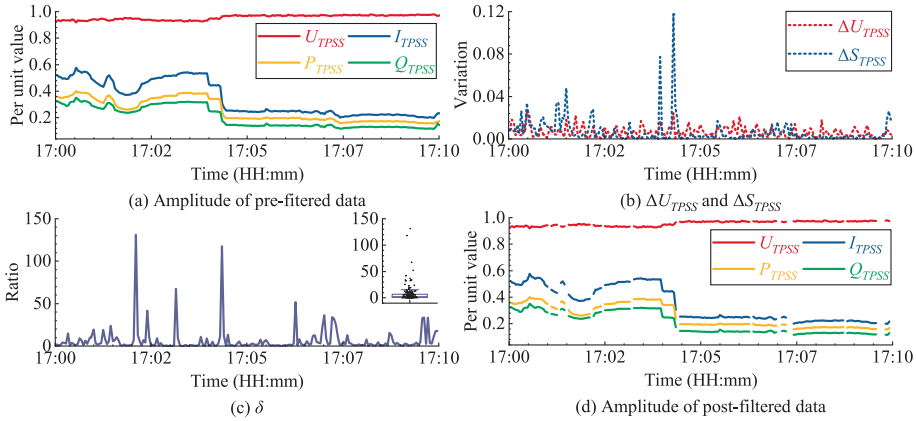


Fig. 8. Pre-filtered data and post-filtered data in a time window.

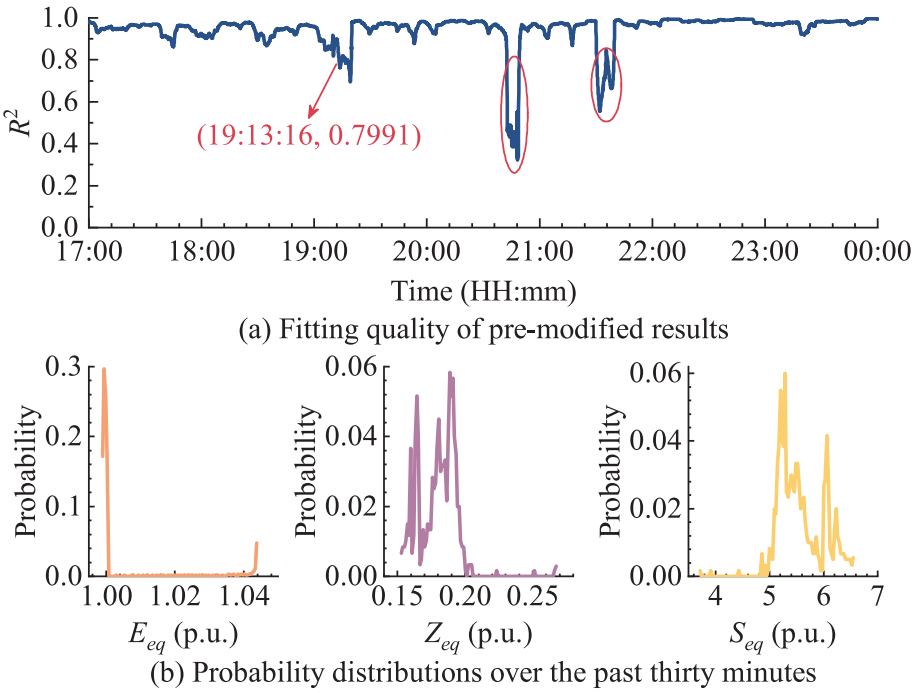


Fig. 9. R^2 and the probability distribution for correction.

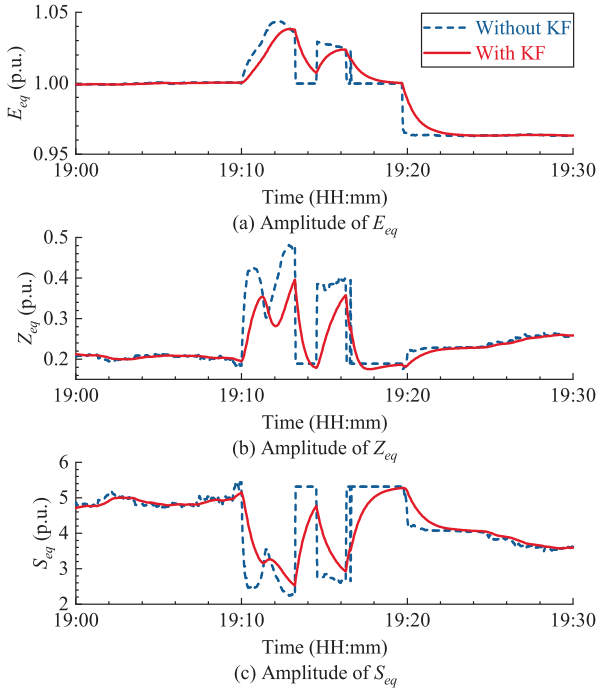


Fig. 10. Identified TEM parameters without and with KF.

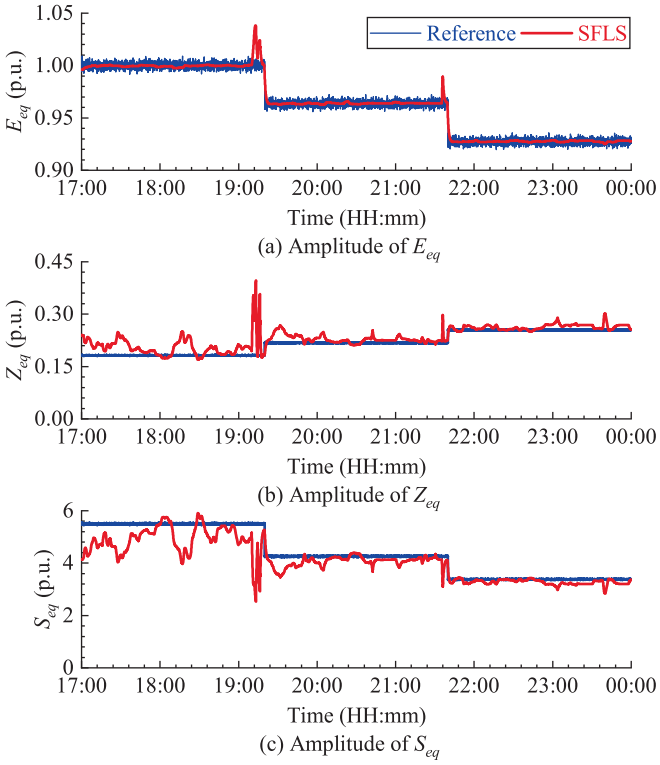


Fig. 11. TEM parameters obtained from the proposed SFLS.

performance indicators could further enhance the evaluation accuracy of system power supply performance. Such an approach may also serve as supplementary validation to indirectly verify the analytical results of individual metrics, which will be addressed in subsequent research to achieve precise assessment of TPSS performance.

Beyond multi-metric analysis, the proposed algorithm is designed for real-time online applications. Thus, the data quality influences analytical accuracy. Asynchronous sampling poses a significant challenge for data-driven algorithms. It typically occurs when measurement data are acquired from different monitoring terminals. However, the time-series samples analyzed in the SFLS are collected from a single monitoring terminal, ensuring high inter-channel synchronization that does not compromise the performance of the proposed algorithm. Besides, the measurement error is also a challenge. Although statistical filtering has been incorporated into the SFLS algorithm to mitigate measurement error impacts, errors remain unavoidable. Enhancing data acquisition quality in monitoring systems will be another key research direction of

Table 2
JSD of algorithms with reference value.

	Parameter	SFLS	LSM	GM	TSM
Case I	E_{eq}	8.766e-6	1.897e-5	1.793	0.064
	Z_{eq}	0.004	0.011	0.969	1.038
	S_{eq}	0.004	0.009	3.664	0.892
Case II	E_{eq}	9.408e-6	1.535e-5	2.905	0.055
	Z_{eq}	0.006	0.015	6.636	1.022
	S_{eq}	0.003	0.007	1.378	0.507

Table 3
Average relative error of algorithms.

	Parameter	SFLS	LSM	GM	TSM
Case I	E_{eq} (%)	0.051	0.066	40.372	1.409
	Z_{eq} (%)	7.378	9.761	4.991e3	1.044e3
	S_{eq} (%)	−6.246	−6.897	−88.942	−46.484
Case II	E_{eq} (%)	0.056	0.062	113.571	−6.162
	Z_{eq} (%)	2.720	4.845	1.756e4	842.611
	S_{eq} (%)	−1.831	−2.548	−71.472	−53.027

future work for the research team.

6. Conclusion

This paper proposes a SFLS algorithm for determining SCC based on the monitoring data of feeders. Targeting issues in TPSSs such as the weak correlation between voltage and power and easily drifting parameters caused by intense fluctuations on both source and load sides, this algorithm constructs a systematic solution. Its core innovation lies in overcoming the applicability limitations of traditional methods in "dual-side fluctuation" scenarios. On one hand, by designing statistical filtering rules based on load power thresholds and box plots, it effectively suppresses errors introduced by asynchronous changes in voltage and power, establishing a foundation for the stable application of the LSM under extreme operating conditions. On the other hand, introducing a real-time correction mechanism based on historical probability distribution solves the issue of missing and inaccurate result in the real-time application. Since the SFLS does not involve complicated differential operations, the time cost of identification is low. The performance of SFLS is validated by simulations with different disturbances. Compared with the traditional method, improvements in accuracy are 9.43 % for *Case I* and 28.18 % for *Case II*, respectively. Besides, identified SCC based on filed data is roughly the same as the actual value with an improvement of 15.76 %. To ensure the applicability in the other systems, the sensitivity analysis of thresholds is carried out. This algorithm not only provides a reliable basis for voltage stability and relay protection analysis in electrified railways, but its integrated "Filtering-Identification-Correction" framework also holds reference value for

Table 4
Identified SCC of algorithms in field application.

Parameter		SFLS	LSM	GM	TSM
Amplitude S_{eq} (p.u.)	Average (p.u.)	6.356	6.252	1.540	1.578
	Standard deviation (p.u.)	0.749	1.248	4.824	2.560
	Relative error	−8.02 %	−9.52 %	−77.71 %	−77.16 %

Table 5
Sensitivity results of the proposed SFLS.

λ (p.u.)	ε	Amplitude of S_{eq} (p.u.)		
		Average	Standard deviation	Relative error
0.01	0.80	6.356	0.749	−8.02 %
0.02	0.80	6.356	0.749	−8.02 %
0.03	0.80	6.339	0.788	−8.26 %
0.04	0.80	6.443	0.761	−6.76 %
0.05	0.80	6.489	0.789	−6.09 %
0.02	0.60	6.307	1.012	−8.73 %
0.02	0.70	6.395	0.845	−7.45 %
0.02	0.80	6.356	0.749	−8.02 %
0.02	0.85	6.370	0.700	−7.81 %
0.02	0.90	5.775	0.702	−16.71 %
0.02	0.95	5.541	0.454	−19.81 %

Table 6
Sensitivity results of the proposed SFLS.

Boundary coefficient	Amplitude of S_{eq} (p.u.)		
	Average	Standard deviation	Relative error
1.0	6.363	0.745	−7.92 %
1.1	6.366	0.751	−7.87 %
1.2	6.354	0.760	−8.04 %
1.3	6.370	0.753	−7.81 %
1.4	6.347	0.748	−8.15 %
1.5	6.356	0.749	−8.02 %
1.6	6.361	0.746	−7.95 %
1.7	6.359	0.750	−7.97 %
1.8	6.358	0.752	−7.99 %
1.9	6.400	0.742	−7.38 %
2.0	6.404	0.742	−7.33 %

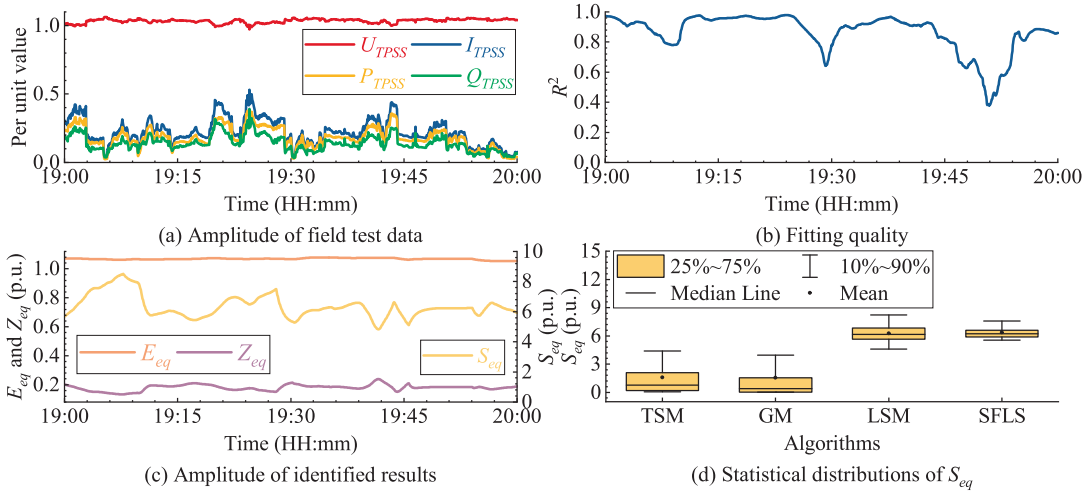


Fig. 12. Identified results of TEM and SCC at the actual substation.

parameter identification in similar volatile systems.

CRediT authorship contribution statement

Shaobing Yang: Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Liang Hu:** Project administration, Funding acquisition, Conceptualization. **Wanqi Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Tingting He:** Writing – review & editing, Writing – original draft, Visualization, Supervision. **Dylan Dah-Chuan Lu:** Writing – review & editing, Writing – original draft. **Mingli Wu:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the China Energy Investment Corporation (**GJNY-21-189**).

Appendix A

Fig. A1 depicts the box plot measuring the data dispersion and variability by constructing quartiles. Samples are divided into four intervals. The interquartile spacing *IQR* is used to distinguish outliers.

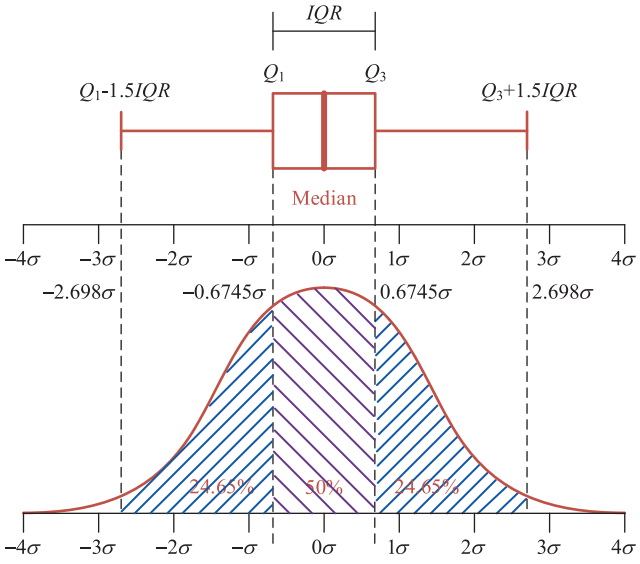


Fig. A1. The schematical diagram of box plots.

Data availability

The authors do not have permission to share data.

References

[1] S. D' Arco, L. Piegari, and P. Tricoli. Comparative analysis of topologies to integrate photovoltaic sources in the feeder stations of AC railways. IEEE Trans. Transp. Electrific., Dec. 2018;4(3):951–60.
[2] Deng W, Dai C, Chen W, Gao S. Experimental investigation and adaptability analysis of hybrid traction power supply system integrated with photovoltaic sources in AC-fed railways. IEEE Trans. Transp. Electrific. Sept. 2021;7(3):1750–64.
[3] Ning F, Ji L, Ma J, Jia L, Yu Z. Research and analysis of a flexible integrated development model of railway system and photovoltaic in China. Renew Energy Sept. 2021;175:853–67.
[4] Liu Y, Chen M, Cheng Z, Chen Y, Li Q. Robust energy management of high-speed railway co-phase traction substation with uncertain PV generation and traction load. IEEE Trans Intell Transp Syst Jun. 2022;23(6):5070–91.
[5] Lu S, Xu Z, Xiao L, Jiang W, Bie X. Evaluation and enhancement of control strategies for VSC stations under weak grid strengths. IEEE Trans Power Syst Mar. 2018;33(2):1836–47.
[6] Yu L, Sun H, Xu S, Zhao B, Zhang J. A critical system strength evaluation of a power system with high penetration of renewable energy generations. CSEE J Power Energy Syst May 2022;8(3):710–20.
[7] Xu Z, Bian Z, Cheng B. An approach to the ultimate goal of power grid development—constant voltage operation. CSEE J Power Energy Syst Dec. 2017;3(4):380–9.
[8] Lin J, Hu S, Li Y, Luo L, Zhang J, Cao Y, et al. A novel power programming strategy for railway power regulation with dynamic exploration. IEEE Trans Smart Grid Jul. 2022;13(4):2798–811.
[9] Chen L, Chen M, Cheng Y, Chen Y, Cheng Y, Zhao N. Modelling and control of a novel AT-fed co-phase traction power supply system for electrified railway. Int J Electr Power Energy Syst Aug. 2021;125(106405):1–9.
[10] Burchett SM, Douglas D, Ghiocel SG, Liehr MWA, Chow JH, Kosterev D, et al. An optimal Thévenin equivalent estimation method and its application to the voltage stability analysis of a wind hub. IEEE Trans Power Syst Jul. 2018;33(4):3644–52.
[11] Hu F, Sun K, Del Rosso A, Farantatos E, Bhatt N. Measurement-based real-time voltage stability monitoring for load areas. IEEE Trans Power Syst Jul. 2016;31(4):2787–98.
[12] Zhang H, Xiao G, Liu Z, Zhou Y. Impedance-based stability analysis and on-site stability evaluation of three-phase active voltage conditioner embedded system. IEEE Trans Power Electron Dec. 2023;38(12):16061–71.
[13] Yun Z, Cui X, Ma K. Online Thevenin equivalent parameter identification method of large power grids using LU factorization. IEEE Trans Power Syst Nov. 2019;34(6):4464–75.
[14] Abdi P, Hamedani-Golshan M-E, Alhelou HH, Milano F. A PMU-based method for on-line Thévenin equivalent estimation. IEEE Trans Power Syst Jul. 2022;37(4):2796–807.
[15] Su H-Y, Liu T-Y. Robust Thevenin equivalent parameter estimation for voltage stability assessment. IEEE Trans Power Syst Jul. 2018;33(4):4637–9.

[16] Cobreces S, Bueno EJ, Pizarro D, Rodriguez FJ, Huerta F. Grid impedance monitoring system for distributed power generation electronic interfaces. IEEE Trans Instrum Meas Sept. 2009;58(9):3112–21.
[17] Fusco G, Losi A, Russo M. Constrained least squares methods for parameter tracking of power system steady-state equivalent circuits. IEEE Trans. Power Del. Jul. 2000;15(3):1073–80.
[18] Peng L, Zhao J, Tang Y, Mili L, Gu Z, Zheng Z. Real-time LCC-HVDC maximum emergency power capacity estimation based on local PMUs. IEEE Trans Power Syst Mar. 2021;36(2):1049–58.
[19] Alinezhad B, Kazemi Karegar H. On-line Thévenin impedance estimation based on PMU data and phase drift correction. IEEE Trans. Smart Grid, Mar. 2018;9(2):1033–42.
[20] Abdelkader SM, Morrow DJ. Online Thévenin equivalent determination considering system side changes and measurement errors. IEEE Trans Power Syst Sept. 2015;30(5):2716–25.
[21] Abdelkader SM, Morrow DJ. Online tracking of Thévenin equivalent parameters using PMU measurements. IEEE Trans Power Syst May 2012;27(2):975–83.
[22] Arefifar SA, Xu W. Online tracking of power system impedance parameters and field experiences. IEEE Trans. Power Del. Oct. 2009;24(4):1781–8.
[23] Mohammadi Y, Miraftabzadeh SM, Bollen MHJ, Longo M. An unsupervised learning schema for seeking patterns in rms voltage variations at the sub-10-minute time scale. Sustain. Energy Grids Netw. Sep. 2022;31(100773):1–16.
[24] Mohammadi Y, Miraftabzadeh SM, Bollen MHJ, Longo M. Seeking patterns in rms voltage variations at the sub-10-minute scale from multiple locations via unsupervised learning and pattern's post-processing. Int J Electr Power Energy Syst Dec. 2022;143(108516):1–24.
[25] Mohammadi Y, Leborgne RC, Polajzer B. Modified methods for voltage-sag source detection using transient periods. Electr Pow Syst Res Jun. 2022;207(107857):1–19.
[26] Matavalam ARR, Ajjarapu V. Sensitivity based Thevenin index with systematic inclusion of reactive power limits. IEEE Trans Power Syst Jan. 2018;33(1):932–42.
[27] Tang Y, Sun H, Yi J, Lin W. "Tracing algorithm for Thevenin equivalent parameters based on complete differential equation," (In Chinese). Proc CSEE May 2009;29(13):48–53.
[28] Wang Y, Xu W, Yong J. An adaptive threshold for robust system impedance estimation. IEEE Trans Power Syst Sept. 2019;34(5):3951–3.
[29] Moghimi Haji M, Xu W. Online determination of external network models using synchronized phasor data. IEEE Trans. Smart Grid, Mar. 2018;9(2):635–43.
[30] Wang S, Gao H, Takyi-Aninakwa P, Guerrero JM, Fernandez C, Huang Q. Improved multiple feature-electrochemical thermal coupling modeling of lithium-ion batteries at low-temperature with real-time coefficient correction. Prot Control Mod Power Syst May 2024;9(3):157–73.
[31] Wang S, Zhang S, Wen S, Fernandez C. An accurate state-of-charge estimation of lithium-ion batteries based on improved particle swarm optimization-adaptive square root cubature kalman filter. J Power Sources Dec. 2024;624(235594):1–11.
[32] Wang S, Dang Q, Gao Z, Li B, Fernandez C, Blaabjerg F. An innovative square root-untraced Kalman filtering strategy with full-parameter online identification for state of power evaluation of lithium-ion batteries. J. Energy Storage Dec. 2024;104(114555):1–20.
[33] Li Y, Cao W, Hu W, Wu M. Abnormality detection for drilling processes based on Jensen–Shannon Divergence and adaptive alarm limits. IEEE Trans. Ind. Inform. Sept. 2021;17(9):6104–13.
[34] Chen H, Xu H, Yang Z. A novel hybrid method to detect arrival times of elastic waves with low SNR based on Jensen–Shannon Divergence and cumulative sum algorithm. IEEE Trans Instrum Meas Oct. 2022;71:1–12.