Electrical Power and Energy Systems 136 (2022) 107613

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

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

Power system anomaly detection using innovation reduction properties of iterated extended kalman filter

Zhaoyang Jin [a], Junbo Zhao [b], Lei Ding [a,*], Saikat Chakrabarti [c], Elena Gryazina [d], Vladimir Terzija [d]

[a] School of Electrical Engineering, Shandong University, 17923 Jingshi Road, Jinan, China
[b] Department of Electrical & Computer Engineering, University of Connecticut, Storrs, CT 06269, United States
[c] Department of Electrical Engineering, Indian Institute of Technology, Kanpur 208016, India
[d] Center for Energy Science and Technology, Skoltech, Moscow, Russian Federation

ARTICLE INFO

Keywords:
Anomaly detection
Forecasting-aided state estimator
IEKF
Normalized innovation
Phasor measurement unit
State estimation

ABSTRACT

The performance of forecasting-aided state estimation can be significantly affected by the presence of anomalies, such as sudden load changes, bad data, or line outage/topology errors. The existing robust alternatives may suppress the influences of some anomalies, but not all of them, while the anomaly detection methods have very low accuracy or even may not detect, e.g., line outages. To bridge this gap, this paper proposes a new anomaly detection framework using the properties of innovation reduction in iterated extended Kalman filter (IEKF) and the normalized residual of static state estimator. The proposed framework can detect and distinguish above-mentioned anomalies with a very high accuracy. It is also robust to different levels of noise, different degrees of measurement redundancies, and different sizes/complexities of the networks. Simulation results carried out on the IEEE 14-, 39-, 57-, and 118-bus test systems demonstrated the effectiveness and accuracy of the proposed method.

1. Introduction

Since its introduction in control rooms in 1970s [1], the static state estimation (SE) [2,3] has been a basic tool for power system operation and control. Besides the static SE, the forecasting-aided state estimator (FASE) has been paid increasing attention, which integrates both state prediction and online measurements for better capabilities in handling anomalies [4–7]. FASE was firstly proposed in [8] but not widely used for practical systems due to its computational requirements. However, this may not be an issue for today's systems and distributed implementations can also be leveraged to mitigate that.

The performance of FASE can be affected by various types of anomalies, including sudden load changes (SLCs), bad data, and sudden topology change caused by line outage. This calls for development of robust FASEs with appropriate detection schemes. In [9], a robust FASE is proposed to suppress outliers by weighting the measurements according to the exponential of the negative absolute values of their residuals [10]. A robust FASE based on a generalized maximum likelihood estimator is later proposed in [11], which enhances the robustness of the FASE using projection statistics and spatial and temporal correlations among the measurements. Recently, a new FASE based on the adaptive H-infinity extended Kalman filter is proposed in [12] to deal with the model uncertainties. The common problem with existing robust FASEs is that they cannot simultaneously handle sudden load and topology changes. As shown in [13], although a robust FASE can suppress the errors caused by SLC, it is much less accurate than the static SE due to the use of erroneous predictions. In the presence of line outages, the situation is even worse; the robust FASEs and the topology processors may not effectively detect them, resulting in persistent large errors.

There are two types of methods to detect and distinguish anomalies, including bad data and SLC. The first one is initially proposed in [14] and has been adopted by recent papers, such as [15]. It uses normalized innovations and their skewness to detect and distinguish bad data and SLC. The second type is proposed in [16], and further tested in [6,17]. It uses both normalized innovations and residuals with fixed thresholds to detect these two types of anomalies. The drawback of the first type is that the skewness is not a reliable criterion to distinguish SLC and bad data, because an SLC in a large network can lead to a large skewness, and

* Corresponding author.
E-mail address: dinglei@sdu.edu.cn (L. Ding).

https://doi.org/10.1016/j.ijepes.2021.107613
Received 5 June 2021; Received in revised form 29 July 2021; Accepted 13 September 2021
Available online 1 October 2021

a group of evenly distributed bad data can lead to small skewness. The latter is against the detection criteria. By contrast, for the second type, it is established under the assumption that the normalized residuals will not exceed a fixed threshold in the SLC case. In fact, if the SLC is very large, the normalized residuals can also be very large, and the distribution of the normalized residuals in presence of SLC has significant overlap with that in the bad data case. Also, none of the methods can detect sudden topology change caused by line outage, another type of anomaly that occurs frequently in power systems.

In this paper, a novel framework based on the innovation reduction properties of iterated extended Kalman filter (IEKF) is proposed to detect and distinguish all three types of anomalies. The contributions are summarized as follows:

- The innovation reduction properties and the properties related to the largest normalized residuals test for the static SE are thoroughly investigated. Although the largest normalized residual method for the detection of bad data is effective [18] and there are several methods developed for topology error detection and [19,20], they cannot distinguish bad data, topology errors, and SLC simultaneously, which is a challenging but open problem. This pushes further the adoption of FASE for practical power system applications.
- Comparison of results with the existing approaches show that our proposed method can achieve much higher accuracy and it can effectively distinguish three types of anomalies under different noise levels, different degrees of network observability, and different networks.

This paper is organized as follows: Section 2 introduces the formulation of IEKF and analyzes its innovation reduction properties. Section 3 presents the proposed anomaly detection algorithm. Section 4 shows the simulation results. Finally, Section 5 concludes the paper.

2. IEKF and its innovation reduction properties

This section presents the mathematical formulation of the system model, IEKF algorithm and investigates its innovation reduction properties in the anomaly cases.

2.1. System state-space model

The power system state evolution model can be described by the following state-space form:

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}_k \boldsymbol{x}_k + \boldsymbol{g}_k + \boldsymbol{q}_k \tag{1}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k, k) + \boldsymbol{r}_k \tag{2}$$

$$E\left[\boldsymbol{q}_{k-1}\boldsymbol{q}_{k-1}^T\right] = \boldsymbol{Q}_{k-1} \tag{3}$$

$$E\left[\boldsymbol{r}_{k-1}\boldsymbol{r}_{k-1}^T\right] = \boldsymbol{R}_{k-1} \tag{4}$$

where $\mathbf{x}$ is the state vector; $\mathbf{F}$ and $\mathbf{g}$ are the transition matrix and input vector; $\mathbf{z}$ is the measurement vector; $\mathbf{h}$ is the nonlinear measurement function; $\mathbf{q}$ and $\mathbf{r}$ are the system process and measurement noise error vectors; their covariance matrices are $\mathbf{Q}$ and $\mathbf{R}$, respectively; the subscript k denotes the instant. By using Holt's smoothing method [21], $\mathbf{F}$ and $\mathbf{g}$ can be estimated via:

$$\boldsymbol{F}_k = \alpha_k(1+\beta_k)\boldsymbol{I} \tag{5}$$

$$\boldsymbol{g}_k = (1+\beta_k)(1-\alpha_k)\boldsymbol{x}_k^- - \beta_k \boldsymbol{a}_{k-1} + (1-\beta_k)\boldsymbol{b}_{k-1} \tag{6}$$

where $\mathbf{I}$ is an identity matrix; α_k and β_k are factors in the range from 0 to 1; $\mathbf{x}_k^-$ is the predicted state at time k. The parameter vectors $\mathbf{a}$ and $\mathbf{b}$ at time k can be estimated as follows:

$$\boldsymbol{a}_k = \alpha_k \boldsymbol{x}_k + (1-\alpha_k)\boldsymbol{x}_k^- \tag{7}$$

$$\boldsymbol{b}_k = \beta_k(\boldsymbol{a}_k - \boldsymbol{a}_{k-1}) + (1-\beta_k)\boldsymbol{b}_{k-1} \tag{8}$$

In this paper, $\alpha_k = 0.8$, $\beta_k = 0.5$ are used [6].

2.2. Iterated extend Kalman filter

Based on the state-space model presented in (1)–(4), the IEKF algorithm can be implemented in three steps: (1) state prediction via (9)–(10); (2) iterative state updating using (11)–(13), and (3) state vector estimation via (15)–(16). Detailed implementation steps are explained in the following algorithm:

Algorithm 1. (*Iterated Extended Kalman Filter*)

Input $\mathbf{a}_0 = \hat{\mathbf{x}}_{-1|-1}$, $\mathbf{b}_0 = 0$, $\hat{\mathbf{x}}_{0|0} = \hat{\mathbf{x}}_{0|-1} = \mathbf{x}_0$, $\mathbf{P}_{0|0} = \mathbf{Q}_0$, $\mathbf{R}_l$, $\mathbf{z}_l$ $(l \geq 1)$.
Output $\hat{\mathbf{x}}_{l|l}$, $\mathbf{P}_{l|l}$,$(l \geq 1)$.
1: $k = 1$.
2: **while** $k \geq 1$ **do**
3: $\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{g}_{k-1}$ (9)
$\boldsymbol{P}_{k|k-1} = \boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_{k-1}^T + \boldsymbol{Q}_k$ □ (10)
5: $i = 1$, $\hat{\boldsymbol{x}}_{k|k}^0 \triangleq \hat{\boldsymbol{x}}_{k|k-1}$, stop = 0.
6: **while** stop==0 **do**
7: $\hat{z}_k^i = \boldsymbol{h}\left(\hat{\boldsymbol{x}}_{k|k}^{i-1}\right)$ (11)
8: $\boldsymbol{H}_k^i = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{s}}(s)\big|_{s=\hat{x}_{k|k}^{i-1}}$. (12)
9: $\boldsymbol{G}_k^i = \boldsymbol{P}_{k|k-1}\left(\boldsymbol{H}_k^i\right)^T\left[\boldsymbol{H}_k^i\boldsymbol{P}_{k|k-1}\left(\boldsymbol{H}_k^i\right)^T + \boldsymbol{R}_k\right]^{-1}$ (13)
10: $\hat{\boldsymbol{x}}_{k|k}^i = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{G}_k^i\left[z_k - \hat{z}_k^i - \boldsymbol{H}_k^i\left(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k}^{i-1}\right)\right]$ (14)
11: **if** $max\left(\left|\hat{\boldsymbol{x}}_{k|k}^i - \hat{\boldsymbol{x}}_{k|k}^{i-1}\right|\right) < 10^{-6}$, **then** stop = 1.
12: $i = i + 1$.
13: **end while**
14: $\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k}^{final}$ □ (15)
15: $\boldsymbol{P}_{k|k} = \left(\boldsymbol{I} - \boldsymbol{G}_k^{final}\boldsymbol{H}_k^{final}\right)\boldsymbol{P}_{k|k-1}$ (16)
16: Return $\hat{\boldsymbol{x}}_{k|k}$and $\boldsymbol{P}_{k|k}$
17: $k = k + 1$.
18: **end while**

In the algorithm, $\mathbf{G}$ is the Kalman gain; $\mathbf{P}$ is the state estimation covariance matrix; "$\hat{}$" denotes estimated value; the superscript "i" represents the ith iteration; $k|k-1$ denotes the prediction for instant k based on all information at instant $k-1$, and $k|k$ denotes the estimate at instant k based on all information.

Remark 1: The IEKF estimate is the final estimate of the weighted least squares (WLS) estimator solved with the Newton-Raphson method, whose measurements include both the regular measurements $\mathbf{z}_k$ and the prediction $\hat{\mathbf{x}}_{k|k-1}$. The estimate of the extended Kalman filter (EKF) is the estimate of the WLS in the first round of iteration. The proof can be found in [22].

Remark 2: The system process error covariance matrix $\mathbf{Q}_k$ is set according to the system condition in the normal case (quasi-steady state condition), and thus is assumed to be constant. This ensures good estimation accuracy for the normal case and high sensitivity for the anomaly cases. In this paper, no correlations between states and that between measurements are considered, so both $\mathbf{Q}_k$ and $\mathbf{R}_k$ are diagonal.

2.3. The innovation reduction properties of IEKF

To introduce the concept of the innovation reduction property, two variables are first defined: (1) the normalized innovation, i.e., the normalized value of the predicted measurement error in (17) (2) the normalized residual, i.e., the normalized value of the estimated measurement error in (18) [6]. Then, the innovation reduction factor, *Inred*, is defined as the ratio between the max value of $\mathbf{r}$ and that of $\boldsymbol{\lambda}$: $Inred = \max(\mathbf{r})/\max(\boldsymbol{\lambda})$.

$$\boldsymbol{\lambda} = \left(\sqrt{\boldsymbol{P}^{zz}_{k|k-1}}\right)^{-1} \left|z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k-1}\right)\right| \tag{17}$$

$$\boldsymbol{r} = \left(\sqrt{\boldsymbol{P}^{zz}_{k|k}}\right)^{-1} \left|z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)\right| \tag{18}$$

where $\boldsymbol{P}^{zz}_{k|k-1}$ and $\boldsymbol{P}^{zz}_{k|k}$ are the covariance matrices of $z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k-1}\right)$ and $z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)$, respectively.

The *Inred* of IEKF has the following properties:

$$Inred \to 1 \text{ for SCADA bad data} \tag{19}$$

$$Inred \to (0, 1] \text{ for PMU bad data} \tag{20}$$

$$Inred \to 0 \text{ for sudden load change} \tag{21}$$

$$Inred > 0 \text{ for line outage} \tag{22}$$

For the SCADA with bad data case, there are large errors in the measurements, i.e., $z_k = z_{k,true} + \boldsymbol{\varepsilon}$, where $z_{k,true}$ is the true measurement vector with large error $\boldsymbol{\varepsilon}$. Since more accurate PMU measurements are given higher weights and the iterative nature of the IEKF can average out the large errors, the residual would be much smaller than the error for the bad data: $|z_{k,true} - h(\widehat{x}_{k|k})| \ll \varepsilon$. It is straightforward to see that $|z_{k,true} - h(\widehat{x}_{k|k})| \ll \varepsilon$, because the prediction is accurate. Thus, (19) can be derived as follows:

$$Inred = \lim_{\varepsilon \to \infty} \frac{max\left(\left|z_{k,true} + \varepsilon - h\left(\widehat{x}_{k|k}\right)\right| \Big/ \sqrt{P^{zz}_{k}}\right)}{max\left(\left|z_{k,true} + \varepsilon - h\left(\widehat{x}_{k|k-1}\right)\right| \Big/ \sqrt{P^{zz}_{k|k-1}}\right)} \approx \frac{max(\varepsilon)}{max(\varepsilon)} = 1 \tag{23}$$

For the PMU with bad data case, $\left|z_{k,true} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k-1}\right)\right| \ll \boldsymbol{\varepsilon}$ still holds since the prediction is still accurate. However, much more weights are given to the PMU measurements so that the estimated measurements are close to the measurements with errors: $z_{k,true} + \boldsymbol{\varepsilon} \to \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)$. The strength of this approximation is decided by the redundancy of PMU measurements and the number of PMU measurements that have bad data. If the redundancy of PMU measurements is low, with many of the PMU measurements having bad data, this approximation will be strong, and $Inred \to 0$. If the redundancy of PMU measurements is high, while there are only a few bad PMU measurements, the residual would be much smaller than the error for the bad data: $\left|z_{k,true} - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)\right| \ll \boldsymbol{\varepsilon}$, so that $Inred \to 1$. Eq. (20) is thus justified.

In the case of a sudden load change, there are large errors in the predictions: $\boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k-1}\right) = \boldsymbol{h}\left(\boldsymbol{x}_{k,true}\right) + \boldsymbol{\varepsilon}$. The iterative nature of the IEKF allows it to average out the large errors, so that $\left|z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)\right| = \boldsymbol{\delta} \ll \boldsymbol{\varepsilon}$, and (19) can be derived as follows:

$$Inred = \lim_{\varepsilon \to \infty} \frac{\max\left(\left|z_k - \boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k|k}\right)\right| \Big/ \sqrt{\boldsymbol{P}^{zz}_{k}}\right)}{\max\left(\left|z_k - \left(\boldsymbol{h}\left(\widehat{\boldsymbol{x}}_{k,true}\right) + \boldsymbol{\varepsilon}\right)\right| \Big/ \sqrt{\boldsymbol{P}^{zz}_{k|k-1}}\right)} = \lim_{\varepsilon \to \infty} \frac{max(\boldsymbol{\delta})}{max(\boldsymbol{\varepsilon})} = 0 \tag{24}$$

In the topology change case caused by line outage, errors might exist in both the predictions and measurements with arbitrary magnitudes. So, *Inred* can be of any value greater than 0.

3. Proposed anomaly detection algorithms

3.1. Anomaly detection

(1) Properties related to the Normalized Residual of Static SE

To facilitate the detection and distinction of different types of anomalies, an additional property related to the normalized residual of the static SE ($\mathbf{r}_{sse}$) is discussed as follows:

- In the bad data and line outage cases, max(rsse) will be significantly increased. In the normal case and SLC case, rsse is only affected by measurement errors.
- The normal case means the situation where no anomalies happen and the results are only affected by the measurement noises. In this paper, the line outage is taken into account only after the circuit breaker opens the branch. Before that, there is no topology change in the network.

(2) Proposed Detection Algorithm

The detection algorithm is shown in Fig. 1. The detection of the normal case is by simply comparing the largest normalized residual of the IEKF, max($\mathbf{r}$), with a fixed statistical threshold. The detection and distinguishing of the other types of anomalies are done according to the innovation reduction properties of the IEKF and the property of the normalized residuals of the static SE.

The detection principles given in Fig. 1 are summarized in Fig. 2 below, where result 1 corresponds to the green area, result 2 corresponds to the blue area, and result 3 corresponds to the red area. The detailed analysis is given as follows:

- Since SLC does not affect the static SE, the normalized residual of the static SE follows normal Gaussian distribution. A threshold t2 can be set to guarantee that an SLC case would not be mis-detected as bad data case for a sufficiently large number of detections.
- An inred threshold t3 can be used to distinguish SLC and bad data based on the following properties: (1) inred in the case of bad SCADA data approximates to 1; (2) bad PMU data have very large max(rsse) due to their large weights, and (3) inred in the case of SLC approximates to 0, as shown in the area under the orange curves.
- In the cases of SLC and bad data, inred lies in the range between 0 and 1. A threshold t4, which is greater than 1, can be set to distinguish line outages from other types of anomalies. However, this detection threshold is optional because there could be a very small number of bad data cases (outliers, less than 0.1%) in which inred > t4. This detection threshold can be adopted if such a rate of incorrect detection can be tolerated.
- As shown in the yellow dashed rectangle, a more reliable method to detect all line outages is to check the values of two successive normalized residuals of the static SE. Because bad data are unlikely to occur at different measurements continuously and sudden load

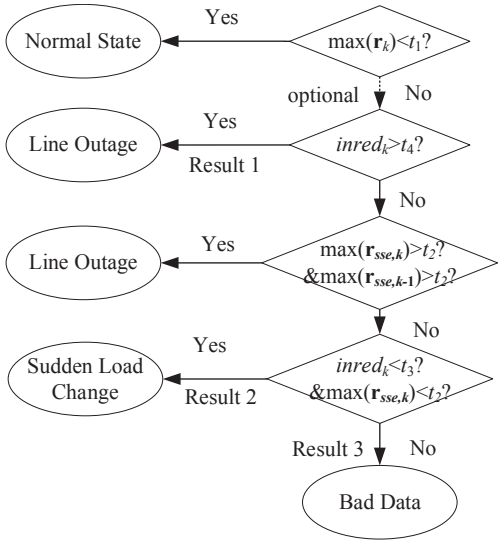

Fig. 1. The flow chart of the detection algorithm, where t_1, t_2, t_3 and t_4 are the detection thresholds.

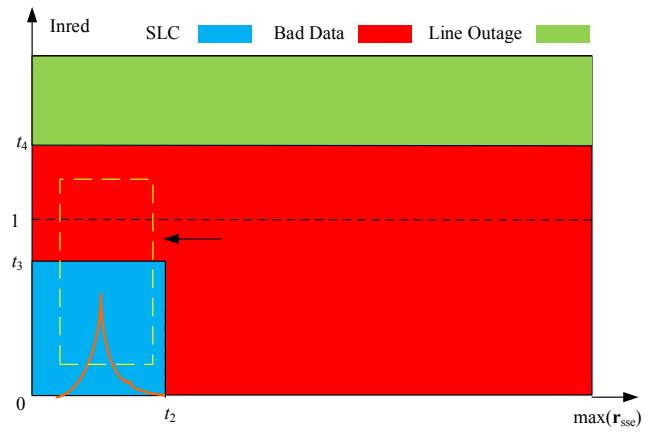

Fig. 2. *Inred* – max($\mathbf{r}_{sse}$) plot of the detection algorithm.

change cannot make max(rsse,k) > t2, if max(rsse,k − 1) > t2 and max(rsse,k) > t2, a line outage case can be confirmed.

where the yellow dashed rectangle represents the most likely distribution area of the *Inred* – max($\mathbf{r}_{sse}$) points for bad data cases at the next instant after the anomaly happens, and the orange curve stands for the most likely distribution area for sudden load change cases (Fig. 3).

(3) Determination of the Detection Thresholds

This subsection discusses the principles for determining detection thresholds and they are listed in Table 1. The proposed algorithm uses 3 major thresholds and 1 optional threshold to detect and discriminate the

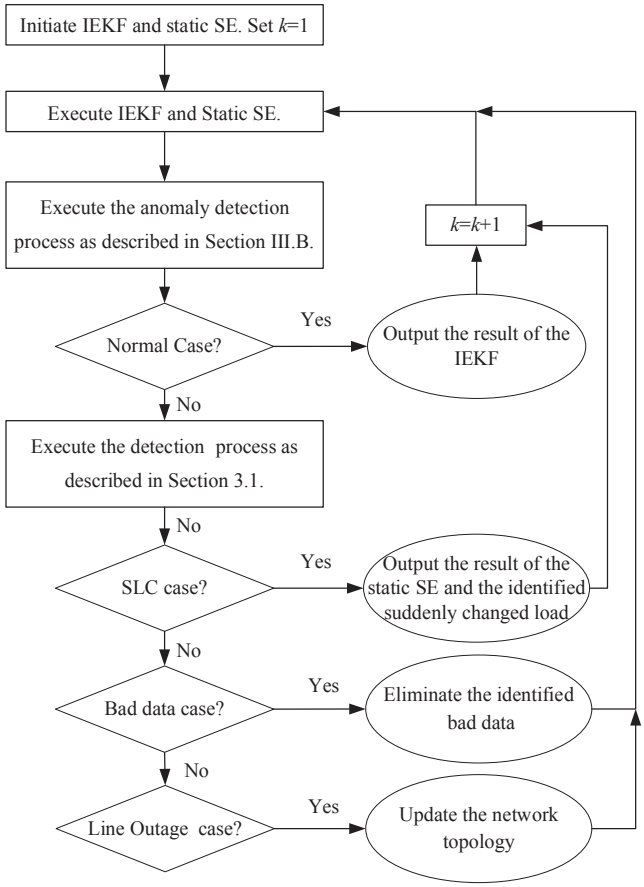

Fig. 3. Flowchart of the proposed FASE algorithm.

Table 1
Detection thresholds for the simulations.

Threshold	t_1	t_2	t_3	t_4
Value	4	7	0.75	1.25

anomalies. Note that the thresholds given in Table 1 is not to eliminate wrong detections but to minimize the rate of wrong detections. The threshold selection methodologies given below are based on conservative statistic analysis, which might not give the optimal threshold but can guarantee good performance in terms of distinguishing different types of anomalies.

The first threshold t_1 is used to distinguish the normal and anomaly cases. In the normal case, where there is only measurement noise, t_1 is usually set to 3 according to the 3-sigma rule. The possibility of any normalized residual of the IEKF being greater than 3 is small (less than 0.3% detection error probability). In the real networks, the number of measurements might be over 1000 or even 10,000, especially when both SCADA and PMU measurements are used for state estimation. This may lead to normal cases being detected as anomaly cases with some probabilities. In this paper, the threshold is set to 4 to further reduce the error detection probability, which is appropriate for large networks.

The second and third thresholds, t_2 and t_3, respectively, are used to distinguish the SLC and bad data cases. To guarantee that for 10,000 detections, SLC would not be detected as bad data cases for a network with 10,000 measurements, the threshold t_2 should be set at the value of 6, since the possibility of any max(rsse) > 6 is about $1/5.068 * 10^8$. Large SLC in the major loads of the network could make the gain matrix ill-conditioned, which leads to larger normalized residuals. t_2 is set to 7 to tackle the challenges in these extreme conditions. However, small bad data in the SCADA measurements could also make max($\mathbf{r}_{sse}$) < t_2. To distinguish them from the SLC cases, the inred values should be compared with t_3. If the strength of the two approximations shown in (19) and (20) were the same, t_3 should be set to 0.5. However, as analyzed in Section 2.3, inred in the SCADA bad data case approximates very close to 1. In contrast, inred in the small SLC case has a weak approximation to 0, since it has a small impact on the system states. Thus, t_3 should be between 0.5 and 1, and is set to be 0.75 in this paper, which is the average value of 0.5 and 1.

The threshold t_4 is for quick detection of line outages. The value of t_4 should be sufficiently larger than 1 so that bad data can be reliably distinguished from line outages, but it should not be much larger than 1 so as to maximize the proportion of line outages that can be detected with just one state estimation run. Thus, t_4 is set to be 1.25.

3.2. Implementation of the proposed FASE algorithm

The proposed FASE algorithm is shown in the following flowchart.

4. Simulation results

4.1. Simulation settings

The validity of the proposed anomaly detection and distinguishing method is demonstrated under various conditions: (i) comparisons with two existing methods in Sections 4.2; (ii) different noise levels in Section 4.3, (iii) different degrees of measurement observability in Section 4.4, and (iv) different networks in Section 4.5. Note that for Sections 4.2 to 4.6, all simulations are performed on IEEE 118-bus system. Specific settings will be shown in each subsection.

All simulations consider four operation scenarios. In the normal operation case, 200 repeated simulations are performed. The SCADA with bad data cases are considered in two sub-cases: For sub-case 1, N bad data are generated randomly by setting N SCADA measurements to be 0, and the number of simulations is set to be 200 for each N. For sub-case 2, 5 bad data are generated randomly by changing 5 SCADA

measurements to different proportions of their original values ranging from 10% to 90%, and the number of simulations is set to be 200 for each percentage (2600 simulations in total). For the PMU bad data case, it is assumed that only a small number of PMU measurements can have bad data because it is more reliable. Thus, 5 PMU measurements are changed to different proportions of their original values ranging from 10% to 90%; and the number of simulations is set to be 200 for each percentage (1800 simulations). The SLC case includes 11 sub-cases, which have different load change ratios (LCR), namely the load after the change divided by the load before the change, ranging from 0 to 0.95. In each sub-case, the simulations change one of the loads to the given LCR, and every load is tested for SLC (2200 simulations). In the line outage case, one of the branches is disconnected, and every branch is tested (186 simulations). Note that for the detection accuracy results summarized in Table 4, Table 6, Table 7 and Table 9, quick detection of line outage using the threshold t_4 is not adopted. Without loss of generality, it is assumed that all anomalies in the simulations happen at instant k, where k can be any integer.

4.2. Validation of the proposed method

In this sub-section, the innovation reduction properties of the IEKF and the property of max($\mathbf{r}_{sse}$) will be assessed first. Then, the detection outcomes will be assessed in the second part and compared with two existing methods. Detailed settings of this simulation are given below. It is assumed that the measurement errors follow Gaussian distribution, and their standard deviation is equal to the absolute value of the measurement times a given coefficient according to different types of measurements. Their coefficients are shown in Table 2. The measurement placement for the IEEE 118-bus test system is shown in Table 3, where the traditional SCADA measurements provide full measurement redundancy while the system is barely observable by 32 PMUs. Therefore, the system measurement redundancy is 3, i.e., there are 3 independent sets of measurements that can make the network observable, where SCADA and PMU data provide two and one sets, respectively.

(1) Assessment of the Innovation Reduction Properties

The innovation reduction properties and the properties related to the normalized residual of the static SE, as stated in Section 2, are demonstrated in Figs. 4–7. For the SLC case, both the average and median values of *inred* are smaller than 0.1 when LCR < 0.6, and they are about 0.2 even when LCR $= 0.95$ (refer Fig. 4). Hence, the first innovation reduction property is validated. For the multiple SCADA bad data case shown in Fig. 5, both the average and median values of *inred* are almost exactly 1 regardless of a few outliers when the number of bad data is small. Therefore, the third innovation reduction property is validated.

The points defined by (*inred*, max($\mathbf{r}_{sse}$)) (I-R) for every anomaly case at the instant when the anomaly happens (t_k) are plotted in Fig. 6. It can be seen that *inred* values of the multiple bad data cases are all in the range of (0,1]. Therefore, the second innovation reduction property is validated. The I-R plot in Fig. 6 also shows that the largest max($\mathbf{r}_{sse}$) in all SLC cases is about 5, while the smallest max($\mathbf{r}_{sse}$) in all SCADA bad data cases is approximately 3. It is therefore shown that it is impossible to distinguish all SLC cases and bad data cases by just comparing the max ($\mathbf{r}_{sse}$) with a fixed threshold, and the possibility of misdetection is very high. This problem can be fully addressed by using both *inred* and max ($\mathbf{r}_{sse}$) properties. For the SCADA bad data, *inred* has been approximated close to 1 according to the second *inred* property; for the SLCs, *inred* approximates to 1 according to the first *inred* property, and large max ($\mathbf{r}_{sse}$) is more likely to happen for smaller *inred* values as a result of Gaussian distribution. It can also be seen from Fig. 6 that the *inred* values for only the line outage cases can be greater than 1.2, which demonstrates the quick detection capability of line outages of the proposed method.

Table 2
Coefficients to multiply with the measurement at a standard muesurement noise level.

SCADA P and Q measurements	0.02
SCADA voltage magnitude	0.002
PMU magnitude measurements	0.002
PMU angle measurements	0.01

The I-R points for all anomaly cases at t_{k+1} are plotted in Fig. 7. It can be seen that max($\mathbf{r}_{sse}$) of only the line outage case is greater than 7 at both instant k (the time instant when the anomaly happens) and instant $k+1$; max($\mathbf{r}_{sse}$) of the SLC case is smaller than 7 at both instants; while for bad data cases, although it can be greater than 7 at instant k, it is smaller than 7 at instant $k+1$. The capability to detect all line outages of the proposed method is therefore demonstrated.

(2) Anomaly Detection Results

The detection accuracies of the proposed method are compared with two existing methods. Method 1 [14,15] uses the normalized innovation and the skewness of normalized innovation to detect anomalies. Its detection accuracy in the SLC case is very poor because the skewness of normalized residual can be very large as long as the load change is not very small. Method 2 [6,16,17] uses the normalized innovation as well as normalized residual (NR) to distinguish SLC and bad data cases based on the assumption that the NR will still be very small in the SLC case. However, the NR in the SLC case can be very large, which makes it difficult to distinguish the SLC and bad data cases.

Table 4 shows that except for 3.68% of the SLC and 0.43% bad data cases not being detected (non-detection: anomaly cases detected as normal cases), the rest of them are all correctly detected by the proposed method that uses the *inred* properties and the property of max($\mathbf{r}_{sse}$). By comparison, the two existing methods can have over 80% of misdetection (one anomaly case detected as another type of anomaly case). The existing methods also wrongly detected almost 30% of the outliers caused by measurement errors and anomaly as they choose $t_1 = 3$ as the threshold. In comparison, the proposed method detected only about 4.5% of the normal cases as anomaly cases. It also shows that all line outage cases are also corrected by the proposed method. The two existing methods have 0 detection accuracy since they have not considered line outage cases.

Two hundred simulations are conducted to obtain the average execution time of the anomaly detection methods and FASE as shown in Table 5. Although the proposed method takes significantly more time than that of the existing methods, they are all negligible compared to the execution time for state estimation. Therefore, the proposed method method is feasible for online/realtime anomaly detection.

4.3. Sensitivity to different measurement noise levels

Two simulations are performed by keeping the other simulation settings the same as the simulation in Section 4.2. However, the measurement noise levels are increased to 3 and 10 times of that listed in Table 2, respectively. The I-R plots for them are shown in Fig. 8 and Fig. 9, respectively, and the detection results are given in Table 6.

The two figures show that as the noise level increases, max($\mathbf{r}_{sse}$) for bad data and line outage cases decreases, and the *inred* approximation effects also decrease. The former result can be explained by the fact that the same amount of error is normalized by a larger standard deviation; while the weaker approximation is because the errors become relatively smaller with the increase of standard deviation. Max($\mathbf{r}_{sse}$) for the SLC case is almost the same since it follows Gaussian distribution as previously explained. Therefore, the detection accuracy will decrease as the measurement noise level increases. However, the number of misdetections is still very small as compared to the correct detections as shown in Fig. 8 and Fig. 9.

Table 3
Measurement placement.

Injection measurements at bus #	1, 4, 7, …, $3[n_{bus}/3]+1$
Flow measurements at branch #	1, 2, 3, …, n_{branch}
Voltage Measurements at bus #	1, 6, 11, …, $5[n_{bus}/5]+1$
PMU placed at bus #	1;5;10;12;13;17;21;25;28;34;37;41;45;49;53;56;62;64;72;73;75;77;80; 85;87;91;94;101;105;110;114;116

Note n_{bus} denotes the number of buses in the network, n_{branch} denotes the number of branches in the network, "…" means "follows the same pattern until", and $[\cdot]$ represents the largest integer not greater than the given real number.

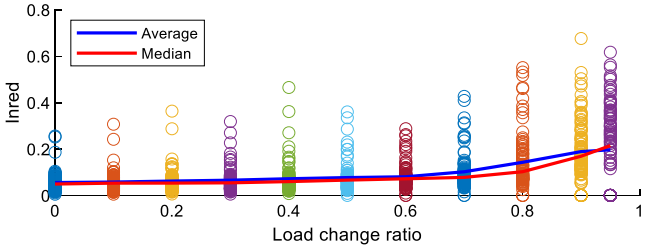

Fig. 4. *Inred* of the IEKF in the sudden load change case with different load change ratios at the instant when the anomaly happens.

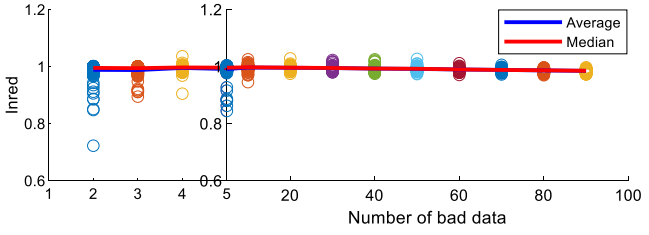

Fig. 5. *Inred* of IEKF in the bad data case with different number of SCADA bad data at the instant when the anomaly happens.

It is shown in Table 6 that even for 10 times the measurement error, the misdetections for the anomalies are only 1.28%, 0.07%, and 5.38% for SLC, bad data, and line outage cases, respectively. However, the non-detection rates have been significantly increased, since larger measurement noises make it harder to distinguish measurement noises and anomalies, especially for the SLC case. Its non-detection rate rises from 3.68% with the original noise level to 10.84% and 24.98% with 3 and 10

Table 4
Detection Accuracy of the Methods (Data in the Bracket Denote the Pecentage of Cases Detected as Normal Cases)

	Normal	SLC	Bad Data	Line Outage
Proposed Method	95.5%	96.32% (3.68%)	99.57% (0.43%)	100%
Method 1 [14,15]	70.5%	57.02% (1.28%)	17.75% (0.19%)	4.84%
Method 2 [6,16,17]	70.5%	11.94(1.28%)	99.81% (0.19%)	4.84%

Table 5
Average execution time of the anomaly detection methods and fase in miliseconds.

	Proposed Method	Method 1 [14,15]	Method 2 [6,16,17]	FASE
Time	0.7330	0.08526	0.06890	13958.5

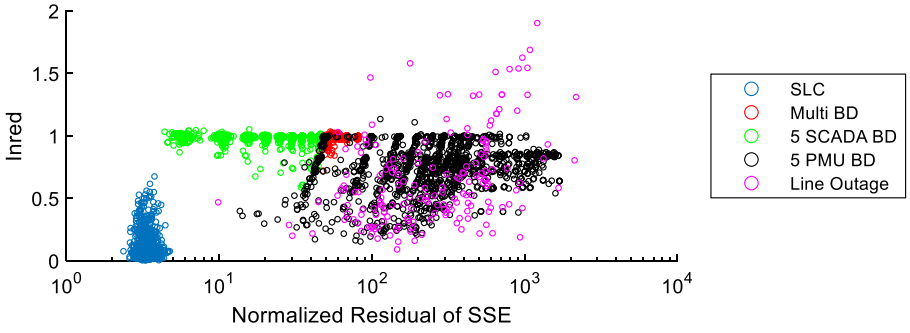

Fig. 6. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases at the instant when the anomaly happens.

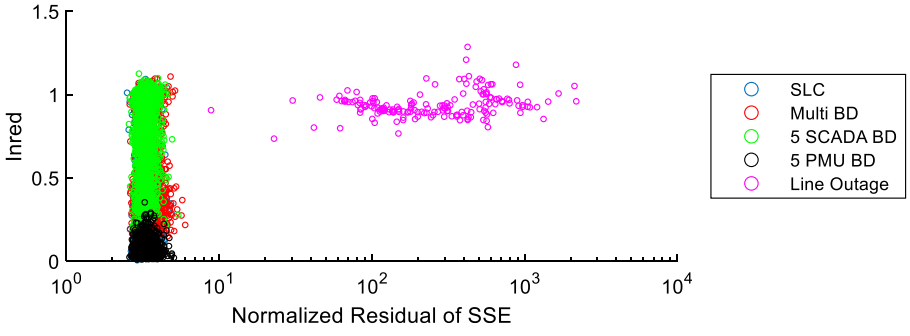

Fig. 7. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases at the next instant when the anomaly happens.

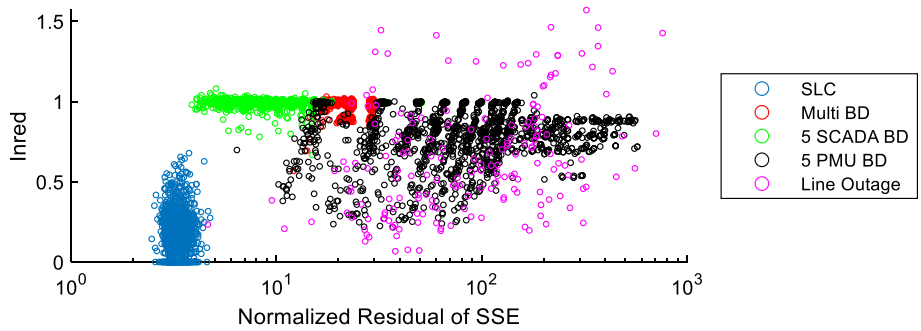

Fig. 8. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases for 3 times the measurement noise level at the instant when the anomaly happens.

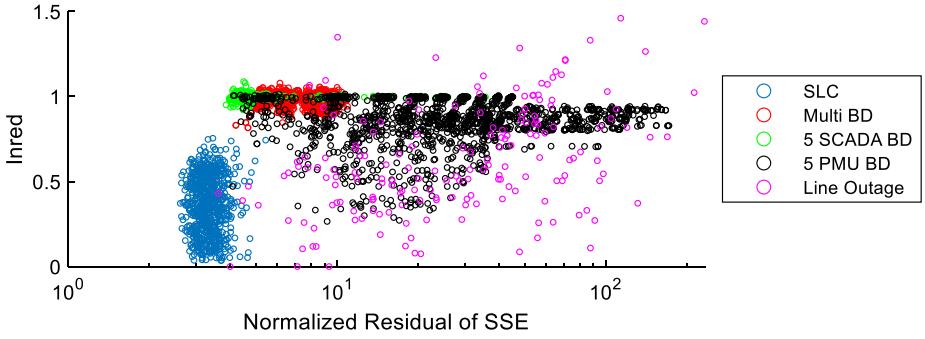

Fig. 9. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases for 10 times the measurement noise level at the instant when the anomaly happens.

Table 6
Detection accuracy of the proposed method with different measurement noise levels.

	Normal	SLC	Bad Data	Line Outage
3 Times	99.5%	88.98% (10.84%)	99.42% (0.54%)	99.46%
10 Times	98%	73.74% (24.98%)	98.34% (1.59%)	94.62%

times the original noise levels, respectively. In general, the proposed method is still reliable for different measurement noise levels as the misdetection rate is still low.

4.4. Sensitivity to different degrees of PMU observability

For the simulations in Section 4.2, the network is assumed to be observable by PMUs. However, many power systems still do not have PMUs, or only parts of the network are observable by PMUs. This subsection investigates the effect of different degrees of PMU observability on the proposed anomaly detection algorithm. For the case where no PMU is placed, the measurement redundancy is 2, which is provided by the SCADA data. For the partial PMU observability case, 18 PMUs are placed at Buses 12, 17, 37, 49, 54, 56, 59, 69, 70, 75, 77, 80, 85, 92, 94, 96, 100, and 105. Table 7 shows the testing results. The measurement redundancy is 2.66, where PMU observes 66% of the buses in the network. For the cases where all buses are equipped with PMUs, the measurement redundancy is 5, where PMUs provide 3 independent sets of measurements that can make the network fully observable.

Table 7
Detection accuracy of the proposed method with different degrees of PMU observability.

	Normal	SLC	Bad Data	Line Outage
118 PMUs	96.5%	99.91% (0.09%)	99.92% (0.08%)	100%
18 PMUs	96.5%	97.61% (2.30%)	99.34% (0.42%)	100%
0 PMU	96%	76.22% (23.42%)	99.46% (0.40%)	100%

The I-R plots in Figs. 10–12 show that as the number of PMUs increases, max($\mathbf{r}_{sse}$) for bad data and line outage cases as well as the *inred* approximation effect increase, and vice versa. This is because the overall measurement accuracy is proportional to the number of PMUs placed in the network. It is notable that when the number of PMUs decreases, the occurrence of outliers (*inred* $>$ 1.2) of SCADA bad data increase; while with a greater number of PMUs, the occurrence of outliers of PMU bad data increases. The former result might be due to the weakened *inred* approximation effect, while the latter could be explained by the stronger correlation between the PMU measurements.

Table 7 shows that the degree of PMU observability has a similar effect on detection accuracy as compared to the change of measurement noise levels, as expected, i.e., the non-detection and misdetection rates of the SLC and bad data cases increase as the number of PMUs decreases/measurement noise level increases. The difference is that the line outage case detection accuracy is 100%, and the detection rate of normal cases is almost unchanged regardless of the number of PMUs; while both are affected by the measurement noise levels. This is because although the number of PMU changes the overall measurement accuracy, it does not change the noise level of each measurement error. In general, the results show that the proposed method is also suitable for SCADA data-based state estimators as it has a very low misdetection rate, but might not be able to detect small SLCs.

4.5. Sensitivity to different networks

Simulation results on other systems are also included to validate the generalizability of the proposed method, including IEEE 14, 39, and 57-bus test systems. The conventional measurements are placed according to Table 3. The PMUs can make the networks barely observable, and the placements are given in Table 8.

The I-R plots of the three networks shown in Figs. 13–15 demonstrate that *inred* and max($\mathbf{r}_{sse}$) are the same in different networks. The reason why in smaller networks the *inred* values are smaller for the SLC case is that fewer simulations are conducted and *inred* has a higher possibility to get values approximated to 0. It is shown in Table 9 that the proposed

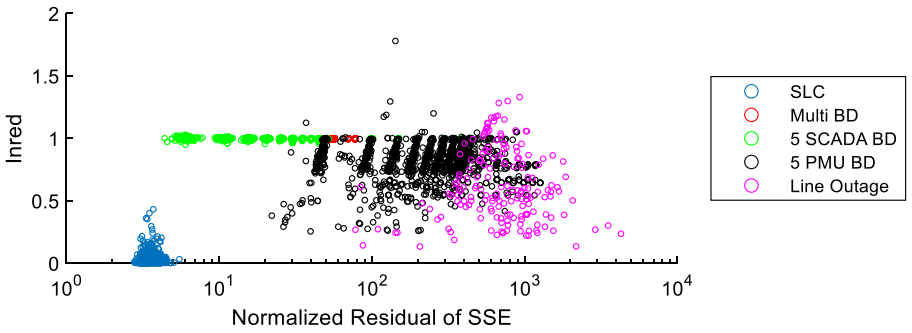

Fig. 10. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases with 118 PMUs at the instant when the anomaly happens.

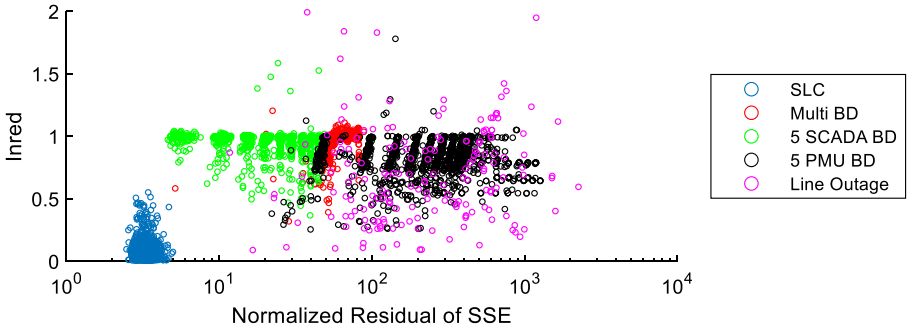

Fig. 11. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases with 18 PMUs at the instant when the anomaly happens.

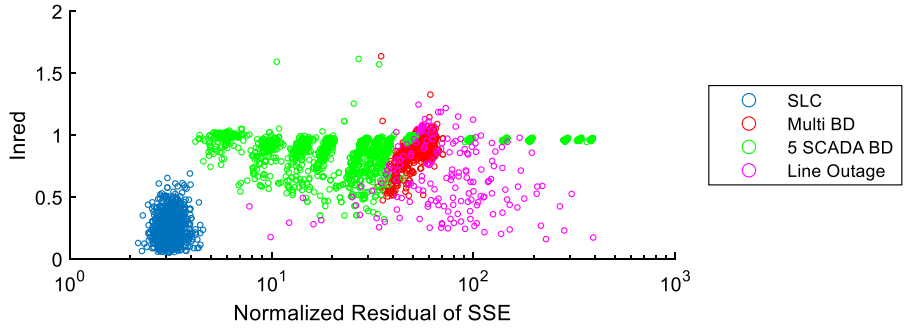

Fig. 12. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases with no PMUs at the instant when the anomaly happens.

Table 8
PMU placement for IEEE 14, 39 & 57 bus networks.

IEEE14	2;6;7;9
IEEE 39	2;6;9;12;14;17;22;23;29;32;33;34;37
IEEE 57	26;12;15;19;22;25;27;32;36;39;41;44;47;50;52;55

method has similar detection accuracy in different networks using the same set of thresholds for the IEEE 118-bus test system. The detection accuracies in the normal, bad data and line outage cases are equal or close to 100% in all three networks, while in the SLC case it is slightly affected. The reason for this is that different networks have different proportions of small loads which is difficult to detect.

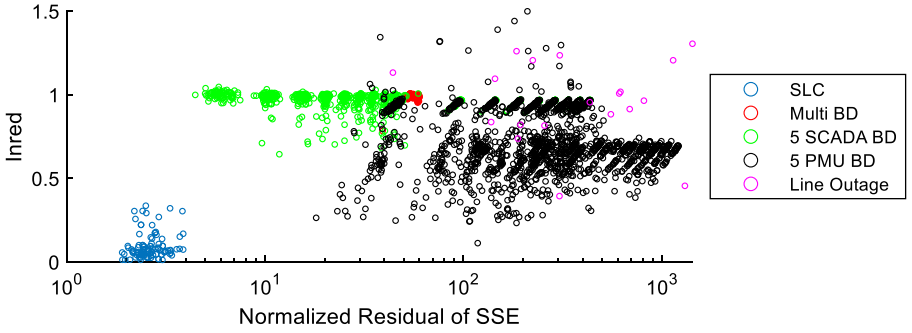

Fig. 13. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases at the instant when the anomaly happens in IEEE 14 bus test system.

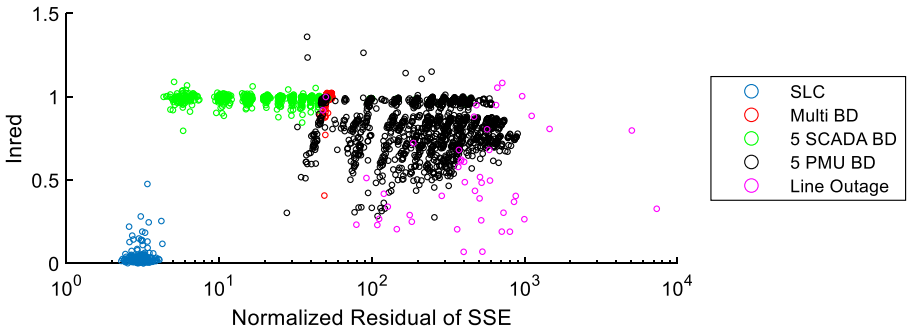

Fig. 14. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases at the instant when the anomaly happens in IEEE 39 bus test system.

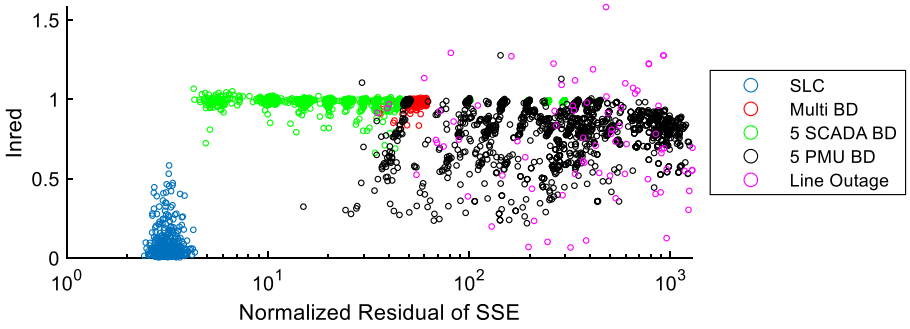

Fig. 15. *Inred* versus max($\mathbf{r}_{sse}$) of IEKF in the anomaly cases at the instant when the anomaly happens in IEEE 57 bus test system.

Table 9
Detection accuracy of the proposed method in IEEE 14, 39 & 57 bus networks.

	Normal	SLC	Bad Data	Line Outage
IEEE 14	98.5%	99.39% (0.61%)	98.00% (2.00%)	100%
IEEE 39	100%	99.04% (0.96%)	99.31% (0.62%)	100%
IEEE 57	97%	95.89% (3.90%)	98.85% (01.12%)	100%

5. Conclusions and future works

This paper proposes a new anomaly detection method that can accurately detect and distinguish three types of anomalies, including sudden load changes, bad data, and line outage. The detection criteria are derived from the innovation reduction properties of the IEKF (*inred*) and the property of the normalized residuals for static SE. Extensive simulation results on different IEEE test systems show that:

- Fixed threshold for the maximum normalized residual of the static SE (max($\mathbf{r}_{sse}$)) cannot reliably distinguish SLC and bad data. This problem is overcome by using the properties of *inred* and max($\mathbf{r}_{sse}$) as proposed in the paper.
- The erroneous detection probability is almost 0 for our proposed method under normal operation, bad data, SLC cases and line outage cases. In comparison, two existing methods have over 30% wrong detections and they could not detect the line outages.
- Given a proper number of SCADA and PMU measurements, line outages can be detected using only the measurements at the time the estimation is performed.
- The proposed method is robust to different levels of noise, different degrees of measurement redundancies and different sizes of the networks. It is found that higher measurement noise levels and low degrees of PMU observability could increase the rate of non-detections but can hardly increase wrong detections.

Our future works will be on testing the proposed method using real systems and data. The developments would include an automated training process that will set the values of the thresholds automatically and keep updating the values as the system condition changes.

CRediT authorship contribution statement

Zhaoyang Jin: Conceptualization, Methodology, Data curation, Investigation, Software, Writing – original draft. **Junbo Zhao:** Methodology, Validation, Visualization, Writing – review & editing. **Lei Ding:** Supervision, Funding acquisition, Project administration, Writing – review & editing. **Saikat Chakrabarti:** Validation, Visualization, Writing – review & editing. **Elena Gryazina:** Validation, Writing – review & editing. **Vladimir Terzija:** Supervision, Funding acquisition, Resources, Writing – review & editing.

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.

Acknowledgement

This work was carried out as a part of the AMPaC Megagrant project supported by The Ministry of Education and Science of Russian Federation, Grant Agreement No 075-10-2021-067, Grant identification code 000000S707521QJX0002.

References

[1] Schweppe FC, Wildes J. Power system static-state estimation, Part I-Part III. IEEE Trans Power Apparatus Syst 1970;PAS-89(1):120–5.
[2] Brandalik Robert, Wellssow Wolfram-H. Power system state estimation with extended power formulations. Int J Electr Power Energy Syst 2020;115. https://doi.org/10.1016/j.ijepes.2019.105443.
[3] Cano Jose M, Arboleya Pablo, Ahmed Mahmoud Rashad, Mojumdar Md Rejwanur R, Orcajo Gonzalo A. Improving distribution system state estimation with synthetic measurements. Int J Electr Power Energy Syst 2021;129. https://doi.org/10.1016/j.ijepes.2020.106751. ISSN 0142-0615.

[4] Ćetenović Dragan N, Ranković Aleksandar M. Optimal parameterization of Kalman filter based three-phase dynamic state estimator for active distribution networks. Int J Electr Power Energy Syst 2018;101:472–81. https://doi.org/10.1016/j.ijepes.2018.04.008. ISSN 0142-0615.
[5] Ji Xingquan, Yin Ziyang, Zhang Yumin, Wang Mingqiang, Zhang Xiao, Zhang Chao, Wang Dong. Real-time robust forecasting-aided state estimation of power system based on data-driven models. Int J Electr Power Energy Syst 2021;125. https://doi.org/10.1016/j.ijepes.2020.106412.
[6] Do Coutto Filho MB, de Souza JCS. Forecasting-aided state estimation-Part I: Panorama. IEEE Trans Power Syst 2009;24(4):1667–77.
[7] Zhao J, et al. Power system dynamic state estimation: motivations, definitions, methodologies, and future work. IEEE Trans Power Syst 2019;34(4):3188–98.
[8] Rousseaux P, Van Cutsem Th, Dy Liacco TE. Whither dynamic state estimation. Int J Elect Power Energy Syst 1990;12(2):104–16.
[9] Shih K, Huang S. Application of a robust algorithm for dynamic state estimation of a power system. IEEE Trans Power Syst 2002;17(1):141–7.
[10] Fernandes Samir Walker, da Rosa Mauro Augusto, Issicaba Diego. A robust dynamic line rating monitoring system through state estimation and bad data analysis. Electr Power Syst Res 2020;189. https://doi.org/10.1016/j.epsr.2020.106648. ISSN 0378-7796.
[11] Zhao J, et al. Robust forecasting aided power system state estimation considering state correlations. IEEE Trans Smart Grid 2018;9(4):2658–66.
[12] Wang Yi, Sun Yonghui, Dinavahi Venkata. Robust forecasting-aided state estimation for power system against uncertainties. IEEE Trans Power Syst 2020;35(1):691–702.
[13] Jin Z, et al. A hybrid robust forecasting-aided state estimator considering bimodal Gaussian mixture measurement errors. Int J Electr Power Energy Syst 2020;120(105962).
[14] Nishiya Ken-Ichi, Takagi Hiroshi, Hasegawa Jun, Koike Toichiro. Dynamic state estimation including detection of innovation process for electric power systems. Elect Eng Jpn 1978;98(1):52–61.
[15] Valverde G, Terzija V. Unscented Kalman filter for power system dynamic state estimation. IET Gener Transmiss Distrib 2011;5(1):29–37.
[16] da Silva AML, Do Coutto Filho MB, Cantera JMC. An efficient dynamic state estimation algorithm including bad data processing. IEEE Trans Power Syst 1987;2(4):1050–8.
[17] Geetha SJ, Chakrabarti S, Rajawat K, Terzija V. An asynchronous decentralized forecasting-aided state estimator for power systems. IEEE Trans Power Syst 2019;34(4):3059–68.
[18] Merrill HM, Schweppe FC. Bad data suppression in power system static state estimation. IEEE Trans Power Syst 1971;Pas-90(6):2718–25.
[19] Souza JCS, Leite da Silva AM, Alves da Silva AP. Online topology determination and bad data suppression in power system operation using artificial neural networks. IEEE Trans Power Syst 1998;13(3):796–803.
[20] Lourenço EM, Simões Costa AJA, Clements KA. A topology error identification method directly based on collinearity tests. IEEE Trans Power Syst, Nov 2006;21(4):1920–9.
[21] Leite da Silva AM, Do Coutto Filho MB, de Queiroz JF. State forecasting in electric power systems. IEE Proc Gen Trans Distrib 1983;130(5):237. https://doi.org/10.1049/ip-c.1983.0046.
[22] Zhao JB, Netto M, Mili L. A robust iterated extended Kalman filter for power system dynamic state estimation. IEEE Trans Power Syst 2017;32(4):3205–16.