Doubly-fed Deep Learning Method for Bad Data Identification in Linear State Estimation

Yingzhong Gu, Senior Member, IEEE, Zhe Yu, Senior Member, IEEE, Ruisheng Diao, Senior Member, IEEE, and Di Shi, Senior Member, IEEE

Abstract—With more data-driven applications introduced in wide-area monitoring systems (WAMS), data quality of phasor measurement units (PMUs) becomes one of the fundamental requirements for ensuring reliable WAMS applications. This paper proposes a doubly-fed deep learning method for bad data identification in linear state estimation, which can: ① identify bad data under both steady states and contingencies; ② achieve higher accuracy than conventional pre-filtering approaches; ③ reduce iteration burden for linear state estimation; ④ efficiently identify bad data in a parallelizable scheme. The proposed method consists of four key steps: ① preprocessing filter; ② online training of short-term deep neural network; ③ offline training of long-term deep neural network; ④ a decision merger. Through delicate design and comprehensive training, the proposed method can effectively differentiate the bad data from event data without relying on real-time topology information. An IEEE 39-bus system simulated by DSATools TSAT and a provincial electric power system with real PMU data collected are used to verify the proposed method. Multiple test scenarios are applied, which include steady states, three-phase-to-ground faults with (un)successful auto-reclosing, low-frequency oscillation, and low-frequency oscillation with simultaneous three-phase-to-ground faults. The proposed method demonstrates satisfactory performance during both the training session and the testing session.

Index Terms—Bad data identification, linear state estimation, preprocessing, deep neural network, wide-area monitoring system (WAMS).

I. INTRODUCTION

NOWADAYS, as artificial intelligence (AI) and big data technologies keep advancing and evolving, more and more data-driven applications are introduced in wide-area monitoring systems (WAMS) [1]. Operation data with massive volume and velocity keep flowing into WAMS to support applications, e.g., 25-60 samples per second, which requires extensive data-parallel processing capability [2]. Among those requirements, data quality is one of the fundamental demands affecting the quality of decision making and accuracy of system monitoring [3]. In traditional WAMS, linear state estimation (LSE) can be used, by leveraging the linear relationship between the system states (i.e., voltage phasors) and the measurements (i.e., voltage and current phasors) to efficiently calculate the system states and perform bad data identifications [4]. The majority of commercial LSE applications are based on weighted least square (WLS) method, which can only identify and remove one bad measurement per iteration. Otherwise, legit measurements can be treated as bad ones and therefore get removed falsely [5], [6]. This limitation forms a significant bottleneck affecting the self-healing capability of data quality in modern WAMS [7]. To fill this technology gap, this paper presents an innovative method by leveraging deep neural networks to process phasor measurement unit (PMU) data for LSE used in a control center.

Many recent research works are reported for bad data identification in LSE, which can be generally divided into two categories: state estimation approaches and data-driven approaches [8]. According to the order of data processing module in the decision-making flow, they can also be classified into the pre-estimation and the post-estimation filtering [9]. Among various state estimation techniques, WLS approach, weighted least absolute value (WLAV) and Bayesian estimation are the popular ones adopted and implemented in WAMS [10].

WLS solves an optimization problem to find the estimated states with the least weighted square error ($L_2$ norm) of the measurement residuals [11]. Then, bad data can be identified based on the residual between actual measurements and estimated measurements [12]. As an extension of WLS, [11] proposed a distributed WLS state estimation method that integrated weight update from multi-area measurements to enhance the capability and accuracy of handling bad data detection. Reference [12] proposed a substation-level bad data detection algorithm by parsing online state of circuit breakers and disconnectors, which can detect bad data from failing current transformers. Reference [6] presented an alternative WLS largest normalized residual (LNR) algorithm that classifies the suspicious measurements into non-interactive groups and then performs simultaneous identification of multiple bad data.

As an alternative to WLS, WLAV solves an optimization problem to find the estimated states that minimize the
weighted absolute value ($L_1$ norm) of the measurement residuals [13]. Extended from WLS and WLAV, the following valuable research works have been conducted. Reference [14] investigated the sparse recovery models for bad data detection and state estimation in power networks where the $L_1$-relaxation model, the multi-stage convex relaxation model and the WLAV were considered. Reference [13] presented a least absolute value based linear state estimator that is computationally efficient and statistically robust.

Different from WLS and WLAV, Bayesian estimation is designed to find the value of states with the highest likelihood [15]. Reference [15] proposed a Bayesian state estimation for unobservable distribution systems via deep learning, which demonstrates the robustness against modeling and estimation errors and the presence of bad and missing data. Reference [16] developed a Bayesian-based harmonic state estimation in distribution systems to calculate states and identify bad data.

Kalman filter is another popular state estimation approach, especially for power system dynamic state estimation [17], [18]. Reference [17] developed an extended Kalman filter (EKF) technique for dynamic state estimation of a synchronous machine using PMU data. Reference [18] developed a robust iterated EKF based approach for estimating power system state dynamics subject to disturbances.

In general, the state estimation based approach can efficiently calculate system states. With accurate and timely updated topology information, e.g., contingencies, breaker status, the identification accuracy of bad data can be quite high. However, the state estimation approaches also suffer from drawbacks in handling bad data [19]. First, these approaches cannot process bad data in critical measurements, critical measurement pair or homogeneously critical measurement groups due to lack of redundancy [20], [21]. Second, the computation time of bad data identification grows linearly to the number of bad data presented and grows exponentially to the system size, or more specifically, the number of measurements and the number of system states [6]. The bad data processing capability for large systems is restrained. Third, the quality of decision making of those approaches relies heavily on accurate and timely updated system topology. Inaccurate or delayed topology information can cause unreliable results in state estimation and bad data identification [19]. Therefore, state estimation based approach alone is not sufficient for online bad data identification in modern WAMS.

The other category of approach is the data-driven approach. Reference [9] proposed a Kalman filter based pre-estimation approach for bad data identification that can detect abrupt changes among consecutive measurements. Reference [22] proposed an online data-driven algorithm to identify low-quality synchrophasor measurements through a density-based local outlier factor (LOF) analysis. Reference [23] proposed a data-driven PMU bad data detection algorithm based on spectral clustering using single PMU data. Reference [24] proposed a feature-based method originating from simple logical method based on observed patterns. Reference [25] presented a wavelet transformation based approach aiming to alleviate the efforts of system operators. Reference [26] proposed a matrix recovering technique that can be utilized to identify and recover bad data via recognizing the low-rank feature of synchrophasors from adjacent channels. Reference [27] presented a low-pass filter for removing spikes in the measurements. Model-based approaches for bad data identification were also proposed that can improve the accuracy of bad data identification [20], [28], [29]. The data-driven approaches generally have a more efficient performance on processing massive volume of data and are usually positioned as a preprocessing filter for an LSE or other WAMS applications. However, the accuracy of the bad data identification via such approaches usually compromises when events or contingencies happen. It is difficult for such approaches to differentiate the bad data from event data accurately.

This paper proposes a double-fed deep learning method for bad data identification that can accurately label the bad data under both normal operation conditions and various types of events. The proposed method has improved robustness against different kinds of system topological changes. The proposed method can handle critical measurements as long as properly labelled data samples are available. These labelled data samples can be generated by leveraging an offline three-phase LSE, dynamic state estimation on historical PMU or supervisory control and data acquisition (SCADA) data and numerical simulations.

The key contributions of this paper are threefold: ① a double-fed deep learning method for bad data processing is proposed to efficiently label bad data for large systems in parallelizable scheme; ② the proposed method can accurately identify bad data under normal operation conditions and various types of events and contingencies; ③ the proposed method has improved robustness against different kinds of system topological changes.

The remainder of the paper is organized as follows. Section II briefly discusses the formulations of LSE. Section III introduces the proposed double-fed deep learning method for bad data preprocessing. Section IV presents the numerical experiments of the proposed method under different scenarios using practical system PMU data. Section V provides the concluding remarks and discusses potential future works.

II. LSE

Since the bad data processing approach is designed for LSE, this section briefly presents the concept of LSE. LSE calculates system states based on phasor measurements by leveraging the linear relationship between the voltage and current phasors [4]. The advantages of LSE includes: ① non-iterative solution for state estimation; ② efficient computation performance compared with conventional state estimation; ③ convergence guarantee compared with Newton method based state estimation [30]. Popular approaches for state estimation include WLS [7], [10], [26], WLAV [13], [14], Bayesian estimation [15], [16] and Kalman filter estimation [17], [18]. In industry applications, WLS is most widely adopted. Therefore, this section takes the WLS as an example to illustrate the LSE.

Assuming a system with $n$ observable nodes, $m_v$ voltage
measurements and $m_t$ branch current measurements, the total number of measurements can be denoted as $m$, which is the summation of $m_v$ and $m_t$. The system state vector $\hat{x} \in \mathbb{C}^{m_{x-1}}$ includes the voltage phasors of all the observable nodes. The measurement vector $\hat{z} \in \mathbb{C}^{m_{x-1}}$ includes the voltage and current phasor measurements of the terminals where PMUs are installed. The measurement model of PMU data can be derived from Ohm’s law as

$$\begin{bmatrix} \hat{V} \\ \hat{i} \end{bmatrix} = \begin{bmatrix} A \\ \hat{Y} \end{bmatrix} \hat{x} = \hat{H} \hat{x}$$

(1)

where $A \in \mathbb{R}^{m \times n}$ is the identity matrix which describes the one-to-one mapping relationship between the state vector $\hat{x}$ and voltage phasor measurement vector $\hat{V}$ for those buses where PMU measurement is available; and $\hat{Y} \in \mathbb{C}^{m_{x-1} \times n}$ is the branch admittance matrix which is derived from Kirchhoff’s voltage law (KVL) to describe the relationship between the branch current phasor $\hat{I}$ and the two voltage phasors at the two terminals. If the voltage phasor of node $j$ is the $r^{th}$ component in the measurement vector of voltage phasors, $A_{jr} = 1$; otherwise $A_{jr} = 0$, where $A_{jr}$ is the element of $A$ on the $i^{th}$ row and $j^{th}$ column. By combining the voltage and current measurements into one formulation, the measurement model of PMU data can be denoted by the complex matrix $\hat{H}$ in (2).

$$\begin{bmatrix} \hat{V} \\ \hat{i} \end{bmatrix} = \begin{bmatrix} A \\ \hat{Y} \end{bmatrix} \hat{x} = \hat{H} \hat{x}$$

(2)

The model in (2) is not numerically linear yet, because its components $\hat{x}$ and $\hat{z}$ are complex numbers. It needs to be further decomposed into the rectangular form in (3). The corresponding measurement model becomes (4).

$$\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} \text{Re}(\hat{x}) \\ \text{Im}(\hat{x}) \\ \text{Re}(\hat{z}) \\ \text{Im}(\hat{z}) \end{bmatrix} \begin{bmatrix} H_r \ H_j \\ -H_j \ H_r \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = H \begin{bmatrix} x \\ z \end{bmatrix}$$

(3)

where $H$ is the matrix which represents the measurement model for LSE in rectangular form; and $H_r$ and $H_j$ are the real and imaginary parts of the $H$, respectively [31].

Based on the formulation in (4), the solution of $x$, namely the vector of estimated system states, can be obtained by (5).

$$\hat{x} = (H^T W^{-1} H)^{-1} H^T W^{-1} z$$

(5)

where $W \in \mathbb{R}^{n \times n}$ is a diagonal matrix, whose diagonal components are the weights of the measurements.

The LSE form (3) - (5) is the positive-sequence LSE. Although this is widely used in practical power industry due to its simplicity and efficiency, it is more useful for large-scale system during steady state. For the purpose of bad data identification, an alternative form of the three-phase representation is preferred because it can correlate and crosscheck all three phases [32].

By extending the positive-sequence voltage and current phasors to three-phase voltage and current forms, (1) can be denoted as

$$\begin{bmatrix} \hat{V}_a \\ \hat{V}_b \\ \hat{V}_c \end{bmatrix} = \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix}$$

$$\begin{bmatrix} \hat{i}_a \\ \hat{i}_b \\ \hat{i}_c \end{bmatrix} = \begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix}$$

(6)

(7)

where $\hat{V}_a$, $\hat{V}_b$, $\hat{V}_c$ are the three-phase voltage phasors; $\hat{i}_a$, $\hat{i}_b$, $\hat{i}_c$ are the three-phase current phasors; $A_{abc} \in \mathbb{R}^{3 (m_{x-1} \times 3)\times 3}$ is the augmented identity matrix; $\hat{Y}_{abc} \in \mathbb{C}^{3 (m_{x-1} \times 3)}$ is the augmented branch admittance matrix; and $\hat{x}_{abc}$ is the three-phase system state vector.

To further illustrate (7), one of the branches can be picked and (7) can be represented as

$$\begin{bmatrix} \hat{i}_{abc} \\ \hat{i}_{avg} \end{bmatrix} = \begin{bmatrix} \hat{Y}_{abc} \hat{V}_{abc} \\ \hat{Y}_{avg} \hat{V}_{avg} \end{bmatrix}$$

(8)

where $\hat{i}_{abc}$ is the from-end current phasor of phase $d$, $d \in \{a, b, c\}$; $\hat{i}_{avg}$ is the to-end current phasor of phase $d$; $\hat{y}_{abc}$ is the branch admittance parameters; $\hat{y}_{avg}$ is the front-end branch to ground susceptance; $\hat{y}_{avg}$ is the to-end branch to ground susceptance; $\hat{V}_{abc}$ is the front-end voltage phasor at phase $d$; and $\hat{V}_{avg}$ is the to-end voltage phasor at phase $d$.

Equation (2) can be extended to

$$\dot{z}_{abc} = \begin{bmatrix} \dot{V}_{abc} \\ \dot{I}_{abc} \end{bmatrix} = \begin{bmatrix} A_{abc} \\ \hat{Y}_{abc} \end{bmatrix} \dot{x}_{abc} = H_{abc} \dot{x}_{abc}$$

(9)

where $\dot{z}_{abc}$ is the phasor vector of three-phase measurements; and $H_{abc}$ is the measurement matrix for the three-phase LSE.

Equations (6) and (7) can be converted to the rectangular form (10), and (9) can be written as (11).

$$\begin{bmatrix} x_{abc} \\ \dot{x}_{abc} \end{bmatrix} = \begin{bmatrix} \text{Re}(\hat{x}_{abc}) \\ \text{Im}(\hat{x}_{abc}) \end{bmatrix}$$

(10)

$$\begin{bmatrix} x_{abc} \\ \dot{x}_{abc} \end{bmatrix} = \begin{bmatrix} \text{Re}(\hat{z}_{abc}) \\ \text{Im}(\hat{z}_{abc}) \end{bmatrix}$$

(11)

where $x_{abc}$ is the scalar vector including both real and imaginary parts of the state variables; $z_{abc}$ is the scalar vector including both real and imaginary parts of the measurement variables; and $H_{r,abc}$ and $H_{i,abc}$ are real and imaginary parts of $H_{abc}$, respectively.

Based on the formulation in (11), the solution of $x_{abc}$ can be obtained by

$$\dot{x}_{abc} = (H_{abc}^{T} W_{abc}^{1} H_{abc})^{-1} H_{abc}^{T} W_{abc}^{1} \dot{z}_{abc}$$

(12)

where $W_{abc} \in \mathbb{R}^{(3m) \times (3m)}$ is a diagonal matrix, whose diagonal components are the weights of the measurements at three phases.

The three-phase gain matrix can be defined as

$$G_{abc} = H_{abc}^{T} W_{abc}^{-1} H_{abc}$$

(13)

The three-phase hat matrix can be defined as

$$K_{abc} = H_{abc} G_{abc}^{-1} H_{abc}^{T} W_{abc}$$

(14)
Then, the three-phase sensitivity matrix can be calculated as (15), where $I$ is the identity matrix.

$$ S_{abc} = I - K_{abc} \quad (15) $$

The residuals of the measurement can be calculated by the difference between the actual and the estimated measurements in (16).

$$ r_{abc} = z_{abc} - \hat{x}_{abc} H_{abc} \quad (16) $$

The covariance matrix of the measurement residual can be calculated as

$$ \Omega = S_{abc} W_{abc}^{-1} \quad (17) $$

The normalized residual for each measurement can be calculated as

$$ r_{N,i} = \frac{|r_i|}{\sqrt{\Omega_i}} \quad i \in [1, m] \quad (18) $$

where $r_i$ is the residual of each measurement; and $\Omega_i$ is the diagonal element of the covariance matrix $\Omega$.

In this paper, the measurement value whose normalized residual is above 3.0 is considered as bad data or abnormal data. In that case, the probability of that measurement is estimated to be less than 0.3% given that the measurement error follows a normal distribution, which is a widely adopted standard in the industry [30]. The three-phase LSE is more effective in identifying bad data because bad data occurrence violating three-phase balancing can be identified [32]. However, due to the problem dimension being at least nine times larger than the positive-sequence state estimation, its online applicability and feasibility in practical systems, especially those large regional systems, are compromised [33].

In this paper, we use the three-phase LSE to generate bad data labels for practical PMU data, which can be employed to train the AI agent using the proposed method. Then online bad data identifications can be performed by a well-trained agent in a distributed manner for large-scale systems in real time.

III. DOUBLY-FED DEEP LEARNING METHOD FOR BAD DATA IDENTIFICATION

The proposed doubly-fed deep learning method for bad data identification consists of two feedback-pipelines: short-term deep neural network and long-term deep neural network. The short-term deep neural network is designed to learn and capture short-term data patterns under variable situations. Online training is performed for the short-term deep neural network to keep updated to the current system operation point. The long-term deep neural network is designed to learn and capture historical patterns and theoretically presumptive patterns such as contingencies or various events generated through transient analysis simulations.

Figure 1 shows the architecture of the proposed method. The green arrow bar represents the incoming PMU data stream. Two moving windows take consecutive data frames and feed them into the corresponding pre-filter. For short-term deep neural network pipeline, the pre-filter S processes the data frames from moving window S and feeds them into the deep neural network S. For long-term deep neural network pipeline, the pre-filter L processes the data frames from the moving window L and feeds them into the deep neural network L. The moving window L has a length twice longer than the moving window S because it needs more data to diagnose whether the current pattern is similar to any of the historical patterns. Both deep neural networks S and L generate the probability of the current data samples being bad data for every channel. The decision functions L and S convert the probability into the corresponding decision scores.

The decision merger generates a decision label for whether the current data sample is a bad data for each channel based on the decision scores from both the short-term and long-term deep neural network pipelines. The decision labels are passed to the linear state estimator, which can remove the bad data before the first-round calculations. The estimated states and measurements are written into the LSE stream, which feeds back to the pre-filters L and S to remove the bad data that may affect the scores for other data samples. The LSE stream also feeds them into the deep neural network S to perform periodically online training.

The data structure of PMU measurements feeding into the moving windows S and L is illustrated in Fig. 2, where $V_o, V_{oa}, V_{ob}, V_{oc}, \theta_{oa}, \theta_{ob}, \theta_{oc}$ are the magnitudes and phase angles of the three-phase voltage phasors, respectively; $I_a, I_b, I_c, \theta_{ia}, \theta_{ib}, \theta_{ic}$ are the magnitudes and phase angles of the three-phase current phasors, respectively; and $f$ is the frequency. PMUs are grouped into clusters for the doubly-fed deep neural networks. There are two principles for the PMU clustering.

1) Correlation principle: the grouped PMU measurements should have a strong correlation to each other governed by the physical laws of nature.
2) Independence principle: the errors and noises of the grouped PMU measurements should have independent probability distributions.

Good examples of clustering include independent PMU measurements at the same substation but for different transmission lines, independent PMU measurements at two terminals of a branch, independent measurements of parallel circuits, and independent measurements at different windings of a transformer. Bad examples of clustering can be irrelevant measurements of elements with weak correlation, correlated measurements sharing the same potential transformers (PTs), current transformers (CTs), or communication channels.

The number of clustered PMU $k$ is suggested to be between one and three. If only one PMU is used, the deep neural network can only learn the cross-correlation between voltages and currents, between phase angles and magnitudes, and among the three phases. With independent PMU measurement incorporated into the deep neural networks, data can be validated across different circuits, PTs, CTs, and terminals, which significantly improve the accuracy of bad data identification. However, if there are too many PMUs clustered in the same deep neural network, the computation performance may decline, and the neural network may not work properly if one of the PMUs is off-line. After the clusters are formed, all 13 measurement quantities in Fig. 2 are integrated into a vector to be fed into the deep neural network as an input vector. The pre-filter S processes time-series data in different PMU channels independently in a parallel manner, while the deep neural network takes pre-filters’ outputs clustered altogether.

The moving window can be denoted as

$$W = \{w_t \mid t \in \{t_0 + t_q - L, ..., t_0, ..., t_0 + t_q\} \}$$ (19)

where $W$ is the set of the data in the moving window; $t$ is the time stamp; $t_q$ is the target time for LSE; $t_0$ is the lead time steps (suggested range is 60-100 ms); and $L$ is the length of the window. The lead time can be set as zero if system operators intend to calculate the LSE solution immediately once the PMU measurement data is available. However, it has been found through many numerical experiments that a non-zero lead time brings the advantage of identifying temporary bad data spikes that can effectively improve the resulting quality of LSE.

The pre-filter can be applied via

$$Q(p) = \inf \{w_t \in \mathbb{R} : p \leq F(w_t), w_t \in W \}$$ (20)

$$P_i(w_t) = \begin{cases} 2g(2(w_i - Q(p_i))) - 1, \\ e^{w_t}/(1 + e^{w_t}) \end{cases}$$ (21)

where $F(\cdot)$ is the cumulative probability distribution function of the data set $W$; $p$ is the threshold probability; $p_i$ is the default threshold probability; $\inf(\cdot)$ is the infimum of the set or the greatest lower bound of the set; $Q(\cdot)$ is the function which calculates the modified current inputs; $g(\cdot)$ is the logistic sigmoid function; and $P_i(\cdot)$ is the function which gives the inputs for the deep neural network. This proposed moving window can leverage the temporal correlation of the data sets to capture temporarily false spikes and mitigate negative impacts of extreme data points on other non-extreme data points.

Based on the numerical experiments, the pre-filter process described in (20) and (21) is essential for the successful functioning of the proposed method. It mitigates the dependency of deep neural network training results on the power system operation conditions and develops a heterogeneous sensitivity toward different levels of measurement errors. The output of $P_i(\cdot)$ is normalized as the interval of $(-1, 1)$.

The outputs of the pre-filter feeding into the deep neural network are depicted in Fig. 3. Assuming that $k$ PMU measurements are grouped in the cluster, the dimension of the inputs to this deep neural network is $13k$, which includes the three-phase voltage magnitudes and angles, three-phase current magnitudes and angles, and the frequency. The number of hidden layers is suggested to be 4, the activation function of the first layer is tanh, the activation function of the rest hidden layers is $relu$, the activation function of the output layer is $softmax$, the number of units for each hidden layer is $10k$, and the number of units for the output layer is $13k$.

For each hidden layer, there is a neuron with no input which represents the bias term for that layer. The output of the deep neural network is one-to-one mapping to the input of the network, which indicates the probability of each input being bad data.

![Deep neural network for bad data identification](image)

Fig. 3. Deep neural network for bad data identification.

$$L = -1/N \sum_{i=1}^{N} \left[ y_i \ln(p(y_i)) + (1-y_i)\ln(1-p(y_i)) \right]$$ (22)

where $y_i$ is the $i$th output of the deep neural network.

The binary cross-entropy function (22) is used as the loss function. Based on the numerical experiments, the cross-entropy function has much better performance in both training and testing phases compared with other popular loss functions such as $L_2$ loss, because the output decision is discrete and the convexity of the binary cross-entropy function makes it easier to find the global optimum [34]. The Adam optimizer is used during the training process, where the learning rate is set to be 0.01.

In the designed architecture presented in Fig. 1, two deep neural networks are used. The short-term neural network generates a probability output using PMU data from the moving window $S$. This neural network is re-trained online every 1 min using the last 2-minute PMU data with LSE labels.

The probability decision function is provided by
\[ D(y_i) = \frac{e^{\beta y_i} - 0.5}{1 + e^{\beta y_i} - 0.5} \] (23)

where \( \beta \) is the squashing coefficient which is suggested to be 10. The decision function makes the probability converge more quickly to 1 (or 0) when the result goes higher (or lower) than the threshold of 0.5, which is to seek a reasonable balance between exploration and exploitation.

The decision merger combines the two probability decisions \( D(y_i) \) and \( D(y_j) \) from the short-term and long-term deep neural network pipelines, as described by

\[ M(y_i) = (D(y_i)D(y_j))^{1/\beta} \] (24)

Different weights can be assigned to demonstrate certain preferences in decision making toward a long-term perspective or short-term perspective. Since LSE has its own bad data identification, it is preferred to be conservative for the neural network to avoid false alarms because any bad data identified by the neural network will be removed directly. For those unidentified bad data, the LSE will process them. Although the proposed method can process most of bad data, some remaining ones (due to small deviations or scenarios not included in the training set) still need to be identified in LSE and are subject to the limitations of LSE.

IV. NUMERICAL EXPERIMENTS

In this paper, the proposed method has been applied to the IEEE 39-bus system using TSAT simulation and one of the largest provincial power systems in China—the Jiangsu power system using the practical PMU data.

A. IEEE 39-bus System

The IEEE 39-bus system is a 10-machine New England power system. In the system, generator 1 represents an aggregation of generators. The detailed system parameters are available at [35]. In the IEEE 39-bus system, generators are in a fourth-order model, while the constant impedance load model is adopted. The system has a single oscillatory mode with natural frequency of \( f_0 = 1.3217 \) Hz. A sinusoidal signal \( A_{\text{ref}} = k \sin(2\pi ft) \) with \( k = 0.6 \) is added to the reference signal of excitation systems. Generator 1 acts as the source of the forced oscillation. The oscillation disturbance is added to the system at \( t = 0 \) s. The low-frequency oscillation event and three-phase-to-ground fault are simulated by DSATools TSAT v18.1.37.

Figure 4 illustrates the current magnitudes of branch 1-2 during a low-frequency oscillation that occurs at generator 1, where the red dot represents the bad data label and the blue bar represents the bad data identification probability of the proposed method (Fig. 5, Fig. 7-11 follow the same legends). Once the blue bar is above 50%, the data sample is labelled as bad data by the proposed method. These samples won’t be taken by LSE as legit measurements in the subsequent computation. The goal of the proposed method is to generate as many blue bars (>50%) to match the red dots at 1.0 as possible and generate as few blue bars whenever the corresponding red dots are at 0 as possible. As shown in Fig. 4, the proposed method successfully avoids labeling oscillation data samples as bad data during the entire oscillation period. Meanwhile, almost all the major bad data at different phases during different periods are successfully labelled, although some identification probabilities are not very large (e.g., phase c at 3 s) and some false alarms are observed (e.g., phase c at 0.6 s). For preprocessing filters, it is acceptable not to label all the bad data because the LSE can process the ones falling through the crack. It is also encouraged not to falsely identify too many good data as bad data because the LSE is unable to correct those false alarms, especially if any of them is critical measurement or in one of the critical pairs.

![Figure 4. Bad data identification probability for three-phase current magnitudes of IEEE 39-bus system during low-frequency oscillation event. (a) Phase a. (b) Phase b. (c) Phase c.](image-url)

Figure 5 illustrates three-phase voltage magnitudes at bus 1 during a low-frequency oscillation at generator 1 plus a temporary three-phase-to-ground fault at branch 2-3 at 0.1 s. After five cycles, the breakers at both terminals of branch 2-3 are opened by the relay, and then the temporary fault is cleared. After 50 cycles, the transmission line is successfully reclosed by the auto-recloser. As shown in Fig. 5, the proposed method successfully avoids labeling the event data as bad data during the oscillation and topology changes. Meanwhile, almost all the major bad data at different phases during different periods are successfully labelled.
The Jiangsu power system has 731 substations and 349 power plants, where 244 of the substations have been equipped with 1138 PMUs. Each PMU measures three-phase voltage phasors, current phasors, and frequency. There are 30365 nodes, 8557 breakers, 22084 disconnectors, 2335 busbars, 2393 transmission lines and 1796 transformers in the node-breaker model of this power system. PMU measurements collected in the system in 2019 are used and have a reporting rate of 25 Hz. Eighty percent of the data sets are used for training for the long-term deep neural network, while the remaining 20% are used for testing. Among the testing data sets, half of the data are used for development phase and the other half are used for testing only. The online training for the short-term deep neural network uses the test- ing data. The true value of practical PMU measurements is unknown. Therefore, an LSE with bad data identification is used to generate the labels for both the training and testing data.

For the proposed method, after training, the average on-line computation time is $19.3995 \mu s$ per sample, which is less than 0.1% of the average computation time of LSE.

To illustrate the effectiveness of the proposed method, a 500 kV double-circuit transmission line is presented in Fig. 6. In the setup of numerical experiment, all the contingencies occur on GL #5322, while LM #5321 is the monitored line for bad data identification and verification. The following scenarios are considered in the numerical experiments.

1) No contingency occurs, and the system is running in steady state with occasional bad data injections.
2) Temporary three-phase-to-ground fault occurs at GL #5322, the relay clears the fault, and the auto-recloser recloses the line successfully.
3) Permanent three-phase-to-ground fault occurs on GL #5322, the relay isolates the fault, the auto-recloser fails to clear the fault, and the relay opens the line permanently.
4) Low-frequency oscillation occurs at the WT plant #3. Figure 7 illustrates the non-contingency scenario.

Fig. 6. Node-breaker model of 500 kV transmission lines LM #5321 and GL #5322.

Fig. 7. Bad data identification probability for three-phase voltage magnitudes during steady state. (a) Phase a. (b) Phase b. (c) Phase c.
It is worth noting that the phase angles measured by PMU are the original phase angles that are not the angle difference to the reference bus. The values of these angles vary across time depending on the deviations of the actual system frequency. Besides, the phase angle is a cyclical variable in the interval of \((-180^\circ, 180^\circ]\). As shown in Fig. 7, the proposed method works pretty well under steady state, and almost all the bad data samples can be labelled correctly.

Figure 8 illustrates the three-phase current magnitudes of LM #5321 when the temporary three-phase-to-ground fault occurs at GL #5322 at 0.1 s. The fault location is 0.1 km from LG plant #2. After five cycles (the base frequency of system is 50 Hz), the breakers at both terminals of GL #5322 are opened by the relay. After that, the temporary fault is cleared, and 50 cycles later, the transmission line is successfully reclosed by the auto-recloser. As is observed in Fig. 8, when the fault occurs, the proposed method successfully avoids labeling the event spikes causing by three-phase-to-ground fault as bad data, which is a common issue faced by many other numerical approaches or topology information-dependent approaches.

In this paper, the fundamental philosophy of the deep neural network that can work here is that every fault in reality has its unique pattern. If the deep neural network can learn these patterns through comprehensive training of historical event data and simulated data, it can distinguish the event data from the bad data. More specifically, for the event that the current magnitude of phase c has a dip of 0.14 p.u. at 2.6 s, it is more likely to be recognized as a bad data instead of faults or other events.

This is because in a practical system, it is very rare to have a fault or an event only causing the change in one phase without any changes in the other two phases. Even a single-phase-to-ground fault affects the magnitudes and phase angles at all three phases. The three-phase patterns can be learned by the deep neural network. In this scenario, the deep neural network does not falsely label any event data as bad data during both the fault and the reclosing periods. Meanwhile, almost all the bad data at different phases during different periods are labelled successfully, although some values are not very large (e.g., phase B at 2.3 s) and some other bad data are not captured due to their small deviations (e.g., phase b at 3.8 s).

Figure 9 illustrates the current phase angles of LM #5321 during the permanent three-phase-to-ground fault that occurs at GL #5322 at 0.1 s.
The fault location is 0.1 km from LG plant #2. After five cycles (the base frequency of system is 50 Hz), the breakers at both terminals of GL #5322 are opened by the relay. The permanent fault fails to be cleared after that. After 50 cycles, the transmission line is reclosed by the auto-recloser, and the fault is activated again. After another 5 cycles, the breaker is triggered by the relay to open the transmission line permanently. As shown in Fig. 9, the proposed method successfully avoids labeling the event data as bad data at the beginning of the fault, at the 1st time of relay operation, at the 1st time of auto-recloser operation, or at the 2nd time of relay operation. Meanwhile, almost all the major bad data at different phases during different periods are labelled successfully, although some values are not very large (e.g., phase a at 4.0 s, phase c at 1.9 s) and some false alarms almost cross the threshold of 0.5 (e.g., phase a at 0.38 s, phase c at 0.2 s). It is also observed that some values have a certain delay in responsiveness (e.g., phase c at 1.1 s).

Figures 10 and 11 present the bad data identification for current magnitudes and current phase angles during a low-frequency oscillation event, respectively. The low-frequency oscillation event occurs at WT plant #3 at 41 s. The location is 82.8 km southeast to LG plant #2. The frequency of the oscillation is 0.66 Hz and the damping factor is 0.99.

As shown in Fig. 10, the deep learning algorithm successfully labels most of the bad data correctly and avoids labeling the event data points caused by the oscillation event as bad data before, after or during the oscillation event.

As shown in Fig. 11, although the impact of the oscillation on current phase angles is not as obvious as the one on current magnitude, most of the bad data are successfully labelled by the proposed method.

Table I illustrates the performance comparison between the proposed method and the benchmark for the training, development and testing data sets. The benchmark is a five-layer (four hidden layers, 53 neurons, the same dimension as the proposed deep neural network) deep neural network without the proposed double-fed framework. Compared with the proposed method, the training accuracy of the benchmark is only 0.8974, and the development and testing accuracies are only 0.7952 and 0.5367, respectively. The precision, recall and $F_1$ scores of the benchmark are only 0.5306, 0.7222, and 0.4194, respectively. As shown Table I, the performance is improved by the proposed method.

Table II presents the comparison of computation efficiency between the proposed method and the method of typical filter + LSE. The LSE is equipped with typical PMU data pre-filters including three-phase balancing check, voltage and current typical range check, topology crosscheck, etc. It
is observed that the pre-filtering time of the proposed method is almost 3 times longer than running these simple pre-filters. In return, the proposed method can help LSE reduce computation time and the number of iterations significantly so that the total computation time is 57.1273% lower. Most of the model enhancements have been discussed in Section III. Based on extensive numerical experiments, it shows that the moving window, PMU clustering, pre-filtering process, decision function, decision merger and the feedback from LSE results to both long-term and short-term deep neural networks are important and need to be carefully configured, which yields an overall well-performing double-fed deep neural network.

### TABLE I

| Method     | Training accuracy | Development accuracy | Testing accuracy | Precision score | Recall score | F1 score |
|------------|------------------|----------------------|-----------------|----------------|-------------|----------|
| Proposed   | 0.9796           | 0.9798               | 0.9577          | 0.9929         | 0.9061      | 0.9475   |
| Benchmark  | 0.8974           | 0.7952               | 0.5367          | 0.5306         | 0.7222      | 0.4194   |

### TABLE II

| Method      | Average LSE iteration | Average pre-filter time (ms) | Average LSE time (ms) | Average total time (ms) |
|-------------|-----------------------|------------------------------|-----------------------|------------------------|
| Proposed    | 3.6144                | 6.3518                       | 25.1415               | 31.4933                |
| Typical filter + LSE | 10.3416            | 2.3251                       | 71.1321               | 73.4572                |

### V. CONCLUSION

As large volumes of PMU data feed into modern control centers, it poses more challenges to the data quality management capability in WAMS applications. Many existing approaches have been developed to handle bad data problems, but it is usually challenging to identify bad data during contingencies, events or topological changes. This paper proposes a doubly-fed deep learning method for bad data identification in LSE. It integrates a long-term deep neural network that learns various historical events/contingencies and a short-term deep neural network that keeps learning the prevailing system operation conditions, and makes reasonably good decisions for labeling bad data. The proposed method is verified using IEEE 39-bus system simulated in TSAT and the real PMU measurements from one of the largest provincial power systems in China. Numerical experiment of normal steady-state and four additional contingencies have illustrated that the proposed method can not only identify the bad data correctly but also not falsely label the event data as bad ones.

Future work can be extended to investigate other machine learning models, e.g., long short-term memory (LSTM), convolutional neural network (CNN), graph neural network (GNN), deep reinforcement learning (DRL), on handling the bad data identification to achieve higher precision and recall scores. It is also worth exploring the potential deep learning applications for dynamic LSEs and full AC state estimations.

### REFERENCES

[1] W. Wang, J. Zhao, W. Yu et al., “FNETVision: a WAMS big data knowledge discovery system,” in Proceedings of 2018 IEEE PES General Meeting (PESGM), Portland, USA, Aug. 2018, pp. 1-5.

[2] P. H. Gadde, M. Biswal, S. Brahma et al., “Efficient compression of PMU data in WAMS,” IEEE Transactions on Smart Grid, vol. 7, no. 5, pp. 2406-2413, Sept. 2016.

[3] A. Sundararajan, T. Khan, A. Moghaddasi et al., “Survey on synchrophasor data quality and cybersecurity challenges, and evaluation of their interdependencies,” Journal of Modern Power Systems and Clean Energy, vol. 7, no. 3, pp. 449-467, May 2019.

[4] L. Zhang, H. Chen, K. Martin et al., “Successful deployment and operational experience of using linear state estimator in wide-area monitoring and situational awareness projects,” IET Generation, Transmission & Distribution, vol. 11, no. 18, pp. 4476-4483, Dec. 2017.

[5] K. D. Jones, A. Pal, and J. S. Thorp, “Methodology for performing synchrophasor data conditioning and validation,” IEEE Transactions on Power Systems, vol. 30, no. 3, pp. 1121-1130, May 2015.

[6] Y. Lin and A. Abur, “A highly efficient bad data identification approach for very large scale power systems,” IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 5979-5989, Nov. 2018.

[7] L. M. Putranto, R. Hara, H. Kita et al., “Series PMU data-based state estimation technique for WAMS application,” in Proceedings of 2016 IEEE PES General Meeting (PESGM), Boston, USA, Aug. 2016, pp. 1-5.

[8] A. Monticelli, “Electric power state system estimation,” Proceedings of the IEEE, vol. 88, no. 2, pp. 262-282, Feb. 2000.

[9] M. Pignati, L. Zanni, S. Sarri et al., “A pre-estimation filtering process of bad data for linear power systems state estimators using PMUs,” in Proceedings of 2014 Power Systems Computation Conference, Wroclaw, Poland, Aug. 2014, pp. 1-8.

[10] T. A. Pegoraro, A. Angioni, M. Pau et al., “Bayesian approach for distribution system state estimation with non-gaussian uncertainty models,” IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 11, pp. 2957-2966, Nov. 2017.

[11] J. Kang and D. Choi, “Distributed multi-area WLS state estimation integrating measurements weight update,” IET Generation, Transmission & Distribution, vol. 11, no. 10, pp. 2552-2561, Jul. 2017.

[12] Y. Wu, Y. Xiao, F. Hohn et al., “Bad data detection using linear WLS and sampled values in digital substations,” IEEE Transactions on Power Delivery, vol. 33, no. 1, pp. 150-157, Feb. 2018.

[13] C. Xu and A. Abur, “A fast and robust linear state estimator for very large scale interconnected power grids,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4975-4982, Sept. 2018.

[14] J. Yang, W. Wu, W. Zheng et al., “Performance analysis of sparse recovery models for bad data detection and state estimation in electric power networks,” in Proceedings of 2016 IEEE PES General Meeting (PESGM), Boston, USA, Jul. 2016, pp. 1-5.

[15] K. R. Mestav, J. Luengo-Rozas, and L. Tong, “Bayesian state estimation for unobservable distribution systems via deep learning,” IEEE Transactions on Power Systems, vol. 34, no. 6, pp. 4910-4920, Nov. 2019.

[16] W. Zhou, O. Ardakianian, H. Zhang et al., “Bayesian learning-based harmonic state estimation in distribution systems with smart meter and DPMU data,” IEEE Transactions on Smart Grid, vol. 11, no. 1, pp. 832-845, Jan. 2020.

[17] I. Ghahremani and I. Kamwa, “Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements,” IEEE Transactions on Power Systems, vol. 26, no. 4, pp. 2556-2566, Nov. 2011.

[18] J. Zhao, M. Netto, and L. Mili, “A robust iterated extended Kalman filter for power system dynamic state estimation,” IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 3205-3216, Jul. 2017.

[19] I. Krestulovic, V. Miranda, A. J. A. S. Costa et al., “Towards an auto-associative topology state estimator,” IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 3311-3318, Aug. 2013.

[20] S. Pal, B. Sikdar, and J. H. Chow, “Classification and detection of pmu data manipulation attacks using transmission line parameters,” IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 5057-5066, Sept. 2018.

[21] X. Wang, D. Shi, J. Wang et al., “Online identification and data recovery for PMU data manipulation attack,” IEEE Transactions on Smart Grid.
Online identification of bad synchrophasor measurements via spatio-temporal correlations,” in Proceedings of 2016 Power Systems Computation Conference (PSCC), Genoa, Italy, Jun. 2016, pp. 1-7.

Z. Yang, H. Liu, T. Bi et al., “Bad data detection algorithm for PMU based on spectral clustering,” Journal of Modern Power Systems and Clean Energy, vol. 8, no. 3, pp. 473-483, May 2020.

C. Han, H. Chen, M. Guo et al., “Wrong data identification and correction for WAMS,” in Proceedings of 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi’an, China, Oct. 2016, pp. 1903-1907.

K. Li, “A method of bad data identification based on wavelet analysis in power system,” in Proceedings of 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Zhangjiajie, China, May 2012, pp. 146-150.

M. Liao, D. Shi, Z. Yu et al., “Estimate the lost phasor measurement unit data using alternating direction multipliers method,” in Proceedings of 2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Denver, USA, Apr. 2018, pp. 1-9.

X. Deng, D. Bian, D. Shi et al., “Impact of low data quality on disturbance triangulation application using high-density PMU measurements,” IEEE Access, vol. 7, pp. 105054-105061, Jul. 2019.

D. Shi, D. J. Tylavsky, and N. Logic, “An adaptive method for detection and correction of errors in PMU measurements,” IEEE Transactions on Smart Grid, vol. 3, no. 4, Vancouver, Canada, Jul. 2012, pp. 1575-1583.

X. Wang, D. Shi, Z. Wang et al., “Online calibration of phasor measurement unit data using density-based spatial clustering,” IEEE Transactions on Power Delivery, vol. 33, no. 3, pp. 1081-1090, Jun. 2018.

A. G. Phadke and J. S. Thorp, Synchronized Phasor Measurements and Their Applications. Berlin: Springer, 2008.

K. D. Jones, J. S. Thorp, and R. M. Gardner, “Three-phase linear state estimation using phasor measurements,” in Proceedings of 2013 IEEE PES General Meeting, Vancouver, Canada, Aug. 2013, pp. 1-5.

T. Yang, H. Sun, and A. Bose, “Transition to a two-level linear state estimator - Part I: architecture,” IEEE Transactions on Power Systems, vol. 26, no. 1, pp. 46-53, Feb. 2011.

Z. Zhang and M. Sabuncu, “Generalized cross entropy loss for training deep neural networks with noisy labels,” in Advances in Neural Information Processing Systems, New York: Curran Associates, Inc., 2018, pp. 8778-8788.

T. Athay, R. Podmore, and S. Virmani, “A practical method for the direct analysis of transient stability,” IEEE Transactions on Power Apparatus and Systems, vol. PAS-98, no. 2, pp. 573-584, Mar. 1979.

Yingzhong Gu received his B.S. degree in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and the Ph.D. degree in electrical engineering at Texas A&M University, College Station, USA, in 2014. He currently works as a Senior Research Scientist at GEIRI North America, San Jose, USA. Previously, he was a Development Lead at GE Energy Consulting in Schenectady, USA. His research interests include linear state estimation, WAMS, dynamic optimization approach to power system operation, assessment of wind generation curtailment, spatio-temporal wind forecasting, and stochastic programming for advanced market operation.

Zhe Yu received the B.E. degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2009, the M.S. degree from the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA, in 2010, and the Ph.D. degree from the School of Electrical and Computer Engineering, Cornell University, Ithaca, USA, in 2016. He joined GEIRI North America, San Jose, USA, in 2017. His current research interests include power system and smart grid, demand response, dynamic programming, and optimization.

Ruisheng Diao received the Ph.D. degree in electrical engineering from Arizona State University, Tempe, USA, in 2009. He is currently the Deputy Department Head, AI System Analytics, GEIRI North America, San Jose, USA. He was a recipient of the 2018 Research and Development 100 Awards. He serves as Editor for IEEE Access, IEEE Transactions on Power Systems, IET Generation, Transmission & Distribution, and IEEE Power Engineering Letters. His research interests include high-fidelity simulation techniques and AI-based methods for power grid operation.

Di Shi received the Ph.D. degree in electrical engineering from Arizona State University, Tempe, USA. He is the Director of Fundamental R&D Center and Department Head of AI System Analytics at GEIRI North America, San Jose, USA. He is an Editor of IEEE Transactions on Smart Grid and Power Engineering Letters. His research interests include WAMS, data analytics, application of AI in power systems, and energy storage systems.