

# Bad Data Detection and Identification of Hybrid AC/DC Power Systems with Voltage Source Converters Using Deep Belief Network and K-Means Clustering

Tong Zhang<sup>1</sup>, Hucheng Li<sup>1</sup>, Yu Huang<sup>2</sup>, Mingyu Zhai<sup>3</sup>, Fei Zeng<sup>1</sup>

<sup>1</sup>State Grid Jiangsu Electric Power Co., Ltd. Research Institute,  
Nanjing, China

<sup>2</sup>School of Electrical Engineering, Southeast University,  
Sipailou Campus of Southeast University, Nanjing, China

<sup>3</sup>NARI Group Corporation, State Grid Electric Power Research Institute,  
Nanjing, China  
huangyu@seu.edu.cn

**Abstract**—The widespread use of Voltage Source Converter based High Voltage Direct Current Transmission (VSC-HVDC) technology has significantly increased the complexity of grid configuration and operation, which demands higher quality of AC/DC state estimation. The bad data detection and identification plays a vital role in ensuring the accuracy of state estimation outcomes. This paper presents a new approach to bad data in hybrid AC/DC grids based on the combined deep belief network (DBN) and K-means clustering method. First, the DBNs are trained separately for active and reactive power given the characteristics of VSC-HVDC by which the bad data can be detected. Then, an improved K-means is used for clustering the DBN outputs by setting the mean and the number of the samples within the clusters as two metrics for bad data identification. Finally, the case study is performed in a modified IEEE 14-bus system and the results demonstrate the effectiveness of the proposed method in terms of both the accuracy and efficiency.

**Index Terms**—AC/DC; VSC; Bad data; Deep belief network; K-means clustering; State estimation.

## I. INTRODUCTION

The technical development of Voltage Source Converter based High Voltage Direct Current Transmission (VSC-HVDC) has become an effective solution to the critical issue of renewable energy integration and utilization [1]. However, the complexity of the hybrid AC/DC network is challenging for the system operation and control, which relies on accurate information of the grid status [2]. Since the bad data is inevitable in the tele-measuring process, it is important to identify them to ensure the reliability of state estimation, and, thus, provide system operators with data-based advanced application module for the AC/DC

system.

A traditional way of identifying bad data is to take the normalized/weighted residuals as eigenvalues, such as Largest normalized residual (LNR) test method [3], [4], which need repeated detection and identification of the bad data in a recursive manner after the state estimation. Specifically, if there are  $k$  detectable bad measurements, the state estimation based on weighted least square (WLS) algorithm along with LNR test should be executed at least  $k$  times before removing all bad data in the entire measurement set. This type of approaches is computationally intensified in the large-scale hybrid AC/DC power systems, given the high-dimensionality of the residual sensitivity matrix.

At present, the data-driven learning algorithms, like support vector machine (SVM) and artificial neural network (ANN), have been gaining increasing popularity in the application of bad data detection and identification [5], [6]. In particular, the research on this field is inspired by the data mining techniques that combine the historical data and prior knowledge. In [7], Chen *et al.* presents a SVM-based approach to identifying the bad measurement data, as well as its application in the generation automatic voltage control (AVC) systems. However, this method is only suitable for small sets of training samples. In other words, it builds up the identification model for one component at a time, e.g., a single generator or a transmission line, which is computationally inefficient and has low accuracy. In [8], a back propagation (BP) neural network is constructed as a filter of the state estimation, and the bad data is identified using the measurement section in the standard working conditions to train the network. Although the BP neural network has strong capabilities of self-learning and fitting the complex and non-linear functions, its network parameters are initialized arbitrarily, which may lead to poor generalization and be easily trapped into local optimization. To circumvent this problem, an alternative way is to enlarge the scale of training for the entire power flow model. As the expansion of

Manuscript received 24 October, 2018; accepted 4 January, 2019.

This research was funded by a grant (J2018049) from the State Grid Jiangsu Electric Power Company Science and Technology Program and supported by the Scientific Research Foundation of Graduate School of Southeast University.

AC/DC grids along with the measurement types and influencing factors, shallow networks may run into the bottleneck in reflecting the complex relationships between input and output variables.

This paper explores an application of using the deep belief network (DBN) approach to the bad data detection and identification in hybrid AC/DC power systems. The training datasets are pre-processed in blocks based on the characteristics of the hybrid system integrating VSCs, and two DBNs are trained respectively for active and reactive power. The DBN outputs are weighted residuals that are used to detect the bad measurements through hypothesis testing. As compared to BP method, the DBN has better initialized network parameters obtained from the unsupervised greedy learning algorithm layer by layer in the pre-training, which could avoid the issue of under-fitting or over-fitting [9]. When the bad data in the measurement section are detected, we use an improved K-means algorithm based on distance cost function to cluster the pre-processed DBN outputs. This clustering approach is to identify the bad data by setting the number of clusters and the distance to the clustering centre. Finally, the effectiveness of the proposed method is verified in a modified IEEE 14-bus test system with multiple VSCs.

The main contribution of the paper is to present an effective data-driven approach to bad data detection and identification in the hybrid AC/DC power systems, which could avoid the issue of computing the high-dimensional and full rank residual sensitivity matrix. It has made a big progress in the aspects of both practical application and detection methodology, which are detailed as:

1. The proposed decoupling strategy and the generalized training model for DC parts are also adaptable to AC/DC network, and, thus, need no additional DC detection modules;
2. The computation burden of WLS estimator could be reduced by enabling the bad data to be detected and eliminated a priori;
3. The two DBNs separately designed for active and reactive power could lower the error detection rate, improve the training efficiency, as well as provide optimized pre-processing data (initial values) for K-means clustering;
4. Two identification indicators are newly defined according to the characteristics of power network. The bad data of different types and of deviation degrees are finally identified using the improved K-means clustering method based on the distance cost function.

The rest of the paper is organized as follows. Section II introduces the state estimation model and the decoupling strategy for the AC/DC system. A brief introduction of DBN is given in Section III, together with its application in bad data detection. In Section IV, the K-means clustering method based on the distance cost function is presented to handle the DBN outputs for bad data identification. Section V shows some comparative case studies for various scenarios, while concluding remarks are made in Section VI.

## II. STATE ESTIMATION MODEL OF THE AC/DC SYSTEM

The VSC is a critical control element in the AC/DC

network, which is functioned as a connection bond between the AC main grid and DC subsystems [10]. Suppose that a hybrid system is coupled via  $n_c$  VSCs. Figure 1 shows the steady-state single-phase equivalent model of the  $i$ th VSC. Note that the harmonic components are neglected in the steady-state operation.

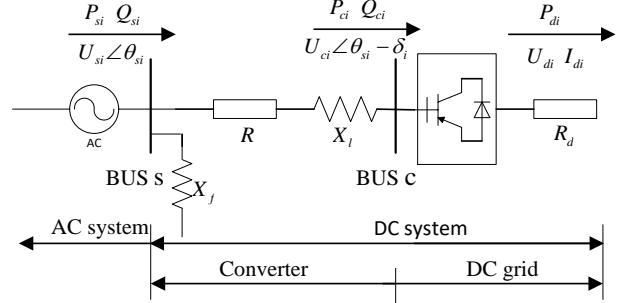


Fig. 1. Steady-state model of single-phase VSC.

As can be seen in Fig. 1,  $U_{s,i}$  is the voltage magnitude at the joint bus, called point of common coupling (PCC). The AC bus is connected successively to the converter bridge and DC network through the filter  $X_{f,i}$  and converter reactor  $X_{l,i}$ .  $U_{c,i}$  is the voltage magnitude of VSC output.  $\delta_i$  is the phase angle of the pulse width modulation (PWM), which represents the difference between the AC voltage angle and the modulating wave [11].

The non-linear measurement equations in state estimation of the AC/DC system can be expressed as:

$$\mathbf{z} = \begin{bmatrix} z_{ac} \\ z_{dc} \end{bmatrix} = \begin{bmatrix} h_{ac}(x_{ac}, x_{dc}) \\ h_{dc}(x_{ac}, x_{dc}) \end{bmatrix} + \begin{bmatrix} \eta_{ac} \\ \eta_{dc} \end{bmatrix}, \quad (1)$$

$$\begin{cases} \mathbf{x}_{ac} = [\theta_i \ U_i]^T, i = 1, 2, \dots, n_a, \\ \mathbf{x}_{dc} = [u_{d,i} \ U_{c,i} \ \delta_i]^T, i = 1, 2, \dots, n_c, \end{cases} \quad (2)$$

where  $\mathbf{z}$ ,  $\boldsymbol{\eta}$  and  $\mathbf{x}$  are vectors of measurement, measurement errors, and state variables, respectively;  $h(x)$  is measurement vector function;  $U_i$  and  $\theta_i$  are voltage magnitude and phase angle of bus  $i$  in the AC grid.  $u_{d,i}$  is the DC voltage.

To enable the decoupling of the active and reactive power in the hybrid system, we need to omit some secondary factors in the Jacobi matrix based on the steady-state characteristics of AC/DC grids. Given that the negative sequence voltage in the grid is confined to 1.5 % of the rated power, to ensure the long-term operation of VSCs, the negative sequence current through the converter reactor  $X_{l,i}$  should be no larger than 5 %~10 % of the rated VSC current [12]. Thus, we have  $X_{l,\min} = 1/2 \cdot \sin(2\delta_{i,\max}) \in [0.15 \sim 0.29]$  p.u. and

$\delta_{i,\max} \in [8.6^\circ \sim 17.4^\circ]$ . It can be deduced that under normal operations the coupling between the active power of VSC ( $P_{s,i}$ ,  $P_{c,i}$ ) and DC state variable ( $U_{c,i}$ ), the reactive power ( $Q_{s,i}$ ,  $Q_{c,i}$ ) and the DC voltage ( $u_{d,i}$ ) and angle ( $\delta_i$ ) are weak and negligible [13]. Besides, the entries  $\partial P / \partial U$  and  $\partial Q / \partial \theta$  in the Jacobi sub-matrix of AC side are close to 0, indicating weak correlations of active and reactive measurements. Therefore, it is reasonable to separate the active and reactive power measurements. The modified measurement and Jacobi matrices are constructed as:

$$\begin{cases}
H(x_p, x_Q) = \begin{bmatrix} H_p(x_p) & 0 \\ 0 & H_Q(x_Q) \end{bmatrix} = \begin{bmatrix} \partial z_p / \partial x_p & 0 \\ 0 & \partial z_Q / \partial x_Q \end{bmatrix}, \\
\mathbf{z}_p = [P_i^m, P_{ij}^m, P_d^m]^T, \\
\mathbf{z}_Q = [Q_i^m, Q_{ij}^m, Q_d^m]^T, \\
\mathbf{x}_p = [\theta, u_d, \delta]^T, \\
\mathbf{x}_Q = [U, U_c]^T,
\end{cases} \quad (3)$$

where  $H_p(x_p)$  and  $H_Q(x_Q)$  are Jacobi matrix regarding the active and reactive power measurements, respectively.  $Z_p$  and  $Z_Q$  are vectors of active and reactive measurements, among which  $P_i^m$  and  $P_{ij}^m$  are measurements for AC active power injections and line flows, respectively;  $P_d^m$  represents DC active power measurements associated with the  $i$ th VSC including  $P_{d,i}$ ,  $P_{s,i}$ ,  $P_{c,i}$ , and  $P_{d,ij}$ .  $Q_i^m$  and  $Q_{ij}^m$  are measurements for AC reactive power injections and line flows, respectively;  $Q_d^m$  represents reactive power measurements associated with the  $i$ th VSC including  $Q_{s,i}$  and  $Q_{c,i}$ .

### III. BAD DATA DETECTION USING DBN

The DBN could well approximate the non-linear mapping of multiple inputs to outputs and has some capability of fault-tolerance [14]. It provides a possible solution to the bad data detection and identification, since it is especially useful for large datasets. The critical problem is how to design the DBN-based data detection model for an effective training.

#### A. Bad Data Detection Model

As is discussed in Section II, the modified sets of active and reactive power measurements are weakly coupled in the AC/DC system integrating VSCs. Therefore, two DBNs are trained separately to detect the bad data, one for active and one for the reactive power to reduce the large training size and improve the efficiency.

Since the high requirements of training samples for reliable outcomes, we take the historical grid measurements as input data and the state estimation results as outputs. Based on the decoupling strategy assumed in Section II, the active and reactive power measurements are sorted into two training datasets as follows:

$$\begin{cases}
z_p = [P_i^m, P_{ij}^m, P_d^m]^T, \\
y_p = [P_i, P_{ij}, P_d]^T,
\end{cases} \quad (4)$$

$$\begin{cases}
z_Q = [Q_i^m, Q_{ij}^m, Q_d^m]^T, \\
y_Q = [Q_i, Q_{ij}, Q_d]^T,
\end{cases} \quad (5)$$

where the active and reactive training datasets are taken from (4) and (5), respectively.  $z_p$ ,  $y_p$  and  $z_Q$ ,  $y_Q$  are the DBN inputs and outputs for active and reactive power, respectively.

The architecture of the proposed DBN for bad data detection is shown in Fig. 2. It is composed of one input layer, one output layer, and several hidden layers. The lower layers aim at extracting features from the input data that are then clamped into higher layers. The output of the top layer is a BP neural network, functioned as the regression layer.  $w_l$  is the

set of connection weights between the  $l$ -1th and the  $l$ th hidden layer.  $b_l$  is the bias vector of the neurons in the  $l$ th layer.

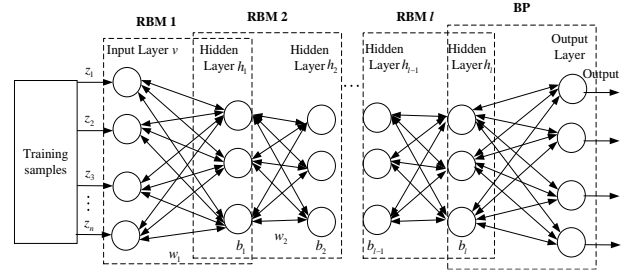


Fig. 2. DBN structure for bad data detection.

#### B. DBN Training

A DBN is piled up by several basic layers, called restricted Boltzmann machine (RBM), which are trained independently to provide the network with better initial parameters ( $w_l$ ,  $b_l$ ), and, thus, solve the intrinsic problem of local optimum for most ANNs [15]. The training process contains two stages: unsupervised pre-training and supervised fine tuning using BP algorithm.

The joint probability of the visible units  $\mathbf{v} = (v_i)$  and hidden units  $\mathbf{h} = (h_j)$  for each RBM is given by [16]

$$P(\mathbf{v}, \mathbf{h}; \boldsymbol{\psi}) = \frac{e^{-E(\mathbf{v}, \mathbf{h}; \boldsymbol{\psi})}}{\sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}; \boldsymbol{\psi})}}, \quad (6)$$

where  $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\psi})$  is the energy function defined as

$$E(\mathbf{v}, \mathbf{h}; \boldsymbol{\psi}) = -\sum_{i=1}^n \sum_{j=1}^m \omega_{ij} v_i h_j - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j, \quad (7)$$

where  $\boldsymbol{\psi} = \{\mathbf{w} = (w_{ij})_{n \times m}, \mathbf{a} = (a_i)_n, \mathbf{b} = (b_j)_m\}$  are sets of network parameters of the RBM.  $w_{ij}$  is the connecting weight between the visible unit  $v_i$  and hidden unit  $h_j$ .  $a_i$ , and  $b_j$  denote the bias of the units  $v_i$  and  $h_j$ .  $n$  and  $m$  are the number of visible and hidden units, respectively.

Since there are no connections between any two hidden units within the same layer, given a particular random input configuration  $\mathbf{v}$ , the probability of the binary unit  $h_j$  set to 1 becomes [17]

$$P(h_j = 1 | \mathbf{v}) = f\left(\sum_{i=1}^n w_{ij} v_i + b_j\right), \quad (8)$$

where  $f(x)$  is the sigmoid function, i.e.,  $f(x) = (1 + e^{-x})^{-1}$ .

Similarly, given the hidden state  $\mathbf{h}$ , the probability of visible unit  $v_i$  set to 1 is given by

$$P(v_i = 1 | \mathbf{h}) = f\left(\sum_{j=1}^m w_{ij} h_j + a_i\right). \quad (9)$$

As for the above RBM model, this paper uses the contrastive divergence (CD- $k$ ) algorithm to optimize the parameters to fit the input training datasets [18]. The update rules for the parameters can be expressed as:

$$\begin{cases} \Delta w_{ij} = \varepsilon_{CD} (\langle v_i h_j \rangle - \langle v_i' h_j' \rangle), \\ \Delta a_i = \varepsilon_{CD} (\langle v_i \rangle - \langle v_i' \rangle), \\ \Delta b_j = \varepsilon_{CD} (\langle h_j \rangle - \langle h_j' \rangle), \end{cases} \quad (10)$$

where  $\varepsilon_{CD}$  represents the learning rate.  $\langle \cdot \rangle$  denotes the expectations for the random variable.  $v_i'$  and  $h_j'$  are the updated states of  $v_i$  and  $h_j$  at the  $k$ th step.

Specifically, the training process of RBM, also called pre-training [19], is unsupervised by nature that each RBM receives the inputs from the previous layer and feeds to the RBM in the next layer. Hence, the parameters of each RBM are updated accordingly, starting by the lower-level RBM and progressively moving up in the hierarchy.

After the pre-training stage, the resulting network is fine-tuned using the BP algorithm to reach the global optimum [20]. We take the sum of square errors of outputs as the loss function defined as

$$C(\boldsymbol{\psi}) = \frac{1}{2N} \sum_x \|y - a_{\boldsymbol{\psi}}(z)\|^2, \quad (11)$$

where  $N$  is the total training samples.  $z$  denotes a specific training sample.  $a_{\boldsymbol{\psi}}$  is the network outputs for the input  $z$ .  $y$  is the associated expected outputs. The parameters  $\boldsymbol{\psi}$  are adjusted in the negative gradient direction that minimize the error  $C(\boldsymbol{\psi})$ .

### C. Bad Data Detection

When the training datasets are large enough, the trained DBN could well approximate the complex non-linear model, and is, thus, suitable for bad data detection. The active and reactive measurements to be detected are input to the DBN, we obtain the weighted output residual  $r_{bw}$ , which can be calculated as

$$r_{bw,i} = \frac{|z_i - o_i|}{\sigma_i}, \quad (12)$$

where  $z_i$  and  $o_i$  are the  $i$ th measurement value and the  $i$ th DBN output, respectively.  $\sigma_i$  is the standard deviation of the  $i$ th measurement point.

The output of the trained DBN can be very close to the expected value, i.e.,  $o_i \approx y_i$ . Since the neural networks are error-tolerated, a few of bad data input to the network will not largely influence the overall training outcomes, and the relationship between the DBN output and the expected value still holds. Based on this, the bad data could be detected by setting the threshold for  $r_{bw}$ . Given that under normal measurement conditions the probability of the measurement error greater than  $3\sigma_i$  is very low [21], we set the following threshold  $\gamma$  as the bad data detection criterion

$$r_{bw,i} > \gamma. \quad (13)$$

Using (13) to detect the AC/DC measurements one by one, i.e., any measurement that exceed this threshold are considered as the suspected bad data. We can come to the remarks that, if all of the output residuals are less than the threshold, the sets of measurements are normal and can be proceeded to the state estimation directly. On the other hand, if there exists  $r_{bw}$  greater than the threshold, output the datasets that contain bad measurements, and includes them to the database for the subsequent bad data identification based on K-means clustering.

## IV. BAD DATA IDENTIFICATION BASED ON K-MEANS CLUSTERING

### A. K-means Clustering Method

The K-means method aims at partitioning the datasets into  $k$  different clusters. Each object in the set is distributed to the cluster, which has the shortest Euclidean distance. Then, we re-calculate the means and form the new clustering centers. This process is repeated until the function in (14) converges [22]:

$$\begin{cases} E = \sum_{k=1}^K \sum_{x \in C_k} \text{dist}(m_k, x)^2, \\ \text{dist}(x_i, x_j) = [\sum_{k=1}^p (x_{ik} - x_{jk})^2]^{1/2}, \end{cases} \quad (14)$$

where  $E$  is the error criterion function, denoting the sum of square errors of the clustering objects  $x$ .  $m_k$  is the mean of the  $k$ th cluster  $C_k$ , representing the center of the cluster.  $\text{dist}(\cdot)$  denotes the Euclidean distance.  $x_i$  and  $x_j$  are two  $p$ -dimensional objects.

The obtained  $k$  clusters, based on the distance error criterion, have the merits of independence and compactability, which could maximize the discrepancy of different clusters and is beneficial to tell apart the bad data from normal measurements. Therefore, the aim of identifying the bad data can be fulfilled by analysing the characteristics of different clusters.

### B. Bad Data Identification

The output residuals of active and reactive power based DBNs are taken as the database for clustering. Since there is little chance that several bad data appear in the same set of measurements at one time based on the intrinsic feature of power systems, we can set the number of clusters  $k_{\max}$  and initialize the value  $k = 2$ .

The ideal number of clusters should be two in this identification issue, i.e., one for sets of normal measurements and one for bad data. However, given that the degree of deviation for the bad data can be different,  $k = 2$  may not be the optimal number of clusters. We, therefore, use the distance cost function  $D$  in (15)–(17) to find the optimal number of clusters, which could reflect two essential indices of the distance in-between and within the clusters [23], i.e.,  $L_g$  and  $L_n$ :

$$L_g = \sum_{k=1}^K |m_k - m|, \quad (15)$$

$$L_n = \sum_{k=1}^K \sum_{x \in C_k} |x - m_k|, \quad (16)$$

$$D = L_g + L_n, \quad (17)$$

where  $m_k$  and  $m$  are the means of the clustered and global objects, respectively.  $L_g$  is the sum of the distances from the clustering center  $C_k$  to the global center  $m$ .  $L_n$  is the sum of distances from each object within the cluster to the associated clustering center  $m_k$ .

The clustering analysis is carried out for  $k_r = 2 \sim k_{\max}$  clusters by calculating and searching for the minimum distance cost function  $D^*$  and outputting the associated  $k_r^*$  clusters. Given that there is a few bad data in the sets of measurements for power systems and they usually deviate large distances from normal measurements, we choose two indicators (cluster center distance  $m_k$  and number of objects within the cluster) as the reference indices for bad data identification. Within the  $k_r^*$  clusters, the clusters of small object numbers along with big center distance  $m_k$  are regarded as the target sets, which contain the bad data and need further identification for each data object.

## V. CASE STUDY

### A. The Bad Data Detection Using DBN

The proposed method is tested on a modified IEEE 14-bus AC/DC system incorporating VSCs (Fig. 3).

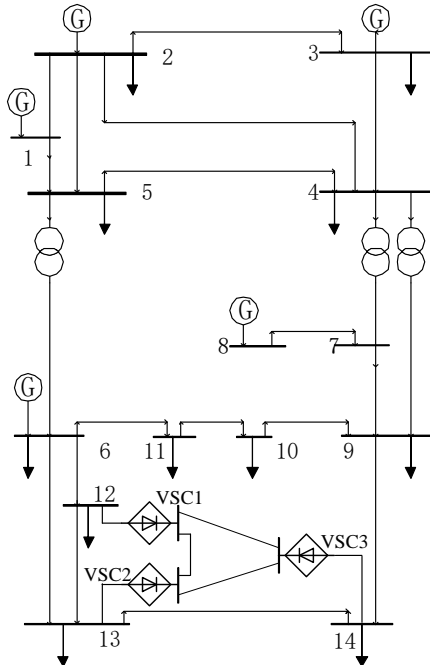


Fig. 3. The modified IEEE 14-bus test system incorporating VSCs.

The training datasets are normalized first with active and reactive AC/DC measurements as DBN inputs and state estimation results as outputs. Meanwhile, to make the test results more applicable to real power system operations, the loads are magnified to simulate the sets of power flow measurements under different conditions. Specifically, the

AC/DC loads are multiplied by a random number  $n$  ( $n=1\sim 4$ ) to generate 4000 sets of power flow measurements. The parameters of converter reactance and resistance of each VSC are set to be 0.15 p.u. and 0.012 p.u., respectively. The DBNs used in the case study have two hidden layers, with 25 and 21 units for the active power, while 24 and 20 units for the reactive power. The learning rate is set to be 0.01 and the number of iteration is 5000.

#### 1) Test results for small bad data

Firstly, we set the small bad data (error magnitude =  $10\sigma$ ) for the power flow measurements ( $P_{s,3}$ ,  $Q_{s,3}$ ,  $P_{14-9}$ , and  $Q_{14-9}$ ) to simulate the bad data in the AC/DC measurement system. Multiple tests on the bad data detection are performed and the results (for active power only) are shown in Table I.

TABLE I. SMALL BAD DATA RESULTS FOR ACTIVE POWER OF DBN OUTPUTS.

Multiple of AC/DC loads	Suspect set	
	Measurement ( $r_{bw}$ )	
1	$P_{14-9}$ (8.8157)	$P_{s,3}$ (9.3919)
2	$P_{14-9}$ (8.2386)	$P_{s,3}$ (9.7690)
3	$P_{14-9}$ (7.9185)	$P_{s,3}$ (8.8965)
4	$P_{14-9}$ (8.6939)	$P_{s,3}$ (8.8127)

Note: the bad data points are  $P_{14-9}$  and  $P_{s,3}$ .

It can be seen that the weighted residual of DBN outputs  $r_{bw}$  for the active power measurements are larger than the predefined threshold  $\gamma = 3$ , and several bad data is successfully detected. Similar conclusions can be made for the reactive power measurements. Therefore, the test results demonstrate that whatever the bad data exist in AC or DC measurements, it is possible to effectively detect and identify the small bad data, as well as their measurement point positions using the DBN.

#### 2) Test results for large bad data

This time we set the large bad data (error magnitude =  $20\sigma$ ) in the same power flow measurements ( $P_{s,3}$ ,  $Q_{s,3}$ ,  $P_{14-9}$ , and  $Q_{14-9}$ ) for multiple trials of bad data detection and identification. The results of the active power outputs are shown in Table II.

TABLE II. LARGE BAD DATA RESULTS FOR ACTIVE POWER OF DBN OUTPUTS.

Multiple of AC/DC loads	Suspect set				
	Measurement ( $r_{bw}$ )				
1	$P_{14}$ (4.0427)	$P_{9-14}$ (3.3176)	$P_{14-9}$ (17.4259)	$P_{s,3}$ (19.2218)	---
2	$P_{14}$ (4.3521)	$P_{14-9}$ (17.9608)	$P_{s,3}$ (19.4976)	---	---
3	$P_{14}$ (4.9251)	$P_{10-11}$ (3.3372)	$P_{13-14}$ (3.9853)	$P_{14-9}$ (17.5451)	$P_{s,3}$ (18.8767)
4	$P_{14}$ (4.2853)	$P_{14-9}$ (18.1639)	$P_{s,3}$ (18.6243)	---	---

It is observed that for large bad data, the detection results are the suspected bad measurement sets. Each set of power flow measurements not only contains the bad data ( $P_{14-13}$  and  $P_{s,3}$ ), but has a few normal measurements. This is because that large bad data may influence the training output, which leads to the errors by misdetecting the normal measurements. Similar conclusions can be made for the reactive power measurements.

3) Comparisons of training performance

Comparative studies are performed using the sets of measurements in Table II between the proposed DBN (active and reactive power model), decoupling strategy, and DBN (network power model) and BP neural network. In order to quantify and compare the accuracy of the training models above, deviation of training outputs based on weighted output residual  $r_{bw}$  is defined below:

$$\text{Deviation of training outputs} = \frac{\text{mean value of } r_{bw} \text{ among the rest meas}}{\text{mean value of } r_{bw} \text{ among suspect bad data}} \quad (18)$$

Note that number of the rest meas = number of meas - number of suspect bad data.

The smaller the deviation of training outputs is, the more accurate the training results are. That is, better-preprocessed database can be provided to K-means clustering, which facilitate identification of bad date. As can be seen from Fig. 4, the training deviation of the proposed DBN (active and reactive power model) is similar to DBN (network power model), and is better than that of BP network. It shows that decoupling training strategy of DBN maintains almost the same accuracy as the conventional DBN (network power model). In addition, comparisons of detection results are shown in Table III. It can be seen that the number of data detected by mistakes is reduced using the DBN compared to those of BP network, and the ratio of error detection is decreased by 18.03 %.

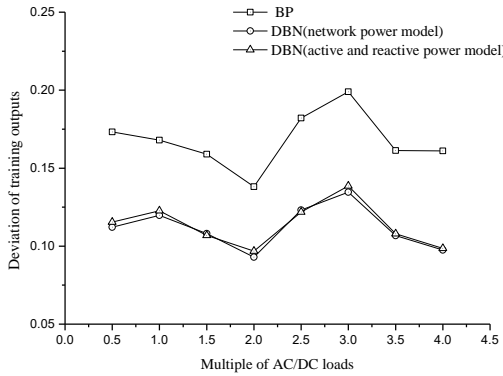


Fig. 4. Deviation of training outputs comparisons.

TABLE III. COMPARISONS OF BAD DATA DETECTION.

		Multiple of AC/DC loads				Average e
		1	2	3	4	
Number of suspect bad data	BP	6	4	7	5	5.5
	DBN (Network power model)	4	3	5	3	3.75
	DBN (Active and reactive power model)	4	3	5	3	3.75
Number of good meas. Identified as bad	BP	4	2	5	3	3.5
	DBN (Network power model)	2	1	3	1	1.75
	DBN (Active and reactive power model)	2	1	3	1	1.75

Finally, the computation time for two training models are compared. One is the traditional network power model, and the other is the proposed classified power model. The results are shown in Table IV.

TABLE IV. THE RESULTS OF TRAINING TIME.

Training model	IEEE 14	IEEE 30
Network power model	100	100
Active and reactive power model	56.21	42.36

It is clear that the classified active and reactive power model is more computationally efficient and takes the advantage when the system is expanded.

B. The Bad Data Identification Using K-means Clustering

The weighted residual of DBN outputs are the pre-processed input data for the K-means clustering, which is used for bad data identification based on the two indicators (mean and number of the sample within the cluster). Take an example of the sets of measurements in Table II with the loads multiplied by 3; we examine the effect of identification for several large bad data of similar deviation degree (Case 1). The results are shown in Table V.

TABLE V. BAD DATA IDENTIFICATION RESULTS FOR K-MEANS CLUSTERING (CASE 1).

Number of clusters	Mean value of the sample within the cluster	Distance cost function D	Bad data
2	18.2109(2), 1.3557(57)	76.4421	Ps,3 P14-9
3	18.2109(2), 2.5296(20), 0.7419(37)	77.7763	Ps,3 P14-9
4	18.2109(2), 3.1513(10), 1.5413(22), 0.5097(25)	79.0238	Ps,3 P14-9
5	18.2109(2), 4.2374(3), 2.2247(17), 1.0809(20), 0.3432(17)	80.3625	Ps,3 P14-9 P14 P2 P13-14

It can be seen that the distance cost function  $D$  reaches the minimum when  $k = 2$ , indicating the best identification result. As compared to the suspected bad datasets in Table II, the bad data points  $P_{14-13}$  and  $P_{14-9}$  are identified, while the normal measurements  $P_{14}$ ,  $P_{10-11}$ , and  $P_{13-14}$  are not detected. This demonstrates that the K-means clustering could effectively reduce the range of the suspected bad datasets and facilitate the bad data identification.

TABLE VI. BAD DATA IDENTIFICATION RESULTS FOR K-MEANS CLUSTERING (CASE 2).

Number of clusters	Mean value of the sample within the cluster	Distance cost function D	Bad data
2	19.2754(1), 1.0270(58)	77.5405	Ps,3
3	19.2754(1), 2.5915(10), 0.7011(48)	78.8680	Ps,3
4	19.2754(1), 7.4193(1), 1.7040(18), 0.5507(39)	77.4446	Ps,3 P14-9
5	19.2754(1), 7.4193(1), 1.8834(13), 0.9505(23), 0.2763(21)	80.1161	Ps,3 P14-9

Set the bad data of different deviation degree in the measurements  $P_{14-9}$  and  $P_{s,3}$  (Error magnitude =  $9\sigma$  and  $20\sigma$ ,

respectively) and keep other conditions unchanged, we can test the effect of identification for bad data with different error magnitude (Case 2) as shown in Table VI.

It is observed that the optimal number of clusters for the bad data with different error magnitude is not the ideal value ( $k = 2$ ). The optimal  $k$  should be 4 for a minimum distance cost function  $D$ , where we have the best identification results.

## VI. CONCLUSIONS

In this paper, we present a new approach for bad data detection and identification using the deep belief network and K-means clustering in the hybrid AC/DC grids. The following conclusions can be drawn as follows:

1. In terms of the effectiveness, the DBN-based approach could detect and identify the position of the small bad data measurements. For large bad data, the suspected sets of measurements can be detected. As compared to the BP neural network, the DBN could effectively reduce the range of bad datasets, as well as the ratio of error detection by 18.03 %;
2. In terms of the training efficiency, two DBNs are trained separately for the active and reactive power based on the characteristics of AC/DC grids. This model is particularly computationally efficient, when the system is expanded;
3. In terms of the applicability, the K-means clustering method is used to identify the exact bad data from the suspected sets of measurements. This method is valid for both AC and DC measurements and would not be affected by the complexity of the system topology.

## REFERENCES

- [1] R. Feldman, M. Tomasini, E. Amankwah, "A hybrid modular multilevel voltage source converter for HVDC power transmission", *IEEE Transactions on Industry Applications*, vol. 49, no. 2, pp. 1577–1588, 2013. DOI: 10.1109/TIA.2013.2257636.
- [2] S. Soltan, M. Yannakakis, G. Zussman, "Power Grid State Estimation Following a Joint Cyber and Physical Attack", *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 499–512, 2018. DOI: 10.1109/TCNS.2016.2620807
- [3] A. Abur, Y. Lin, "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, 2018. DOI: 10.1109/TPWRS.2018.2826980.
- [4] E. Caro, A. J. Conejo, R. Mínguez, "Multiple bad data identification considering measurement dependencies", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 1953–1961, 2011. DOI: 10.1109/TPWRS.2011.2157366.
- [5] M. Hossam, M. E. El-Hawary, "Optimized Neural Network Parameters Using Stochastic Fractal Technique to Compensate Kalman Filter for Power System-Tracking-State Estimation", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 1–10, 2018. DOI: 10.1109/TNNLS.2018.2839101.
- [6] C. Tu, X. He, Z. Shuai, "Big data issues in smart grid—A review", *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1099–1107, 2017. DOI: 10.1016/j.rser.2017.05.134.
- [7] B. Chen, Y. Liu, Z. Jing, "Fake Measurement Detection in Automatic Voltage Control using Support Vector Machine", in *Proc. Power and Energy Engineering Conference (APPEEC)*, 2010 Asia-Pacific IEEE, 2010, pp. 1–4. DOI: 10.1109/APPEEC.2010.5448729.
- [8] S. J. Huang, J. M. Lin, "Enhancement of power system data debugging using GSA-based data-mining technique", *IEEE Transactions on power systems*, vol. 17, no. 4, pp. 1022–1029, 2002. DOI: 10.1109/TPWRS.2002.804992.
- [9] G. E. Hinton, S. Osindero, Y. W. Teh, "A fast learning algorithm for deep belief nets", *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006. DOI: 10.1162/neco.2006.18.7.1527.
- [10] L. Zhang, L. Harnefors, H. P. Nee, "Modeling and control of VSC-HVDC links connected to island systems", *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 783–793, 2011. DOI: 10.1109/TPWRS.2010.2070085.
- [11] Y. N. Liu, Y. Xu, H. C. Han, "Load Balancing of Multi-terminal Back-to-back Flexible DC Interconnection System", *Electric Power Engineering Technology*, vol. 37, no. 3, pp. 84–90, 2018 [Online]. Available: CNKI:SUN:JSDJ.0.2018-03-017.
- [12] Z. Xu, H. Xiao, Z. Zhang, "Design of Main Circuit Parameters of Modular Multilevel Converters", *High Voltage Engineering*, vol. 41, no. 8, pp. 2514–2527, 2015. DOI: 10.13336/j.1003-6520.hve.2015.08.005.
- [13] T. Zhang, Y. Wang, M. Zhai, "Fast Decoupling Algorithm of State Estimation for Hybrid AC/DC Power Systems with Voltage Source Converters", *Automation of Electric Power Systems*, vol. 42, no. 21, pp. 70–76, 2018. DOI: 10.7500/AEPS20170911012.
- [14] N. Lopes, B. Ribeiro, "Towards adaptive learning with improved convergence of deep belief networks on graphics processing units", *Pattern recognition*, vol. 47, no. 1, pp. 114–127, 2014. DOI: 10.1016/j.patcog.2013.06.029.
- [15] T. Kuremoto, S. Kimura, K. Kobayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines", *Neurocomputing*, vol. 137, no. 15, pp. 47–56, 2014. DOI: 10.1016/j.neucom.2013.03.047.
- [16] Y. Huang, Q. Xu, X. Jiang, "Modelling Correlated Forecast Error for Wind Power in Probabilistic Load Flow", *Elektronika ir Elektrotehnika*, vol. 23, no. 5, pp. 61–66, 2017. DOI: 10.5755/j01.eie.23.5.19244.
- [17] G. Hinton, L. Deng, D. Yu, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012. DOI: 10.1109/MSP.2012.2205597.
- [18] E. R. Merino, F. M. Castrillejo, J. D. Pin, "Neighborhood-Based Stopping Criterion for Contrastive Divergence", *IEEE Transactions on Neural Networks & Learning Systems*, vol. 29, no. 7, pp. 2695–2704, 2018. DOI: 10.1109/TNNLS.2017.2697455.
- [19] Y. Bengio, "Learning deep architectures for AI", *Foundations and trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009. DOI: 10.1561/22000000006.
- [20] H. A. Soodi, A. M. Vural, "STATCOM Estimation Using Back-Propagation, PSO, Shuffled Frog Leap Algorithm, and Genetic Algorithm Based Neural Networks", *Computational Intelligence & Neuroscience*, vol. 3, no. 1, pp. 1–17, 2018. DOI: 10.1155/2018/6381610.
- [21] A. Gomez-Exposito, A. Abur, "Power system state estimation: theory and implementation", in CRC press, pp. 1–154, 2004.
- [22] Y. Huang, Q. Xu, Y. Yang and X. Jiang, "Numerical method for probabilistic load flow computation with multiple correlated random variables," *IET Renew. Power Gen.*, vol. 12, no. 11, pp. 1295–1303, 2018. DOI: 10.1049/iet-rpg.2017.0067.
- [23] Y. S. Li, S. L. Yang, X. J. Ma, "Optimization study on K value of spatial Clustering", *Journal of System Simulation*, vol. 18, no. 3, pp. 573–576, 2006. DOI: 10.3969/j.issn.1004-731X.2006.03.013.