Data-driven Robust State Estimation Through Off-line Learning and On-line Matching

Yanbo Chen, Hao Chen, Yang Jiao, Jin Ma, and Yuzhang Lin

Abstract-To overcome the shortcomings of model-driven state estimation methods, this paper proposes a data-driven robust state estimation (DDSE) method through off-line learning and on-line matching. At the off-line learning stage, a linear regression equation is presented by clustering historical data from supervisory control and data acquisition (SCADA), which provides a guarantee for solving the over-learning problem of the existing DDSE methods; then a novel robust state estimation method that can be transformed into quadratic programming (QP) models is proposed to obtain the mapping relationship between the measurements and the state variables (MRBMS). The proposed QP models can well solve the problem of collinearity in historical data. Furthermore, the off-line learning stage is greatly accelerated from three aspects including reducing historical categories, constructing tree retrieval structure for known topologies, and using sensitivity analysis when solving QP models. At the on-line matching stage, by quickly matching the current snapshot with the historical ones, the corresponding MRBMS can be obtained, and then the estimation values of the state variables can be obtained. Simulations demonstrate that the proposed DDSE method has obvious advantages in terms of suppressing over-learning problems, dealing with collinearity problems, robustness, and computation efficiency.

Index Terms—Robust state estimation, historical snapshot, offline learning, on-line matching, collinearity.

I. INTRODUCTION

MART grids make the huge amounts of system status data available; however, at the same time, the requirements for the data reliability are higher than ever before [1], [2]. Achieving a comprehensive, real-time, and accurate perception of smart grids is a prerequisite for the intelligent dispatch and control [3]. Therefore, the performance of the The performance of MDSE methods has been greatly improved during the past fifty years. However, in the emerging smart grid environments, MDSE methods have the following shortcomings [8].

1) MDSE methods usually use only the current measurement snapshot while ignoring the massive historical data collected in smart grids. When the gross errors and the parameter errors exist simultaneously, the identification ability of MDSE methods drops [9], [10].

2) The grid parameters themselves have a certain degree of uncertainty [11]. The grid parameters stored in the database of power dispatching control centers may differ from their actual values, which will affect the performance of the real-time MDSE.

3) In smart grids, the power generation and loads often become intermittent and much more uncertain [12], and the topology also changes frequently, resulting in the significant state shifts. This will cause the widely-used weighted least square (WLS) to converge to a local optimal solution without any physical meaning [13]-[16], as WLS usually uses the estimation result of the last snapshot as the initial guess.

4) Using large amounts of historical data, the malicious attackers are more likely to accurately know the state of power grids and launch a malicious data injection attack that cannot be identified by MDSE [17], [18].

Recently, data-driven SE (DDSE) methods have been proposed to address the above problems. In [19], a data-driven scalable approach is proposed to monitor distribution systems by using artificial neural networks (ANNs) for its SE. Reference [16] proposes an architecture to clean historical data and conduct supervised learning, and then the nonlinear relationship between the current measurements and the state vector are estimated by using the historical data. The robust data-driven Kalman filter approaches are used to estimate the rotor angles and the angular velocities in [20], [21]. In [22], a "PaToPaEM" framework is proposed to estimate the topology and parameters simultaneously with the historical data. However, in some power grids, the number of phasor measurement unit (PMU) measurements may not be sufficient to meet the observability requirement of this method.

JOURNAL OF MODERN POWER SYSTEMS AND CLEAN ENERGY

Manuscript received: November 30, 2020; accepted: March 10, 2021. Date of CrossCheck: March 10, 2021. Date of online publication: June 4, 2021.

This work was supported in part by National Natural Science Foundation of China (No. 52077076), and in part by the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (No. LAPS2021-18).

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

Y. Chen (corresponding author), H. Chen, and Y. Jiao are with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Electrical & Electronic Engineering, North China Electric Power University, Beijing 102206, China (e-mail: chenyanbo@ncepu. edu. cn; 1092026803@qq.com; jiaoasdww@163.com).

J. Ma is with the School of Electrical and Information Engineering, University of Sydney, Sydney 2006, Australia (e-mail: j.ma@sydney.edu.au).

Y. Lin is with the Department of Electrical and Computer Engineering, University of Massachusetts, Lowell, MA 01852, USA (e-mail: yuzhang_lin@uml. edu).

DOI: 10.35833/MPCE.2020.000835

state estimation (SE) is crucial in the development of smart grid [4]-[6]. The traditional SE methods require the establishment of measurement equations based on the node admittance matrix, and then construct a mathematical optimal model to estimate the state variables [7]. These methods are called model-driven SE (MDSE) methods.

In [23], [24], data-driven methods are used to enhance the observability of SE for distribution networks. In [25], Gauss-Newton method is provided with a good initial value through a shallow neural network by using the historical or simulation-derived data.

In summary, compared with MDSE methods, the general DDSE methods include the following characteristics.

1) A large amount of historical data is stored in the historical database of SE, and these data include measurement vectors and the corresponding estimation values of state vectors given by historical MDSE methods. DDSE methods try to use the massive historical data to overcome the shortcomings of MDSE methods.

2) DDSE methods do not need to know the measurement equations of the current snapshot exactly like MDSE methods, nor do they need to build an optimal estimation model of the state vector of the current snapshot based on the measurement equations like MDSE methods.

3) DDSE methods generally need to use a large amount of historical data or simulation data as the sample data, and construct a learning model (such as a regression model, etc.) to find the internal mapping relationship between the measurements and the state variables (MRBMS); and then the corresponding MRBMS is used to calculate the state vector of the current snapshot.

Although the development of DDSE methods has been in steady progress, the over-learning problem and the low-learning efficiency are the main shortcomings or even big obstacles [19], which can be attributed to the use of nonlinear MRBMS. Considering the fact that so far, measurements provided by the supervisory control and data acquisition (SCA-DA) system still account for the absolute majority, if a linear MRBMS can be constructed based on historical data of SE and SCADA, it is expected to solve the over-learning problem of the existing DDSE methods, and the computation efficiency will be greatly improved. This motivates us to propose a DDSE method based on the linear regression equation (LRE).

In addition to the above general characteristics, the proposed DDSE method also has the following unique characteristics.

1) It includes off-line learning stage and on-line matching stage. The former performs off-line analysis and processing of historical data, while the latter performs on-line calculation based on the current measurement snapshot.

2) The off-line learning stage only needs to run once, but it can be used multiple times; whereas the on-line matching stage requires online real-time periodic operation.

3) At the off-line learning stage, an LRE suitable for DDSE is formed by clustering the historical data; then based on this LRE, a novel robust estimation method is proposed to filter the historical data to obtain the MRBMS.

4) At the on-line matching stage, by quickly matching the current snapshot with the historical data (QMCH), the corresponding MRBMS can be obtained quickly, and then the estimated state variables of the current snapshot can be obtained quickly based on this MRBMS.

The contributions of this paper mainly include four aspects.

1) By clustering historical data of SE, an LRE is presented, which provides a guarantee for solving the over-learning problem of the existing DDSE methods.

2) A novel robust estimation method that can be transformed into QP models is proposed to obtain the MRBMS at the off-line learning stage, and the proposed QP-based method can solve the collinearity problem in historical data.

3) The computational efficiency of the off-line learning stage is sped up by the reduction of historical categories (RHC), the establishment of a tree search structure for known historical topologies, and the usage of sensitivity algorithm for QP models.

4) A method of QMCH is proposed, thereby greatly improving the computation efficiency of the on-line matching stage, which is beneficial to the on-line application of the proposed DDSE method.

The remainder of this paper is organized as follows. The LRE for the DDSE method is presented in Section II. To obtain the MRBMS, Section III proposes a robust estimation method that can be transformed into QP models based on the proposed LRE. In Section IV, the off-line learning stage is sped up from three aspects. In Section V, the on-line matching stage is presented. The performance of the proposed DDSE method is tested in Section VI. Conclusions are presented in Section VII.

II. FORMULATION OF LRE FOR DDSE METHOD

A. Review of Exact Linear Measurement Equations (ELMEs) for SE

Reference [26] proposes that the nonlinear measurement equations in traditional SE can be exactly linearized by a coordinate transformation on the measurements and the state variables, resulting in the ELMEs as:

$$\tilde{z} = Jy + \omega \tag{1}$$

where $\tilde{z} = [u_i, P_i, Q_i, P_{ij}, Q_{ij}, I_{ij}^2]^T \in \mathbb{R}^m$ is the auxiliary measurement vector, *i* and *j* are the node numbers, $u_i = v_i^2$ is the square of the voltage amplitude measurement, P_i and Q_i are the measurements of active and reactive injection power at node *i*, respectively, P_{ij} and Q_{ij} are the measurements of active and reactive power flow from bus *i* to bus *j*, I_{ii}^2 is the square of the line current magnitude measurement from bus *i* to bus *j*, *m* is the total number of measurements; y = $[u_i, R_{l,l_i}, K_{l,l_i}] \in \mathbb{R}^n$ is the auxiliary state vector, n = N + 2b, N and b are the numbers of nodes and branches, respectively, $R_{l,l_i} = v_{l_i}v_{l_i}\cos\theta_{l_i}$, $K_{l_i} = v_{l_i}v_{l_i}\sin\theta_{l_i}$, l_i and l_j are the terminal buses of the branch l, v_{l_i} and v_{l_i} are the voltage magnitudes of buses l_i and l_j , respectively, θ_{l_i} and θ_{l_i} are the voltage angles of buses l_i and l_j , respectively, $\theta_{l_i l_j} = \theta_{l_i} - \theta_{l_j}$ is the angle difference between buses l_i and l_i ; ω is the *m*-dimensional vector of measurement error with variance \tilde{R} (an $m \times m$ diagonal matrix); and $J \in \mathbb{R}^{m \times n}$ is a constant matrix, in which all elements are determined by the network topology and network parameters. As for the details of (1), please refer to [26]. Equation (1) illustrates that the auxiliary measurement vector is an exact linear function of the auxiliary state vector.

B. Clustering of Historical Data from SCADA

1) Problem Description

The general modeling method of DDSE is shown in Fig. 1. As shown in Fig. 1, to construct a DDSE learning model, a large number of historical snapshots stored in the historical database of SE need to be used as the sample data. These historical snapshots include the historical measurement vectors and the corresponding estimation values of historical state vectors given by the historical MDSE, of which the historical measurement vectors are the sample input, and the estimation values of historical state vectors are the sample output.



Fig. 1. General modeling method of DDSE.

A very important observation that can be made from (1) is that the constant Jacobian matrix of the ELMEs changes only with respect to the changes in topological structure. Therefore, the clustering of the historical snapshots can be based on their topologies, and those snapshots with the same topology should be in the same category. Since the above process only needs to be performed off-line, we call it off-line historical data clustering (OHDC).

Note that when the DDSE learning model is built, the available historical data generally include the historical measurement vectors and the corresponding estimation values of historical state vectors given by historical MDSE. The corresponding historical topology may be known or unknown, and correspondingly, different OHDC methods need to be constructed, which are presented in the following subsections.

2) OHDC When Historical Topologies Are Known

In general, the vast majority (almost all) of different operation modes of the studied power grid are stored in the historical database of SE, i.e., the historical data include almost all the possible topologies with a high probability. As a result, when the historical topologies are known, it is theoretically possible to directly cluster the historical data according to the topologies. However, the drawbacks of direct clustering according to the original topologies are as follows.

1) Since each topology leads to a category, the direct clustering method will result in too many categories caused by too many different topologies for the studied power grid, which will affect the computation efficiency.

2) A large number of historical snapshots are needed to form the datasets for the corresponding topologies, which af-

fects the practicability of the algorithm.

3) In extreme cases, it may be difficult to find multiple historical snapshots with the same topology as the current snapshot in the historical database of SE. To this end, a spanning tree method is proposed to solve the above problems through the following seven steps.

Step 1: for the studied distribution network, assuming that all the branches are put into operation, the network has N nodes and b edges (here, multiple edges connected in parallel between two nodes are treated as one edge). The number of all spanning trees of this network is assumed to be T, the value of which can be determined by Kirchhoff's matrix tree theorem (KMTT) and will be reduced in Section IV.

Step 2: for each spanning tree, create a corresponding structure including three fields. The first field C_{links} stores all the links corresponding to this spanning tree (in ascending order according to link numbers), which is the flag of the corresponding spanning tree. The number of elements in each C_{links} is $l_c = b - N + 1$. The second field $C_{snapshots}$ stores the multiple historical snapshots (the measurement vectors and the corresponding estimation values of state vectors) with the same spanning tree. And the third field $C_{mapping}$ is used to store the corresponding two MRBMSs, which will be introduced in detail below. When each spanning tree is formed, the corresponding links are determined, so the first field C_{links} in the corresponding structure is easy to be determined. The formation of the second field C_{snapshots} requires processing large numbers of historical snapshots, and the processing method is given in Steps 3-7. The formation of the third field $C_{mapping}$ corresponding to each spanning tree will be given in Section III.

Step 3: assuming that there are S historical snapshots available, take out the i^{th} (the initial value of *i* is 1) historical snapshot, including the topology, the measurement vector, and the estimation value of state vector.

Step 4: for the ith historical snapshot, select a spanning tree in the distribution network. If the distribution network is radial, the corresponding spanning tree is itself; if the distribution network is meshed, take one of its spanning trees that has not been selected so far. Store the corresponding links of this spanning tree into a collection (in ascending order according to link numbers) and mark it as $C_{i,meas}$, and then match $C_{i,meas}$ with all C_{links} in Step 2. If $C_{i,meas}$ and one of C_{links} in Step 2 are identical, store the i^{th} historical snapshot into the corresponding collection $C_{snapshots}$ in Step 2. The specific storage method is illustrated as follows: 1) take out all the measurements and the corresponding estimation values of the state vector in the i^{th} historical snapshot; (2) according to the definition in (1), calculate the historical auxiliary measurement vector of all the measurements and the auxiliary state vector associated with the spanning tree, then store them into the corresponding field $C_{snapshots}$ in Step 2.

Step 5: let i = i + 1. If $i \le S$, go to Step 4; else, go to Step 6.

Step 6: if all the second fields $C_{snapshots}$ in Step 2 store at least s historical snapshots (the value of s will be analyzed and given in Section III), then go to Step 7; otherwise, take new historical data and return to Step 3.

Step 7: for each spanning tree, there are s historical snap-

shots in $C_{snapshots}$, which have the same unknown constant Jacobian matrix according to (1).

For each snapshot i (i = 1, 2, ..., s), mark the historical auxiliary measurement vectors in $C_{snapshots}$ as $\tilde{z}_i \in \mathbb{R}^m$ and the historical auxiliary state vectors associated with the spanning tree as $y_i \in \mathbb{R}^n$, n = N + 2(N - 1) = 3N - 2 (for the spanning tree, b is equal to N - 1). Further, the auxiliary measurement vectors and the auxiliary state vectors associated with the spanning tree of all the s historical snapshots are aggregated into the following matrix forms.

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \dots & \boldsymbol{y}_s \end{bmatrix} \in \mathbb{R}^{n \times s}$$
(2)

$$\boldsymbol{Z} = \begin{bmatrix} \tilde{\boldsymbol{z}}_1 & \tilde{\boldsymbol{z}}_2 & \dots & \tilde{\boldsymbol{z}}_s \end{bmatrix} \in \mathbb{R}^{m \times s}$$
(3)

In most cases, multiple historical snapshots with the same topology as the current snapshot are available in the historical database of SE. These historical snapshots and the current snapshot have the same MRBMS. At this time, $Z \in \mathbb{R}^{m \times s}$ and $Y \in \mathbb{R}^{n \times s}$ can be used as the sample input and the sample output to construct the DDSE learning model, respectively. In this circumstance, all the measurements in each historical measurement snapshot are used in the DDSE learning model.

In a few extreme cases, historical snapshots with the same topology as the current snapshot are not available in the historical database of SE. Therefore, there are no historical snapshots with the same MRBMS as the current snapshot. However, we can always find multiple historical snapshots with the same spanning tree as the current snapshot; these historical/current auxiliary measurement vectors associated with the spanning tree have the same MRBMS. At this time, the sample output is still $Y \in \mathbb{R}^{n\times s}$, but the sample input of the DDSE learning model should be historical auxiliary measurement vectors associated with the spanning tree, and the matrix form is formulated as:

$$\boldsymbol{Z}_{cut} = \begin{bmatrix} \tilde{\boldsymbol{z}}_1^- & \tilde{\boldsymbol{z}}_2^- & \dots & \tilde{\boldsymbol{z}}_s^- \end{bmatrix}$$
(4)

where $\tilde{z}_i^- \in \mathbb{R}^{m_{cut}} (m_{cut} \le m)$ is the historical auxiliary measurement vector associated with the spanning tree for the *i*th snapshot; and $Z_{cut} \in \mathbb{R}^{m_{cut} \times s}$.

At the off-line learning stage, two MRBMSs should be formed for each spanning tree. For the first MRBMS, the input and output of the corresponding DDSE learning model are $Z \in \mathbb{R}^{m \times s}$ and $Y \in \mathbb{R}^{n \times s}$, respectively, which are suitable for most cases where the same topologies of historical snapshots as the current snapshot are available. For the second MRBMS, the input and output of the corresponding DDSE learning model are $Z_{cut} \in \mathbb{R}^{m_{cut} \times s}$ and $Y \in \mathbb{R}^{n \times s}$, respectively, which are suitable for a few extreme cases where the same topologies of historical snapshots as the current snapshot are not available. For the convenience of expression, the following analysis only takes the formation of the first MRBMS as an example. By using the same method, it is easy to get the second MRBMS.

Note that these historical auxiliary measurement vectors associated with the spanning tree for the second MRBMS include the measurements of all node voltage amplitudes, power flow measurements on the twigs of the spanning tree, and injection power measurements of those nodes that are not connected to any links.

When the historical topologies are known, the advantages of the proposed OHDC method based on the spanning tree include four aspects.

1) In theory, the number of spanning trees is less than that of all possible original topologies of the studied power grid. Therefore, the total number of modes that need to be processed at the off-line learning stage can be reduced by using the spanning tree method.

2) Any operational topology of the studied power grid is included in all T structures of the field C_{links} , and the corresponding historical snapshots needed are stored in the field $C_{snapshots}$, which lays the data foundation for DDSE.

3) The proposed spanning tree method has no special requirements for the distribution of historical measurements and all the historical measurements can be used in the auxiliary measurement vectors in most cases, whereas the number of auxiliary state variables associated with the spanning tree is n=N+2(N-1)=3N-2, which is smaller than the number of auxiliary state variables corresponding to a mesh network (N+2b). The accuracy of the DDSE model is improved by using the proposed spanning tree method.

4) It should be emphasized that the off-line learning stage of the proposed DDSE method is to learn the MRBMSs (i.e., the mapping matrix H) corresponding to all the different spanning trees when the historical topologies are known. As a result, even if a new topology appears in the current snapshot in extreme cases, that is to say, there are no historical snapshots with the same topology as the current snapshot, we can still find historical snapshots with the same spanning tree as the current snapshot. These historical/current auxiliary measurement vectors associated with the spanning tree have the same MRBMS, thus the proposed DDSE method can still work.

3) OHDC When Historical Topologies Are Unknown

When the historical topologies are unknown, the corresponding topologies can only be inferred according to the measurement vectors.

Taking out any two historical snapshots, the corresponding auxiliary measurement vectors $\tilde{z}_1 \in \mathbb{R}^m$ and $\tilde{z}_2 \in \mathbb{R}^m$ can be calculated according to (1). If \tilde{z}_1 and \tilde{z}_2 have a strong linear correlation, they can be considered to have the same topology [16]. The Spearman's rank correlation coefficient (RCC) in statistics can be used to measure the correlation between the two vectors. The RCC of \tilde{z}_1 and \tilde{z}_2 is calculated as:

$$RCC(\tilde{z}_1, \tilde{z}_2) = 1 - \frac{6\sum_{i=1}^m d_i^2}{m(m^2 - 1)}$$
(5)

where $d_i = \operatorname{rg}(\tilde{z}_{1,i}) - \operatorname{rg}(\tilde{z}_{2,i})$ is the difference between the two ranks of \tilde{z}_1 and \tilde{z}_2 , $\operatorname{rg}(\cdot)$ represents the consecutively ranking number from small (starting from 1) to large, and $\tilde{z}_{1,i}$ and $\tilde{z}_{2,i}$ are the *i*th elements of \tilde{z}_1 and \tilde{z}_2 , respectively; and $RCC(\tilde{z}_1, \tilde{z}_2)$ is the RCC of \tilde{z}_1 and \tilde{z}_2 .

The criterion is that, if $R_{threshold} \leq RCC(\tilde{z}_1, \tilde{z}_2) \leq 1$, then it is considered that \tilde{z}_1 and \tilde{z}_2 have a strong linear correlation, and their corresponding topologies can be considered the same. Obviously, the choice of the threshold $R_{threshold}$ is very

important. In a large number of simulation experiments, we have found that even if $R_{threshold}$ is 0.9, the above method can still be used to correctly identify the historical measurement snapshots with the exact same topology. It should be pointed out that the reason why we use RCC instead of Pearson correlation coefficient (PCC) is that RCC does not require the auxiliary measurement vectors to conform to the normal distribution, and RCC is robust when gross errors exist in historical data.

As an analogue to the method proposed in Section II-B, all historical snapshots can be clustered according to the RCC between any two historical snapshots. Here, the three fields stored in the structure corresponding to each topological structure should be RCC, $C_{snapshots}$, and $C_{mapping}$. With sufficient historical snapshots, theoretically, all the topologies of the studied power grid can be obtained. For each topology, assuming that there are *s* snapshots, the aggregation method of the auxiliary measurement vectors and the auxiliary state vectors is the same as (2)-(4).

C. LREs for DDSE

1) Expression for MRBMS

According to (1), the true value of y must be a linear function of \tilde{z} , so we have

$$y = H\tilde{z} + v \tag{6}$$

where $H \in \mathbb{R}^{n \times m}$ is the unknown mapping matrix from the true value of \tilde{z} to the true value of y; and $v \in \mathbb{R}^n$ is the error vector.

According to (6), Y must also be a linear function of Z.

$$Y = HZ + V \tag{7}$$

where $V \in \mathbb{R}^{n \times s}$ is the error matrix.

According to (1) and (6), as long as the local topologies of different snapshots are the same, these local topologies have the same matrix J and matrix H. This is the reason why different topologies sharing the same spanning tree can be clustered into one category.

Obviously, the unknown matrix H corresponds to the third field $C_{mapping}$ in OHDC. If H can be estimated according to (7) by the given Y and Z, the MRBMS can be obtained. Further, when the current measurement snapshot is given, by matching the current snapshot with the historical snapshots, we can get the MRBMS corresponding to the current snapshot, and then get the estimation value of the state variables, thereby constructing a DDSE method.

2) Number of Required Historical Snapshots

The task of the off-line learning stage of the proposed DDSE method is to estimate H based on (7). Obviously, the most direct method is to use the WLS method. At this time, a unique estimation value of H can be obtained only when the matrix Z^{T} is column full rank. Therefore, for each historical category, it is preferable to ensure that $s(s \ge m)$ historical snapshots are available when the WLS method is used, which is obviously a necessary condition. The necessary and sufficient condition is ensuring that s auxiliary measurement vectors are linearly independent.

In practical systems, we may not be able to obtain so many historical snapshots, and worse, there may be collinearity problems among different historical snapshots, so it is not ideal to use the WLS method to estimate H directly. This issue will be addressed in Section III.

3) Vectorization of Mapping Relationships

Considering that the variables to be solved in mathematical planning are often vectors, two methods can be adopted to transform the matrix H into a vector.

The first method is to stack the elements of the matrix variable H into a vector, then we have

$$\boldsymbol{\Gamma} = \boldsymbol{\Phi}\boldsymbol{\beta} + \boldsymbol{\Xi} \tag{8}$$

where $\Gamma = \operatorname{vec}(Y^{\mathsf{T}})$ and $\Xi = \operatorname{vec}(V^{\mathsf{T}})$ are the column vectors with *F* elements, $F = s \times n$, and $\operatorname{vec}(\cdot)$ represents the vectorization of the matrix by the column vectors; $\boldsymbol{\beta} = \operatorname{vec}(H^{\mathsf{T}})$ is the column vector to be solved with *M* elements, $M = m \times n$; and $\boldsymbol{\Phi} = \operatorname{diag} \{\underbrace{Z^{\mathsf{T}}, Z^{\mathsf{T}}, \dots, Z^{\mathsf{T}}}_{n}\}$ is the block diagonal matrix com-

posed of *n* matrices Z^{T} , $\boldsymbol{\Phi} \in \mathbb{R}^{F \times M}$.

The estimation value of β can be obtained based on (8) by using the historical data, and then β can be compressed into the matrix H. However, in this method, Φ is a diagonal matrix with a very large order, and is highly sparse, thus the memory required in the estimation process might be very large, which may affect the practicability of the algorithm.

The second method is to use the historical snapshots to solve each column of H^{T} . The regression equation for solving the *i*th (*i*=1, 2,..., *n*) column of H^{T} is as follows.

$$\boldsymbol{\Theta} = \boldsymbol{\Psi} \boldsymbol{\alpha} + \boldsymbol{\Omega} \tag{9}$$

where $\boldsymbol{\Psi} = \boldsymbol{Z}^{\mathrm{T}} \in \mathbb{R}^{s \times m}$; and $\boldsymbol{\Theta} = \boldsymbol{Y}_{i}^{\mathrm{T}} \in \mathbb{R}^{s}$, $\boldsymbol{\alpha} = \boldsymbol{H}_{i}^{\mathrm{T}} \in \mathbb{R}^{m}$, and $\boldsymbol{\Omega} = V_{i}^{\mathrm{T}} \in \mathbb{R}^{s}$ are the *i*th columns of $\boldsymbol{Y}^{\mathrm{T}}$, $\boldsymbol{H}^{\mathrm{T}}$, and $\boldsymbol{V}^{\mathrm{T}}$, respectively.

Solving *n* models based on (9) by using the historical snapshots could give each column of H^{T} . Obviously, the computer memory required to solve each model is very small. Although this method needs to solve (9) *n* times, the total off-line calculation takes much less time than the first method, which supports the practical application of the algorithm. Equation (8) or (9) is the LRE for the proposed DDSE method.

Compared with the existing nonlinear regression equations and the approximate LREs of existing DDSE methods, the establishment of the LRE in this paper avoids the over-fitting and ill-conditioned problems in the existing DDSE methods, and lays a foundation for the establishment of new DDSE methods with good robustness.

III. OFF-LINE LEARNING OF MRBMS BY SOLVING QP MODEL

The most important task at the off-line learning stage of the proposed DDSE method is to estimate H based on the LRE, which will be given in this section.

A. Motivation

For each $C_{mapping}$, it is necessary to estimate the value of $\boldsymbol{\alpha} = \boldsymbol{H}_i^{\mathrm{T}}(i=1,2,...,n)$ based on (9). An intuitive idea is using WLS, which is very simple in principle. When the noise conforms to the normal distribution, the WLS method is the optimal estimation; however, it is not a robust method and cannot handle ill-conditioned situations (e.g., there is a collinearity problem in the historical snapshots).

Considering that the least absolute value (LAV) estimation has good robustness [13], and the ridge regression (RR) has good adaptability to the ill-conditioned situations, a new robust estimation method is proposed by combing linear least squares (LS), LAV, and RR. In this paper, we call this new robust method the LSAVRR method.

B. LSAVRR Model for Estimating MRBMSs

Based on (9), the estimation value of α , denoted as $\hat{\alpha}$, can be obtained by solving the following LSAVRR model.

$$\begin{cases} \min_{\boldsymbol{\alpha},t} \sum_{i=1}^{s} \left(t_{i}^{2} + \eta_{1} \left| t_{i} \right| \right) + \eta_{2} \sum_{j=1}^{m} \alpha_{j}^{2} \\ \text{s.t.} \quad t_{i} = \Theta_{i} - \boldsymbol{\Psi}_{i} \boldsymbol{\alpha} \quad i = 1, 2, \dots, s \end{cases}$$
(10)

where Θ_i is the *i*th element of Θ ; Ψ_i is the *i*th row of Ψ ; $\eta_1 \ge 0$ and $\eta_2 \ge 0$ are tuning parameters; and $t \in \mathbb{R}^s$ is residual vector, t_i is the *i*th element of t.

When both η_1 and η_2 are equal to 0, model (10) is the LS method; when η_1 is equal to 0, model (10) is the RR method; when η_1 is sufficiently large, model (10) approaches the LAV method. Based on the extensive simulations, we recommend taking $\eta_1 = 1$ and $\eta_2 = 1 \times 10^{-6}$, respectively. Note that the proposed model (10) is different from the elastic net regression method [27].

C. Equivalent Model of LSAVRR and Solution Method

1) Equivalent Model of LSAVRR

Note that the objective function of the proposed model (10) is non-differentiable, so model (10) cannot be solved directly using the gradient-based method. According to the same method in [14], model (10) can be rewritten as:

$$\begin{aligned} &\min_{\boldsymbol{a},\boldsymbol{u},\boldsymbol{v}} \sum_{i=1}^{s} \left[\left(u_{i} + v_{i} \right)^{2} + \eta_{1} \left(u_{i} + v_{i} \right) \right] + \eta_{2} \sum_{j=1}^{m} \alpha_{j}^{2} \\ \text{s.t. } \boldsymbol{\Theta} - \boldsymbol{\Psi} \boldsymbol{a} - \boldsymbol{u} + \boldsymbol{v} = \boldsymbol{0} \\ &\boldsymbol{u} \ge \boldsymbol{0}, \boldsymbol{v} \ge \boldsymbol{0} \end{aligned}$$
(11)

where $\boldsymbol{u} \in \mathbb{R}^s$ and $\boldsymbol{v} \in \mathbb{R}^s$ are two auxiliary vectors, and u_i and v_i are the *i*th elements of \boldsymbol{u} and \boldsymbol{v} , respectively. 2) Solution Method

2) Solution Method

Model (11) can be further transformed into a standard form of QP model as:

$$\begin{cases} \hat{X}(\hat{\alpha}) = \arg\min\left(\frac{1}{2}X^{\mathsf{T}}QX + X^{\mathsf{T}}c\right) \\ \text{s.t. } AX = \Theta \\ CX \ge 0 \end{cases}$$
(12)

where $X = [u_1, v_1, u_2, v_2, ..., u_s, v_s, \alpha_1, \alpha_2, ..., \alpha_m]^T$ is a column vector with 2s+m elements; $\hat{X}(\hat{a})$ is the estimation value of X; $Q = \text{diag}(\underline{q}_1, \underline{q}_1, ..., \underline{q}_1, q_2)$ is the semi-definite block diago-

nal matrix, $\boldsymbol{q}_1 = \begin{bmatrix} s & 2 \\ 2 & 2 \end{bmatrix}$, \boldsymbol{q}_2 represents the identity matrix (order is *m*) multiplied by $2\eta_2$; $\boldsymbol{c}^T = [\mathbf{1}_{2s}, \mathbf{0}_m]$, $\mathbf{1}_{2s}$ is the 2*s*-dimensional row vector, whose elements are all η_1 , $\mathbf{0}_m$ is the 0-dimensional row vector, whose elements are all 0; $\boldsymbol{A}_i = [\boldsymbol{D}_i, \boldsymbol{\Psi}_i]$ is the *i*th row of the matrix \boldsymbol{A} with *s* rows and 2s + m columns, \boldsymbol{D}_i is a row vector with 2s elements, of which the $(2i - 1)^{\text{th}}$ element is 1, the $(2i)^{\text{th}}$ element is -1, other elements are all 0; $C = [I_{2s}, 0]$ is a matrix with 2s rows and 2s + m columns, I_{2s} is the identity matrix (order is 2s), the other elements in C are all 0.

The standard QP model (12) can be solved by mature commercial software, such as GROUBI to obtain the estimation value $\hat{\alpha}$. Z^{T} and $Y_{i}^{T}(i=1,2,...,n)$ are respectively introduced into model (11), then the estimation values of each column of H can be obtained, thereby obtaining the MRBMS corresponding to each historical category. That is, the third field $C_{mapping}$ is obtained.

IV. Speeding UP Off-line Learning Stage of DDSE Method

The complete off-line learning stage of the proposed DDSE method has been given above. Further research reveals that the off-line learning stage can be accelerated through the following three aspects.

A. RHC

1) When Historical Topologies Are Known

As shown in Section II, when the historical topologies are known, all possible spanning trees of the studied power network have been stored in T fields C_{links} , so the spanning tree of the current snapshot needs to be matched with T fields C_{links} . This may cause the following problems.

1) Although some topologies exist in theory, the probability that they appear in the actual operation of the power network is very low. Considering all possible spanning trees will cause the value of T too large (for example, T is 3909 for IEEE 14-bus system), leading to a very heavy computational load for off-line learning.

2) Likewise, it will also cause too much calculation by considering all spanning trees at the on-line matching stage.

To this end, we propose a specific method for the RHC: (1) calculate the probability of each spanning tree in the actual operation of the power grid; (2) consider only those spanning trees with a relatively high probability of occurrence, so the value of T can be greatly reduced, and the reduced value is set to be T_{cur} .

2) When Historical Topologies Are Unknown

When the historical topologies are unknown, we propose to consider only T_{cut} categories by the RHC, i.e., the RCC values of all categories from large to small are sorted, and then only the first T_{cut} categories are retained.

B. Tree Retrieval Structure for Known Historical Topologies

When historical topologies are known, for T_{cut} spanning trees, a tree retrieval structure is established based on the following steps to improve the efficiency of the on-line matching stage.

Step 1: for each C_{links} , establish a tree retrieval structure by taking $l_c = b - N + 1$ elements (link numbers) in C_{links} as the l_c nodes, where the smallest number in C_{links} is the root node, and each node in the tree structure (TS) corresponds to a link number in C_{links} . Since the elements in C_{links} have been arranged in the ascending order, the child node is always larger than its parent node. At this time, each node has only one child node (except for the leaf node); the number of lay-

ers in the TS is l_c . A total of T_{cut} TSs can be obtained corresponding to all the C_{links} ; let i = 1.

Step 2: if there are TSs with the same value of the i^{th} layers, the i^{th} nodes of these TSs are merged into one node and the number of TSs is reduced.

Step 3: i=i+1; if $i \le l_c$, go to Step 2; otherwise, go to Step 4.

Step 4: denote T_{merge} as the number of TSs; let j = 1.

Step 5: for the j^{th} TS, let k = 2.

Step 6: sort the k^{th} layer of the j^{th} TS in the ascending order according to the value of the k^{th} node.

Step 7: k=k+1; if $k \le l_c$, go to Step 6; otherwise, go to Step 8.

Step 8: j=j+1; if $j \le T_{merge}$, go to Step 5; otherwise, go to Step 9.

Step 9: end.

C. Sensitivity Algorithm for QP Models

To get MRBMSs, the estimation values of the columns of H can be obtained by solving QP models n times. It is apparent that only the values of $\Theta = Y_i^T$ are different in these n QP models; therefore, the sensitivity algorithm for QP models in [28] can be used to improve the efficiency of obtaining H. By adopting the above sensitivity algorithm, the calculation efficiency of the off-line learning stage can be further improved.

V. ON-LINE MATCHING STAGE OF DDSE METHOD

The task of the on-line matching stage of the proposed DDSE method is the QMCH to obtain the corresponding H matrix. That is to say, if the current snapshot and historical snapshots have the same spanning tree (when historical topologies are known) or the same original topology (when historical topologies are unknown), then the current snapshot and historical snapshots have the same H matrix; and further, the original state variables of the current snapshot can be estimated based on H.

A. On-line Matching When Current Topology Is Known

When the current topology is known, the method of QMCH is as follows.

Step 1: choose a spanning tree arbitrarily in the network of the current snapshot; store all links in a collection C_c with an ascending order by link numbers. Note that the open or overhauled branches in the current snapshot should also be considered as links. When selecting the spanning tree, avoid selecting branches that appear in one tree retrieval structure as twigs at the same time.

Step 2: take the first element in C_c , and denote it as $C_c(1)$; find the TS whose root node equals $C_c(1)$, and denote it as TS_c .

Step 3: according to the depth-first search (DFS) algorithm, find a path from the root node to the leaf node in TS_c , and the nodes on this path need to be the same as all the elements in C_c . The C_{links} corresponding to this path stores the historical topology that matches the current topology. If the original topologies and the spanning tree of the historical snapshots are the same as those of the current snapshot, the

first MRBMS in the corresponding $C_{mapping}$ stores the target matrix H; if only the spanning tree of the historical snapshots is the same as that of the current snapshot, and their original topologies are not the same, the second MRBMS in the corresponding $C_{mapping}$ stores the target matrix H.

B. On-line Matching When Current Topology Is Unknown

When the current topology is unknown, the auxiliary measurement vector of the current snapshot can be used to match historical snapshots based on (5). Once the RCC between the historical measurement vectors and the current measurement vector satisfies $0.9 \le RCC \le 1$, the corresponding MRBMS, i.e., the target matrix **H**, is obtained.

C. Estimation Value of State Variables

After *H* is obtained, the estimation value of the auxiliary state vector for the current snapshot, \hat{y}_{c} , can be obtained by:

$$\hat{\boldsymbol{y}}_{c} = \boldsymbol{H}\tilde{\boldsymbol{z}}_{c} \tag{13}$$

where \tilde{z}_c is the auxiliary measurement vector of the current snapshot.

If the current topology is unknown, the power flow can be further obtained and the estimation process ends; if the current topology is known, the estimation value of the original state variables $\hat{\mathbf{x}}_c$ (the voltage amplitudes and angles of all nodes) of the current snapshot can be further obtained [26], [29]-[31]. The structural framework of the proposed DDSE method is shown in Fig. 2.

The advantages of the proposed DDSE method are as follows: (1) the MRBMSs are obtained based on the LRE, which avoids the over-learning problems; 2) the proposed LSAVRR method can well solve the problem of collinearity in historical data, and it also has good robustness; (3) the proposed DDSE method does not require nonlinear iterations, and therefore does not require an initial guess, so that the proposed method has a strong adaptability to the uncertainty of power generation and load in smart grids; (4) the proposed DDSE method does not need to know any network parameters, so the uncertainty of network parameters does not have any impacts on the proposed method; meanwhile, the problems of convergence and leverage gross errors caused by the network parameters also have no effects on the proposed method; (5) the proposed DDSE method can be implemented with both known and unknown topologies, so it is adaptive to the frequent changes in the topology of smart grids; 6 in extreme cases, it may be difficult to find multiple historical snapshots with the same topology as the current snapshot in the historical database of SE, but the proposed DDSE method can still work by using the spanning tree method.

VI. CASE STUDIES

This section tests the performance of the proposed DDSE method on IEEE benchmark systems. In the tests, the load used in the simulation gradually changes from 90% to 110% of the base case. One measurement snapshot is generated every 10 s to mimic the sampling period of SCADA and MDSE runs every 3 minutes to generate the sample data.



Fig. 2. Structural framework of DDSE method.

The generation method of historical sample data is as follows: first calculate the power flow, and then superimpose the normal distribution random errors (the standard deviation is 10^{-3}) on the result of the power flow to simulate the measurements. For the historical SE, the widely-used WLS method is used to obtain estimation values of the state variables, and the largest normal residual (LNR) method is used to identify gross errors. The generated historical data are stored for testing. The tests are performed on an Intel[®] CoreTM i5 PC, with 2.20 GHz processor and 8 GB RAM.

A. Tests on IEEE 4-bus System

The IEEE 4-bus system is firstly used to show the calculation process of the proposed DDSE method in detail.

1) When Current Topology Is Known

The topology and measurement configuration of the IEEE 4-bus system are shown in Fig. 3, where 1-4 represent the node numbers and (1)-(5) represent the branch numbers. The numbers of nodes, branches and links are N=4, b=5 and $l_c=2$, respectively. Here it is assumed that only the measurement vectors and estimation values of state vectors corresponding to this topology are available in the historical database of SE.



Fig. 3. IEEE 4-bus system with N = 4, b = 5, and $l_c = 2$.

1) Off-line learning stage

According to the KMTT, T=8 can be obtained. According to the proposed OHDC method, all the C_{links} can be obtained as: $\{3, 5\}, \{3, 4\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, and \{2, 5\}$. The corresponding tree retriev-

al structures before the RHC are shown as Fig. 4(a). According to the RHC method, there are three categories ($T_{cut} = 3$) for the most common topologies of this grid, and the corresponding C_{links} are {(3), (4)}, {(1), (3)}, and {(2), (4)}. The corresponding tree retrieval structures after the RHC are shown as Fig. 4(b), and they correspond to three spanning trees, and the three $C_{mapping}$ (i.e., six H matrices) corresponding to these three spanning trees need to be solved based on the historical data. Here, we take the number of historical snapshots as 12, i.e., s = 12.

Fig. 4. Tree retrieval structures. (a) Before RHC. (b) After RHC.

For the first MRBMS, the numbers of auxiliary measurements and auxiliary state variables in this test system are m = 12 and n = 10, respectively. Then, the first three $H \in \mathbb{R}^{10 \times 12}$ matrices corresponding to Fig. 4(b) can be obtained based on the LSAVRR method by solving (12).

For the second MRBMS, the number of auxiliary state variables in this test system are n=10; whereas the number of auxiliary measurements corresponding to TS₁, TS₂, and TS₃ in Fig. 4(b) are m=6, m=8, and m=8, respectively. Therefore, the three mapping matrixes corresponding to TS₁, TS₂, and TS₃ in Fig. 4(b) are $H \in \mathbb{R}^{10\times 6}$, $H \in \mathbb{R}^{10\times 8}$, and $H \in \mathbb{R}^{10\times 8}$, respectively. Then, the second three mapping matrices corresponding to Fig. 4(b) can also be obtained based on the LSAVRR method by solving (12).

2) On-line matching stage

Suppose there are two topologies of the current snapshot, as shown in Fig. 3 (Topology A) and Fig. 5 (Topology B), respectively. We will show how to use the on-line matching stage of the proposed DDSE method to estimate the state vector.



When the topology of the current snapshot is the same as Fig. 3 (Topology A), select a spanning tree arbitrary, suppose the twigs of the selected spanning tree are (2), (4), and (5), and the corresponding C_c is $\{(1), (3)\}$. After quickly matching with the tree retrieval structures in Fig. 4(b), it can be found that it corresponds to TS₁, and then the first $H \in \mathbb{R}^{10 \times 12}$ matrix corresponding to TS_1 is the target matrix. All the 12 measurements of the current snapshot can be used at the on-line matching stage. The estimation values of the original state variables (voltage amplitudes and phase angles of all nodes) of the current snapshot can be further obtained easily, as shown in Table I. For comparison, Table I also gives the true values of the original state variables as well as the estimation values given by WLS. It can be seen from Table I that the estimation values obtained by the proposed DDSE method is in good agreement with the true values, thus proving the correctness of the proposed DDSE method.

 TABLE I

 Ture and Estimation Values of State Variables for Topology A

| Node | True value of state variable | Estimation value of state variable | | |
|------|---------------------------------|------------------------------------|-------------------------|--|
| | | WLS | Proposed DDSE | |
| 1 | 1∠0 | 1∠0 | 1∠0 | |
| 2 | $0.9797 \angle -0.0223$ | $0.9797 \angle -0.0223$ | $0.9797 \angle -0.0223$ | |
| 3 | $0.9739 \angle -0.0292$ | $0.9739 \angle -0.0292$ | $0.9739 \angle -0.0292$ | |
| 4 | 1.0200∠0.0245 | 1.0200∠0.0245 | 1.0200∠0.0245 | |

When the topology of the current snapshot is the same as Fig. 5 (Topology B), there is no historical snapshot with the same topology as the current snapshot. According to the proposed QMCH method, a spanning tree can be selected for the current snapshot. Suppose the twigs of the selected spanning tree are (1), (2), and (5), and the corresponding C_c is $\{(3), (4)\}$. After quickly matching with the tree retrieval structures in Fig. 4(b), it can be found that it corresponds to TS₃, and then the second $H \in \mathbb{R}^{10\times 8}$ matrix corresponding to TS₃ is the target matrix. The estimation values of the auxiliary state variables of the current snapshot (denoted as \hat{y}_c) are given by:

$$\hat{\boldsymbol{y}}_{c} = \boldsymbol{H}\tilde{\boldsymbol{z}}_{cut} \tag{14}$$

where $\tilde{z}_{cul} = [u_1, u_4, P_{12}, P_{43}, Q_{12}, Q_{43}, P_1, Q_1]^T$ includes the auxiliary measurements associated with the selected spanning tree.

Since the twigs of the selected spanning tree (1), 2), and (5) do not include (4), P_4 and Q_4 are not associated with this spanning tree, and therefore, P_4 and Q_4 are not included in \tilde{z}_{cut} .

As shown in Table II, the estimation values of the original state variables of the current snapshot can be further obtained. For comparison, Table II also gives the true values of the original state variables as well as and the estimation values given by WLS. It can be observed from Table II that even if there is no historical snapshot with the same topology as the current snapshot, the proposed DDSE method can still obtain high-precision SE results. Note that for this case, the existing DDSE methods [16], [19] cannot work.

TABLE II TRUE AND ESTIMATION VALUES OF STATE VARIABLES FOR TOPOLOGY B

| Node | True value of state variables | Estimation value of state variable | | | |
|------|-------------------------------|------------------------------------|-------------------------|--|--|
| | | WLS | Proposed DDSE | | |
| 1 | 1∠0 | 1∠0 | 1∠0 | | |
| 2 | $0.9824 \angle -0.0170$ | $0.9824 \angle -0.0170$ | $0.9823 \angle -0.0171$ | | |
| 3 | $0.9690 \angle -0.0327$ | $0.9690 \angle -0.0327$ | $0.9689 \angle -0.0328$ | | |
| 4 | 1.0200∠0.0266 | 1.0200∠0.0266 | 1.0199∠0.0266 | | |

2) When Current Topology Is Unknown

When the historical topologies of the IEEE 4-bus system are unknown, the performance of the proposed DDSE method is also tested. It is assumed that the measurement vectors and the estimation values of state vectors given by historical MDSE in the historical database of SE correspond to the three topological operation modes of the IEEE 4-bus system. In the test, 1000 historical measurement snapshots are taken out. The RCCs are calculated and the results show that these historical measurement snapshots can be clustered into 3 categories, and the intra-cluster RCC values for these 3 categories are all between 0.9975 and 1; while the average values of inter-cluster RCC are less than 0.8.

Take ten of the historical auxiliary measurement vectors, and the RCCs among them are shown in Fig. 6. It can be observed from Fig. 6 that ten historical measurement snapshots can be clustered into 3 categories, i.e., \tilde{z}_1 , \tilde{z}_7 , \tilde{z}_8 , and \tilde{z}_{10} belong to the first category; \tilde{z}_2 , \tilde{z}_4 , and \tilde{z}_9 belong to the second category; and \tilde{z}_3 , \tilde{z}_5 , and \tilde{z}_6 belong to the third category. The identification results are completely consistent with the actual topologies, thereby proving the correctness of the clustering using RCC. After using the RCC for on-line matching, the estimation values of the auxiliary state variables and the power flow of the current snapshot can be further obtained, and the error between the estimation results given by the proposed DDSE method and those given by WLS is less than 10^{-4} , thereby proving the correctness of the proposed DDSE method when the historical topologies are unknown.

B. Tests on Other IEEE Benchmark Systems

1) When Historical Topologies Are Known

1) The number of spanning trees before and after RHC

Before and after the RHC, the number of spanning trees in each IEEE benchmark system is shown in Table III. It can be known from Table III that, before the RHC, although the spanning trees can include all possible topological operation states of the power grid, the number of spanning trees is too large, which substantially increases the time of offline learning and on-line matching. After the RHC, the most common topological operation modes of the power grid is retained, which can greatly improve the computation efficiency of the proposed DDSE method.



Fig. 6. RCC correlation matrix of ten historical measurement snapshots.

TABLE III Number of Spanning Trees of Each IEEE Benchmark System Before and After RHC

| System | Number of spanning trees | | | |
|---------------|--------------------------|-----------|--|--|
| System | Before RHC | After RHC | | |
| IEEE 9-bus | 6 | 3 | | |
| IEEE 14-bus | 3909 | 5 | | |
| IEEE 30-bus | 7824000 | 10 | | |
| IEEE 39-bus | 421380 | 10 | | |
| IEEE 57-bus | 61946380490028 | 12 | | |
| IEEE 118-bus | 9223372036854775807 | 20 | | |
| IEEE 300-bus | 10 ²⁰ | 25 | | |
| IEEE 2746-bus | 10^{30} | 100 | | |

2) Estimation accuracy with different numbers of historical snapshots

Obviously, the number of historical snapshots s will affect the estimation accuracy of the matrix H at the off-line learning stage, which, in turn, will affect the estimation accuracy of the proposed DDSE method.

To test the influence of the number of historical snapshots on the estimation accuracy of the proposed DDSE method, we denote Ratio = s/m. Here, the measurement redundancy is taken to be 2.5 for each test system. With different *Ratio* (from 0.1 to 1.2), the mean absolute error of voltage magnitude (denoted as $|d\theta|_m$) and the mean absolute error of phase angle (denoted as $|d\theta|_m$) between the true values and the estimation values of the proposed DDSE method are given in Figs. 7 and 8, respectively. It can be observed from Figs. 7 and 8 that as *Ratio* increases, both $|dv|_m$ and $|d\theta|_m$ gradually decrease; when *Ratio*=1, both $|dv|_m$ and $|d\theta|_m$ obtained by the proposed DDSE method are smaller than 10^{-4} for all the test systems, thereby proving the correctness of the proposed DDSE method. Note that even when *s* is less than *m* (i.e., Ratio < 1), the estimation accuracy obtained by the proposed DDSE method is still acceptable; whereas at this time, the insufficient observability prevents from using WLS for estimating *H*, which demonstrates the good performance of the proposed LSAVRR method.



Fig. 7. $|dv|_m$ with different *Ratio* for each IEEE benchmark system.



Fig. 8. $|d\theta|_m$ with different *Ratio* for each IEEE benchmark system.

3) Estimation accuracy with different measurement redundancy

As we all know, the measurement redundancy affects the estimation accuracy of SE.

To test the influence of the measurement redundancy on the estimation accuracy of the proposed DDSE method, let s=m. With different measurement redundancy from 1.5 to 3, $|dv|_m$ and $|d\theta|_m$ between the true values and the estimation values of the proposed DDSE method are given in Figs. 9 and 10, respectively. It can be seen from Figs. 9 and 10 that as the measurement redundancy increases, both $|dv|_m$ and $|d\theta|_m$ gradually decrease; when the measurement redundancy is larger than 2, both $|dv|_m$ and $|d\theta|_m$ obtained by the proposed DDSE method are smaller than 10^{-4} for all the test systems. The above test results prove the practicability of the proposed DDSE method considering the measurement redundancy of the actual systems.



Fig. 9. $|dv|_m$ with different redundancy for each IEEE benchmark system.



Fig. 10. $|d\theta|_m$ with different redundancy for each IEEE benchmark system.

4) Comparison of estimation accuracy with other SE methods

When s is equal to m (*Ratio* = 1) and the measurement redundancy is 2.5, the estimation accuracies of WLS, the existing DDSE method in [16], and the proposed DDSE method are compared in Table IV. As can be observed from Table IV, among the above three SE methods, the estimation accuracy of WLS is the highest, followed by the proposed DDSE method. The estimation accuracy of the proposed DDSE method can meet the requirements of engineering applications.

 TABLE IV

 COMPARISON OF ESTIMATION ACCURACY WITH DIFFERENT SE METHODS

| Sector | WLS | | DDSE in [16] | | Proposed DDSE | |
|---------------|--------------------|------------------------|--------------------|------------------------|--------------------|------------------------|
| System | $ dv _m$ | $ \mathrm{d}\theta _m$ | $ dv _m$ | $ \mathrm{d}\theta _m$ | $ dv _m$ | $ \mathrm{d}\theta _m$ |
| IEEE 9-bus | 2×10 ⁻⁸ | 1×10^{-8} | 4×10 ⁻⁵ | 1×10 ⁻⁵ | 9×10 ⁻⁶ | 3×10 ⁻⁵ |
| IEEE 14-bus | 2×10 ⁻⁸ | 1×10^{-8} | 3×10 ⁻⁵ | 6×10 ⁻⁵ | 4×10 ⁻⁶ | 1×10^{-5} |
| IEEE 30-bus | 4×10 ⁻⁸ | 1×10^{-8} | 7×10 ⁻⁵ | 9×10 ⁻⁵ | 3×10 ⁻⁵ | 3×10 ⁻⁵ |
| IEEE 39-bus | 7×10 ⁻⁸ | 2×10 ⁻⁸ | 9×10 ⁻⁵ | 9×10 ⁻⁵ | 5×10 ⁻⁵ | 6×10 ⁻⁵ |
| IEEE 57-bus | 6×10 ⁻⁸ | 6×10 ⁻⁸ | 9×10 ⁻⁵ | 8×10^{-5} | 6×10 ⁻⁵ | 4×10 ⁻⁵ |
| IEEE 118-bus | 4×10 ⁻⁸ | 2×10 ⁻⁷ | 6×10 ⁻⁵ | 4×10 ⁻⁵ | 3×10 ⁻⁶ | 2×10 ⁻⁶ |
| IEEE 300-bus | 1×10^{-8} | 3×10 ⁻⁸ | 2×10 ⁻⁶ | 4×10 ⁻⁶ | 2×10 ⁻⁷ | 6×10^{-7} |
| IEEE 2746-bus | 1×10^{-8} | 1×10^{-8} | 1×10 ⁻⁶ | 2×10 ⁻⁶ | 1×10^{-7} | 4×10 ⁻⁷ |

5) Robustness of proposed DDSE method

The robustness of the proposed DDSE method is also tested. For each IEEE benchmark system, the percentage of bad data (PBD) from 0% to 10% in each snapshot is randomly selected, and then added with 10% relative error. At this time, the number of the selected historical measurement snapshots meets s=1.1m (*Ratio*=1.1). In 100 tests, the changes of the average values of $|dv|_m$ and $|d\theta|_m$ with the changes of PBD obtained by the proposed DDSE method are shown in Figs. 11 and 12, respectively. As can be observed from Figs. 11 and 12, with the gradual increase of PBD, the estimation accuracy of the proposed DDSE method decreases slowly; when the PBD in historical data is as high as 10%, the proposed DDSE method still suppresses the gross errors well and obtains highly accurate estimation results.

6) Computation efficiency test

The computation efficiency of the on-line matching stage of the proposed DDSE method directly determines its engineering usability.



Fig. 11. $|dv|_m$ with different PBD for each IEEE benchmark system.



Fig. 12. $|d\theta|_m$ with different PBD for each IEEE benchmark system.

In order to measure the efficiency of the proposed DDSE method, the on-line calculation time of the proposed DDSE method is compared with that of WLS, as shown in Fig. 13. It can be observed from Fig. 13 that the on-line computation efficiency of the proposed DDSE method is much higher than that of WLS, so the proposed DDSE method is very suitable for online applications of large-scale systems.



Fig. 13. Comparison of calculation efficiency.

2) When Historical Topologies Are Unknown

For the IEEE benchmark systems, when historical topologies are unknown, a large number of historical measurement snapshots are clustered based on the RCC. The clustering results show that the historical measurement snapshots have obvious clustering phenomena, and most of the measurement snapshots belong to the most common topological categories. And even if there are gross errors in the current snapshot, the matching method based on the proposed RCC also achieves the correct matching results. This proves the rationality of clustering using the RCC and the necessity of RHC. The final estimation error is less than 10⁻⁴, which proves the correctness of the proposed DDSE method. Due to page limitations, the specific results of tests are omitted here.

VII. CONCLUSION

Aiming at resolving the shortcomings of the traditional MDSE methods, this paper proposes a DDSE method which includes off-line learning stage and on-line matching stage. The off-line learning stage targets to cluster historical data and develop the linear MRBMS; while the on-line matching stage obtains the current MRBMS by QMCH, and further quickly obtains the estimation values of the state variables of the current snapshot. The proposed DDSE method does not need to know the parameters of the network, and has good robustness and very high computation efficiency, making it very suitable for the on-line applications of large-scale systems.

In low-voltage distribution networks, the number of measurements is very limited (often not enough to ensure observability), and the topology information is difficult to obtain accurately. Next, we will study the DDSE method for lowvoltage distribution networks. Also, the proposed DDSE method can be extended to the integrated energy systems (IESs) so as to realize the comprehensive, real-time and accurate perception of IES in uncertainty circumstances.

REFERENCES

- Nature. (2008, Sept.). Big data (specials). [Online]. Available: http:// www.nature.com/news/specials/bigdata/index.html
- [2] R. Qiu and P. Antonik, Smart Grid and Big Data. Hoboken: Wiley, 2014.
- [3] X. He, Q. Ai, R. C. Qiu *et al.*, "A big data architecture design for smart grids based on random matrix theory," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674-686, Mar. 2017.
- [4] F. C. Schweppe and J. Wildes, "Power system static state estimation, part I: exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120-125, Jan. 1970.
- [5] F. C. Schweppe and D. B. Rom, "Power system static state estimation, part II: approximate model," *IEEE Transactions on Power Apparatus* and Systems, vol. PAS-89, no. 1, pp. 125-130, Jan. 1970.
- [6] F. C. Schweppe, "Power system static state estimation, part III: implementation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 130-135, Jan. 1970.
- [7] A. Abur and A. G. Expósito, Power System State Estimation: Theory and Implementation, New York: Marcel Dekker, 2004, pp. 157-184.
- [8] K. Dehghanpour, Z. Wang, and J. Wang, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2312-2322, Mar. 2019.
- [9] Y. Chen, F. Liu, G. He et al., "A Seidel-type recursive Bayesian approach and its applications to power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1710-1711, Aug. 2012.
- [10] Y. Chen, F. Liu, S. Mei et al., "An improved recursive Bayesian approach for transformer tap position estimation," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2830-2841, Aug. 2013.
- [11] G. Sivanagaraju, S. Chakrabarti, S. C. Srivastava *et al.*, "Uncertainty in transmission line parameters: estimation and impact on line current differential protection," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 6, pp. 1496-1504, Jun. 2014.
- [12] Y. Xiang, J. Liu, Y. Liu *et al.*, "Robust energy management of microgrid with uncertain renewable generation and load," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1034-1043, Mar. 2016.
 [13] Y. Chen, F. Liu, S. Mei *et al.*, "A robust WLAV state estimation using
- [13] Y. Chen, F. Liu, S. Mei *et al.*, "A robust WLAV state estimation using optimal transformations," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 2190-2191, Jul. 2015.
- [14] Y. Chen, J. Ma, P. Zhang et al., "Robust state estimator based on maximum exponential absolute value," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1537-1544, Jul. 2017.
- [15] Y. Chen, Z. Zhang, H. Fang *et al.*, "Generalised-fast decoupled state estimator," *IET Generation, Transmission & Distribution*, vol. 12, no. 22, pp. 5928-5938, Sept. 2018.
- [16] Y. Weng, R. Negi, C. Faloutsos et al., "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no.

4, pp. 1956-1967, Jul. 2017.

- [17] J. Kim, L. Tong, R. J. Thomas *et al.*, "Subspace methods for data attack on state estimation: a data driven approach," *IEEE Transactions* on Signal Processing, vol. 63, no. 5, pp. 1102-1114, Mar. 2015.
- [18] J. Zhang, Z. Chu, L. Sankar *et al.*, "Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems?" *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4775-4786, Sept. 2018.
- [19] M. Ferdowsi, A. Benigni, and A. Löwen, "a scalable data-driven monitoring approach for distribution systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 5, pp. 1292-1305, May 2015.
- [20] M. Netto and L. Mili, "A robust data-driven koopman kalman filter for power systems dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7228-7237, Nov. 2018.
- [21] W. S. Rosenthal, A. M. Tartakovsky, and Z. Huang, "Ensemble kalman filter for dynamic state estimation of power grids stochastically driven by time-correlated mechanical input power," *IEEE Transactions* on *Power Systems*, vol. 33, no. 4, pp. 3701-3710, Jul. 2018.
- [22] J. Yu, Y. Weng, and R. Rajagopal, "PaToPaEM: a data-driven parameter and topology joint estimation framework for time-varying system in distribution grids," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 1682-1692, May 2019.
- [23] K. Dehghanpour, Y. Yuan, Z. Wang *et al.*, "A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 5942-5951, Nov. 2019.
- [24] Y. Yuan, K. Dehghanpour, and F. Bu, "A multi-timescale data-driven approach to enhance distribution system observability," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3168-3177, Jul. 2019.
- [25] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learningbased optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796-4805, Nov. 2019.
- [26] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458-1459, Aug. 2006.
- [27] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society*, vol. 67, no. 5, pp. 301-320, Mar. 2005.
- [28] A. G. Hadigheh, "Sensitivity analysis in convex quadratic optimization: simultaneous perturbation of the objective and right-hand-side vectors," *Algorithmic Operations Research: Series B*, vol. 2, pp. 94-111, May 2007.
- [29] C. Gomez-Quiles, A. Villa Jaen, and A. Gomez-Exposito, "A factorized approach to WLS state estimation," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1724-1732, Aug. 2011.
- [30] Y. Chen, J. Ma, F. Liu et al., "A bilinear robust state estimator," International Transactions on Electrical Energy Systems, vol. 26, no. 7, pp. 1476-1492, Jul. 2016.
- [31] Y. Chen, Y. Yao, and Y. Zhang, "A robust state estimation method based on SOCP for integrated electricity-heat system," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 810-820, Jan. 2021.

Yanbo Chen received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007, the M.S. degree in electrical engineering from China Electric Power Research Institute, Beijing, China, in 2010, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013. He is currently an Associate Professor of North China Electric Power University, Beijing, China. He serves as the Associate Editor of IET Generation, Transmission & Distribution and IET Renewable Power Generation. His research interests include state estimation and power system analysis and control.

Hao Chen received the B.S. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2018, where he is currently studying for the M.S. degree. His research interests include state estimation and power system analysis and control.

Yang Jiao received the B.S. degree in electrical engineering from Shanxi University, Taiyuan, China, in 2019. He is currently pursuing the M.S. degree at North China Electric Power University, Beijing, China. His research interests include state estimation and power system analysis and control.

Jin Ma received the B.S. and M.S. degrees in electrical engineering from

Zhejiang University, Hangzhou, China, in 1997 and 2000, respectively, and the Ph. D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2004. He is currently with the School of Electrical and Information Engineering, University of Sydney, Sydney, Australia. He is the Member of CIGRE W.G. C4.605 "Modeling and aggregation of loads in flexible power networks" and the Corresponding Member of CIGRE Joint Workgroup C4-C6/CIRED "Modeling and dynamic performance of inverter based generation in power system transmission and distribution studies". He is a Registered Chartered Engineer. His research interests include load modeling, nonlinear control system, dynamic power system, and power system economics.

Yuzhang Lin received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2012 and 2014, respectively, and the Ph.D. degree in Electrical Engineering from Northeastern University, Boston, USA, in 2018. He is currently an Assistant Professor of Department of Electrical and Computer Engineering, University of Massachusetts, Lowell, USA. His research interests include state estimation and power system analysis and control.