A Hybrid-Learning Algorithm for Online Dynamic State Estimation in Multimachine Power Systems

Guanyu Tian, *Student Member, IEEE*, Qun Zhou, *Member, IEEE*, Rahul Birari, Junjian Qi, *Senior Member, IEEE*, and Zhihua Qu, *Fellow, IEEE*

***Abstract*— With the increasing penetration of distributed generators in the smart grids, having knowledge of rapid real-time electromechanical dynamic states has become crucial to system stability control. Conventional Supervisory Control and Data Acquisition (SCADA)-based dynamic state estimation (DSE) techniques are limited by the slow sampling rates, while the emerging phasor measurement units (PMUs) technology enables rapid real-time measurements at network nodes. Using generator bus terminal voltages, we propose a hybrid-learning DSE (HL-DSE) algorithm to estimate the synchronous machine rotor angle and speed in real time. The HL-DSE takes the power system model into account and trains neuroestimators with real-time data in an online manner. Compared with traditional DSE methods, the HL-DSE overcomes limitations by using a data-driven approach in conjunction with the physical power system model. The time efficiency, accuracy, convergence, and robustness of the proposed algorithm are tested under noises and fault conditions in both small- and large-scale test systems. Simulation results show that the proposed HL-DSE is much more computationally efficient than widely used Kalman filter (KF)-based methods while maintaining comparable accuracy and robustness. In particular, HL-DSE is over 100 times faster than square-root unscented KF (SR-UKF) and 80 times faster than extended KF (EKF). The advantages and challenges of the HL-DSE are also discussed.**

***Index Terms*— Dynamic state estimation (DSE), hybrid learning, neural networks, phasor measurement units (PMUs), synchronous generators.**

I. Introduction

THE traditional electric power grid has been dependent on synchronous generators. In recent years, distributed generators (DGs), such as solar PV, wind turbines, and storage devices, are penetrating the power grid rapidly [1]. While providing the system with benefits of clean and inexpensive energy, DGs also come with extra uncertainties. Having accurate knowledge of system state dynamics in real time is vital to reliable operations of the power grid [2]. Rotor angles and rotational frequencies of synchronous generators are the most important variables for transient stability and control.

Traditionally, Supervisory Control and Data Acquisition (SCADA)-based approaches are unable to capture the rapidly changing electromechanical dynamics of the system due to the slow sampling rate, typically 0.25–0.5 Hz. With the advent of phasor measurement units (PMUs) in tandem with wide-area measurement system (WAMS), high-frequency real-time sampled synchronized measurement data from a wide area system are available. Measurements acquired from PMUs can be phasors of voltage, current, and power. Measurement data are transmitted in time-stamped packets to the control center and processed in a synchronous manner in order to avail the dynamic states of the system. The sampling rate of PMUs is much higher than traditional SCADA systems, normally ranging from 30 to 120 Hz. With the faster measurements, estimation of power systems dynamic states becomes feasible.

Dynamic state estimation (DSE) plays an important role in achieving real-time wide-area monitoring, protection, and control of power systems [3]. For example, the estimated sates are used for monitoring of oscillations and tuning of power system stabilizers to suppress any detected oscillations [4]. Using the estimated dynamic states, model-free transient stability monitoring can be achieved [5]. Besides, the estimated dynamic states can be utilized to initiate effective generator out-of-step protection and control generator excitation systems [6], [7]. DSE could also greatly facilitates economic dispatch considering generator dynamics [8]–[10].

A. Literature Review

The Kalman filter (KF) is a DSE method that has been widely used in various engineering fields, including power engineering, transportation, and aerospace engineering. It is a two-stage estimation process commonly used for linear systems. In the first stage, KF predicts the state vector and covariance matrix based on the previous estimation of system dynamics. The prediction in the first stage is called prior estimation. In the second stage, the prior estimation is updated based on measurement residuals to obtain a posterior estimation. The correction coefficient of the measurement residual is the Kalman gain, which optimally minimizes estimation covariance [11], [12].

The classic KF is fast and accurate in linear systems but cannot be directly applied to complex nonlinear systems, such

Manuscript received July 1, 2018; revised December 31, 2018, May 27, 2019, and September 17, 2019; accepted January 11, 2020. This work was supported by Leidos. *(Corresponding author: Qun Zhou.)*

Guanyu Tian, Qun Zhou, Junjian Qi, and Zhihua Qu are with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL 32816 USA (e-mail: tiang@knights.ucf.edu; qun.zhou@ucf.edu; junjian.qi@ucf.edu; qu@ucf.edu).

Rahul Birari is with Software Motor Company, Sunnyvale, CA 98223 USA (e-mail: rahulbirari91@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2020.2968486

as power systems. To implement KF in nonlinear systems, extended KF (EKF) methods are introduced [12], [13]. In the prediction stage of EKF, prior estimation is built upon the nonlinear state transition function $f(\cdot)$, and the covariance matrix is estimated with F_k, the Jacobian matrix of $f(\cdot)$. In the update stage, the Jacobian matrix H_k of the measurement model $h(\cdot)$ is adopted. In this way, the nonlinear system state estimation problem is transformed to a linear estimation by linearizing the model at the estimation point. Many modifications have been introduced based on EKF to improve its performance [14], [15].

While different versions of EKFs are capable of solving some nonlinear DSE problems, estimation accuracy decreases with the increase of system nonlinearity [16]. The deteriorated accuracy is due to approximation errors between the linearized and actual models. To address this, unscented KF (UKF) is introduced to avoid linear approximation and improve accuracy on highly nonlinear systems [17], where the nonlinearity is captured by unscented transformation (UT). UT requires generating a set of sigma points that are symmetrically distributed near the current estimation point. Each sigma point has weights on mean and covariance, respectively. Accordingly, the estimated mean and covariance are the weighted average of all sigma points. The values of sigma points are determined by the square root of the covariance matrix, which needs to be positive semidefinite to maintain numerical stability.

A square-root UKF (SR-UKF) method is modified from the classic UKF with better numerical stability. It avoids the semidefinite constraint of the covariance matrix by directly estimating the square root of the covariance matrix for generating sigma points [18]. The simulation results show that it has both the lowest estimation mean and standard deviations of errors among many state-of-the-art DSE methods.

However, UKF-based methods require computational time to generate sigma points, limiting real-time applications on large-scale systems. The execution of DSE algorithms needs to be faster than the sampling rate of measurements to keep in track with the real-time system dynamics. In other words, state estimation is time-sensitive, and hence, delays should be minimized to obtain meaningful results.

B. Proposed Method

In recent years, with the development of machine learning techniques, many engineering problems are seeking data-driven approaches. Exclusively data-driven approaches may lead to incorrect decisions as they do not incorporate useful information from first-principle-based physical models. In contrast, a hybrid-learning model that combines real-time data and physical models is a promising solution to overcoming the limitations [19]. Therefore, this article proposes a hybrid-learning DSE (HL-DSE) algorithm that takes into account the power system model and estimates dynamic states in real time.

The HL-DSE algorithm proposed in this article is an online adaptive estimation method that uses artificial neural networks (ANNs) as the estimator. ANNs, as one of the most advanced machine learning techniques, have been implemented in different areas of power system studies, such as system stability control, fault detection, energy forecasting, and state estimation [20]–[22]. ANNs have been shown to have powerful universal function approximation capability, which is much needed in power system DSE. Efficient algorithms can be designed to untap its potential in accurately estimating states of highly nonlinear systems. ANN-aided EKFs and UKFs have been introduced [23]. Neural networks are integrated as a part of the estimator to capture the system nonlinearity, and the results were shown to be more accurate and robust. However, most of them rely on off-line supervised training to tune network parameters, which limits the application of real-time DSE.

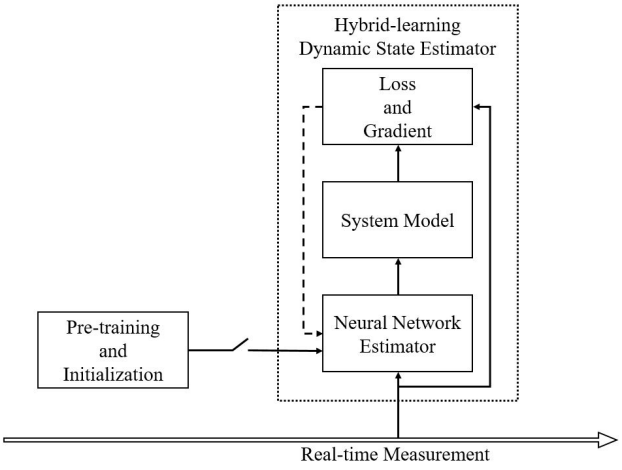


Fig. 1. Overall structure of the hybrid-learning dynamic state estimator.

The HL-DSE proposed in this article considers the power system dynamic model and measurement model and online trains the neuroestimator to minimize the error between estimated measurements and actual measurements from PMUs. Fig. 1 depicts the proposed HL-DSE algorithm. The hybrid-learning structure can be viewed as an autoencoder with the decoder replaced by the physical system model. The neural network estimator is initialized during the steady state. During system dynamics, the neural network learns the system model and converts the measurements to estimated states that are then fed back to the system model to generate estimated measurements. The losses are calculated between the actual and estimated measurements to evaluate and update the neuroestimator. Compared with KF-based methods, this hybrid-learning algorithm combines the advantages of data-driven and model-based approaches. It is executed very fast in real time while achieving comparable accuracy with KF-based methods in large-scale system DSE.

The contributions of this article are threefold.

1) *Neural networks are applied in a novel and critical application in the smart grid:* As the wide-area monitoring system is gradually adopted, close monitoring of generator dynamics is the key to detecting anomalies and instability in the smart grid. A fast online estimation algorithm is greatly needed. To the best of our knowledge, this is the first work that applies neural networks to dynamically track generator states in emergency conditions.

2) *An innovative hybrid dual-net architecture is designed for DSE:* In particular, the two networks are tailored for the second-order power system dynamics, representing the rotor angle and speed. The hybrid-learning structure is both data-driven and first-principle-based. The architecture leverages real-time data and power system domain knowledge to achieve good performance.
3) *Proposed algorithm substantially improves the computational efficiency, making online estimation possible:* DSE is a time-critical application that requires fast runtime performance corresponding to the high sampling rate of PMU data. Using backpropagation with stochastic gradient descent, the proposed HL-DSE is able to achieve high computational efficiency while maintaining comparable accuracy with KF-based methods.

This article is organized as follows. Section II briefly introduces the DSE in power systems and the related models. Section III presents the proposed HL-DSE algorithm. In Section IV, the proposed method is tested on the WSCC three-machine nine-bus system and the NPCC 48-machine 140-bus system. Numerical results are compared with state-of-the-art EKF, UKF, and SR-UKF. Section V discusses challenges and future research directions. Section VI provides concluding remarks.

II. Power System Dynamic State Estimation

The goal of power system DSE is to estimate dynamic states of rotor angle and speed of synchronous generators based on the dynamic model and measurement model. The formulation of a general nonlinear power system dynamic model is given in the following [24]:

$$\begin{cases} x_k = f(x_{k-1}, u_k) + q_k \\ z_k = h(x_k, u_k) + r_k \end{cases} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state vector at time k that contains all the state variables of synchronous generators, and $z_k \in \mathbb{R}^m$ is the measurement vector that contains voltage phasors from PMUs. $q_k \sim \mathcal{N}(0, Q)$ is the Gaussian processing noise, and $r_k \sim \mathcal{N}(0, R)$ is the Gaussian measurement noise, where Q and R are the corresponding covariance matrices. Two consecutive discrete time steps are denoted by k and $k-1$. The nonlinear transition function of state variables is denoted by $f(\cdot)$, while the measurement model is denoted by $h(\cdot)$ [25]. In power system DSE, the transition function is the swing model of synchronous machines, and the measurement model of PMUs is represented by power flow equations.

A. State Variable Transition Model

The state transition function in power system DSE is the swing model of synchronous machines. The classic generator model is expressed by

$$\begin{cases} \dot{\delta}_{i,k} = \omega_{i,k} - \omega_0 \\ \dot{\omega}_{i,k} = \dfrac{1}{2H_i}(P_{mi,k} - P_{ei,k}) \end{cases} \tag{2}$$

where i is the generator index, and δ and ω represent the machine angle and machine speed, respectively. ω_0 is the rated speed that equals $2\pi f_0$, where f_0 is the rated frequency. H is the inertia constant of the generator i, and P_m and P_e are the mechanical power and the electric power, respectively. The formulation of P_e is [26], [27]

$$P_{ei} = E_i^2 G_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{g} E_i E_j (G_{ij}\cos(\delta_i - \delta_j) + B_{ij}\sin(\delta_i - \delta_j)) \tag{3}$$

where G and B are the real and imaginary parts of the admittance matrix of a reduced network that only has internal generator buses. E is the generator internal voltage.

$\dot{\delta}$ and $\dot{\omega}$ in (2) represent the derivative of machine angle and machine speed in the continuous-time domain. In the discrete-time domain, the incremental angle and speed at each time step are discretized as follows:

$$\begin{cases} \Delta\delta_{i,k} = \Delta t \times \dot{\delta}_{i,k} \\ \Delta\omega_{i,k} = \Delta t \times \dot{\omega}_{i,k}. \end{cases} \tag{4}$$

The discretized transition function of state variables is formulated in the following equation:

$$f(x_k, u_k) = \begin{cases} \delta_{i,k} = \delta_{i,k-1} + \Delta\delta_{i,k} \\ \omega_{i,k} = \omega_{i,k-1} + \Delta\omega_{i,k}. \end{cases} \tag{5}$$

The state vector x and the input vector u are defined in (6) and (7), where the state variables are machine angles and speed. The inputs are mechanic power and generator internal voltages

$$x = [\delta^T \quad \omega^T]^T \tag{6}$$

$$u = \left[P_m^T \quad E^T\right]^T. \tag{7}$$

B. Measurement Model

The measurement model of voltage phasors for the multimachine power system is derived from the following equation [28]:

$$\mathbf{Y_{exp}V} = \begin{bmatrix} Y_{\mathrm{GG}} & Y_{\mathrm{GL}} \\ Y_{\mathrm{LG}} & Y_{\mathrm{LL}} \end{bmatrix} \begin{bmatrix} E\angle\delta \\ V\angle\theta \end{bmatrix} = \begin{bmatrix} I_G\angle\delta \\ 0 \end{bmatrix} \tag{8}$$

where $\mathbf{Y_{exp}}$ is the expanded system admittance matrix consisting of the admittance matrix between the generators (Y_{GG}), the admittance matrix between the generators and loads (Y_{GL} and Y_{LG}), and the admittance matrix between the loads (Y_{LL}). $\mathbf{V}$ is the expanded voltage vector, consisting of $E\angle\delta$ and $V\angle\theta$, which are the vectors of generator internal voltages and bus voltages, respectively. The product of $\mathbf{Y_{exp}}$ and $\mathbf{V}$ is the current injection of the corresponding buses. $I_G\angle\delta$ denotes the generator current injection. Since the current injection of loads is zero, the lower part of the current injection vector is zero, and the upper part is the generator current injection. Expanding (8), the relation between the generator internal voltages and bus voltages is expressed by

$$V\angle\theta = (-Y_{\mathrm{LL}})^{-1} Y_{\mathrm{LG}} E\angle\delta = R_v E\angle\delta \tag{9}$$

where the R_v matrix is the voltage reconstruction matrix, derived from Y_{LL} and Y_{LG}.

Note that the measured voltages are phasors with real and imaginary parts. Relating to (1), the measurement vector z consists of two parts as follows:

$$z = \begin{bmatrix} V_{re}^T & V_{\text{imag}}^T \end{bmatrix}^T. \tag{10}$$

III. Hybrid-Learning Dynamic State Estimation

In this section, the HL-DSE algorithm is proposed. This HL-DSE method incorporates the power system dynamic model into the data-driven neural network to mimic the dynamic responses of a power system and estimate the states during system transients. This is an online algorithm that can be used in real time with very fast execution time while achieving accurate estimates of system states.

A. Neural Network Structure

Neural networks have been widely known for its universal function approximation capability. This article uses three-layer feedforward neural networks containing the input layer, the hidden layer, and the output layer. This structure is selected due to its computational simplicity and fast runtime performance, which meets the time-critical requirement of DSE.

In the proposed HL-DSE, the inputs are machine terminal voltage phasors measured by PMUs. The hidden layer extracts additional information from the inputs. The output layer has the same dimension with the training target, and the error can be calculated from a specified loss function. The design of the loss function is mostly related to the learning task. For instance, the loss function for classification problems is the cross entropy, while the loss function for regression is the 2-norm distance [29]. In the context of DSE, the outputs are specifically designed in a way to mimic power system dynamics, which is detailed in III-B.

The training of an ANN involves updating weights and biases by backpropagating the error from the loss function. The backpropagation algorithm is commonly used for off-line training in conjunction with gradient descent. This is also called batch training, and the weights are updated in an off-line manner using the batch gradient. Nevertheless, in our application, backpropagation is used with stochastic gradient descent, which trains the neural network in an online manner. At each time step, the loss is calculated based on the measurement received. Then, the loss is backpropagated, and the weights are updated. This process is repeated sequentially for every new data point. The backpropagation with stochastic gradient descent is introduced in [30] and [31]. The proposed HL-DSE algorithm using this technique is suitable for real-time state estimation.

B. HL-DSE Architecture

In the proposed HL-DSE, we specifically designed a dual-neural-net architecture to complete the task of DSE. The design is inspired by the observation that only one system state δ is reflected in the measurement model in (9), while the other state ω is only linked to δ without entering the measurement model. The proposed dual neural nets can effectively capture the unique system dynamics.

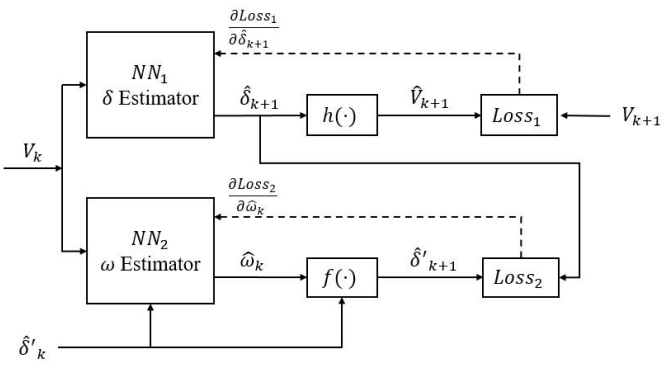


Fig. 2. HL-DSE architecture.

Fig. 2 shows the hybrid-learning architecture of the coordination between the neural nets and system models. Solid lines indicate feedforward paths, and dashed lines indicate backpropagation paths. δ and ω are estimated by two neural nets. Specifically, NN_1 is intended to learn the measurement model $h(\cdot)$ and estimate δ, while NN_2 is to learn the dynamic model $f(\cdot)$ and estimate ω.

Note that the hybrid dual-net architecture represents prediction and estimation, which appropriately mimics power system dynamics. The observed voltage measurements serve as inputs to both neural nets. NN_1 predicts δ at time $k+1$, while NN_2 estimates ω at time k. The objective for NN_1 is straightforward, minimizing the difference between the predicted and actual terminal voltages

$$\text{Min} \quad \text{Loss}_1 = \frac{1}{2}\sum_{i=1}^{n}(\hat{V}_{i,k+1} - V_{i,k+1})^2. \tag{11}$$

The resulting $\hat{\delta}_{k+1}$ from NN_1 can be viewed as the posterior estimate of machine angles.

To evaluate the effectiveness of the second neural net NN_2, we derive prior estimate of $\hat{\delta}'_{k+1}$ from the estimated $\hat{\omega}_k$ in the dynamic model $f(\cdot)$ and then compare the difference between posterior estimate $\hat{\delta}_{k+1}$ and prior estimate $\hat{\delta}'_{k+1}$. Therefore, NN_2 has the following objective:

$$\text{Min} \quad \text{Loss}_2 = \frac{1}{2}\sum_{i=1}^{n}\left(\hat{\delta}'_{k+1} - \hat{\delta}_{k+1}\right)^2. \tag{12}$$

By implementing this architecture, the posterior estimate $\hat{\delta}$ is able to converge to the real δ iteratively by minimizing measurement differences. At the same time, $\hat{\omega}$ is able to converge as well, as any deviations from real w_k would increase the loss function between prior $\hat{\delta}'_{k+1}$ and posterior $\hat{\delta}_{k+1}$. The gradients of the two estimators are

$$\begin{cases} \nabla\hat{\delta}_{k+1} = \dfrac{\partial \text{Loss}_1}{\partial \hat{V}_{k+1}}\dfrac{\partial \hat{V}_{k+1}}{\partial \hat{\delta}_{k+1}} \\ \nabla\hat{\omega}_k = \dfrac{\partial \text{Loss}_2}{\partial \hat{\delta}'_{k+1}}\dfrac{\partial \hat{\delta}'_{k+1}}{\partial \hat{\omega}_k}. \end{cases} \tag{13}$$

The detailed HL-DSE is given in Algorithm 1, where W_{init} and b_{init} are the initial NN parameters, continuously trained in the steady state. NN$(\cdot)$ represents the feedforward of the estimator. BP$(\cdot)$ represents the backpropagation of the NN.

Algorithm 1 HL-DSE

begin

note: inputs and outputs are in vector form

INITIALIZATION

$W_{NN_1}, b_{NN_1} \longleftarrow W_{1,init}$

$W_{NN_2}, b_{NN_2} \longleftarrow W_{2,init}$

$\hat{\delta}_1 = \delta_1$

$\hat{\omega}_1 = \omega_1$

while $k >= 1$ **do**

STEP-1 (ESTIMATE STATES)

$\hat{\delta}_{k+1} = NN_1(V_k)$

$\hat{\omega}_k \ = NN_2(V_k, \hat{\delta}_k)$

STEP-2 (PREDICT NEXT STATES)

$\hat{\delta}'_{k+1} = f(\hat{\delta}_k, \hat{\omega}_k)$

STEP-3 (ESTIMATE MEASUREMENT)

$\hat{V}_{k+1} = h(\hat{\delta}_{k+1})$

STEP-4 (CALCULATE LOSSES)

$loss_1 = \frac{1}{2}\sum_{i=1}^{n}(\hat{V}_{i,k+1} - V_{i,k+1})^2$

$loss_2 = \frac{1}{2}\sum_{i=1}^{n}(\hat{\delta}'_{k+1} - \hat{\delta}_{k+1})^2$

STEP-5 (CALCULATE GRADIENTS)

$\nabla\hat{\delta}_{k+1} = BP(loss_1)$

$\nabla\hat{\omega}_k = BP(loss_2)$

STEP-6 (UPDATE NERUAL NETWORK)

$W_{NN_1} = W_{NN_1} - \eta_1 \times \nabla W_{NN_1}$

$W_{NN_2} = W_{NN_2} - \eta_2 \times \nabla W_{NN_2}$

$b_{NN_1} = b_{NN_1} - \eta_1 \times \nabla b_{NN_1}$

$b_{NN_2} = b_{NN_2} - \eta_2 \times \nabla b_{NN_2}$

$k \longleftarrow k + 1$

η is the learning rate. The estimation at each time step contains multiple steps: estimate the states based on measurements, predict the next step machine angle and estimated measurements, calculate losses and gradients, and update the neural network estimator.

C. Neural Network Initialization

The initialization of the neural network is completed in the steady state before the fault transients. To further illustrate the initialization process, an example is given in Fig. 3, which shows a typical time-domain rotor speed curve in two phases. Phase I (0–5 s) is the steady-state phase, where the electric torque equals to the mechanic torque. As a result, the states of generators and the PMU measurements are constant. In this phase, the neural networks are first randomly initialized and continuously trained using the steady-state PMU measurements. The neural networks are able to quickly converge to the steady state, and hence, the initialization is considered as complete. Phase II (5–10 s) is the dynamic phase, where the HL-DSE is deployed to track the changing rotor angle and speed. The weights and biases obtained in Phase I are used as initial values for the neural network in Phase II.

D. Adaptive Learning Rate

Due to the nonlinearity and time-variant system dynamics, a fixed learning rate may not best suit the neural networks.

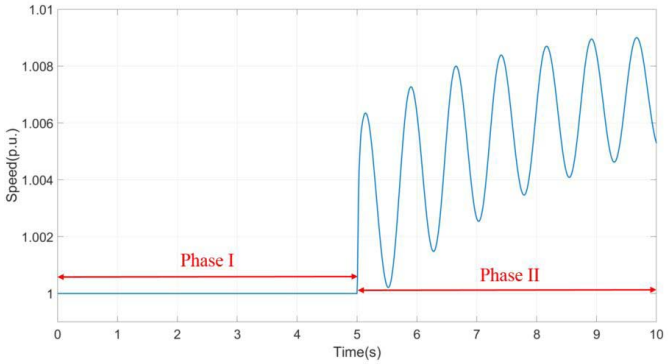


Fig. 3. Typical rotor speed before and after fault.

Therefore, an adaptive learning rate at different period of estimation is proposed to enhance the robustness of the neuroestimator. When the system enters into a dynamic state, the estimation error is at the peak, and the network needs to be adjusted with large steps. Therefore, the learning rate should be relatively high. After the network is approximating the system dynamics, large updating steps will introduce overshoot and oscillation to the output; hence, the learning rate should be relatively lower. We implement the sigmoid function-based adaptive learning rate method. The calculation of learning rate is formulated in the following equation:

$$\eta_k = (\eta_{\text{init}} - \eta_{\min}) \times \frac{1}{1 + e^{k - T_{\text{set}}}} + \eta_{\min} \tag{14}$$

where η_k is the learning rate at a time step k. η_{init} is the initial learning rate that satisfies the fast updating of NN to catch up with the system dynamics. $\eta_{\min}$ is the minimum learning rate that keeps the network stable. T_{set} is the time when the estimator starts tracking system dynamics.

In this way, the learning rate is large enough at the beginning period of DSE without too much decaying and small enough to keep the network stable without becoming idle.

IV. Simulation Results

The proposed HL-DSE algorithm is tested and compared with EKF, UKF, and SR-UKF methods in the case studies. Comprehensive performance comparison on accuracy and time efficiency is conducted on both small- and large-system test cases. The small-system test is performed on the WSCC three-machine nine-bus system, and the large-system test is performed on the NPCC 48-machine 140-bus system [32].

Note that in reality, the DSE is done in a control room, as shown in Fig. 4. Generator terminal voltages are sampled by PMUs and sent back with timestamps. In the control room, an energy management system (EMS) calls the DSE function to estimate generator states, including frequencies and angles, which are inputs to other EMS functions for control. The estimation time is critical since any unstable generator states should be identified as early as possible for system operators to take control actions to prevent system collapse. Considering communication latency and control room operating latency, at the minimum, the average execution time must be lower than 8.33 ms, which is the sampling interval of 120-Hz PMU [33].

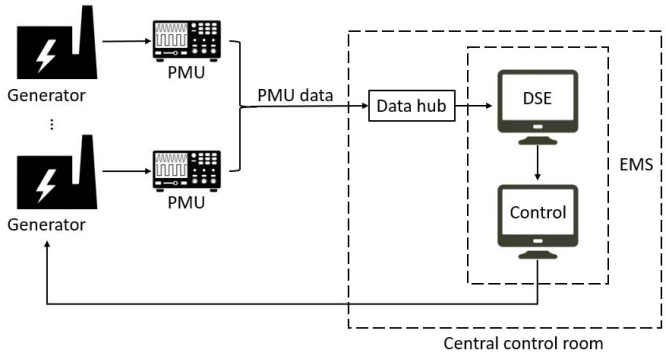


Fig. 4. Architecture of real-time implementation of DSE algorithms.

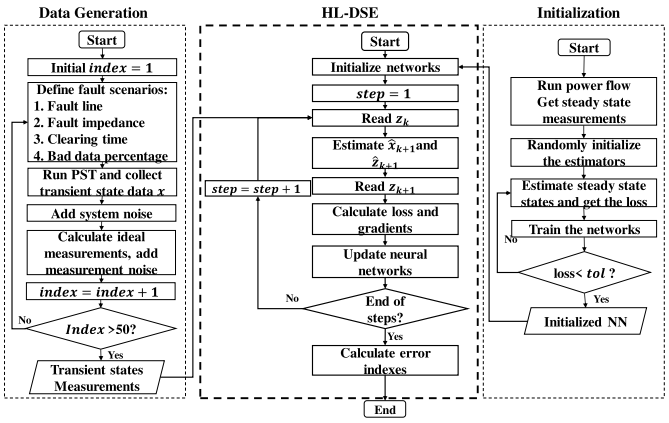


Fig. 5. Flowchart of HL-DSE algorithm and simulation.

To mimic the real-world implementation, we construct a real-time simulation structure, as shown in Fig. 5. The test data set is generated from PST, a MATLAB-based package that simulates power system fault transient [34]. Different scenarios are generated by setting the fault starting time, clearing time, fault impedance, and bad data ratios. Each scenario contains a 10-s simulated PMU data stream with 1200 data points. The PMU data are fed into the HL-DSE algorithm, and the execution time is counted for all 1200 steps. Both HL-DSE and the KF-based algorithms are implemented in MATLAB 2018a under the Microsoft Windows 10 environment on a computer with Intel Core i5-6400 CPU@2.70-GHz Processor and 8-GB RAM.

A. Settings

The simulation data are generated as follows.

1) The simulation step size and the sampling rate of PMUs are set to 120 Hz.
2) Three-phase faults are tested. Six scenarios are tested in the WSCC system, and 50 scenarios are tested on NPCC system. The fault of each scenario is a nongenerator-tripping fault [18]. The fault lines are selected based on the steady-state power flow.
3) The fault lasted for 0.0333 s (two cycles) before it is cleared at the near-end bus. The far-end-bus opens 0.075 s (4.5 cycles) after the fault happened.
4) The testing window is a 10-s postfault period.

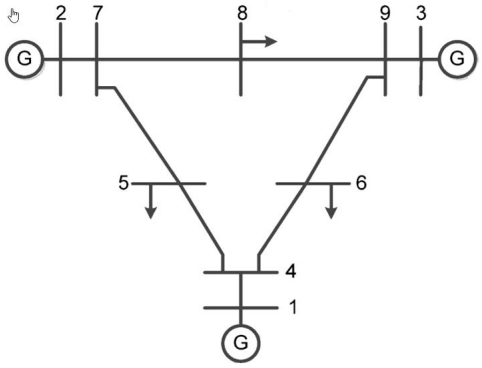


Fig. 6. Topology of the WSCC three-machine nine-bus system.

5) The processing noise and measurement noise are the independent Gaussian noise.
6) The initial values of the state vector in EKF, UKF, and SR-UKF are the steady states. The initial neural network parameters are online trained from the steady states.
7) The adaptive learning rate for the WSCC test system is designed as follows: $\eta_{\text{init}} = 0.045$ and $\eta_{\min} = 0.01$. For the NPCC system where the fault has less impact on the larger system, the learning rate can be set to a constant, $\eta = 0.03$. The values are selected based on the randomly generated training data set. A good range of learning rates can be found at 10^{-2} level.
8) The input layer size and output layer size are related to the dimension of measurements and states, respectively. The number of neurons in the hidden layer is set to three times of machine number to satisfy the approximation complexity. Hence, for the WSCC system, the neural nets have six input neurons, nine hidden neurons, and three output neurons. For the NPCC system, the neural nets have 96 input neurons, 144 hidden neurons, and 48 output neurons.
9) The error index that quantitatively assesses the estimation accuracy is a root-mean-squared error (RMSE)

$$e_x = \sum_{j=1}^{S} \sqrt{\frac{\sum_{i=1}^{g} \sum_{t=1}^{T_s} (\hat{x}_{i,t} - x_{i,t})^2}{g T_s}} \tag{15}$$

where g is the number of generators, T_s is the number of time steps, and S is the number of scenarios. $\hat{x}_{i,t}$ and $x_{i,t}$ are the estimated state and the actual state of generator i at time step t, respectively [18].

B. WSCC Three-Machine System

The WSCC three-machine nine-bus system topology is shown in Fig. 6. The selected fault lines are nongenerator-tripping lines, representing six scenarios in Table I, where the state-steady lines flow and losses are also shown.

1) Computational Efficiency and Accuracy: The performance of the four methods is listed in Table II, including average error indices for generator angle and speed, as well as average time consumption of all six scenarios. It is shown that in this small system, HL-DSE has comparable accuracy

TABLE I
SIX FAULT SCENARIOS

Line number	From bus	To bus	Line flow(pu)	Loss(pu)
1	5	7	0.8508	0.0195
2	7	8	0.7638	0.0027
3	6	9	0.6097	0.0248
4	4	5	0.4690	0.0923
5	4	6	0.3072	0.0401
6	8	9	0.2438	0.0983

TABLE II
SUMMARY OF ESTIMATION PERFORMANCE

Estimator	$\bar{e}_\delta$	$\bar{e}_\omega$	Total time(s)
EKF	0.0187	0.7182	0.1832
UKF	0.0137	0.2524	2.8187
SR-UKF	0.0130	0.2357	3.1469
HL-DSE	0.0200	0.2372	0.1550

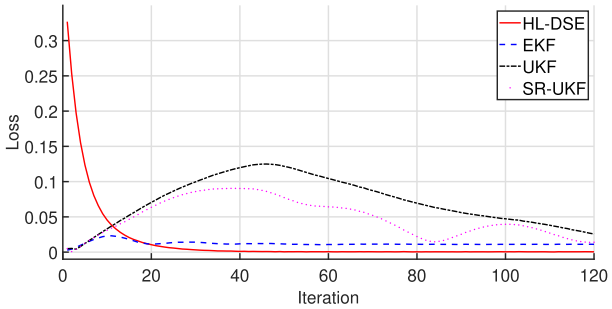


Fig. 7. Convergence of the loss in the steady state for the WSCC system.

TABLE III
ITERATIONS TO CONVERGENCE IN WSCC SYSTEM CASE STUDY

Estimator	Steady state	Dynamic state
EKF	60	58
UKF	261	114
SR-UKF	281	42
HL-DSE	42	35

with the other three methods, with slightly higher e_δ and lower e_ω. In terms of time consumption, HL-DSE and EKF are over ten times faster than UKF and SR-UKF. Obviously, HL-DSE is the most time-efficient algorithm in this small test case.

2) Convergence: The convergence is examined for both the steady-state phase and the dynamic phase. The loss tolerance is set at 10^{-3}. The dual neural networks are initialized and online trained to the steady-state values. For the steady state, the losses of the four estimators are compared for the first 120 iterations (see Fig. 7). For the dynamic state, the average convergence time is calculated for all six scenarios. The results are summarized in Table III.

One can see that HL-DSE yields the fastest convergence. In the steady state, HL-DSE and EKF are the two fastest algorithms to converge, taking 42 and 60 iterations, respectively, while UKF and SR-UKF take much longer to converge (i.e., 261 and 281 iterations).

In the dynamic state, Fig. 8 shows an example of the estimation trajectory. The fault is on line 2 from bus 7 to bus 8. The estimation trajectory shows consistent results with

TABLE IV
ESTIMATION PERFORMANCE—THE BASE FAULT

Estimator	$\bar{e}_\delta$	$\bar{e}_\omega$	Time(s)	Time per iteration(ms)
EKF	0.0691	0.9212	5.5588	4.6323
UKF	-	-	-	-
SR-UKF	0.0366	0.2046	83.8934	69.911
HL-DSE	0.0404	0.1041	0.7152	0.596

Table II, where HL-DSE has competitive results on angle estimation while outperforming EKF and UKF on speed estimation. For the convergence, on average, HL-DSE converges at 35 iterations, while SR-UKF follows at 42 iterations.

Note that in this small system, the execution time of EKF, UKF, and SR-UKF is within a 10-s testing window, indicating the feasibility for online estimation. HL-DSE is the most time-efficient algorithm, and this advantage will be more significant in the larger system as seen in Section IV-C.

C. NPCC 48-Machine System

The NPCC 48-machine 140-bus system is shown in Fig. 9. The total number of lines is 233, and 50 lines with the largest line flow without system oscillations are selected in fault scenarios.

1) Computational Efficiency and Accuracy: Three fault cases are considered: 1) the base fault case with settings in Section IV-A; 2) faults with nonzero impedance; and 3) faults with delayed clearing time. The UKF method failed this test due to a positive semidefiniteness of P in the iteration. EKF and SR-UKF are the baseline methods.

a) Base fault case: Performance of estimation accuracy and time consumption are shown in Table IV. In terms of estimation accuracy, HL-DSE is the second accurate method on δ and the most accurate method of ω. In terms of time consumption, HL-DSE is the fastest method, over 100 times faster than SR-UKF and 80 times faster than EKF. Note that the average time consumption per iteration of SR-UKF is larger than 8.33 ms, making the online application questionable. EKF is sufficiently fast but with unsatisfactory accuracy. The HL-DSE algorithm yields competitive accuracy with the least time consumption, which is suitable for real-time DSE.

The estimation trajectories of machine 3 and machine 7 are shown in Fig. 10, which indicates the same results with the above-mentioned numerical comparison.

b) Faults with nonzero impedance: The fault impedance is set at 0.5 (per unit), and the estimation performances are compared in Table V. The observations from base fault scenarios remain the same. HL-DSE is the second accurate method at δ and the most accurate method in ω. The computational efficiency of HL-DSE is again noteworthy. HL-DSE consumes less than 1% of the time used by SR-UKF. It is approximately eight times faster than the EKF.

c) Faults with delayed clearing time: Normally, the fault clearing time of protecting relays is around two cycles (0.0333 s). Delayed clearing time would lead to large state oscillations and make it more difficult to track generator states. To compare the performance of this severe scenario, the fault clearing time is doubled to 0.0666 s.

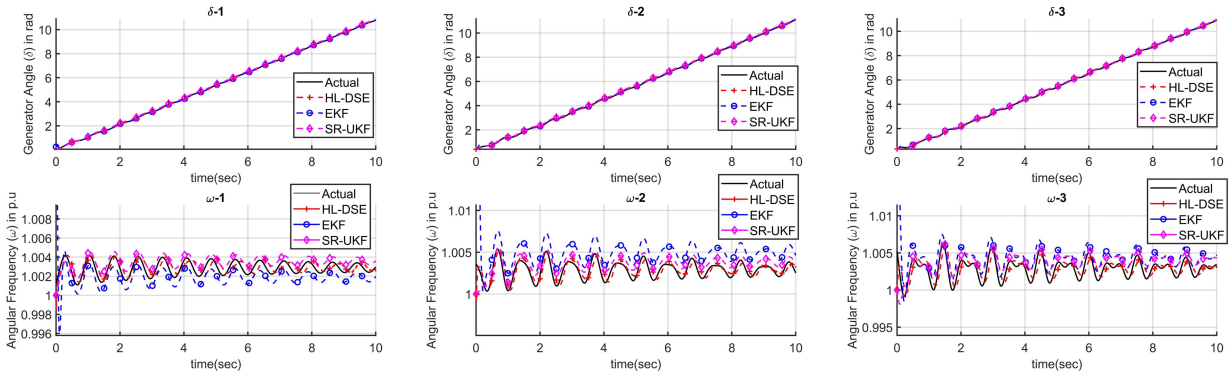


Fig. 8. Estimation trajectory of scenario 2 in the WSCC test.

TABLE V
ESTIMATION PERFORMANCE—NONZERO IMPEDANCE

Estimator	$\bar{e}_\delta$	$\bar{e}_\omega$	Time(s)	Time per iteration(ms)
EKF	0.0705	0.9420	5.3363	4.447
UKF	-	-	-	-
SR-UKF	0.0263	0.2470	82.0163	68.347
HL-DSE	0.0413	0.1239	0.7139	0.595

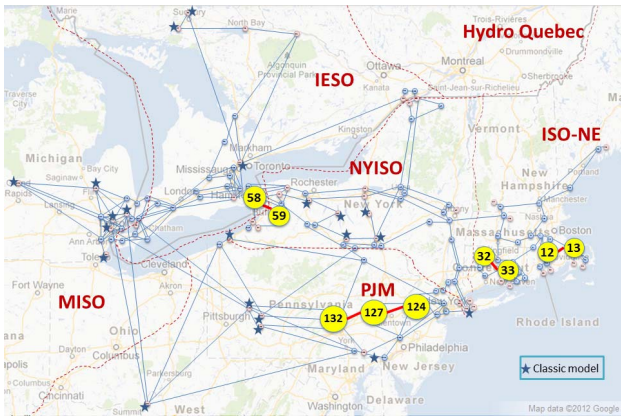


Fig. 9. Map of the NPCC 48-machine 140-bus system.

TABLE VI
ESTIMATION PERFORMANCE—DELAYED CLEARING TIME

Estimator	$\bar{e}_\delta$	$\bar{e}_\omega$	Time(s)	Time per iteration(ms)
EKF	0.7933	1.1886	5.5362	4.614
UKF	-	-	-	-
SR-UKF	0.0248	0.2375	70.5772	58.814
HL-DSE	0.0431	0.1873	0.7549	0.629

We summarize the results for all converged scenarios, and results are shown in Table VI. The results reflect the same conclusion as in the other test cases, where HL-DSE achieves good accuracy using much less time. Here, we notice that there is one nonconverging scenario for HL-DSE, while EKF has six nonconverging scenarios, and SR-UKF has none. This indicates that HL-DSE may be less robust in extreme cases. More will be discussed in Section IV-C3 on the robustness test.

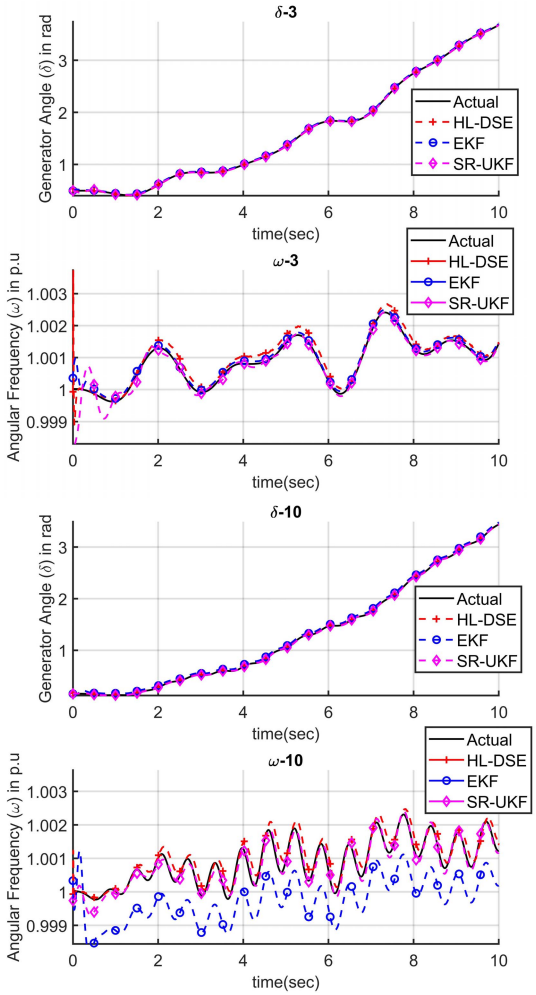


Fig. 10. Machines 3 and 10 estimation trajectory.

d) Machinewise comparison: With a large number of generators and the long distances between them, the estimation accuracy in the NPCC system differs from the generator to generator. Therefore, it is important to have a stable estimation among all machines. The average machinewise error index is shown in Fig. 11. The standard deviation of machinewise estimation errors is given in Table VII to compare their

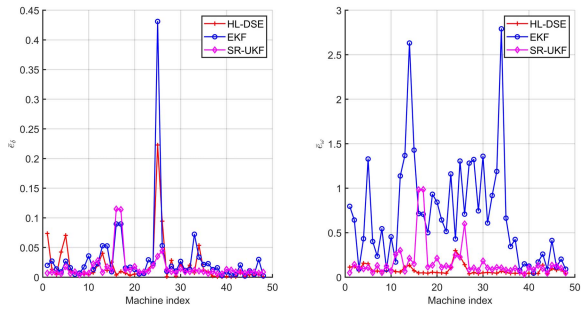


Fig. 11. Machinewise error.

TABLE VII
STANDARD DEVIATION OF MACHINEWISE ERROR

Estimator	$std(e_\delta)$	$std(e_\omega)$
EKF	0.0632	0.6063
UKF	-	-
SR-UKF	0.0370	0.1294
HL-DSE	0.0361	0.0568

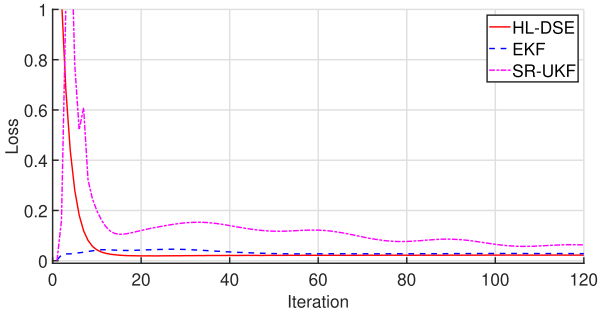


Fig. 12. Loss convergence in the steady state for the NPCC system.

TABLE VIII
ITERATIONS TO CONVERGENCE IN NPCC SYSTEM CASE STUDY

Estimator	Steady state	Dynamic state
EKF	63	812
UKF	-	-
SR-UKF	374	364
HL-DSE	25	225

estimation stability. It can be seen that HL-DSE results in the lowest standard deviations in both angle and speed across all machines. Therefore, HL-DSE can be considered as the most stable and consistent method compared to the other three.

2) Convergence: The convergence for the three methods is compared for the steady and dynamic states. The first 120 iterations in the steady state are shown in Fig. 12. For the dynamic state, the required iterations are averaged across all 50 scenarios. The results are reported in Table VIII.

Consistently, HL-DSE converges fastest to the steady states. With a tolerance level at 10^{-3}, HL-DSE converges at 25 iterations, while EKF converges at 63 iterations and SR-UKF at 374 iterations.

For the dynamic state, Fig. 10 shows an example. One can see that HL-DSE again converges very fast to the true states. Specifically, HL-DSE converges at 225 iterations, while SR-UKF converges at 364 iterations.

TABLE IX
SUMMARY OF ROBUSTNESS TEST PERFORMANCE (100 REPETITIONS)

Filter	$\bar{e}_\delta$	$std(e_\delta)$	$\bar{e}_\omega$	$std(e_\omega)$
EKF	0.0686	0.0633	0.9137	0.6136
UKF	-	-	-	-
SR-UKF	0.0380	0.0361	0.2203	0.1856
HL-DSE	0.0404	0.0313	0.1041	0.0638

TABLE X
PERFORMANCE UNDER FAULTED PMU DATA

Estimator	$\bar{e}_\delta$	$\bar{e}_\omega$
EKF	0.0705	0.9336
UKF	-	-
SR-UKF	0.0270	0.2523
HL-DSE	0.0490	0.2495

3) Robustness: Three cases are tested: 1) robustness against random noises; 2) robustness against faulted PMU data; and 3) robustness against model uncertainty.

a) Robustness against random noises: Due to randomization and model noises, the estimation performance would be different in a series of simulations. To test the robustness of all methods, we perform DSE on one scenario for 100 repetitions and calculate the average error index as well as the standard deviation. The result is given in Table IX. In the robustness test, EKF is doubtless the least accurate, the least robust method as having the highest average error and highest deviation. For the comparison between SR-UKF and HL-DSE, SR-UKF is 6% more accurate on δ, and HL-DSE is 53.97% more accurate on ω, while HL-DSE has lower standard deviation than SR-UKF on both states.

b) Robustness against faulted PMU data: PMU data can be faulted due to communication issues. To test the robustness under this scenario, we randomly select 50 measurements and double the magnitude to simulate a large deviation from the true values. The results are shown in Table X. Note that HL-DSE has one nonconverging scenario, while EKF has four nonconverging scenarios, and SR-UKF has none. For all converged scenarios, HL-DSE is still the second accurate method in δ while performing slightly better than SR-UKF in ω. This again indicates that the robustness of HL-DSE lies between EKF and SR-UKF.

c) Robustness against model uncertainty: We have adopted the classical model from (2) and (3) in our testing so far. Synchronous machine models have been studied extensively, and the IEEE recommends to use 1.1 or 2.2 models [35]. It is necessary to use high fidelity models to obtain close proximity to the real world. Therefore, a robustness test against model uncertainty is critically needed.

Assume that the synchronous machines are represented using the IEEE 1.1 or 2.2 models, but the system operators do not have exact parameters and model information of the true system. Indeed, model validation and parameter calibration of synchronous machines present challenges and require intensive labor and testing [36]. Thus, we study the impact of model uncertainty using the IEEE 1.1 and 2.2 models for simulated

TABLE XI
PERFORMANCE UNDER MODEL UNCERTAINTIES

Estimator	$\bar{e}_\delta$		$\bar{e}_\omega$	
	1.1	2.2	1.1	2.2
EKF	0.1075	0.3620	1.1272	1.6162
UKF	-	-	-	-
SR-UKF	0.1406	0.2432	1.0055	1.4998
HL-DSE	0.0518	0.3634	0.4977	0.5836

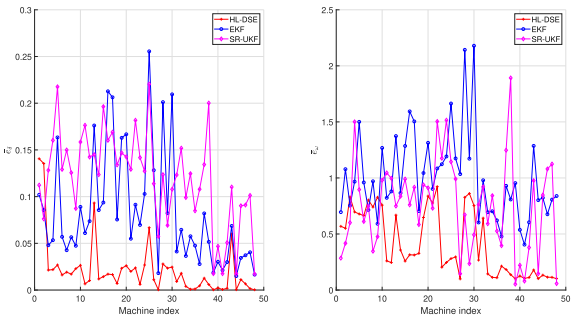


Fig. 13. Comparison of machinewise error index under the IEEE 1.1 model.

data, which are then fed into the proposed HL-DSE as well as the KF-based methods.

During the testing, it is found that under the 1.1 model, HL-DSE and EKF have five nonconverging cases, while SR-UKF has 11; under the 2.2 model, HL-DSE has five nonconverging cases, EKF has 27, and SR-UKF has 23. It is observed that the convergence performance is deteriorated under model uncertainties. Nevertheless, HL-DSE maintains relatively high robustness in terms of convergence compared with the KF-based methods.

Table XI summarizes the estimation results for all converging cases under the IEEE 1.1 and 2.2 models. Comparing the results with the previous testing, one can see that the estimation errors are increased due to inconsistent models for simulation and estimation. The overall performance of HL-DSE is still better than that from EKF and SR-UKF. This robustness of HL-DSE is attributed to the leverage of real-time data correlations to compensate for model uncertainties.

Fig. 13 depicts the averaged errors for all 48 machines under the IEEE 1.1 model. One can see that the errors from the HL-DSE are the lowest for most synchronous machines. Therefore, the proposed HL-DSE appears more robust to model inconsistency than KF-based methods.

d) Note on nonconverging cases: It is noteworthy that in the robustness tests, all three algorithms experience nonconvergence to some extent. Hence, it is necessary to establish backup solutions to mitigate nonconvergence. For example, abnormality detection can be added to HL-DSE by comparing consecutive measurement estimation errors. If the errors are trending higher, early stopping criteria with a minimum learning rate can be deployed. Alternatively, a baseline model that is less accurate but with better convergence could be explored.

V. DISCUSSION

This work is among the first few applying neuroadaptive hybrid learning to estimate power system dynamic states. It is aimed to pave the way for the machine learning to be applied in wide-area monitoring systems in the smart grid.

The dual-net hybrid-learning algorithm maintains a good balance between computational efficiency, accuracy, convergence, and robustness. The significantly improved computational time makes it feasible to estimate generator states in real time. We also notice that the proposed HL-DSE is robust against random noises but needs improvement in extreme cases. This could be potentially addressed by better tuning the learning rates and neural network sizes using more extreme training data set. To improve the estimation performance under IEEE recommended models, the future work will be focused on expanding the neural network architecture and input–output dimensions so that higher-fidelity models could be integrated.

Following this preliminary work, we also intend to explore deep learning and reinforcement learning to enhance estimation accuracy and computational efficiency. For example, a three-layer feedforward neural network configuration is used due to its computational efficiency for the time-critical application. In our future work, we intend to explore deep neural networks and investigate the tradeoff between estimation accuracy and computational efficiency. Another direction is reinforcement learning. Instead of taking historical measurement sequentially, the HL-DSE could take current observations and random selections of historical observations, as introduced in [37], to enhance the learning capability.

VI. CONCLUSION

Power system DSE is enabled by emerging PMU technology. The estimates of generator dynamic states provide essential information for online mitigation of faults and errors. In this article, an HL-DSE is proposed and compared with three widely accepted KF-based DSE methods in both the small WSCC and large NPCC systems. The simulation results show that HL-DSE achieves comparable accuracy in the small system while outperforms KF-based methods in the large system. The advantage of extremely fast execution time makes it desirable for online use. The proposed HL-DSE method is the first implementation of the adaptive neural network into power system DSE. For future work, we seek to improve the robustness of the algorithm. Deep learning and reinforcement learning will also be explored.

ACKNOWLEDGMENT

The authors are grateful to the editor and reviewers for their constructive suggestions and comments that have greatly helped to improve the presentation of our hybrid-learning dynamic state estimation algorithm and our case study findings.

REFERENCES

[1] X. He, J. Yu, T. Huang, and C. Li, "Distributed power management for dynamic economic dispatch in the multimicrogrids environment," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1651–1658, Jul. 2019.

[2] N. Zhou, D. Meng, Z. Huang, and G. Welch, "Dynamic state estimation of a synchronous machine using PMU data: A comparative study," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 450–460, Jan. 2015.

[3] A. K. Singh and B. C. Pal, *Dynamic Estimation and Control of Power Systems*. New York, NY, USA: Academic, 2018.

[4] N. Zhou, J. W. Pierre, and D. Trudnowski, "A stepwise regression method for estimating dominant electromechanical modes," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1051–1059, May 2012.

[5] S. Wei, M. Yang, J. Qi, J. Wang, S. Ma, and X. Han, "Model-free MLE estimation for online rotor angle stability assessment with PMU data," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2463–2476, May 2018.

[6] E. Farantatos, R. Huang, G. J. Cokkinides, and A. P. Meliopoulos, "A predictive generator out-of-step protection and transient stability monitoring scheme enabled by a distributed dynamic state estimator," *IEEE Trans. Power Del.*, vol. 31, no. 4, pp. 1826–1835, Aug. 2016.

[7] A. K. Singh and B. C. Pal, "Decentralized control of oscillatory dynamics in power systems using an extended LQR," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1715–1728, May 2016.

[8] C. Li, X. Yu, T. Huang, and X. He, "Distributed optimal consensus over resource allocation network and its application to dynamical economic dispatch," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2407–2418, Jun. 2018.

[9] X. He, D. W. C. Ho, T. Huang, J. Yu, H. Abu-Rub, and C. Li, "Second-order continuous-time algorithms for economic power dispatch in smart grids," *IEEE Trans. Syst., Man, Cybern.*, vol. 48, no. 9, pp. 1482–1492, Sep. 2018.

[10] C. Li, X. Yu, W. Yu, T. Huang, and Z.-W. Liu, "Distributed event-triggered scheme for economic dispatch in smart grids," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1775–1785, Oct. 2016.

[11] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[12] E. Ghahremani and I. Kamwa, "Dynamic state estimation in power system by applying the extended Kalman filter with unknown inputs to phasor measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2556–2566, Nov. 2011.

[13] N. Zhou, D. Meng, and S. Lu, "Estimation of the dynamic states of synchronous machines using an extended particle filter," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4152–4161, Nov. 2013.

[14] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, Jul. 2017.

[15] E. A. Blood, B. H. Krogh, and M. D. Ilic, "Electric power system static state estimation through Kalman filtering and load forecasting," in *Proc. IEEE Power Energy Soc. Gen. Meeting-Convers. Del. Electr. Energy 21st Century*, Jul. 2008, pp. 1–6.

[16] S. Wang, W. Gao, and A. P. S. Meliopoulos, "An alternative method for power system dynamic state estimation based on unscented transform," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 942–950, May 2012.

[17] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[18] J. Qi, K. Sun, J. Wang, and H. Liu, "Dynamic state estimation for multi-machine power system by unscented Kalman filter with enhanced numerical stability," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1184–1196, Mar. 2018.

[19] A. Del Angel, P. Geurts, D. Ernst, M. Glavic, and L. Wehenkel, "Estimation of rotor angles of synchronous machines using artificial neural networks and local PMU-based quantities," *Neurocomputing*, vol. 70, nos. 16–18, pp. 2668–2678, Oct. 2007.

[20] D. Xu, J. Liu, X.-G. Yan, and W. Yan, "A novel adaptive neural network constrained control for a multi-area interconnected power system with hybrid energy storage," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6625–6634, Aug. 2018.

[21] S. Mehraeen, S. Jagannathan, and M. L. Crow, "Power system stabilization using adaptive neural network-based dynamic surface control," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 669–680, May 2011.

[22] W. Liu *et al.*, "Neural network based decentralized controls of large scale power systems," in *Proc. IEEE 22nd Int. Symp. Intell. Control*, Oct. 2007, pp. 676–681.

[23] R. Zhan and J. Wan, "Neural network-aided adaptive unscented Kalman filter for nonlinear state estimation," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 445–448, Jul. 2006.

[24] B. Wang and K. Sun, "Power system differential-algebraic equations," 2015, *arXiv:1512.05185*. [Online]. Available: https://arxiv.org/abs/1512.05185

[25] F. Schweppe and J. Wildes, "Power system static-state estimation, part I: Exact model," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 1, pp. 120–125, Jan. 1970.

[26] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*. Hoboken, NJ, USA: Wiley, 2008.

[27] J. Qi, K. Sun, and W. Kang, "Optimal PMU placement for power system dynamic state estimation by using empirical observability Gramian," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2041–2054, Jul. 2015.

[28] A. Gomez-Exposito and A. Abur, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC Press, 2004.

[29] S. Haykin and N. Network, "A comprehensive foundation," *Neural Netw.*, vol. 2, p. 41, Feb. 2004.

[30] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 9–48.

[31] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[32] D. Kosterev, C. Taylor, and W. Mittelstadt, "Model validation for the August 10, 1996 WSCC system outage," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 967–979, Aug. 1999.

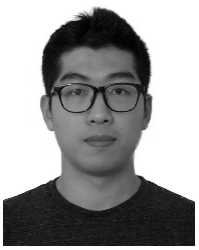
[33] A. Phadke and J. Thorp, "Communication needs for wide area measurement applications," in *Proc. 5th Int. Conf. Crit. Infrastruct. (CRIS)*, Sep. 2010, pp. 1–7.

[34] J. Chow and K. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE Trans. Power Syst.*, vol. 7, no. 4, pp. 1559–1564, Nov. 1992.

[35] P. Dandeno *et al.*, *IEEE Guide For Synchronous Generator Modeling Practices and Applications in Power System Stability Analyses*, IEEE Standard 1110–2002, Dec. 2003, p. 1–72.

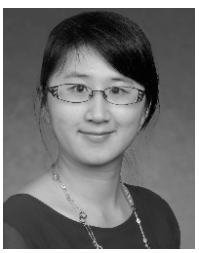
[36] R. Huang *et al.*, "Calibrating parameters of power system stability models using advanced ensemble Kalman filter," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2895–2905, May 2018.

[37] M. Cheng, J. Li, and S. Nazarian, "DRL-cloud: Deep reinforcement learning-based resource provisioning and task scheduling for cloud service providers," in *Proc. 23rd Asia South Pacific Design Autom. Conf. (ASP-DAC)*, Jan. 2018, pp. 129–134.



Guanyu Tian (Student Member, IEEE) received the B.Sc. degree in electrical engineering from Shandong University, Jinan, China, in 2015, and the M.S. degree in electrical engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 2016. He is currently pursuing the Ph.D. degree with the University of Central Florida, Orlando, FL, USA.

His research interests include state estimation, power system data analytics, commercial building modeling, and optimization and demand response.



Qun Zhou (Member, IEEE) received the Ph.D. degree in electrical engineering from Iowa State University, Ames, IA, USA.

She was a Power System Engineer with Genscape, Boston, MA, USA, and GE Grid Solutions, Redmond, WA, USA. She is currently an Assistant Professor with the University of Central Florida, Orlando, FL, USA, where she is also the Director of the UCF Smart Infrastructure Data Analytics Lab. She is devoted to improving energy efficiency and customer engagement through data analytics, probabilistic modeling, and efficient pricing schemes. She focuses on grid-edge resources, including smart buildings, rooftop PVs, and batteries and investigates their interactions with the grid.

Rahul Birari received the B.E. degree in electronics and telecommunication engineering from the University of Mumbai, Mumbai, India, in 2012, and the M.S. degree in electrical engineering from the University of Central Florida, Orlando, FL, USA, in 2017, with a focus on power electronics, and power systems and controls.

He is currently working with Software Motor Company, Sunnyvale, CA, USA, in the areas of motor and drive controls.

Junjian Qi (Senior Member, IEEE) received the B.E. degree in electrical engineering from Shandong University, Jinan, China, in 2008, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013.

He was a Visiting Scholar with Iowa State University, Ames, IA, USA, in 2012, a Research Associate with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA, from 2013 to 2015, and a Post-Doctoral Appointee with the Energy Systems Division, Argonne National Laboratory, Lemont, IL, USA, from 2015 to 2017. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Central Florida, Orlando, FL, USA. His research interests include cascading blackouts, microgrid control, state estimation, synchrophasors, and cybersecurity.

Dr. Qi was a recipient of the NSF CAREER Award in 2020. He is also the Secretary of the IEEE Working Group on Energy Internet and an Associate Editor of IEEE ACCESS.

Zhihua Qu (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in June 1990.

Since then, he has been with the University of Central Florida (UCF), Orlando. FL, USA, where he is currently the SAIC Endowed Professor with the College of Engineering and Computer Science, a Pegasus Professor and the Chair of Electrical and Computer Engineering, and the Director of the FEEDER Center. His areas of expertise are nonlinear systems and control, with applications to autonomous systems and energy/power systems. His recent work focuses upon control, optimization, and plug-and-play resilient operation of networked systems.