IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

Data-driven Dynamic State Estimation Framework Using a Koopman Operator-Based Linear Predictor

D. Y. Yang[1], Member, IEEE, Han Gao[1], Zhe Chen[2], Fellow IEEE, Yanling Lv[1], and Lixin Wang[3]

[1] School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin 150080, China
[2] Department of Energy Technology, Aalborg University, Aalborg, Denmark
[3] School of Electrical Engineering, Northeast Electrical Power University, Jilin 132012, China

Corresponding author: D. Y. Yang (e-mail: dyyang@hrbust.edu.cn).

This work was supported in part by "Excellent Young Teacher Basic Research Support Program" for provincial undergraduate institutions in Heilongjiang Province under Grant YQJH2023248.

ABSTRACT Dynamic state estimation (DSE) is a fundamental task in many fields, including control systems, robotics, and signal processing. Traditional DSE methods, which rely on mathematical models to describe system dynamics, are often limited in their applicability to real-world scenarios due to inaccuracies and assumptions. In this paper, we propose a purely data-driven DSE framework based on a Koopman operator-based linear predictor. The Koopman operator is a powerful tool in dynamical systems theory that allows us to analyze and predict the behavior of nonlinear systems. By leveraging the Koopman operator extracted solely from measured input–output data through the extended dynamic mode decomposition with control (EDMDc) method, a linear predictor that can accurately estimate the state variables of a dynamic system is developed. Introducing the extended Kalman filter (EKF) as an estimation method, the learned Koopman operator-based linear predictor is then used to estimate the current state of the system given only the past and present input–output measurements. To evaluate the effectiveness of the proposed framework, we conduct experiments on both simulated and real-world datasets of complex power systems. The results demonstrate that the proposed data-driven approach outperforms traditional model-based approaches in terms of accuracy and robustness. Moreover, the proposed framework is capable of handling nonlinear and time-varying systems, making it applicable to a wide range of practical cases.

INDEX TERMS Koopman operator, data-driven, linear predictor, extended dynamic mode decomposition with control (EDMDc), extended Kalman filter (EKF), dynamic state estimation (DSE)

I. INTRODUCTION

A. Background and Motivation

In many real-world systems, such as power systems [1,2], manufacturing processes [3], and transportation systems [4], obtaining accurate knowledge of the system's dynamic states is crucial for efficient and safe operation. Dynamic state estimation provides valuable insights into the behavior, performance, and health of the system, enabling effective decision-making and control actions [5-7].

Traditional model-based DSE techniques heavily rely on mathematical models that describe the dynamics of the nonlinear system. However, these models often include simplifications and assumptions that may not fully capture the complexity and variability of real-world systems, such as large-scale power systems. Moreover, they require extensive computational resources and are prone to modeling errors and uncertainties [8].

In contrast, data-driven DSE approaches offer a promising alternative [9,10]. By utilizing the abundant data available from sensors, measurements, and historical records, these approaches can directly learn and estimate the system's dynamic states without explicit dependence on complex mathematical models. This data-driven approach improves the accuracy and robustness of state estimation, particularly in the face of dynamic changes, uncertainties, and disturbances that occur in nonlinear dynamic systems.

B. Literature Review

The two most important tasks in DSE are establishing an appropriate dynamic model (including obtaining accurate model parameters) and choosing efficient filtering algorithms [11]. Compared to the challenge of developing foundational dynamic models, current research on DSE mainly emphasizes the selection and application of filtering algorithms.

The Kalman filter (KF) is a well-known and widely used algorithm in DSE [12]. It is known for its simplicity and efficiency in estimating the state of a linear system in the presence of Gaussian noise. The main limitation of the Kalman filter in DSE is its reliance on linearity and Gaussian assumptions, which may lead to suboptimal performance or divergence in the presence of nonlinearities and non-Gaussian noise. To address these challenges, several variations of the Kalman filter have been developed, with representative examples including the extended Kalman filter (EKF) [13], unscented Kalman filter (UKF) [14], and cubature Kalman filter (CKF) [15]. Additionally, continuous improvements are being made to the EKF, UKF, and CKF [16-21]. Another widely used approach in DSE is particle filtering (PF) [22], also known as the sequential Monte Carlo (SMC) filter. Similar to the Kalman filter, improved versions of particle filtering, such as the extended particle filter (EPF) [23] and unscented particle filter (UPF) [24], have also been developed. Other filtering algorithms applied in DSE include the Bayesian filter [25], Gaussian mixture model filter [26], and maximum likelihood ensemble filter (MLEF) [27].

Each filtering algorithm possesses specific advantages that depend on the characteristics of the dynamic system at hand. However, achieving desirable performance with these filtering algorithms in DSE relies on the assumption that the system model and its parameters are accurate and reliable.

In fact, the model and parameters are also crucial for DSE. References [28] and [29] discussed the effects of applying detailed models in DSE. However, even with a unified research object, it is challenging to establish a unified model specifically for DSE, particularly for dynamic systems that involve multiple auxiliary control devices, such as the synchronous generator (SG) and the doubly fed induction generator (DFIG) wind turbine.

C. Contribution and Organization

In recent years, data-driven modeling techniques and methods have received significant attention in the field of nonlinear dynamic system monitoring and control. In particular, Koopman operator-based data-driven modeling methods have experienced rapid development. References [10] and [30] conducted exploratory research on data-driven DSE based on the Koopman operator. However, neither of these studies considered the role of controllers. In this paper, a purely data-driven DSE framework is constructed by introducing a linear predictor [31] based on the data-driven Koopman operator. The developed Koopman operator-based linear predictor not only captures the dynamic characteristics of the system but also takes into account the role of the controller. The main contributions of this paper are summarized as follows.

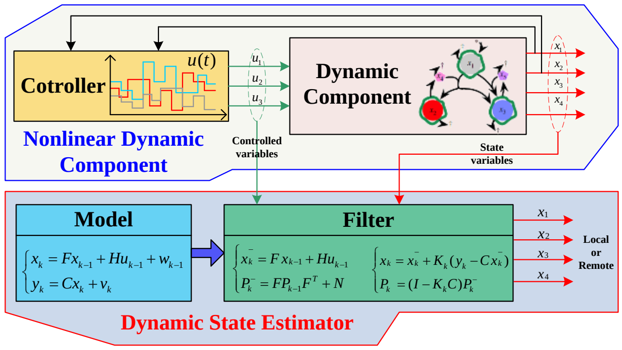

Fig. 1. Dynamic state estimation process flowchart for dynamic system

1) Utilization of the Koopman operator for behavior analysis and prediction: The Koopman operator, a powerful tool in dynamical systems theory, enables us to analyze and predict the behavior of nonlinear systems. By incorporating the Koopman operator into our DSE framework, we can accurately estimate the current state of the system based on past and present input–output measurements.

2) Integration of the extended Kalman filter (EKF) for state estimation: We employ the EKF as an estimation method within our framework. The learned Koopman operator-based linear predictor, combined with the EKF, enhances the accuracy and robustness of state estimation.

3) Proposal of a purely data-driven dynamic state estimation (DSE) framework: We introduce a novel approach that does not rely on mathematical models to estimate the state variables of a dynamic system. Instead, we leverage the Koopman operator-based linear predictor, which is derived solely from input–output data using the extended dynamic mode decomposition with control (EDMDc) method.

4) Evaluation of the proposed framework on simulated and real-world datasets: We conduct experiments on complex power system datasets to assess the effectiveness of our data-driven approach. The results demonstrate superior performance compared to that of traditional model-based approaches in terms of accuracy and robustness. Furthermore, our framework exhibits versatility in handling nonlinear and time-varying systems, making it suitable for various practical applications.

The remainder of this paper is structured as follows: Section II presents the classical unified DSE framework. Section III presents the proposed purely data-driven DSE based on a linear predictor extracted by employing EDMDc. The performance verifications are presented based on numerical simulations and real-world datasets of a complex power system in Section IV. The conclusions of this paper are presented in Section VI.

II. Unified Dynamic State Estimation Framework

DSE refers to the estimation of a system's future states

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535521

over a certain period of time based on observed data and a model, as shown in Fig. 1. It has significant applications in various fields, such as robot navigation, traffic management, power systems, and financial forecasting.

For complex nonlinear dynamical systems, DSE is a complex and crucial process that requires the consideration of factors such as model selection, data processing, and optimization algorithms to provide accurate and reliable state estimation results. It plays a vital role in system monitoring, identification, and control.

DSE is essentially the design of a nonlinear observer, where the construction of the system model and the choice of filtering algorithm are two key factors for achieving accurate and efficient estimation.

A. The state-space model for dynamic estimation

In general, the mathematical model of a time-varying nonlinear dynamic component/system can be described as follows:

$$\begin{cases} \dot{x} = f(x,u) \\ y = h(x) \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$ is the state variable vector, $u \in \mathbb{R}^l$ is the control variable vector, $y \in \mathbb{R}^m$ is the observation variable vector, $f(\cdot)$ is an n-dimensional nonlinear state equation, and $h(\cdot)$ is an m-dimensional nonlinear measurement equation.

In practical applications, it is necessary to discretize Equation (1). The resulting equation after discretization can be represented as:

$$\begin{cases} x_{k+1} = f(x_k, u_k) + w_k \\ y_{k+1} = h(x_{k+1}) + v_{k+1} \end{cases} \tag{2}$$

where $w \in \mathbb{R}^n$ is the process noise vector and $v \in \mathbb{R}^m$ is the measurement noise vector. It is typically assumed that $w_k \sim N(0,Q)$ and $v_k \sim N(0,R)$, where Q and R are the covariance matrices of the process noise and measurement noise, respectively.

Furthermore, we describe the discrete model of the system in the state-space form as follows:

$$\begin{cases} x_{k+1} = Fx_k + Gu_k + w_k \\ y_{k+1} = Hx_{k+1} + v_{k+1} \end{cases} \tag{3}$$

where *F*, *G*, and *H* are the state matrix, control matrix, and observation matrix, respectively.

Equation (3) represents the general state-space model of the component or system used for nonlinear dynamic state estimation. Based on this model, the appropriate filtering algorithm can be chosen to perform dynamic estimation of the state variables.

Remark: A precise dynamic component/system model is essential for designing effective observers and optimizing their parameters. The observers utilize the dynamic component/system model in conjunction with observed data to perform state estimation. The design of an observer relies on the dynamic component/system model, including selecting appropriate observer types and determining observer gain matrices and noise covariance matrices, among other parameters. Therefore, an accurate and applicable dynamic component/system model is a fundamental prerequisite for achieving efficient and precise state estimation. The component/system models traditionally used in DSE are generally mathematical models based on physical principles and behavioral laws, which help researchers understand their physical behavior and mechanisms. However, for complex components and systems, establishing accurate mathematical models based on physical properties can become very complex and challenging. Moreover, such models typically involve numerous parameters that need to be estimated through experiments or other means. The process of parameter estimation can be influenced by measurement errors, limitations in data collection, and incomplete information, which in turn affect the accuracy and reliability of the model. Additionally, these types of models do not consider dynamic interactions and influences between components. In recent years, data-driven modeling methods have developed rapidly because data-driven models can effectively avoid the issues of poor reliability in physically driven models and their parameters. This paper primarily focuses on data-driven modeling methods applicable to dynamic state estimation, with the objective of achieving model- and parameter-independent dynamic state estimation under only data-driven conditions.

B. The selection of filtering method

Among various algorithms, the Kalman filtering algorithm is the most classical method in DSE. Based on the model given in Equation (3), the KF performs dynamic estimation of state variables through the prediction-correction process.

1) *Prediction step*

$$\begin{cases} \hat{x}_{k+1}^- = F\hat{x}_k + Bu_k \\ P_{k+1}^- = FP_kF^T + V_{k+1} \end{cases} \tag{4}$$

2) *Correction (Filtering) step*

$$K_k = P_k^- H^T (HP_k^- H^T + W_k)^{-1} \tag{5}$$

$$\begin{cases} \hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - H\hat{x}_{k+1}^-) \\ P_{k+1} = (I - K_{k+1}H)P_{k+1}^- \end{cases} \tag{6}$$

In Equations (4)-(6), *P* is the covariance matrix of the state variables, and *K* is the gain coefficient matrix. W_k and V_k are the state and observation error covariance matrices, respectively. Typically, w_k and v_k are assumed to be independent Gaussian white noise; $\mathrm{E}[w_k] = 0$, $\mathrm{E}[w_k w_k^T] = W_k$, $\mathrm{E}[v_k] = 0$, and $\mathrm{E}[v_k v_k^T] = V_k$, where $\mathrm{E}[\cdot]$ denotes the expectation operator. Reference [12] provides a

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535521

detailed and comprehensive introduction to the Kalman filter.

Remark: The selection and application of filtering algorithms are common areas of innovation in current research on dynamic state estimation. In DSE, the classical filtering algorithms that are commonly used include the KF and PF, as well as variations and improvements of these two algorithms. Every filtering approach provides ideal estimation results within its applicable scenario, but this relies on accurate and reliable basic models and parameters for the system/components. This paper primarily focuses on the construction of nonlinear component/system models in DSE. The chosen filtering algorithm is the classical extended Kalman filter.

This paper takes a data-driven approach and extracts a Koopman operator-based linear predictor from measurement data, following the same form as Equation (3), for dynamic state estimation.

III. Purely Data-driven Linear Predictor Based Dynamic State Estimation Scheme

Dynamic state estimation focuses on the prediction of the trajectory of the state variables in Equation (1) given a condition x and control inputs u. In a purely data-driven environment, the key is to use known state variables and control variables to extract linear predictors in the form of Equation (3). It is a common approach to assume the investigated linear predictors are a linear dynamical system with control, and the form of this system is as follows [31].

$$\begin{cases} \dot{z} = Az + Bu \\ \hat{x} = Cz \end{cases} \tag{7}$$

$$z = \varphi(x) \tag{8}$$

where $z \in \mathbb{R}^N$ (with $N \geq n$) is the lifting state variable vector, $\hat{x}$ is the prediction (estimation) of x, $\varphi(\cdot)$ is a lifting function that is specified by the user, $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times l}$ and $C \in \mathbb{R}^{n\times N}$.

As process noise w and measurement noise v exist independently, they can be directly incorporated into the linear predictor presented in Equation (7). By incorporating filtering algorithms, such as the extended Kalman filter, it is possible to perform dynamic estimation of the state variables. Next, our primary task is to extract the linear predictor from the measurement data.

A. Koopman operator for dynamic with control

The Koopman operator is a mathematical tool commonly used in nonlinear dynamical systems, providing a linear representation of the system's temporal evolution. The Koopman operator maps the state variables of the system to functions in an infinite-dimensional Hilbert space, where linearity characterizes the dynamics, enabling the construction of the linear predictor (7). This transformation converts the originally complex nonlinear system into a linearly solvable problem, facilitating further analysis and control.

Specifically, Koopman is an infinite-dimensional linear operator that acts on a scalar observable g: $\mathbb{R}^{n+l} \to \mathbb{R}$, which belongs to $\mathcal{H}$ and provides the expected value evolution in the state space. In this paper, referring to the rigorous and practical Koopman operator proposed in reference [31] for a dynamical system with control, we consider the Koopman operator $\mathcal{K}$:$\mathcal{H} \to \mathcal{H}$ for dynamic system (1) as:

$$[\mathcal{K}g](x_k, u_k) = g(f(x_k, u_k), u_{k+1}) \tag{9}$$

with $x_k = x(k)$ and $u_k = u(k)$. Including control in (9) renders the Koopman operator nonautonomous.

In theory, the Koopman operator (9) is a linear operator that fully describes the nonlinear dynamical system (1). For a detailed review of the Koopman operator and its applications, see [31] and [32].

To obtain a linear predictor of the form (7), it is highly suitable for the objectives of this paper to employ the data-driven method of extracting a finite-dimensional approximation to the infinite-dimensional Koopman operator $\mathcal{K}$.

Dynamic mode decomposition (DMD) is a typical and effective data-driven method used for extracting finite-dimensional approximations of the Koopman operator. DMD [33] offers a way to project the Koopman operator onto the space of linear observables, while extended dynamic mode decomposition (EDMD) [34] provides more accurate approximations by incorporating nonlinear observables.

Algorithm 1 Extended dynamic mode decomposition with control

Inputs: Data matrices X, X_f, U; Lifting function $\psi(\cdot)$

Outputs: Koopman model matrices $\hat{A}$, $\hat{B}$, and $\hat{C}$

Step **1:** Choose a truncation value r

Step **2:** Perform *SVD*: $[\psi(X)^T \quad U^T]^T = U_r S_1 V_1^*$

Step **3:** Bipartite $U_r = [U_{r1}^T \quad U_{r2}^T]^T$ using the number of observables

Step **4:** Perform *SVD*: $\psi(X_f) = U_2 S_2 V_2^*$

Step **5:** Solve (13) to get C

Step **6:** Compute the reduced-order model matrix

$\hat{A} \leftarrow U_2^T \psi(X_f) V_1 S_1^{-1} U_{r1}^T U_2$

$\hat{B} \leftarrow U_2^T \psi(X_f) V_1 S_1^{-1} U_{r2}$

$\hat{C} \leftarrow C U_2$

Step **7:** Return $\hat{A}$, $\hat{B}$, $\hat{C}$

B. Extended dynamic mode decomposition with control

Following [32], it is assumed that the measurement data are collected in the form of snapshots as:

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535521

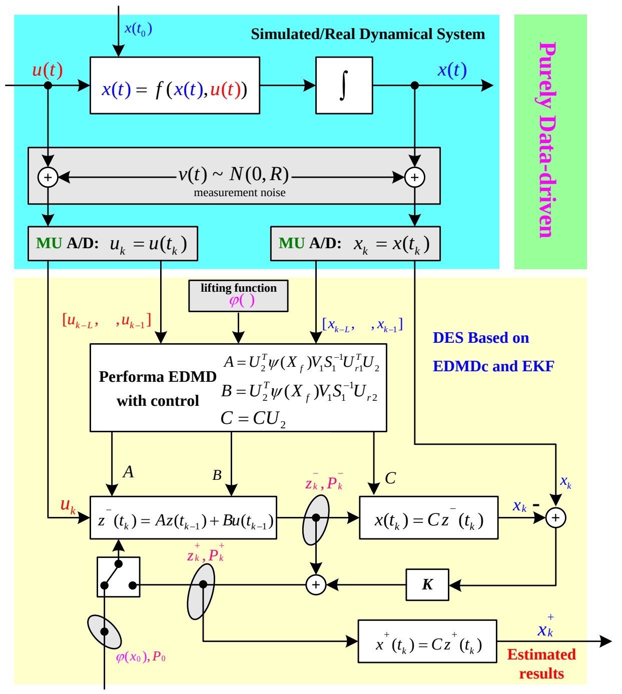

Fig. 2. Signal flow diagram of the proposed data-driven DES scheme based on EKF

$$X=[x^1,\ldots,x^L],\quad X_f=[x_f^1,\ldots,x_f^L],\quad U=[u^1,\ldots,u^L] \tag{10}$$

where $x^i, x_f^i \in \mathbb{R}^n$, $u^i \in \mathbb{R}^l$, and $x_f^i = f(x^i,u^i)$. Therefore, we do not need to make any assumptions or reset the temporal ordering of the data as $x_f^i = x^{i+1}$. The lifting function's action can then be expressed as:

$$\psi(X)=[\varphi(x^1),\ldots,\varphi(x^L)] \tag{11}$$

where $\varphi(x)=[\varphi_1(x),\ldots\varphi_N(x)]$ is a given dictionary of nonlinear lifting functions.

Given the data X, X_f, and U in Equation (10), we obtain the matrices A, B, and C in Equation (7) as the optimal linear one-step predictor in the lifted space using least-squares regression, and these matrices are derived as the solutions to the optimization problem:

$$\min_{A,B}\left\|\psi(X_f)-A\psi(X)-BU\right\|_F \tag{12}$$

$$\min_{C}\left\|X-C\psi(X)\right\|_F \tag{13}$$

where the symbol $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

By solving the normal equations for equations (12) and (13), we can obtain A, B, and C.

In practical applications, it is desirable to have a low-dimensional model for fast dynamic state estimation and real-time control. To achieve this, the basic DMD scheme can be extended to incorporate exogenous effects and utilize truncated POD modes for order reduction [35], allowing approximation through linear observables. Here, EDMD is developed to establish a reduced-order Koopman representation with control. To achieve this objective, the EDMD algorithm begins with a singular value decomposition (*SVD*):

$$[\psi(X)^T \quad U^T]^T = U_r S_1 V_1^* \tag{14}$$

where U_r and V_1 are the left and right singular vectors respectively; and S_1 is the singular value matrix, $U_r=[U_{r1}^T \quad U_{r2}^T]^T$.

Next, the *SVD* is performed on $\psi(X_f)$ as follows:

$$\psi(X_f)=U_2 S_2 V_2^* \tag{15}$$

where the truncation value is r and $r<n$; thus, a low dimensional Koopman model of order r is established. The reduced-order model matrix can be computed using the following formulas:

$$\mathring{A}=U_2^T\psi(X_f)V_1S_1^{-1}U_{r1}^TU_2 \tag{16}$$

$$\mathring{B}=U_2^T\psi(X_f)V_1S_1^{-1}U_{r2} \tag{17}$$

$$\mathring{C}=CU_2 \tag{18}$$

The linear predictor (2) thus reduces to the coordinate $z=U_2^T\varphi(x)$ by replacing A, B, and C with $\mathring{A}$, $\mathring{B}$, and $\mathring{C}$. In this way, we utilize the first r Koopman modes for constructing a reduced-order dynamic system representation, summarized in **Algorithm 1**

C. Purely data-driven dynamic state estimation scheme

Based on the linear predictor extracted from the measurement data by employing EDMDc, a purely data-driven dynamic state estimation strategy is designed, in which the extended Kalman filter, proven to be effective in dynamic state estimation, is selected. The key of the proposed purely data-driven dynamic state estimation method is to extract linear predictors from the data, which are collected using measurement units (MUs), such as phasor measurement units (PMU) in power systems. The signal flow diagram of the proposed purely data-driven DSE scheme based on the EKF is shown in Fig. 2.

It should be noted that in the extracted Koopman-based linear predictors, a lifting state variable $\boldsymbol{z=\varphi(x)}$ is introduced, and the system state variables are reconstructed as "observed

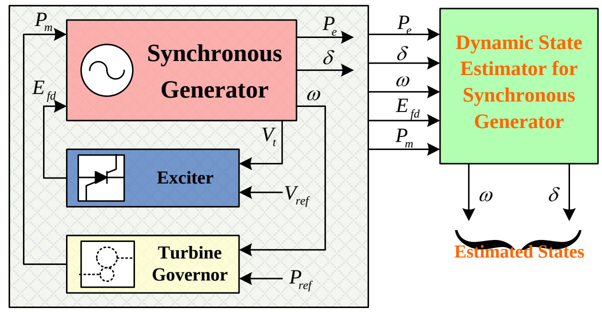

Fig. 3. Schematic diagram of DES for synchronous generator.

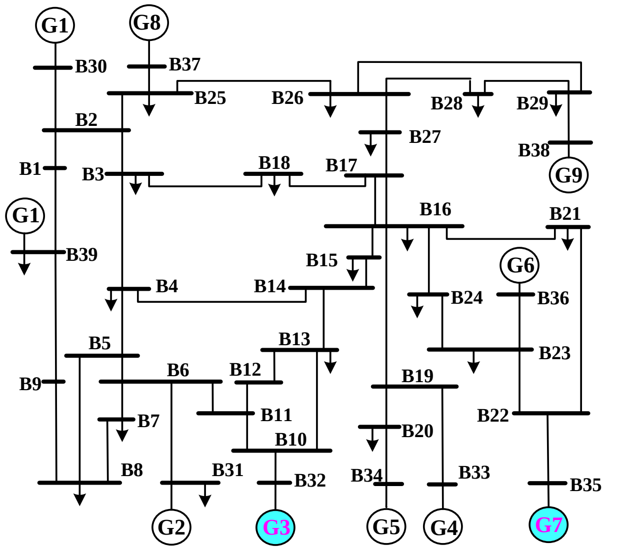

Fig. 4. Structural diagram of IEEE 10- generator 39-bus test system.

variables". Therefore, in the proposed DSE scheme, by utilizing the EKF, we obtain the lifting state variable z. To further obtain an estimation of the actual system state variables, we need to utilize the observation equations in the extracted linear predictors. This is different from the traditional EKF approach

II. CASE STUDIES

To validate the purely data-driven DSE scheme proposed in this paper, we conduct calculations and analysis using the example of DSE for a synchronous generator s in a complex power system. This section presents the results of our investigation and demonstrates the effectiveness of our data-driven DSE scheme in this specific scenario.

Modern power systems are typical large-scale nonlinear dynamic systems, with the dynamic processes primarily driven by synchronous generators. The DSE of synchronous generators in a complex power system, as shown in Fig. 3, poses unique challenges due to the inherent complexities involved. First, synchronous generators are highly nonlinear and time-varying systems, which makes it challenging to accurately capture their dynamic characteristics and model their dynamic behavior. Additionally, synchronous generators are interconnected with other components in the power system, such as transmission lines, transformers, and control devices. This interdependency introduces complexities in the dynamics of the system, as variations in one component can propagate and affect the behavior of the synchronous generator. Thus, the DSE of synchronous generators in a complex power system is a challenging task that requires robust algorithms, advanced modeling techniques, and intelligent data-driven approaches to handle the intricacies posed by nonlinearity, coupling effects, synchronization requirements, measurement uncertainties, and the large-scale nature of the system.

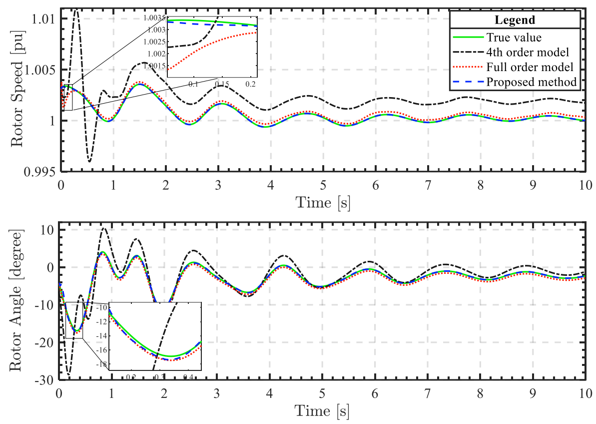

Fig.5. Estimation results of G3 under the scenario of model mismatch

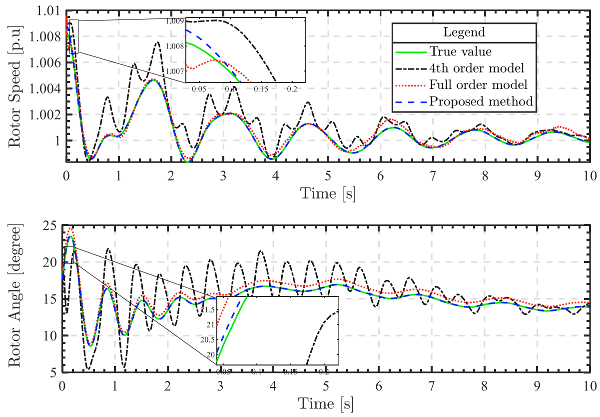

Fig. 6. Estimation results of G7 under the scenario of model mismatch.

In the numerical simulations, we first compared the proposed data-driven DSE scheme with the model-based DSE methods. Next, we analyzed the impact of parameter uncertainty on model-based DSE methods and the proposed data-driven method. Given the data-driven nature of the proposed scheme, we utilized actual PMU recording data to investigate and evaluate the effectiveness and applicability of the proposed DSE scheme, which is based on the Koopman operator-based linear predictor.

All calculations were performed in MATLAB on a workstation with an Intel Core i9-14900K CPU and 16 GB memory.

A. Numerical simulation results

In this study, the simulation system is the IEEE 10-generator, 39-bus test system (as illustrated in Fig. 4), in which all synchronous generators assumed for numerical simulation are described using the detailed two-axis generator model with an IEEE-DC1A exciter and a TGOV1 turbine governor, whose parameters are taken from [36]. A three-phase short-circuit fault with a duration of 0.1 s was applied to bus 2, and the power system toolbox (PST) [37] was used to generate the simulation results, which were selected as the actual values. It was assumed that all generators were equipped with PMUs, which were used to collect measurements of the rotor angle, rotor speed, active power, *et.al*.

In the construction of linear predictors using EDMDc, the selection of the state variables is an open problem that can be adjusted based on the requirements of the research object and the scenario. Here, we focus on the electromechanical dynamic response of synchronous generators. Therefore, $x = [\delta,\ \omega]^T$ was selected as the state variable. Without loss of generality, the control variable was chosen as $u=[P_m, E_{fd}]^T$. In the proposed data-driven DSE method, another key aspect is the selection of the lifting functions $\boldsymbol{\varphi(x)}$. For different research objects and different problems, there is no unique choice for the lifting function. In this work, we chose $\mathbf{z}=\boldsymbol{\varphi}(\boldsymbol{x})=[\ \delta,\ \omega,\ \sin(\delta),\ \sin(\omega),\ E_{fd},\ P_e]^T$ as the lifting function, where $\delta,\ \omega,\ E_{fd},$ and P_e are all measurable electrical quantities. (δ and ω are the rotor angle and speed of the synchronous generator, respectively; E_{fd} is the field voltage; and P_m and P_e are the mechanical power and active electrical power, respectively.)

A random Gaussian variable, $N(0, 0.01)$, is selected to simulate system process noise and measurement noise, and the covariance matrix is $10^{-6}I$.

We tested and compared three DSE methods, namely, a *4th-order model* in which a 4th-order synchronous generator model was used, a *full-order model* in which a 6th-order synchronous generator model with an IEEE-DC1A exciter and a TGOV1 turbine-governor was used, and the proposed method in which the extracted linear predictor was used. For convenience in analysis, the EKF is used as the filtering algorithm for all three methods.

1) Model mismatch

After selecting an efficient filtering algorithm, the model is the key to obtaining accurate results in dynamic state estimation. In current research on DSE for synchronous generators, commonly used models are the reduced-order models of synchronous generators. In this paper, generators G3 and G7 were selected to demonstrate the estimation performance of the proposed method, the 4th-order model, and the full-order model. Note the different numbers of state variables associated with different models. In this paper, we focused on analyzing two commonly observed state variables, namely, the rotor angle δ and rotor speed ω.

The estimated results for generators G3 and G7 are displayed in Fig. 5 and Fig. 6, respectively. The estimation performance of the 4th-order generator model is significantly poorer in terms of both the state variables, rotor angle δ and rotor speed ω. In particular, for generator G7, the estimation results of the 4th-order generator model exhibited noticeable oscillation with a significant deviation from the true values. The same phenomenon also occurred in the initial stage of the estimation results for generator G3. The main reason for this is that the reduced-order model used for dynamic state estimation does not include the detailed models of the excitation and turbine-governor systems that are present in the actual numerical simulation.

Additionally, as shown in Fig. 5 and Fig. 6, the estimation results of the proposed method and the full-order model show good agreement with the true dynamic process of the generator. However, there is still a certain deviation between the results of the full-order model and the true values. In comparison, the estimation results of the proposed method perfectly match the true values, except for a small deviation during the initial moment. The response data can accurately reflect the dynamic process of the system and components. The linear predictor extracted from the dynamic response data using EDMDc is also closer to the actual physical process of the synchronous generator. This is the most direct reason why the estimation results of the proposed method are more accurate. Although the full-order model also uses the same detailed model considering the excitation and turbine-governor systems as in the numerical simulation, it neglects the interaction between the synchronous generators within the system, leading to a deviation between its estimation results and the true values.

Remarks: Accurate models serve as the foundation for achieving precise dynamic state estimation. The numerical results indicate that model mismatch not only leads to significant estimation errors but also tends to cause numerical oscillations in the estimation process. The linear predictor, which is extracted from dynamic response data using EDMDc, can more accurately describe the dynamic characteristics of grid-connected synchronous generators under disturbances. The DSE results based on the proposed linear predictor-based scheme are also more accurate.

2) Parameters uncertainty

For component mathematical models built on physical principles, there are numerous parameters involved. For example, the detailed synchronous generator model with the IEEE-DC1A exciter and IEEE TGOV1 turbine governor, which was used in this paper, has 22 parameters, and the numerical values of these parameters are influenced by various factors, introducing a level of uncertainty. This uncertainty in the model parameters will directly impact the results of dynamic state estimation.

To investigate the impact of the model parameters on the dynamic state estimation results, two testing conditions were set as follows:

Case A: We assume 20% uncertainties in the inertia and damping constants for generator G3.

Case B: We assume 15% uncertainties in the transient reactance of the generator and the gain of the excitation system for generator G7.

Note: In these two cases, to investigate the advantages of the proposed data-driven DSE method under parameter uncertainty conditions more clearly, the discussion of the 4th-order generator model is not included. The proposed DSE method is model- and parameter-free, while the full-order model requires the inclusion of model and parameter information.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3535521

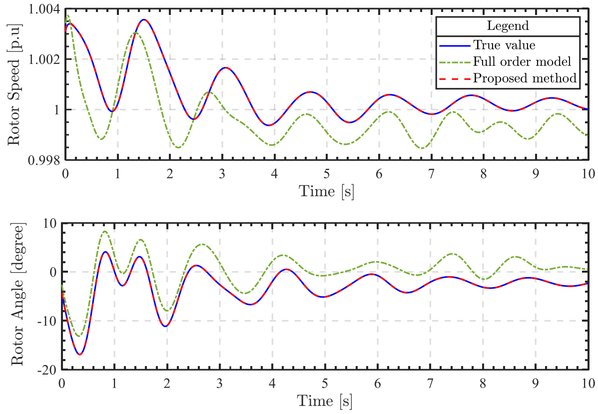

Fig.7. Estimation results of G3 under the scenario of parameters uncertainty

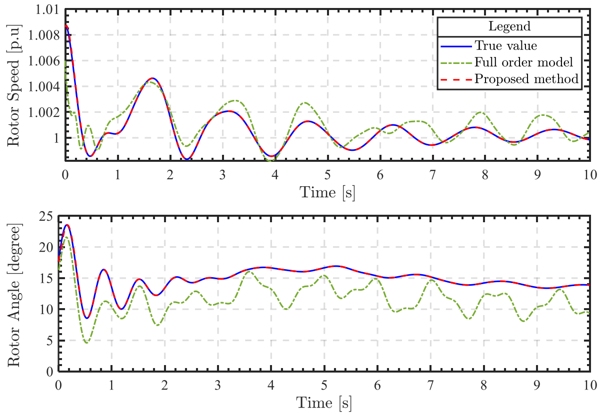

Fig.8. Estimation results of G7 under the scenario of parameters uncertainty

The estimated rotor speed and rotor angle obtained by the proposed method and the full-order model for Case A and Case B are shown in Fig. 7 and Fig. 8. As depicted in Fig. 7, when there is 20% uncertainty in the inertia and damping constants used in the state estimation model compared to the true values, the estimation results based on the full-order model not only have a slow tracking speed but also deviate significantly from the actual rotor speed and rotor angle. It is evident from Fig. 8 that the estimation results based on the full-order model are poor, especially in the case of rotor angle, when there is a 15% uncertainty in the transient reactance of the generator and the gain of the excitation system.

In summary, the issues found in Case A and Case B highlight the sensitivity of model-based DSE methods to parameter uncertainties. In contrast, the proposed DSE method, which utilizes data-driven modeling, is not influenced by parameter uncertainty. The advantages of the proposed data-driven DSE scheme based on a linear predictor are further evident in this context.

TABLE I
STATISTICAL RESULTS OF ESTIMATION ERROR FOR G3

State variable	Category	NRMSE	RAE	R^2
Rotor Speed	A	**0.0066**	**0.0027**	**0.9996**
	B	**0.3209**	**0.3469**	**0.8970**
	C	0.8879	0.9065	0.1365
	D	0.9627	0.9968	0.0341
Rotor Angle	A	**0.0014**	**0.0016**	**0.9999**
	B	**0.1491**	**0.2196**	**0.9077**
	C	0.6071	0.7566	0.24278
	D	0.9496	0.9653	0.09816

TABLE II
STATISTICAL RESULTS OF ESTIMATION ERROR FOR G7

State variable	Category	NRMSE	RAE	R^2
Rotor Speed	A	**0.0205**	**0.0072**	**0.9995**
	B	**0.1819**	**0.2124**	**0.86691**
	C	0.6671	0.7514	0.47838
	D	0.8720	0.8083	0. 04391
Rotor Angle	A	**0.0411**	**0.0143**	**0.9983**
	B	**0.3188**	**0.4865**	**0.8983**
	C	0.7998	0.7882	0.3602
	D	0.9284	0.9064	0.0871

3) Estimation error statistics

To objectively analyze the impact of model mismatch and parameter uncertainty on the results of dynamic state estimation, this study employed three dimensionless error statistical indicators, namely:

Normalized Root-Mean-Squared Error (NRMSE): NRMSE is based on the root-mean-squared error (RMSE) and is calculated by dividing it by a scalar value [38]. It overcomes scale dependency and simplifies the comparison between models or datasets of different scales. The range of NRMSE is [0, 1], and a value closer to 0 indicates a smaller error between the calculated result and the true value.

Relative Absolute Error (RAE): The RAE is calculated by dividing the total absolute error by the absolute difference between the mean value and the actual value [38]. The values of RAE range from 0 to 1. Zero is the optimal value for a model.

R^2 (Coefficient of determination): R^2 is a statistical analysis indicator used to measure the degree of correlation between variables and is commonly used to assess the goodness of fit of a model [39]. The range of R^2 is from 0 to 1. The closer the value is to 1, the stronger the model's fitting ability. In this study, it is used to evaluate the performance of the DSE results under different scenarios.

To analyze the estimation errors under different operating conditions more clearly, we categorize the estimation results under model mismatch and parameter uncertainty into four categories:

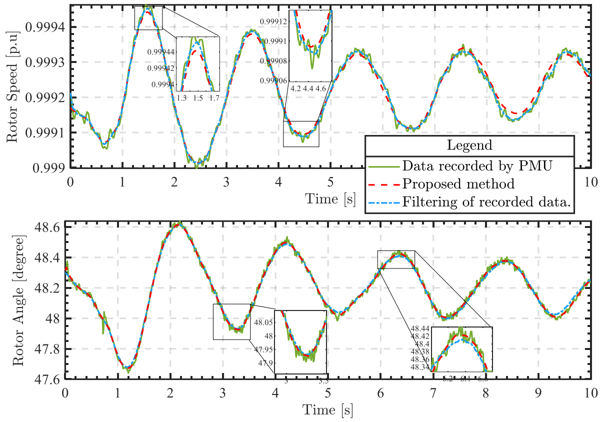

Fig.11. Estimation results based on real-world data

Category A: The results estimated using the proposed method.

Category B: The results estimated using the full-order model.

Category C: The results estimated using the 4th-order model.

Category D: The results estimated using the full-order model under the condition of parameter uncertainty.

The statistical results of the deviation between the estimated values and the true values of the rotor speed and rotor angle under the four scenarios are shown in Tab. I and Tab. II. In Fig. 9 and Fig. 10, the violin plot is employed to illustrate the distribution of rotor angle estimation errors for generators G3 and G7 under the four different categories.

Upon careful observation of the data in Tab. I and Tab. II, along with the violin plots shown in Fig. 9 and Fig. 10, it is seen that the values of NRMSE and RAE in Scenario D are the highest. Additionally, the values of NRMSE and RAE in Categories C and D are significantly higher than those in Categories A and B. Moreover, the values of R^2 in Categories C and D are clearly smaller than those in Categories A and B. According to the definitions of NRMSE, RAE and R^2, the estimation results in Categories C and D are relatively poor. Additionally, as depicted in Fig. 9 and Fig. 10, the distribution of estimation errors for Categories C and D is more dispersed over a larger range than those of Categories A and B. The above analysis indicates that Categories C and D correspond precisely to model mismatch and parameter uncertainty, respectively. This further confirms the crucial importance of the model and its parameters in dynamic state estimation.

The data-driven DSE scheme based on a linear predictor proposed in this paper, corresponding to Category A, exhibits the smallest values of NRMSE and RAE and the highest value of R^2. Here, the superior performance of the proposed data-driven scheme in DSE is verified from the perspective of estimation errors.

TABLE III
STATISTICAL RESULTS OF ESTIMATION ERROR BASED ON REAL-WORLD DATA

State variable	Reference	NRMSE	RAE	R^2
Rotor Speed	Measurement data	0.1242	0.1114	0.9845
	Filtered measurement	0.0888	0.0773	0.9921
Rotor Angle	Measurement data	0.0877	0.0823	0.9929
	Filtered measurement	0.0512	0.0476	0.9973

B. Real-world measurements

To further validate the practical engineering value of the proposed data-driven DSE scheme based on a linear predictor, actual data of the synchronous generator rotor angle and speed recorded by PMUs in a real power grid after a fault event were selected to analyze the estimation performance of the proposed method. The rotor angle and speed were estimated for 10 seconds after the fault, and the results are shown in Fig. 11. Note that the model and corresponding parameters of the generator and its control system, including the exciter and turbine-governor, are unknown. Therefore, we did not analyze the DSE methods based on either full-order or reduced-order models. The method proposed in this paper allows DSE for the synchronous generator with measurements only.

Due to the presence of measurement noise, the data curves recorded by the PMUs contain irregular "spikes". Filtering techniques were applied to smooth out the data curves for a more refined representation. The DSE results obtained using the proposed data-driven method were compared with the recorded data from the PMUs, as well as the filtered data, as shown in Fig. 11, which demonstrates that the data-driven method proposed in this paper is capable of effectively tracking and predicting the rotor angle and rotor speed, especially with the filtered data.

Similarly, the statistical analysis of the estimation errors for the dynamic state estimation results based on the data recorded by the PMUs was conducted and is presented in Tab. III. As shown in Tab. III, the values of NRMSE and RAE are small and approach 0, while the value of R^2 is close to 1.0. This result further validates the effectiveness of the proposed method for the measurement data recorded by the PMUs, which highlights the significant engineering application value and potential of the proposed method.

IV. Conclusion

In this paper, we present a purely data-driven dynamic state estimation (DSE) framework based on the Koopman operator-based linear predictor. The framework offers an alternative approach to traditional DSE methods by leveraging the power of the Koopman operator in analyzing and predicting the behavior of nonlinear dynamic systems.

We evaluated the effectiveness of our framework through experiments on both numerical simulations and real-world datasets of complex power systems. The results demonstrate that our data-driven approach outperforms traditional model-based approaches in terms of accuracy and robustness. The proposed framework exhibits superior performance in handling nonlinear dynamic systems, making it highly applicable to a wide range of practical cases. The error analysis of the estimation results, validated from a statistical perspective, provides compelling evidence for the superiority of the proposed method over traditional model-based DSE methods.

Our approach opens up new opportunities for analysis and prediction in fields such as control systems, robotics, and signal processing. Future work could focus on further improving the computational efficiency of the framework and exploring its applicability to other domains beyond power systems. Additionally, investigating the performance of the framework in the presence of noisy or incomplete data would be valuable for real-world implementation.

III. REFERENCE

[1] M. Z. El-Sharafy, S. Saxena and H. E. Farag, "Optimal Design of Islanded Microgrids Considering Distributed Dynamic State Estimation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1592-1603, March. 2021

[2] M. Rostami and S. Lotfifard, "Distributed Dynamic State Estimation of Power Systems," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3395-3404, Aug. 2018

[3] Zhihui Hong, Luping Xu, Junghui Chen, "Particle filter combined with data reconciliation for nonlinear state estimation with unknown initial conditions in nonlinear dynamic process systems," *ISA Transactions*, vol. 103, pp. 203-214, 2020

[4] Y. Xing and C. Lv, "Dynamic State Estimation for the Advanced Brake System of Electric Vehicles by Using Deep Recurrent Neural Networks," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 11, pp. 9536-9547, Nov. 2020

[5] Jun Hu, Zidong Wang, Guo-Ping Liu, Chaoqing Jia, Jonathan Williams, "Event-triggered recursive state estimation for dynamical networks under randomly switching topologies and multiple missing measurements," *Automatica*, vol. 115, 108908, 2020

[6] Harikumar Kandath, Md Meftahul Ferdaus,"PASE: An autonomous sequential framework for the state estimation of dynamical systems," *Expert Systems with Applications*, vol. 215, 2023

[7] J. Zhao et al., "Roles of Dynamic State Estimation in Power System Modeling, Monitoring and Operation," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2462-2472, May 2021

[8] Z. Zheng, J. Zhao, L. Mili and Z. Liu, "Robust Unscented Unbiased Minimum-Variance Estimator for Nonlinear System Dynamic State Estimation With Unknown Inputs," *IEEE Signal Processing Letters*, vol. 27, pp. 376-380, 2020

[9] Sesidhar D.V.S.R., Chandrashekhar Badachi, Robert C. Green II, "A review on data-driven SOC estimation with Li-Ion batteries: Implementation methods & future aspirations," *Journal of Energy Storage*, vol. 72, Part C, 2023, 108420

[10] 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

[11] J. Zhao et al., "Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188-3198, July 2019

[12] Bar-Shalom Y, Li X R, Kirubarajan T. Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. New York: John Wiley and Sons, 2001

[13] A. Paul, I. Kamwa and G. Jóos, "Centralized Dynamic State Estimation Using a Federation of Extended Kalman Filters With Intermittent PMU Data From Generator Terminals," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6109-6119, Nov. 2018.

[14] Y. Zhang, M. Li, Y. Zhang, Z. Hu, Q. Sun and B. Lu, "An Enhanced Adaptive Unscented Kalman Filter for Vehicle State Estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-12, 2022

[15] M. Zhu, H. Liu, J. Zhao, B. Tan, T. Bi and S. S. Yu, "Dynamic State Estimation for DFIG with Unknown Inputs Based on Cubature Kalman Filter and Adaptive Interpolation," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 4, pp. 1086-1099, July 2023

[16] J. Zhao, "Dynamic State Estimation With Model Uncertainties Using H∞ Extended Kalman Filter," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1099-1100, Jan. 2018.

[17] A. H. Chughtai, U. Akram, M. Tahir and M. Uppal, "Dynamic State Estimation in the Presence of Sensor Outliers Using MAP-Based EKF," *IEEE Sensors Letters*, vol. 4, no. 4, pp. 1-4, April 2020.

[18] Y. Wang, Z. Yang, Y. Wang, Z. Li, V. Dinavahi and J. Liang, "Resilient Dynamic State Estimation for Power System Using Cauchy-Kernel-Based Maximum Correntropy Cubature Kalman Filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-11, 2023.

[19] J. Zhao and L. Mili, "A Decentralized H-Infinity Unscented Kalman Filter for Dynamic State Estimation Against Uncertainties," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 4870-4880, Sept. 2019.

[20] Shengyong Liu, Dan Deng, Shunli Wang, "Dynamic adaptive square-root unscented Kalman filter and rectangular window recursive least square method for the accurate state of charge estimation of lithium-ion batteries," *Journal of Energy Storage*, vol. 67, 2023, 107603

[21] Jing Zhang, Tianshu Bi, Hao Liu, "Dynamic state estimation of a grid-connected converter of a renewable generation system using adaptive cubature Kalman filtering," *International Journal of Electrical Power & Energy Systems*, vol. 143, 2022, 108470

[22] Stone L D, Streit R L, Anderson S L. Introduction to Bayesian Tracking and Particle Filters. Springer Nature, 2023.

[23] S. S. Yu, J. Guo, T. K. Chau, "An Unscented Particle Filtering Approach to Decentralized Dynamic State Estimation for DFIG Wind Turbines in Multi-Area Power Systems," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2670-2682, July 2020.

[24] N. Zhou, D. Meng and S. Lu, "Estimation of the Dynamic States of Synchronous Machines Using an Extended Particle Filter," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4152-4161, Nov. 2013

[25] K. Li and J. C. Príncipe, "Functional Bayesian Filter," *IEEE Transactions on Signal Processing*, vol. 70, pp. 57-71, 2022

[26] Zhang, Bin, and Yung C. Shin. "A Gaussian mixture filter with adaptive refinement for nonlinear state estimation." *Signal Processing*, vol. 201, 2022, 108677.

[27] B. Uzunoğlu and M. A. Ülker, "Maximum Likelihood Ensemble Filter State Estimation for Power Systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 9, pp. 2097-2106, Sept. 2018

[28] Y. Cui and R. Kavasseri, "A Particle Filter for Dynamic State Estimation in Multi-Machine Systems With Detailed Models," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3377-3385, Nov. 2015

[29] B. Tan, J. Zhao and M. Netto, "A General Decentralized Dynamic State Estimation With Synchronous Generator Magnetic Saturation," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 960-963, Jan. 2023

[30] A. Surana and A. Banaszuk, "Linear observer synthesis for nonlinear systems using Koopman Operator framework", *IFAC-Papers OnLine*, vol. 49, no. 18, pp. 716-723, 2016.

[31] Milan Korda, Igor Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149-160, 2018.

[32] Ali Tavasoli, Teague Henry, Heman Shakeri, "A purely data-driven framework for prediction, optimization, and control of networked processes," *ISA Transactions*, vol. 138, pp. 491-503, 2023.

[33] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: Theory and applications," *Journal of Computational Dynamics*, vol. 1, no. 2, pp. 391–421, 2014.

[34] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A Data–Driven Approximation of the Koopman Operator: Extending Dynamic Mode

Decomposition," *Journal of Nonlinear Science*, vol. 25, no. 6, pp. 1307– 1346, Dec 2015.

[35] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic Mode Decomposition with Control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142-161, 2016.

[36] IEEE PES TF on Benchmark System for Stability Controls, Benchmark systems for small-signal stability analysis and control; 2015.

[37] "Power System Toolbox," in *Power System Dynamics and Stability: With Synchrophasor Measurement and Power System Toolbox* 2e, 2017, pp. 305-325.

[38] Jadon, Aryan, Avinash Patil, and Shruti Jadon. "A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting." arXiv preprint arXiv:2211.02989 (2022).

Di Bucchianico, Alessandro. "Coefficient of determination (R^2)." Encyclopedia of statistics in quality and reliability (2008)

Deyou Yang (M'16) received the B.S. and M.S. degrees from Northeast Electric Power University, Jilin, China, in 2005 and 2009, respectively, and the Ph.D. degree from North China Electric Power University, Beijing, China, in 2014, all in electrical engineering.

He is currently a Professor of electrical engineering with Harbin University of Science and Technology. From 2009 to 2010, he was a Research Assistant with the Hong Kong Polytechnic University, Hong Kong. From 2014 to 2023, he was a Professor of electrical engineering with Northeast Electric Power University. His current research interests include power system stability analysis and control.

Han Gao received the B.S. degree and the Ph.D. degree from Northeast Electric Power University, Jilin, China, in 2017 and 2023 respectively.
She is currently an assistant of electrical engineering with Harbin University of Science and Technology. Her research interests include the area of machine learning, and power system stability analysis and control.

Zhe Chen (Fellow, IEEE) received the B.Eng. and M.Sc. degrees in electrical engineering from the Northeast China Institute of Electric Power Engineering, Jilin, China, the M.Phil. degree in power electronic from Staffordshire University, Stoke-on-Trent, U.K., and the Ph.D. degree in power and control from the University of Durham, Durham, U.K.

He is currently a Full Professor with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. He is the Leader of Wind Power System Research Program, Department of Energy Technology, Aalborg University and the Danish Principle Investigator for the Wind Energy of Sino Danish Centre for Education and Research.

Yanling Lv was born in Heilongjiang, China, in 1975. She received the Ph.D. degree in electrical engineering from the Harbin University of Science and Technology (HUST), Harbin, China, in 2010, respectively. Currently, she is a Professor with HUST. Her research interests include the operation and protection of large generators, generators and its system operation analyses, power system fault, and protection analyses.

Lixin Wang received the B.S., M.S. and Ph.D. degrees from Northeast Electric Power University, Jilin, China, in 2014, 2017, and 2021, respectively. Her research focuses on power system stability analysis and control.