Deep Learning Based Time-Varying Parameter Identification for System-Wide Load Modeling

Mingjian Cui, *Senior Member, IEEE*, Mahdi Khodayar, *Student Member, IEEE*, Chen Chen, *Member, IEEE*, Xinan Wang, *Student Member, IEEE*, Ying Zhang, *Student Member, IEEE*, and Mohammad E. Khodayar, *Senior Member, IEEE*

***Abstract*—The integration of uncertain power resources is causing more challenges for traditional load modeling research. Parameters identification of load modeling is impacted by a variety of load components with time-varying characteristics. This paper develops a deep learning based time-varying parameter identification (TVPI) model for composite load modeling (CLM) with ZIP load and induction motor (IM). A multimodal long short-term memory (M-LSTM) deep learning method is used to estimate all the time-varying parameters of CLM considering system-wide measurements. It contains a multimodal structure that makes use of different modalities of the input data to accurately estimate time-varying load parameters. An LSTM network with flexible number of temporal states is defined to capture powerful temporal patterns from the load parameters and measurements time series. The extracted features are further fed to a shared representation layer to capture the joint representation of input time series data. This temporal representation is used in a linear regression model to estimate time-varying load parameters at the current time. Numerical simulations on the 23- and 68-bus systems verify the effectiveness and robustness of the proposed M-LSTM method. Also, the optimal lag values of parameters and measurements as input variables are solved.**

***Index Terms*—Composite load model, deep learning, long short-term memory, parameter identification.**

NOMENCLATURE

A. Parameter and State Variables:

$a_{P,t}, b_{P,t}, c_{P,t}$	Percentages for ZIP active power at t.
$a_{Q,t}, b_{Q,t}, c_{Q,t}$	Percentages for ZIP reactive power at t.
$i_{d,t}, i_{q,t}$	d- and q-axis stator current at t.
$P_{IM,t}$	Active power of induction motor at t.
$P_{ZIP,0}$	Base active power of ZIP load.
$P_{ZIP,t}$	Active power of ZIP load at t.
P_t, Q_t	Measured active and reactive power at t.
$Q_{IM,t}$	Reactive power of induction motor at t.
$Q_{ZIP,0}$	Base reactive power of ZIP load.
$Q_{ZIP,t}$	Reactive power of ZIP load at t.
$r_{R,t}, r_{S,t}$	Rotor and stator resistance at t.
s_t	Rotor slip at t.
$u_{d,t}, u_{q,t}$	d- and q-axis bus voltage at t.
V_0, V_t	Nominal and measured voltage magnitude at t.
$v'_{d,t}, v'_{q,t}$	d- and q-axis transient voltage at t.
$x_{m,t}, x'_t$	Magnetizing and short circuit reactance at t.
$x_{R,t}, x_{S,t}$	Rotor and stator reactance at t.
H_t	Inertia constant of induction motor at t.
d_ξ, d_M	Dimensions of parameters and measurements.
$\varepsilon_{\xi M}, \varepsilon_{\xi\xi}$	Estimation error/noise.
k_ξ, k_M	Lag values considered for the time series of parameter and measurement variables.
N	Total number of load buses in the system.
$C(i)$	Updated state of the memory cell at round i.
$\widetilde{C}(i)$	Temporal updated state of the memory cell at round i.

B. Sets, Vectors, Matrices, and Functions:

$\boldsymbol{\xi}^t_{d_\xi}, \boldsymbol{M}^t_{d_M}$	Time-varying parameter variables and measurement variables at time t.
$\widehat{\boldsymbol{\xi}}^t_{d_\xi}$	Estimations of parameter variables at time t.
$\boldsymbol{M}^{i,t}_{d_\xi}$	Measurement vector at time t on load bus i.
$\boldsymbol{\xi}_{d_\xi \times k_\xi}$	Parameter variables with the lag k_ξ.
$\boldsymbol{M}_{d_M \times (k_M+1)}$	Measurement variables with the lag k_M.
$\boldsymbol{f}_{\xi M}$	Nonlinear function between parameter and measurement variables.
$\boldsymbol{F}_{\xi M}$	Nonlinear function between parameter and measurement variables for one targeted bus.
$\boldsymbol{F}_M$	Nonlinear function between measurements on the targeted bus and other system-wide buses.
$\mathbb{F}_{\xi M}$	Nonlinear function between time-varying parameters and measurements on system-wide load buses.
$\boldsymbol{h}_\xi(i)$,$\boldsymbol{h}_M(i)$	Latent variable vector obtained from the LSTM at the ith round of network update.
u^t_ξ, u^t_M	Average of temporal latent variables at time t.
$\boldsymbol{I}(i)$,$\boldsymbol{f}(i)$,$\boldsymbol{o}(i)$	Vectors of the input, forget, and output gates.
$T(\cdot)$	Nonlinear sigmoidal activation function.
$\widetilde{T}(\cdot)$	Hyperbolic tangent function.
$\widehat{T}(\cdot)$	Nonlinear tangent hyperbolic activation for mapping the memory content to the latent space.
$\boldsymbol{J}$	Tunable weight matrices and $\boldsymbol{J} = \{W_i, W_f, W_o, U_i, U_f, U_o\}$.
$\boldsymbol{B}$	Bias vectors and $\boldsymbol{B} = \{b_i, b_f, b_o\}$.
$\boldsymbol{S}$	Tunable parameters of the memory cell and $\boldsymbol{S} = \{W_c, U_c, b_c\}$.
$\boldsymbol{\varphi}$	Total parameters of M-LSTM.
$g(\cdot)$	Linear regression model.
$E(\cdot)$	Supervised loss function.

M. Cui, M. Khodayar, X. Wang, Y. Zhang, and M. E. Khodayar are with the Department of Electrical and Computer Engineering, Southern Methodist University, Dallas, TX, 75275 USA (e-mail: {mingjiancui, mahdik, xinanw, yzhang1, mkhodayar}@smu.edu).

C. Chen is with the Energy Systems Division, Argonne National Laboratory, Argonne, IL, 60439 USA (e-mail: morningchen@anl.gov).

Manuscript received, 2018.

I. Introduction

ACCURATE time-varying load modeling is becoming more and more important due to the increasing integration of uncertain power resources. The common load modeling structures consist of static model, dynamic model, and composite model. The composite load modeling (CLM) with specific parameters has been widely used since it considers both the static and the dynamic characteristics of static model and dynamic model [1], [2]. The more accurate load modeling can even supplement the conventional load forecasting [3]–[5] under some particular circumstance with missing data. However, due to the high frequency changes caused by uncertain power resources, parameters of CLM present more and more time-varying characteristics [6].

Both statistical and heuristic techniques have been widely used to identify load parameters in recent years. For statistical techniques, Hiskens [7] used nonlinear least-square based method to estimate parameters by the best fit between measurements and model response. Zhao *et al*. [8] utilized maximum likelihood approach to estimate parameters for power system state dynamics. Kock *et al*. [9] used the gradient method to estimate parameters of the induction motor. Other techniques such as Kalman filter and vector fitting have also been used in the literature for the parameter identification of dynamic load modeling. Rouhani and Abur [10] used an unscented Kalman filter to track the unknown parameters of the exponential dynamic load model. Zhang *et al*. [11] used the trajectory sensitivity vector fitting to obtain a feature vector for parameters in the time-varying response of the load. Ma *et al*. [12] utilized the support vector fitting of load data for load modeling with a strong generalization capability to describe the time-varying characteristics of the whole dataset.

For heuristic techniques, neural networks [13], [14] were used to describe the complicated time-varying behavior of load modeling. The simulated annealing algorithm [15] was used to find the global optimum for identifying load parameters. However, most of these algorithms do not consider the time-varying characteristics of load parameters. Ignoring the time-varying characteristics of load modeling can lead to erroneous results in the transient stability studies of power system operations. In other words, the estimated parameters at the current time are significantly related with not only the current measurements (e.g., bus voltage, active power, and reactive power) but also the previous parameters of load modeling.

Recently, time-varying load modeling has been focused on by worldwide researchers [16]. For example, Wang *et al*. [6] proposes a robust time-varying parameter identification technique for CLM in a batch-model regression form. Huang *et al*. [17] developed time-varying load models to determine the photovoltaic penetration level in a distribution network. Wang *et al*. [18] developed a time-varying exponential load model to assess conservation voltage reduction effects. However, all the aforementioned methods identify load parameters by only considering measurements on a corresponding bus where this load is connected with. Since all the loads are integrated into one power system, a perturbation in the system can cause different impacts on the parameter identification of each load. Thus, there exists a relationship between the targeted load and other loads that are connected with the same system. In addition, the conventional load modeling research cannot characterize the impacts of ambient noises, which may reduce the estimation accuracy of load parameters. In a prescribed power system, the ambient noises of estimated parameters are related with not only measurements of the targeted load but also those of other system-wide loads. By considering ambient noises, it is expected to improve the estimation accuracy of load parameters.

The combined time-varying parameter identification (TVPI) of the system-wide load modeling is usually too complicated to be expressed by the component-based parametric model. The sophisticated behavior and attendant challenge with identifying accurate time-varying parameters causes insufficient and inaccurate results in load modeling problems. It is still challenging to use a specific mathematical or physical model to characterize both the time-varying and system-wide relationship.

With the rapid development of Artificial Intelligence (AI) [19], it is possible to use deep learning techniques to represent the complicated nonlinear relationship of the combined time-varying characteristics of system-wide load modeling. Though several artificial neural networks (ANN) methods have been explored in this topic, such as feed-forward neural networks [13], [14], [20], [21] and fuzzy neural networks [22], [23], these models cannot efficiently learn the temporal behavior of time series data corresponding to time-varying load parameters and measurements due to the lack of a recurrent structure. Moreover, such models contain shallow neural architectures with few hidden layers. Hence, the generalization capability of these approaches are relatively low. From the aspect of compact representation, it is shown that deep hypothesis spaces have exponential advantages over shallow spaces in terms of representing target functions [24]. More specifically, when an implicit assumption in a hypothesis space is approximately satisfied by the target function (for example, the deep composition of piece-wise linear transformations), deep networks yield much higher accuracy compared to shallow models.

To bridge the gap in load modeling, we seek to address two critical questions for power system operators. Is it possible to use deep learning techniques to consider both the time-varying and system-wide relationship for load modeling? Can deep learning techniques really improve the accuracy of TVPI compared with other methods? To this end, this paper develops a novel two deep learning architectures for TVPI of load modeling, which is capable of capturing meaningful temporal patterns from the time series of load parameters and measurements. The main contributions of this paper include: (i) developing a deep neural architecture with high generalization capacity and expressivity for the TVPI of load modeling; (ii) presenting a multi-modal structure with a joint representation layer for the first time in the area of load modeling; and (iii) applying the multi-modal long short-term memory (M-LSTM) method to find the optimal lag values of load parameters and measurements.

The organization of this paper is as follows. In Section II, the system-wide TVPI modeling of CLM is briefly introduced.

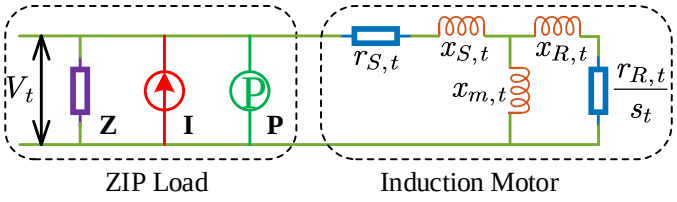


Fig. 1. Equivalent circuit of CLM consisting of ZIP and IM model.

Section III presents the detailed deep learning technique based on the multi-modal long short-term memory. Case studies and result analysis are discussed in Section IV. Concluding remarks are summarized in Section V.

II. System-Wide TVPI Modeling

A. Time-Varying Parameters

1) Time-Varying ZIP Model: One of the typical static load model is the ZIP model, which mainly consists of three parts. The first part is the constant impedance (Z) component. The second part is the constant current (I) component. The third part is the constant power (P) component. Percentages of three components are assumed to be time-varying due to whether conditions or customer behaviors. The ZIP model is mathematically formulated by:

$$P_{ZIP,t} = a_{P,t}\left(V_t/V_0\right)^2 + b_{P,t}\left(V_t/V_0\right) + c_{P,t} \tag{1}$$

$$Q_{ZIP,t} = a_{Q,t}\left(V_t/V_0\right)^2 + b_{Q,t}\left(V_t/V_0\right) + c_{Q,t} \tag{2}$$

where percentage parameters satisfy $a_{P,t}+b_{P,t}+c_{P,t}=1$ and $a_{Q,t}+b_{Q,t}+c_{Q,t}=1$ at any time period t.

2) Time-Varying IM Model: The time-varying induction motor (IM) models are defined with a similar approach to the synchronous machine. In this paper, the typical singe-cage rotor model is used by additionally considering the time-varying impacts of weather conditions and customer behaviors. The simplified electrical circuit used for the single-cage induction motor is shown in Fig. 1. The differential algebraic equations (DAEs) of state variables are formulated by:

$$\frac{dv'_{d,t}}{dt} = \frac{-r_{R,t}}{x_{R,t}+x_{m,t}}\left(v'_{d,t} + \frac{x_{m,t}^2}{x_{R,t}+x_{m,t}} i_{q,t}\right) + s_t v'_{q,t} \tag{3}$$

$$\frac{dv'_{q,t}}{dt} = \frac{-r_{R,t}}{x_{R,t}+x_{m,t}}\left(v'_{q,t} - \frac{x_{m,t}^2}{x_{R,t}+x_{m,t}} i_{d,t}\right) - s_t v'_{d,t} \tag{4}$$

$$\frac{ds_t}{dt} = \frac{1}{2H_t}\left[T_{m0}\left(1-s_t\right)^2 - v'_{d,t} i_{d,t} - v'_{q,t} i_{q,t}\right] \tag{5}$$

where the d-axis stator current $i_{d,t}$ and the q-axis stator current $i_{q,t}$ are given by:

$$i_{d,t} = \frac{r_{S,t}\left(u_{d,t} - v'_{d,t}\right) + x'_t\left(u_{q,t} - v'_{q,t}\right)}{r_{S,t}^2 + x_t'^2} \tag{6}$$

$$i_{q,t} = \frac{r_{S,t}\left(u_{q,t} - v'_{q,t}\right) - x'_t\left(u_{d,t} - v'_{d,t}\right)}{r_{S,t}^2 + x_t'^2} \tag{7}$$

where the quadratic sum of the d- and q-axis bus voltages $u_{d,t}$ and $u_{q,t}$ should equal the square of the measured voltage, given by:

$$V_t^2 = u_{d,t}^2 + u_{q,t}^2 \tag{8}$$

The short circuit reactance x'_t is formulated by:

$$x'_t = x_{S,t} + \frac{x_{m,t}x_{R,t}}{x_{m,t}+x_{R,t}} \tag{9}$$

The active and reactive power of the IM model can be formulated by using the states, time-varying parameters, and bus voltage variables, given by:

$$\begin{aligned} P_{IM,t} = \big[r_{S,t}\big(u_{d,t}^2 + u_{q,t}^2 - u_{d,t}v'_{d,t} - u_{q,t}v'_{q,t}\big) - \\ x'_t\left(u_{d,t}v'_{q,t} - u_{q,t}v'_{d,t}\right)\big]/\left(r_{S,t}^2 + x_t'^2\right) \end{aligned} \tag{10}$$

$$\begin{aligned} Q_{IM,t} = \big[x'_t\big(u_{d,t}^2 + u_{q,t}^2 - u_{d,t}v'_{d,t} - u_{q,t}v'_{q,t}\big) - \\ r_{S,t}\left(u_{d,t}v'_{q,t} - u_{q,t}v'_{d,t}\right)\big]/\left(r_{S,t}^2 + x_t'^2\right) \end{aligned} \tag{11}$$

Based on models (1)–(2) and (10)–(11), the active and reactive power of the time-varying composite ZIP and IM model can be formulated by:

$$P_t = P_{ZIP,t} + P_{IM,t} \tag{12}$$

$$Q_t = Q_{ZIP,t} + Q_{IM,t} \tag{13}$$

3) Identifying Time-Varying Parameters: Assuming that the sets of time-varying parameter variables and measurement variables are denoted as:

$$\boldsymbol{\xi}_{d_\xi}^t = [r_{S,t}, x_{S,t}, x_{m,t}, x_{R,t}, r_{R,t}, H_t, a_{P,t}, b_{P,t}, a_{Q,t}, b_{Q,t}] \tag{14}$$

$$\boldsymbol{M}_{d_M}^t = [P_t, Q_t, V_t] \tag{15}$$

where d_ξ and d_M are the dimensions of parameter variables $\boldsymbol{\xi}$ and measurement variables $\boldsymbol{M}$, i.e., d_ξ=10 and d_M=3.

Based on the mathematical models in (1)–(13), the highly complex nonlinear relationship between parameter variables and measurement variables can be generalized as:

$$\boldsymbol{\xi}_{d_\xi}^t = \boldsymbol{f}_{\xi M}\left(\boldsymbol{M}_{d_M}^t\right) + \varepsilon_{\xi M} \tag{16}$$

Normally, values of time-varying parameter variables fluctuate slightly within a short time period. Thus, the relationship between adjacent time intervals can be expressed as:

$$\boldsymbol{\xi}_{d_\xi}^t = \boldsymbol{\xi}_{d_\xi}^{t-1} + \varepsilon_{\xi\xi 1} = \boldsymbol{\xi}_{d_\xi}^{t-2} + \varepsilon_{\xi\xi 2} = \cdots = \boldsymbol{\xi}_{d_\xi}^{t-k_\xi} + \varepsilon_{\xi\xi k} \tag{17}$$

where k_ξ is the maximum lag value considered for the time series of parameter variables $\boldsymbol{\xi}$. Similarly, considering the time-series characteristics of measurements, the model in (16) can be expanded as:

$$\begin{aligned} \boldsymbol{\xi}_{d_\xi}^t &= \boldsymbol{f}_{\xi M}\left(\boldsymbol{M}_{d_M}^t\right) + \varepsilon_{\xi M} = \boldsymbol{f}_{\xi M1}\left(\boldsymbol{M}_{d_M}^{t-1}\right) + \varepsilon_{\xi M1} \\ &= \cdots = \boldsymbol{f}_{\xi Mk}\left(\boldsymbol{M}_{d_M}^{t-k_M}\right) + \varepsilon_{\xi Mk} \end{aligned} \tag{18}$$

where k_M is the maximum lag values considered for the time series of measurements $\boldsymbol{M}$.

Based on (17) and (18), a nonlinear function $\boldsymbol{F}_{\xi M}$ is used to represent this relationship with lag values:

$$\boldsymbol{\xi}_{d_\xi}^t = \boldsymbol{F}_{\xi M}\left(\boldsymbol{\xi}_{d_\xi}^{t-1}, \cdots, \boldsymbol{\xi}_{d_\xi}^{t-k_\xi}, \boldsymbol{M}_{d_M}^t, \cdots, \boldsymbol{M}_{d_M}^{t-k_M}\right) \tag{19}$$

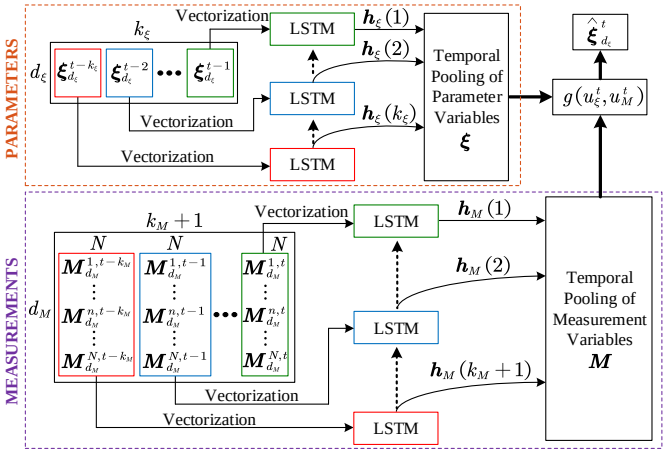


Fig. 2. Structure of the proposed multi-modal LSTM.

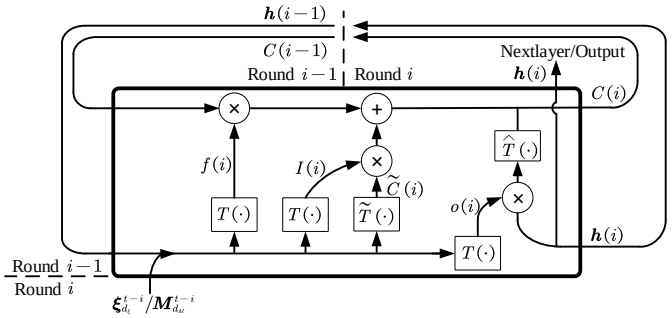


Fig. 3. Structure of the proposed LSTM block.

Normally, k_ξ and k_M cannot be too large since the current parameters have strong relationships with those in a short-term due to the time series characteristics. In this paper, a supervised learning algorithm is used to find the optimal values of k_ξ and k_M. When k_ξ=1 and k_M=0, the model considering (17) and (18) is returned to the conventional TVPI model [6] which only considers the previous parameters at time step t-1 and the current measurements at time step t.

B. System-Wide TVPI

All of the parameters and measurements in Section II-A are considered on the targeted bus where the load parameters are to be estimated. Normally, measurements on the targeted bus are related with those on system-wide load buses based on the power flow analysis. A nonlinear function $\boldsymbol{F}_M$ is used to represent this relationship:

$$\boldsymbol{M}_{d_M}^{i,t} = \boldsymbol{F}_M\left(\boldsymbol{M}_{d_M}^{1,t}, \cdots, \boldsymbol{M}_{d_M}^{N,t}\right) \tag{20}$$

where i is the index of the targeted load bus. N is the total number of load buses in the system.

By integrating (19) and (20), the system-wide TVPI model can be expressed by a highly nonlinear function $\mathbb{F}_{\xi M}$ in (21). In this function, symbols of $\boldsymbol{\xi}_{d_\xi \times k_\xi}$ and $\boldsymbol{M}_{d_M \times (k_M+1)}$ are respectively given by:

$$\boldsymbol{\xi}_{d_\xi \times k_\xi} = \left(\boldsymbol{\xi}_{d_\xi}^{t-1}, \boldsymbol{\xi}_{d_\xi}^{t-2}, \cdots, \boldsymbol{\xi}_{d_\xi}^{t-k_\xi}\right) \tag{22}$$

$$\boldsymbol{M}_{d_M \times (k_M+1)} = \left(\cdots, \boldsymbol{M}_{d_M \times (k_M+1)}^{n}, \cdots, \boldsymbol{M}_{d_M \times (k_M+1)}^{N}\right) \tag{23}$$

$$\boldsymbol{M}_{d_M \times (k_M+1)}^{n} = \left(\boldsymbol{M}_{d_M}^{n,t}, \boldsymbol{M}_{d_M}^{n,t-1}, \cdots, \boldsymbol{M}_{d_M}^{n,t-k_M}\right) \quad n = 1, 2, \cdots, N \tag{24}$$

where $\boldsymbol{\xi}_{d_\xi}^{t-1}$ denotes the time-varying parameters vector of the targeted load at the previous time step t-1. $\boldsymbol{M}_{d_M \times (k_M+1)}^{n}$ denotes the measurements vector on bus n with the lag k_M. Its element $\boldsymbol{M}_{d_M}^{n,t}$ denotes the measurements vector at time t on bus n. However, this function cannot be analytically formulated and solved to estimate time-varying load parameters. Since the deep learning technique is significantly powerful in representing complex nonlinear relationship, it is used and introduced in the following description.

III. Deep Learning Based TVPI

Essentially, the TVPI of CLM is a mathematical representation of the highly nonlinear relationship between current parameters and previous ones as well as current measurements, i.e., bus voltage, active power, and reactive power. The task of time-varying parameter identification is to constitute a representation that describes time-varying characteristics especially for the transient period. Towards this end, the M-LSTM technique is developed to characterize the TVPI for system-wide load modeling.

The M-LSTM is a pattern recognition neural architecture capable of capturing highly nonlinear temporal features from multiple high dimensional time series. At each time step t, the M-LSTM takes 2-D tensors of load parameters vector $\boldsymbol{\xi}_{d_\xi \times k_\xi}$ and system-wide measurements vector $\boldsymbol{M}_{d_M \times (k_M+1)}$ as different modalities of the temporal input data. The structure of M-LSTM is shown in Fig. 2. An LSTM network [25] is assigned to each element of the historical parameters vector $\boldsymbol{\xi}$. At each training time step t, the LSTM observes $k_\xi \times d_\xi$ temporal element in $\boldsymbol{\xi}$ consecutively starting from the first one, that is $\boldsymbol{\xi}_{d_\xi}^{t-1}$ and going through all the elements until the last observations of $\boldsymbol{\xi}_{d_\xi}^{t-k_\xi}$. The latent variable vector obtained from the LSTM at the ith round of network update is denoted by $\boldsymbol{h}_\xi(i)$ which is the ith element of the parameter temporal feature tensor $\boldsymbol{h}_\xi$. When all the input vectors in $\boldsymbol{\xi}$ are observed,

$$\boldsymbol{\xi}_{d_\xi}^{t} = \mathbb{F}_{\xi M}\Big(\underbrace{\boldsymbol{\xi}_{d_\xi}^{t-1}, \cdots, \boldsymbol{\xi}_{d_\xi}^{t-k_\xi}}_{\boldsymbol{\xi}_{d_\xi \times k_\xi}}, \underbrace{\underbrace{\boldsymbol{M}_{d_M}^{1,t}, \cdots, \boldsymbol{M}_{d_M}^{1,t-k_M}}_{\boldsymbol{M}_{d_M \times (k_M+1)}^{1}}, \underbrace{\boldsymbol{M}_{d_M}^{2,t}, \cdots, \boldsymbol{M}_{d_M}^{2,t-k_M}}_{\boldsymbol{M}_{d_M \times (k_M+1)}^{2}}, \cdots, \underbrace{\boldsymbol{M}_{d_M}^{N,t}, \cdots, \boldsymbol{M}_{d_M}^{N,t-k_M}}_{\boldsymbol{M}_{d_M \times (k_M+1)}^{N}}}_{\boldsymbol{M}_{d_M \times (k_M+1)}}\Big) \tag{21}$$

the average of temporal latent variables at time step t is computed by:

$$u_{\xi}^{t}=\frac{1}{k_{\xi}}\sum_{i=1}^{k_{\xi}}\boldsymbol{h}_{\xi}\left(i\right) \tag{25}$$

Similarly with the aforementioned process of the parameter vector $\boldsymbol{\xi}$, the system-wide measurement vector $\boldsymbol{M}$ is repeatedly disposed. The average of temporal latent variables for measurements at time step t is computed by:

$$u_{M}^{t}=\frac{1}{k_{M}+1}\sum_{i=1}^{k_{M}+1}\boldsymbol{h}_{M}\left(i\right) \tag{26}$$

where $\boldsymbol{h}_{M}\left(i\right)$ is the latent variable vector for measurements obtained from the LSTM at the ith round of network update.

To describe the principle of the proposed LSTM in Fig. 2, Fig. 3 shows the structure of LSTM at the ith update round with the historical d_{ξ} dimensional input $\boldsymbol{\xi}_{d_{\xi}}^{t-i}$ and the corresponding hidden values of $\boldsymbol{h}_{\xi}\left(i\right)$ by taking parameters as an example. The same structure of LSTM can also be applied to the system-wide measurements. As can be seen, this network consists of memory blocks with memory cell units having self-connections aimed to maintain the temporal state corresponding to each update round. In addition to the memory cells, the LSTM blocks contain special multiplicative computational units named as the input, output, and forget gate. The input activation that is fed to the memory cell is controlled by the input gate. The magnitude of the output flow corresponding to the cell activations is decided by the output gate. The forget gate is capable to scale the current temporal LSTM state before adding it back to the LSTM memory using its self-recurrent weights. Here, $\boldsymbol{\xi}_{d_{\xi}}^{t-i}$ is the input at the update round i, and $\boldsymbol{h}_{\xi}\left(i\right)$ is the output of the current LSTM block. The latent vector corresponding to the LSTM state at round i-1, $\boldsymbol{h}_{\xi}\left(i-1\right)$, is fed to the LSTM in the next round in order to make a recurrent model to capture high-level temporal abstractions in the temporal data. Symbols of $\boldsymbol{I}(i)$, $\boldsymbol{f}(i)$, and $\boldsymbol{o}(i)$ represent the vectors of the input, forget, and output gates. The gates' outputs are calculated as:

$$\boldsymbol{I}(i)=T\left[W_{i}\boldsymbol{\xi}_{d_{\xi}}^{t-i}+U_{i}\boldsymbol{h}_{\xi}\left(i-1\right)+b_{i}\right] \tag{27a}$$

$$\boldsymbol{f}(i)=T\left[W_{f}\boldsymbol{\xi}_{d_{\xi}}^{t-i}+U_{f}\boldsymbol{h}_{\xi}\left(i-1\right)+b_{f}\right] \tag{27b}$$

$$\boldsymbol{o}\left(i\right)=T\left[W_{o}\boldsymbol{\xi}_{d_{\xi}}^{t-i}+U_{o}\boldsymbol{h}_{\xi}\left(i-1\right)+b_{o}\right] \tag{27c}$$

where $T[\cdot]$ is the nonlinear activation function and usually considered as a sigmoidal function for the input, output, and forget vectors. Here, the parameter space has a set of tunable weight matrices $\boldsymbol{J}=\{W_{i},W_{f},W_{o},U_{i},U_{f},U_{o}\}$ and bias vectors $\boldsymbol{B}=\{b_{i},b_{f},b_{o}\}$. The updated state of the memory cell at round i is denoted by $C(i)$. To obtain $C(i)$, the model partially forgets the existing memory content stored in $C(i-1)$ and adds the temporal update value $\widetilde{C}\left(i\right)$ which can be computed by:

$$\widetilde{C}\left(i\right)=\widetilde{T}\left[W_{c}\boldsymbol{\xi}_{d_{\xi}}^{t-i}+U_{c}\boldsymbol{h}_{\xi}\left(i-1\right)+b_{c}\right] \tag{28a}$$

$$C\left(i\right)=\boldsymbol{f}\left(i\right)C\left(i-1\right)+\boldsymbol{I}\left(i\right)\widetilde{C}\left(i\right) \tag{28b}$$

The set of tunable parameters of the memory cell is defined as $\boldsymbol{S}=\{W_{c},U_{c},b_{c}\}$ where W_{c} and U_{c} are the weights while b_{c} is a bias vector. Here, a hyperbolic tangent function is employed as the nonlinear activation $\widetilde{T}[\cdot]$ for the memory content update value in (28). Based on (28), the latent variable resulted by the LSTM at the ith round is computed as $\boldsymbol{h}\left(i\right)=\boldsymbol{o}\left(i\right)\times\widehat{T}\left[C\left(i\right)\right]$ where $\widehat{T}[\cdot]$ is a nonlinear tangent hyperbolic activation for mapping the memory content $C(i)$ to the latent space. To train the LSTM, the set of its total parameters $\boldsymbol{\varphi}=\boldsymbol{J}\cup\boldsymbol{B}\cup\boldsymbol{S}$ is trained by using the stochastic gradient descent algorithm at each time step t.

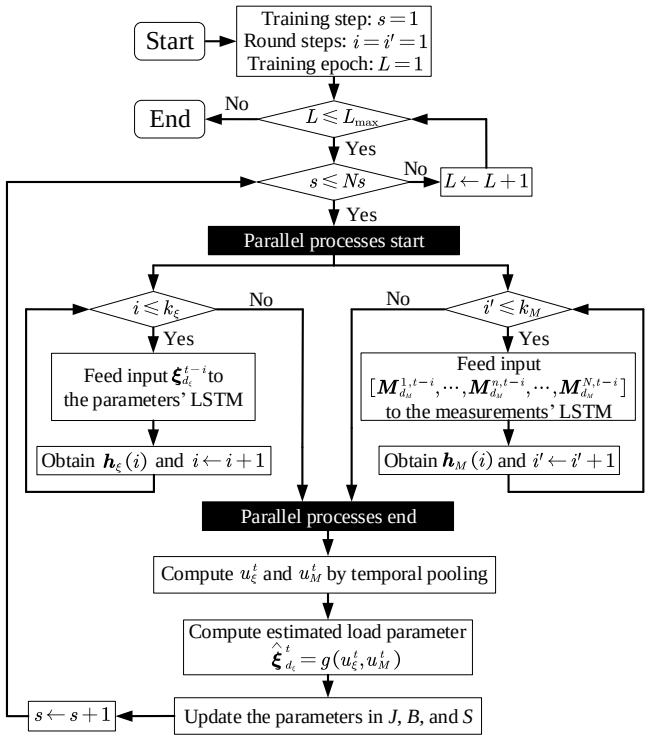


Fig. 4. Supervised learning algorithm of the proposed multi-modal LSTM architecture for time-varying TVPI.

Similar temporal feature extraction model is also applied for measurements $\boldsymbol{M}$ with $k_{M}+1$ temporal columns for each training step. To update the corresponding LSTM parameters, at each round i', a new representation vector $\boldsymbol{h}_{M}(i')$ is computed. The total temporal feature u_{M}^{t} in (26) is obtained by taking a temporal average over all values of the temporal representation vector $\boldsymbol{h}_{M}$ during $(k_{M}+1)$ rounds.

Based on (25) and (26), the temporal pooling is correspondingly generated for parameters and measurements. Finally, the temporal averages are put into a linear regression model $g(u_{\xi}^{t},u_{M}^{t})$ at each training step t to estimate the time-varying parameter values at the current time step t, i.e., $\widehat{\boldsymbol{\xi}}_{d_{\xi}}^{t}$, by using the historical parameter and measurement time series.

Fig. 4 demonstrates the flow chart of the supervised algorithm to update the proposed M-LSTM architecture. The feed-forward algorithm is executed on the LSTM models corresponding to the parameters and measurements in k_{ξ} and $k_{M}+1$ rounds, respectively. In this stage, vectors $\boldsymbol{h}_{\xi}(i)$ and $\boldsymbol{h}_{M}(i)$ are obtained. Then, the temporal pooling layer obtains the average temporal latent u_{ξ}^{t} and u_{M}^{t}. The linear regression is employed to output $\widehat{\boldsymbol{\xi}}_{d_{\xi}}^{t}$ as an estimation for the actual $\boldsymbol{\xi}_{d_{\xi}}^{t}$

at the current time step t. During the aforementioned process, the supervised loss function $E = \left[\boldsymbol{\xi}_{d_\xi}^t - \widehat{\boldsymbol{\xi}}_{d_\xi}^t\right]^2$ is used with the stochastic gradient descent (SGD) with Ns training samples and maximum epoch $L_{\max}$ to tune the proposed M-LSTM network.

IV. Case Studies

To evaluate the performance of the proposed methodology, it is tested on two test cases: 23- and 68-bus systems. The composite ZIP and IM loads are used to validate the proposed M-LSTM method. Four benchmarks including the Time Delay Neural Network (TDNN) [26], Nonlinear Autoregressive Neural Network with External Input (NARX) [27], Gated Recurrent Unit (GRU) network [28], and single model LSTM (S-LSTM) are used to compare with the proposed M-LSTM method. The learning rate is set to 10^{-3}. Moreover, to avoid overfitting, L2 regularization with the regularization coefficient $\lambda = 5\times10^{-3}$ is applied by using the error R_θ for each parameter θ of M-LSTM, added to the supervised loss function E:

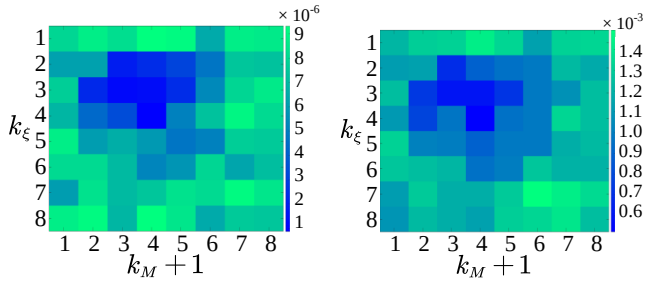
$$R_\theta = \frac{\lambda}{2}\|\theta\|^2, \quad \theta \in \boldsymbol{\varphi} \tag{29}$$

All case studies using the deep learning technique are carried out using a high-level neural networks API, Keras [29], on an Intel core-i7 CPU desktop with 32 GB of RAM memory and Geforce GTX 1080 graphic card. Six metrics are used to evaluate the performance of different methods, including: correlation coefficient, root mean square error (RMSE), maximum absolute error (MaxAE), mean absolute percentage error (MAPE), fourth root mean quartic error (4RMQE), and 95th percentile. A smaller value indicates a better estimation performance for most of the metrics, except for the correlation coefficient. Detailed information about these metrics can be found in [30], [31].

In Section IV-A and IV-B, the lag values k_M and k_ξ are first picked from the predefined domain. Then, the LSTM is trained on the training dataset using the chosen lag values. For both of 23- and 68-bus systems, the training set contains 80% of collected data samples. The optimal LSTM with the least validation error is chosen as the optimal model. The model is further evaluated on the testing dataset to assess the performance of LSTM with the chosen k_M and k_ξ in terms of RMSE and MAPE. The estimated load parameters using the LSTM model in the validation and testing datasets consist of 10% of the dataset gathered from the 23- and 68-bus systems, respectively. Error metrics using different methods for estimating load parameters in the validation dataset are quantitatively compared in Section IV-A and IV-B, respectively. In Section IV-C, the estimated load parameters in the testing dataset calculated in Section IV-B are put into the 68-bus system to estimate measurements, i.e., active power, reactive power, and voltage. The comparison between estimated measurements and real measurements can verify the effective usage of estimated time-varying parameters.

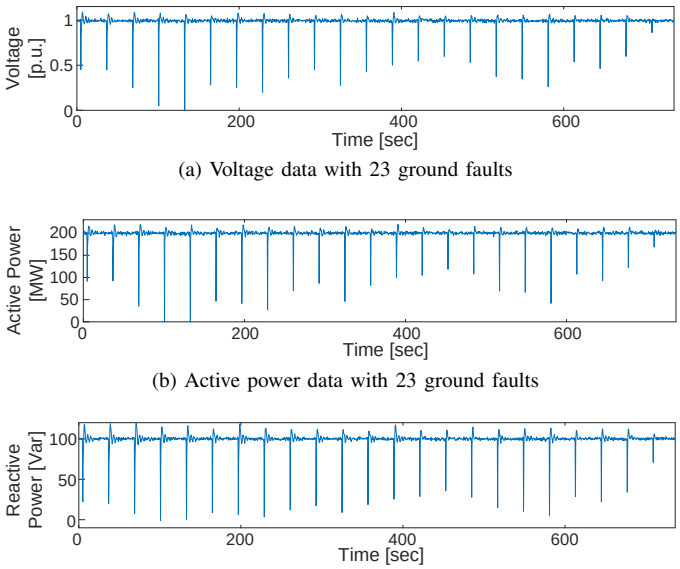
A. 23-bus System

The data of the first case is generated by the Siemens PSS/E 23-bus system [32], [33]. There is only one CLM with



(a) Validation RMSE (b) Validation MAPE

Fig. 5. Validation (a) RMSE and (b) MAPE of M-LSTM for the 23-bus system.



(a) Voltage data with 23 ground faults

(b) Active power data with 23 ground faults

(c) Reactive power data with 23 ground faults

Fig. 6. Time domain curves of voltage, active power, and reactive power on Bus 153 (the first load bus) with 23 ground faults in the 23-bus system.

time-varying parameters connected into this system. The mean values of these parameters $r_{S,t}$, $x_{S,t}$, $x_{m,t}$, $x_{R,t}$, $r_{R,t}$, H_t, $a_{P,t}$, $b_{P,t}$, $a_{Q,t}$, $b_{Q,t}$ are set as 0.05, 0.2, 4, 0.2, 0.01, 20, 0.3, 0.5, 0.2, and 0.4. For each parameter, a Gaussian random variable with zero mean and standard deviation of a hundredth of the mean value of the corresponding parameter is added to simulate parameter changes [6]. In this system, there are five load buses. Thus, the dimension of output variables is 10 (=10×1) and that of input variables is $5\times3\times(k_M+1)+10\times k_\xi$. To obtain sufficient data for training, 23 ground fault events are simulated on each bus. The sampling time is set as 0.4 s. For each event, the time horizon is set as 32 s. Time domain curves of voltage, active power, and reactive power on Bus 153 (the first load bus) with 23 ground faults are shown in Fig. 6.

Fig. 5 shows the validation RMSE and MAPE of M-LSTM for the 23-bus system, as a function of the number of temporal training steps, i.e., the lag numbers k_ξ and k_M, respectively. The maximum lag values of k_ξ and k_M are set as 8 and 7, respectively. The deeper color means the smaller RMSE/MAPE and a better performance of the lag value. As can be seen in Fig. 5a, in the dark blue area, the M-LSTMs

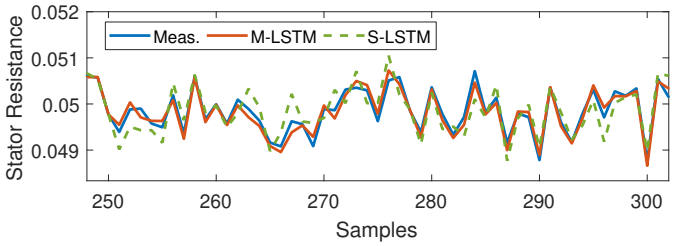


Fig. 7. An example of $r_{S,t}$ with several samples randomly selected in the 23-bus system.

TABLE I
NORMALIZED RMSE OF ALL PARAMETERS USING FIVE METHODS IN 23-BUS SYSTEM

Parameters	Estimation Methods				
	TDNN	NARX	GRU	S-LSTM	M-LSTM
r_S	10.16	9.32	6.82	6.26	2.78
x_S	5.73	5.26	3.84	3.53	1.57
x_m	5.10	4.68	3.42	3.14	1.40
x_R	4.32	3.97	2.90	2.66	1.18
r_R	6.79	6.23	4.56	4.19	1.86
H	3.43	3.15	2.31	2.12	0.94
a_P	3.74	3.43	2.51	2.31	1.02
b_P	3.14	2.88	2.11	1.93	0.86
a_Q	5.01	4.60	3.36	3.09	1.37
b_Q	3.66	3.36	2.46	2.26	1.00

TABLE II
ERROR METRICS AND RUNNING TIME OF FIVE METHODS TO ESTIMATE x_R IN 23-BUS SYSTEM

Metrics	Estimation Methods				
	TDNN	NARX	GRU	S-LSTM	M-LSTM
Correlation coeff.	0.91	0.92	0.96	0.96	0.99
RMSE ($\times 10^{-4}$)	8.88	8.15	5.96	5.47	2.43
MaxAE ($\times 10^{-3}$)	1.54	1.41	1.03	0.95	0.42
MAPE ($\times 10^{-4}$)	7.73	7.09	5.19	4.76	2.11
4RMQE ($\times 10^{-3}$)	1.02	0.94	0.68	0.63	0.28
95th perc. ($\times 10^{-3}$)	1.44	1.32	0.96	0.88	0.39
Running Time for Each Input Sample [s]	0.009	0.027	0.063	0.095	0.115

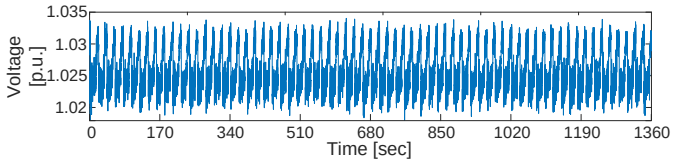
with the lag values of k_ξ (2–4) and k_M (1–4) shows the smallest RMSE values. The same observation can also be found in Fig. 5b, where the best performance of M-LSTM is with the same range of lag values. For the 23-bus system, this observation can be used by power system operators when choosing input variables for the proposed M-LSTM.

In order to evaluate the generalization capability and show the capability of our proposed architecture to avoid the over-fitting issues, Table I compares normalized RMSE values of all parameters using five estimation methods. As can be seen, the proposed M-LSTM method performs best and shows the smallest normalized RMSE values for all parameters compared with other benchmark methods. For the detailed statistical analysis, Table II compares the numerical performance of different methods for the rotor reactance x_R. As shown in this table, the proposed M-LSTM method can significantly increase the correlation coefficient metric and lower other metrics. Taking RMSE and MAPE as an example, for the benchmark methods, TDNN method has the largest RMSE and MAPE. This is because it does not have a recurrent structure and can only capture low-level abstractions of the training data. However, the proposed M-LSTM method can improve RMSE by 72.64% and MAPE by 72.71% compared with TDNN. Compared with S-LSTM, M-LSTM method can also improve RMSE by 55.58% and MAPE by 55.67%. For the sake of simplicity, Fig. 7 shows an example of $r_{S,t}$ with 50 samples randomly selected to compare the performance of M-LSTM and S-LSTM. As can be seen, M-LSTM shows a more fitting accuracy compared with S-LSTM. This is mainly due to the multi-modal structure of the proposed model that can make full use of different modalities of the input data to accurately estimate the underlying target function. However, S-LSTM does not provide the joint representation of different modalities of the input data. Hence, it is not able to use the data effectively.

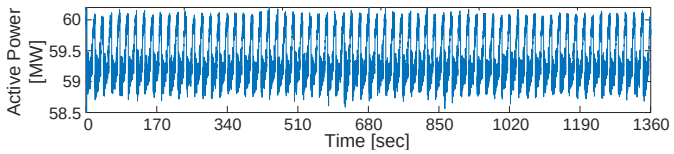
B. 68-bus System

The data of the second case is generated by the 5-area 16-machine 68-bus test system (also known as the New England and New York Interconnected system) [34]. There are four CLM with time-varying parameters connected on buses 27, 39, 49, and 60. The mean values of these parameters $r_{S,t}$, $x_{S,t}$, $x_{m,t}$, $x_{R,t}$, $r_{R,t}$, H_t, $a_{P,t}$, $b_{P,t}$, $a_{Q,t}$, $b_{Q,t}$ are set as 0.01, 0.15, 5, 0.15, 0.05, 3, 0.35, 0.45, 0.25, 0.45. In this system, there are 35 load buses. Thus, the dimension of output variables is 40 (=10×4) and that of input variables is $35\times3(k_M+1)$+$40\times k_\xi$. In the 68-bus system the changes in the network topology is captured by outages in transmission line (i.e., breaker opening). It is not expected that the topology of the system changes after an outage of the transmission line in the simulation period. Here, the proposed M-LSTM model incorporates the data collected from original network topology as well as the reconfigured network (after transmission line outage) to estimate the time-varying load parameters. Once the grid topology changes, the data collected at each time step with the new network topology is also used to determine the load parameters for the next time step. The sampling time is set as 0.1 s. For each event, the time horizon is set as 20 s. Time domain curves of voltage, active power, and reactive power on Bus 17 (the first load bus) with 68 transmission line outages are shown in Fig. 8.

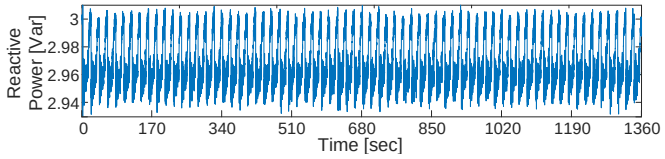
Fig. 9 shows the validation RMSE and MAPE of M-LSTM for the 68-bus system, as a function of the number of temporal training steps, i.e., the lag numbers k_ξ and k_M, respectively. The maximum lag values of k_ξ and k_M are set as 8 and 7, respectively. Fig. 9a shows the smallest RMSE values with the lag values of k_ξ (3–6) and k_M (5–7). The similar observation can also be found in Fig. 9b with the same range of lag values. Compared with the results in the 23-bus system, the lag value of k_ξ corresponding to the smallest RMSE and MAPE has a slight increase while that of k_M has a significant increase. This is mainly because the M-LSTM of the 68-bus system leads to better generalization using more input variables and output variables. Another interesting finding is that the optimal lag value of k_M increases larger than that of k_ξ. This phenomenon indicates that power system operators need to gather more information of system-wide measurements of active power,



(a) Voltage data with 68 transmission line outages

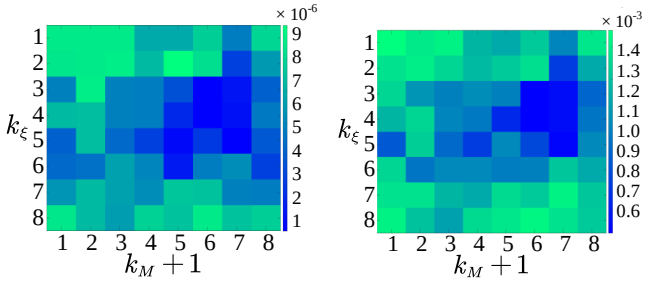


(b) Active power data with 68 transmission line outages



(c) Reactive power data with 68 transmission line outages

Fig. 8. Time domain curves of voltage, active power, and reactive power on Bus 17 (the first load bus) with 68 transmission line outages in the 68-bus system.



(a) Validation RMSE (b) Validation MAPE

Fig. 9. Validation (a) RMSE and (b) MAPE of M-LSTM for the 68-bus system.

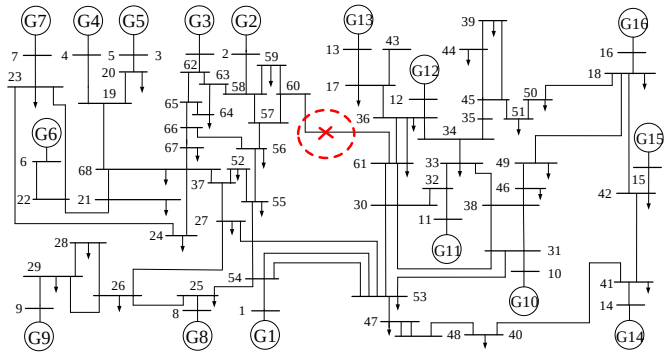


Fig. 10. Topology of the 68-bus system with an example of one transmission line outages.

reactive power, and voltage, in order to estimate more accurate time-varying parameters.

Table III compares normalized RMSE values of all parameters using five estimation methods. As can be seen, the proposed M-LSTM method performs best and shows the smallest normalized RMSE values for all parameters compared with other benchmark methods. For the detailed statistical

TABLE III
NORMALIZED RMSE OF ALL PARAMETERS USING FIVE METHODS IN 68-BUS SYSTEM

Parameters	Estimation Methods				
	TDNN	NARX	GRU	S-LSTM	M-LSTM
r_S	17.16	17.16	15.70	14.60	10.56
x_S	16.96	16.93	15.41	14.85	10.30
x_m	17.24	16.57	15.30	14.51	10.29
x_R	17.16	16.93	15.92	14.70	10.40
r_R	17.26	15.97	15.84	14.76	10.38
H	16.80	16.74	15.44	14.68	10.31
a_P	17.41	16.70	15.06	14.11	10.17
b_P	17.55	17.22	16.26	14.93	10.60
a_Q	17.14	17.08	15.40	14.75	10.50
b_Q	16.88	17.19	15.14	14.84	10.26

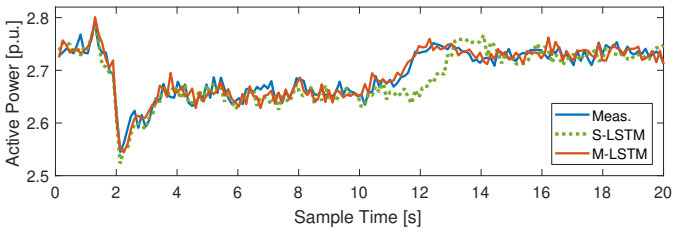
TABLE IV
ERROR METRICS AND RUNNING TIME OF FIVE METHODS TO ESTIMATE r_S IN 68-BUS SYSTEM

Metrics	Estimation Methods				
	TDNN	NARX	GRU	S-LSTM	M-LSTM
Correlation coeff.	0.018	0.006	0.016	0.029	0.306
RMSE ($\times 10^{-4}$)	1.77	1.77	1.62	1.51	1.09
MaxAE ($\times 10^{-4}$)	8.98	8.20	7.38	6.43	3.70
MAPE ($\times 10^{-4}$)	1.17	1.15	1.15	1.13	0.87
4RMQE ($\times 10^{-4}$)	2.78	2.73	2.44	2.17	1.43
95th perc. ($\times 10^{-4}$)	4.02	3.98	3.52	3.11	2.14
Running Time for Each Input Sample [s]	0.024	0.039	0.078	0.104	0.129

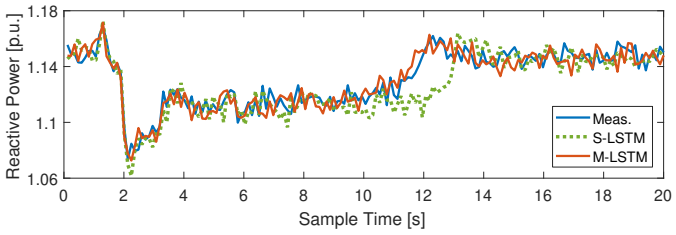
analysis, Table IV compares the numerical performance of different methods for the stator resistance r_S in the 68-bus system. As can be seen, the proposed M-LSTM method can significantly increase the correlation coefficient metric and lower other metrics. Taking RMSE and MAPE metrics as an example, for the benchmark methods, TDNN method yields poor results with the largest RMSE and MAPE values. The S-LSTM improves RMSE by 14.69% and MAPE by 3.42% compared with TDNN method. However, the proposed M-LSTM can significantly improve RMSE by 38.42% and MAPE by 25.64% compared with TDNN method. Compared with S-LSTM, M-LSTM method can improve RMSE by 27.81% and MAPE by 23.01%. These aforementioned findings are consistent with those in Section IV-A and can verify the robustness of the proposed method. This is mainly because it can provide the high-level abstraction and contain a multi-modal structure.

C. Effectiveness of Estimated Time-Varying Parameters

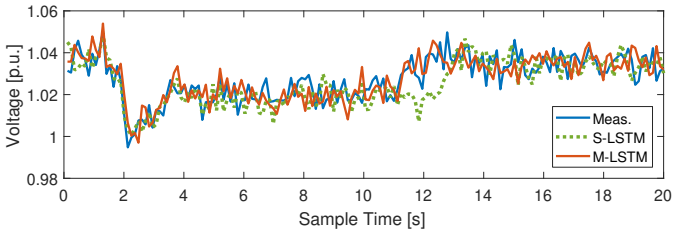
To verify the effectiveness of estimated time-varying parameters, the results obtained by S-LSTM and M-LSTM methods are respectively put into CLM to get the estimated measurements: active power, reactive power, and voltage. A transmission line outage on transmission line 60–61 in the 68-bus system is simulated to gather the estimated measurements, which can be seen in Fig. 10. Fig. 11 compares the estimated measurements with the real data by using S-LSTM and M-LSTM. As can be seen, the results estimated by M-LSTM method (the orange line) are closer to the real measurements (the blue line). The slight difference between the real data and the estimated measurements by M-LSTM is mainly caused by



(a) Active Power



(b) Reactive Power



(c) Voltage

Fig. 11. Comparison of the estimated active power (a), reactive power (b), and voltage (c) obtained by S-LSTM and M-LSTM methods.

the noise. However, the results estimated by M-LSTM method (the dashed green line) are farther from the real measurements, especially for the time interval during 10–14 s.

Table V shows the performance of five methods considering the RMSE and MAPE metric of estimated active power, reactive power, and voltage. As shown in this table, both S-LSTM and M-LSTM methods outperform TDNN, NARX, and GRU leading to smaller values of RMSE and MAPE. For TDNN, NARX, and GRU, the RMSEs of three measurements are in the larger range of 0.03–0.05, 0.015–0.03, and 0.01–0.02, respectively. For S-LSTM and M-LSTM methods, RMSEs of three measurements are in the smaller range of 0.01–0.03, 0.005–0.015, and 0.005–0.01, respectively. A similar observation is made for the MAPE metrics. Comparing L-LSTM with M-LSTM, for the estimated active power, M-LSTM method can improve RMSE by 48.48% and MAPE by 38.46% compared with S-LSTM method. For the estimated reactive power, M-LSTM method can improve RMSE by 34.86% and MAPE by 24.24%. For the estimated voltage, M-LSTM method can improve RMSE by 17.28% and MAPE by 9.84%. This is mainly because S-LSTM does not consider the joint representation of modalities of input variables. M-LSTM can estimate the most accurate time-varying parameters which can be used to get measurements with the best RMSE and MAPE metrics to the real data. This finding can verify the effectiveness of the proposed M-LSTM method.

TABLE V
ERROR METRICS OF ESTIMATED ACTIVE POWER, REACTIVE POWER, AND VOLTAGE IN FIVE METHODS

Metrics	Estimation Methods	Measurements		
		Active Power	Reactive Power	Voltage
RMSE	TDNN	0.0421	0.0205	0.0187
	NARX	0.0384	0.0196	0.0144
	GRU	0.0357	0.0182	0.0103
	S-LSTM	0.0264	0.0109	0.0081
	M-LSTM	0.0136	0.0071	0.0067
MAPE	TDNN	0.0101	0.0099	0.0095
	NARX	0.0094	0.0091	0.0082
	GRU	0.0087	0.0088	0.0079
	S-LSTM	0.0065	0.0066	0.0061
	M-LSTM	0.0040	0.0050	0.0055

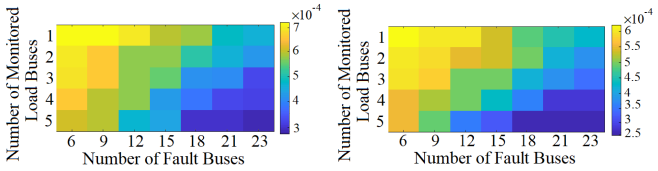
D. Impact of a Limited Number of Monitored Load Buses and System Events

To validate the effectiveness of the proposed M-LSTM method with limited data, two types of limitations are considered: a limited number of monitored load buses and a limited number of system events. The priority on monitoring of the load bus is determined based on the value of its active power. That is to say, the operators give higher priority to the buses with larger active power consumption for monitoring. For example, assuming that only measurements on three load buses can be monitored by operators, we choose to monitor the first three load buses with largest active power consumption. Based on the above consideration, limited number of load buses are considered for being monitored.

Similarly, a limited number of system buses and transmission lines are selected for ground faults and outages respectively. In this context, higher priority is given to the load bus with larger active power consumption as well as the transmission lines that transmit the larger volume of active power. For example, assuming that the operators can only simulate nine buses with ground faults in the whole system, the first nine system buses with the largest active power consumption are chosen as the limited number of ground fault buses. This is because the ground fault on the load bus with larger active power consumption could significantly affect the stability of the power grid compared to the faults on load buses with smaller active power consumption.

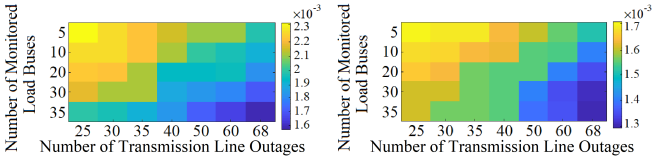
Therefore, higher priority is given to the load bus with larger active power consumption as well as the transmission lines that transmit the larger volume of active power. Similarly, the outage of the transmission lines that carry a large volume of active power could significantly impact the stability of the power system. Considering the limited number of monitored load buses and system events (i.e., a ground fault on buses or transmission line outages), the proposed M-LSTM model is trained and the corresponding RMSE and MAPE metrics of the estimated time-varying parameters are computed.

For the 23-bus system with ground faults, Fig. 12 shows the RMSE and MAPE of the estimated parameter $a_{P,t}$ in the ZIP load model. The vertical axis presents the change in the number of monitored load buses, while the horizontal axis shows the change in the number of ground faulted buses. As shown in this figure, with the increasing number of monitored load buses (from top to bottom of the vertical axis) and



(a) RMSE of $a_{P,t}$ (b) MAPE of $a_{P,t}$

Fig. 12. An example of RMSE and MAPE results of parameter $a_{P,t}$ in the 23-bus system.



(a) RMSE of $x_{R,t}$ (b) MAPE of $x_{R,t}$

Fig. 13. An example of RMSE and MAPE results of parameter $x_{R,t}$ in the 68-bus system.

system buses with ground faults (from left to right in the horizontal axis), both RMSE and MAPE metrics are reduced. This reflects the better accuracy of M-LSTM when more measurements and system events are provided.

For the 68-bus system with outages in transmission lines, Fig. 13 shows the RMSE and MAPE for the estimated parameter $x_{R,t}$. Similar observations were made in this case, i.e., increasing the number of monitored load buses (from top to bottom of the vertical axis) and system buses with ground faults (from left to right in the horizontal axis) decreases both RMSE and MAPE metrics and yields more accurate estimates of the parameters.

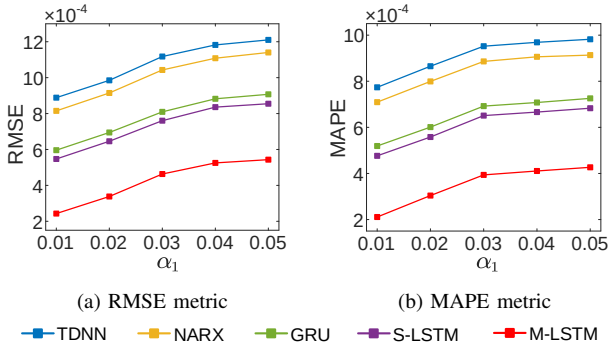
Overall, Figs. 12 and 13 indicate that the estimation accuracy of the developed M-LSTM method is increased when more information on measurements and system events (e.g., ground faults and transmission line outages) is provided to the system operator.

E. Sensitivity Analysis Considering the Variabilities of Parameters

To efficiently simulate the time-varying behavior of load modeling in practice, different variability measures for the load parameters are considered to capture their impacts on the estimation accuracy of the load parameters. The time-varying parameters are generated using (30) in which μ_ξ is the mean value of time-varying parameters, and $\mathcal{N}(\cdot)$ is the normal distribution function with mean zero and the standard deviation $\mu_\xi \times \alpha_1$. In this case, five values of α_1 are chosen from 0.01, 0.02, 0.03, 0.04, and 0.05, respectively.

$$\boldsymbol{\xi} \sim \mu_\xi + \underbrace{\mathcal{N}(0,\ \mu_\xi \times \alpha_1)}_{Gaussian\ Random\ Variable} \tag{30}$$

Fig. 14 shows the accuracy measures for five methods with the increase of α_1. It is shown that for all benchmarks, both RMSE and MAPE metrics increase with the proportion α_1. Therefore, higher variability of parameters would lead to lower estimation accuracy. However, considering the relatively small



(a) RMSE metric (b) MAPE metric

Fig. 14. RMSE and MAPE of five methods with different variabilities of load parameters.

RMSE $[2, 6]\times10^{-4}$ and MAPE $[2, 5]\times10^{-4}$, the proposed M-LSTM method still outperforms other four estimation methods, i.e., TDNN, NARX, GRU, and S-LSTM with a higher input load parameter error. This is due to better generalization and the capability of making use of the measurement information.

F. Computational Time Analysis

1) Execution Time: Tables II and IV show the running time for each input sample in the 23- and 68-bus systems, respectively. As shown in these tables, TDNN and NARX have fast computation performance as they are shallow architectures with low number of computational layers. However, these methods are error-prone. Compared to TDNN and NARX, the GRU, S-LSTM, and M-LSTM methods are deep neural architectures with better accuracy due to multiple computational layers. Among deep networks, GRU has smaller time complexity due to a smaller set of parameters. However, LSTM variations (S-LSTM and M-LSTM) yield better accuracy due to a larger number of parameters that help these models to have higher generalization capability. M-LSTM has slightly more time complexity compared to S-LSTM due to its additional temporal pooling and regression layers. Considering the trade-off between time complexity with accuracy, the M-LSTM model is more preferable compared to other benchmarks due to significantly higher accuracy. In addition, the execution time of the M-LSTM method is small enough to be employed for the problem of estimating time-varying parameters for load modeling.

2) Time Complexity Analysis (Asymptotic Analysis): The developed M-LSTM model consists of four major computational steps:

- ***Step 1 – Temporal Feature Extraction From Load Parameters***: In this step, the temporal load parameter data $\left(\boldsymbol{\xi}_{d_\xi}^{t-1}, \boldsymbol{\xi}_{d_\xi}^{t-2}, \cdots, \boldsymbol{\xi}_{d_\xi}^{t-k_\xi}\right)$ is observed by an LSTM network in k_ξ time steps. At each time step i, the temporal feature $\boldsymbol{h}_\xi(i)$ is obtained using recurrent feed-forward formulations in (27a)–(28b).
- ***Step 2 – Temporal Feature Extraction From Measurement Data***: This step extracts temporal feature $\boldsymbol{h}_M(i)$ at each time step $1 \le i \le k_M + 1$ from the measurement data $M(i) = \left\{\boldsymbol{M}_{d_M}^{1,t-i}, \boldsymbol{M}_{d_M}^{2,t-i}, \cdots, \boldsymbol{M}_{d_M}^{N,t-i}\right\}$, using (27a)–(28b).

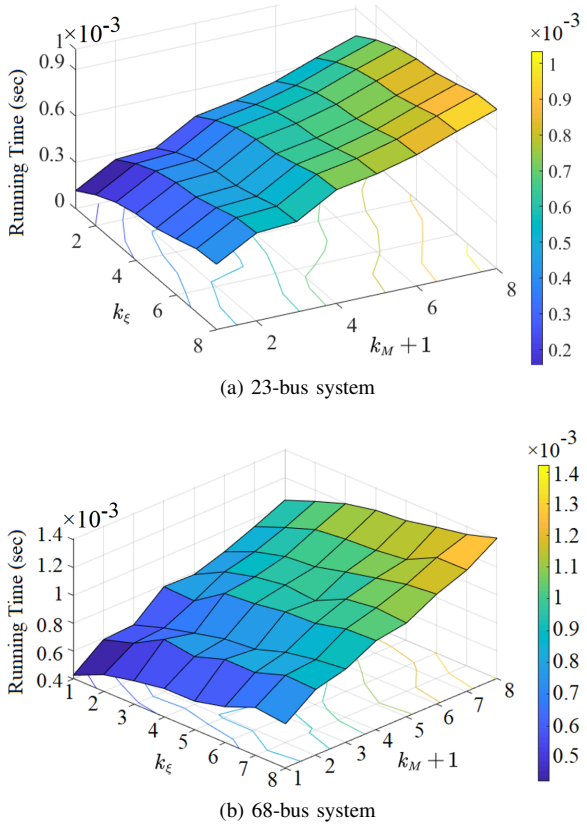
- ***Step 3 – Temporal Pooling***: By using (25) and (26), the average of $\boldsymbol{h}_{\xi}(i)$, $1 \leq i \leq k_{\xi}$, and $\boldsymbol{h}_{M}(i')$, $1 \leq i' \leq k_{M}+1$, is computed.
- ***Step 4 – Regression***: $\widehat{\boldsymbol{\xi}}_{d_{\xi}}^{t} = g\left(u_{\xi}^{t}, u_{M}^{t}\right)$ is computed as a linear regression model.

As the first two steps are being executed simultaneously (see Fig. 4), in order to find the time complexity, one needs to compute the execution time only for the LSTM with the largest number of time steps. Each time step in the LSTM consists of several feed-forward calculations formulated in (27a)–(28b). The total number of LSTM parameters K is computed using the O notation:

$$\begin{aligned} K &= 4 \times n_c \times n_c + 4 \times n_c \times n_i + n_c \times n_o + 3 \times n_c \\ &= O\left(n_c^2 + n_c\left(n_i + n_o\right)\right) \end{aligned} \tag{31}$$

where n_c is the number of memory cells, n_i is the number of input variables, and n_o is the number of output features. As n_i and n_o are small constants, asymptotic complexity is further formulated as $K = O\left(n_c^2 + n_c\right) = O\left(n_c^2\right)$. Thus, the first two steps would take $T(1) = O\left(Cn_c^2\right)$ with $C = \max\left(k_{\xi}, k_{M}\right)$, as both LSTMs are run in parallel (see Fig. 4). The third step takes $T(2) = O\left(\max\left(k_{\xi}, k_{M}+1\right)\right) = O(1)$ as time lags k_{ξ} and k_{M} are small constants. The fourth step is a regression implemented by the multiplication of W by the temporal latent vector $u = [u_{\xi}^{t}, u_{M}^{t}]$ to obtain $\widehat{\boldsymbol{\xi}}_{d_{\xi}}^{t}$. As the output, i.e., $\widehat{\boldsymbol{\xi}}_{d_{\xi}}^{t}$, is 10-dimensional, and the weight matrix $W \in \mathbb{R}^{10 \times \dim(u)}$, one can write the time complexity of the regression as $T(3) = O\left(10\dim(u)\right) = O\left(\dim(u)\right) = O(1)$. Hence, the asymptotic time complexity of the proposed model is $T_{total} = O\left(T(1) + T(2) + T(3)\right) = O\left(n_c^2 + 1 + 1\right) = O\left(n_c^2\right)$. The complexity of the proposed model is a low degree polynomial of the number of memory cells of the LSTM. This shows that the running time of the proposed model is as low as a simple feed-forward neural network with n_c number of hidden units at each hidden layer. As a result, the proposed methodology can be efficiently used for the time-varying load modeling problem.

Fig. 15 shows the average computational time of the proposed M-LSTM model on the samples of the validation data for various hyper-parameters $1 \leq k_{\xi} \leq 8$ and $1 \leq k_{M}+1 \leq 8$ in the 23- and 68-bus system. As shown in Fig. 15, the running time increases with the growth in time lags. The maximum running time of the 23-bus system is 1.034 ms while the 68-bus system has 1.409 ms maximum running time for each validation sample.



(a) 23-bus system

(b) 68-bus system

Fig. 15. M-LSTM running time (in seconds) for various time lags in two test cases.

V. Conclusion

This paper develops a deep learning based time-varying parameter identification model considering system-wide measurements, including active power, reactive power, and voltage. A multi-modal long short-term memory (M-LSTM) method is developed to estimate the time-varying parameters. An LSTM network using flexible number of temporal states is defined over the load parameters and measurements time series in order to capture deep features with high abstractions. The proposed multi-modal model is able to be tuned in an end-to-end fashion using differentiable computational components. The linear regression is employed to output the estimation of time-varying parameters at the current time. Compared with the benchmark methods, the effectiveness and robustness of the developed M-LSTM method is verified by numerical simulations on the 23- and 68-bus system. It is shown that our model outperforms shallow methodologies due to the provided high-level abstraction, and deep networks due to its multi-modal structure. Some universal and common lessons are shown as follows:

(i) The proposed M-LSTM method performs best and shows the smallest metrics for all parameters compared to shallow pattern recognition architectures that exist in the literature.

(ii) In contrast to previous methodologies, by using flexible number of temporal states, our model is capable of searching for the optimal lag number when learning temporal features for the parameter estimation task. Optimal lag values of parameters and measurements are in the ranges of 2–4 and 1–4 for the 23-bus system, and in the larger ranges of 3–6 and 5–7 for the 68-bus system.

(iii) Based on estimated time-varying parameters, the measurements obtained by M-LSTM are closer to the real data with smaller RMSE and MAPE metrics, due to the high generalization capability and the recurrent structure of the proposed model, as well as the multi-modal structure capable of providing a joint representation for the underlying load parameters and measurements.

In the future work, this study can be extended to incorporate

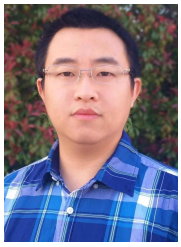
with supervised discriminative learning approaches including decision trees and support vector machines, in order to learn nonlinear decision surfaces to differentiate renewable generation from the time-varying load. While the proposed approach could be extended to estimate the parameters of renewable generation units, such effort could be considered as an extension of the current research effort.

ACKNOWLEDGMENT

Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, under contract DE-AC02-06CH11357.

REFERENCES

[1] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling–a review," *IEEE Trans. Smart Grid*, 2017, in press.
[2] D. Karlsson and D. J. Hill, "Modelling and identification of nonlinear dynamic loads in power systems," *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 157–166, 1994.
[3] M. Cui, J. Wang, and M. Yue, "Machine learning based anomaly detection for load forecasting under cyberattacks," *IEEE Trans. Smart Grid*, 2019, in press.
[4] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, 2018.
[5] C. Li, Y. Li, Y. Cao, J. Ma, Y. Kuang, Z. Zhang, L. Li, and J. Wei, "Credibility forecasting in short-term load forecasting and its application," *IET Gener. Transm. Distrib.*, vol. 9, no. 13, pp. 1564–1571, 2015.
[6] C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust time-varying parameter identification for composite load modeling," *IEEE Trans. Smart Grid*, 2017, in press.
[7] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 702–710, 2001.
[8] 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, 2017.
[9] J. De Kock, F. Van Der Merwe, and H. Vermeulen, "Induction motor parameter estimation through an output error technique," *IEEE Trans. Energy Convers.*, vol. 9, no. 1, pp. 69–76, 1994.
[10] A. Rouhani and A. Abur, "Real-time dynamic parameter estimation for an exponential dynamic load model," *IEEE Trans. Smart Grid*, vol. 7, no. 3, pp. 1530–1536, 2016.
[11] K. Zhang, H. Zhu, and S. Guo, "Dependency analysis and improved parameter estimation for dynamic composite load modeling," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3287–3297, 2017.
[12] J. Ma, R. He, and D. J. Hill, "Load modeling by finding support vectors of load data from field measurements," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 726–735, 2006.
[13] T. Hiyama, M. Tokieda, W. Hubbi, and H. Andou, "Artificial neural network based dynamic load modeling," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1576–1583, 1997.
[14] B.-Y. Ku, R. J. Thomas, C.-Y. Chiou, and C.-J. Lin, "Power system dynamic load modeling using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1868–1874, 1994.
[15] V. Knyazkin, C. A. Canizares, and L. H. Soder, "On the parameter estimation and modeling of aggregate power system loads," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1023–1031, 2004.
[16] J. Zhao, Z. Wang, and J. Wang, "Robust time-varying load modeling for conservation voltage reduction assessment," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3304–3312, 2018.
[17] D. Q. Hung, N. Mithulananthan, and K. Y. Lee, "Determining PV penetration for distribution systems with time-varying load models," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 3048–3057, 2014.
[18] Z. Wang and J. Wang, "Time-varying stochastic assessment of conservation voltage reduction based on load modeling," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2321–2328, 2014.
[19] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C. Kang, "Deep learning-based socio-demographic information identification from smart meter data," *IEEE Trans. Smart Grid*, 2018, in press.
[20] A. Keyhani, W. Lu, and G. T. Heydt, "Composite neural network load models for power system stability analysis," in *Proc. IEEE PES Power Syst. Conf. and Expo.*, 2004, pp. 1159–1163.
[21] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 422–433, Apr. 2015.
[22] M. Khodayar, J. Wang, and M. Manthouri, "Interval deep generative neural network for wind speed forecasting," *IEEE Trans. Smart Grid*, 2018, in press.
[23] X. Li, M. Peng, H. He, and T. Liu, "Dynamic load modeling for power system based on GD-FNN," in *Proc. 3rd Int. Conf. Digital Manuf. Autom.*, 2012, pp. 339–342.
[24] K. Kawaguchi, L. P. Kaelbling, and Y. Bengio, "Generalization in deep learning," *arXiv preprint arXiv:1710.05468*, 2017.
[25] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy*, 2018, in press.
[26] H. Meng, N. Bianchi-Berthouze, Y. Deng, J. Cheng, and J. P. Cosmas, "Time-delay neural network for continuous emotional dimension prediction from facial expression sequences," *IEEE T. Cybern.*, vol. 46, no. 4, pp. 916–929, 2016.
[27] F. da Costa Lopes, E. H. Watanabe, and L. G. B. Rolim, "A control-oriented model of a PEM fuel cell stack based on NARX and NOE neural networks," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5155–5163, 2015.
[28] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, 2018.
[29] Keras framework. [Online]. Available: https://github.com/fchollet/keras
[30] M. Cui, C. Feng, Z. Wang, and J. Zhang, "Statistical representation of wind power ramps using a generalized Gaussian mixture model," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 261–272, Jan. 2018.
[31] M. Cui, Z. Wang, C. Feng, and J. Zhang, "A truncated Gaussian mixture model for distributions of wind power ramping features," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, 2017, pp. 1–5.
[32] Siemens. (2010) PSSE 32.0.5 Program Operational Manual.
[33] M. Cui, J. Wang, J. Tan, A. Florita, and Y. Zhang, "A novel event detection method using PMU data with high precision," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 454–466, Jan. 2019.
[34] Report on the 68-bus, 16-machine, 5-area system. [Online]. Available: https://eioc.pnnl.gov/benchmark/ieeess/NETS68/New_England_New_York_68_Bus_System_study_report.pdf



Mingjian Cui (S'12–M'16–SM'18) received the B.S. and Ph.D. degrees from Wuhan University, Wuhan, Hubei, China, all in Electrical Engineering and Automation, in 2010 and 2015, respectively.

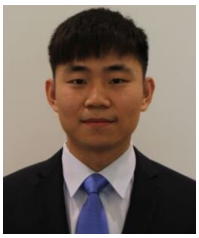
Currently, he is a Postdoctoral Research Associate at Southern Methodist University, Dallas, Texas, USA. He was also a Visiting Scholar from 2014 to 2015 in the Transmission and Grid Integration Group at the National Renewable Energy Laboratory (NREL), Golden, Colorado, USA. His research interests include power system operation, wind and solar forecasts, machine learning, data analytics, and statistics. He has authored/coauthored over 50 peer-reviewed publications. Dr. Cui serves as an Associate Editor for the journal of IET Smart Grid. He is also the Best Reviewer of the IEEE Transactions on Smart Grid for 2018.

Mahdi Khodayar (S'17) received his B.Sc. degree in computer engineering and the M.Sc. degree in artificial intelligence from Khajeh Nasir Toosi University of Technology, Tehran, Iran, in 2013 and 2015, respectively.

He is currently working toward the Ph.D. degree in Electrical Engineering at Southern Methodist University, Texas, United States. His main research interests include machine learning and statistical pattern recognition. He has served as a reviewer for reputable journals including IEEE Transactions on Industrial Informatics, IEEE Transactions on Fuzzy Systems, IEEE Transactions on Sustainable Energy, and IEEE Transactions on Power Systems. He is currently focused on deep learning methodologies for pattern recognition in power systems.

Chen Chen (M'13) received the Ph.D. degree in electrical engineering from Lehigh University, Bethlehem, PA, USA, in 2013.

During 20132015, he worked as a Postdoctoral Researcher at the Energy Systems Division, Argonne National Laboratory, Argonne, IL, USA. He is currently a Computational Engineer in the Energy Systems Division, Argonne National Laboratory. His primary research interests include optimization, communications and signal processing for smart electric power systems, cyber-physical system modeling for smart grids, and power system resilience. He is an Editor of the IEEE TRANSACTIONS ON SMART GRID and the IEEE POWER ENGINEERING LETTERS.



Xinan Wang (S'15) received the B.S. degree in electrical engineering from Northwestern Polytechnical University, Xian, China, in 2013, and the M.S. degree in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2016.

During Aug. 2016-May. 2017, he worked as a research assistant in the Advanced Power System Analytics Group at GEIRI North America, Santa Clara, CA, USA. He is currently pursuing his PhD degree in electrical engineering at Southern Methodist University, Dallas, TX, USA. His research interests include WAMS related application in power system, data driven load monitoring, and renewable energy integration.

Ying Zhang (S'18) received the B.S. degree and M.S. degree, in electric engineering from Shandong University, Jinan, China, in 2014 and 2017, respectively.

She is currently pursuing the Ph.D. degree in the Department of Electric Engineering at Southern Methodist University, Dallas, Texas, USA. Her research interests include power system analysis and control, and state estimations in distribution systems.

Mohammad E. Khodayar (SM'17) received the B.Sc. degree from Amirkabir University of Technology, Tehran, Iran; the M.S. degree from Sharif University of Technology, Tehran; and the Ph.D. degree from the Illinois Institute of Technology, Chicago, IL, USA, in 2012, all in electrical engineering.

He was a Senior Research Associate with the Robert W. Galvin Center for Electricity Innovation, Illinois Institute of Technology. He is currently an Assistant Professor in the Department of Electrical Engineering, Southern Methodist University, Dallas, TX, USA and an associate editor of the IEEE Transactions on Sustainable Energy and IEEE Transactions on Vehicular Technology. His current research interests include power system operation and planning.