



Vaasan yliopisto
UNIVERSITY OF VAASA

OSUVA Open
Science

This is a self-archived – parallel published version of this article in the publication archive of the University of Vaasa. It might differ from the original.

Wide-Area Composite Load Parameter Identification Based on Multi-Residual Deep Neural Network

Author(s): Afrasiabi, Shahabodin; Afrasiabi, Mousa; Jarrahi, Mohammad Amin; Mohammadi, Mohammad; Aghaei, Jamshid; Javadi, Mohammad Sadegh; Shafie-Khah, Miadreza; Catalão, João P. S.

Title: Wide-Area Composite Load Parameter Identification Based on Multi-Residual Deep Neural Network

Year: 2023

Version: Accepted manuscript

Copyright ©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Please cite the original version:

Afrasiabi, S., Afrasiabi, M., Jarrahi, M. A., Mohammadi, M., Aghaei, J., Javadi, M. S., Shafie-Khah, M. & Catalão, J. P. S. (2023). Wide-Area Composite Load Parameter Identification Based on Multi-Residual Deep Neural Network. *IEEE Transactions on Neural Networks and Learning Systems* 34(9), 6121-6131.
<https://doi.org/10.1109/TNNLS.2021.3133350>

Wide-Area Composite Load Parameter Identification based on Multi-Residual Deep Neural Network

Shahabodin Afrasiabi, *Graduate Student Member, IEEE*, Mousa Afrasiabi, Mohammad Amin Jarrahi, Mohammad Mohammadi, *Member, IEEE*, Jamshid Aghaei, *Senior Member, IEEE*, Mohammad Sadegh Javadi, *Member, IEEE*, Miadreza Shafie-Khah, *Senior Member, IEEE*, and João P.S. Catalão, *Senior Member, IEEE*

Abstract—Accurate and practical load modeling plays a critical role in the power system studies including stability, control and protection. Recently, wide-area measurement systems (WAMS) are utilized to model the static and dynamic behavior of the load consumption pattern in real-time, simultaneously. In this paper, a WAMS based load modeling method is established based on a multi-residual deep learning structure. To do so, a comprehensive and efficient load model founded on combination of impedance-current-power and induction motor is constructed at the first step. Then, a deep learning-based framework is developed to understand the time-varying and complex behavior of the composite load model. To do so, a residual convolutional neural network (ResCNN) is developed to capture the spatial features of the load at different location of the large-scale power system. Then, gated recurrent unit (GRU) is used to fully understand the temporal features from highly variant time-domain signals. It is essential to provide a balance between fast and slow variant parameters. Thus, the designed structure is implemented in a parallel manner to fulfill the balance and moreover, weighted fusion method is used to estimate the parameters, as well. Consequently, an error-based loss function is reformulated to improve the training process as well as robustness in the noisy conditions. The numerical experiments on IEEE 68-bus and Iranian 95-bus systems verify the effectiveness and robustness of the proposed load modeling approach. Furthermore, a comparative study with some relevant methods demonstrate the superiority of the proposed structure. The obtained results in the worst case scenario show error lower than 0.055% considering noisy condition and at least 50% improvement comparing the several state-of-art methods.

Index Terms—Composite load model, deep learning, gated recurrent unit (GRU), pseudo-Huber loss function, residual convolutional neural network (ResCNN).

I. INTRODUCTION

A. Motivation

ELECTRICAL load modeling is essential in the analysis of the traditional and reconstructed power systems. In load

modeling studies, the main goal is to construct a mathematical description to describe the consumption pattern throughout a specific time interval [1]–[3]. Due to emergence of new concepts such as smart grids, renewable energies, active distribution networks, demand-side management, etc, electrical load modeling has faced new challenges such as complicated characteristics and random time variance trends [4]–[6]. This study aims to resolve these challenges that still require a comprehensive and general solution. In this regard, the desired approach must have the following features:

- It must be fast and accurate to track highly variant load characteristics.
- It must be robust to handle highly noisy conditions.
- It must have the ability to understand the spatial and temporal features for providing suitable knowledge on load consumption pattern.
- It must be capable to identify a large number of unknown parameters with regards to the emergence of advanced measurement devices such as phasor measurement units (PMUs) and smart meters provide.

So, the main motivation of this paper is to develop a multi residual deep network to establish a general framework for the wide-area load parameter identification.

B. Brief Literature Review

Load modeling has been conducted in two main stages, including 1) choosing a proper load model based on measurement-based or physical based models, 2) designing a parameter estimation approach to identify the selected load model parameters [7], [8]. Physical-based models provide a detail description of physical behavior and functioning of electrical devices, however they are not applicable in practical conditions due to lack of required detailed information [9], [10]. Therefore, measurement-based load models are more preferred as they offer more practical models. The main principle in measurement-based models is collecting a dataset from measurement equipment and directly implemented for load modeling. Based on the behavior, load can be modeled as static type or can be formulated as dynamic model. However, composite models involve the patterns of both static and dynamic models [11].

Figure 1 presents a classification of load modeling approaches. In this figure, various schemes for each group of load modeling techniques are depicted. The load models in

the static category have the ability to represent the active and reactive consumption as the functions of bus frequency and voltage. Several models such as impedance-current-power (ZIP), exponential and frequency dependent models are most common static load models [12], [13]. Dynamic load models can model the time-variant relationship between active and reactive power based on bus voltage throughout a time interval [14]–[16]. Induction motor (IM) and exponential recovery load model (ERLM) are two widely used dynamic load models in previous literature [17]. However, one single static/dynamic model cannot fully represent a behavior of actual electrical load. Composite load model (CLM) aggregates dynamic and static characteristics which can sufficiently represent the actual load model [18]. Therefore, CLM is one of the most preferable load model in previous studies [19], [20]. To this end, this paper proposes an approach to identify parameters of CLM.

In term of identification approach, the CLM parameter identification approaches can be divided into three following groups: 1) optimization-based techniques, 2) state-space model based techniques, and 3) data-driven based approaches. In the following, it is tried to introduce these groups in more details with the highlights from some papers in each area. The summary of this classification are tabulated in Table I. In this table, the cons and pros of each group are declared.

- *Optimization-based techniques*

The methods of this group firstly determine an objective function founded on error that is the difference between measurement and estimated values. Then, they employ different techniques such as Lagrangian based algorithm [9], particle swarm optimization [21] and heuristic search algorithm [22] to optimize the objective function. The main disadvantages of these approaches are related to high computational complexity and considering only measurement data at current time. As load consumption is a time series data that depend on the previous time steps, ignoring last data can adversely affects on the accuracy.

- *State-space model based techniques*

The approaches in this group attempt to estimate the parameters using measurement data and state-space load model. For instance, a CLM parameter identification is formulated as weighted least square problem in [23]. Moreover, Kalman filter (KF) based approaches such as extended KF (EKF) and unscented KF (UKF) are also presented in [24] and [25] for CLM and dynamic load models parameter identification, respectively. Generally, the correlation between the loads of the system at various locations cannot be taken into account in this group. So, this shortcoming considerably affects the performance of the methods. That is why, the correlation between loads must be observed in the methods to improve their accuracy.

- *Data-driven based techniques*

Data-driven based approaches are fast and capable of considering the impact of load at different location, which are included two main subcategory i.e. shallow and deep structures. The shallow based structure such as support vector machine (SVM) [14], artificial neural network (ANN) [26], and fuzzy logic [27] are developed in recent years for load parameter

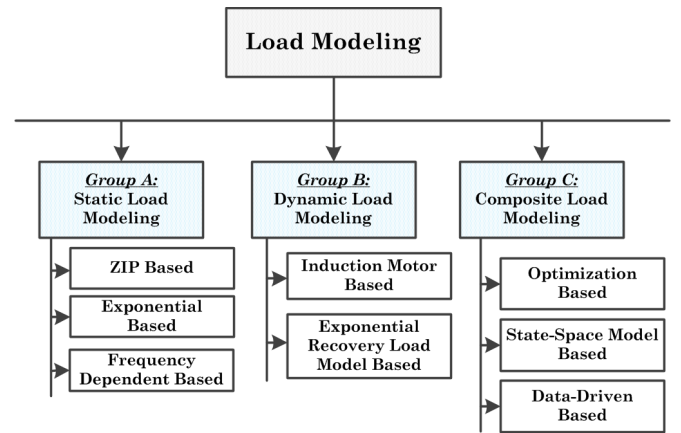


Fig. 1. Load modeling approaches
Table. I. Descriptions of CLM methods

Techniques	References	Attributes (+/-)
Optimization Based	[9], [21], [22]	+ Easy to implement - High computational burden - Considering only data of current time
State-Space Model Based	[23-25]	+ Detailed modeling - Does not consider the correlation between loads - Complicated with huge parameters
Data-Driven Based	[2], [3], [4], [6], [14], [26-28]	+ Highly Adaptive + Does not need load profile - Sensitive to measurement noises

estimation. However, these structures are not able to properly realize the nature and characteristics of the measurement raw data. Also, another major shortcomings of these approaches that makes them unusable due to their small hypothesis space are highly sensitivity to measurement noise and suffering from lack of generality [28]. Moreover, a large share of previously presented approaches can only be utilized for a single bus, therefore, they are not suitable. Deep learning emergence as an evolutionary concept in machine learning has attracted a large number of researchers due to the ability in capturing complex and nonlinear features from the raw time signals [2], [3], [6]. A deep recurrent neural network based approach, namely long short-term memory (LSTM) is developed in [4]. Although it offers suitable performance, it does not have proper efficiency in learning spatial features and noisy environments. Moreover, single or multimodal LSTM networks are not fully compatible with different dynamic behaviors for tracking various load patterns.

Considering the presented discussion, this paper focuses on presenting a data-driven based technique for CLM parameter identification.

C. Contributions

This paper aims to provide an accurate and fast CLM parameter identification approach, proposing a deep residual based structure to fully understand robust, spatial, and temporal features. In the designed structure, convolutional neural network (CNN) is used to realize the spatial features, however, in contrast with RNN-based networks, CNN is weakened to learn spatial features, especially in long-tailed time series associated with high variations. To this end, firstly a CNN

unit converts to residual CNN units, then a gated recurrent unit (GRU) as a time efficient and stronger temporal feature learner than LSTM is added to the designed network. Then, to make a balance between different dynamic behaviors of the load parameter in CLM, the designed network converted into three parallel networks and also improve the computational efficiency. Consequently, an error-based loss function is adopted to improve the method performance in noisy conditions. Furthermore, it can help the method to increase its training ability.

To sum up, the main novelties of the current work are presented as follows:

- A multi-residual deep network is developed to accurately identify the CLM parameters through a strong spatial and temporal learners.
- An error-based loss function is suggested to enhance the method capability in training process which can provide accurate estimation for CLM parameters in noisy environments.
- The slow/fast dynamic behavior of CLM parameters are realized using three parallel deep networks, which can improve the time efficiency, as well.

D. Organizations

The organization for the following contents is as follows: the mathematical descriptions of the wide-area CLM model are provided in section II. Section III describes the developed multi-residual deep network in details. The numerical results and experiments are given in section IV. At the end, the conclusions are drawn in section V.

II. WIDE-AREA LOAD MODELING

A. Composite Load Model

The composite load model involves the static and dynamic components which are modeled as *ZIP* and *IM* models, respectively. The *ZIP* model consists of three main parts which are constant impedance shown as *Z*, constant current shown as *I*, and constant power as *P*, that are formulated as follows [4]:

$$P_t^{ZIP} = \alpha_t^P \left(\frac{V_t}{V_b} \right)^2 + \beta_t^P \left(\frac{V_t}{V_b} \right) + \gamma_t^P \quad (1)$$

$$Q_t^{ZIP} = \alpha_t^Q \left(\frac{V_t}{V_b} \right)^2 + \beta_t^Q \left(\frac{V_t}{V_b} \right) + \gamma_t^Q \quad (2)$$

where P_t^{ZIP} and Q_t^{ZIP} represent active and reactive power in the model, and $\alpha_t^{P/Q}$, $\beta_t^{P/Q}$ and $\gamma_t^{P/Q}$ are constants showing the percentage associated with active/reactive power, subjected to $\alpha_t^{P/Q} + \beta_t^{P/Q} + \gamma_t^{P/Q} = 1$. Based on *ZIP* model, the static reflection of loads is dependent on the voltage variation.

The dynamic component of the composite load model is formulated as a three-order *IM* model with taking into account of the meteorological impacts and consuming load patterns [11]:

$$\dot{v}_d^t = \frac{-r_R^t}{x_R^t + x_m^t} \left(v_d^t + \frac{(x_m^t)^2}{x_m^t + x_R^t} i_q^t \right) + s^t v_q^t \quad (3)$$

$$\dot{v}_q^t = \frac{-r_R^t}{x_R^t + x_m^t} \left(v_q^t + \frac{(x_m^t)^2}{x_m^t + x_R^t} i_d^t \right) + s^t v_d^t \quad (4)$$

$$\dot{s}^t = \frac{1}{2H^t} \left(T_m (1 - s^t)^2 - v_d^t i_d^t - v_q^t i_q^t \right) \quad (5)$$

where $v_{d/q}^t$, s^t show d/q-axis transient voltage and rotor slip as state variables, while parameters r_R^t , x_R^t , x_m^t , H^t and T_m represent rotor resistance, rotor reactance, magnetizing reactance, motor inertia, and mechanical torque, respectively. The *IM* model is described by voltage, current, and slip of *IM* as the state variables, while the parameters need to be identified are resistance/reactance of the stator/rotor, motor inertial, and magnetizing reactance.

In addition, d/q-axis stator currents are shown by $v_{d/q}^t$ and computed, as:

$$i_d^t = \frac{r_S^t (u_d^t - v_d^t) + x_{sc}^t (u_q^t - v_q^t)}{(r_S^t)^2 + (x_{sc}^t)^2} \quad (6)$$

$$i_q^t = \frac{r_S^t (u_q^t - v_q^t) + x_{sc}^t (u_d^t - v_d^t)}{(r_S^t)^2 + (x_{sc}^t)^2} \quad (7)$$

where d/q- axis bus voltage, stator resistance, and short-circuit reactance are respectively depicted by $u_{d/q}^t$, r_s^t and x_{sc}^t . The measured bus voltage V_b^t consists of d and q-axis component, as $(V_b^t)^2 = (u_d^t)^2 + (u_q^t)^2$. Furthermore, x_{sh}^t obtains algebraically, as

$$x_{sh}^t = x_S^t + \frac{x_m^t x_R^t}{x_m^t + x_R^t} \quad (8)$$

The consuming active and reactive power of the *IMs* as time-varying parameters is modeled as:

$$P_{IM}^t = \frac{\left\{ r_S^t \left[(u_d^t)^2 + (u_q^t)^2 - u_d^t v_d^t - u_q^t v_q^t \right] - x_{sh}^t (u_d^t v_q^t - u_q^t v_d^t) \right\}}{\left[(r_S^t)^2 + (x_{sh}^t)^2 \right]} \quad (9)$$

$$Q_{IM}^t = \frac{\left\{ x_{sh}^t \left[(u_d^t)^2 + (u_q^t)^2 - u_d^t v_d^t - u_q^t v_q^t \right] - r_S^t (u_d^t v_q^t - u_q^t v_d^t) \right\}}{\left[(r_S^t)^2 + (x_{sh}^t)^2 \right]} \quad (10)$$

Consequently, the consuming active and reactive power in the composite load model are:

$$P_{com}^t = P_{ZIP}^t + P_{IM}^t \quad (11)$$

$$Q_{com}^t = Q_{ZIP}^t + Q_{IM}^t \quad (12)$$

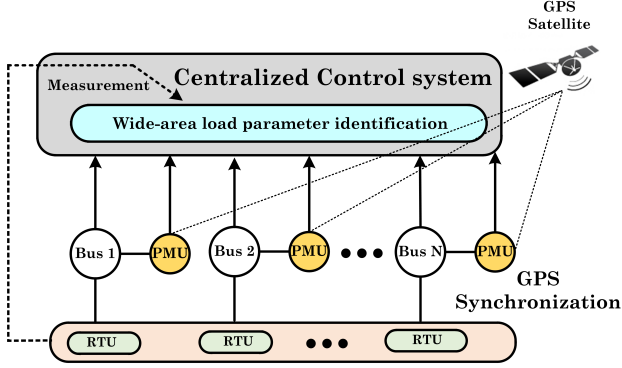


Fig. 2. Overall schematic on the practical implementation of the wide-area measurement based load parameter identification

B. Conventional Load Parameter Identification

To estimate the time-varying load parameters, two sets of parameters including load and measurable variables are considered which denoted as Θ_d^t and Υ_d^t , respectively. In this regard, these parameters are $\Theta_{d^\Theta}^t = [r_S^t, x_S^t, x_m^t, x_R^t, r_R^t, H^t, \alpha_t^P, \alpha_t^Q, \beta_t^P, \beta_t^Q]$ and $\Upsilon_{d^Y}^t = [P_{com}^t, Q_{com}^t, V_b^t]$. The dimension of CLM and measured parameters by WAMS are respectively represented by d^Θ and d^Y .

In overall, the composite load model shows a function depended on the measured variables associated with measurement/process noises, $\Theta_{d^\Theta}^t = f(\Upsilon_{d^Y}^{t-r}) + e_Y$.

In the k time interval, CLM parameters can be as follows:

$$\Theta_{d^\Theta}^t = \Theta_{d^\Theta}^{t-1} + e_{\Theta 1} = \Theta_{d^\Theta}^{t-2} + e_{\Theta 2} = \dots = \Theta_{d^\Theta}^{t-k_\Theta} + e_{\Theta k_\Theta} \quad (13)$$

where k_Θ represent the estimation time horizon of CLM parameters. Similarity, the dependency of CLM parameters and measurement variables can be describes as:

$$\Theta_{d^\Theta}^t = f(\Upsilon_d^t) + e_Y = f_1(\Upsilon_d^{t-1}) + e_{Y1} = \dots = f_{k_Y}(\Upsilon_d^{t-k_Y}) + e_{Y k_Y} \quad (14)$$

where k_Y is a window length of measurement.

C. Wide-Area Load Parameter Identification

From the practical point of view, wide-area load modeling is carried out based on the wide-area measurement systems. Firstly, the information in the whole power system should be send to the centralized control center. Then, the control center organized the received data as an input for the designed data-driven based load parameter identification block. The input data activates the trained network. Formerly, the load parameters of the system are estimated by the designed network. The mentioned procedure is shown in Fig. 2.

The correlation of the electrical load consumption has taken into account in the load parameter identification via WAMS. Therefore, CLM load models are described based on $F_{d^Y}^t$ as:

$$\chi_{d_x, i}^t = F_{d^Y}^t (\chi_{d_x, 1}^t, \dots, \chi_{d_x, N}^t) \quad (15)$$

The bus number is represented by i , while N depicts the total number of the buses in the network.

Considering (14) and (15), CLM based on WAMS can be formulated as follows [4]:

$$\Theta_{d^\Theta}^t = \mathcal{F}_X \left(\begin{array}{c} \Theta_{d^\Theta}^{t-1}, \dots, \Theta_{d^\Theta}^{t-k_\Theta} \\ \underbrace{\chi_{d_x \times (k_Y+1), 1}^t, \dots, \chi_{d_x \times (k_Y+1), 1}^{t-k_Y}}_{\chi_{d_x \times (k_Y+1)}}, \underbrace{\chi_{d_x \times (k_Y+1), 2}^t, \dots, \chi_{d_x \times (k_Y+1), 2}^{t-k_Y}}_{\chi_{d_x \times (k_Y+1)}}, \dots, \underbrace{\chi_{d_x \times (k_Y+1), N}^t, \dots, \chi_{d_x \times (k_Y+1), N}^{t-k_Y}}_{\chi_{d_x \times (k_Y+1)}} \end{array} \right) \quad (16)$$

where $\Theta_{d^\Theta \times k_\Theta}$, $\chi_{d_x \times (k_Y+1)}$ and $\chi_{d_x \times (k_Y+1)}$ are:

$$\Theta_{d^\Theta \times k_\Theta} = \left(\Theta_{d^\Theta}^{t-1}, \Theta_{d^\Theta}^{t-2}, \dots, \Theta_{d^\Theta}^{t-k_\Theta} \right) \quad (17)$$

$$\chi_{d_x \times (k_Y+1)} = \left(\dots, \chi_{d_x \times (k_Y+1), i}, \dots, \chi_{d_x \times (k_Y+1), N} \right) \quad (18)$$

$$\chi_{d_x \times (k_Y+1), i} = \left(\chi_{d_x, i}^t, \chi_{d_x, i}^{t-1}, \dots, \chi_{d_x, i}^{t-k_Y} \right) \quad (19)$$

An analytical model cannot estimate the parameter in (16), therefore, this paper proposes a data-driven method to identify the time-varying parameters.

III. WIDE-AREA CLM MODELING BASED ON MULTI-TASK DEEP LEARNING APPROACH

In the data-driven load modeling, historical data including $\Theta_{d^\Theta \times k_\Theta, i} \in R^{d^\Theta \times k_\Theta}$ and $\chi_{d_x \times (k_Y+1)} \in R^{d_x \times (k_Y+1)}$ is the basis to approximate a function with the output of $\Theta'_{d^\Theta \times k_\Theta, i} \in R^{d^\Theta \times k_\Theta}$. The data-driven structure should be able to generate the output with minimum difference between $\Theta_{d^\Theta \times k_\Theta, i}$ and $\Theta'_{d^\Theta \times k_\Theta, i}$. The input dataset defines as $X = \left\{ \left(\Theta_{d^\Theta}^{t-k_\Theta}, \dots, \Theta_{d^\Theta}^{t-1} \right), \left(\chi_{d_x, i}^{t-k_Y}, \chi_{d_x, i}^{t-1} \right) \right\}$ and the output is $Y = \left\{ \left(\Theta_{d^\Theta}^t \right) \right\}$ in the data-driven load parameter estimation. To design a structure to estimate CLM parameters, the squared error loss function is usually used.

The conventional loss function is a squared error loss function. The main shortcoming of this loss function is the possibility of mean biased or estimation of minimum variance. To address this issue, this paper reformulates the loss function with a novel approach entitled as Pseudo-Huber loss function as follows [29]:

$$f_{\text{loss}}^{pH}(Y) = \sum_{t=1}^T \left[C^2 \sqrt{\frac{c^2 + (Y_t)^2}{C^2}} \right] - C^2 \quad (20)$$

where f_{loss}^{pH} represents loss function and control parameter, respectively. This loss function outputs the values lose to $\frac{(Y)^2}{2}$. Therefore, it is prevented the large values due to producing the straight line with the slope $\frac{(Y)^2}{2}$. Furthermore, the performance of the suggested loss function would not be significantly affected by external factors including noises. Thus, the learning weights θ_l of each l^{th} layer can be obtained as:

$$\theta_l^{pH}(X) = \frac{\theta_l(X)}{\sqrt{1 + \left(\frac{\tilde{Y}_{pH} - Y_l}{C} \right)^2}} \quad (21)$$

Moreover, the estimator of proposed pseudo-Huber loss function, \tilde{Y}_{pH} , is defined as follows:

$$\tilde{Y}_{pH} = \frac{\sum_{t=1}^T \theta_l^{pH}(X) Y_l}{\sum_{t=1}^T \theta_l^{pH}(X)} \quad (22)$$

The efficient and well-known gradient descent based, Adam [30] [31] is utilized to find the optimal learning weights. That is why, the developed loss function have high degree of robustness against various noises such as measurement and process ones in CLM.

Although a loss function plays a key role in designing a strong approach in time-varying CLM, it is essential to design a network that can understand Spatio-temporal features of time signals in the power systems. Thus, this paper presents a multi-task residual Spatio-temporal deep consists of three parallel parts and each part consists of four main blocks i.e. residual CNN, GRU, fully-connected networks (FCNs), and weight fusion blocks. The proposed structure is depicted in Fig. 3.

A. Arrangement of Input Dataset

In the first step, the measurement and load parameters obtained by a generated dataset is normalized based on [32]:

$$X^i = \frac{x^i - X_{\min}^i}{X_{\max}^i - X_{\min}^i} \quad (23)$$

the maximum and minimum input values are shown as X_{\max}^i and X_{\min}^i , respectively.

The $1D$ time series is not sufficient for complete feature extraction. Therefore, the representation learning technique is applied in which the initial dataset, which is extensively assessed using real input, is converted to $2D$ -vectors through the following equation [33]:

$$X^i = \begin{bmatrix} \Theta_d^{t-1} & \chi_{d_x}^{t-1} \\ \Theta_{d^\theta}^{t-2} & \chi_{d_x}^{t-2} \\ \vdots & \vdots \\ \Theta_{d^\theta}^{t-k_\theta} & \chi_{d_x}^{t-k_\tau} \end{bmatrix} \quad (24)$$

B. Residual CNN

In the first blocks of the designed deep network, raw data is fed into a CNN block. The measurement and parameter data in the previous time intervals construct a sparse and noisy input dataset. In the first block, convolutional layers and max-polling layers are used to extract time-varying patterns and spatial dependency between measurement signals from different busses in the power system. The convolutional layers consist of multiple kernels with a size of $\underbrace{wi}_{\text{width}} \times \underbrace{he}_{\text{height}}$. The output of l^{th} convolution layer is as follows [34]:

$$O_{l,i}^{Conv} = f^{act} (\omega_{l,i}^{Conv} \otimes X_i + B_{l,i}) \quad (25)$$

where $O_{l,i}^{Conv}$, $\omega_{l,i}^{Conv}$ and $B_{l,i}$ represent output, weight matrix, and bias matrix of the l^{th} convolutional layer, respectively. Also, function f^{act} and operator \otimes show activation function

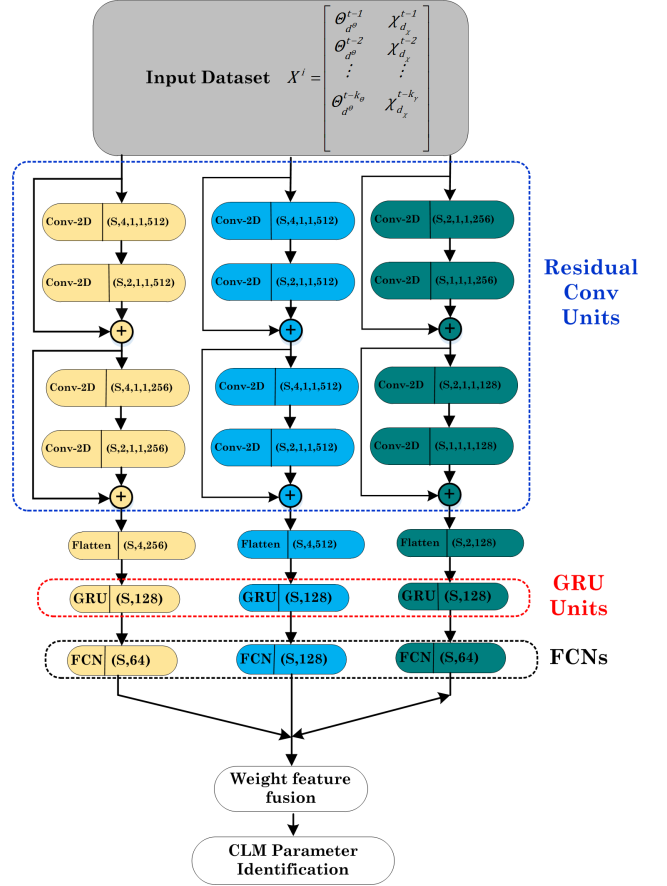


Fig. 3. Proposed multi-residual deep network structure for CLM parameter estimation

and convolutional operator. As can be seen from (25), the input measurement and load parameters are filtered based on the activation function for the several times. Also, the input data convolved to extract the inherent features including spatial features (loads at different locations).

To enhance the CNN performance in capturing spatial features, we added a residual mapping after two convolutional layers, therefore, the output after passing through a residual network is:

$$O_i^{res} = f^m + X_i^{res} \quad (26)$$

where the output of the residual unit, mapping element, and input of the residual units are shown by O_i^{res} , f^m , and X_i^{res} respectively.

C. GRU

In the second blocks of three parallel structures, GRU units are added to the designed multi-residual deep network to capture fully temporal feature as a time-efficient version of memory-based recurrent neural networks. GRU units as the modified version of LSTM consist of update and reset gates. The update gate stores the importance features throughout a signal and realize long-term dependency, while reset gate removes features with low weight importance by resetting memory [3].

D. FCNs

The FCN layers are used to control the dimension of the GRU, the three parallel FCN layers are trained in an end-to-end manner and improve the training performance without further machinery [34].

E. Weighted Fusion Block

In the time-varying composite model, different parameters in static and dynamic model have a different impact on the parameter identification performance. For example, in ZIP load model, the parameters might be almost constant, in particular during the normal operational conditions, while dynamic parameters could highly change during an ultra-short-term period, such in the millisecond time period. Thus, different influential factors could highly influence on the parameter estimation with different behavior [35]. To this end, firstly we designed a three parallel network, then, a fusion method is used to construct final outputs as:

$$O^{fu} = \omega^{P_1} \odot O^{P_1} + \omega^{P_2} \odot O^{P_2} + \omega^{P_3} \odot O^{P_3} \quad (27)$$

where O^{fu} , $\omega^{P_1/P_2/P_3}$ and $O^{P_1/P_2/P_3}$ show output of weight fusion block, each parallel part of the designed Multi-residual network weights and outputs, respectively. The output of weight fusion block is the final output of the proposed deep approach for CLM parameters Θ_d^t with aim of minimum difference with with aim of minimum difference with Θ_d^t .

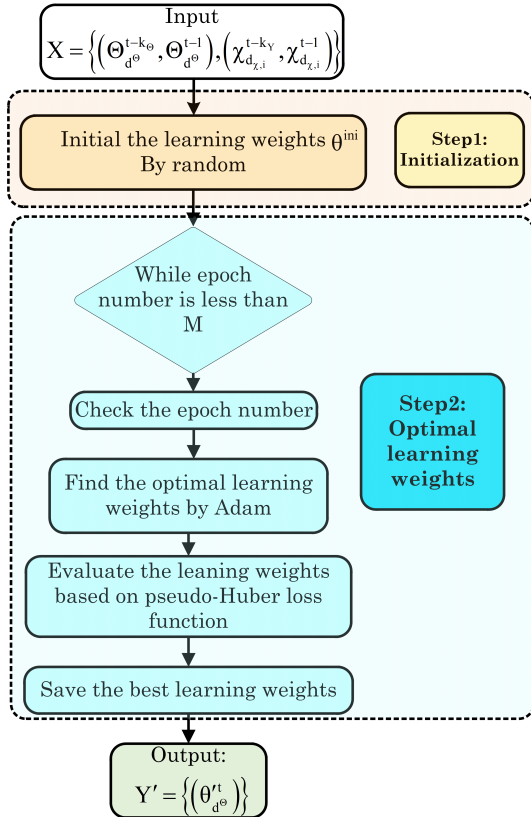


Fig. 4. The training process of the multi-residual deep network

F. Training Process

The training process is given in Fig. 4. As can be seen, training of the designed deep learning based CLM parameter identification is simple. Firstly, the designed network is fed by a set of historical data, then an iterative process is conducted to obtain a set of optimal learning weights by optimizing a pseudo-Huber loss function and Adam algorithm.

G. Design Procedure

To provide a balance between the slow and fast variants in the composite load model (CLM) parameters identification, the designed structure composes of three parallel branches as follows:

- 1) The first branch is trained based on the measurement parameters and parameters of the ZIP model $(\alpha_t^P, \beta_t^P, \alpha_t^Q, \alpha_t^Q)$.
- 2) The second branch is trained using measurement parameters, rotor, and stator parameters of the IM $(x_S^t, x_R^t, r_S^t, r_r^t)$.
- 3) The last branch is trained based on the H^t and x_m^t .

As can be seen from Fig 1, each branch includes two residual CNN, one GRU, and one FCN layer. To this end, the proposed network is designed based on the following steps:

- Step 1: Initialization

Step 1.1: The input data (measurement and parameters) is normalized based on (23).

Step 1.2: The normalized data is converted from 1D to 2D-signals.

- Step 2: Residual CNN

Step 2.1: As above-mentioned, three parallel blocks have been considered to provide a balance between the slow and dynamic behavior of the parameters in the static and dynamic model. Each block consists of two residual CNN and the first convolutional layer converts the input to (S,4,1,1,512) (S: samples) into the two parallel branches and the output of the last branch (including measurement data and historical data of H^t and x_m^t) form is (S,2,1,1,512). Furthermore, about 15% of the layers are dropped out in the first residual CNN layers. The dropout approach is a beneficial and widely used technique to prevent overfitting as well as improve the computational burden of the designed network.

Step 2.2: In the second residual CNN, the output of the first residual CNN converts to the outputs with sizes (S,2,1,1,512) into the two parallel layers (including IM rotor, IM stator, and ZIP parameters), while the form of the outputs in the last parallel layer is (S,2,1,1,512). In the second residual CNN layers, about 25% of the layers are dropped.

- Step 3: GRU

Step 3.1: The outputs of residual CNN blocks are 2D signals. Firstly, 2D signals convert to the 1D signals by using flatten layer. Therefore, flatten layers are used to convert 2D signals to 1D signal with the size of (S,4,256), (S,4,512), and (S,2,218) in the first, second, and third parallel layers.

Step 3.2: The 1D outputs of flattening layers are fed into the three sets of GRU layers, each branch consists of one of the

GRU layers. The output of GRU layers is the time series with the size of (S,128).

- Step 4: FCN

Step 4.1: The output of GRU is fed into the FCN blocks. FCN block converts the outputs of GRU into the vectors with size (S,64), (S,128), and (S,64) in the first, second, and third parallel layers, respectively.

- Step 5: Output

Step 5.1: The outputs of the FCN blocks are fed into the weight fusion block. The weight fusion block is used to concatenate the three parts and then construct the final outputs. These outputs are the results of CLM parameter identification.

IV. NUMERICAL EXPERIMENTS

The numerical results of the designed deep-based CLM parameter identification structure is given in this section. The robustness and effectiveness of the proposed multi-residual deep network are validated using two benchmark case studies with regards to noise impacts. For sake of the comparison, four different data-driven approaches including multi-LSTM [4], single LSTM [4] and 1D-CNN [36] as previously presented deep networks and SVM [14] as shallow-based network have been considered. Besides, to show the GRU units impacts on the performance of the proposed approach, we consider the designed network without GRU units and as a state-of-the-art approach, multi-residual convolutional (MRC) approach is also used for assessing the performance of the developed technique.

The designed network is coded in Python (TensorFlow package) and the two datasets are gathered from MATLAB program in a PC system with Core I7 CPU @3.00 GHz RAM.

A. Data Generation and Description

The first benchmark system considered for this study is IEEE 68-bus system. This system involves 86 lines and 16 generators. To simulate the network, the power system toolbox (PST) with 0.01 samples per second has been utilized [37]. Two various methods are deliberated to generate the datasets. To illustrate the generation technique with an example, consider that IEEE 68-bus system has 34 different loads which are disconnected from the system each a time, while the data is recored. Then, this trend is taken into account for the lines. In other words, they also disconnected from the system while the data is being captured. This procedure leads to generating more than 59500 various samples from IEEE 68-bus system. This obtained data is divided to three subsections including training, validation and testing samples. In this regard, 70% of the data is assigned to training stage, while 15% of samples are considered for the validation and the remaining 15% of this dataset is devoted to testing process. It must be noted that the noisy data is generated based on Normal distribution function with the mean values equaled to mean values of original data and standard deviation equaled to 10% of the mean values. These produced noisy samples are also considered in the evaluation of the developed technique.

B. Evaluation Metrics

Four important and critical error-based metrics are considered to assess the performance of the designed network from numerical aspects. The formulation of these indices are brought as follows:

- Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y'_i - Y_i)^2} \quad (28)$$

- Normalized root mean square error (NRMSE):

$$NRMSE = \frac{1}{Y_i^{\max}} \sqrt{\frac{1}{N} \sum_{i=1}^N (Y'_i - Y_i)^2} \quad (29)$$

- Mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y'_i - Y_i| \quad (30)$$

- Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y'_i - Y_i}{Y_i} \right| \quad (31)$$

C. Discussion on Results

Figures 5 and 6 show the comparative results of the proposed method with the actual values and estimated parameters by the MCR approach. In these figures, it can be seen that the proposed approach follows the pattern of measured α_p^t and x_m^t with high degree of accuracy. This results indicate that the proposed approach has high capabilities in this regard.

The performance of the proposed approach in terms of four different accuracy indices with regard to the defined equations (28)-(31) are given in Fig. 7. As can be seen from this figure, the results validate the high accuracy of the proposed approach.

The comparative study between the mentioned methods and the suggested technique for the estimation of α_p^t and x_m^t are given in Figs. 8 and 9. From these figures, it is clear that the proposed approach is far more accurate than the state-of-the-art approaches. For instance, based on Fig. 8, the proposed approach has a higher accuracy percentage in terms of MAPE than MRC, MLSTM, LSTM, 1D-CNN, and SVM which are 73.14%, 77.70%, 90.42%, 93.73% and 98.35%, respectively. Also, Fig. 9 denotes that the developed scheme is more accurate in terms of NRMSE than SVM which its performance is 94.80%. Also the developed scheme has better performance than MLSTM, LSTM, and 1D-CNN which have approximately 76.50%, 89.51%, 89.63% percentage in terms of NRMSE respectively. The comparison between the proposed approach and MRC shows the effectiveness of the GRU units with improving accuracy significantly, about 74.20% and 69.24% based on NRMSE values in estimation of parameters, α_p^t and x_m^t , respectively.

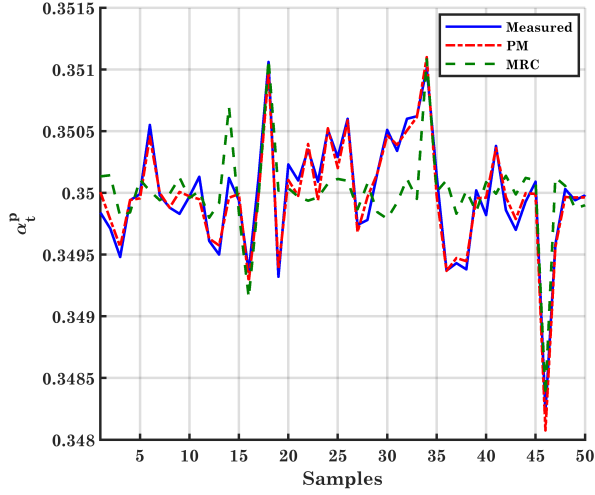


Fig. 5. The estimation results of α_t^p obtained by proposed method (PM) and MRC approaches versus actual data

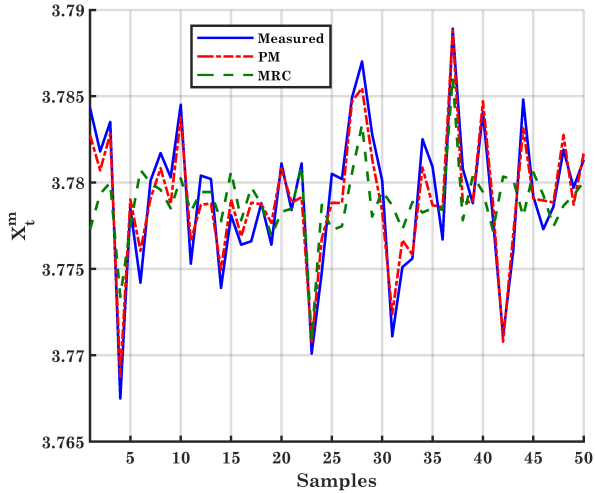


Fig. 6. The estimation results of x_t^m obtained by proposed method (PM) and MRC approaches versus actual data

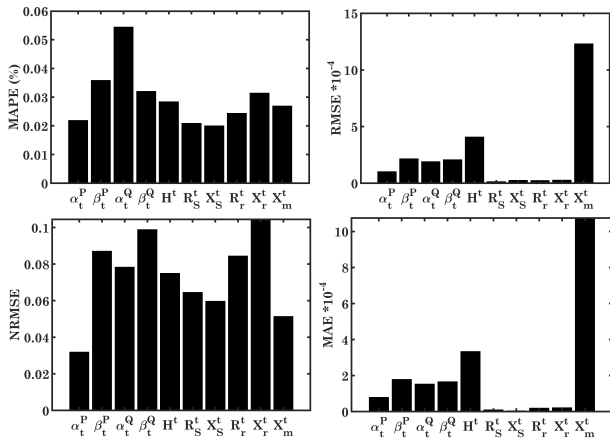


Fig. 7. Obtained metrics by the proposed method in wide-area load modeling

D. Sensitivity Analysis on Parallel Structure

To analyse the effectiveness of the proposed approach with different parallel parts, a sensitivity study is conducted.

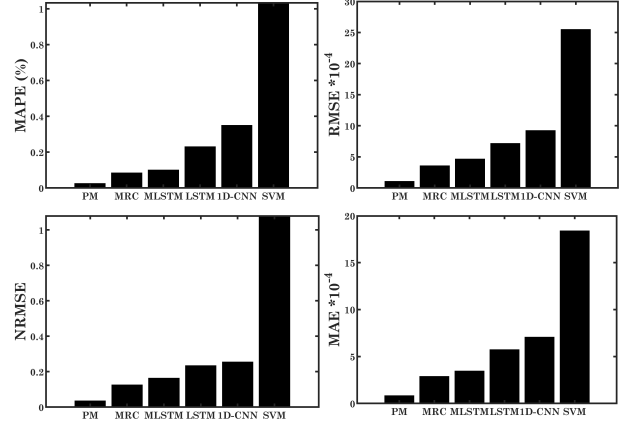


Fig. 8. The estimation results of α_t^p obtained by proposed method (PM) in parameter identification of CLM with MRC, MLSTM, LSTM, 1D-CNN and SVM schemes

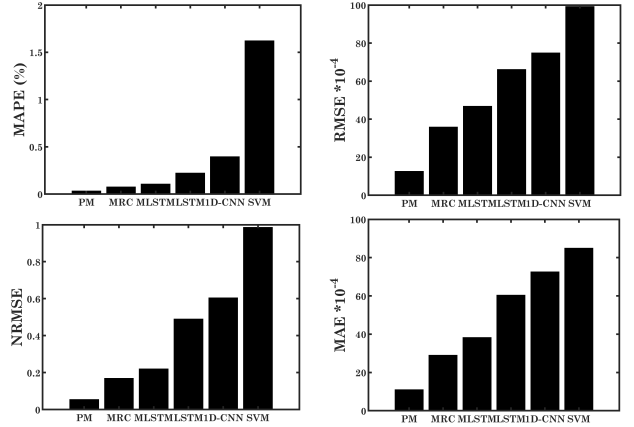


Fig. 9. The estimation results of x_t^m obtained by proposed method (PM) in parameter identification of CLM with MRC, MLSTM, LSTM, 1D-CNN and SVM schemes

The results of time-varying load parameter identification for parameter x_m^t are depicted in Fig. 10 with regards to the mentioned four different indices. As can be realized, the performance of the proposed approach with three different parts are more accurate than the scheme with 1, 2, 4, and 5 parts. With increasing number of the parts from 1 to 3, the accuracy is enhanced. However the designed deep network with 5 different parts shows lower accuracy than with 3 and 4 parts. For example, the proposed approach with 3 parallel layer is significantly reduced the error in terms of MAPE associated with 1, 2, 4, and 5 parallel layers approximately 80.32%, 63.37%, 32.41%, and 72.83%, respectively. It can be stated that the higher number of parallel branches leads to overfitting.

E. Sensitivity Analysis on Noise

The robustness of the proposed approach is discussed in this subsection through a comparison with other approaches. To this end, five different noise signals are added to the actual values; each of them follows a Gaussian distribution with zero mean and standard deviation of 0.005, 0.01, 0.015, 0.02, 0.025,

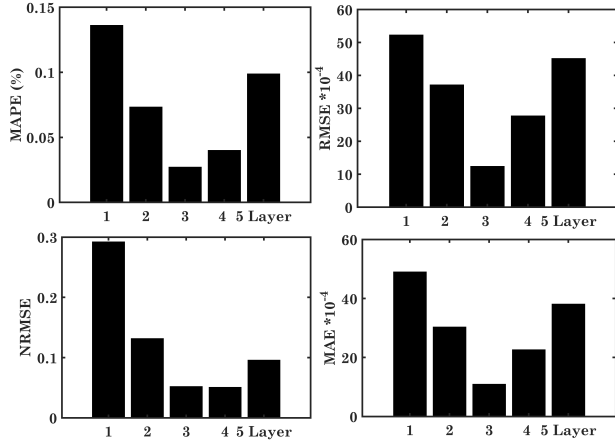


Fig. 10. Comparison of different structures of the proposed method in parameter identification of CLM with 1-5 parallel layers in estimation of x_t^m

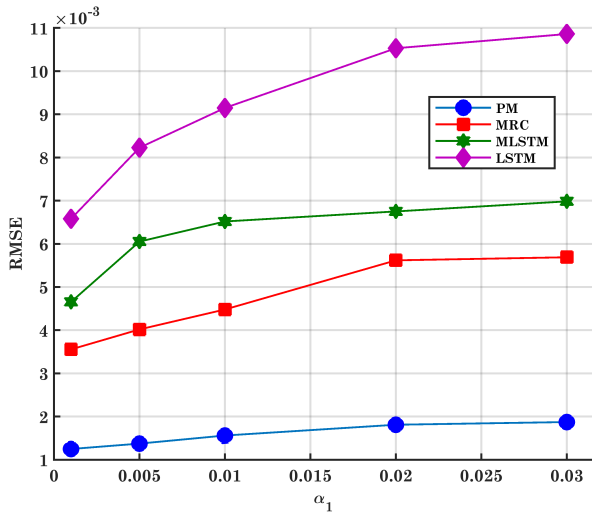


Fig. 11. Proposed multi-residual deep network structure for CLM parameter Comparison of the proposed method (PM), MRC, MLSTM, and LSTM results on x_t^m estimation in terms of RMSE

and 0.03 (shown by α_1). Figures 11 and 12 compared the results for estimation of x_t^m obtained by the proposed approach, MRC, MLSTM, and LSTM in terms of RMSE and MAPE indices. As can be seen, the proposed approach are far more accurate than other approaches and the results demonstrated the robustness of the proposed approach in noisy conditions. For instance, in highly noise condition with standard deviation 0.03, the proposed approach is more accurate than MRC, MLSTM, and LSTM, about 64.61%, 80.91%, and 86.91%, respectively, based on Fig. 11. From Fig. 12 it is clear that, MAPE of the designed network, MCR, MLSTM, and LSTM are 0.001872, 0.005689, 0.006983, and 0.01086, respectively, in highly noise condition ($\alpha_1 = 0.03$). These MAPE value shows that proposed approach has improved accuracy of multi residual convolutional neural network based CLM parameter identification significantly, about four times better. Besides, the error of the proposed approach in estimation of x_t^m are at least 5 and 7 times less than MLSTM and LSTM, respectively.

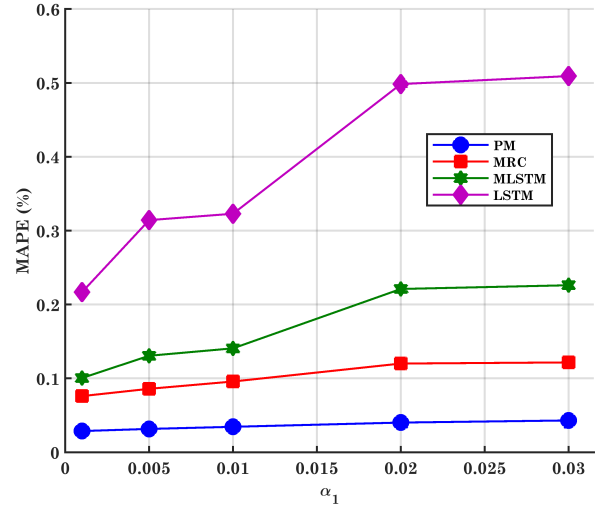


Fig. 12. Proposed multi-residual deep network structure for CLM parameter Comparison of the proposed method (PM), MRC, MLSTM, and LSTM results on x_t^m estimation in terms of MAPE

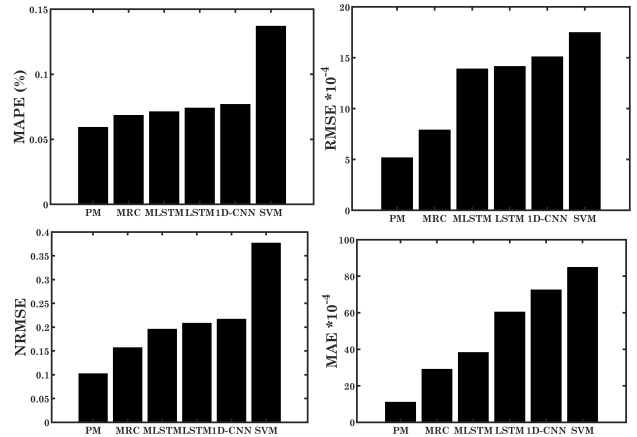


Fig. 13. The estimation results of H^t obtained by proposed method (PM) in parameter identification of CLM with MRC, MLSTM, LSTM, 1D-CNN and SVM schemes in Iranian 95-bus system

F. Further Results: Iranian 95-bus Network

To verify the generality of the proposed method, the real network of Iranian 95-bus network is considered [38]. To this end, the proposed approach is tested on the data gathered from this network, in which, the results obtained by the proposed on the estimation of the H^t of the load connected to the bus 63 (more information is provided in [38]). The results are compared based on some metrics shown in Fig. 13. As can be seen, it is clear that the designed multi-residual deep network is superior over the compared methods. Therefore, based on attained results on the IEEE-68 bus system and Iranian 95-bus systems, the effectiveness as well as generality of the proposed approach is demonstrated.

V. CONCLUSION AND FUTURE WORKS

In the power system studies, it is essential to model the loads with highest accuracy to achieve more precise results. The proper load modeling procedure has two consecutive stages which are a selecting practical load model and an

developing a powerful parameter identification scheme. In this paper, the wide-area CLM is selected as a realistic and practical load model. Furthermore, a fast and accurate deep neural network, namely multi-residual deep neural network is designed to achieve the unknown parameters. The proposed approach benefits from a residual convolutional layers to capture spatial and robust features of measurement signal and used GRU unit to realize fully temporal features. To make a balance between learning dynamic and static behavior of CLM parameters, the proposed network is designed in three parallel manner. The training ability and noise immunity of the designed technique is boosted by a reformulated error loss function named pseudo-Huber loss function. The methodology is assessed on two different case studies including IEEE 68-bus and Iranian 95-bus. The results network verifies the superiority of the proposed approach through comparison with previous approaches such as MLSTM, LSTM, and SVM, with at least more than 70% accuracy improvement. To address the GRU application in the designed network, the proposed approach is compared with MCR (the proposed approach without GRU units) and shows at least 50% accuracy improvement. The sensitivity analysis on noise shows the proposed is almost robust in different noisy conditions and shows at least 3 times less error compared with other approaches. Finally, a three parallel structure efficiency is demonstrated by a comparative analysis with different structure of the designed network.

The investigations on the proposed CLM parameter estimation based on WAMS reveal that further explorations in the following directions would be worthwhile:

- 1) Estimation of the full statistical information of the electrical loads in from of probability density function (PDF) instead of point estimations of load parameters.
- 2) Developing parameter identification models for composite demand side models such as combination of load, renewable power generations, and energy storage devices. In this case, the number of the parameters and the model would be change and the uncertainty associated with renewable generations can be a serious challenge in the modern power systems.

REFERENCES

- [1] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, "Load modeling—a review," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5986–5999, 2017.
- [2] X. Wang, Y. Wang, D. Shi, J. Wang, and Z. Wang, "Two-stage wecc composite load modeling: A double deep q-learning networks approach," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 4331–4344, 2020.
- [3] F. Bu, Z. Ma, Y. Yuan, and Z. Wang, "Wecc composite load model parameter identification using evolutionary deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5407–5417, 2020.
- [4] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6102–6114, 2019.
- [5] Z. Ma, B. Cui, Z. Wang, and D. Zhao, "Parameter reduction of composite load model using active subspace method," *IEEE Transactions on Power Systems*, 2021.
- [6] M. Khodayar and J. Wang, "Probabilistic time-varying parameter identification for load modeling: A deep generative approach," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 1625–1636, 2020.
- [7] Y. Lin, Y. Wang, J. Wang, and D. Shi, "Tensor-based parameter reduction of dynamic load models with variable frequency drive," *IEEE Transactions on Power Systems*, 2021.
- [8] G. Dudek, P. Pełka, and S. Smył, "A hybrid residual dilated lstm and exponential smoothing model for midterm electric load forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [9] M. Kabiri and N. Amjady, "A hybrid estimation and identification method for online calculation of voltage-dependent load parameters," *IEEE Systems Journal*, vol. 13, no. 1, pp. 792–801, 2018.
- [10] C. Zheng, S. Wang, Y. Liu, C. Liu, W. Xie, C. Fang, and S. Liu, "A novel equivalent model of active distribution networks based on lstm," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2611–2624, 2019.
- [11] J.-K. Kim, K. An, J. Ma, J. Shin, K.-B. Song, J.-D. Park, J.-W. Park, and K. Hur, "Fast and reliable estimation of composite load model parameters using analytical similarity of parameter sensitivity," *IEEE transactions on Power Systems*, vol. 31, no. 1, pp. 663–671, 2015.
- [12] C. Wang, Z. Wang, and S. Ma, "Svm-based parameter identification for static load modeling," in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*. IEEE, 2018, pp. 1–5.
- [13] Y. Lin, Y. Wang, J. Wang, S. Wang, and D. Shi, "Global sensitivity analysis in load modeling via low-rank tensor," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2737–2740, 2020.
- [14] C. Wang, Z. Wang, J. Wang, and D. Zhao, "Svm-based parameter identification for composite zip and electronic load modeling," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 182–193, 2018.
- [15] A. Ellis, D. Kosterev, and A. Meklin, "Dynamic load models: Where are we?" in *Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES*, no. 2. IEEE Brighton, England, 2006, pp. 1320–1324.
- [16] E. O. Kontis, T. A. Papadopoulos, A. I. Chrysochos, and G. K. Papagiannis, "Measurement-based dynamic load modeling using the vector fitting technique," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 338–351, 2017.
- [17] X. Zhang, D. J. Hill, and C. Lu, "Identification of composite demand side model with distributed photovoltaic generation and energy storage," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 1, pp. 326–336, 2019.
- [18] Q. Huang, R. Huang, B. J. Palmer, Y. Liu, S. Jin, R. Diao, Y. Chen, and Y. Zhang, "A generic modeling and development approach for wecc composite load model," *Electric Power Systems Research*, vol. 172, pp. 1–10, 2019.
- [19] S. Yu, S. Zhang, and X. Zhang, "Two-step method for the online parameter identification of a new simplified composite load model," *IET Generation, Transmission & Distribution*, vol. 10, no. 16, pp. 4048–4056, 2016.
- [20] M. E.-N. Jahromi and M. T. Ameli, "Measurement-based modelling of composite load using genetic algorithm," *Electric Power Systems Research*, vol. 158, pp. 82–91, 2018.
- [21] P. Regulski, D. Vilchis-Rodríguez, S. Djurović, and V. Terzija, "Estimation of composite load model parameters using an improved particle swarm optimization method," *IEEE Transactions on Power Delivery*, vol. 30, no. 2, pp. 553–560, 2014.
- [22] Y. Wang, C. Lu, and X. Zhang, "Applicability comparison of different algorithms for ambient signal based load model parameter identification," *International Journal of Electrical Power & Energy Systems*, vol. 111, pp. 382–389, 2019.
- [23] C. Wang, Z. Wang, J. Wang, and D. Zhao, "Robust time-varying parameter identification for composite load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 967–979, 2017.
- [24] A. M. Najafabadi and A. T. Alouani, "Real time parameter identification of composite load model," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.
- [25] A. Rouhani and A. Abur, "Real-time dynamic parameter estimation for an exponential dynamic load model," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1530–1536, 2015.
- [26] A. Keyhani, W. Lu, and G. T. Heydt, "Composite neural network load models for power system stability analysis," in *IEEE PES Power Systems Conference and Exposition, 2004*. IEEE, 2004, pp. 1159–1163.
- [27] H. Kassaei, A. Keyhani, T. Woung, and M. Rahman, "A hybrid fuzzy, neural network bus load modeling and predication," *IEEE Transactions on Power Systems*, vol. 14, no. 2, pp. 718–724, 1999.
- [28] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 865–876, 2019.
- [29] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4331–4339.

- [30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [31] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Designing a composite deep learning based differential protection scheme of power transformers," *Applied Soft Computing*, vol. 87, p. 105975, 2020.
- [32] S. Afrasiabi, M. Afrasiabi, M. Mohammadi, and B. Parang, "Fault localisation and diagnosis in transmission networks based on robust deep gabor convolutional neural network and pmu measurements," *IET Generation, Transmission & Distribution*, vol. 14, no. 26, pp. 6484–6492, 2020.
- [33] S. Afrasiabi, M. Mohammadi, M. Afrasiabi, and B. Parang, "Modulated gabor filter based deep convolutional network for electrical motor bearing fault classification and diagnosis," *IET Science, Measurement & Technology*, vol. 15, no. 2, pp. 154–162, 2021.
- [34] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, "Advanced deep learning approach for probabilistic wind speed forecasting," *IEEE Transactions on Industrial Informatics*, 2020.
- [35] S. M. H. Rizvi, "Time series deep learning for robust steady-state load parameter estimation using 1d-cnn," *Arabian Journal for Science and Engineering*, pp. 1–14, 2021.
- [36] M. Afrasiabi, M. Mohammadi, M. Rastegar, and S. Afrasiabi, "Deep learning architecture for direct probability density prediction of small-scale solar generation," *IET Generation, Transmission & Distribution*, vol. 14, no. 11, pp. 2017–2025, 2020.
- [37] J. H. Chow and K. W. Cheung, "A toolbox for power system dynamics and control engineering education and research," *IEEE transactions on Power Systems*, vol. 7, no. 4, pp. 1559–1564, 1992.
- [38] M. Amini, A. Jalilian, and M. R. P. Behbahani, "A new method for evaluation of harmonic distortion in reconfiguration of distribution network," *International Transactions on Electrical Energy Systems*, vol. 30, no. 6, p. e12370, 2020.

