Leveraging Deep Learning to Improve Performance of Distributed Optimal Frequency Control under Communication Failures

Siyu Xie, Masoud H. Nazari*, *Senior Member, IEEE,* Farinaz Nezampasandarbabi, *Student Member, IEEE,* Le Yi Wang, *Life Fellow, IEEE*

***Abstract*—This paper proposes a deep learning approach to overcome the impacts of communication failures on the performance and convergence rate of the distributed optimal frequency control (DOFC) for power systems. Novel features of the proposed framework are fourfold. First, the nonlinear model of the DOFC is developed to consider for nonlinearities of power flows. Second, the long short-term memory (LSTM) algorithm is used for dynamic model estimation during communication failures. Next, the LSTM-based DOFC method is introduced to cope with the impact of communication failures on the performance of the distributed control strategy. Finally, we prove the convergence of LSTM-DOFC and show that the algorithm has superior performance compared to the linearized prediction methods, such as autoregressive–moving-average models. Simulations on two real-world power systems are carried out to demonstrate the effectiveness of the proposed framework.**

***Index Terms*—Communication failures, deep learning, distributed optimal frequency control, fast convergence, multi-agent network, prosumer.**

NOMENCLATURE

$ARMA$:	Autoregressive–moving-average
$\mathcal{A}$:	Quasi-steady-state system matrix for frequency control
B_{ij} :	Equivalent susceptance between prosumer i and j
$\mathcal{B}$:	Quasi-steady-state control matrix for frequency control
$DOFC$:	Distributed optimal frequency control
$\Delta\omega_i$:	Variations of angular frequency
ΔP_i :	Power deviations
$\Delta\delta_i$:	Variations of the voltage angle
$\Delta\delta_{ij}$:	$\Delta\delta_i - \Delta\delta_j$
$\delta(k)$:	$[\delta_1(k), \cdots, \delta_n(k)]^\top$
d_1^h :	Prediction error from the LSTM model at iteration h
d_2^h :	Gradient error at iteration h
$DERs$:	Distributed energy resources
EVs :	Electric vehicles
$\mathcal{F}^{(q)}(\cdot)$:	Nonlinear function of the q-layers of LSTM
γ_i :	Characteristics of the internal quasi-steady-state dynamics of prosumer i
Γ :	$\text{diag}\{\gamma_1, \cdots \gamma_n\}$, i.e., the diagonal matrix of control parameters
$H_p(\delta(k))$:	Network matrix
$LSTM$:	Long short-term memory
MPS :	Modern power systems
M_i :	Cost of frequency deviations for prosumer i
N_i :	Cost of control for prosumer i
$\mathcal{N}_i$:	Set of prosumers electrically adjacent to prosumer i
ODE :	Ordinary differential equation
$Prosumer$:	Hybrid producer-consumer agent
$P(k)$:	$[P_1(k), \cdots, P_n(k)]^\top$
σ_i :	Droop constant of prosumer i
Σ :	$\text{diag}\{\sigma_1, \cdots \sigma_n\}$, i.e., the diagonal matrix of equivalent droop constants
T_s :	Sample time for frequency control
$\theta_i^{(q)}$:	Parameter vector of the q-layers of LSTM
$u(k)$:	$[u_1(k), \cdots, u_n(k)]^\top$
u_i :	Frequency control variable of prosumer i
$u_i(k)$:	Control action of prosumer i at time instant k
$\widehat{u}_i^{(q)}(k)$:	Estimation of control action of prosumer i at time instant k using q-layers of LSTM
$\widehat{u}_i(k)$:	Estimation of control action of prosumer i at time instant k using the ARMA model
$u^*(k)$:	Optimal control action
V_i :	Voltage at prosumer i

Masoud H. Nazari* (Corresponding author), Farinaz Nezampasandarbabi, and Le Yi Wang are with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA. <masoud.nazari@wayne.edu, farinaz.ne@wayne.edu, lywang@wayne.edu>

Siyu Xie is with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China. <syxie@uestc.edu.cn>

This research was supported in part by the Junior Faculty Start-up Fund index number 176822.

I. INTRODUCTION

Modern power systems are fundamentally networked systems, involving DERs, prosumer agents, and advanced sensing and communication technologies [1]–[3]. Traditionally, power system operating tasks, such as frequency control, are implemented in a centralized architecture. While the centralized architecture works well for the traditional power system, it may not be scalable to massive integration of DERs and prosumers in MPS due to: 1) single point of cyber-physical failure, 2) slow response to stochastic variations caused by emerging renewable energy resources, 3) lack of flexibility for prosumers integration, 4) low privacy for controlling non-utility owned assets, such as EVs and residential appliances, 5) dependence of dedicated communication links that are not appropriate to networks with large numbers of users or system

components [4]. In this regard, the idea of moving towards a distributed architecture is proposed in [5], [6] and has been extended recently in [7]–[13].

DOFC is an important problem in the MPS operation, which involves bringing frequency deviations from 60 Hz or 50 Hz to zero while optimizing power sharing among agents [14]–[21]. Although frequency is a global variable, optimizing power sharing among prosumers to restore frequency requires coordination [14]. Two general approaches have been offered for DOFC: 1) Distributed Averaging, and 2) Primal-Dual Methods. The former scheme is based on distributed averaging integral control, which equalizes the marginal prices to ensure correct steady-state frequency and a fair sharing of power generation [22]–[28]. This method requires communication between the local controllers. Note that fully decentralized integral control is not robustness to measurement bias and clock drifts [29]–[32]. Furthermore, the decentralized integral controllers cannot generally obtain an efficient allocation of generation resources.

Primal-dual optimization methods have been emerged as an alternative to integral-control strategies. These optimization methods can be implemented online as frequency controllers and can directly solve the optimal generator dispatch problem. They are typically based on primal-dual gradient methods [33], [34] and seek the saddle points of the Lagrangian function of the underlying optimization problem.

To our best of knowledge, the aforementioned works require perfect communication among controllers to reach consensus. This assumption is strong and cannot always hold. For instance, in [31] it has been demonstrated that the distributed schemes can be sensitive to faults and misbehavior of agents. In other words, communication failures, caused by disconnection of communication links, cyber-attacks, delay and other channel imperfections, can jeopardize the performance of distributed algorithms and lead to system-level reliability and efficiency problems [35]–[40]. The state-of-the-art asynchronous distributed optimization methods are only valid under bounded delay conditions [41]–[44]. If delay exceeds the pre-determined bound, the asynchronous algorithms cannot converge to the optimal solution.

In this paper, we address this challenge by leveraging deep learning based on LSTM to estimate the states of other prosumers and predict their future actions when delays exceed the limit or the communication links between agents are lost. An LSTM unit is composed of a cell, an input gate, an output gate and a forget gate (see Fig. 1). The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. By stacking multiple LSTM units, we can form a multilayer deep learning model. LSTM uses nonlinear functions to capture the nonlinear behavior of the system.

The LSTM-based estimation and prediction uses the historical time series data and can be implemented on DOFC to ensure efficiency and reliability of the frequency control under communication failures. The LSTM-based DOFC can bypass convergence in the cyber network during communication failures and can be directly implemented on the physical grid. In our earlier work, we have used the linearized model of power flows and applied the ARMA model to estimate missing data during communication failures [45]. The ARMA model uses a polynomial approximation which forms a linear model for estimation and prediction. If the linearized power flow equations are valid, the ARMA model can be an effective approach for modeling and prediction. Note that the linearized power flow model is valid when power deviations are small. In this model, $\Delta\delta_i$ is assumed to be smaller than 10 degree and V_i is assumed to be close to 1 p.u. However, when power deviations are large, the linearized model may obtain infeasible solutions.

In this paper, we extend our previous works and design the control strategy based on nonlinear power flows. We will show that due to system nonlinearities the LSTM algorithm has superior performance for state estimation, prediction, and algorithm convergence compared with the ARMA model and other linearized models. The main contributions of this paper are:

1) Developing the distributed LSTM-based state estimation and prediction algorithm for prosumer-based MPS by considering the nonlinearities of power flow equations.
2) Embedding the LSTM model into the DOFC algorithm, and laying a foundation for learning-based DOFC to mitigate the impact of communication failures.
3) Quantitatively characterizing the fundamental relationship between the estimation errors and convergence rates of the LSTM-based DOFC algorithm to develop a practical criterion for securing reliability of optimal frequency control under communication uncertainties.
4) Illustrating that the LSTM-based DOFC algorithm under communication failures asymptotically approaches the same convergence rate of the normal condition.

The rest of the paper is organized as follows. Section II develops the nonlinear model of DOFC considering nonlinear frequency ω/real-power relationships. In Section III the dynamic system estimation method based on the LSTM model is proposed to estimate missing data during communication failures. In Section, IV, the convergence analysis of the LSTM-based DOFC algorithm is conducted and the main results including the error bounds, strong convergence, and asymptotic optimality errors are derived. Simulation studies on two real-world power systems are carried out in Sections V. The paper concludes with discussions of the overall findings and future directions in Section VI.

II. Distributed Optimal Frequency Regulation Formulation

We first formulate the general structure of the DOFC problem for systems with n prosumers, where each prosumer represents a balancing area [46]:

$$\min_{u_1,\cdots u_n} \quad \frac{1}{2}\sum_{i=1}^{n} \Delta\omega_i(k)^\top M_i \Delta\omega_i(k) + u_i(k)^\top N_i u_i(k) \tag{1a}$$

$$\text{s.t.} \quad \Delta\omega_i(k) = \gamma_i u_i(k) - \sigma_i \Delta P_i(k) \tag{1b}$$

$$\Delta\delta_i(k+1) = \Delta\omega_i(k) + T_s \Delta\delta_i(k) \tag{1c}$$

$$\Delta P_i(k) = \sum_{j\in\mathcal{N}_i} V_i(k)V_j(k) B_{ij}(\sin(\delta_{ij}(k)+\Delta\delta_{ij}(k)) - \sin(\delta_{ij}(k))) \tag{1d}$$

The objective function of (1a) is to minimize quasi-steady-state frequency deviations at the prosumer-level while minimizing the system-wide control effort at time interval k. M_i and N_i are two positive constants. Constraint (1b) is the three-way droop equation which relates the variations of angular frequency at the prosumer level to the frequency control and power deviations. Constraint (1c) represents the relationship between voltage angle and frequency given the sample time T_s. The discrete time model is used as is appropriate for the quasi-static models and is standard in the frequency control literature [47]. Constraint (1d) illustrates the nonlinear power flow equations.

It is well known that power systems are inherently nonlinear. For example, in frequency regulation problems, the power flow equations are nonlinear functions:

$$\Delta P_i(k) = \sum_{j \in \mathcal{N}_i} g_{ij}(k) c_{ij}(k) \Delta \delta_{ij}(k), \tag{2}$$

where

$$\begin{aligned} c_{ij}(k) &= V_i(k) V_j(k) B_{ij}, \\ g_{ij}(k) &= \frac{\sin(\delta_{ij}(k) + \Delta\delta_{ij}(k)) - \sin(\delta_{ij}(k))}{\Delta \delta_{ij}(k)}. \end{aligned} \tag{3}$$

Furthermore, due to inaccurate or unknown physical model of transmission lines including transformers, compensation devices, protection systems, and unmodeled inductance and capacitance, the true power systems models are nonlinear and their parameters cannot be accurately and easily derived from line distance only. In this paper, we employ the LSTM model to represent the unknown physical system. The model parameters in LSTM will be learned from either historical or real-time operational data. Since the LSTM models are unstructured, they are flexible in representing real systems.

Generally, the LSTM models entail a parameterized nonlinear mapping. In our applications, the control action of neighboring prosumers is estimated through nonlinear q-th order dynamic models as follows:

$$\begin{aligned} \widehat{u}_i^{(q)}(k) = \mathcal{F}^{(q)}(&u_i(k-1), \ldots, u_i(k-q), \\ &P_i(k), \ldots, P_i(k-q+1), \theta_i^{(q)}), \end{aligned} \tag{4}$$

The parameter vector $\theta_i^{(q)}$ must be learned from operational data. For notational simplicity, we suppress the dependence of the following models and optimization problems on the unknown parameter vector $\theta_i^{(q)}$. The learning algorithms for $\theta_i^{(q)}$ and the convergence properties are given in Section IV.

Power systems contain a physical network for power flow interaction and a cyber network for information exchange. In this paper, we assume that the cyber networks and the prosumer-based power networks are the same. The set of nonlinear power flow equations has the following network analogy [48], [49]:

$$\Delta P(k) = H_p(\delta(k)) \Delta \delta(k), \tag{5}$$

$H_p(\delta(k))$ is the network matrix and is defined as follows:

$$H_{p_{ij}}(\delta(k)) = \begin{cases} g_{ij}(k) c_{ij}(k) & \text{if } i = j \\ -\sum\limits_{j \in \mathcal{N}_i, i \neq j} c_{ij}(k) g_{ij}(k) & \text{if } i \neq j \\ 0 & \text{Otherwise.} \end{cases} \tag{6}$$

Combining the constraints in (1) with (5) and assuming that $H_p(\delta(k+1)) \approx H_p(\delta(k))$, the following analogy to (1) is obtained:

$$\begin{aligned} \Delta P(k+1) =& H_p(\delta(k+1)) \Delta \delta(k+1) \\ =& H_p(\delta(k)) \Delta \delta(k) + H_p(\delta(k)) T_s \Delta \omega(k) \\ =& \Delta P(k) + H_p(\delta(k)) T_s \Gamma u(k) \\ & - H_p(\delta(k)) T_s \Sigma \Delta P(k) \\ =& (I - T_s H_p(\delta(k)) \Sigma) \Delta P(k) \\ & + T_s H_p(\delta(k)) \Gamma u(k), \end{aligned} \tag{7}$$

where I is the identity matrix, $\Sigma = \text{diag}\{\sigma_1, \cdots \sigma_n\}$ is the diagonal matrix of equivalent droop constants of prosumers, and $\Gamma = \text{diag}\{\gamma_1, \cdots \gamma_n\}$ is the matrix of control parameters. Note that the elements of the network matrix are dominated by the susceptance matrix $[B_{ij}]$. Thus, we can assume that in the small-time interval between k and $k+1$, the topology of the power grid is fixed and the elements of the network matrix will not change.

Under the quasi-steady-state frequency deviations, it is common that the DOFC problem and its solutions are simplified by approximation around rated frequency and nominal power flows. Following this approach, by removing Δ for simplicity, the quasi-steady-state dynamic model of prosumer-based power network can be approximated by:

$$P(k+1) = \mathcal{A}(\delta(k)) P(k) + \mathcal{B}(\delta(k)) u(k), \tag{8}$$

where

$$\mathcal{A}(\delta(k)) = I - T_s H_p(\delta(k)) \Sigma, \tag{9a}$$

$$\mathcal{B}(\delta(k)) = T_s H_p(\delta(k)) \Gamma. \tag{9b}$$

This simplification leads to the following DOFC problem which contains a cost coefficient for its power deviation and that for its control action:

$$\begin{aligned} \min_{u(k)} \mathcal{J}(u(k)) =& \min_{u(k)} \frac{1}{2} [P(k+1)^\top Q P(k+1) + u^\top(k) R u(k)] \\ =& \min_{u(k)} \frac{1}{2} [(\mathcal{A}(\delta(k)) P(k) + \mathcal{B}(\delta(k)) u(k))^\top Q \\ & \cdot (\mathcal{A}(\delta(k)) P(k) + \mathcal{B}(\delta(k)) u(k)) \\ & + u^\top(k) R u(k)], \end{aligned} \tag{10}$$

where Q and R are diagonal and positive definite cost matrices. The matrices $\mathcal{A}(\delta_k) = [\mathcal{A}_{ij}(\delta_k)]$ and $\mathcal{B}(\delta_k) = [\mathcal{B}_{ij}(\delta_k)]$ are related to the network topology, i.e., they are sparse. To

find a distributed solution, we denote that for prosumer i the following performance index can be defined:

$$\begin{aligned}\mathcal{J}_i(u(k)) &= \frac{1}{2}\Bigg[R_i u_i^2(k) + Q_i\Bigg(\sum_{j=1}^{n}\mathcal{A}_{ij}(\delta(k))P_j(k) \\ &\qquad + \sum_{j=1}^{n}\mathcal{B}_{ij}(\delta(k))u_j(k)\Bigg)^2\Bigg] \\ &= \frac{1}{2}\Bigg[R_i u_i^2(k) + Q_i\Bigg(\sum_{j\in\mathcal{N}_i\cup\{i\}}\mathcal{A}_{ij}(\delta(k))P_j(k) \\ &\qquad + \sum_{j\in\mathcal{N}_i\cup\{i\}}\mathcal{B}_{ij}(\delta(k))u_j(k)\Bigg)^2\Bigg]. \qquad (11)\end{aligned}$$

Then, we have $\mathcal{J}(u(k)) = \sum_{i=1}^{n}\mathcal{J}_i(u(k))$, and the optimal solution

$$u^*(k) = -G^{-1}(\delta(k))\mathcal{B}^\top(\delta(k))Q\mathcal{A}(\delta(k))P(k), \qquad (12)$$

where $G(\delta(k)) = R + \mathcal{B}^\top(\delta(k))Q\mathcal{B}(\delta(k))$.

We should emphasize that this simplified optimization represents a valid approximation due to smoothness of the model structures (they are continuously differentiable) and robustness of the problem that can be verified by the properties of solutions to the Lyapunov and Riccati equations. As a result, in the convergence analysis for optimization algorithms, this simplified model is employed. However, the LSTM models are nonlinear and their learning algorithms have the advantage of flexibility in representing different nonlinear functions and accurate estimation of their parameters. Consequently, they perform better than linear models, such as the ARMA model in our recent work [45]. All our case studies employ true nonlinear models of power systems.

A. Distributed Prediction Framework

The optimal solution in (12) is not feasible in distributed strategies since we have to calculate the inverse of the matrix $G(\delta(k))$. Next, we develop a distributed algorithm to search for the optimal solution. Note that

$$\nabla_u\mathcal{J}(u) = \begin{pmatrix}\nabla_{u_1}\mathcal{J}(u)\\ \nabla_{u_2}\mathcal{J}(u)\\ \vdots\\ \nabla_{u_n}\mathcal{J}(u)\end{pmatrix} \in \mathbb{R}^n, \qquad (13)$$

where

$$\nabla_{u_i}\mathcal{J}(u) = \sum_{\ell\in\mathcal{N}_i\cup\{i\}}\nabla_{u_i}\mathcal{J}_\ell(u). \qquad (14)$$

At each time instant k, prosumer i can adopt the following gradient algorithm to track the optimal solution $u_i^*(k)$ at iteration h:

$$\begin{aligned}u_i^{h+1}(k) &= u_i^h(k) - \mu^h\nabla_{u_i}\mathcal{J}(u^h(k)) \\ &= u_i^h(k) - \mu^h\sum_{\ell\in\mathcal{N}_i\cup\{i\}}\nabla_{u_i}\mathcal{J}_\ell(u^h(k)), \qquad (15)\end{aligned}$$

where the step size $\mu^h = 1/h^\gamma$, $1/2 < \gamma < 1$, and from (11) we know that

$$\begin{aligned}&\nabla_{u_i}\mathcal{J}_\ell(u^h(k)) \\ &= \begin{cases} R_\ell u_\ell^h(k) + Q_\ell\mathcal{B}_{\ell\ell}(\delta(k))\Bigg(\sum\limits_{j\in\mathcal{N}_\ell\cup\{\ell\}}\mathcal{A}_{\ell j}(\delta(k))P_j(k) \\ \qquad + \sum\limits_{j\in\mathcal{N}_\ell\cup\{\ell\}}\mathcal{B}_{\ell j}(\delta(k))u_j^h(k)\Bigg) & \text{if } \ell = i, \\ Q_\ell\mathcal{B}_{\ell i}(\delta(k))\Bigg(\sum\limits_{j\in\mathcal{N}_\ell\cup\{\ell\}}\mathcal{A}_{\ell j}(\delta(k))P_j(k) \\ \qquad + \sum\limits_{j\in\mathcal{N}_\ell\cup\{\ell\}}\mathcal{B}_{\ell j}(\delta(k))u_j^h(k)\Bigg) & \text{if } \ell \in \mathcal{N}_i. \end{cases}\end{aligned}$$

Thus, this algorithm is strictly distributed, since for prosumer i it only requires the gradient information from its neighbors, i.e., $\nabla_{u_i}\mathcal{J}_\ell(u^h(k))$ where $\ell \in \mathcal{N}_i \cup \{i\}$. Since $P_j(k)$ can be measured via tie-line flows and is the physical power imbalance, only $u_j^h(k)$ needs to be shared with neighbors.

However, during communication failures prosumer i may not be able to receive $u_j^h(k)$. The remedy to this problem is the main contribution of this paper. In the next section, we propose the LSTM method to estimate $u_j^h(k)$ based on the historical data $(u_j(k-1), u_j(k-2), ...)$.

III. Dynamic Model Estimation Methods

There are different types of dynamic models. We first use simple examples, and then extend them to general expressions. ARMA model is more effective when the power system represents a linear behavior. For nonlinear power flow models, deep machine learning methods are more effective. LSTM is the method used in this paper for dynamic model estimation. Note that all the parameters in LSTM must be learned from data. This is similar to identification.

A. LSTM Method and LSTM-based DOFC Algorithm

LSTM networks are a special kind of Recurrent Neural Network (RNN), capable of learning long-term dependencies. Remembering information for long periods of time is practically their default behavior. Also, LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. A gate in a neural network acts as a threshold for helping the network to distinguish when to use normal stacked layers versus an identity connection. An identity connection uses the output of lower layers as an addition to the output of consecutive layers. They are composed of a $sigmoid$ neural net layer and a pointwise multiplication operation. $Sigmoid$ is a smooth nonlinear activation function, which maintains the values between 0 and 1. This helps the network to update or forget the data.

In this paper, the gate functions are $sigmoid$ and $tanh$, tangent hyperbolic activation function. These are typical smooth nonlinear functions; consequently, the LSTM models are continuously differentiable.

Each prosumer is trained offline using LSTM. The historical power deviation data is different for each prosumer; therefore,

individual training is required. For prosumer i, LSTM first looks at the current data on $u_i(k-1)$ and $P_i(k)$:

$$F_i(k) = \sigma\big(W_i^{FP} P_i(k) + W_i^{FU} u_i(k-1) + b_i^F\big). \quad (16)$$

Next, it decides which new information needs to be stored in the cell. This has two parts. First part, a *sigmoid* layer called the "input gate" decides which values are updated. Second, a *tanh* layer creates a vector of new candidate values, $\hat{S}_i(k)$ that could be added to the state:

$$I_i(k) = \sigma\big(W_i^{IP} P_i(k) + W_i^{IU} u_i(k-1) + b_i^I\big), \quad (17)$$
$$G_i(k) = \phi\big(W_i^{GP} P_i(k) + W_i^{GU} u_i(k-1) + b_i^G\big), \quad (18)$$
$$O_i(k) = \sigma\big(W_i^{OP} P_i(k) + W_i^{OU} u_i(k-1) + b_i^O\big). \quad (19)$$

Next, the old cell state $S_i(k-1)$ is updated into the new cell state $S_i(k)$, and the old state is multiplied by $F_i(k)$:

$$S_i(k) = G_i(k) \odot I_i(k) + S_i(k-1) \odot F_i(k) \quad (20)$$

The last part is obtaining the output, which is the estimate of control action of neighboring prosumers.

$$\widehat{u}_i^{(1)}(k) = \phi_i(S_i(k)) \odot O_i(k). \quad (21)$$

The goal is to search for the parameters $W_i^{FP}, W_i^{FU}, b_i^F$, etc. for (11)-(14) by minimizing the difference between $u_i(k)$ and the output value $\widehat{u}_i^{(1)}(k)$. The structure of the LSTM block is shown in the following figure.

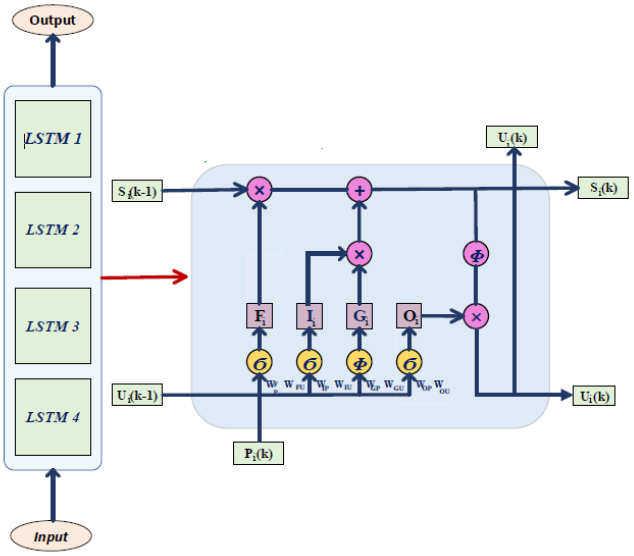


Fig. 1: Schematics of the LSTM algorithm with 4 layers.

The aforementioned equations are for one layer of the LSTM model. We can use the following notation to simplify the statement for the estimate of control action of neighboring prosumers.

$$\widehat{u}_i^{(1)}(k) = \mathcal{F}(u_i(k-1), P_i(k), \theta_i^{(1)}), \quad (22)$$

where $\mathcal{F}(\cdot)$ denotes the nonlinear function of the one layer LSTM model, $\theta_i^{(1)}$ denotes the unknown parameters of the one layer LSTM model which needs to be estimated based on existing data.

For recurring function, we can use this function and it is based on historical data. If needed more layers of LSTM can be added. The generalized format of q-layers LSTM model for prosumer i is provided in (4).

For the DOFC algorithm in (15), if prosumer i does not receive $u_j^h(k)$ at iteration h, an online estimate based on LSTM model will be used. Figure 2 shows the flowchart of the LSTM implementation on the DOFC algorithm.

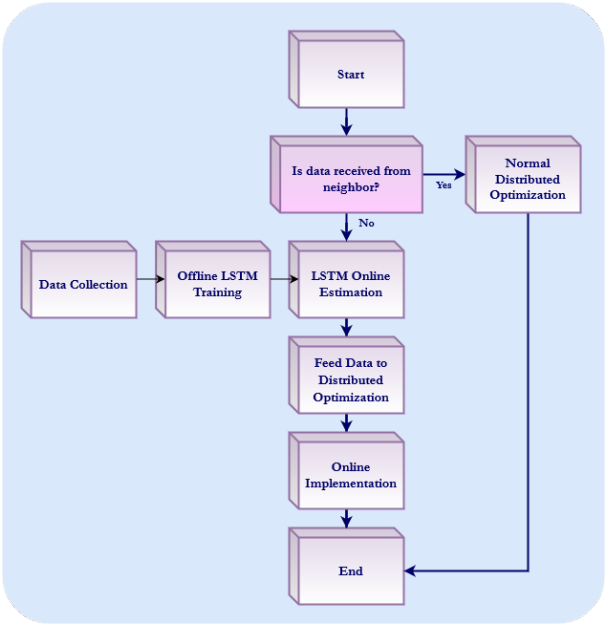


Fig. 2: Flowchart of LSTM integration with the online DOFC algorithm.

In our previous work [45], the below linear ARMA model is used to predict the missing data.

$$\widehat{u}_i(k) = \sum_{l=1}^{q} \phi^l u_i(k-l) + \sum_{l=1}^{q} \theta^l P_i(k-l+1), \quad (23)$$

However in practical situations, the true dynamic of $\widehat{u}_i(k)$ may be nonlinear. Thus, the prediction error from the ARMA model is quite large for nonlinear cases, see Section V for details. In order to make the prediction more accurate, we use the aforementioned q-layers LSTM nonlinear function (4).

IV. Convergence Analysis

A. Convergence of the parameter estimation

For (4), the unknown parameter $\theta_i^{(q)}$ in the LSTM model must be learned from data by using identification algorithms. There are many gradient-based algorithms and their modifications, such as Newton's method, Nesterov accelerated gradient, Adam, AdaGrad, AdaDelta, RMSProp, etc.

For prosumer i, to minimize the following error criteria

$$\begin{aligned} e(\theta_i^{(q)}, k) =& \|u_i(k) - \widehat{u}_i^{(q)}(k)\| \\ =& \|u_i(k) - \mathcal{F}^q(u_i(k-1), \ldots, u_i(k-q), \\ & P_i(k), \ldots, P_i(k-q+1), \theta_i^{(q)})\|^2, \end{aligned} \quad (24)$$

the gradient-based stochastic approximation algorithm is used to search for the parameter $\theta_i^{(q)}$:

$$\theta_i^{(q)}(k+1) = \theta_i^{(q)}(k) - \beta(k)[\nabla_{\theta_i^{(q)}} e(\theta_i^{(q)}(k), k) + d(k)], \quad (25)$$

where $\beta(k) > 0$ is the step size satisfying $\beta(k) \to 0$ as $k \to \infty$ and $\sum_k \beta(k) = \infty$. The term $d(k)$ is the observation/computation noise for the gradient which is a random variable such that $\mathbb{E}[d(k)] = 0, \mathbb{E}[d(k)d^\top(k)] = \Sigma_d$. Also, Σ_d is symmetric and positive definite.

Assumption 4.1: The error criteria (24) for the LSTM model is smooth, namely continuously differentiable to any order needed, and is convex with respect to $\theta_i^{(q)}$.

Remark 4.1: By using smooth gate functions such as $sigmoid$, LSTM are naturally smooth. Thus, *Assumption 4.1* includes convex functions. At present, convergence properties for optimal parameter estimation algorithms do require convex conditions for global convergence. As a result, these assumptions are not restrictive.

Under *Assumption 4.1*, the convergence properties of (25) can be analyzed by using the limit ordinary differential equation (ODE) method [50]. For prosumer i, denote $\tilde{\theta}_i^{(q)}(k) = \theta_i^{(q)}(k) - \theta_i^{(q)}$ as the optimization error, then we have

$$\tilde{\theta}_i^{(q)}(k+1) = \tilde{\theta}_i^{(q)}(k) - \beta(k)[\nabla_{\theta_i^{(q)}} e(\theta_i^{(q)}(k), k) + d(k)].$$

For simplicity, we omit the detailed proof of the following theorem and refer the reader to [50].

Theorem 4.1: For prosumer i, assume that $\nabla_{\theta_i^{(q)}} e(\theta_i^{(q)}, k)$ is positive definite and *Assumption 4.1* holds, then $\theta_i^{(q)}(k) \to \theta_i^{(q)}$ with probability one (w.p.1) as $k \to \infty$.

While the actual proof of *Theorem 4.1* will be skipped, the main ideas can be summarized as follows. The limit ODE is $\dot{\tilde{\theta}}_i^{(q)} = -\nabla_{\theta_i^{(q)}} e(\theta_i^{(q)}, k)$, and it has a unique equilibrium point, which is the optimal solution $\theta_i^{(q)}$. By *Assumption 4.1*, and using the Lyapunov method, the equilibrium point $\theta_i^{(q)}$ is asymptotically stable. Using the ODE method in stochastic approximation [50], the convergence result can be obtained.

B. Convergence of the optimization process

In this part, we prove that the gradient-based DOFC algorithm (15) based on the LSTM prediction can improve the convergence rate of the optimization process. Also, it has superior performance compared with the linearized prediction models, such as ARMA used in [45]. For simplicity, we omit k in the remaining part. Thus, the update algorithm with the prediction error and gradient error can be written in the following vector form:

$$u^{h+1} = u^h - \mu^h[G(\delta)u^h + \mathcal{B}^\top(\delta)Q\mathcal{A}(\delta)P + d_1^h + d_2^h], \quad (26)$$

where $d_1^h \in \mathbb{R}^n$ denotes the prediction error from the LSTM model at iteration h, and $d_2^h \in \mathbb{R}^n$ denotes the gradient error at iteration h. Next, we make the following basic assumption for the theoretical analysis.

Assumption 4.2: The prediction error $\{d_1^h\}$ and the gradient error $\{d_2^h\}$ are two mutually independent sequences of independent and identically distributed (i.i.d.) random variables such that $\mathbb{E}[d_1^h] = \mathbb{E}[d_2^h] = \mathbf{0}_{n\times 1}$, $\mathbb{E}[d_1^h(d_1^h)^\top] = \Sigma_{d_1}$, $\mathbb{E}[d_2^h(d_2^h)^\top] = \Sigma_{d_2}$, where $\Sigma_{d_1} \in \mathbb{R}^{n\times n}$ and $\Sigma_{d_2} \in \mathbb{R}^{n\times n}$ are symmetric positive definite.

Thus, we can obtain the strong convergence result in Theorem 4.2. Note that the prediction and gradient noises are assumed to be stochastic, and their statistical coefficients are known, including their means and variances. When the means are non-zero, the noises will introduce a bias on optimal solutions, which can be corrected in the following way: If we know the mean d, a simple subtraction from the observation data will lead to a modified observation error sequence whose mean is then zero. Such transformation techniques are widely used in control systems and related fields. Also, if the mean is unknown, estimation algorithms may be added, leading to adaptive optimization algorithms. The topic is beyond the scope of this paper and will not be further discussed here.

Theorem 4.2: Under *Assumption 4.2*, the iterates $\{u^h\}$ generated by (26) converge to the optimal solution $u^h \to u^*$ w.p.1 as $h \to \infty$.

For simplicity, we omit the detailed proof and refer the reader to [50, Chapters 5 and 6]. The main idea is summarized here. Using the ODE method in stochastic approximation [50], define $t^h = \sum_{j=0}^{h-1} \mu^h$, $\varpi(t) = \max\{h : t^h \le t\}$, the piecewise constant interpolation $u_0(t) = u^h$ for $t \in [t^h, t^{h+1})$, and the shift sequence $u_h(t) = u_0(t + t^h)$. If the step size $\mu^h = 1/h^\gamma$ $(1/2 < \gamma < 1)$, the interpolated sequence $\{u_h(\cdot)\}$ is uniformly bounded and equicontinuous. By [50], we can extract a subsequence $\{u_{h_\ell}(\cdot)\}$, which converges to $u(\cdot)$ on any compact intervals w.p.1 such that $u(\cdot)$ is a solution. The limit ODE is $\dot{u} = G(\delta)u + \mathcal{B}^\top(\delta)Q\mathcal{A}(\delta)P$, whose unique equilibrium point is precisely the optimal solution $u^* = -G^{-1}(\delta)\mathcal{B}^\top(\delta)Q\mathcal{A}(\delta)P$ which is an asymptotically stable point. This theoretical result leads to the desired property of *Theorem 4.2*.

Next, we demonstrate the convergence rate. Define $v^h = (u^h - u^*)/\sqrt{\mu^h}$, which takes a continuous-time interpolation as $v_0(t) = v^h$ for $t \in [t^h, t^{h+1})$, and define $v_h(t) = v_0(t + t^h)$. As in [50, Chapter 10], we can show that $v_h(\cdot)$ converges weakly to $v(\cdot)$ such that $v(\cdot)$ is a solution of an appropriate stochastic differential equation. The scaling factor $\sqrt{\mu_k}$ together with the asymptotic covariance gives the desired rate of convergence. The standard central limit theorem argument yields that $\frac{1}{\sqrt{h}} \sum_{j=\kappa}^{\kappa+h-1} (d_1^j + d_2^j)$ converges weakly to $N(0, \Sigma_{d_1} + \Sigma_{d_2})$. Here, we consider the following averaging form:

$$\bar{u}^h = \sum_{j=0}^{h-1} u^j / h.$$

Thus, the following results can be obtained:

Theorem 4.3: Under *Assumption 4.2*, $\sqrt{h}(\bar{u}^h - u^*)$ converges weakly to a normal random variable with zero mean and asymptotic covariance

$$\Sigma^* = G^{-1}(\delta)(\Sigma_{d_1} + \Sigma_{d_2})G^{-1}(\delta). \quad (27)$$

Remark 4.2: Here we omit the proof and refer interested readers to [50, Chapter 11]. Note that $\bar{u}^h - u^*$ is asymptotically normal (Gaussian distributed) with zero mean and covariance Σ^*/h. Also, we use the error covariance matrix Σ^* to evaluate

how fast convergence to the optimal solution can be achieved. It can be observed that by increasing the accuracy of the estimate $\widehat{u}_j^h$, the prediction error variance Σ_{d_1} decreases, thus the algorithm converges faster to the optimal solution. By [51], we know that if the packet delivery ratio or the prediction error is small enough, the gradient algorithm with prediction will converge to the optimal solution faster than the original gradient algorithm with packet loss.

The ARMA-based prediction error is large when the true dynamic is nonlinear. But the LSTM-based prediction algorithm can provide good results for nonlinear systems, see Section V. In this scenario, Σ_{d_1} is smaller for LSTM-based prediction than the ARMA model. This means that the convergence rate for LSTM is faster than that for ARMA.

V. Simulation Results

In this section, we first compare the accuracy of the LSTM-based prediction with the ARMA model for two real-world power systems. The first system is the power grid on Flores Island and the second system is the power grid on Sao Miguel Island. The PC used for all simulations and case studies has Intel(R) Core(TM) i7-7700HQ CPU @ 2.80 GHz with installed RAM of 16GB, and 64-bit operating system. Next, we illustrate the performance of the LSTM-based DOFC under different communication contingency scenarios.

A. LSTM-based Prediction Performance

1) Flores Island: Flores is one of the smaller islands of the Azores Archipelago. It has three small power plants supplying the island's electrical demand with average of 2 $MWhr$. The generators have three technologies: 1) diesel generators, 2) hydro plants, and 3) wind plants [52]. The electrical network of Flores is clustered into three prosumers, where each prosumer has a generator to regulate frequency. The topological structure of Flores' electrical network is shown in Fig. 3 and the prosumer-based electric power grid on Flores is shown in Fig. 4. The detailed electrical characteristics of Flores can be found in [52], [53].

In this case study, the historical power flow data of prosumers consisting of 4463 data points are used. This is one-month power flow data for January 2019, which is publicly available in [54]. 90% of the historical data is used for training and 10% is used for testing the LSTM model.

For better data fitting and preventing training divergence, we standardized the training data to have zero mean and unit variance. We then standardized the test data using the same training data parameters. The next step was creating an LSTM network and training the network for forecasting. The prediction horizon is one step ahead based on the previous q steps historical data. This is consistent with the q layers of the LSTM model. In other words, each layer uses the previous value as input to the function. The last part of the algorithm is to update the network state with the observed value, which predicts the next time step.

For this case study, the LSTM model has 4 layers and the epoch number is 250. Note that the epoch number is the number of iterations multiplied by the number of batches. The number of iterations relates to the convergence of the LSTM model to find the accurate prediction value. Also, batch size is the number of training data processed at a particular time. In this case study, the batch size is fixed and equals to one.

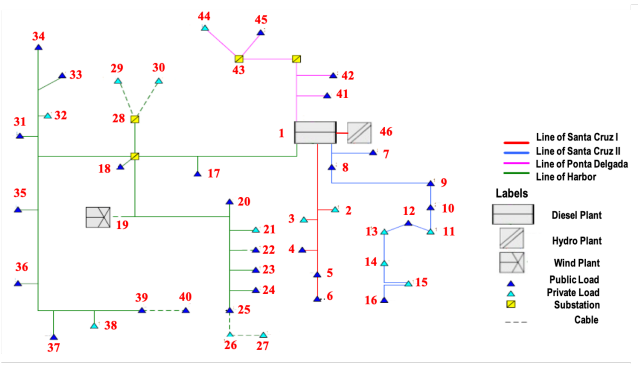


Fig. 3: The electrical network on Flores Island.

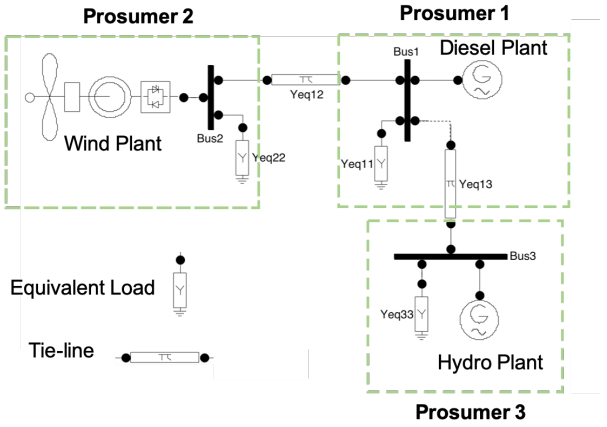


Fig. 4: The prosumer-based electrical network on Flores Island.

An epoch occurs when the full set of the training data is passed/forward propagated and then backpropagated through the neural network. The LSTM layers are mathematical functions that are stacked on top of each other so that the output of the first layer is the input to the second LSTM layer and so on. The root mean square error of the LSTM model is 0.44%. This implies that the prediction error is small. In other words, the smaller the root mean square error, the better is the power flow prediction. Also, having less error means the LSTM model is accurate and can be implemented on the DOFC algorithm. The results of the LSTM-based prediction is presented in Fig. 5.

Next, we apply the ARMA-based prediction method, which is based on the linearized model, to the same data set. Due to nonlinearities of the model the ARMA method has a significant percentage of error (8.05%) as shown in Fig. 6. More importantly, by increasing the nonlinearity of the model, the ARMA method becomes unstable and does not converge. Note that the ARMA method has an acceptable performance for linear models and has less computational complexity.

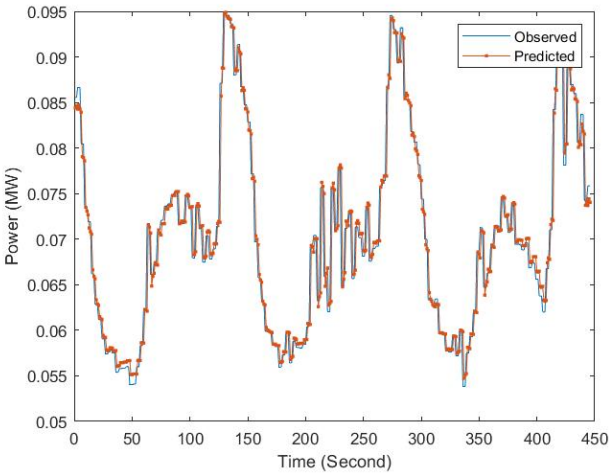


Fig. 5: LSTM-based prediction for Flores Island.

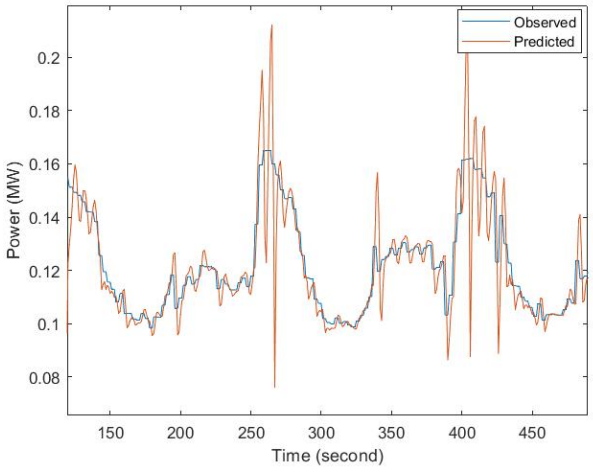


Fig. 6: ARMA-based prediction for Flores Island.

B. Sao Miguel Island

Sao Miguel Island is the capital of the Azores Archipelago. The electric power grid on the Sao Miguel Island consists of fifteen generators. These generators are geothermal, diesel, and hydro plants. We assume that the island is clustered into fifteen prosumers, where each prosumer is a balancing area for frequency control [52]. The network topology of the prosumer-based electric power grid on Sao Miguel is shown in Figure 7. The detailed electrical characteristics of Sao Miguel is provided in [52], [53]. Similar to the previous case study, historical power flow data for January 2019 is used for the training and testing of the LSTM model. The historical data includes 4463 data points [54]. 90% of the data is used for training and 10% is used for testing the LSTM model.

The LSTM method has 4 hidden layers and the epoch number is 250. All the training process took 51 seconds. The root mean square error is 1.55% which implies that the LSTM prediction is very accurate. Note that the training and testing process can be done offline. All weights and parameters will be tuned offline and will be updated periodically. Once the

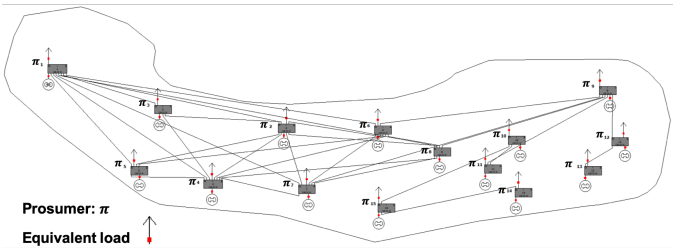


Fig. 7: The prosumer-based electrical network on Sao Miguel Island.

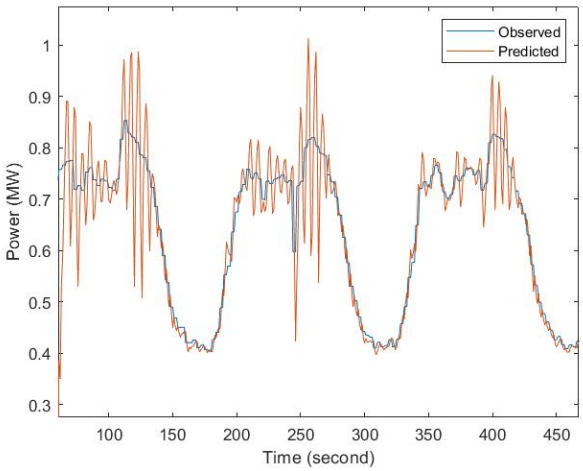


Fig. 8: ARMA for SMG Island time prediction

parameters are obtained, the LSTM model can be used for online estimation and prediction in the DOFC algorithm. The result of the LSTM-based prediction for the Sao Miguel Island is presented in Fig. 9.

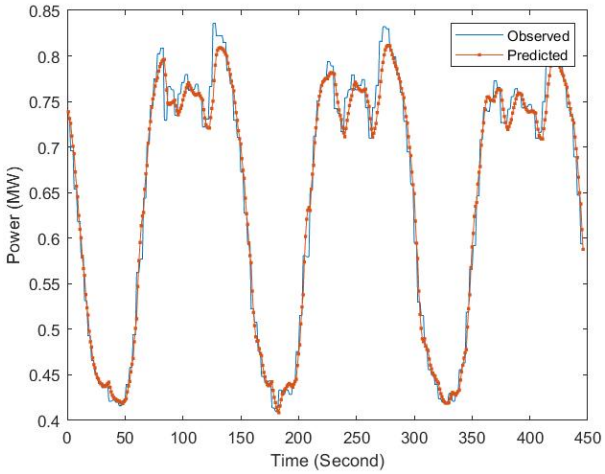


Fig. 9: LSTM for SMG Island time prediction

In order to increase the speed of the LSTM algorithm, we can reduce the number of epochs. For instance, when the number of epochs decreases to 100, the prediction results are still acceptable as shown in Fig. 10. In this case, the root mean

square error is 3.02% (two times larger than the case with 250 epochs). However, the computation complexity decreases. Fig 10 shows the training progress for 250 and 100 epochs for the same set of historical data.

Based on our simulation studies, 4 layer is the minimum number of layers that we could use for this case study. Reducing the number of layers to 3 or 2 can significantly impact the accuracy of the prediction. Depending on the system model and historical power used, the number of required LSTM layer could vary. Having more layers improves the prediction accuracy, but it increases the complexity of the LSTM model. As DOFC is an online algorithm and needs to be updated every few seconds, there can be a trade off between the number of LSTM layers and the speed of algorithm. This can be the topic of future research endeavor.

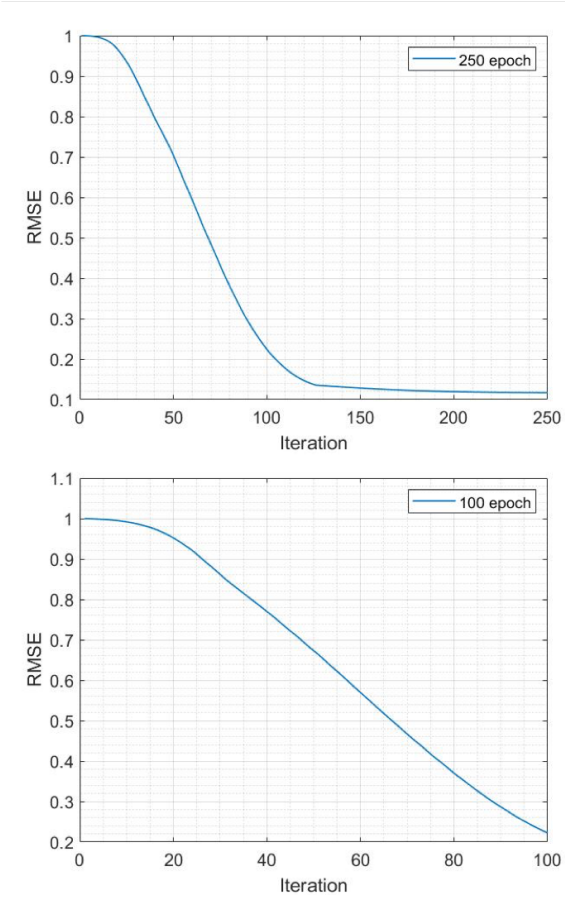


Fig. 10: Training progress for 100 vs 250 epoch

C. LSTM-based DOFC Simulation

1) Flores Island: We simulate the performance of DOFC after a power imbalance equal to 25 kW in the island. We study three communication/computation scenarios:

- Normal communication system, using model-based DOFC based on distributed ADMM optimization. This is the benchmark scenario.
- Complete communication disconnection at prosumer 3 (disconnection of the link between prosumer 1 and 3), using model-based DOFC based on ADMM. In this scenario, the missing data cannot be estimated, instead the previous received data is used.
- Communication disconnection, but using the LSTM model to predict the states of prosumers 1 and 3.

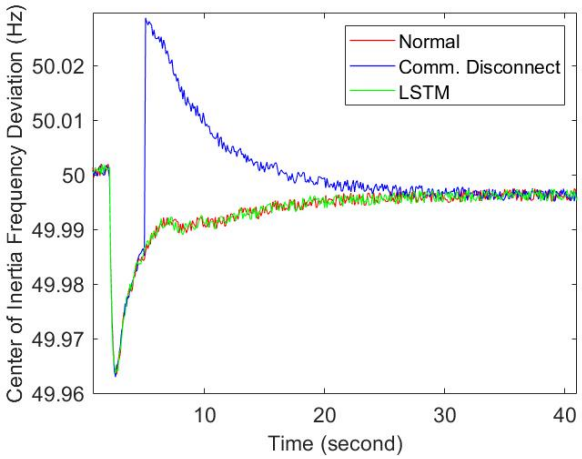


Fig. 11: Flores Island center of inertia

Fig. 11 shows the quasi-steady-state dynamics of the center of inertia for the three scenarios. As shown in the figure, the LSTM-based DOFC has a very similar performance compared with the normal condition. In other words, the LSTM model allows the disconnected prosumers to accurately predict the states of their neighbors and continue performing DOFC. However, the second scenario has a poor performance.

2) Sao Miguel Island: The historical power flow data of fifteen prosumers are used for the simulation study of the Sao Miguel Island. We simulate the quasi-steady-state dynamics of DOFC after a power imbalance of 0.69 MW. Similar to the previous case study, we investigate three communication/computation scenarios:

- Normal communication system, using model-based DOFC based on distributed ADMM optimization.
- Complete communication disconnection at prosumer 15. In this scenario model-based DOFC based on ADMM is used. Also, the previous received data is used when there is missing data.
- Complete communication disconnection at prosumer 15, but using the LSTM model to predict the states of neighboring prosumers.

Fig. 12 shows the quasi-steady-state dynamics of the center of inertia for the three scenarios. As shown in the figure, the LSTM-based DOFC has a very similar performance to the normal condition.

VI. Conclusions

This paper proposed a deep learning method-based LSTM to mitigate the impacts of communication failures on the performance of the DOFC algorithm. We considered the nonlinearities of the power flow equations in designing the control strategy. The proposed LSTM-based DOFC is novel because it can estimate the states of prosumers and predict their future control actions during communication failures,

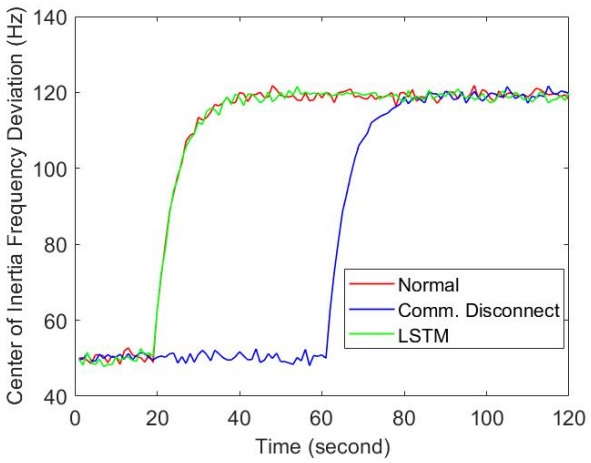


Fig. 12: SMG Island center of inertia

thus it can bypass convergence in the cyber network and is scalable to large-scale power systems. In other words, as the method only requires learning the behavior of electrically adjacent prosumers, when the power grid increases in size and scale, the learning and training will be still limited to neighbor-to-neighbor interaction.

We proved the convergence properties of the LSTM-based DOFC and showed that the algorithm has superior performance compared with the ARMA model and other linearized prediction methods. The theoretical results prove that the LSTM-based DOFC asymptotically reaches the same convergence rate of the normal condition (no communication contingencies). The simulation studies on two real-world power grids validate the theoretical findings.

The proposed LSTM-based prediction method can be implemented on distributed economic dispatch and unit commitment algorithms to increase their resilience against communication failures and improve their convergence rate for large-scale systems. This is the topic of future research endeavors.

REFERENCES

[1] A. G. Madureira and J. P. Lopes, "Coordinated voltage support in distribution networks with distributed generation and microgrids," *IET renewable Power generation*, vol. 3, no. 4, pp. 439–454, 2009.

[2] J. M. Guerrero, J. Matas, L. Garcia de Vicuna, M. Castilla, and J. Miret, "Decentralized control for parallel operation of distributed generation inverters using resistive output impedance," *IEEE Transactions on Industrial Electronics*, vol. 54, no. 2, pp. 994–1004, 2007.

[3] K. Young, C. Wang, L. Wang, and K. Strunz, *Electric Vehicle Battery Technologies, Electric Vehicle Integration into Modern Power Networks*. Edited by Garcia-Valle, Rodrigo and Lopes, João A Peças.

[4] Z. Wang, F. Liu, J. Z. F. Pang, S. H. Low, and S. Mei, "Distributed optimal frequency control considering a nonlinear network-preserving model," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 76–86, 2019.

[5] S. Grijalva, M. Costley, and N. Ainsworth, "Prosumer-based control architecture for the future electricity grid," in *Control Applications (CCA), 2011 IEEE International Conference on*, pp. 43–48, IEEE, 2011.

[6] T. Ramachandran, Z. Costello, P. Kingston, S. Grijalva, and M. Egerstedt, "Distributed power allocation in prosumer networks," in *IFAC Necsys*, 2012.

[7] X. He, J. Yu, T. Huang, and C. Li, "Distributed power management for dynamic economic dispatch in the multimicrogrids environment," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 4, pp. 1651–1658, 2019.

[8] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2192–2203, 2018.

[9] Q. Li, D. W. Gao, H. Zhang, Z. Wu, and F.-y. Wang, "Consensus-based distributed economic dispatch control method in power systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 941–954, 2019.

[10] Q. Lü, X. Liao, H. Li, and T. Huang, "Achieving acceleration for distributed economic dispatch in smart grids over directed networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1988–1999, 2020.

[11] M. Hosseinzadeh, L. Schenato, and E. Garone, "A distributed optimal power management system for microgrids with plug&play capabilities," *Advanced Control for Applications: Engineering and Industrial Systems*, vol. 3, no. 1, p. e65, 2021.

[12] C. Zhao, J. He, P. Cheng, and J. Chen, "Analysis of consensus-based distributed economic dispatch under stealthy attacks," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 6, pp. 5107–5117, 2017.

[13] L.-Y. Lu, H. J. Liu, H. Zhu, and C.-C. Chu, "Intrusion detection in distributed frequency control of isolated microgrids," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6502–6515, 2019.

[14] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Transactions on Smart Grid*, vol. 8, pp. 2941–2962, Nov 2017.

[15] N. Ainsworth and S. Grijalva, "Design and quasi-equilibrium analysis of a distributed frequency-restoration controller for inverter-based microgrids," in *2013 North American Power Symposium (NAPS)*, pp. 1–6, Sep. 2013.

[16] A. Jokić, M. Lazar, and P. P. van den Bosch, "Real-time control of power systems using nodal prices," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 9, pp. 522–530, 2009.

[17] R. Mudumbai and S. Dasgupta, "Distributed control for the smart grid: The case of economic dispatch," in *2014 Information Theory and Applications Workshop (ITA)*, pp. 1–6, Feb 2014.

[18] A. D. Dominguez-Garcia and C. N. Hadjicostis, "Coordination and control of distributed energy resources for provision of ancillary services," in *2010 First IEEE International Conference on Smart Grid Communications*, pp. 537–542, Oct 2010.

[19] A. Kiani Bejestani, A. Annaswamy, and T. Samad, "A hierarchical transactive control architecture for renewables integration in smart grids: Analytical modeling and stability," *IEEE Transactions on Smart Grid*, vol. 5, pp. 2054–2065, July 2014.

[20] D. Cai, E. Mallada, and A. Wierman, "Distributed optimization decomposition for joint economic dispatch and frequency regulation," *IEEE Transactions on Power Systems*, vol. 32, pp. 4370–4385, Nov 2017.

[21] M. Ilic, L. Xie, and Q. Liu, *Engineering IT-Enabled Sustainable Electricity Services*. Springer, 2013.

[22] C. Zhao, E. Mallada, and F. Dörfler, "Distributed frequency control for stability and economic dispatch in power networks," in *2015 American Control Conference (ACC)*, pp. 2359–2364, July 2015.

[23] J. W. Simpson-Porco, F. Dorfler, and F. Bullo, "Synchronization and power sharing for droop-controlled inverters in islanded microgrids," *Automatica*, vol. 49, no. 9, pp. 2603 – 2611, 2013.

[24] L. Lu, "Consensus-based p-f and q-v droop control for multiple parallel-connected inverters in lossy networks," in *2013 IEEE International Symposium on Industrial Electronics*, pp. 1–6, May 2013.

[25] D. Burbano and M. di Bernardo, "Consensus and synchronization of complex networks via proportional-integral coupling," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1796–1799, June 2014.

[26] M. Andreasson, D. V. Dimarogonas, K. H. Johansson, and H. Sandberg, "Distributed vs. centralized power systems frequency control," in *2013 European Control Conference (ECC)*, pp. 3524–3529, July 2013.

[27] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Breaking the hierarchy: Distributed control and economic optimality in microgrids," *IEEE Transactions on Control of Network Systems*, vol. 3, pp. 241–253, Sep. 2016.

[28] J. W. Simpson-Porco, Q. Shafiee, F. Dörfler, J. C. Vasquez, J. M. Guerrero, and F. Bullo, "Secondary frequency and voltage control of islanded microgrids via distributed averaging," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 7025–7038, Nov 2015.

[29] M. Andreasson, D. V. Dimarogonas, H. Sandberg, and K. H. Johansson, "Distributed control of networked dynamical systems: Static feedback, integral action and consensus," *IEEE Transactions on Automatic Control*, vol. 59, pp. 1750–1764, July 2014.

[30] M. Andreasson, D. V. Dimarogonas, H. Sandberg, and K. H. Johansson, "Distributed pi-control with applications to power systems frequency control," in *2014 American Control Conference*, pp. 3183–3188, June 2014.

[31] F. Dorfler and S. Grammatico, "Gather-and-broadcast frequency control in power systems," *Automatica*, vol. 79, pp. 296 – 305, 2017.

[32] J. Schiffer, C. A. Hans, T. Kral, R. Ortega, and J. Raisch, "Modeling, analysis, and experimental validation of clock drift effects in low-inertia power systems," *IEEE Transactions on Industrial Electronics*, vol. 64, pp. 5942–5951, July 2017.

[33] T. Kose, "Solutions of saddle value problems by differential equations," *Econometrica, Journal of the Econometric Society*, pp. 59–70, 1956.

[34] G. W. Brown and J. Von Neumann, "Solutions of games by differential equations," *Contributions to the Theory of Games*, vol. 1, no. 73-79, p. 43, 1950.

[35] C. Lenzen and R. Wattenhofer, "Distributed algorithms for sensor networks," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1958, pp. 11–26, 2012.

[36] T. Ramachandran, M. H. Nazari, S. Grijalva, and M. Egerstedt, "Overcoming communication delays in distributed frequency regulation," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2965–2973, 2016.

[37] M. H. Nazari, L. Y. Wang, S. Grijalva, and M. Egerstedt, "Communication-failure-resilient distributed frequency control in smart grids: Part i: Architecture and distributed algorithms," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1317–1326, 2020.

[38] M. H. Nazari, L. Y. Wang, S. Grijalva, and M. Egerstedt, "Communication-failure-resilient distributed frequency control in smart grids: Part ii: Algorithmic implementation and system simulations," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3192–3202, 2020.

[39] M. H. Nazari, S. Grijalva, and M. Egerstedt, "Distributed frequency regulation in the presence of communication failure," in *2016 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2016.

[40] M. H. Nazari, S. Xie, and L. Y. Wang, "Mimo architecture for fast convergence of distributed online optimization in smart grids," *International Journal of Electrical Power & Energy Systems*, vol. 142, p. 108206, 2022.

[41] T.-H. Chang, W.-C. Liao, M. Hong, and X. Wang, "Asynchronous distributed admm for large-scale optimization-part ii: Linear convergence analysis and numerical performance," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3131–3144, 2016.

[42] S. Magnússon, G. Qu, and N. Li, "Distributed optimal voltage control with asynchronous and delayed communication," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3469–3482, 2020.

[43] Q. Zhang, Y. Guo, Z. Wang, and F. Bu, "Distributed optimal conservation voltage reduction in integrated primary-secondary distribution systems," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 3889–3900, 2021.

[44] G. Chen and Z. Zhao, "Delay effects on consensus-based distributed economic dispatch algorithm in microgrid," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 602–612, 2018.

[45] M. Nazari, S. Xie, F. Nezampasandarbabi, and L. Y. Wang, "Prediction-based distributed optimal frequency control for resilience against communication failures," *in International Journal of Electrical Power & Energy Systems*, under review, 2022.

[46] M. H. Nazari, Z. Costello, M. Feizollahi, S. Grijalva, and M. Egerstedt, "Distributed frequency control of prosumer-based electric energy systems," *IEEE Transactions on Power Systems*, vol. 29, pp. 2934–2942, Nov 2014.

[47] N. Popli and M. Ilic, *Modeling and Control Framework to Ensure Intra-dispatch Regulation Reserves, in Chapter 14 of Engineering IT-Enabled Sustainable Electricity Services*. Springer, 2013.

[48] M. Ilic and J. Zaborszky, *Chapter 5 of Dynamics and Control of Large Electric Power Systems*. Wiley-IEEE Press, 2000.

[49] M. Ilic, "Network theoretic conditions for existence and uniqueness of steady state solutions to electric power circuits," in *1992 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 6, pp. 2821–2828 vol.6, 1992.

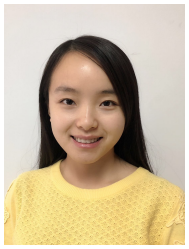
[50] H. Kushner and G. Yin, *Stochastic approximation and recursive algorithms and applications*. 2nd ed. New York, NY, USA: Springer-Verlag, 2003.

[51] M. H. Nazari, S. Xie, L. Yi Wang, G. Yin, and W. Chen, "Impact of communication packet delivery ratio on reliability of optimal load tracking and allocation in dc microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 2812–2821, 2021.

[52] M. H. Nazari, *Electrical Networks of the Azores Archipelago, in Chapter 3 of Engineering IT-Enabled Sustainable Electricity Services*. Springer, 2013.

[53] M. H. Nazari, *Small-Signal Stability Analysis of Electric Power Systems on the Azores Archipelago, in Chapter 17 of Engineering IT-Enabled Sustainable Electricity Services*. Springer, 2013.

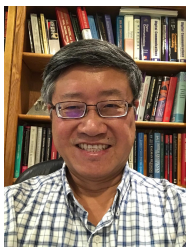
[54] Historical data of Flores Island and Sao Miguel Island. https://github.com/FarinazNe/historical-data.



Siyu Xie received the B.S. degree in information and computing science (systems control) from Beijing University of Aeronautics and Astronautics in 2013, and Ph.D. degree in control theory from Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2018. She was a postdoctoral fellow from 2019 to 2022 at the Department of Electrical and Computer Engineering, Wayne State University, USA. She is currently an Associate Professor in the School of Aeronautics and Astronautics with the University of Electronic Science and Technology of China. She received the IEEE CSS Beijing Chapter Young Author Prize in 2015, China National Scholarship for graduate students in 2017, the President Scholarship of Chinese Academy of Sciences in 2018, and Excellent Doctoral dissertation of Chinese Academy of Sciences in 2019. Her research interests include networked systems, distributed optimization problems for power systems, distributed adaptive filters, machine learning, identification, and distributed control.

Masoud H. Nazari received the MSc degree in Electrical Engineering (Energy Systems) from Sharif University of Technology in 2005, the MSc degree in Engineering & Public Policy from Carnegie Mellon University, Pittsburgh, PA, USA, in 2010 and dual Ph.D. degree in Electrical & Computer Engineering, and Engineering & Public Policy from Carnegie Mellon University in 2012. He was a Postdoctoral Fellow from 2013 to 2015 in the Electrical & Computer Engineering Department at the Georgia Institute of Technology. He was an Assistant Professor of Electrical Engineering with California State University, Long Beach, USA. He is currently an Assistant Professor of Electrical & Computer Engineering with Wayne State University, Detroit, MI, USA. He is the Chair of IEEE Smart Buildings-Loads-Customer Systems (SBLC) Architecture Subcommittee, Editorial Board of IEEE Technology Conferences, and Global Learning Faculty Fellow at WSU. He has been the Primary Investigator of several research projects, such as the $2.5 million California Energy Commission Project to develop an innovative building energy management system. Dr. Nazari is the IEEE Senior Member in the Power and Energy Society (PES) and was the recipient of the Best Paper Award in the 2017 North American Power Symposium.

Farinaz Nezampasandarbabi received the BSc degree in Electrical Engineering from Wayne State University in 2020. She is currently a PhD student in the Electrical & Computer Engineering Department at Wayne State University. Her research interest is in smart grid optimization based on machine learning methods. She received the Three-Minute Thesis Competition Award in the College of Engineering at Wayne State University in 2020, and Research Opportunities for Engineering Undergraduates (ROEU) Awards at Wayne State University in 2019 and 2018.



Le Yi Wang (S'85-M'89-SM'01-F'12) received the Ph.D. degree in electrical engineering from McGill University, Montreal, Canada, in 1990. Since 1990, he has been with Wayne State University, Detroit, Michigan, where he is currently a Professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of complexity and information, system identification, robust control, information processing and learning, as well as medical, automotive, communications, power systems, and computer applications of control methodologies. He was a plenary speaker in many international conferences. He serves on the IFAC Technical Committee on Modeling, Identification and Signal Processing. He was an Associate Editor of the IEEE Transactions on Automatic Control and several other journals.