IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 22, 2020, accepted July 26, 2020, date of publication July 30, 2020, date of current version August 13, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013165

Very Short-Term Power System Frequency Forecasting

OGUN YURDAKUL[1], (Student Member, IEEE), FATIH ESER[2], FIKRET SIVRIKAYA[3], AND SAHIN ALBAYRAK[1]

[1]Department of Electrical Engineering and Computer Science, Technical University of Berlin, 10587 Berlin, Germany
[2]Department of Electrical and Electronics Engineering, Middle East Technical University, 06800 Ankara, Turkey
[3]German-Turkish Advanced Research Center for ICT, 10587 Berlin, Germany

Corresponding author: Ogun Yurdakul (yurdakul@tu-berlin.de)

This work was supported in part by the German Research Foundation, and in part by the Open Access Publication Fund of TU Berlin.

ABSTRACT Power system frequency plays a pivotal role in ensuring the security, adequacy, and integrity of a power system. While some frequency response services are automatically delivered to maintain the frequency within the stipulated limits, certain cases may require that system operators (*SO*s) manually intervene—against the clock—to take the necessary preventive or corrective actions. As such, *SO*s can be greatly aided by practical tools that afford them greater temporal leeway. To this end, we propose a methodology to forecast the power system frequency in the subsequent minute. We perform an extensive analysis so as to identify the factors that influence power system frequency. By effectively exploiting the identified factors, we develop a forecasting methodology that harnesses the long short-term memory model. We demonstrate the effectiveness of the proposed methodology on Great Britain transmission system frequency data using comparative assessments with selected benchmarks based on various evaluation metrics.

INDEX TERMS Forecasting, frequency control, frequency response, long short-term memory (LSTM), recurrent neural network (RNN).

NOMENCLATURE

$\mathbb{R}^n$	set of real-valued $n-$dimensional column vectors
$\mathbb{R}^{m\times n}$	set of $m-$by$-n$ real-valued matrices
$\mathcal{T}$	length of input sequence, *look-back* window
$\mathcal{M}$	number of input features
$\mathcal{B}$	number of *LSTM* blocks
$\mathcal{L}$	learning rate
$\mathcal{N}$	number of training samples
$\boldsymbol{f}_t \in \mathbb{R}^{\mathcal{T}}$	vector of frequency measurements for the previous $\mathcal{T}$ time periods from time period t
$\tilde{\boldsymbol{f}}_t \in \mathbb{R}^{\mathcal{T}}$	vector of values obtained after the application of min-max scaling to each corresponding element of $\boldsymbol{f}_t$
$\boldsymbol{p}_t \in \mathbb{R}^{\mathcal{T}}$	vector of power system load in the previous $\mathcal{T}$
$\tilde{\boldsymbol{p}}_t \in \mathbb{R}^{\mathcal{T}}$	vector of values obtained after the application of min-max scaling to each corresponding element of $\boldsymbol{p}_t$
h_t	the hour of the day of the forecast time period t
$\tilde{\boldsymbol{h}}_t \in \mathbb{R}^{24}$	one-hot encoding representation of h_t
d_t	the day of the week of the forecast time period t
$\tilde{\boldsymbol{d}}_t \in \mathbb{R}^7$	one-hot encoding representation of d_t
i	subscript that denotes the input gate of an *LSTM* block
φ	subscript that denotes the forget gate of an *LSTM* block
c	subscript that denotes the memory cell of an *LSTM* block
o	subscript that denotes the output gate of an *LSTM* block
$(\cdot)^\mathsf{T}$	transposition
$\sigma(\cdot)$	logistic sigmoid function
$\tanh(\cdot)$	hyperbolic tangent function
$\odot$	Hadamard product
$\ell = 1, \ldots, \mathcal{T}$	index of an *LSTM* sequence time step
t	index of a forecast time period

The associate editor coordinating the review of this manuscript and approving it for publication was Ruisheng Diao.

$(\boldsymbol{x}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell, \ldots, \boldsymbol{x}_{\mathcal{T}})$	*LSTM* input sequence
$\boldsymbol{x}_\ell \in \mathbb{R}^{\mathcal{M}}$	input vector provided to the *LSTM* network in time step ℓ
$(\boldsymbol{c}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_\ell, \ldots, \boldsymbol{x}_{\mathcal{T}})$	sequence of *LSTM* network memory cell states
$\boldsymbol{c}_\ell \in \mathbb{R}^{\mathcal{B}}$	vector of memory cell states in time step ℓ
$(\boldsymbol{\kappa}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_\ell, \ldots, \boldsymbol{\kappa}_{\mathcal{T}})$	*LSTM* output sequence
$\boldsymbol{\kappa}_\ell \in \mathbb{R}^{\mathcal{B}}$	vector of *LSTM* block outputs in time step ℓ
$\boldsymbol{W}_{xi}, \boldsymbol{W}_{x\varphi}, \boldsymbol{W}_{xc}, \boldsymbol{W}_{xo} \in \mathbb{R}^{\mathcal{B}\times\mathcal{M}}$	input weight matrices
$\boldsymbol{W}_{\kappa i}, \boldsymbol{W}_{\kappa\varphi}, \boldsymbol{W}_{\kappa c}, \boldsymbol{W}_{\kappa o} \in \mathbb{R}^{\mathcal{B}\times\mathcal{B}}$	recurrent weight matrices
$\boldsymbol{w}_{ci}, \boldsymbol{w}_{c\varphi}, \boldsymbol{w}_{co} \in \mathbb{R}^{\mathcal{B}}$	peephole weight vectors
$\boldsymbol{b}_i, \boldsymbol{b}_\varphi, \boldsymbol{b}_c, , \boldsymbol{b}_o \in \mathbb{R}^{\mathcal{B}}$	bias vectors
$\boldsymbol{X}_t := [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell, \ldots, \boldsymbol{x}_{\mathcal{T}}]^{\mathsf{T}} \in \mathbb{R}^{\mathcal{T}\times\mathcal{M}}$	*LSTM* network input matrix for the forecast time period t

I. INTRODUCTION

A large-scale power system is—for all intents and purposes—a massive and highly complex machine, with generators rotating in steady-state synchronism to supply electricity to meet the system load. The speed of this rotation is referred to as frequency and measured in *cycles/second* or *Hertz* (*Hz*).

We refer to the frequency at which a power system is designed to operate as target frequency. While the target frequency is 50 *Hz* in the larger part of the world including continental Europe and Great Britain (*GB*), it is 60 *Hz* in North America.

The reliability of a power system is described by two fundamental concepts: power system security[1] and power system adequacy. Power system security is the ability of a power system to withstand sudden disturbances. The frequency of a power system is a key aspect that needs to be monitored to ensure the security of a power system, as large frequency excursions may result in equipment damages. Power system adequacy is defined as the ability of a power system to supply the aggregate electric power and energy requirements at all times [1]. The frequency of a power system has a vital role in power system adequacy, since large frequency excursions may trigger under-frequency or over-frequency relays that automatically interrupt loads and/or require manual load shedding [2].

The frequency of a power system may change instantaneously based on the total power injection vis-à-vis withdrawal in a power system. Power systems are continually subject to imminent disturbances, such as sudden changes in load, generator outages, and equipment failures, which may bring about large frequency excursions. To ensure a secure and reliable power system operation at all times, various frequency response services are harnessed to keep the frequency within the stipulated limits.

In the time frame of seconds, the frequency response services are delivered mainly by governors and certain loads. Generators are equipped with governors, which sense the changes in speed of rotation and accordingly adjust the input valve position to modulate the mechanical power output. If the generators in a power system slow down due to declining frequency, governors ensure that more energy is supplied to the prime movers of the generators, thereby counteracting frequency decline [3]. Further, rotating and inductive loads automatically consume less power as frequency decreases, and so analogously restrain frequency excursions [4]. While these frequency response services may arrest frequency increase or decline, it is required to utilize frequency response services with longer time frames to restore the frequency in the event of a large frequency excursion [5].

The frequency response services delivered in the time scale of minutes involve automatic generation control and the manual actions of system operators (*SO*s), including phone calls to generators and purchases and sales of electricity [5]. In the event of a large frequency excursion, an *SO* may further need to perform reserve deployment or take corrective actions, such as manual load shedding, to ensure that the system transitions to a restorative state from an emergency state [3], [6].

*SO*s perform such actions in a fire-fighting mentality so as to forestall the onset of cascading outages. As such, time is of cardinal importance to *SO*s, and *SO*s can be greatly aided by a tool that forecasts the frequency in the subsequent minute.

In this article, we propose such a methodology to forecast the power system frequency in the subsequent minute. The proposed methodology is based on long short-term memory (*LSTM*), which is an artificial recurrent neural network (*RNN*) model with the salient capability to capture long-term temporal dependencies [7]. To design the proposed methodology, we conduct a thorough assessment of the factors that influence power system frequency.

We design our model generic enough to permit its application to any power system. This notwithstanding, a specific application of the designed model would be remiss, if it were to disregard the salient characteristics of the system to which it is applied. On the grounds that we illustrate the application of the proposed methodology to forecast the *GB* power system frequency, we pay special attention to the frequency limits and frequency response services in *GB*. We further investigate the influence of system load and system inertia on frequency excursions and analyze the variation of system load and system inertia over time so as to gain insights into the underlying phenomena that affect power system frequency. These insights are explicitly leveraged in the design of the proposed forecasting methodology.

It is worth noting that we specifically aim at forecasting the frequency in the subsequent time minute, because the discussion above made clear that the frequency response

[1]Since 2001, the North American Electric Reliability Corporation has been using the term *operational reliability* instead of security due to the possible misinterpretations of the word *security*.

services effective in the time frame of seconds are delivered in an automatic way. On the flip side, the manual interventions of *SO*s are typically exercised in the time frame of minutes. As such, it makes sense to develop a tool that forecasts the frequency in the subsequent minute. *SO*s can capitalize on such a tool to judiciously decide on and effectuate the necessary course of actions with an extra minute at their disposal. In the literature, forecasting with a lead time of between a few seconds and one hour is referred to as very short-term forecasting [8]. Nevertheless, the proposed approach can be easily extended to perform forecasts with different lead times.

A. RELATED WORK

In the literature, very few studies have been conducted to forecast power system frequency. In [9], an auto-regressive integrated moving average model has been proposed to forecast power system frequency. In contrast to our approach, [9] forecasts the frequency with a 15-minute lead time, does not take into account the influence of relevant exogenous variables on frequency variation, and harnesses statistical methods rather than *RNN*s. While [10] and [11] utilize artificial neural networks to forecast power system frequency, they do not address the influence of system load or system inertia on frequency response and only consider historical values. We further note the work conducted in [12], which investigates the frequency response services in *GB* and provides a comparative analysis of different forecasting approaches to forecast the frequency-corrected demand of *GB*.

A line of research similar to forecasting power system frequency is forecasting power system imbalance, that is, difference between system demand and supply. Quantile Random Forecasts are utilized in [13] to forecast the power system imbalance over the next two hours, where historical imbalances and temporal information are incorporated as features. In [14], artificial neural networks are used to forecast daily imbalance medians. Hourly imbalances have been forecast in [15] using random forests, where the proposed methodology includes a wide variety of features, such as wind forecast, load forecast, and temperature.

B. CONTRIBUTIONS AND STRUCTURE OF THE ARTICLE

The general contributions and novel aspects of this article are as follows:

1) We develop a novel *LSTM*-based methodology to forecast power system frequency.[2] We demonstrate the effectiveness of the proposed methodology on real-world *GB* power system frequency data using various benchmarks and evaluation metrics. Given that only a handful of studies focused on power system frequency forecasting thus far, the present article may serve as a jumping-off point for future studies in this field.
2) We provide valuable insights into the influence of system load and system inertia—as well as their variations over different time horizons—on power system frequency. We effectively exploit the gained insights to select the relevant features of the designed *LSTM* network. As such, unlike the existing methods in the literature that solely consider historical frequency, we additionally capitalize on exogenous variables in our forecasting methodology.
3) Our study presents a novel application area for *LSTM* networks. While *LSTM* networks have found a wide range of applications—load forecasting [17], speech recognition [18], handwriting recognition [19], and music composition [20] to name but a few—this article serves as the first study that utilizes *LSTM* networks to forecast power system frequency.
4) We investigate the influence of hyperparameters and conduct an analysis of the impact of number of *LSTM* blocks, look-back window, and learning rate on forecasting performance, which serves as a useful guideline for their efficient tuning.
5) *SO*s can utilize the developed methodology as a practical tool to assess whether their manual intervention may be required in the subsequent minute to keep the frequency within the stipulated limits. As such, the developed methodology makes a dent in the development of a decision support tool that can provide additional temporal leeway to *SO*s.

The remainder of the article consists of four sections. In Section II, we investigate the factors that influence power system frequency and identify the relevant features of the proposed forecasting methodology. We briefly describe *LSTM* in Section III and present the proposed forecasting methodology. In Section IV, we illustrate the application of the proposed methodology to forecast the *GB* power system frequency in the subsequent minute and discuss the results. We present concluding remarks and discuss the scope of further work in Section V.

II. FREQUENCY RESPONSE SERVICES

We devote this section to the discussion of the steps undertaken so as to pinpoint the pertinent features of a methodology to forecast the frequency of a power system. Since we illustrate the application of the proposed frequency forecasting methodology in Section IV on the *GB* transmission system, we pay special attention to study the key details on the frequency limits and frequency response services in *GB*.

We start out by the investigation of the factors that influence the power system frequency. The frequency of a power system may change around the clock based on the total power injection vis-à-vis withdrawal in a power system. In fact, if the total power injection to a power system is greater (resp. less) than the total power withdrawal from a power system, then the power system frequency increases (resp. decreases). While generator outputs and inter-regional power imports contribute to the power injection to a system, the power withdrawal from a system can be ascribed to system load, inter-regional power exports as well as losses [5], [21].

[2]The source code and simulation scripts are available at [16].

The frequency of a power system is typically subject to various statutory and operational limits. In *GB*, while the statutory limits stipulate that the frequency must be kept between 49.5 *Hz* and 50.5 *Hz*, the operational limits are stricter and state that the frequency must be between 49.8 *Hz* and 50.2 *Hz* [22]. National Grid Electricity System Operator (*NGESO*)—the entity responsible for the around-the-clock power balance in the *GB* transmission system—maintains the frequency within the stipulated limits by leveraging various frequency response services, *viz.*: primary response, secondary response, and high response, as depicted in Fig. 1 [4]. These frequency response services are rendered over various time frames so as to attenuate frequency excursions and bring frequency toward the target frequency. While primary response and secondary response are delivered to increase the frequency by an automatic increase in generation or decrease in load, high response is delivered to decrease the frequency by an automatic decrease in generation or increase in load [23].

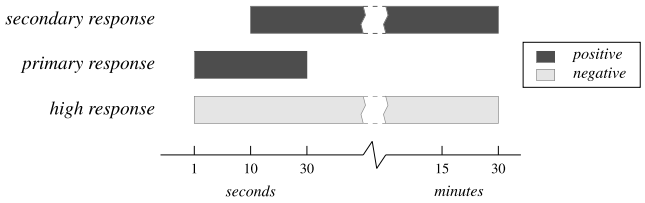


FIGURE 1. Time frames of the frequency response services in *GB*.

Central to the discussion of frequency excursion is the frequency response characteristic, or β, defined by the North American Electric Reliability Corporation (*NERC*) as "the total of all initial responses to a frequency excursion" [5]. β represents the actual initial *MW* contribution deployed to stabilize frequency following a disturbance in *MW/0.1Hz* and has a negative sign [6]. A disturbance with a specific *MW* size causes a larger frequency excursion in the time periods of lower β than those of higher β. As such, it would behoove us to closely examine the factors that influence β and draw on these factors in the design of the proposed methodology.

One such factor is system load: β rises as system load increases [5]. To elucidate the relationship between system load and β, we probe the β values for four Interconnections in North America, *viz:* Eastern Interconnection, Western Interconnection, Texas Interconnection, and Quebec Interconnection. Each of the four Interconnections operates as a frequency-independent island and has a distinct frequency response characteristic. We present in Table 1 the mean β values for the four Interconnections along with their net energy for load values in 2018 [24], [25]. The reported β values bring out the general trend that the magnitude of β increases with system load. Specifically, while a 0.02-*Hz* deviation in frequency prompts a 482-*MW* response in the Eastern Interconnection, it elicits a 172-*MW* response in the Quebec Interconnection.

Further, β is favorably impacted by the system inertia, which depends on the total energy stored in the rotating masses of a system, such as synchronous generators and synchronous electric motors. β attains higher values in the time periods of higher system inertia than those of lower system inertia [5]. The governor response is a major contributor to frequency response characteristic, which involves the adjustment of the generator output based on a ratio that depends on the deviation of the frequency from the target frequency. This ratio is referred to as speed regulation and quantifies the percentage change in frequency that causes the generator output to change from 0% to 100% so as to counteract frequency deviation [26]. This notwithstanding, the ability of a generator to deliver governor response at any moment hinges on the capacity at which it operates. In the event of a frequency decline, a generator must have the capability to increase its output so as to remedy the decline in frequency.

TABLE 1. Frequency response characteristics and loads of interconnections in North America.

Interconnection	mean β (*MW*/0.1*Hz*)	net energy for load (*TWh*)
Eastern	-2,411	3,182.5
Western	-1,789	877.2
Texas	-940	376.4
Quebec	-862	188.5

In addition to synchronous generators, the inertia of loads such as synchronous electric motors contributes to frequency response characteristic. As the frequency falls below the target frequency, electric motors slow down and consume less power, thereby restraining the frequency decline. The decrease in motor load takes effect immediately after a decline in frequency and typically, for every 1% decline in frequency, the motor load decreases by 3% [27].

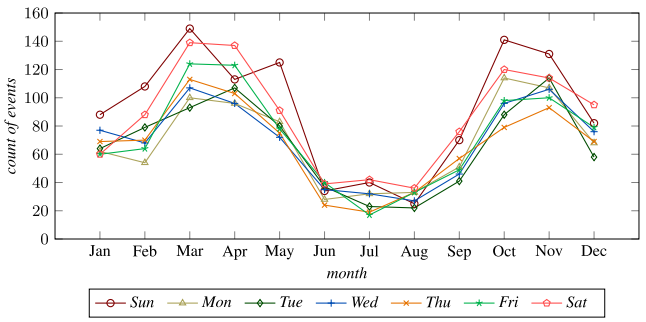


FIGURE 2. Number of frequency excursions greater than 28 *mHz* in the Eastern Interconnection by the month of the year and the day of the week, reconstructed from [5].

The empirical studies corroborate the tight coupling between the magnitude of frequency excursions, β, and system load and inertia. We observe from Fig. 2 that the number of frequency excursions greater than 28 *mHz* in the Eastern Interconnection is lower in summer and higher on the weekend. These observations jibe with the fact that the system load—ergo the synchronous generation and system inertia—are higher in summer and on the weekdays, thereby

yielding higher β values. Further, these observations reinforce the need that the proposed methodology effectively capture intertemporal relationships and beget the idea that the proposed methodology can be greatly aided by the incorporation of the system load and the temporal information of the forecast time period as features.

III. FREQUENCY FORECASTING METHODOLOGY

The analysis in Section II highlights the key factors that influence power system frequency and punctuates the necessity of the consideration of time in the development of a methodology to forecast power system frequency. Indeed, a major requirement of such a methodology is the ability to effectively grasp the intertemporal relationships among data points. To this end, we harness the *LSTM* model, whose salient capability is to learn long-term temporal dependencies.

A. LONG SHORT-TERM MEMORY MODEL

We devote this subsection to the delineation of the *LSTM* model. Fig. 3 depicts the *LSTM* block employed in our model.

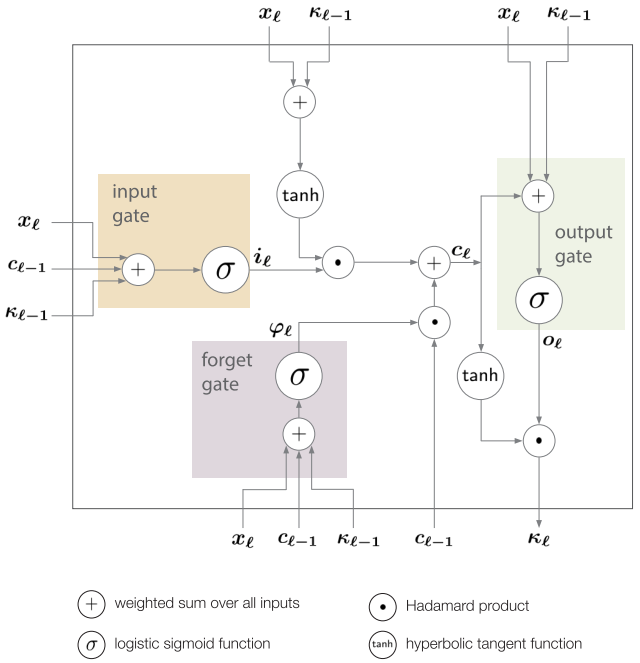


FIGURE 3. Graphical depiction of the *LSTM* block.

An *LSTM* block includes three *gates*: the input gate, the forget gate, and the output gate, which control the information flow of an *LSTM* block. A salient feature of an *LSTM* block is the *memory cell*, which enables an *LSTM* block to remember important information over time, subject to the regulation of the *LSTM* gates [7].

We next describe analytically the flow of information in an *LSTM* network and the operation of the *LSTM* gates. We consider the input sequence $(\boldsymbol{x}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_\ell, \ldots, \boldsymbol{x}_{\mathcal{T}})$. Each element of the sequence $(\boldsymbol{x}_\ell)_{\ell=1}^{\mathcal{T}}$ is an $\mathcal{M}-$dimensional vector, where $\mathcal{M}$ is the number of input features. We define by $(\boldsymbol{\kappa}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{\kappa}_1, \ldots, \boldsymbol{\kappa}_\ell, \ldots, \boldsymbol{\kappa}_{\mathcal{T}})$ the output sequence, and by $(\boldsymbol{c}_\ell)_{\ell=1}^{\mathcal{T}} := (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_\ell, \ldots, \boldsymbol{x}_{\mathcal{T}})$ the sequence of memory cell states corresponding to the input sequence $(\boldsymbol{x}_\ell)_{\ell=1}^{\mathcal{T}}$. Each element of $(\boldsymbol{c}_\ell)_{\ell=1}^{\mathcal{T}}$ and $(\boldsymbol{\kappa}_\ell)_{\ell=1}^{\mathcal{T}}$ is a $\mathcal{B}-$dimensional vector, where $\mathcal{B}$ is the number of *LSTM* blocks [7].

The forward pass of an *LSTM* network can be represented by iterating the equations (1)-(5) for $\ell = 1, \ldots, \mathcal{T}$.

$$\boldsymbol{i}_\ell = \sigma\big(\boldsymbol{W}_{xi}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa i}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{w}_{ci} \odot \boldsymbol{c}_{\ell-1} + \boldsymbol{b}_i\big), \tag{1}$$

$$\boldsymbol{\varphi}_\ell = \sigma\big(\boldsymbol{W}_{x\varphi}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa\varphi}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{w}_{c\varphi} \odot \boldsymbol{c}_{\ell-1} + \boldsymbol{b}_\varphi\big), \tag{2}$$

$$\boldsymbol{c}_\ell = \boldsymbol{\varphi}_\ell \odot \boldsymbol{c}_{\ell-1} + \boldsymbol{i}_\ell \odot \tanh\big(\boldsymbol{W}_{xc}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa c}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{b}_c\big), \tag{3}$$

$$\boldsymbol{o}_\ell = \sigma\big(\boldsymbol{W}_{xo}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa o}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{w}_{co} \odot \boldsymbol{c}_\ell + \boldsymbol{b}_o\big), \tag{4}$$

$$\boldsymbol{\kappa}_\ell = \boldsymbol{o}_\ell \odot \tanh\big(\boldsymbol{c}_\ell\big), \tag{5}$$

where the $\boldsymbol{W}$ and $\boldsymbol{w}$ terms represent the weights and the $\boldsymbol{b}$ terms represent the biases [7], [18], [28]. Further, $\sigma(\cdot)$ denotes the logistic sigmoid function, $\tanh(\cdot)$ the hyperbolic tangent function, and $\odot$ the Hadamard product.

The operation of the input gate is expressed by (1), which certifies that the input gate evaluates a wide range of information, both from the previous time step and the current time step, to decide on the extent to which the input in the current step is incorporated to the *LSTM* network. The input gate is provided with the input in time step ℓ, the memory cell state in time step $\ell-1$, and the *LSTM* block output in time step $\ell-1$ that go through a hyperbolic tangent function to generate a value between zero and one. Higher input gate values ensure that the input is more heavily incorporated to the memory cell state and the *LSTM* block output calculation in time step ℓ [17].

A key architectural element of the *LSTM* network in terms of propagating information across time steps is the forget gate, which signifies the influence of the memory cell state in time step $\ell-1$ on the memory cell state and the *LSTM* block output calculation in time step ℓ, as expressed in (2). The forget gate attains a value between zero and one; the closer is the value of the forget gate to zero, the more liable is the memory cell state in $\ell-1$ to be forgotten [28].

The memory cell state, in conjunction with the operation of the gates, spearheads the storage of valuable information in an *LSTM* network. To compute the memory cell state in time step ℓ, the input and forget gate values in time step ℓ are jointly evaluated with the memory cell state and output in time step $\ell-1$, as expressed by (3). The memory cell state is utilized by the output gate, which further evaluates the input in time step ℓ and the *LSTM* block output in time step $\ell-1$ to generate a value between zero and one, as represented by (4). Finally, the output gate value and the memory cell state in time step ℓ are assessed in (5) to compute the *LSTM* block output, which attains a value between -1 and 1 due to the utilization of the hyperbolic tangent function.

The description of the *LSTM* model presented in this subsection spells out the extensive interrelationships among the input, memory cell state, and output of different time steps as per the regulation of the *LSTM* gates. Next, we present the

specific details as to the utilization of the *LSTM* model in the proposed forecasting methodology.

B. PROPOSED METHODOLOGY

The proposed methodology harnesses an *LSTM* network that has four input features: the frequency measurements for the previous $\mathcal{T}$ time periods, the power system load in the previous $\mathcal{T}$ time periods, the day of the week of the forecast time period, and the hour of the day of the forecast time period. The term $\mathcal{T}$ denotes the so-called *look-back* window. Since *LSTM* networks are sensitive to the scale of the input data, we start out by preprocessing the data of the input features.

We denote by $\boldsymbol{f}_t \in \mathbb{R}^{\mathcal{T}}$ the vector of frequency measurements for the previous $\mathcal{T}$ time periods, where the subscript t denotes the forecast time period. The elements of $\boldsymbol{f}_t$ largely hover around 50 *Hz* and so we apply *min-max scaling* to scale each element of $\boldsymbol{f}_t$ to the range $[-1, 1]$. We define by $\tilde{\boldsymbol{f}}_t \in \mathbb{R}^{\mathcal{T}}$ the vector of values obtained after the application of min-max scaling to each corresponding element of $\boldsymbol{f}_t$. Let $\boldsymbol{p}_t \in \mathbb{R}^{\mathcal{T}}$ denote the column vector of power system load in the previous $\mathcal{T}$ time periods. We similarly apply min-max scaling to scale each element of $\boldsymbol{p}_t$ to the range $[-1, 1]$ and denote by $\tilde{\boldsymbol{p}}_t \in \mathbb{R}^{\mathcal{T}}$ the vector of values obtained after min-max scaling. Let h_t denote the hour of the day and d_t denote the day of the week of the forecast time period t. For instance, if the frequency is forecast for a time period from 11:00 a.m. up to but not including 12:00 p.m of a Tuesday, then $h_t = 12$ and $d_t = 2$.

We use one-hot encoding to express h_t and d_t. One-hot encoding refers to the representation of a variable that can take M categorical values by a vector in the M-dimensional Euclidean space, where only the j^{th} element of the vector corresponding to the category j is one and the remaining elements of the vector are zero. We express h_t and d_t in one-hot encoding by the column vectors $\tilde{\boldsymbol{h}}_t \in \mathbb{R}^{24}$ and $\tilde{\boldsymbol{d}}_t \in \mathbb{R}^{7}$, respectively. We construct the matrix $\tilde{\boldsymbol{H}}_t := \left[\tilde{\boldsymbol{h}}_t, \cdots, \tilde{\boldsymbol{h}}_t\right]^{\mathsf{T}} \in \mathbb{R}^{\mathcal{T}\times 24}$ where each row of the matrix $\tilde{\boldsymbol{H}}_t$ is $(\tilde{\boldsymbol{h}}_t)^{\mathsf{T}}$. Analogously, we construct the matrix $\tilde{\boldsymbol{D}}_t := \left[\tilde{\boldsymbol{d}}_t, \cdots, \tilde{\boldsymbol{d}}_t\right]^{\mathsf{T}} \in \mathbb{R}^{\mathcal{T}\times 7}$.

We define by $\boldsymbol{X} := \left[\tilde{\boldsymbol{f}}_t, \tilde{\boldsymbol{p}}_t, \tilde{\boldsymbol{H}}_t, \tilde{\boldsymbol{D}}_t\right] \in \mathbb{R}^{\mathcal{T}\times 33}$ the input matrix of the proposed forecasting methodology. Each row of the matrix $\boldsymbol{X}$ is input sequentially in $\mathcal{T}$ steps to the proposed neural network, whose construction is described next.

We utilize *LSTM* blocks to form an *LSTM* layer and denote by $\mathcal{B}$ the number of *LSTM* blocks in each *LSTM* layer. Owing to the sequential nature of *LSTM*s, we may stack one or multiple *LSTM* layers to construct an *LSTM* network. We elaborate on the process to tune the network hyperparameters in Section IV. We feed the output of the *LSTM* blocks of the topmost layer to a conventional feedforward neural network with a single neuron, whose output is the power system frequency forecast for the subsequent time period. We provide the graphical depiction of the proposed forecasting methodology in Fig. 4.

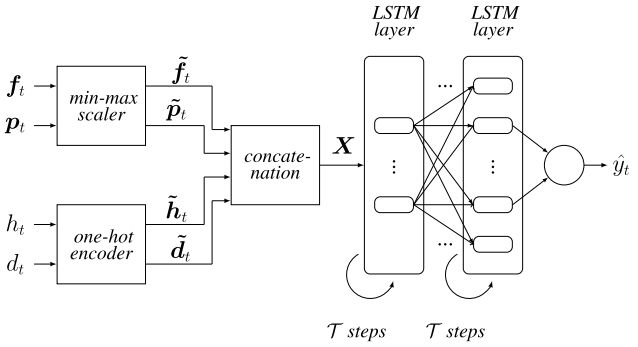


FIGURE 4. Graphical depiction of the proposed forecasting methodology.

We denote by $\hat{y}_t$ the power system frequency forecast for the subsequent time period t. Let y_t denote the actual power system frequency measurement for the time period t. We train the constructed neural network with $\mathcal{N}$ training samples to minimize the loss function:

$$\sum_{t=\mathcal{T}+1}^{\mathcal{T}+\mathcal{N}} \left(y_t - \hat{y}_t\right)^2. \tag{6}$$

We denote by $\mathcal{L}$ the learning rate of the optimizer with which we train the constructed neural network. The selection of the optimizer and the assignment of specific values to $\mathcal{L}$ are discussed in Section IV.

IV. CASE STUDY AND RESULTS

We illustrate the application of the proposed methodology to forecast the *GB* transmission system frequency in the subsequent time period.

A. DATASET

The original dataset [29] contains historical frequency measurements collected by the *NGESO* at one-second resolution and we take the measurements collected in two time windows so as to form two separate datasets for our experiments. We first consider the frequency measurements collected from August 11, 2017 to October 9, 2017 and apply downsampling by a factor of 60 to obtain a dataset at one-minute resolution, which we refer to as Dataset I. We split Dataset I into training (70%), validation (15%), and test (15%) sets.

To ascertain the generality of the model, we study its performance to forecast the power system frequency in another year. To this end, we consider the *GB* transmission system frequency measurements collected from July 2, 2018 to July 10, 2018 and apply downsampling by a factor of 60 to obtain a dataset at one-minute resolution, which we refer to as Dataset II. We utilize Dataset II solely as a test set. We further use the *GB* transmission system demand data [30] for the time periods covered by the time windows of Datasets I and II.

B. BENCHMARKS

Due to the scarcity of existing methods in the literature to forecast the power system frequency, we investigate

TABLE 2. Forecasting performances.

model	*MSE*				*MAE*				*MAPE*			
	Dataset I test set		Dataset II		Dataset I test set		Dataset II		Dataset I test set		Dataset II	
	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)
PM	0.12038	0.25410	0.15030	0.29104	2.61378	2.28178	2.89446	2.57918	5.22767	4.56283	5.78866	5.15766
GRU	0.12278	0.25280	0.15093	0.29152	2.65569	2.28605	2.90049	2.58475	5.31159	4.57141	5.80092	5.16926
SRN	0.12366	0.25319	0.15268	0.28951	2.66651	2.29247	2.93139	2.58350	5.33345	4.58495	5.86318	5.16787
NPM	0.13532	0.36003	0.16612	0.34973	2.70257	2.49551	2.95907	2.80278	5.40517	4.99003	5.91809	5.60541
SMM	0.46118	0.64188	0.50500	0.64527	5.47453	4.01845	5.85340	4.02960	10.94900	8.03323	11.70590	8.05572

different *RNN* models and empirical benchmarks with which we compare our proposed methodology. To this end, we harness the simple recurrent network (*SRN*) model, which is a rudimentary *RNN* architecture void of the gate structures of more complex *RNN* models, such as the *LSTM* model. We further utilize the gated recurrent unit (*GRU*) model, which is an *RNN* model furnished with forget and reset gates. We present the details of the operation and internal architecture of the *SRN* and *GRU* models in Appendix.

It is often customary to compare very short-term forecasts with empirical benchmarks [8]. For this purpose, we use the *naïve persistence model* (*NPM*), which forecasts that the value measured for the current time period will also be observed in the subsequent time period. We further utilize the *statistical mean model* (*SMM*), which generates a forecast that is equal to the statistical mean of the historical measurements collected in the same hour of the day and on the same day of the week of the forecast time period.

C. EXPERIMENTAL RESULTS

We utilize Tensorflow [31] and Keras [32] to train, validate, and test the constructed neural network. We use the validation set to tune the hyperparameters and settle on the architecture with 1 *LSTM* layer comprising 48 *LSTM* blocks, i.e., $\mathcal{B} = 48$. We select $\mathcal{T} = 3$ and use the Adam [33] optimizer with $\mathcal{L} = 3\,(10^{-4})$. We similarly use the validation set to tune the hyperparameters of the *GRU* and *SRN* models and decide on 1 layer comprising 48 neurons for the *GRU* model and 1 layer comprising 48 neurons for the *SRN* model.

We provide in Table 2 the mean squared error (*MSE*), the mean absolute error (*MAE*), and the mean absolute percentage error (*MAPE*) on Dataset I test set and Dataset II of the proposed model (*PM*) as well as those of the *GRU* model, the *SRN* model, and the empirical benchmarks *NPM* and *SMM*. Each ''mean'' (resp. ''std'') column indicates the unweighted average (resp. standard deviation) of all forecast errors based on the corresponding metric.

The results show that, on both test sets, the proposed model yields lower mean errors than do the benchmark models based on all metrics. Specifically, the mean errors on Dataset II attest to the capability of the proposed methodology to generate successful forecasts for a year different from that for which it is trained. We further note that, in each but one test, the standard deviation of the forecast errors obtained via the proposed model is lower than that of the other benchmark models, which indicates that the forecast errors of the proposed model are more clustered around the mean values and less spread out compared with the benchmark models. On the heels of the proposed model is the *GRU* model, which produces lower mean errors than do the remaining benchmark models based on all metrics. In fact, the standard deviation of the *GRU* model forecast errors on Dataset I test set based on the *MSE* metric is also lower than that of the proposed model. These results pinpoint the capability of the *GRU* model to effectively learn temporal relationships. While the *SRN* model mostly yields lower mean errors compared with the empirical benchmarks, it results in higher mean errors compared with the proposed model and the *GRU* model. Such results are representative of the shortcomings of the *SRN* model in learning long-term dependencies.

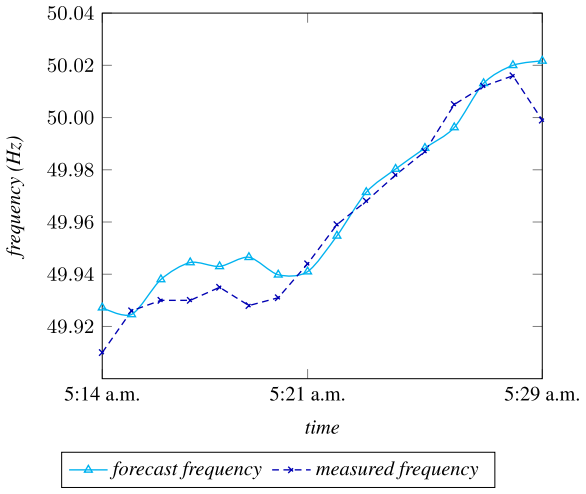


FIGURE 5. Forecast and measured frequency for the interval 5:14 a.m.-5:29 a.m. on August 13, 2017 on Dataset I training set.

Figs. 5, 6, and 7 present the *GB* transmission system frequency forecasts for a representative interval in Dataset I training set, Dataset I test set, and Dataset II, respectively, along with the actual corresponding *GB* transmission system frequency measurements. Figs. 5-7 visually verify that the

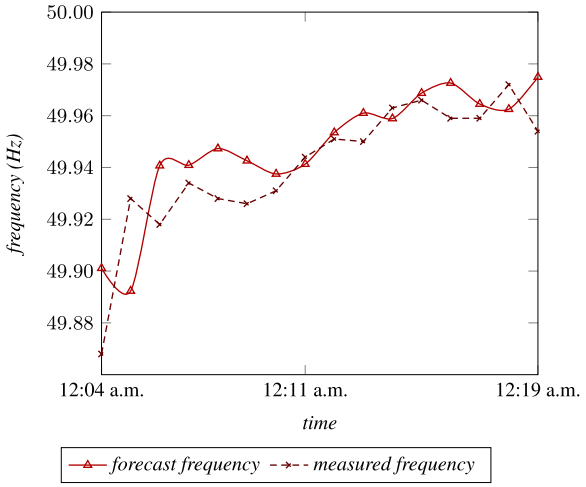


FIGURE 6. Forecast and measured frequency for the interval 12:04 a.m.-12:19 a.m. on October 2, 2017 on Dataset I test set.

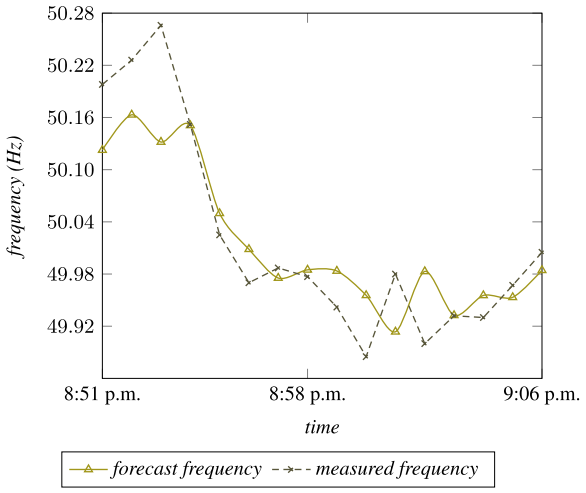


FIGURE 7. Forecast and measured frequency for the interval 8:51 p.m.-9:06 p.m. on July 5, 2018 on Dataset II.

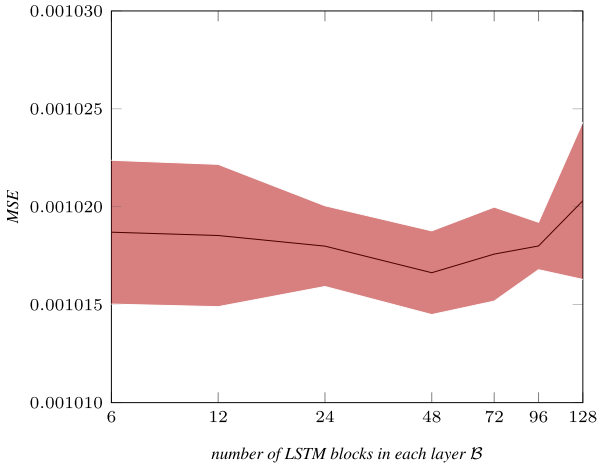


FIGURE 8. Influence of the number of *LSTM* blocks in each layer on forecasting performance.

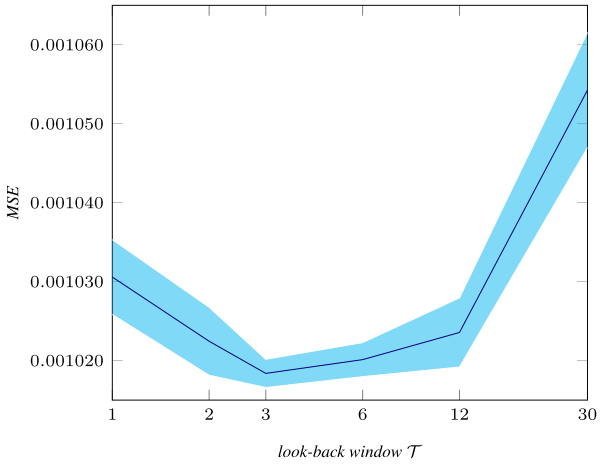


FIGURE 9. Influence of the look-back window on forecasting performance.

proposed forecasting methodology is able to generate forecasts that follow the general trend of the actual frequency measurements.

D. INFLUENCE OF THE HYPERPARAMETERS

We delve into the influence of $\mathcal{B}$, $\mathcal{T}$, and $\mathcal{L}$ on the forecasting performance. To this end, we probe the influence of each hyperparameter individually, while keeping the value of the remaining two hyperparameters constant at the values described in the previous subsection, i.e., $\mathcal{B} = 48$, $\mathcal{T} = 3$, and $\mathcal{L} = 3\,(10^{-4})$. Since it is impractical to traverse the entire hyperparameter space, we investigate a finite set of values for each hyperparameter. We conduct seven experiments for each investigated value of each hyperparameter so as to take into the account the influence of the random initialization of the neural network weights. We conduct experiments with the following number of *LSTM* blocks in each layer: $\mathcal{B} = 6, 12, 24, 48, 72, 96, 128$. The look-back window values with which we perform experiments are: $\mathcal{T} = 1, 2, 3, 6, 12, 30$. We further perform experiments with the following learning rates: $\mathcal{L} = 10^{-4}, 3\,(10^{-4}), \sqrt{10}\,(10^{-4}), 10^{-3}, \sqrt{10}\,(10^{-3}), 10^{-2}$.

Figs. 8, 9, and 10 illustrate the influence of $\mathcal{B}$, $\mathcal{T}$, and $\mathcal{L}$, respectively, on the performance of the forecasts on Dataset I validation set. The solid lines in Figs. 8-10 indicate the mean *MSE* of the seven experiments conducted for each investigated hyperparameter value. Further, the shaded regions around the solid lines depict one standard deviation around the mean, computed similarly from the performed seven experiments for each investigated hyperparameter value.

The influence of $\mathcal{B}$ on the forecasting performance is illustrated in Fig. 8. The results show that as $\mathcal{B}$ increases from 6 to 48, the mean *MSE* value decreases, which may be attributed to the improved capability of the neural network to grasp the underlying structure. Nevertheless, the assignment of increasingly large values to $\mathcal{B}$ hurts the forecasting performance.

TABLE 3. Forecasting performances in the presence of frequency measurement errors.

model	*MSE*				*MAE*				*MAPE*			
	modified Dataset I test set		modified Dataset II		modified Dataset I test set		modified Dataset II		modified Dataset I test set		modified Dataset II	
	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)	mean (%)	std (%)
PM	0.12091	0.25909	0.15127	0.29294	2.61952	2.28674	2.90733	2.58342	5.2391	4.5726	5.81442	5.16617
% change	*0.44065*	*1.96461*	*0.64301*	*0.65224*	*0.21947*	*0.21736*	*0.44469*	*0.16446*	*0.21886*	*0.21507*	*0.44496*	*0.16509*
GRU	0.12388	0.25365	0.15246	0.29382	2.66639	2.29739	2.91798	2.59441	5.33279	4.59359	5.83592	5.18866
SRN	0.12517	0.25276	0.15290	0.28803	2.68086	2.30876	2.95273	2.56351	5.36214	4.61750	5.90580	5.12764
NPM	0.13562	0.36016	0.16624	0.34972	2.70580	2.49820	2.96266	2.80121	5.41164	4.99542	5.92526	5.60228
SMM	0.46128	0.64191	0.50481	0.64520	5.47544	4.01843	5.85214	4.02911	10.95083	8.03318	11.70336	8.05471

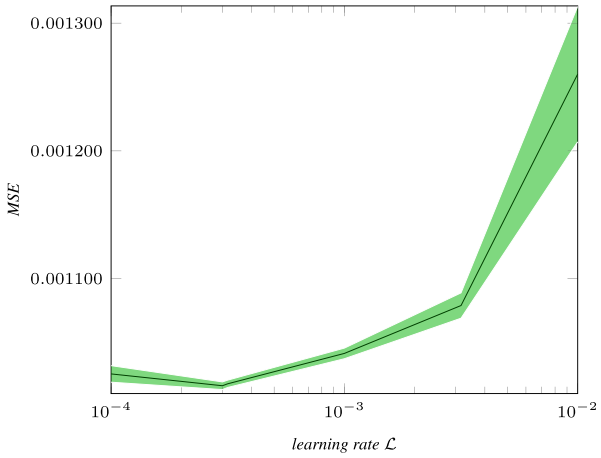


FIGURE 10. Influence of the learning rate on forecasting performance.

We present in Fig. 9 the influence of $\mathcal{T}$ on the forecasting performance. Clearly, $\mathcal{T}$ is a key determinant of forecasting performance. We observe that as $\mathcal{T}$ increases from 1 to 3, the *MSE* decreases, which demonstrates the advantage of explicitly providing multiple historical values. On the flip side, very large $\mathcal{T}$ values hurt the forecasting performance, which can be ascribed to the fact that as $\mathcal{T}$ increases, the complexity of the input data also increases. Fig. 10 illustrates that the learning rate $\mathcal{L}$ has a vital role on the forecasting performance. Among the investigated values, the learning rate of $\mathcal{L} = 3(10^{-4})$ yields the lowest mean *MSE* and especially high $\mathcal{L}$ values result in markedly high *MSE*.

E. INFLUENCE OF THE FREQUENCY MEASUREMENT ERRORS

The analyses conducted so far in this section relied on the premise that the frequency measurements utilized in training, validating, and testing the proposed methodology were void of any errors. The application of the proposed methodology in a real-life environment, however, may entail the utilization of data that contain frequency measurement errors. For this purpose, we devote this subsection to the investigation of the influence of frequency measurement errors on the performance of the proposed methodology, as well as that of the benchmark models.

The IEEE Standard for Synchrophasor Measurements for Power Systems, IEEE Std C37.118.1, spells out requirements for steady-state frequency measurements and specifies the maximum frequency measurement error as 0.005 *Hz* [34]. To explicitly take into account the influence of measurement errors on the proposed methodology, we model the measurement errors by a Gaussian random variable with mean 0 and standard deviation $1.667(10^{-3})$, i.e., $N\left(0, 1.667(10^{-3})\right)$. Such a measurement error model assigns a probability of 0.9973 (resp. 0.0027) to the event that the measurement error is less (resp. greater) than or equal to 0.005 *Hz*.

We follow the approach described in [35] to incorporate the frequency measurement errors to the proposed methodology. To this end, we augment Dataset I training and validation sets with data obtained by adding an error term that follows the distribution $N\left(0, (1.667)10^{-3}\right)$ to each measurement reported in Dataset I training and validation sets. We refer to the resulting datasets as augmented Dataset I training set and augmented Dataset I validation set.

By explicitly including data with, as well as without, error terms in the augmented Dataset I training set, we aim to ensure that the proposed methodology is resilient to measurement errors. We recognize that, in the event of the deployment of the proposed methodology in the control room of an *SO*, the proposed methodology may be provided with measurements containing errors. As such, if the proposed methodology were to generate poor forecasts in the presence of errors, its feasibility would be severely hampered.

We train the model with the loss function (6) using the $2\mathcal{N}$ samples of the augmented Dataset I training set.We similarly follow the approach in [35] to evaluate the performance of the models. We modify the two test sets by adding an error term that follows the distribution $N\left(0, (1.667)10^{-3}\right)$ to each measurement in Dataset I test set and Dataset II and refer to the resulting test sets as modified Dataset I test set and modified Dataset II, respectively. The utilization of these test sets enables us to assess and compare the ramifications of the measurement errors on the performance of each model. We test the performance of each model on modified Dataset I test set and modified Dataset II based on the *MSE*, *MAE*, and *MAPE* metrics and tabulate the results in Table 3. For the proposed methodology, we further report in Table 3 the

percentage changes in the forecast error values due to the inclusion of measurement errors.

In line with our expectations, we remark upon the slight increase in forecast errors of the proposed methodology with the incorporation of the measurement errors. The results further illustrate that the proposed model yields lower mean forecast errors than do the benchmark models on both modified Dataset I test set and modified Dataset II based on all metrics. In a similar vein, the standard deviation of the forecast errors obtained by the proposed methodology is mostly lower than that of the benchmark models. On these grounds, we maintain that the proposed methodology, by and large, performs well in the presence of errors and may lend itself as a powerful tool for *SO*s in real-life applications.

F. COMPUTATIONAL PERFORMANCE

The capability of the proposed forecasting methodology to assist *SO*s in a real-life environment is contingent upon its computation time. For this purpose, we measure the computation time of the proposed forecasting methodology in generating a forecast for the subsequent minute.

We observe that it takes 1.2303 milliseconds (*ms*) for the proposed methodology to generate a single forecast on Intel(R) Xeon(R) *CPU* E5-2699A v4 @ 2.40 *GHz* with 768 *GB RAM*. We further test the performance of the proposed methodology on a personal computer with a four-core AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx and 16 *GB RAM* and obtain a computation time of 1.4679 *ms*. In light of the fact that the proposed methodology aims to provide *SO*s with a frequency forecast for the subsequent minute, a computation time in the time frame of milliseconds renders the proposed tool feasible in real-life applications.

V. CONCLUSION

In this article, we propose a methodology to forecast the frequency of a power system in the subsequent minute. The proposed methodology can aid the *SO*s in the assessment of whether their manual intervention will be required to maintain the frequency within the stipulated limits in the subsequent minute. Since such manual interventions are typically exercised in a race against time, the proposed methodology can provide the *SO*s with an additional leeway of one minute for them to decide on and take the necessary course of actions. We illustrate the application of the proposed methodology on the *GB* power system frequency data. The results attest to the superiority of the performance of the proposed methodology over that of selected benchmarks based on various evaluation metrics. Our extensive investigations provide valuable insights into the influence of the number of *LSTM* blocks in each layer, look-back window, and learning rate on forecasting performance. The proposed methodology's forecasting performance in the presence of measurements errors, coupled with its computation time in the time frame of milliseconds, certifies its feasibility in real-life applications.

While the proposed methodology has been demonstrated on the *GB* transmission system, it can be easily extended to any transmission system independent of its scale. Our analysis in Section II makes clear that tailoring the proposed methodology for a specific system calls for a detailed analysis of the frequency response services effective in that system jointly with the time frames over which they are delivered. The thorough examination of the system load, as well as its diurnal, weekly, and seasonal variation, is further warranted and needed due to the influence of system load on frequency response characteristic. One must also make a conscientious effort to take into account the electricity generation mix so as to investigate the inherent inertia of the system. Finally, the findings of these studies need to be effectively exploited in the adaptation of the input features of the proposed model.

A natural extension of the presented work is to forecast the power system frequency over a time period of multiple minutes. Another area for further research is to forecast an imminent frequency anomaly, i.e., the event that the frequency will attain a value outside a range of pre-specified limits.

APPENDIX
SIMPLE RECURRENT NETWORK AND GATED RECURRENT UNIT ARCHITECTURES

We devote this appendix to the discussion of the simple recurrent network (*SRN*) model and the gated recurrent unit (*GRU*) model, which are leveraged in Section IV as benchmarks in the performance evaluation of the proposed *LSTM*-based forecasting methodology.

The *SRN* model is an *RNN* architecture equipped with a rudimentary recurrent connection. While the recurrent connection enables an *SRN* network to pass information over time, the *SRN* model lacks the gate structures inherent in more sophisticated *RNN* architectures, such as the *LSTM* model. Due to its primitive recurrent connection, the *SRN* model prefers more recently provided data over data provided earlier. As such, the gradients associated with data observed further in the past, as important as they may be, are liable to vanish quickly—a phenomenon known as the *vanishing gradient* problem [36]. We provide the graphical depiction of the *SRN* model in Fig. 11.

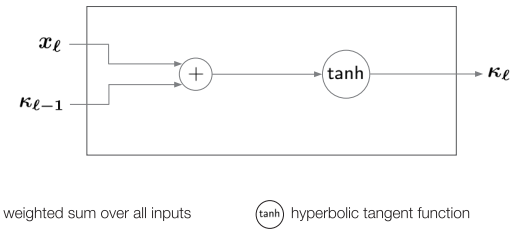


FIGURE 11. Graphical depiction of the *SRN* model.

We consider the input sequence $(\boldsymbol{x_\ell})_{\ell=1}^{\mathcal{T}} := (\boldsymbol{x_1}, \ldots, \boldsymbol{x_\ell}, \ldots, \boldsymbol{x_\mathcal{T}})$ of length $\mathcal{T}$. Each element of the sequence $(\boldsymbol{x_\ell})_{\ell=1}^{\mathcal{T}}$ is an $\mathcal{M}$-dimensional vector, where $\mathcal{M}$ is the number of input features. The forward pass of an *SRN* model can be

expressed by the following equation for $\ell = 1, \ldots, \mathcal{T}$:

$$\boldsymbol{\kappa}_\ell = \tanh\left(\boldsymbol{W}_{x\kappa}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa\kappa}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{b}_\kappa\right), \tag{7}$$

where the $\boldsymbol{W}$ terms represent the weight matrices, the $\boldsymbol{b}$ term the bias vector, $\boldsymbol{\kappa}_\ell$ the output in time step ℓ, and $\tanh(\cdot)$ the hyperbolic tangent function [37], [38]. Equation (7) illustrates that the output of the *SRN* model in time step ℓ is jointly influenced by the input in time step ℓ and the output in time step $\ell-1$, which permits the *SRN* model to carry information across time steps.

We next turn to the description of the *GRU* model, which is an *RNN* architecture that has more internal structural elements—and so has the capability to garner more underlying temporal information—than does the *SRN* model. Similar to the *LSTM* model, the *GRU* model is equipped with gates, *viz.*: update gate and reset gate, which control the flow of information among time steps and the integration of the input at each time step. Despite its less intricate internal structure compared with the *LSTM* model, the *GRU* model is widely touted as performing on par with the *LSTM* model [39]. We present the internal architecture of the *GRU* model in Fig. 12.

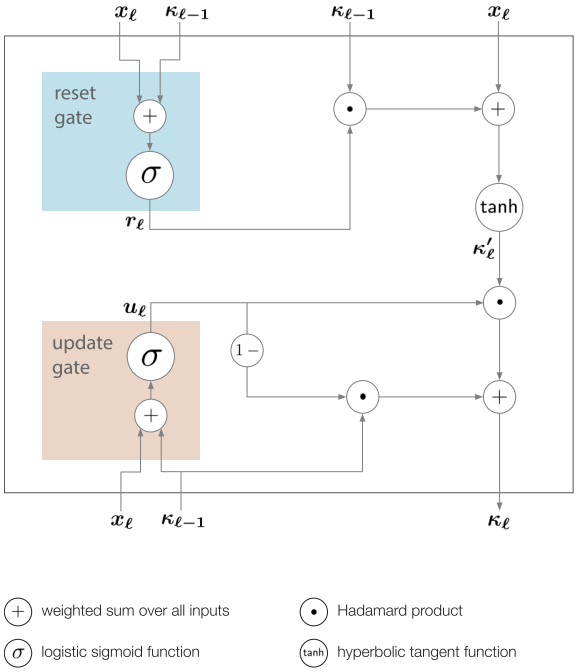


FIGURE 12. Graphical depiction of the *GRU* model.

We next analytically describe the internal operation of the *GRU* model. We denote the update gate and reset gate of the *GRU* model by $\boldsymbol{u}$ and $\boldsymbol{r}$, respectively. We denote by $(\boldsymbol{x}_\ell)_{\ell=1}^{\mathcal{T}}$ the input sequence and state the forward pass of a *GRU* network for $\ell = 1, \ldots, \mathcal{T}$ as follows:

$$\boldsymbol{u}_\ell = \sigma\left(\boldsymbol{W}_{xu}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa u}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{b}_u\right), \tag{8}$$

$$\boldsymbol{r}_\ell = \sigma\left(\boldsymbol{W}_{xr}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa r}\boldsymbol{\kappa}_{\ell-1} + \boldsymbol{b}_r\right), \tag{9}$$

$$\boldsymbol{\kappa'}_\ell = \tanh\left(\boldsymbol{W}_{x\kappa'}\boldsymbol{x}_\ell + \boldsymbol{W}_{\kappa\kappa'}(\boldsymbol{r}_\ell \odot \boldsymbol{\kappa}_{\ell-1}) + \boldsymbol{b}_{\kappa'}\right), \tag{10}$$

$$\boldsymbol{\kappa}_\ell = \left(\boldsymbol{u}_\ell \odot \boldsymbol{\kappa'}_\ell\right) + \left((1 - \boldsymbol{u}_\ell) \odot \boldsymbol{\kappa}_{\ell-1}\right), \tag{11}$$

where the $\boldsymbol{W}$ terms represent the weight matrices, the $\boldsymbol{b}$ terms the bias vectors, $\boldsymbol{\kappa}_\ell$ the output in time step ℓ, $\sigma(\cdot)$ the logistic sigmoid function, $\tanh(\cdot)$ the hyperbolic tangent function, and $\odot$ the Hadamard product [39], [40].

The update gate controls the integration of the information from time step $\ell-1$ and the input in time step ℓ, to the computations in time step ℓ, as expressed by (8). The hallmark of the update gate is its ability to explicitly ensure the propagation of information across time steps, thereby availing the *GRU* model to overcome the vanishing gradient problem. [40]. The operation of the reset gate is expressed by (9). The objective of the reset gate is to adjust the incorporation of information from time step $\ell-1$ while evaluating the input in time step ℓ.

ACKNOWLEGMENT

We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

REFERENCES

[1] (Dec. 2007). *Definition of 'Adequate Level of Reliability'*. [Online]. Available: https://www.nerc.com/docs/pc/Definition-of-ALR-approved-at-Dec-07-OC-PC-mtgs.pdf

[2] E. Ela, M. Milligan, and B. Kirby. (Aug. 2011). *Operating Reserves and Variable Generation*. [Online]. Available: https://www.nrel.gov/docs/fy11osti/51978.pdf

[3] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation and Control*, 3rd ed. Hoboken, NJ, USA: Wiley, 2014.

[4] *Future Requirements for Balancing Services*. Accessed: Jun. 17, 2020. [Online]. Available: https://www.nationalgrideso.com/document/88586/download

[5] (Jan. 2011). *Balancing and Frequency Control*. [Online]. Available: https://www.nerc.com/docs/oc/rs/NERC%20Balancing%20and%20Frequency%20Control%20040520111.pdf

[6] (Apr. 2004). *Frequency Response Standard Whitepaper*. [Online]. Available: https://www.nerc.com/pa/Stand/Project%20200712%20Frequency%20Response%20DL/Frequency_Response_White_Paper.pdf

[7] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[8] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78–111, Oct. 2016.

[9] M. K. Tikariha. *Frequency Forecasting Using Time Series Arima Model*. Accessed: Jun. 17, 2020. [Online]. Available: http://210.212.126.93:8000/ft/IPS/Session-10C/04.FREQUENCY%20FORECASTING.pdf

[10] D. Chourey, A. Kumar, H. Gupta, J. Kumar, A. Kumar, and A. Mishra, "An artificial neural network based approach for forecasting the frequency of power grid to achieve optimal frequency stabilization," *Int. J. Adv. Res. Electr., Electron. Instrum. Eng.*, vol. 6, no. 2, pp. 886–893, 2017.

[11] N. V. Tomin, "Intelligent monitoring and forecasting of the expected operating conditions of electric power system," in *Proc. 3rd Int. Youth Conf. Energetics (IYCE)*, Jul. 2011, pp. 1–8.

[12] J. W. Taylor and M. B. Roberts, "Forecasting frequency-corrected electricity demand to support frequency control," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1925–1932, May 2016.

[13] T. S. Salem, K. Kathuria, H. Ramampiaro, and H. Langseth, "Forecasting intra-hour imbalances in electric power systems," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 9595–9600.

[14] M. P. Garcia and D. S. Kirschen, "Forecasting system imbalance volumes in competitive electricity markets," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 240–248, Jan. 2006.

[15] C. Contreras, "System imbalance forecasting and short-term bidding strategy to minimize imbalance costs of transacting in the Spanish electricity market," M.S. thesis, Dept. Electr. Power Ind., Comillas Pontifical Univ., Madrid, Spain, 2016.

[16] O. Yurdakul, F. Eser, F. Sivrikaya, and S. Albayrak. (2020). *Short-Term Power System Frequency Forecasting*. [Online]. Available: https://github.com/afeser/Short-Term-Power-System-Frequency-Forecasting

[17] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[18] A. Graves, N. Jaitly, and A.-R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 273–278.

[19] R. Messina and J. Louradour, "Segmentation-free handwritten chinese text recognition with LSTM-RNN," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2015, pp. 171–175.

[20] D. Eck and J. Schmidhuber, "Finding temporal structure in music: Blues improvisation with LSTM recurrent networks," in *Proc. 12th IEEE Workshop Neural Netw. Signal Process.*, Sep. 2002, pp. 747–756.

[21] O. Yurdakul, "Analysis and performance evaluation of coordinated transaction scheduling," M.S. thesis, Dept. Elect. Comput. Eng., Univ. Illinois Urbana-Champaign, Urbana, IL, USA, 2018. [Online]. Available: http://hdl.handle.net/2142/101097

[22] *Mandatory Frequency Response*. Accessed: Jun. 17, 2020. [Online]. Available: https://www.nationalgrideso.com/document/92441/download

[23] (Dec. 2017). *Firm Frequency Response*. [Online]. Available: https://www.nationalgrid.com/sites/default/files/documents/Firm%20Frequency%20Response%20%28FFR%29%20Interactive%20Guidance%20v1%200_0.pdf

[24] North American Electric Reliability Corporation. (2019). *Electricity Supply & Demand*. [Online]. Available: https://www.nerc.com/pa/RAPA/ESD/Pages/default.aspx

[25] (2019). *2019 State of Reliability*. [Online]. Available: https://www.nerc.com/pa/RAPA/PA/Performance%20Analysis%20DL/NERC_SOR_2019.pdf

[26] H. D. Vu and J. C. Agee. (1998). *Wecc Tutorial on Speed Governors*. [Online]. Available: https://www.wecc.org/Reliability/Governor%20Tutorial.pdf

[27] (2018). *Power System Elements*. [Online]. Available: https://www.pjm.com/-/media/training/nerc-certifications/trans-exam-materials-2018/power-system-elements/lesson-1-system-loads-2018.ashx

[28] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.

[29] National Grid Electricity System Operator. (2020). *Historic Frequency Data*. [Online]. Available: https://www.nationalgrideso.com/balancing-services/frequency-response-services/historic-frequency-data

[30] National Grid Electricity System Operator. (2020). *Historic Demand Data*. [Online]. Available: https://www.nationalgrideso.com/balancing-data/data-finder-and-explorer#tab-1

[31] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: https://arxiv.org/abs/1603.04467

[32] F. Chollet. *Keras Documentation*. Accessed: 2015. [Online]. Available: Keras.io

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[34] *IEEE Standard for Synchrophasor Measurements for Power Systems*, IEEE Standard C37.118.1-2011 (Revision IEEE Std C37.118-2005), 2011, pp. 1–61.

[35] X. Li, Z. Liu, and Z. Huang, "Attention-based radar PRI modulation recognition with recurrent neural networks," *IEEE Access*, vol. 8, pp. 57426–57436, 2020.

[36] O. Orojo, J. Tepper, T. M. McGinnity, and M. Mahmud, "A multi-recurrent network for crude oil price prediction," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2019, pp. 2940–2945.

[37] J. L. Elman, "Finding structure in time," *Cognit. Sci.*, vol. 14, no. 2, pp. 179–211, Mar. 1990.

[38] H.-G. Kim, G.-J. Jang, H.-J. Choi, M. Kim, Y.-W. Kim, and J. Choi, "Recurrent neural networks with missing information imputation for medical examination data prediction," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2017, pp. 317–323.

[39] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.

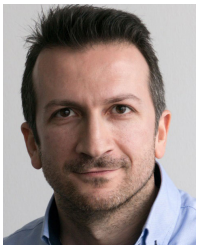
[40] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

OGUN YURDAKUL (Student Member, IEEE) received the Bachelor of Science degree in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, in 2016, and the Master of Science degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, IL, USA, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering and computer science with the Technical University of Berlin, Berlin, Germany.

He is currently the Head of the Energy Group, DAI Laboratory, Technical University of Berlin. Since 2018, he has been working as a Researcher and a Lecturer with the Technical University of Berlin. He was previously employed by the University of Illinois at Urbana–Champaign, as a Research Assistant and a Teaching Assistant. His research interests include power system operations, planning, and economics, forecasting and its applications in power systems, and optimization under uncertainty.

Mr. Yurdakul was a Fulbright Scholar during his studies at the University of Illinois at Urbana–Champaign.

FATIH ESER is currently pursuing the Bachelor of Science degree in electrical and electronic engineering and the Bachelor of Science degree in computer engineering with Middle East Technical University. He has worked as a Research Assistant with the German-Turkish Advanced Research Center for ICT, Berlin, Germany, with Erasmus+ Mobility Support, in summer 2019. His current research interests include power system security, machine learning, parallel processing, and computer vision.



FIKRET SIVRIKAYA received the bachelor's degree in computer engineering from Boğaziçi University, Istanbul, Turkey, in 2000, and the Ph.D. degree in computer science from the Rensselaer Polytechnic Institute (RPI), NY, USA, in 2007. He was a Research and a Teaching Assistant with RPI, during his Ph.D. studies. He has served as the Director of the Competence Center Network and Mobility, DAI Laboratory, TU Berlin, Berlin, Germany, from 2008 to 2015. Since 2016, he has been the Research Director of the German-Turkish Advanced Research Center for ICT (GT-ARC gemeinnützige GmbH), an affiliated Institute of TU Berlin. His research interests include future mobile networks, wireless communications, multi-hop ad-hoc networks, the Internet of Things, and smart cities.

SAHIN ALBAYRAK received the Ph.D. degree in computer science and the Habilitation degree from Technische Universität Berlin, Berlin, Germany, in 1992 and 2002, respectively.

He is currently a Professor and the Head of the Chair Agent Technologies in Business Applications and Telecommunication, TU Berlin, where he is also the Founder and the Head of the Distributed Artificial Intelligence Laboratory, currently employing about 120 researchers. He is also the Founder of the Deutsche Telekom Innovation Laboratories and the German-Turkish Advanced Research Centre for Information and Communication Technology and the Founding Director of the Connected Living Association. He serves as an Advisor with several government authorities and large-scale companies in both Germany and Turkey. His research interests include next-generation telecommunication services and infrastructures, service-centric architectures, service engineering, autonomous systems, agent-oriented modeling, agent architectures, agent programming languages, mobility-supporting services, supply chain management, smart home and smart cities, preventive health, and cybersecurity.