Recurrent Neural-Network-Based Maximum Frequency Deviation Prediction Using Probability Power Flow Dynamic Tool

SUNGYOON SONG1, (Member, IEEE), YOONGUN JUNG2, (Student Member, IEEE), CHANGHEE HAN2, (Student Member, IEEE), SEUNGMIN JUNG2, (Member, IEEE), MINHAN YOON4, (Member, IEEE), AND GILSOO JANG2, (Senior Member, IEEE)

1Energy ICT Convergence Research Department, Korea Institute of Energy Research, Daejeon 34129, South Korea
2Department of Electrical Engineering, Korea University, Seoul 02841, South Korea
3Department of Electrical Engineering, Hanbat National University, Daejeon 305-719, South Korea
4Department of Electrical Engineering, Kangwon University, Seoul 01897, South Korea

Corresponding author: Gilsoo Jang (gjang@korea.ac.kr)

This work was supported in part by the National Research Foundation under Grant 2017K1A4A3013579, and in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) under Grant 20191210301890.

ABSTRACT This paper proposes a recurrent neural network (RNN)-based maximum frequency deviation forecasting model for power systems with high photovoltaic power (PV) penetration. The proposed RNN model extracts the nonlinear features and invariant structures exhibited in regional PV power output data and time-variable frequency data in case of contingency. To capture the regularity and random characteristics of PV power output, a probability power flow-dynamic tool (PPDT) for uncertain power system modeling has been developed. This tool considers all possible combinations of PV power generation patterns, even those with low probability, such as those caused by passing clouds. The results are verified by a comparison of various artificial intelligence methods using case studies from the South Korean power system. An online dispatch algorithm that considers the frequency constraints for a designated contingency can be implemented by using the proposed model.

INDEX TERMS RNN, frequency stability, probability power flow, randomness.

NOMENCLATURE

\( Y \) PV power output vector
\( x, y, z \) Input vectors of irradiance, temperature, humidity
\( f_1 \) linear equation for PV power calculation
\( f_k(y) \) Probability density function
\( \mu \) Mean or expectation of the distribution
\( \mathcal{R} \) Kernel function
\( n \) Sample size
\( k \) Smoothing parameter
\( \sigma \) Standard deviation
\( A \) Regional correlation matrix
\( R \) Total area
\( w \) Coefficient values of matrix \( A \)
\( \hat{y}_l \) Total PV power of \( l \)th hour
\( x_i \) PV rated value in the \( i \)th area
\( b_{l,i} \) Regional PV power at \( l \)th hour in the \( i \)th area

\( h_t \) Hidden state in RNN model
\( U \) Weight matrix between input and hidden layers
\( W \) Weight matrix between hidden and hidden layers
\( V \) Weight matrix between output and hidden layers
\( b, c \) Bias vectors
\( x_l \) Current observation
\( a_l \) Temporary variable
\( \mathcal{L} \) Total cost of all time sequences
\( \eta \) Learning rate of the RNN
\( k_{l,m} \) Simulation value in a total N series of datasets
\( k_{l,m} \) Predicted value in a total N series of datasets
\( \hat{y}_l \) Frequency output of the \( \hat{y} \) in case of contingency
\( N \) The number of iterations in the PPDT
\( \hat{x} \) Normalized value between 0 and 1
\( x_{max} \) Maximum of the simulation data
\( x_{min} \) Minimum of the simulation data
\( R^2 \) Coefficient of determination
\( RMSE \) Root mean square error
\( MAPE \) Mean absolute percentage error

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
I. INTRODUCTION

A. MOTIVATION

Owing to the global push toward sustainable energy systems in recent years, much effort has been focused on the development of converter-interfaced renewable energy generators [1]. In particular, due to fundamental changes in the South Korean legal framework, there have been mandatory targets set by the Korean government for the amount of renewable energy, and it has been mandated that 50 GW of renewable energy resources must be installed by 2030 [2]. However, when many small nonsynchronous power generation units replace synchronous power generation units, the total rotational inertia of the synchronous generators decreases significantly.

This causes large frequency variations that can result in an unstable and insecure grid. For instance, a high rate of change of frequency (RoCoF) of 6 Hz/s was recorded during the South Australian blackout of September 28, 2016, which was caused by the “lightness” of the South Australian power system. In the UK, the Hornsea offshore wind farm and Little Barford gas power station suffered failures owing to a lightning strike, and the system automatically disconnected customers in the network. Approximately 5% of the electricity demand was turned off to protect the remaining 95% [3]. A maximum frequency deviation of 48.8 Hz was recorded on August 9, 2019. Compared to the blackout in the UK in 2008, the RoCoF increased as the total synchronous generator capacity decreased from 37.5 GW to 16.5 GW. As the penetration of variable renewable energy generation into power systems increases, planning for the effects of contingency events will become increasingly important. Thus, in Australia, successful frequency control in the Australian energy market requires that RoCoF should not exceed ±0.5 Hz/s, and the maximum allowed frequency deviations for contingency events is ±0.5 Hz [4].

From the perspective of the power system operator, frequency response forecasting for contingencies plays a crucial role in power systems with high renewable penetration. Using the forecasted values, the maximum frequency deviation can be calculated to dispatch schedules for generators or synchronous condensers through real-time or day-ahead bids [5]. Recently, power systems containing mostly generation units with long start-up times and low ramp rates have found it more difficult to integrate variable generation successfully [6]. Thus, a frequency response-forecasting model can be fully integrated into existing system planning and operation techniques. The proposed forecasting model in power systems can be used to manage rapid and large frequency deviations arising from sizeable faults and can ensure that planned systems are capable of managing an unexpected contingency event.

B. LITERATURE REVIEW

With regard to prediction techniques, more sophisticated solar photovoltaic (PV) prediction models have been investigated to evaluate various PV power uncertainties in energy management systems (EMS) [7] or to maximize the prices in an interday electricity market [8]. Based on earlier studies of the PV power prediction method, models can be classified into three categories: physical, statistical, and artificial intelligence. The statistical method extracts features from historical samples to predict future behavior via an error minimization. Artificial intelligence techniques can solve the problem of nonlinear function estimation; thus, they are excellent tools to forecast the renewable energy generation and load. With the rapid development of artificial intelligence algorithms and their excellent performance in many fields, they are proving superior to other models in dealing with nonlinear problems that include strong uncertainties [9].

With regard to intelligence algorithms, several authors proposed short-term PV power prediction models using simple artificial neural networks (ANNs) based on meteorological information [10]. Wang, S. et al. proposed a backpropagation neural network (BPNN) algorithm using solar radiation and the module temperature [11]. In [11], an improved adaptive BPNN prediction model was established to forecast PV power output. Some prediction techniques based on the Tree [12] or Random Forest methods [13], that predict the PV power output by adding an additional layer of randomness to bagging were also proposed. Accurate prediction results were reported by considering time-series features, as in the historical similar mining (HISIMI) model [14]. Finally, a recurrent neural network (RNN) model in which there is a fully-connected single- or multi-layered network with complex neurons, to capture PV power output sequence patterns [15], and a more developed version of RNN called the long short-term memory (LSTM)-based model and the gated recurrent unit (GRU) were proposed [16]. In recent days, a high-precision deep neural network model named PVPNet has been also proposed to forecast PV system output power [17]. The methodology behind the proposed model is based on deep neural networks, and the model is able to generate a 24-h probabilistic.

The RNN employed in this study is used as a tool for time-series frequency prediction. The proposed RNN can remember the network topology changes and changing frequency responses derived from regional PV power, because of having, a recurrent architecture and memory units. Previous studies proved that an RNN model could be used to approximate a highly nonlinear dynamic system and forecast the system state by establishing a nonlinear prediction model.

Nevertheless, there is still no published research that considers both probable combinations of PV power generation profiles and maximum frequency deviations. A mathematical model that uses the swing equation exhibits low accuracy because it contains several equivalent terms while research into data-driven-method-based frequency prediction has not been addressed owing to a dearth of event data. Focusing on simulation data from power system analysis programs such as PSS®E, PowerFactory, and PowerWorld, numerous iterations for power flow and dynamic simulations using uncertain
renewable power system modeling are required. Simulation by using representative values for loads, renewable generation, and network topology leads to only one operating condition result.

To overcome these issues, an automatic probability power flow-dynamic tool (PPDT) is developed in this study. The PPDT that considers uncertain factors such as loads and PV power is integrated with a Newton-Raphson power flow function on a server running Power Transmission System Planning Software (PSS®E). To calculate all reliable maximum frequency deviations in power systems, the most important consideration is to consider low-probability scenarios. Thus, in this study, all possible PV power generation cases are considered by the PPDT to create a reliable RNN-based prediction model. By using the PPDT, each RNN model has a different contingency scenario while including numerous operating points under variable PV power, as shown in Fig. 1.

FIGURE 1. Proposed frequency forecasting RNN model structure in the power system.

C. CONTRIBUTION
To overcome above issues, the problem statement and main contributions of this paper are as follows:

1) Voltage changes from regional PV power generation cause loads to deviate from the original power consuming status under a fault condition [18] and this influences frequency response in the dynamic simulation.

2) Thus, the PPDT written in Python script was developed and linked to a PSS®E server for simulating automatic probability power flow-dynamics.

3) A probable combination of PV power data in correlation with adjacent regions to capture the regularity and random characteristics is presented by the PPDT, which can consider the low probability of PV power interruptions that may occur with passing cloud covers.

4) For the first time, an RNN-based maximum frequency deviation prediction model is proposed to improve existing power system planning and operation techniques.

5) The RNN model is tailored fully to extract high-level nonlinear features hidden in the regional PV power output and time-domain-based frequency values.

The performance of the model is validated considering the planned PV power rated capacity in South Korea in 2030. The system frequency and PV power data in one day are generated by PPDT considering the probability characteristic. The results show that the proposed methodology can extract the invariant structures exhibited in a large quantity of PV output power data, and accurately predict the frequency response in the case of a contingency event. The application of such a model driven by the proposed RNN method provides higher certainty for decision-making. To perform a detailed analysis, the developed PPDT with PV power data is discussed in Section II. The general RNN and the RNN-based forecasting models are introduced in Section III. Section IV and V present a case study and discussion, respectively.

II. PROBABILITY POWER FLOW DYNAMIC TOOL FOR DATASET CONFIGURATION

A. OVERVIEW OF THE PROBABILITY POWER FLOW DYNAMIC TOOL
The main object of this research is to predict the maximum frequency deviation and response in a contingency event, i.e., to minimize the error between the simulation frequency and the predicted frequency response. The frequency deviation is determined by generator trip, load change, system inertia, and load damping by considering the block diagram of swing equation. Regarding load damping constant, voltage changes from probabilistic PV power generation cause loads to deviate from the original power consuming status under a fault condition and this influences frequency response in the dynamic simulation. Thus, we focus on the load damping constant which is a nonlinear relationship with the network bus voltage. Therefore, large quantities of power flow and dynamic simulation data are required to train and validate the proposed RNN model. However, performing the load flow and dynamic simulation for every possible combination of uncertainty factors is time-consuming and highly inefficient. For a network with a buses and b different amounts of PV generation for each bus, the total number of all possible arrangements is a to the b, i.e., a^b and deterministic power flow and dynamic calculations will be required [19].

To address this issue, the Python-script-based PPDT, which is linked to a PSS®E server, was developed as shown in Fig. 2. The program can create all probable combinations of PV power patterns (such as for passing clouds) since the system operator should prepare for all possible PV power generation scenarios. In addition, the PPDT provides simple and rapid calculations for users without requiring modifications to the power flow and dynamic equations because it provides an iteration function. The graphic user interface is implemented using the Python Tkinter library.

The PPDT that yields random outputs for randomly input PV power generation variables was implemented for uncertain power system modeling as analytical methods could lead to incorrect results. In the PPDT, the set of electrical equations is considered a black box, and every effort is made to determine how the inputs should be fed into this box to obtain the desired level of precision using a minimum number of possible states. However, the procedure for selecting the minimum number of states is not trivial [19]. To address this issue, a regional correlation matrix derived
by the probabilistic density function (PDF) was proposed. These methods can manage the regional correlations between PV power generation uncertainties and reduce the size of the input dataset. The scenario generation procedure using the PPDT is as follows:

1) Input historical day sample records from the $i$th regional site, and generate hourly PV and load data using a PDF and normal distribution, respectively.
2) Input the data of the regional maximum and minimum output power of PV based on the regional correlation matrix.
3) Generate random variables within the random range at each $i$th area and perform $N$ iterations.
4) Perform economic dispatch based on the net load and merit order of generators. To consider network congestion problems, a must-run-generator list can be implemented in the PPDT.
5) Solve for the power flow with random samples of PV power using the Newton-Raphson method.
6) Check the power flow convergence and save the results of $N$ power flow data (*.sav).
7) Solve the dynamic simulation with $N$ power flow data (*.sav) for the predefined contingency event and save the frequency results in Excel.

**B. CALCULATION OF REGIONAL CORRELATION MATRIX**

Note that wind power is considered as a constant value among contingency events. The normal distribution, which is a continuous probability distribution, was used for load modeling. This function provides the probability of observing any true boundaries that fall between two real numbers, as the curve approaches zero on both sides. The PDF of a normal distribution is given by [19]

$$f_n(s|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s-\mu)^2}{2\sigma^2}}. \quad (1)$$

where $\mu$ represents the mean or expectation of the distribution and $\sigma$ is the standard deviation of $s$. By considering the input form of the PPDT, historical daily curves of irradiance should be converted into active power. This can be expressed by

$$Y = f_i(x, y, z). \quad (2)$$

where $Y$ denotes the output PV power vector and $x$, $y$, and $z$ are the input vectors of irradiance, temperature, and humidity, respectively. $f_i$ is the linear equation for PV power calculations [20]. It was assumed that there are numerous data during the $i$th hour and a probability density analysis is required to capture the stochastic characteristics. Therefore, the samples of stochastic data representing the randomness of PV power are modeled by the nonparametric kernel density function. Let $Y_i$ denote $n$ samples of PV power at the $i$th hour; thus, the probability density function $f_k(y)$ can be estimated as follows:

$$f_k(y) = \frac{1}{nk} \sum_{k=1}^{n} R(y - \frac{Y_i}{h}). \quad (3)$$

where $R$ is the kernel function, which is a determinant of several types of kernel function and $k$ is a smoothing parameter called the bandwidth. $\sigma$ is the standard deviation of power samples and $n$ is the sample size. Accordingly, the probability distributions can be acquired from the historical samples, and they provide hourly PV power probability data. The probability density functions of these variables provide a useful means for simulating the frequency distributions of the power variable. The reason for not using a histogram is that it results in a larger random range owing to its characteristic discontinuity. The histogram has a disadvantage in that a discontinuity appears at the bin boundary and the probability density varies depending on the bin size; the kernel estimation method was used to overcome above issues. Furthermore, the increasing rate of the bad data has given rise to concerns regarding the inexact histogram model and individual bad data increases the power flow and dynamic calculation time. To avoid excessive demand of measured bad data and computation efforts, it is appropriate to use the continuous probability model such as kernel function.

To obtain a regional correlation matrix, the maximum values in the continuous probability density curves for the $i$th area at the $l$th hour are required. The coefficients of the regional correlation matrix $A$ for the $i$th region at the $l$th hour can be represented by

$$A = \left[w_{i,l}\right]_{R \times 24} = \left[\frac{\max\{f_{h,i}(p_i)\}}{\max\{\max\{f_{h,i}(p_i)\}\}}\right]_{R \times 24}, \quad (4)$$

where $R$ is the total area, and $\max\{f_{h,i}(p_i)\}$ is the maximum value of the probability density curves of PV power in the $i$th area at the $l$th hour. $\max\{\max\{f_{h,i}(p_i)\}\}$ is the maximum value among all kernel functions of region $R$. It was assumed that there is a region $R$ in a bulk power system and the $R \times 24$ coefficient values of $w$ can be derived. For example, $w_{1,9}$ and $w_{2,9}$ at 9:00 a.m. can be calculated as shown in Fig. 3.
By using the matrix coefficients the regional PV output power can be calculated by
\[
\hat{y}_i = w_{i,l} x_i^T = \begin{bmatrix} w_{i,l}, \cdots, w_{R,l} \end{bmatrix} \begin{bmatrix} x_i \\ \vdots \\ x_R \end{bmatrix}, \tag{5}
\]
\[
b_{i,l} = w_{i,l} x_i^T. \tag{6}
\]
where \( \hat{y}_i \) is the total PV power in the grid at the \( l \)th hour and \( x_i \) is the PV rated value in the \( i \)th area. To select the minimum number of states in the PPDT, the correlation index \( w_{i,l} \) is multiplied by PV rated \( x_i \) to limit the random range. Thus, the regional PV power \( b_{i,l} \) is randomly generated within the range of 0 to \( w_{i,l} x_i^T \) in the \( i \)th region at the \( l \)th hour. The return value \( b_{i,l} \) is limited in the range of the kernel function and correlation matrix. By using a random variable, \( N \) scenarios produced from the PPDT and \( N \) power flow and dynamic simulations to acquire the frequency time-variable data are performed by tripping the two largest generation units, as shown in step 6 of Fig. 2. Note that \( b_i \) has a range of \( b_i \) to \( b_i + \sigma \) during the simulation timeframe; thus, we can configure the RNN model with datasets derived by the PPDT and the return frequency data is calculated by capturing the inherent daily regularity and randomness of the regional PV power supply.

III. PROPOSED RECURRENT NEURAL NETWORK MODEL

A. RNN MODEL [22]

By using the \( N \) series of PV power and frequency data, the RNN model can be configured. Fig. 4 shows the general structure of the RNN model. This model uses a dynamic system where the hidden state \( h_t \) is not only dependent on the current observation \( x_t \) but also relies on the previous hidden state \( h_{t-1} \). The modeling procedure is similar to the approach used in [22], which is briefly described here. From Fig. 4, we can represent
\[
h_t = f(Ux_t + b) + Wh_{t-1}, \tag{7}
\]
\[
o_t = Vh_t + c, \tag{8}
\]
\[
y_t = f(o_t). \tag{9}
\]
where \( U \) is the weight matrix between the input and hidden layers, \( W \) is the weight matrix between the hidden and hidden layers, \( V \) is the weight matrix between the output and the hidden layers, and \( h_t \) is the hidden layer state at step \( t \) that serves as a memory function. \( b \) and \( c \) are bias vectors, and \( o_t \) is a temporary variable. \( f = \text{ReLU} \) is the activation function of the hidden and output layers. Among many popular activation functions, we adopted the rectified linear unit (ReLU) for all nodes in the hidden and output layers. ReLU is known for providing enhanced training performance and preventing gradient vanishing and radiant explosion.

To learn the parameters \( U, V, W, b, \) and \( c \), the backpropagation through time (BPTT) approach, which is applied to the sequence data, is generally used. BPTT begins by unfolding a recurrent neural network in time [11]. The unfolded network contains inputs and outputs, but every copy of the network shares the same parameters \((U, V, \) and \( W)\). Then, the backpropagation algorithm is used to find the gradient of the cost with respect to all network parameters. The cost function of the RNN can be represented by
\[
\mathcal{L} = \sum_{t=1}^{T} \left( \frac{1}{2} \sum_{m=1}^{N} \left( \hat{k}_{t,m} - k_{t,m} \right)^2 \right), \tag{10}
\]
where \( \mathcal{L} \) is the total cost of all time sequences and \( \hat{k}_{t,m} \) and \( k_{t,m} \) are the simulation value and predicted value in a total \( N \) series of datasets, respectively. The total cost is the sum of the subcosts at each time step. The hidden state gradient of step \( t \) is defined by
\[
\sigma_t = \frac{\partial \mathcal{L}}{\partial h_t}, \tag{11}
\]
From the RNN model, it can be seen that \( \sigma_t \) is determined by the subcosts at current steps \( t \) and \( t + 1 \). Thus, \( \sigma_t \) is related to the output temporary variable \( o_t \) and the hidden layer state \( s_t + 1 \). According to the chain rules, \( \sigma_t \) can be represented by
\[
\sigma_t = \frac{\partial \mathcal{L}}{\partial o_t} \frac{\partial o_t}{\partial h_t} + \frac{\partial \mathcal{L}}{\partial h_{t+1}} \frac{\partial h_{t+1}}{\partial h_t}, \tag{12}
\]
The gradient of network parameters at step \( t \) are calculated systematically by backpropagation. Then, the gradients of \( U, V, W, b, \) and \( c \) can be expressed by the following
formulas:
\[
\begin{align*}
\frac{\partial L}{\partial c} &= \sum_{i=1}^{T} \left( \frac{\partial L}{\partial o_t} \frac{\partial o_t}{\partial c} \right), \\
\frac{\partial L}{\partial V} &= \sum_{i=1}^{T} \left( \frac{\partial L}{\partial o_t} \frac{\partial o_t}{\partial V} \right), \\
\frac{\partial L}{\partial b} &= \sum_{i=1}^{T} \left( \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial b} \right), \\
\frac{\partial L}{\partial W} &= \sum_{i=1}^{T} \left( \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial W} \right), \\
\frac{\partial L}{\partial U} &= \sum_{i=1}^{T} \left( \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial U} \right).
\end{align*}
\]

The final gradients of the network parameters are the sum of the subgradients at each time step. Therefore, the updated rules for these parameters are as follows:
\[
\begin{align*}
b_{t+1} &= b_t - \eta \frac{\partial L}{\partial b}, \\
c_{t+1} &= c_t - \eta \frac{\partial L}{\partial c}, \\
V_{t+1} &= V_t - \eta \frac{\partial L}{\partial V}, \\
W_{t+1} &= W_t - \eta \frac{\partial L}{\partial W}, \\
U_{t+1} &= U_t - \eta \frac{\partial L}{\partial U}.
\end{align*}
\]
where $\eta$ is the learning rate of the RNN and superscript $\eta$ stands for the BPTT iteration times. The partial derivatives of the cost function with respect to the disturbances of $U, V, W, b,$ and $c$ can be inferred. Whenever the proposed model processes the given training instances that involve both the PV power information and the simulated frequency output for a particular hour, its weights are gradually updated through an iterative learning process.

**B. DATASET CONFIGURATION**

To configure the RNN model, the training data are generally divided into time series and ordered data structures. The time-series data change over time and remain consistent in adjacent clips such as timeframes for frequency or voltage variations in case of a contingency event. However, the variation of voltage is more dependent on the local generator or static synchronous compensator status and not the frequency data. A large number of generators or other reactive power resources that actively participate in voltage regulation are closely correlated to the voltage. Therefore, it is difficult to find a close correlation between the time-series-based voltage data and frequency responses. This reduces the accuracy of the RNN model. In addition, the line flow does not correlate with the frequency. Thus, the time-domain frequency and PV power generation data are used in time-series data. If necessary, any other input variables in a power system can be considered in future work.

For the ordered data, the inertial changes are implemented as listed in Table 1. The expensive generators located in the load center are generally turned off, based on the economic dispatch. Thus, inertia is changed by the PV out power. This kind of dataset improves the accuracy of the forecasting model. In short, we use seven input datasets to forecast the maximum frequency deviation and frequency response.

**TABLE 1. Input data for RNN.**

| Item                      | Data num. |
|---------------------------|-----------|
| Time-series data          | Frequency (Hz) 1,000,000 |
| Ordered data              | System Inertia (H) 1,000 |
| Total PV power in Metro (MW) | [b1,1 to b1,1 + σ] 1,000,000 |
| Total PV power in Yangnam (MW) | [b2,1 to b2,1 + σ] 1,000,000 |
| Total PV power in Changnam (MW) | [b3,1 to b3,1 + σ] 1,000,000 |
| Total PV power in Honam (MW) | [b4,1 to b4,1 + σ] 1,000,000 |
| Total PV power in Gangwon (MW) | [b5,1 to b5,1 + σ] 1,000,000 |

The total amount of time per step is 10 s (total simulation time) $\div 0.01$ (simulation time step) = 1,000. A total of 1,000 frequency datasets is generated for one contingency. Thus, a dynamic simulation of $N = 1,000$ generates 1,000,000 frequency datasets. To obtain the RNN model, datasets can be represented by the following:
\[
x_t = [b_{i,1}(t + 1), \cdots, b_{R,1}(t + 1)],
\]
\[
x_{t+1} = [b_{i,1}(t-a), \cdots, b_{i,1}(t)],
\]
\[
\hat{y}_i = \sum_{i=0}^{R} b_{i,t}, \quad y_{t+1} = F_{\hat{y}_t} (t + 1).
\]
where $b_{R,1}$ is the PV output power of region $R$ based on (6) and $b_{i,1}(t)$ is the PV power at time $t$. $\hat{y}_t$ is the total PV output power at $t$th hour and $y_{t+1}$ is the frequency output of the $\hat{y}_t$ in case of contingency. The PV power output variable is $b_i = rand(0, w_i x_i^T)$; thus, $b_i$ has a range from $b_i$ to $b_i+\sigma$ during the simulation period because PV power continues to change due to environmental factors. $N$ represents the number of iterations that are chosen in the PPDT; thus, we have $N$ series of the regional PV output and frequency data and $N$ inertia data. For our purposes, $\hat{y}_i (t + 1)$ is the predicted frequency for the total PV power outputs and $y_{t+1}$ is the expected output, as shown in Fig. 5.

Based on the RNN model, we can explore a more accurate prediction performance by considering the PV power data of adjacent areas. In (23) to (24), the trend information hidden in the PV power for adjacent areas can be extracted by the RNN. This will be the input of the next step for the frequency response. In (25), the predicted frequency $\hat{y}_i (t + 1)$ is also related to $b_{i,1}(t)$ and $b_{R,1}(t + 1)$ in the forecasting model.

Note that the unit of input training data derived by the PPDT is different. The range of the regional PV power is from...
zero to thousands and the range of input data of the system inertia is from zero to tens. Since the RNN is a nonlinear model, it easily causes a gradient explosion when dealing with large-value data. Thus, unrefined input data will also reduce the learning efficiency of the RNN. A classic method to solve this problem is data normalization and the power data can be limited to a normalized value between 0 and 1 to reduce the regression error. The mathematical formula of min-max scaling is as follows:

\[
\hat{x} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}. \tag{26}
\]

where \(x\) is the original data and \(x_{\text{max}}\) and \(x_{\text{min}}\) are the maximum and minimum of the simulation data.

**C. FREQUENCY FORECASTING MODEL**

Owing to the complex nature of differential equations for the system dynamics, the maximum frequency deviation data depend on various factors including PV power, economic dispatch results, and system size. The proposed RNN-based predictor can consider all probable combinations of PV output power correlated to adjacent regions because the model has multiple input steps and the current hidden layer input step includes the state of the previous step’s hidden layer. This means that short-term memory attributes based on the RNN can better mine the potential links in the time-domain frequency response and improve the accuracy of the maximum frequency deviation prediction. A final flowchart for the proposed forecasting model is shown in Fig. 6 and our approach includes three steps, as follows:

1) The suggested multilayer RNN-based model with multiple hidden layers and an increased number of parameters is an attempt to estimate more accurate frequency outputs than the ANN-based model.

2) A performance difference occurs depending on the number of hidden layers and after repeated experiments, we chose the four layers with the best performance.

3) Following the suggestions in [23], through many experimental trials, we determined the epochs that minimized the over- and underfitting problems while maximizing the generalization ability of the proposed models in terms of the loss value obtained.

Note that the LSTM model as trying to learn long and short-term relationships occurs an error, which is not particularly desirable for this forecasting model because the frequency dataset is only depend upon the short-term relationship. Our proposed model has a good efficiency in a short-term based multi-step prediction. Moreover, the values of \(R^2\) (coefficient of determination) deteriorate when the forecasting horizon increases. However, the maximum time of frequency response in transient state is 10 s. So, the proposed model is regardless with the horizon issue. In conclusion, the proposed forecasting model with its recurrent architecture and memory units can effectively learn the frequency changes. And, the PV output power data vs. frequency was divided into a training dataset and a testing dataset. The training dataset and testing dataset were normalized and then arranged in several input sequences. A multi horizon forecasting model based on the RNN was established and trained using the BPTT algorithm to predict the frequency response as accurately as possible.

These models are similar to those mentioned in previous studies [15], [16], and [22], but the RNN-based model proposed in this study has a significant difference: it attempts to estimate the frequency response depending upon the regional PV output power pattern without performing inefficient deterministic simulations. By using the proposed model, the experimental results using a simulated PV power output and frequency dataset showed that the proposed models could successfully predict the frequency response when compared to conventional prediction models.
IV. SIMULATION RESULT

The characteristics of the output response of the frequency drop are nonlinear and unpredictable. These characteristics can cause many problems in power systems and affect power quality, generation control, overload, and protection such as with an underfrequency relay. Thus, all possible combinations of output patterns of PV power generators should be analyzed. In this study, the regional PV power data used were collected by Korea Electric Power Corporation (KEPCO) of South Korea. The planned PV power rated capacity and maximum PV power in 2030 are considered in this scenario, as shown in Table 2. To obtain a regional correlation matrix, the probability density functions of the same PV rated systems in five areas are represented in Fig. 7. Thus, the input random range of the five areas is restricted by regional correlation coefficients and PDF.

TABLE 2. Maximum PV power at 2:00 p.m.

| Area     | Planned Rated PV power in 2030 | Coefficient $w_i$ of matrix $A$ | Maximum PV power at 9:00 a.m. |
|----------|-------------------------------|---------------------------------|-------------------------------|
| Metro    | 1.0 GW                        | 0.91                            | 0.91 GW                       |
| Yungnam  | 3.8 GW                        | 0.94                            | 3.572 GW                      |
| Chungnam | 1.0 GW                        | 1                               | 1.0 GW                        |
| Honam    | 5.0 GW                        | 0.87                            | 4.35 GW                       |
| Gangwon  | 1.8 GW                        | 0.78                            | 1.404 GW                      |

For the maximum frequency calculation, $N = 1,000$ dynamic simulations are performed by tripping the two largest units of the generator (approximately 3.0-GW generation). The contingency event is shown in Fig. 8. In addition, a probable combination of the PV power patterns for the five regions is implemented to train the RNN network, as shown in Fig. 9. The testing dataset is used to evaluate the prediction performance of the proposed method.

For the extracted combination of the PV power patterns, the frequency forecasting results of the proposed method for a total of 200 cases and randomly selected cases 1 and 2 and case 3 are shown in Fig. 10 and 11, respectively. It was assumed that the PV output power in each area could be considerably different because of the transient overcast event. In the simulation, four kinds of forecasting models [24] were chosen as benchmarks to illustrate the improvement of the proposed forecasting model in terms of forecasting quality.

In each case, the forecasting model was adapted to each period through independent training and testing. Fig. 10 depicts how well the proposed method produced robust performance by the linear regression of the 200 total cases and we can conclude that the result of proposed model is the most accurate because the data points are well concentrated along the linear line. The Linear Regression method also exhibits high accuracy because of the output curve characteristic. In addition, the real and predicted frequency curves are indicated as blue and red lines, respectively, as shown in Fig. 11. Tests and the proposed method provide simple and intuitive frequency forecasting results based on the amount of regional PV power generation. Note that the Korean power system has a large operating power reserve; the frequency response is close to being linear because of the initial governor droop control.
To evaluate the performance, the widely used indices of root mean square error (RMSE) and the mean absolute percentage error (MAPE) are implemented as follows:

\[
\text{MAPE} = \frac{1}{N} \sum_{i=m}^{N} \left| \frac{k_m - \bar{k}_m}{k_{\text{mean}}} \right| \times 100\%, \tag{27}
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{k}_m - k_m)^2}. \tag{28}
\]

where \(k_{\text{mean}}\) is the average of the measured frequency data.

The indices confirm that the proposed method is superior to the other benchmark methods for all cases. Compared to earlier works, on average both the MAE and RMSE indices were improved, as listed in Table 3. Thus, our proposed model is efficient for multistep prediction.

Moreover, the values of \(R^2\) (coefficient of determination) deteriorate when the forecasting horizon increases. However, the frequency response time in the transient state is maximally 10 s. Thus, the proposed model is unaffected by the horizon issue. Unlike with traditional forecasting methods, we feed the PV power of adjacent areas and the time-series frequency data into the RNN model, where the similarity and correlation are verified against actual frequency data. The proposed forecasting model with its recurrent architecture and memory units can effectively learn the frequency changes. In conclusion, by exploiting short-term patterns hidden in the past frequency time variable values performance by the proposed RNN-based model is significantly improved compared to all other models considered. The results demonstrate that the MAPE and RMSE of the proposed method outperform the benchmarks within the 10-s forecasting horizon.

To address this issue, a probabilistic power system analysis that considers the uncertain generation of PV power supply is required. To capture the regularity and random characteristics of PV power output, a probability power flow-dynamic tool (PPDT) for uncertain power system modeling has been developed. By using the PPDT, this paper proposes a recurrent neural network (RNN)-based maximum frequency deviation forecasting model for power systems with high PV power penetration. The proposed RNN model extracts the nonlinear features and invariant structures.
exhibited in regional solar photovoltaic (PV) power output data and time-variable frequency data in case of contingency.

VI. DISCUSSION

Earlier works such as k-means clustering, two-point estimation, Harr and Hong’s point estimate, multilinearization methods, and Fourier transform methods required some approximation or neglected some aspect of probability to reduce the calculation time. In this study, a tool developed to further consider the low probability of PV power generation acquired all probable combinations of PV power data. The proposed RNN model can be utilized in existing system planning and operational techniques, as shown in Fig. 12. In the flowchart of the Australia economic dispatch algorithm example, a feedback loop was introduced to evaluate the synchronous inertia adequacy of a dispatch result via the rapid dynamic simulation of an N-1 contingency event. If the post-contingency frequency response of a PV power supply cannot satisfy the constraints of RoCoF and frequency deviation, then the dispatch algorithm will find a new solution until the simulated frequency response is acceptable [25].

In general, the system operator uses a 1-min or 5-min central dispatch cycle consisting of generation and dispatches to balance instantaneously the generation and load in the most cost-effective manner, as shown in Fig. 13. However, a dynamic simulation requires significant computational time to solve the state-space matrix when the power system is large. Thus, system operators generally use the equivalent frequency calculation model for rapid N-1 dynamic simulation. The proposed RNN model can be utilized in this procedure instead of using a rapid N-1 dynamic simulation in renewable power systems. Thus, the 1-min dispatch algorithm that considers frequency constraints for a designated contingency can also be implemented.

Furthermore, the algorithm can be fully utilized in a microgrid system, because real-time microgrid operation via model predictive control has been realized [26]. The microgrid operation schedule is determined by predictions for future PV power generation and power consumption. The proposed model can reduce the calculation time and generate more accurate frequency results for more precise central dispatching to be achieved.

REFERENCES

[1] J. I. Lewis and R. H. Wise, “Fostering a renewable energy technology industry: An international comparison of wind industry policy support mechanisms,” Energy Policy, vol. 35, no. 3, pp. 1844–1857, Mar. 2007.
[2] S. Kim, H. Lee, H. Kim, D.-H. Jang, H.-J. Kim, J. Hur, Y.-S. Cho, and K. Hur, “Improvement in policy and proactive interconnection procedure for renewable energy expansion in South Korea,” Renew. Sustain. Energy Rev., vol. 98, pp. 150–162, Dec. 2018.
[3] M. J. Bryant, R. Ghanbari, M. Jali, P. Sokolowski, and L. Meegahapola, “Frequency control challenges in power systems with high renewable power generation: An Australian perspective,” RMIT Univ., Melbourne, VIC, Australia, Tech. Rep. 1, 2019.
[4] R. Yan, Nahid-Al-Masood, T. K. Saha, F. Bai, and H. Gu, “The anatomy of the 2016 South Australia blackout: A catastrophic event in a high renewable network,” IEEE Trans. Power Syst., vol. 33, no. 5, pp. 5374–5388, Sep. 2018.
[5] H. Gu, R. Yan, and T. K. Saha, “Minimum synchronous inertia requirement of renewable power systems,” IEEE Trans. Power Syst., vol. 33, no. 2, pp. 1533–1543, Mar. 2018.
[6] E. Lannoye, D. Flynn, and M. O’Malley, “Evaluation of power system flexibility,” IEEE Trans. Power Syst., vol. 27, no. 2, pp. 922–931, May 2012.
[7] C. L. Nge, I. U. Ranaweera, O.-M. Midtgård, and L. Norum, “A real-time energy management system for smart grid integrated photovoltaic generation with battery storage,” Renew. Energy, vol. 130, pp. 774–785, Jan. 2019.
[8] W. Meng, X. Lin, and M. Pedram, “A Stackelberg game-based optimization framework of the smart grid with distributed PV power generations and data centers,” IEEE Trans. Energy Convers., vol. 29, no. 4, pp. 978–987, Dec. 2014.
[9] W.-M. Lin, F.-S. Cheng, and M.-T. Tsay, “Nonconvex economic dispatch by integrated artificial intelligence,” IEEE Trans. Power Syst., vol. 16, no. 2, pp. 307–311, May 2001.
[10] I. Drezga and S. Rahman, “Input variable selection for ANN-based short-term load forecasting,” IEEE Trans. Power Syst., vol. 13, no. 4, pp. 1238–1244, Nov. 1998.
[11] S. Wang, “The unpredictability of standard back propagation neural networks in classification applications,” Manage. Sci., vol. 41, no. 3, pp. 555–559, Mar. 1995.
[12] M. W. Ahmad, M. Mourshed, and Y. Riezgui, “Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression,” Energy, vol. 164, pp. 465–474, Dec. 2018.
[13] P. H. Chiang, S. P. V. Chiluvuri, S. Dey, and T. Q. Nguyen, “Forecasting of solar photovoltaic system power generation using wavelet decomposition and bias-compensated random forest,” presented at the 9th Annu. IEEE Green Technol. Conf. (GreenTech), 2017, pp. 260–266.
[14] C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte, “Short-term power forecasting model for photovoltaic plants based on historical similarity,” Energies, vol. 6, no. 5, pp. 2624–2643, May 2013.
[15] D. Lee and K. Kim, “Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information,” Energies, vol. 12, no. 2, p. 215, Jan. 2019.
[16] M. Abdel-Nasser and K. Mahmoud, “Accurate photovoltaic power forecasting models using deep LSTM-RNN,” Neural Comput. Appl., vol. 31, no. 7, pp. 2727–2740, Jul. 2019.
[17] C.-J. Huang and P.-H. Kuo, “Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting,” IEEE Access, vol. 7, pp. 74822–74834, 2019.

[18] K. Hatipoglu, I. Fidan, and G. Radman, “Investigating effect of voltage changes on static zip load model in a microgrid environment,” in Proc. North Amer. Power Symp. (NAPS), Sep. 2012, pp. 1–5.

[19] S. Song, C. Han, S. Jung, M. Yoon, and G. Jang, “Probabilistic power flow analysis of bulk power system for practical grid planning application,” IEEE Access, vol. 7, pp. 45494–45503, 2019.

[20] A. Selim, S. Kamel, and F. Jurado, “Efficient optimization technique for multiple DG allocation in distribution networks,” Appl. Soft Comput., vol. 86, Jan. 2020, Art. no. 105938.

[21] R. N. Allan, A. M. Da Silva, and R. C. Burchett, “Evaluation methods and accuracy in probabilistic load flow solutions,” IEEE Trans. Power App. Syst., vol. PAS-100, no. 5, pp. 2539–2546, May 1981.

[22] G. Li, H. Wang, S. Zhang, J. Xin, and H. Liu, “Recurrent neural networks based photovoltaic power forecasting approach,” Energies, vol. 12, no. 13, p. 2538, Jul. 2019.

[23] M. Li, T. Zhang, Y. Chen, and A. J. Smola, “Efficient mini-batch training for stochastic optimization,” presented at the 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2014, pp. 661–670.

[24] K. Uludağ and Ö. Korçak, “Energy and rate modeling of data download over LTE with respect to received signal characteristics,” presented at the 27th Int. Telecommun. Netw. Appl. Conf. (ITNAC), Nov. 2017, pp. 1–6.

[25] M. J. Bryant, R. Ghanbari, M. Jalili, P. Sokolowski, and L. Meegahapola, “Frequency control challenges in power systems with high renewable power generation: An Australian perspective,” Tech. Rep.

[26] Y. Du, W. Pei, N. Chen, X. Ge, and H. Xiao, “Real-time microgrid economic dispatch based on model predictive control strategy,” J. Mod. Power Syst. Clean Energy, vol. 5, no. 5, pp. 787–796, Sep. 2017.

SUNGYOON SONG (Member, IEEE) received the B.S. degree in electrical engineering from Soongsil University and the Ph.D. degree from the Department of Electrical Engineering, Korea University, Seoul, South Korea. He is currently a Senior Engineer with the Korea Institute of Energy Research (KIER). His research interests include power system operation and control including voltage source converter and renewable energy resources.

YOONGUN JUNG (Student Member, IEEE) is currently pursuing the B.S. degree in electrical engineering with Korea University, Seoul, South Korea. His research interests include deep learning application on power systems dynamics and control.

CHANGHEE HAN (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2015 and 2017, respectively, where he is currently pursuing the Ph.D. degree. His research interests include DC application on power system and distribution system operation.

SEUNGMIN JUNG (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea. He worked at the School of Electrical Engineering, Korea University, as a Research Professor, for seven months. Since 2017, he has been with the Department of Electrical Engineering, Hanbat National University, Daejeon, South Korea, where he is an Associate Professor. His research interests include renewable energy resources and energy management system.

MINHAN YOON (Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, Korea University, Seoul, South Korea. He was a Postdoctoral Research Associate with Seoul National University, Seoul, in 2015, and a Senior Engineer with the Korea Electrotechnology Research Institute (KERI), from 2015 to 2017. In 2020, he joined the Department of Electrical Engineering, Kwangwoon University, Seoul, as an Assistant Professor. His research interests include power system operation and control including HVDC, FACTS, and DERs.

GILSOO JANG (Senior Member, IEEE) received the B.S. and M.S. degrees from Korea University, South Korea, and the Ph.D. degree from Iowa State University, USA, in 1997. He worked at the Electrical and Computer Engineering Department, Iowa State University, as a Visiting Scientist for one year, and as a Researcher with the Korea Electric Power Research Institute for two years. He is currently a Professor with the School of Electrical Engineering, Korea University. His research interests include power system dynamics and control.

***