KEPCO Journal on Electric Power and Energy
p-ISSN 2465-8111, e-ISSN 2466-0124
Volume 8, Number 1, June 2022, pp. 21-29
DOI 10.18770/KEPCO.2022.08.01.21

Renewable Power Generation Forecasting Method for Distribution System: A Review

배전시스템 운영계획을 위한 신재생에너지원 발전량 예측 방법

Jintae Cho, Hongjoo Kim, Hosung Ryu, Youngpyo Cho
조진태, 김홍주, 류호성, 조영표

Abstract

Power generated from renewable energy has continuously increased recently. As the distributed generation begins to interconnect in the distribution system, an accurate generation forecasting has become important in efficient distribution planning. This paper explained method and current state of distributed power generation forecasting models. This paper presented selecting input and output variables for the forecasting model. In addition, this paper analyzed input variables and forecasting models that can use as mid-to long-term distributed power generation forecasting.

Keywords: Distribution power system planning, Distributed power generation forecasting, Hybrid forecasting model

I. INTORDUCTION

As renewable energy supply expands in Korea, an accurate generation forecasting has become important in stable system management and efficient planning. The growing proportion of energy production from photovoltaics and wind turbines is an increasing concern about uncertain energy generation. Therefore, the power system must manage this uncertainty to maintain customer service levels. The International Energy Agency (IEA) specifies renewable energy proportions of 3 to 15% as Phase 2 and recommends improving individual photovoltaic and wind turbine generation forecasting by constructing an enhanced forecasting system [1].

Korean photovoltaic and wind turbine systems are producing about 3.6% of energy needs as of 2020, which qualifies as Phase 2. Power generated from renewable energy has continuously increased recently. This year, it increased by 26% compared to last year. Therefore, to keep pace with this growth, there is an increasing need to construct the forecasting system for the individual photovoltaic and wind turbine. Specifically, as the distributed generation connected with customers at the end of the distribution system, issues began to occur regarding the management of the gap with the transmission system. In addition, the distribution system needs to have capacity for electric vehicles (EV) and energy storage systems (ESS). This issue highlights the need for research in distribution planning for tomorrow's supply and demand challenges.

However, since the existing distribution planning was created when the proportion of distributed power was not significant, and the equipment's planned capacity assumed continuous increase loads in the distribution line, it could not manage the growth of distributed energy generation. Therefore, it is necessary to control of the interconnection of distributed generation in the power supply. Small generation such as household distributed power system is not be monitored, so mid-to long-term forecasting models must be prepared to accurately manage the amount of power in distribution lines from distributed generation. Highly accurate distributed generation forecasting will make it possible to establish distribution planning that considers the linkage of renewable energy sources and forecasts loads, which will prevent overinvestment in unneeded distribution facilities. It will also be possible to construct more reliable grid by applying the scenarios generated by the forecasting model to obtain the stability of the distribution network.

This study configured in the following order. First, the distributed generation forecasting models and input variables were analyzed. Next, the mid-to long-term distributed generation models for the distribution plan were analyzed, and the pros and cons were discussed. Lastly, the hybrid forecasting models were described.

II. DATA AND MODEL PROPERTY FOR FORECASTING THE DISTRIBUTED GENERATION

Distributed generation refers to small-scale generation equipment that can be distributed and operated around a power consumption area, unlike large-scale concentration power sources. Generally, this term refers to the generating equipment using new and renewable

Article Information

Manuscript Received September 6, 2021, Accepted April 28, 2022, Published online December 30, 2020

The authors are with KEPCO Research Institute, Korea Electric Power Corporation, 105 Munji-ro Yuseong-gu, Daejeon 34056, Republic of Korea.

Correspondence Author: Jintae Cho (jintae.cho@kepco.co.kr)

energy. Representative new energies include tidal energy, fuel cell, coal liquefaction, and hydrogen energy, and renewable energy sources are photovoltaics, wind turbine, hydropower, waste, and bioenergy.

When carrying out forecasting based on power generation and generating equipment capacity data for renewable energy, it assumed that the rate of growth of renewable energy generation and equipment capacity is constant. However, since the capacity of renewable energy varies for each area, the predicted and the actual measured quantity are not equal when the forecasting carried out with applying the increasing rate to each area. Therefore, the model must learn the increasing rate and patterns specific to each area. In addition, to enhance the accuracy of the forecasting model needs to analyze and reflect identified variables in mid-to long-term forecasts.

The application scheme and forecasting method vary depending on the time cycle, factors determining the result, the type of data pattern, and other various aspects. Previous power generation and meteorological data are important factors in the model for predicting power generation with significant variation. Therefore, it is important to find the events causing the changes or identify the factors responsible for changing the pattern.

The core process in the forecasting is to identify outliers and biases that should not be learned from historical data and obtain more accurate patterns. Normally, it assumed that an amount of change in numbers is constant. However, data that is not refined or preprocessed often have volatile changes. Therefore, finding the cause and method of changing the values and examining simple external trends is necessary to design the most accurate and effective forecasting model.

Classification of the distributed generation forecasting techniques is shown in Fig. 1 [2]. First, the short-term forecasting method predicts the power generation of the distributed generation for 1 hour, several hours, a day, or up to 7 days. Such a type is useful for designing an integrated energy management system and improving the grid operation's security. Further, this method allows smoother starting and shutdown plans for the generator, better scheduling, and more economic power supply. Second, the mid-term forecasting method used to make forecasting for one week to one month or more, useful for establishing power system planning and maintenance schedules.

Lastly, the long-term forecasting method makes predictions from one month to one year. The long-term forecasting is useful for power generation, transmission, and distribution planning in addition to energy bidding and stable operation. In some papers, the forecasting period of the power generation divided into four categories by adding the ultra-short-term forecasting. The ultra-short-term forecasting method considers units of seconds and minutes and carries out for power smoothing, real-time power supply, and optimal reserve power [3].

Fig. 2 shows the frequency of algorithm utilization for each forecasting period. Since the ultra-short forecasting method has not published in many studies, it is uneasy to select useful algorithms. For short-term forecasting, models are used in this order: the ANN (Artificial Neural Network) model, the time series model, and the SVM model. Mid-term forecasting also seems not to be used as forecasting cycles as often as the ultra-short-term forecasting. The regression technique is mostly used in long-term forecasting, and ANN models, the time series model, and the bottom-up model seem to be applicable to some degree.

Fig. 3 shows the input variable utilization frequency for each forecasting period. As in the previous graph, there are few research samples related to the ultra-short-term and mid-term forecasting. In short-term forecasting, data are used in the order of precedence of

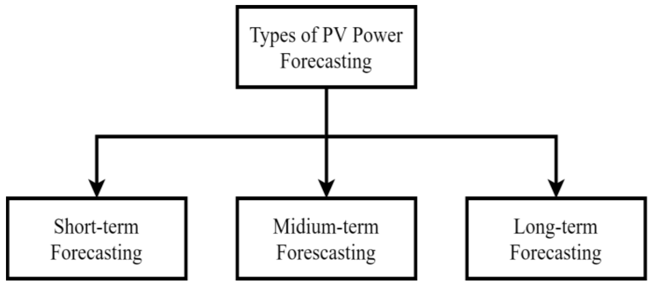

Fig. 1. Distributed power generation forecasting classification over time

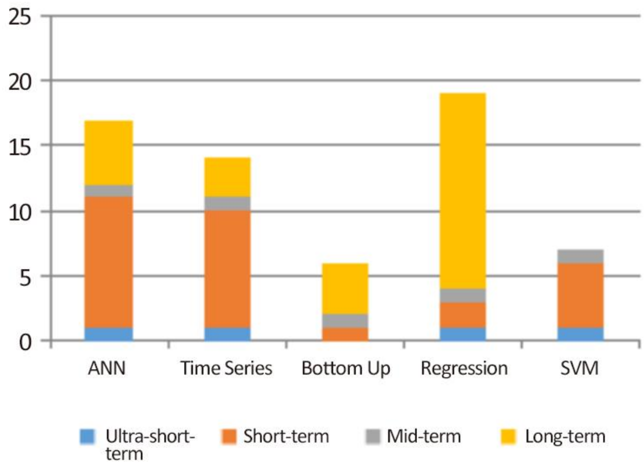

Fig. 2. Frequency of algorithm by forecasting period

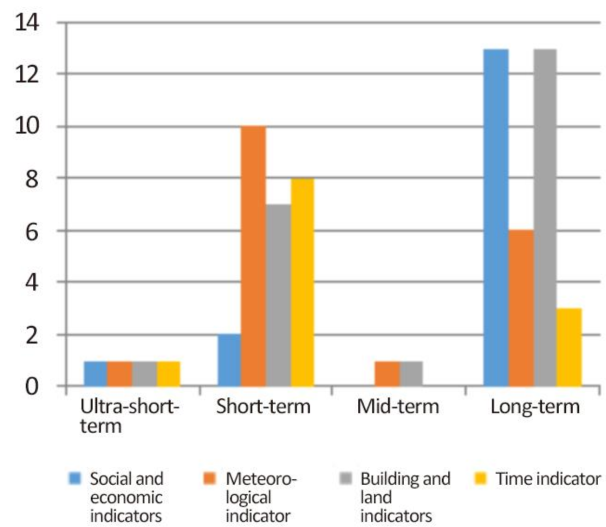

Fig. 3. Frequency of input variable by forecasting period

meteorological data, time indicators, and building and land use data. The social and economic variables are not well used in short-term forecasting, which means that variables for the distant future are not used since the forecasting period is short. On the other hand, the utilization of the meteorological variables and the time indicator decreased in the long-term forecasting. In contrast, the utilization of social and economic variables and building and land use data was increased. The long-term forecasting model considers the economic growth and analyzes the possibility of introducing the new load and

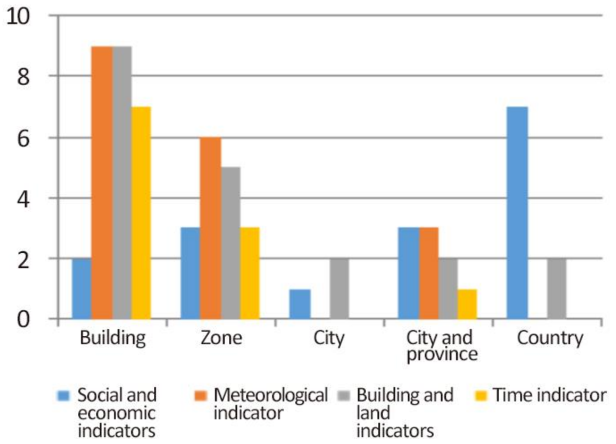

Fig. 4. Frequency of input variable by forecasting region

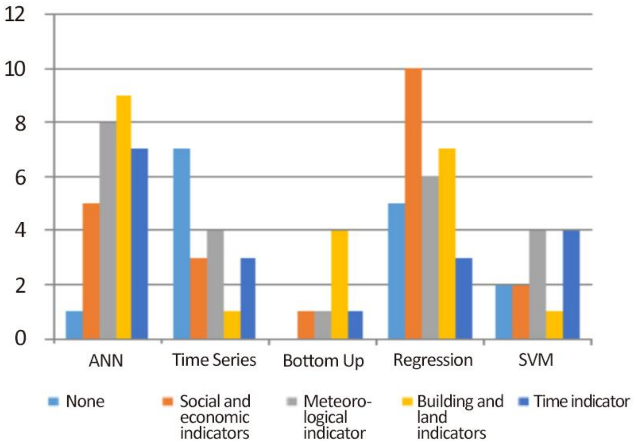

Fig. 5. Frequency of algorithm by forecasting method

distributed generation, so it seems to use the variables of the social and economic variables and the building and land use data.

The graph in Fig. 4 shows the frequency of utilization of the input variables for each size of the forecasting zone. The frequency of research that studies a small forecasting area unit like a building or district is high. In the load forecasting for a corresponding range, meteorological data, building and land use data, and time indicator data are frequently used. Few research studies involve forecasting area units the size of a city, and even province-sized forecasting does not focus on specific input variables in Korea. The forecasting in country-sized units is estimated to consider the proportion of buildings and land use constituting the country, and social and economic indicators are used most often.

Fig. 5 is a graph showing the frequency of utilization of the input variables for each forecasting technique. The ANN technique utilizes various input variables. Alternatively, it can also be determined that the ANN technique itself has a high utilization rate. The time series technique makes heavy use of meteorological variables. As seen in the preceding graph, the time series technique is commonly used for short-term forecasting. It is presumed that this is because meteorological variables have a great influence on the time series technique. For this reason, there is a shortage of research utilizing building and land use. A literature review finds scant use of the bottom-up technique, but building and land use are prevalent. This indicates that the forecasting for the entire zone is carried out through small-scale forecasting due to the properties of the bottom-up technique. This appears to be true since a detailed analysis of the building and land use is required. Further, since the regression technique with the highest utilization has been used a lot in the long-term forecasting, most published studies use social and economic variables, but other variables are also used appropriately. The SVM model is not frequently used for the forecasting, but the availability of meteorological variables and time indicators was high.

III. APPLICATION OF DISTRIBUTED GENERATION FORECASTING MODEL

Algorithms used for predicting the existing load and distributed generation have various algorithm techniques from regression to artificial intelligence (AI). Since the algorithm performance varies depending on the forecasting period and purpose, it is important to select an appropriate algorithm by confirming the pros and cons of each algorithm. Forecasting algorithms can be classified as Persistence, statistical models, and artificial intelligence models, and each group uses various algorithms.

The Persistence method is the simplest and uses the previous day's performance as the predicted value for the day. This method is the reference model used in many research studies. The most popular statistical schemes used in these studies are regression analysis and various time series-based analytic techniques. Multiple models are under development based on ANN for machine learning techniques in hybrid forms by combining statistical methods. The hybrid forecasting technique couples the advantages of the time series technique and the artificial intelligence forecasting model [4].

3.1 Statistical model

The statistical model is a mechanical model that carries out the statistical estimation based on the given data. Representative statistical models include ARMA, the ARIMA model considering the differential, and the AutoRegressive Moving Average with eXogeneous inputs (ARMAX) model the external variables and the time series data. The ARMA and ARIMA models have high forecasting accuracy for the time series data but cannot consider the external variables affecting the forecasting target. Therefore, it is expected that ARMAX or ARIMAX models will be highly utilized for distributed generation forecasting models that need to consider various external variables.

$$\Delta^d S_t(t) = c + \varphi_1 \Delta^d S_{t-1} + \varphi_2 \Delta^d S_{t-2} + \cdots + \varphi_p \Delta^d S_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{1}$$

Here, Δ_t^{dS} refers to the d-order differential time series, d refers to the number of certain differentials, and when the d is 0, ARIMA (p, d, q) can be represented as ARIMA (p, q). When the ARIMA model is used for predicting the generation quantity of the photovoltaics, the maximum output data of the generation of photovoltaics for each season can be collected and confirmed. However, in the ARIMA model, it is difficult to identify which factor generates the maximum output of the generation of photovoltaics. In other words, when predicting the generation quantity of photovoltaics using the corresponding model, it is not easy to clarify the relationship between the meteorological data and the maximum generation quantity data of photovoltaic power generation.

In reference [5], the SARIMA model was proposed to predict insolation in Seoul based on insolation data per hour collected for 37 years (1981 to 2017) from the meteorological administration. The

Root-mean-square error (RMSE) and the coefficient of determination (R^2) were used for evaluating the forecasting accuracy and were compared with the Monte Carlo model. In the paper, the insolation per month was predicted using the SARIMA(4,1,1) model and the insolation per day was predicted using the ARIMA(1,1,2) model. As a result, in the SARIMA forecasting model, RMSE=33.18 and R^2=79%, and in the ARIMA forecasting model, RMSE=104.26 and R^2=68%, thereby achieving excellent performance.

In reference [6], the forecasting performance was compared using the ARIMA model, the ARIMAX model adding the independent variables to the ARIMA model, and the multiple regression model to predict the insolation based on the meteorological data collected from the meteorological administration. As shown in Table 1, the mean absolute error (MAE) was utilized to compare the forecasting accuracy. The number 1 after the forecasting model name refers to the model using all four variables (cloud cover, temperature, humidity, and extra-atmospheric solar radiation). The number 2 refers to the model using only three meteorological variables (cloud cover, temperature, and humidity). There is a difference of about 0.16 to 0.23 depending on whether the extra-atmospheric solar radiation is reflected. Further, as a result of additional comparison of the actual values with the predicted values in the ARIMAX and multiple regression models, the result showed that the ARIMAX model was more suitable for the pattern of actual solar radiation than the multiple regression method.

TABLE 1
Comparison of forecast accuracy

Forecasting model	MAE
ARIMA	0.4121
ARIMAX 1	0.1895
ARIMAX 2	0.4137
Multiple regression 1	0.1553
Multiple regression 2	0.3146

The Vector Auto-Regression (VAR) model is a probabilistic process, represented as linear functions, that considers the variables dependent on the predictive variables and the past predictive values in multivariate analysis. The structural equation by the typical regression model describes the dependent variable Y using several independent variables $\{X_1, X_2, X_{3\,...}\}$ through the causality between the variables. However, in the regression model, it is assumed that the influence of the variable is always constant even when time t is changed, such that if structural changes progress rapidly. Therefore, although the variables' influence is changed, such a situation may not be appropriately reflected. Further, the structural model is constructed using economic principles. Therefore, there is a disadvantage in that the selection of the variables and the selection of exogenous variables in the model, that is, the input and the output, are determined by the subjectivity of the model designer. The ARIMA model was suggested to overcome the rigidity and subjectivity for the time. However, the ARIMA model predicts the future value assuming that the current measured values Z_t are reproduced by some regularity with the past, and the regularity will be maintained in the future. The ARIMA model is easily set but ignores the interaction between variables and thereby faces the limitation of univariate analysis. A model that supplements the regression model and the limitations of the time series analysis is the VAR model [7].

The VAR model is similar to the system of simultaneous equations but has analytical features. First, it is possible to confirm the dynamic effect in which a change in any one variable affects the endogenous variables through Impulse Response Analysis. Second, it is possible to analyze the relative magnitude of the portion where these variables contribute to the entire variation among the variations of the respective endogenous variables through the Variance Decomposition. Reference [8] models the annual data for the quantity of renewable electricity generated and carbon dioxide emitted via VAR. This study confirmed the influence of increases in the occupation rate of renewable energy sources on economic growth by Impulse Response Analysis. Further, the impact of the low occupation rate of renewable energy sources was verified by describing the economic growth and changes in the predictive error in carbon dioxide emissions through Variance Decomposition.

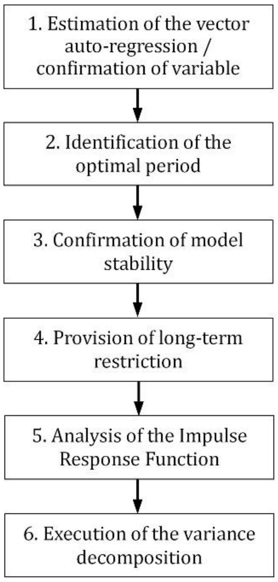

Fig. 6. Flowchart of VAR model based on renewable energy and economic indicators

GP regression is a technique with high utilization among non-linear regressions and provides a forecasting distribution other than a simple point forecasting with the probabilistic model. In the GP, it is assumed that simple target data is distributed with multivariate Gaussian and Gaussian noise at each point. In this assumption, the forecasting distribution is made into a Gaussian distribution [9].

An object of the Gaussian process regression technique is to identify and model the relationship between the input data and the output data through the learning data. The Gaussian process model can express the nonlinearity simpler than the artificial intelligence model and, therefore, is used for energy-related time series data analysis and modeling. The average function of the Gaussian process is shown in the equation below [10].

$$M(x) = E[f(x)] \tag{2}$$

The covariance function of the Gaussian process is expressed by the equation below.

$$k(x, x') = E\left[\left(f(x) - m(x)\right)\left(f(x') - m(x')\right)^T\right] \tag{3}$$

When the forecasting $y_0 = f(x_0)$ is made using the Gaussian process model, y0 and Y follow the Gaussian distribution, which can be expressed as follows [11].

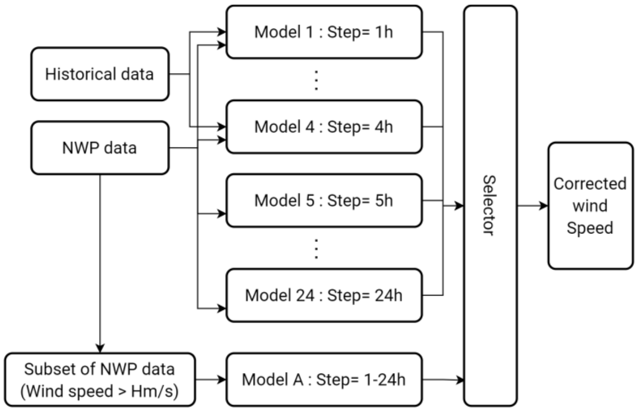

Fig. 7. Flowchart of suggested GP model

$$\begin{bmatrix} y_0 \\ Y \end{bmatrix} \sim N\left(0, \begin{bmatrix} A_p & B_p \\ B_p^T & C_p \end{bmatrix}\right)$$
$$A_p = cov_p(y_0, y_0) \tag{4}$$
$$B_p = \left(cov_p(y_0, y_1), \cdots, cov_p(y_o, y_N)\right)$$
$$C_p(i,j) = cov_p(y_i, y_j)$$

Reference [12] shows that wind power generation data with numerical date forecasting models included, data analysis measured by SCADA systems, and Gaussian processes can be used to predict wind power generation over up to a day. A flowchart of the proposed Gaussian process is shown in Fig. 7.

The wind turbine generating quantity in the Kansu region in western China was predicted using the Gaussian process proposed above. The forecasting accuracy was evaluated using multiple indicators, including RMSE, NRMSE, MAE, NAMPE. When comparing the accuracy according to the number of times of learning with the actual and the predicted wind turbine generation quantity, it can be confirmed that as the number of times of learning increases, the accuracy of the forecasting of wind turbine generation increases, and the RMSE is significantly decreased compared to the multilayer perceptron (MLP) model.

The wind turbine generation quantity obtained for the site can be incorrect due to inoperative sensors, measurement errors, or malfunction. Therefore, the problem can also be solved by combining the Gaussian process with multiple substitutions. Generally, as errors in techniques increase, the information loss rate also increases. However, it can be seen that the method using the Gaussian process instead of other models has the smallest error. Therefore, the Gaussian process method provides the highest reliability when information is lost due to various causes.

3.2 Artificial intelligence model

The artificial intelligence model is applied in various fields depending upon the improvement in the computer operation performance and is actively applied to the load and distributed generation forecasting, i.e., power demand. The ANN model is very comprehensive among the artificial intelligence models, and all models using artificial neural networks belong to the ANN model. Forecasting using artificial neural networks is a mathematical model developed based on how the biological neural system works. A typical artificial neural network structure is composed of an input layer, a hidden layer, and an output layer and interconnections of neurons, as shown in Fig. 8 [13].

An artificial neural network is a model that learns the relationship between input variables and output variables. First,

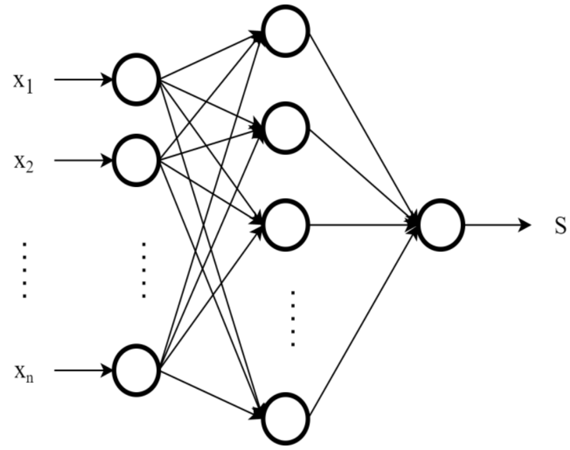

Fig. 8. Block diagram of ANN

various external variables are input to the input layer of the model composed of the artificial neural network. Then, the relationship between the input and hidden layers is automatically calculated, and the weighted values between the nodes are selected. Finally, the relationship between the hidden and output layers is automatically calculated inside the artificial intelligence technique. As a result, the output data are derived from the input variables, but the model developer cannot know from which function the predicted value is derived. The model incapable of knowing the intermediate process until these output variables are derived is referred to as the black-box model and is one of the disadvantages of artificial intelligence models.

In reference [14], the generation quantity from photovoltaics was predicted using the ANN. First, a Colorado region was selected that had an 805kW photovoltaic system installed. Then, the real-time data were learned using the multilayer perceptron-based ANN. Input variables were time, insolation, temperature, humidity, and wind speed, and outputs were DC voltage and current. Further, forecast errors were evaluated using MSE, which is expressed as follows.

$$MSE = \sum_{i=1}^{P} \sum_{k=1}^{N} \left(\widehat{y_{i,k}} - y_{i,k}^{P}\right)^2 \tag{5}$$

The actual generation quantity and the predicted generation quantity in the Colorado site were compared by applying ANN. It can be confirmed that the average percentage error was obtained at about 1.8%, and the MSE value was 0.9827. These results indicated that the improved ANN model had an accuracy of 93.9%. This type of research enables efficient modeling of photovoltaic panels.

In reference [15], an efficient forecasting system was produced using big data and ANN. The proposed method is composed of two steps, that is, the process of making the received data in the desired form and the process of learning the data using the BP algorithm of ANN.

Since the existing neural network independently learns the input data, there is a time limitation for learning the information. However, since the RNN model has a parallel structure by copying several identical neural networks, the learned result of the previous neural network can be used for training the current neural network. In other words, this is a very effective algorithm capable of predicting the current situation using previously acquired data [16].

Fig. 9 shows the structure of the RNN model, and the RNN structure variables are shown in Table 2. The RNN model is configured in a recursive structure in which a Hidden State Node is connected to the edge with a direction. Therefore, the input is made

Table 2
Internal variable of RNN

x_t	Input value at the time t
W_x	Weight for encoding (vector value)
H_t	Weight of the right edge at
W_h	$H_t$the number of certain differentials
W_y	Weight for decoding (vector value)
y_t	The output value of the decoding result at the time t (predicted value)

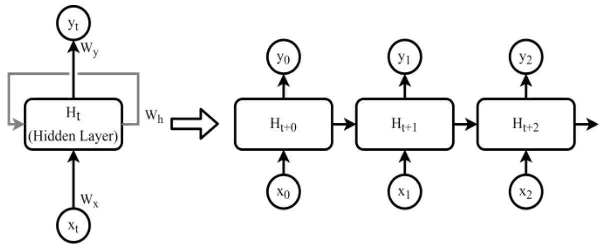

Fig. 9. Block diagram of RNN

Table 3
Comparison of decision parameter of each stage

Case	Coefficient of determination	15 minutes	30 minutes	45 minutes	60 minutes	75 minutes	90 minutes
Case 1	R^2	0.9994	0.9975	0.9938	0.9881	0.98	0.9699
Case 2	R^2	0.9954	0.9828	0.9619	0.9296	0.8864	0.8359

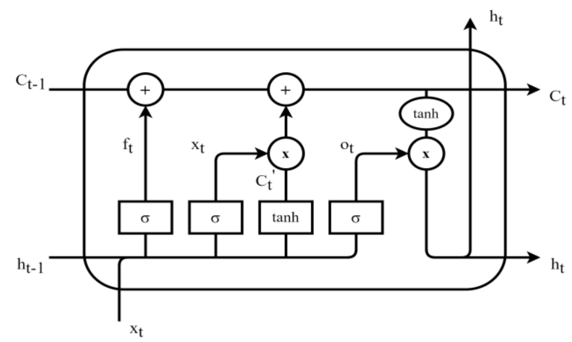

Fig. 10. Block diagram of LSTM

after calculating the state, and the previous data affects the next data. When the time information (t) and the weight (W) value are added, the previous data are affected next time through the W value. Further, the set of the weighted values for learning is in one cell and shared by all cells. In the RNN model, h_t is shown in the equation below.

$$h_t = f_{\omega}(h_{t-1}, x_t) \tag{6}$$

Here, h_t is an equation calculating the following new state, f_w refers to a function having a parameter ω, x_t refers to a vector value as an input at a certain time, and h_{t-1} refers to a value of the previous state. Various forms of the RNN are developed and provided, and the improved representative model is the LSTM and the Gated Recurrent Unit (GRU).

In the reference [17], the simulation was carried out based on various forecasting ranges to evaluate the forecasting performance of the proposed RNN model. The forecasting range was set between 15 minutes and 90 minutes, and the linear regression analysis result for the predicted value and the output value according to the forecasting range was suggested. The deviation of the predicted value from the actual measured value increased depending on the forecasting range. In particular, if the forecasting range is 60 minutes or more, a large error occurs in the linear regression analysis result. Further, compared to case 1, case 2 is a type that belongs to the winter season, and in case 2, the stability of the insolation was decreased, and the temperature was low, and therefore, the performance was relatively decreased. Therefore, the deviation of the forecasting results in case 1 was less than the deviation in case 2.

Further, in Table 3, the coefficient of determination R^2 according to the forecasting range by case was calculated. It can be determined that the closer the coefficient of determination of the RNN forecasting model is to 1, the more efficient it is. In Table 3, for the RNN model, R^2 was 0.9699~0.9994 in case 1, and R^2 was a value between 0.8359~0.9954 in case 2. The value of the R^2 was decreased depending on the forecasting range, which indicates that the efficiency of the RNN model is affected by time. Further, the R^2 in case 2 was lower than that of case 1, which indicates that the performance of the RNN model can be affected by the season of the year.

In reference [18], the generation quantity of the wind turbine generator was predicted based on the wind speed data using the RNN forecasting model. The RNN forecasting model uses only previously obtained wind speed data for predicting the wind speed, thereby shortening the forecasting time. The actual measured quantity and the wind speed forecasting results in 2003 of the Feed-forward Neural Network (FNN) model and the RNN model were compared using Mean Absolute Percentage Error (MAPE). First, for the RNN, it was 4.87% before 3 hours, 5.19% before 6 hours, 7.31% before 9 hours, 19.53% before a day, 23.89% before 2 days, and 28.22% before 3 days. For the FNN, it was 7.14% before 3 hours, 7.71% before 6 hours, 9.71% before 9 hours, 21.17% before a day, 26.42% before 2 days, and 30.07% before 3 days. These results indicate that the forecasting error in wind speed forecasting increases as the forecasting time increases. In particular, the mean square error of the FNN model gets unstable because the time-series information is destroyed, and the forecasting error increases. In other words, the use of the RNN showed better performance than the FNN model in predicting wind turbine generation quantity.

The RNN has a problem in that the learning ability is lowered as the distance between the related information and the section using that information increases. To solve this problem, the LSTM model uses memory instead of neurons in the RNN. The LSTM structure is shown in Fig. 10. The LSTM cell is divided into 4 layers, in which Layer 1 is a forget gate layer that determines whether to retain the memory values of the h_{t-1} of the previous state from the cell state and the current input value X_t as values of "0" and "1" or delete them through the sigmoid(σ) function. Layer 2 is an input gate layer and a step of determining the data to be stored in each cell state by a product of the value obtained by the function of sigmoid(σ) and the tanh function. Layer 3 is a cell state update, i.e., a step to store the information obtained from step 2 in the cell state. Layer 4 is an output gate layer to determine the final output value by the product of the sigmoid(σ) function result of the previous output and the current input and the tanh function result of the updated cell state. Therefore, the LSTM structure improved the learning ability to provide information continuity by storing, deleting, and maintaining the information [19].

Table 4
Performance comparison between LSTM and Other Models

Forecasting technique	Seasons	MSE	nRMSE	nMBE	R
Persistence	Spring	1857	0.110	0.000	0.942
	Summer	2332	0.124	0.000	0.924
	Autumn	1080	0.084	0.000	0.948
DNN	Spring	1154	0.087	0.010	0.965
	Summer	1545	0.101	0.006	0.950
	Autumn	847	0.075	0.022	0.967
SVR	Spring	1260	0.091	0.006	0.960
	Summer	1536	0.100	0.004	0.949
	Autumn	948	0.079	0.029	0.964
LSTM	Spring	1747	0.1049	−0.029	0.949
	Summer	2204	0.1178	−0.022	0.928
	Autumn	958	0.077	0.0071	0.953

Reference [17] presents the results of comparing the performances of the LSTM model with other forecasting techniques, which are shown in Table 4. Among the single models tested, the DNN model had the best performance. For the LSTM model, it is necessary to improve the performance using the ensemble modes. Therefore, in reference [17], the hybrid model was suggested, thereby improving the performance of the LSTM model.

In reference [18], the LSTM model was utilized to predict the output of the wind turbine generation site located on Jeju island. The performance was compared by implementing the models, such as ANN and SVR, often applied to existing forecasting models. To predict the output of the wind turbine generation site, the wind speed and output data obtained at 15-minute intervals from January 1 to August 31, 2018, were used for model learning, and RMSE, MAPE, etc., were used to determine the forecasting accuracy. The October, November, and December simulations were carried out to compare each forecasting technique's wind turbine generation quantity. Referring to the entire forecasting results in November, the ANN and LSTM models had results similar to those in September and October. Still, the SVR model was slightly low compared to those in September and October as MAPE=8.60%. This means that the SVR model constitutes the formula that best predicts the output data for all input data, so the smaller the output variation rate, the better the forecasting results. As a result, RMSE and the MAPE had similar forecasting performances by month while the simulation was carried out. However, the LSTM forecasting model had a better performance than the ANN and SVR models.

The GRU is a proposed model that reduces computations while maintaining the advantages of LSTM's recurrent neural networks. There is also the evaluation that there is no significant difference in terms of performance because there is no significant difference from the recurrent neural network of the LSTM in terms of structure. However, there is a performance difference depending on the data used and applications. For example, the GRU does not have memory for storing and memorizing the cell state and, therefore, can show worse results than that of the recurrent neural network of the LSTM when the amount of data increases. However, when a small amount of time series data is used for the operation, the GRU has a smaller number of operations than the LSTM and, therefore, has a better result in terms of time and accuracy [20].

The GRU algorithm simplifies the structure of the LSTM algorithm designed for solving the long-term dependency problem of the recurrent neural network. The LSTM is composed of three gates, that is, input, output, and forget. In contrast, the GRU is composed of

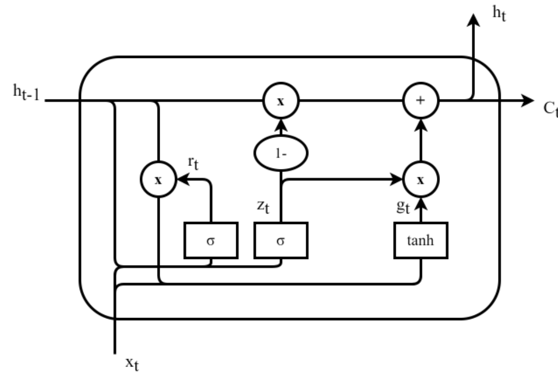

Fig. 11. Block diagram of GRU

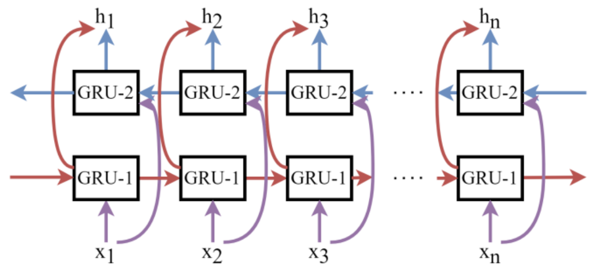

Fig. 12. Block diagram of bidirectional GRU

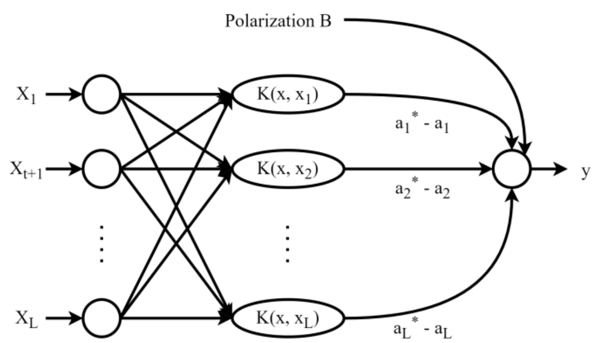

Fig. 13. Block diagram of SVR

the update gate and reset gate coupled with the input and forget gates to maintain the long-term dependency problem solution ability of the LSTM algorithm while keeping the advantages of the same forecasting performance and fast learning speed [21].

The GRU can be mathematically expressed as follows [20].

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ r_t &= \sigma(W_r x_t + W_r h_{t-1}) \\ h_t &= \tanh(W_{x_t} + U(r_t \circ h_{t-1})) \\ h_t &= (1 - z_t) \circ h_{t-1} + h_{t-1} + z_t \circ h_t \end{aligned} \tag{7}$$

When predicting the wind speed value and comparing the actual value with the predicted value using the GRU model, it can be seen that the RMS value is 0.98, and the predicted value and the actual value are statistically similar, along with the trends. Therefore, when the algorithm is used, it is expected that a similar trend can be confirmed in other cases for wind speed.

The performance of the GRU cannot reach the level theoretically analyzed and predicted. Furthermore, the GRU has the disadvantages of accessing previously acquired information and is incapable of accessing future information. The bidirectional GRU was devised to overcome such limitations. In the bidirectional GRU model, the final output affects the data at (t+1) and the data at (t-1).

The SVR is a technique that generalizes the Support Vector Machine (SVM) among the machine learning algorithms and finds the optimal plane, including as much data as possible to predict the data. The model that carries out the forecasting using regression analysis among the SVMs is the SVR. The SVR is used in many fields, such as power demand forecasting, meteorology, economic activity, and finance, and goes through preparing data and estimating the function to learn the SVR [22]. The structure of the SVR is shown in Fig. 13.

The problem of the SVR is to determine f(x) that is a function approximating an unknown function. The SVR technique is based on the calculation of the linear regression function in the high-dimensional characteristic space and can be shown as follows.

$$f(x) = \left(w \cdot \Phi(x)\right) + b \tag{10}$$

In reference [23], the reliability of the result was evaluated by introducing MAPE and RMSE to evaluate the photovoltaic generation output forecasting, and the time for generation of photovoltaics output was set as 8 hours. Furthermore, the insolation of the photovoltaic generation site where the data for insolation are not provided was predicted by applying the Kriging technique for predicting photovoltaic generation quantity. The forecasting was carried out using the SVR model having a fast calculation speed, and the result can be confirmed that the average MAPE is 7.44%, and the average RMSE is 0.86MW. Compared to other machine learning techniques, the result was confirmed as a decrease of about 14%.

V . EXTENTION OF ENSEMBLE MODEL

The ensemble model refers to the mixed model that sorts out the high accuracy of each model. Fig. 14 shows the features of the ensemble model. The figure shows the model's performance for each forecasting period, and several models have different performances depending on the forecasting period. Thus, introducing the ensemble model makes it possible to suggest an improved model compared to using just one model.

Fig. 14 shows an error graph of the model for each forecasting period, and pros and cons depending on the features of the model in addition to the forecasting period are shown. The ensemble model combines the advantages of each model into one model. Recently in the field of forecasting, it is rare to present a single model. Therefore, after testing various forecasting models, the final model is an ensemble of models with the highest accuracy.

In reference [24], the accuracy was evaluated by predicting the Daily Global Solar Radiations (DGSRs) of the hybrid model combining the advantages of the SARIMA model and the SVM model and the existing SARIMA model. RMSE, Normalized Root Mean Square Error (NRMSE), Mean Absolute Percent Error (MAPE), etc., were used to evaluate the above. Table 5 shows the results of comparing the performance of the hybrid model and the SARIMA model. The hybrid model utilizing the two forecasting techniques could improve the forecasting performance of the uncertain DGSR and solve the somewhat random phenomenon.

In the reference mentioned above [17], the forecasting accuracy was evaluated by predicting the insolation of the hybrid model combining the advantages of the LSTM model and the Deep Neural Network (DNN) model and various forecasting models. The indicators, such as MSE, NRMSE, NMBE, and R, were used to evaluate

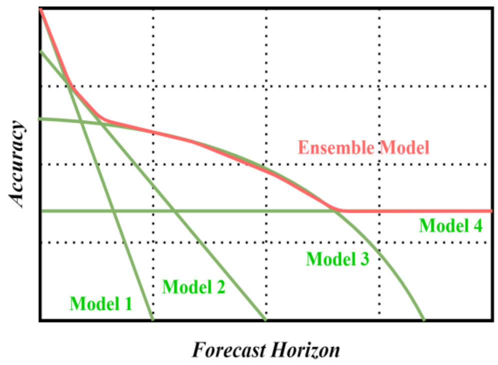

Fig. 14. Accuracy of each forecasting model and ensemble model

Table 5
Comparison of SARIMA and Hybrid Model

Forecasting technique	RMSE [Wh/m^2]	NRMSE [%]	MAPE [%]	R
SARIMA	889.768	14.913	14.139	0.867
Hybrid	866.814	14.529	13.820	0.874

Table 6
Performance comparison between hybrid models based on LSTM and Other Models

Forecasting technique	Seasons	MSE	nRMSE	nMBE	R
LSTM	Spring	1747	0.1049	−0.029	0.949
	Summer	2204	0.1178	−0.022	0.928
	Autumn	958	0.077	0.0071	0.953
WT+BPNN	Spring	1891	0.111	0.006	0.939
	Summer	2145	0.119	0.002	0.928
	Autumn	1011	0.081	0.021	0.966
WT+RBFNN	Spring	2782	0.135	0.001	0.912
	Summer	3610	0.154	0.005	0.887
	Fall	1921	0.112	0.010	0.903
Hybrid WT + LSTM-DNN	Spring	973	0.080	0.016	0.971
	Summer	762	0.071	0.005	0.975
	Fall	660	0.066	0.015	0.971

the above. As shown in Table 6, the hybrid model combining LSTM and DNN had excellent performance in all aspects compared to other techniques. However, this technique can significantly decrease the values of MSE, MAPE, and NRMSE when compared to other models, such as Persistence, WT + RBFNN, WT + BPNN, and LSTM. The DNN model also had excellent forecasting performance but was inferior to the performance of the hybrid model. Further, it is simple to predict the photovoltaic generation quantity using LSTM, but it has a lower performance than the hybrid model.

VI . CONCLUSION

In these days, distributed generations are critical factors in distribution planning. It is important for distribution planning to forecast distributed generations. This paper analyzed various forecasting techniques for predicting the distributed generation and use cases of forecasting models. And this paper discussed the input variables that must be considered. Since the distribution planning considers the distribution equipment of the line in terms of the distribution grid in the mid-to long term, it is important to identify the amount of distributed generation introduced in the mid-to long term rather than predicting the generation quantity in short time units. Regression, ANN, time series, and bottom-up methods are often used in mid-to long-term forecasting algorithms. The ensemble model had a higher forecasting accuracy than the single model according to the use cases of the models. Therefore, it is determined that the mid-to long-term distributed generation forecasting model carries out the forecasting using the bottom-up method and the ensemble model consisting of a combination of the algorithms of the time series-based ANN model and the regression model will be most effective.

References

[1] Renewable Energy 3020 Implementation Plan, Ministry of Trade, Industry and Energy, 2017.

[2] Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Deventer, W.V., Horan, B., Stojcevski, A., Forecasting of photovoltaic power generation and model optimization: A review. Renew. Sustain. Energy Rev., vol. 81, pp. 912–928, 2018.

[3] W. Li, T. Logenthiran, W. L. Woo, "Multi-GRU prediction system for electricity generation's planning and operation," in IET Generation, Transmission & Distribution, vol. 13, no. 9, pp. 1630-1637, 2019.

[4] Akhter, M. N., Mekhilef, S., Mokhlis, H., & Shah, N. M., Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques, IET Renewable Power Generation, vol. 13, no. 7, pp. 1009-1023, 2019.

[5] Alsharif MH, Younes MK, Kim J., Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation: The Case Study of Seoul, South Korea. Symmetry, vol. 11, pp. 240-257, 2019.

[6] Sahm Kim, "A study on solar irradiance forecasting with weather variables", The Korean Journal of Applied Statistics, 2017.

[7] Oryani, B., Koo, Y., Rezania, S., "Structural Vector Autoregressive-Approach to Evaluate the Impact of Electricity Generation Mix on Economic Growth and CO2 Emissions in Iran," Energies, vol. 13, no. 4268, 2020.

[8] Y. Li et al., "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," Renewable Energy, vol. 66, pp. 78-89, June 2014.

[9] Jae Hyun Yoo, "Gaussian Process Model-based Reinforcement Learning," Journal of Institute of Control, Robotics and Systems, vol. 25, no. 8, pp. 746-751, 1997.

[10] A. G. Abo-Khalil and D. Lee, "SVR-based Wind Speed Estimation for Power Control of Wind Energy Generation System," 2007 Power Conversion Conference, Nagoya, pp. 1431-1436, 2007.

[11] J. Yan, K. Li, E. Bai and A. Foley, "Special condition wind power forecasting based on Gaussian Process and similar historical data," 2015 IEEE Power & Energy Society General Meeting, Denver, pp. 1-5, 2015.

[12] N. Chen, Z. Qian, I. T. Nabney and X. Meng, "Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction," in IEEE Transactions on Power Systems, vol. 29, no. 2, pp. 656-665, March 2014.

[13] S. Bae, Research Trends of Photovoltaic Generation Forecasting, KIEE: World of electric, vol. 67, no. 12, pp. 16-25, 2018.

[14] A. Almadhor, "Performance prediction of distributed PV generation systems using Artificial Neural Networks (ANN) and Mesh Networks," International Conference on Smart Grid, Nagasaki, Japan, pp. 88-91, 2018.

[15] M. N. Rahman, A. Esmailpour, "An Efficient Electricity Generation Forecasting System Using Artificial Neural Network Approach with Big Data," IEEE First International Conference on Big Data Computing Service and Applications, Redwood City, CA, pp. 213-217, 2015.

[16] Gwon-Yoon Lee, Sang-Boo Lee, Universal Prediction SystemRealization Using RNN, Journal of KII, vol. 16, no. 10, pp. 11-20, 2018.

[17] Ospina, Juan, AlviNewaz, M. Omar Faruque, Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model, IET Renewable Power Generation, vol. 13, no. 7, pp. 1087-1095, 2019.

[18] Tomonobu, S., Atsushi, Y., Naomitsu, U., Toshihisa, F., "Application of Recurrent Neural Network to Long-Term-Ahead Generating Power Forecasting for Wind Power Generator," IEEE Power Systems Conference and Exposition, pp. 1260-1265, 2006.

[19] Jungin Lee, Il-Woo Lee. "Technology trends of Renewable energy generation forecasting based on ICT," The Journal of The Korean Institute of Communication Sciences, vol. 36, no. 11, pp. 3-8, 2019.

[20] K. W. Kim, G. S. Jang, S. M. Lim, I. K. Ahn, J. Park, H. C. Oh, "GRU-based Activity Recognition from Early-stage Motion," The Institute of Electronics and Information Engineers, pp. 2016-2019, 2020.

[21] Yong-jin Jung, Kyoung-woo Cho, Jong-sung Lee, Chang-heon Oh, Particulate Matter (PM10) Concentration Forecasting model using GRU, The Korea Institute of Information and Communication Engineering, vo. 23, no. 2, pp. 644-646, 2019.

[22] Ju-Hwan Ham, Sung-Yul Kim, "A Study on the Comparison of Kernel Functions Appropriate to the SVR-based Power Demand Prediction Algorithms," The Korean Institute of Electrical Engineers, Workshop, pp. 158-159, 2020.

[23] Mun, Gwon-Sun, Understanding of VAR model, KOSTAT, Research of statistics analysis, pp. 23-56, 1997.

[24] Boualit, S.B.; Mellit, A., "SARIMA-SVM hybrid model for the prediction of daily global solar radiation time series," In Proceedings of the 2016 International Renewable and Sustainable Energy Conference (IRSEC), IEEE, Marrakesh, Morocco, pp. 712–717, PP. 14–17, November 2016.