Photovoltaic Generation Forecasting Using Artificial Neural Networks Model with Input Variables and Model Parameters Selection Algorithm in Korea

Fauzan Hanif Jufri and Jaesung Jung

Abstract—One of the challenges in the photovoltaic (PV) system integration to the utility grid is its uncertainty which may affect the grid stability. Consequently, PV generation forecasting becomes one of the most important roles in the development of the grid-connected PV systems. This paper presents a PV generation forecasting using Artificial Neural Networks (ANN) with input variables and model parameters selection algorithm for a PV system located in Jangseong-gun, South Korea. It uses the ASHRAE Clear-Sky model to solve the unavailability of the irradiance data in this area. Additionally, the weather forecasting information is considered to compensate the uncertainty of the sky condition. Pearson Correlation Coefficient is employed to decide which weather data mostly influence the amount of solar irradiance. Cross-validation technique is used to select model parameters so that the model can fit to forecast PV generation in any conditions. As a result, it is shown that the model accuracy is improved compared to the model without input variables and model parameters selection.

Index Terms—ASHRAE clear-sky model, photovoltaic generation forecasting, artificial neural networks, input variable selection.

I. INTRODUCTION

Electric power generation using fossil fuels has caused an environmental problem such as greenhouse gas effect due to its carbon dioxide (CO_2) emission. Many countries including South Korea are led to a collaborative effort to reduce the CO_2 emission by limiting the fossil fuel based power generation. Photovoltaic generation is largely utilized as an alternative to fulfill the electricity demand. However, PV generation is not as easy to be controlled as fossil fuel based power generation due to its high uncertainty and intermittency. Therefore, the assurance of availability through the higher accuracy of the forecasting technique becomes one of the crucial factors when high penetration of PV generation is integrated into the electricity grid.

Many studies have been conducted to develop the methodology to forecast the PV generation with higher accuracy. In [1], the accuracy of the forecasting was enhanced by including the PV short circuit current and open circuit voltage in developing the model. A division of bulk data by using various techniques was applied in [2] to improve the performance of the neural networks for PV generation prediction. The other effort for the forecasting improvement was proposed by [3] where solar zenith angle and solar azimuth angle were included to develop the model. Additionally, other literature discussed the improved methods for one-day-ahead PV generation or irradiance forecasting such as in [4]-[8].

A comparison of PV generation forecasting models was studied in [9], [10]. This comparison showed that no absolute model could be applied in all area. The accuracy of PV generation forecasting depends on the area where the PV is installed. Even though one methodology may fit at one place, this may poorly perform when applied at another place. Moreover, some PV generation location may not have sufficient facilities to provide the required weather information. Therefore, the forecasting study must be carried out by considering case by case condition of the site to get a more accurate result.

In this paper, the Artificial Neural Networks with input variables and model parameters selection algorithm is utilized to forecast PV generation in Jangseong-gun, South Korea. The input variables include estimated irradiance, weather information, and the previous PV generation. The model parameters cover the number of nodes in the hidden layer, regularization parameter, and learning rate.

Irradiance forecasting information is not always available at the target location. This fact forces the PV systems operators to employ a model to estimate the irradiance. In [11], twelve models to simulate the solar irradiance were reported. A widely used model to estimate the irradiance, ASHRAE Clear-Sky model [12]-[17], is utilized in this paper. Then, Pearson Correlation Coefficient is employed to select the most correlated weather information with the PV generation before they are used to develop the forecasting model.

A suitable number of nodes in the hidden layer, regularization parameter, and learning rate are selected through validation technique. This technique is applied to make sure that the model can fit to forecast any given data. Final model parameters are decided after the validation process so that it can be used to forecast PV generation in any weather conditions.

The paper is organized as follows: the input variables preparation and selection are explained in Section II. Section III discusses the algorithm for the model parameters selection. Section IV shows the result, and lastly, the conclusions are given in Section V.

II. INPUT VARIABLES PREPARATION AND SELECTION

A. Irradiance Estimation by ASHRAE Clear-Sky Model

ASHRAE Clear sky model estimates the solar radiation under a clear sky (cloudless) as a function of the site latitude, solar position (azimuth and altitude angle), solar declination, atmosphere condition, and time. This model is developed based on the extraterrestrial solar irradiance, which is the intensity of solar radiation that arrives perpendicularly at the outer surface of the earth's atmosphere. This value differs depending on the distance between the sun and the earth, and this distance varies throughout the year. The extraterrestrial solar irradiance is attenuating when it reaches the earth's surface due to the depth of the atmosphere and gasses and particulate matters in it. By considering this, the portion of the extraterrestrial solar irradiance on the earth's surface (I_B) can be calculated as follows [18]:

$$I_B = A \exp(-km) \tag{1}$$

$$A = 1160 + 75 \times \sin[(360/365) \times (n - 275)] \tag{2}$$

$$k = 0.174 + 0.035 \times \sin[(360/365) \times (n - 100)] \tag{3}$$

$$m = h_2 / h_1 = \sqrt{(708 \times \sin \beta)^2 + 1417} - 708 \times \sin \beta \tag{4}$$

where A is "apparent" extraterrestrial solar irradiance which varies every day throughout the year, k is optical depth due to the energy transmission through atmosphere, m is the air mass ratio which is the ratio of the actual length of the path of the sun's rays when they pass through the atmosphere (h_2) to the shortest possible length of that path (h_1), n is the day number and β is the solar altitude angle. The solar altitude angle (β) is calculated by the following equation:

$$\sin \beta = (\cos L \times \cos \delta \times \cos H) + (\sin L \times \sin \delta) \tag{5}$$

where L is the site latitude, δ is the solar declination, and H is the hour angle.

The solar irradiance estimation is based on the solar time. Therefore, two adjustments are required to calculate the hour angle (H) so that this model can be applied for local time. The first one is longitude adjustment and the second one is related to the variation of the length of the day in one year.

$$ST = CT + 4 \times (LTM - LL) + E \tag{6}$$

$$E = 9.87 \times \sin 2B - 7.53 \times \cos B - 1.5 \times \sin B \tag{7}$$

$$B = (360/364) \times (n - 81) \tag{8}$$

where ST is the solar time, CT is local clock time, LTM is local time meridian, LL is local longitude, and E is the equation of time. Hence, the hour angle (H) must be adjusted following the (6), (7), and (8).

The amount of solar irradiance on the collector is determined as the summation of the direct beam, the diffuse irradiance, and the reflected irradiance projected onto the collector [18]. The reflected irradiance is usually neglected because the amount is relatively very small compared to others. So, it can be mathematically expressed as below:

$$I = I_{BC} + I_{DC} \tag{9}$$

$$I_{BC} = I_B \times \cos \theta \tag{10}$$

$$I_{DC} = I_B \times C \times ((1 + \cos \Sigma)/2) \tag{11}$$

where I_{BH} and I_{DH} are the direct beam irradiance and diffuse irradiance on collector's surface, respectively. The angle of incidence θ is defined as the angle between the collector's normal line and the incoming beam, whereas Σ is the tilt angle of the collector. C is a sky-diffuse factor, which is given by the following equation:

$$C = 0.095 + 0.04 \times \sin[(360/365) \times (n - 100)] \tag{12}$$

It is worth to note that if the collector is set to fully facing the sky (horizontal), then the tilt angle value (Σ) is zero. Therefore, the calculation of the solar irradiance can be simplified. This case is generally known as Global Horizontal Irradiance (GHI) [19].

To perform the irradiance calculation, the parameters that are related to the site of the study case are shown in Table I.

TABLE I: PV LOCATION INFORMATION

Site latitude (L)	35.25 °
Local Time Meridian (LMT)	135 °
Local Longitude (LL)	126.72 °

B. Weather Data Selection by Pearson Correlation Coefficient

The estimated irradiance is only fit to forecast the PV generation in clear sky condition which is not always happened. Therefore, additional information representing the sky and weather condition is required. In Korea, the Dong-Nae Forecast in Korea Meteorological Administration (KMA) provides 48 hours weather forecasts in every 3 hours [20]. In this paper, the weather information is retrieved from the weather station which is 15 km away from the location of the installed PV.

The available weather forecasts are temperature, humidity, rain fall, wind speed, dew point temperature, and wind direction. The inclusion of all of the weather data may not always guarantee better forecasting results. First, it will lead to longer computation time and the second one is overfitting problem. Therefore, it is necessary to analyze the weather data which one is highly correlated with the forecasted value which is the amount of generated power here.

Pearson Correlation Coefficient is used to analyze the correlation between the weather data and the generated power [21]. This coefficient can be calculated by the following equation:

$$R_m = \frac{n \times (\sum X_m Y) - (\sum X_m) \times (\sum Y)}{\sqrt{(n \times \sum X_m{}^2 - (\sum X_m)^2 \times (n \times \sum Y^2) - (\sum Y)^2)}} \tag{13}$$

where R_m is the Pearson Correlation Coefficient for the weather data m, n is number of data, X_m is the weather data (i.e., temperature, humidity, etc.), and Y is the generated power.

The Pearson Correlation Coefficient (R) value is between -1 and 1. The positive sign means that the two variables have a

positive relationship, on the other hand, the negative sign means that the two variables have a negative relationship. A strong relationship is shown when the value of R is close to 1 (strong positive relationship) or -1 (strong negative relationship). If the value of R is zero or close to zero, it means that two variables have no relationship or weak relationship.

C. Sky Conditions

Solar irradiance received on the collector varies depending on the sky condition. A lot of clouds may block the solar irradiance from the sun to the earth's surface. The ASHRAE Clear-Sky model is developed to estimate the irradiance in cloudless condition. The PV generation forecasting by including this estimated irradiance is only suitable if there are a few clouds in the sky. However, the estimated irradiance is not legitimate if the sky is cloudy. Generalizing the sky condition in the model will lead to incorrect forecasting results.

One approach to solve this problem is by building a model for different sky conditions. The Dong-Nae Forecast provides the sky condition in four different classes. Therefore, four forecasting models are developed based on the sky conditions in this paper. Those are Class-1 to indicate the "clear", Class-2 to indicate "partly cloudy", Class-3 for "mostly cloudy", and Class-4 for "cloudy" sky condition. The sky classes are determined based on the number of clouds given by the weather station. This model separation can minimize the effect of inaccurately estimated irradiance since each model is trained based on the sky condition.

III. Artificial Neural Network and Model Parameters Selection

The ANN is constructed by the input layer, hidden layer, and output layer. The architecture of the ANN is shown in Fig. 1. The input layer consists of the irradiance data retrieved from ASHRAE Clear-Sky model, selected weather data, and observed or forecasted power generation of the previous time points. The hidden layer is composed of one layer with several nodes that will activate the input variables through an activation function and then, the output layer shows the forecasting result. Each node in every layer is connected to the nodes in the next layer through a certain value of weight.

There are some parameters that have to be determined in the ANN model development process. Those are the number of nodes in the hidden layer, regularization parameter, and learning rate. These parameters must be properly selected so that the model may fit in any conditions.
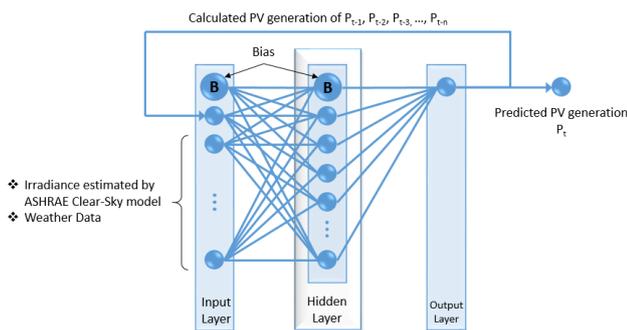

Fig. 1. Architecture of artificial neural networks.

The model that is acquired from the training algorithm may show a higher accuracy, but it may result in a lower accuracy if it is applied to other data set which is not akin with the train data set. The model parameters selection in this paper uses cross-validation techniques. The data set is divided into three parts called train data set, cross-validation (CV) data set, and test data set. In this technique, the model given by the train data set is not directly used as a forecasting model. Instead, it must be checked if it can fit to forecast PV generation in any given irradiance, sky condition, or weather condition.

The overall step of the algorithm is shown in Fig. 2. First, the initial input variables such as weather information (W) and a number of previous PV generation (TL), and model parameters such as the number of nodes in the hidden layer (HL), regularization parameter (λ), and learning rate (LR) are set. These variables and parameters are used in the train algorithm to obtain the initial model. The initial model is validated by using the CV data set and then, the error is checked. If the error is not acceptable, the input variables and model parameters are adjusted until the error is satisfied. The model with the accepted error is picked as the final model to forecast PV generation.
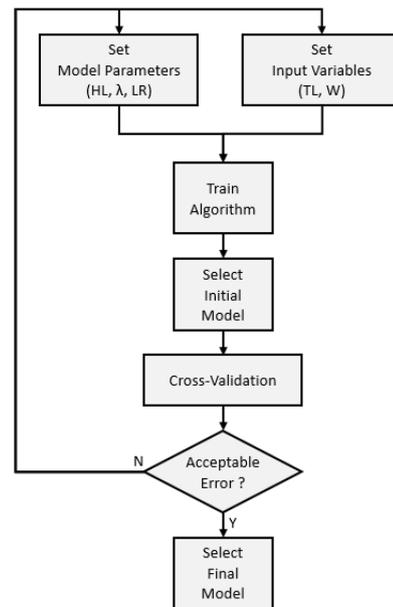

Fig. 2. The forecasting model development flowchart.

IV. Results

A. Input Variables Selection

The main variable that highly influences the PV generation is solar irradiance since it has a direct correlation with the amount of generated power. In this paper, solar irradiance is estimated by using (1) to (12) together with the information provided in Table I.

The other required variables are the weather condition. As explained above, it is necessary to select the most suitable weather data in developing the ANN model to avoid an overfitting model. Pearson correlation coefficient is used to do this task. The result of the correlation analysis is presented in Table II and Fig. 3.

Table II shows that the ambient temperature and humidity

are more correlated with PV generation compared to other weather parameters. Ambient temperature has a moderately positive relationship with the PV generation (0.5238), and humidity has a moderately negative relationship with PV generation (-0.4881). Whereas, the other parameters are less correlated with PV generation. This result is also shown in Fig. 3. The ambient temperature and humidity give steeper slope than the others which show an almost flat slope (R close to zero).

TABLE II: RESULT OF PEARSON CORRELATION COEFFICIENT

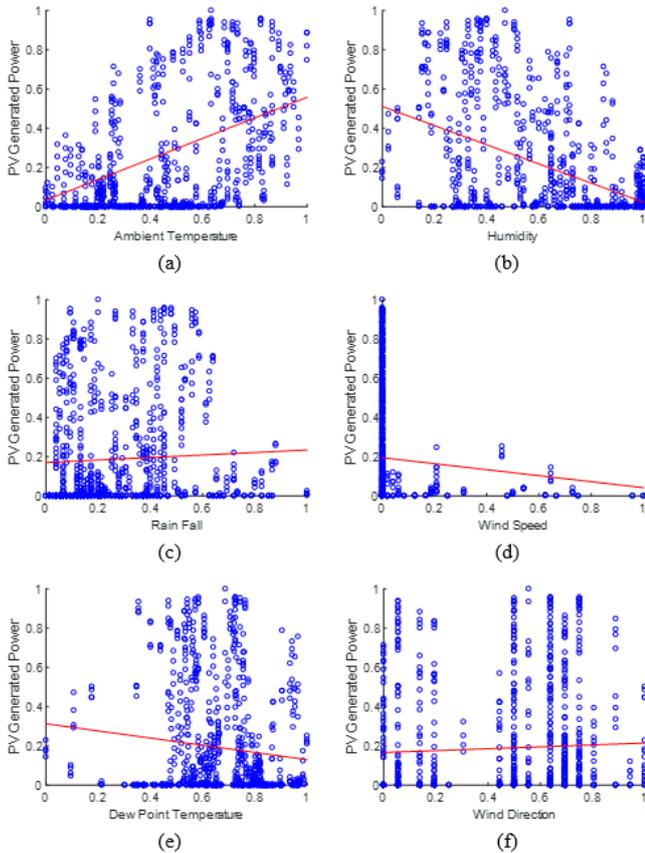| Weather Parameter	R Value
Ambient Temperature (T)	0.5238
Humidity (H)	-0.4881
Rain Fall (RF)	0.0636
Wind Speed (WS)	-0.1530
Dew Point Temperature (DT)	-0.1829
Wind Direction (WD)	0.0482

Fig. 3. Pearson Correlation Coefficient. (a) Ambient temperature, (b) Humidity, (c) Rain fall, (d) Wind speed, (e) Dew point temperature, (f) Wind direction.

According to the Pearson Correlation Coefficient, the ambient temperature and humidity are selected as an additional input to develop the ANN model. A further test is required since it is not always true that including all parameters in the forecasting model development will increase the precision. The model which uses a combination of these two weather parameters is developed and tested to check which combination provides a better accuracy.

B. The Forecasting Result

Three models are developed to show the accuracy comparison between the proposed model and the others. The

accuracy is measured by using Mean Absolute Percentage Error (MAPE) on the test data set.

First, the model without input variables and model parameters selection is built (Model-1). All variables are considered in this model and default parameters are used. The second model is built by using input variables selection algorithm but without model parameter selection (Model-2). This model uses the ambient temperature and humidity as the weather variables. It uses the previous 3-hour of PV generation.

At last, the proposed model (Model-3) is developed with input variables and model parameters selection. The weather variables used in this model is only ambient temperature. The PV generation of the previous 3-hour is also included. Table III shows the comparison of this three models.

TABLE III: RESULT COMPARISON

	Model-1	Model-2	Model-3
Selected variables	T, H, RF, WS, DT, WD	T, H	T
Previous PV generation	3-hr	3-hr	3-hr
MAPE (%) in CV data set	16.05	11.81	11.93
MAPE (%) in test data set	12.97	13.04	12. 49

Table III shows the result of three models when they are applied to the CV data test and the test data set. It is shown that in Model-1, even though all variables are included in the model development, it does not guarantee to result in a good model, which is indicated by higher MAPE (16.05%).

Model-2 is developed by considering the input variables selection but without model parameters selection. The model includes the ambient temperature and humidity as additional weather variables and also the previous 3-hr of the PV generation. It has the lowest MAPE value (11.81%) and may be regarded as the best forecasting model. However, when it is applied to the test data set, it gives a higher error (13.04%). This indicates an overfitting model, which means the model is only relatively suitable to forecast the data that is similar to its train data set.

Model-3 has slightly higher MAPE value (11.93%) compared to Model-2. But since the model parameters have been adjusted so that it has assured the model can fit in any given condition regardless the similarities of the data set with its train data set. It is shown that Model-3 can provide a lower MAPE compared to Model-2 in the test data set (12.49%).

Thus, it can be concluded that the suitable forecasting model is obtained by using input variables including irradiance, ambient temperature, and 3-hr data of previous PV generation. In addition, the selected model parameters are ten nodes in the hidden layer, regularization parameter of 0.1, and optimization learning rate of 0.3. The MAPE of this model in the test data set is 12.49%.

The results of this model on the test data set are shown in Fig. 4 to Fig. 6. The blue color indicates the actual PV generation, the red color indicates the forecasted PV generation, and the green color indicates the sky class index. Fig. 4 shows the results for the day during 'clear' sky condition (Class-1). It can be seen that the error between the actual and forecasted result is very small, which means the model provides an accurate forecasting during the "clear" sky

condition.

Fig. 6 shows the results for the day when the sky is varied from "clear", "partly cloudy" and "mostly cloudy". Even though there is more error when compared with "clear" sky, it is still acceptable to be used as a reference for PV generation forecasting. The pattern of the forecasting result in these sky conditions is mimicking the actual PV generation. This means that the model keeps its consistency as high as possible to forecast PV generation. Fig. 6 shows the result for the cloudy day without any "clear" day information. This result also shows the consistency of the forecasting model.

It can also be observed that the actual PV generation in these three days have an almost similar trend with small fluctuation even though the sky condition is varied. This is because the KMA forecast covers broad forecasting area so that the sky will not always be same at the PV site.
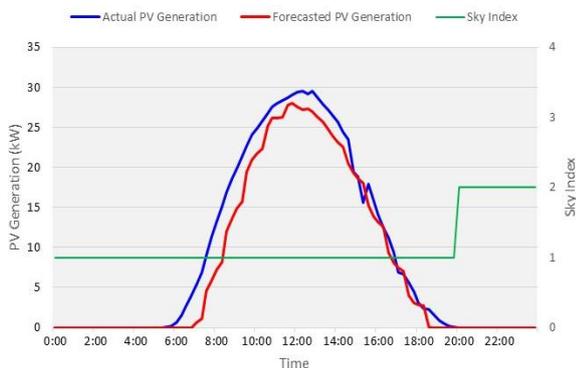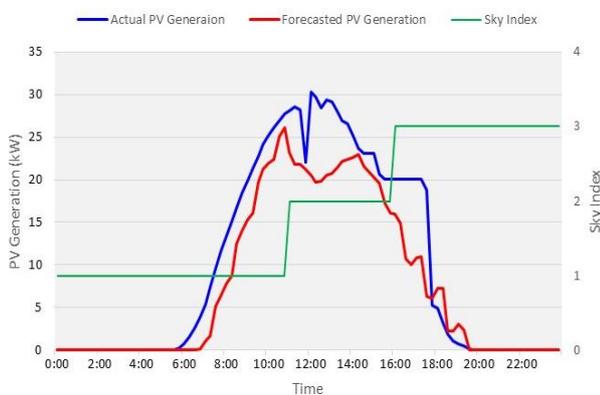

Fig. 4. Forecasting result of test day-1.

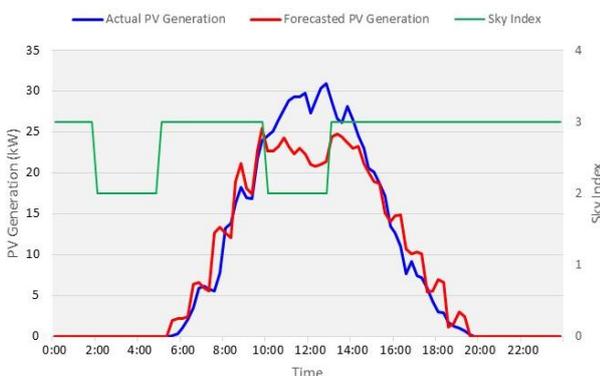Fig. 5. Forecasting result of test day-3.

Fig. 6. Forecasting result of test day-4.

V. CONCLUSION

The ANN algorithm is used to build a PV generation forecasting model. The development of ANN model for PV generation forecasting requires various model parameters and allows any variables as its input

The input variables used in the forecasting model are solar irradiance, weather information, and the previous 3-hr PV generation. The unavailability of the forecasted irradiance is solved by using ASHRAE Clear-Sky model. The weather information which is selected by using Pearson Correlation Coefficient is validated within the cross-validation data set to get the final decision of which weather information is required to develop the model. It remains that only ambient temperature can give a better accuracy on the test data.

The model parameters are selected through cross-validation technique. The parameters of the model that gives lowest MAPE in the cross-validation data set are chosen as the ANN parameters to forecast PV generation. It is shown that this model fits and suitable to forecast in the test data set.

This paper has shown that the proposed algorithm improves the accuracy of the forecasting model compared to the model without input variables and model parameters selection algorithm. The model can be used regardless the sky condition within an acceptable error.

REFERENCES

[1] K. Al Khuffash and L. A. Lamont, *Using Artificial Neural Networks to Enhance the Accuracy of the Photovoltaic Simulation Model*, pp. 0-4, 2017.
[2] S. Pulipaka and R. Kumar, "Performance enhancement of neural network training using hybrid data division technique for photovoltaic power prediction," in *Proc. - 2017 17th IEEE Int. Conf. Environ. Electr. Eng. 2017 1st IEEE Ind. Commer. Power Syst. Eur. EEEIC / I CPS Eur. 2017*, 2017.
[3] C. Huang, A. Bensoussan, M. Edesess, and K. L. Tsui, "Improvement in artificial neural network-based estimation of grid connected photovoltaic power output," *Renew. Energy*, vol. 97, pp. 838-848, 2016.
[4] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and C. H. Kim, "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," *2007 Int. Conf. Intell. Syst. Appl. to Power Syst. ISAP*, 2007.
[5] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Math. Comput. Simul.*, vol. 131, pp. 88–100, 2017.
[6] C. Voyant, P. Randimbivololona, M. L. Nivet, C. Paoli, and M. Muselli, "Twenty four hours ahead global irradiation forecasting using multi-layer perceptron," *Meteorol. Appl.*, vol. 21, no. 3, pp. 644–655, 2014.
[7] A. Ahmad, T. N. Anderson, and T. T. Lie, "Hourly global solar irradiation forecasting for New Zealand," *Sol. Energy*, vol. 122, pp. 1398-1408, 2015.
[8] C.-M. Huang, C.-J. Kuo, S.-J. Chen, and S.-P. Yang, "One-day-ahead hourly forecasting for photovoltaic power generation using an intelligent method with weather-based forecasting models," *IET Gener. Transm. Distrib.*, vol. 9, no. 14, pp. 1874-1882, 2015.
[9] M. Kayri and I. Kayri, "The performance comparison of multiple linear regression, random forest and artificial neural network by using photovoltaic and atmospheric data," in *Proc. 2017 14th International Conference on Engineering of Modern Electric Systems (EMES)*, 2017, pp. 1-4.
[10] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78-111, 2016.
[11] J. A. Davies and D. C. McKay, "Evaluation of selected models for estimating solar radiation on horizontal surfaces," *Sol. Energy*, vol. 43, no. 3, pp. 153-168, 1989.
[12] K. Bakirci, "Estimation of solar radiation by using ASHRAE clear-sky model in Erzurum, Turkey," *Energy Sources, Part A Recover. Util. Environ. Eff.*, vol. 31, no. 3, pp. 208-216, 2009.

[13] J. B and M. Khan, "Estimation of clear-sky solar radiation using ASHRAE Model for Aligarh, India," *Int. J. Eng. Res. Technol.*, vol. 7, no. 3, pp. 227-236, 2014.

[14] S. A. Al-Sanea, M. F. Zedan, and S. A. Al-Ajlan, "Adjustment factors for the ASHRAE clear-sky model based on solar-radiation measurements in Riyadh," *Appl. Energy*, vol. 79, no. 2, pp. 215-237, 2004.

[15] S. Alsadi and Y. F. Nassar, "Correction of the ASHRAE clear-sky model parameters based on solar radiation measurements in the Arabic countries," vol. 5, no. 4, pp. 1-16, 2016.

[16] M. Abouhashish, "Applicability of ASHRAE clear-sky model based on solar-radiation measurements in Saudi Arabia," vol. 140001, 2017.

[17] P. Amarananwatana and C. Sorapipatana, "An assessment of the ASHRAE clear sky model for irradiance prediction in thailand nuntiya," *Asian Journal on Energy and Environment*, vol. 8, no. 2. pp. 523-532, 2007.

[18] G. M. Masters, *Renewable and Efficient Electric Power Systems*, 2nd Ed. Wiley-IEEE Press, 2013.

[19] M. J. Reno and C. Hansen, *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*, Sandia National Laboratories, August 2016, 2014.

[20] Korea Meteorological Administration. [Online]. Available: http://web.kma.go.kr/eng/biz/forecast_01.jsp

[21] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise Reduction in Speech Processing*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1-4.

Fauzan Hanif Jufri received the B.S. degree in electrical engineering from Universitas Indonesia, Indonesia, in 2008.

He has work experience as an electrical design engineer with multinational engineering companies in Indonesia and South Korea. He is currently pursuing M.S. degree at the Department of Energy System Research, Ajou University, Suwon, South Korea. His research interests include smart grid and microgrids, Distributed Energy Resources (DER) operation and control, machine learning, and grid resilience.

Jaesung Jung received the B.S. degree in electrical engineering from Chungnam National University, Korea; the M.S. degree in electrical engineering from the North Carolina State University, Raleigh, NC; and the Ph.D. degree in electrical engineering from the Virginia Tech, Blacksburg, VA.

He is currently a faculty member in the Department of Energy Systems Research at Ajou University, Korea. His research interests include the development and deployment of renewable and sustainable energy technologies.