Prediction Model for PV Performance With Correlation Analysis of Environmental Variables

Gyu Gwang Kim, Jin Ho Choi, So Young Park, Byeong Gwan Bhang, Woo Jun Nam, Hae Lim Cha, NeungSoo Park, *Member, IEEE*, and Hyung-Keun Ahn, *Member, IEEE*

***Abstract*—With increasing installations of photovoltaic (PV) systems, interest in power forecasting has also increased. Inaccurate forecasts would result in substantial economic losses and system reliability issues. The correlation between weather variables and PV power is critical to ensure the efficient use of energy in PV systems. A key step toward accurate power forecasting is estimating the output from a PV system based on known environmental input data. In this research, all available weather data are used to predict the PV power. Meteorological and power data are then analyzed using a statistical approach to identify the order of significance of the input variables. Then, a predictive model is suggested as a function of irradiance, ambient temperature, wind speed, and relative humidity. The model produces a root mean square error of 4.957% and a mean absolute percentage error of 5.468% during the measurement period and over the entire range of irradiation.**

***Index Terms*—Correlation analysis, correlation coefficient, mean absolute percentage error (MAPE), prediction model, regression analysis, weather variable.**

I. Introduction

In 2017, the top ten countries in terms of installed photovoltaic (PV) power system capacity shared a total installed capacity of 344.5 GW. For the past three years, the number of PV power generation systems worldwide has increased because of a surge in China's installations [1].

Therefore, the importance of establishing reliability in PV power systems is increasing as the use of PV power generation systems is expanding. The performance of solar power plants and plans for energy distribution can be reviewed through accurate forecasting of solar power systems. PV power is affected by a variety of weather and environmental conditions, including geography, solar radiation, temperature, relative humidity, and wind speed. The stability of solar power generation systems is generally evaluated in terms of the performance ratio. Usually, the performance ratio of PV power generation is in the range of $0.7-0.9$ [2]. Insolation and temperature are the primary factors affecting PV power generation. Increasing the temperature of a solar module causes a decrease in its power output. Generally, PV power output decreases by 0.4% for every 1 °C rise in temperature [3].

Although various methods have been proposed for calculating PV power output, there are basically two ways of assessing the maximum amount of power produced by solar modules. The first approach calculates the instantaneous peak power based on the *I–V* curve under certain conditions, such as the standard test condition (STC). PV module manufacturers evaluate their products under the STC, which uses a standard spectrum of 1000 W/m^2 irradiation, a cell temperature of 25 °C, and an air mass of AM1.5 (equivalent to ASTM G173-03). The second approach uses regression analysis of long-term data on PV module power generation. These data can be used to build models of module operations at different meteorological and solar radiation values. The rated power is calculated under certain reference conditions. The photovoltaics for utility scale applications (PVUSA) test conditions (PTC) and the nominal operating cell temperature test (NOCT) are examples of methods that use regression. In the mid-1990s, under the guidance of the U.S. National Renewable Energy Laboratory, a series of test conditions were developed for measuring solar cell panel performance under real-world conditions. The conditions were referred to as solar light generation, or PTC, for larger scale application test conditions. The PTC represents the test conditions developed to test and compare PV systems as part of the PVUSA project [4]. The PTC utilizes a solar radiation intensity of 1000 W/m^2, a temperature of 20 °C, and a wind speed of 1 m/s at 10 m above the ground. The criteria for these methods are listed in Table I [5], [6].

In general, the PTC is thought to provide realistic measurements, because it gives a better approximation of actual solar and climatic conditions than the STC. The main difference between the STC and PTC is temperature; the STC assumes the surrounding environment is at 25 °C. Once the PV module reaches its maximum power output under the reference condition, the energy rating can be determined.

In this study, different methods for predicting the power output of PV modules were compared. In addition, the PV module

Manuscript received November 6, 2018; revised December 13, 2018 and January 22, 2019; accepted February 5, 2019. Date of publication February 26, 2019; date of current version April 19, 2019. This work was supported in part by the 2017 Open R&D Program of Korea Electric Power Corporation under Grant R17XH02, and in part by the New and Renewable Energy Technology Program of the Korea Institute of Energy Technology Evaluation and Planning granted financial resources by the Ministry of Trade, Industry and Energy, Republic of Korea (20183010014260). *(Corresponding author: Hyung-Keun Ahn.)*

G. G. Kim, J. H. Choi, S. Y. Park, B. G. Bhang, W. J. Nam, H. L. Cha, and H.-K. Ahn are with the Next Generation Photovoltaic Module and Power System Research Center, Konkuk University, Seoul 05029, South Korea (e-mail: rbrhkd00@konkuk.ac.kr; shorev@konkuk.ac.kr; sheyen@konkuk.ac.kr; bbk0627@konkuk.ac.kr; ska4034@cbtp.or.kr; haelim@konkuk.ac.kr; hkahn@konkuk.ac.kr).

N. Park is with the Department of Computer Science and Engineering, Konkuk University, Seoul 05029, South Korea (e-mail: neungsoo@konkuk.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JPHOTOV.2019.2898521

TABLE I
TEST CONDITIONS FOR PV MODULE POWER OUTPUT [7]–[10]

Method	Irradiance [W/m^2]	Temperature [°C]	Spectrum	Wind speed [m/s]
STC	1000	25 (module)	AM 1.5	-
PTC	1000	20 (ambient)	-	1
NOCT	800	20 (ambient)	-	1

generation forecast model was presented and analyzed using the correlation between weather elements and power generation. The different regression models that were reviewed are described in Table II.

Gianolli-Rossi and Krebs described the energy ratings method, which is used only to calculate regression coefficients for irradiation data above 500 W/m^2 [11]. The Farmer model with PTC parameters determines the power rating based on regression analysis using data for 30 days [12]. The cumulative data collection periods and solar radiation are at least 30 days and 10 kWh/m^2, respectively. This method uses solar irradiance at 500 W/m^2 or greater, ambient temperature, and wind speed at 10 m above the ground surface. The model developed by Rosell and Ibanez extends the benefits of the *I–V* curve method [13]. Based on adjusted *I–V* curves, this method computes a set of maximum power outputs using different irradiances and temperatures.

Furthermore, this model set is fitted by a nonlinear multivariable regression equation and determines the maximum output under operating conditions using variables $D_j (j = 1 - 4)$ and *m* parameters. The model proposed by Yang *et al.* uses irradiance and cell temperature, where *G* is the incident solar radiation (W/m^2), α is a temperature coefficient, and β is a calibration constant [14]. Fuentes *et al.* proposed one of the simplest known models, which is described in detail by Osterwald [15]. However, the version of this model shown in Table II has been widely used since the 1970s. In this model, $P_{\max.\mathrm{ref}}$ is the maximum reference PV power under STC, $G_{T.\mathrm{ref}}$ is the incident global irradiance under STC (1000 W/m^2), G_T is the incident global irradiance (W/m^2), and γ is the cell maximum power temperature coefficient (°C^{-1}) [16]. The model by Huld, which is based on standardized irradiance and the PV module temperature, uses empirical coefficients [17]. This model, which is a modification of the model developed by King [18], [19], expresses the power output of the PV module in terms of empirical coefficients $P_{\mathrm{STC},m}$ and k_j $(j = 1 - 6)$. As shown in Table II, this model can be calibrated to data containing only the measured power at given values of *G* and *T*.

II. CORRELATION ANALYSIS BETWEEN WEATHER VARIABLES AND POWER GENERATION

A. *Outdoor Exposure Testing Site*

Monitoring of power generation from a system of PV modules was conducted on a rooftop, as shown in Fig. 1. The geographical coordinates of this test site location are 36° 54′ 08.3″ N and 127° 32′ 26.4″ E. Fig. 2 shows the flowchart of the data measurements at the test site.

Weather data were obtained for global horizontal irradiance (GHI), ambient temperature (T_a), wind speed (WS), and relative humidity (RH). The temperature of the PV modules (T_m), plane of array irradiance (POA) and maximum PV power output ($P_{\max}$) were measured separately. Information on the PV modules used for this study is shown in Table III.

The weather data were measured every five minutes over a period of approximately two years, and the $\mathrm{P_{max}}$ values were measured at the same times.

Table IV shows the monthly average values obtained from these measurements. The average temperatures were in the range of 10−15 °C annually, 23−26 °C in August (the hottest month), and –6−3 °C in January (the coldest month). At the test site, the GHI and POA levels were the highest in May. The T_m and T_a values were the highest in August and the lowest in January. The WS values indicate that the winds were the strongest in the spring months and the weakest in the fall and winter months

B. *Correlation Coefficient Analysis*

The correlation factors were analyzed to determine the relationships between $P_{\max}$ and the individual weather factors. The correlation coefficient *r*, which indicates the degree of association between two variables, x_i and y_i, is expressed as follows [20]–[22]:

$$\mathrm{r} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2 \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}} \tag{1}$$

and

$$\bar{x}_i = \frac{1}{N}\sum_{i=1}^{N} x_i, \bar{y}_i = \frac{1}{N}\sum_{i=1}^{N} y_i. \tag{2}$$

By applying (2) to (1), the following equation can be driven:

$$r = r_{x_i y_i} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{3}$$

where $\{\bar{x}_i,\ \bar{y}_i\}$ and *n* are the mean and sample size, respectively, and $\{x_i, y_i\}$ are the individual sample points indexed by *i*.

There are two different methods for estimating the correlation (and the correlation coefficients) between the two variables: Pearson and Spearman. The Pearson correlation analysis method evaluates the linear relationship between two variables. If a change in one variable is proportional to a change in the other variable, then there is a linear relationship between these variables. The Spearman correlation analysis method evaluates the simple (ordinal or rank) relationship between two variables. In such simple relationships, the two variables tend to change together, but not necessarily in a proportional manner. The Spearman correlation coefficient is not based on the raw data but on the ranked values for each variable. This study used the

TABLE II
REGRESSION MODELS FOR PV POWER OUTPUT

Correlation	Comments	Reference
$P = C_1 + (C_2 + C_3T_a)G_T + (C_4 + C_5V_f)G_T{}^2$	C_j regression coefficients	Taylor (1986) [6]
$P = C_1G_T + C_2G_T{}^2 + C_3G_T \ln G_T$	Above 500 W/m^2	Gianolli-Rossi & Krebs (1988) [11]
$P = G_T(b_1 + b_2G_T + b_3T_a + b_4V_f)$	Above 500 W/m^2, b_j regression coefficients, V_f 10 m above ground	Farmer (1992) [12]
$P_{DC(G,T_c)} = D_1G + D_2T_c + D_3(\ln G)^m + D_4T_c(\ln G)^m$	D_j (j = 1–4), m parameters	Rosell & Ibanez (2006) [13]
$P = (\alpha T_c + \beta)G$	α = temperature coefficient, β = calibration constant	Yang et al. (2000) [14]
$P_{max} = P_{max.ref}\frac{G_T}{G_{T.ref}}[1 + \gamma(T_c - 25)]$	γ = -0.0035 (range - 0.005°C^{-1} to - 0.003°C^{-1}, T_c in °C)	Fuentes et al. (2007) [15]
$P_{(G',T')} = G'(P_{STC,m} + k_1(\ln G') + k_2(\ln G')^2 + k_3T' + k_4T'(\ln G') + k_5T'(\ln G')^2 + k_6T'^2)$	$G' \equiv G/G_{STC}$, $T' \equiv T_{mod} - T_{STC}$, k_j regression coefficients	Huld (2011) [17]

Fig. 1. Test site for outdoor experiment.

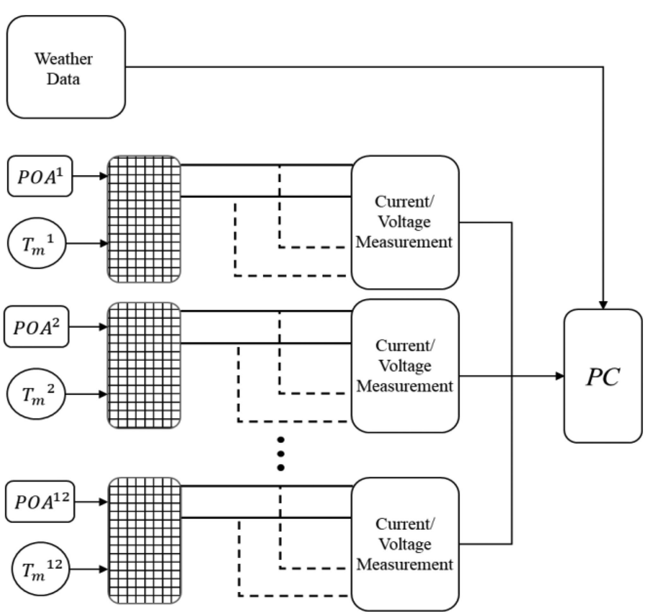

Fig. 2. Measurement flow for measuring weather data.

TABLE III
SPECIFICATION OF PV MODULES USED IN TEST

c-Si	Contents
P_{max} [W]	260
V_{oc} [V]	38.0
I_{SC} [A]	9.16
V_{mpp} [V]	30.7
I_{mpp} [A]	8.62
Efficiency [%]	16.1
Temperature coefficient of P_{max} [%/°C]	-0.4

Pearson correlation method to analyze the correlation between $P_{\max}$ and the environmental variables. Fig. 3 shows the results of this correlation analysis using the R studio program.

The histograms shown in the diagonal plots in Fig. 3 depict the frequency distributions of $P_{\max}$ and the environmental data presented in Table IV. The numbers above the diagonal are the correlation coefficients, which range from -1 to $+1$. The direction of the relationship is indicated by the sign of the coefficient; the "+" sign indicates a positive correlation and the "–" sign a negative correlation. The higher the absolute value of the correlation coefficient, the stronger is the association between the two variables. The graphs below the diagonal are the scatter plots of the measured data, which depict the correlation between the corresponding variable pairs in the orthogonal coordinate system. Red line represents the simple multiple regression line between the two variables [23], [24].

TABLE IV
AVERAGE ENVIRONMENTAL DATA DURING THE MEASUREMENT PERIOD

Month	POA [W/m²]	GHI [W/m²]	Tm [°C]	Ta [°C]	WS [m/s]	RH [%]
Jan.	414	345	18.1	-1.9	0.84	55
Feb.	534	397	26.7	1.1	1.04	51
Mar.	558	472	35.5	9.4	0.78	45
Apr.	550	501	42.0	16.5	0.97	46
May	660	583	51.0	21.4	1.01	45
Jun.	514	492	50.7	25.1	0.83	54
Jul.	447	450	49.8	27.1	0.85	63
Aug.	530	477	54.9	28.5	0.73	59
Sep.	504	434	48.6	23.5	0.64	58
Oct.	519	426	43.1	18.1	0.67	58
Nov.	343	332	26.8	8.6	0.49	61
Dec.	360	319	20.4	1.9	0.60	62

C. Correlation Analysis Between PV Power Generation and Environmental Variables

1) PV Module Power Versus Irradiation (POA and GHI): A PV module generates the maximum amount of electricity (i.e., $P_{\max}$) when the incident light is orthogonal to the plane of the module. Therefore, based on this analysis of PV modules installed at an angle of 30°, $P_{\max}$ has a stronger correlation to POA than GHI. Specifically, $P_{\max}$ and POA show a correlation coefficient of 1.0, but the correlation coefficient between $P_{\max}$ and GHI is 0.77. Thus, for an installed PV module, POA has a stronger influence on $P_{\max}$ than GHI. The correlation between POA and GHI is 0.79, which also indicates a high degree of association between these variables [25].

2) PV Module Power Versus Temperature (T_a and T_m): The correlation analysis suggests that an increase in T_m is likely to cause an increase in $P_{\max}$. However, in reality, this correlation is negative. Under constant solar radiation conditions, power generation decreases as the temperature of the photovoltaic modules increases. The power output of crystalline PV modules typically reduces by 0.4% per 1 °C increase in the temperature, but at installation sites, both $P_{\max}$ and T_m increase with the irradiation. Thus, the increase in T_m caused by the increase in irradiation has a stronger correlation with $P_{\max}$ than with T_a. The correlation coefficient between $P_{\max}$ and T_m is 0.71, whereas, that with T_a is 0.13. The correlation coefficient between T_m and T_a is 0.73, which is the same as the coefficient of correlation between POA and T_m. [26]

3) PV Module Power Versus Relative Humidity (RH): Relative humidity affects PV power generation in a similar way as dust accumulation does. Water vapor particles reduce the amount of insolation, and light that hits water droplets is scattered by refraction, reflection, or diffraction [27], [28]. In previous studies, relative humidity, ambient temperature, and power output were measured for crystalline and amorphous PV modules. These studies found that increasing the ambient temperature or relative humidity reduced the output efficiency of the PV module significantly.

When identical PV modules with the same efficiency were compared, the effect of relative humidity on variations in maximum power efficiency were found to be 50% greater than that of the ambient air temperature [29]. In this study, the correlation coefficient between RH and $P_{\max}$ was estimated as –0.46, which shows that a reduction in relative humidity is associated with an increase in power output. Furthermore, Fig. 4 shows the time series of RH and POA measurements, which indicates that, in outdoor conditions, humidity decreases as irradiation increases, which consequently increases $P_{\max}$.

4) PV Module Power Versus Wind Speed: There is a weak positive correlation of 0.19 between WS and $P_{\max}$. Studies have shown that an increase in wind speed causes a reduction in module surface temperature, which in turn results in an increase in power generation [30]. In the mid-1980s, heat models were developed at the Sandia National Laboratories to study temperature and PV performance. The temperature of the PV module can be determined within an accuracy of ±5 °C in terms of the plane of array irradiance, ambient temperature, and wind speed [31].

III. PREDICTION MODEL OF PV POWER

The process flow for selecting the input variables of the PV power predictive model is shown in Fig. 5. In this process flow, the numbers are the correlation coefficients (estimated in Section II) between $P_{\max}$ and the environmental variables.

Based on these correlation coefficients, the variables shown in each branch of the process flow were selected to create the six different regression models shown in Table V. The environmental variables were included based on POA, which has the highest correlation coefficient with $P_{\max}$. If $|r|$ is greater than 0.5, the variable itself is used in the regression model, otherwise the variable is multiplied by POA. Model P_1 uses T_m and POA with the highest correlation; P_2 uses POA and RH; P_3uses POA, T_m, and RH, which had the next highest correlation coefficient after POA; P_4 uses POA, WS, and T_a, which had the lowest correlation coefficient; P_5 uses POA, T_a, WS, and RH; and P_6 uses POA, T_m, WS, and RH.

Based on the correlation analysis, a model of PV power generation was obtained through regression analysis by selecting factors that affect PV power generation [32]. The coefficients of the regression analysis model are estimated next. For a multi-regressive model, the dependent variables are described by

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + e_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + e_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \beta_3 x_{3n} + \cdots + e_n. \end{aligned} \tag{4}$$

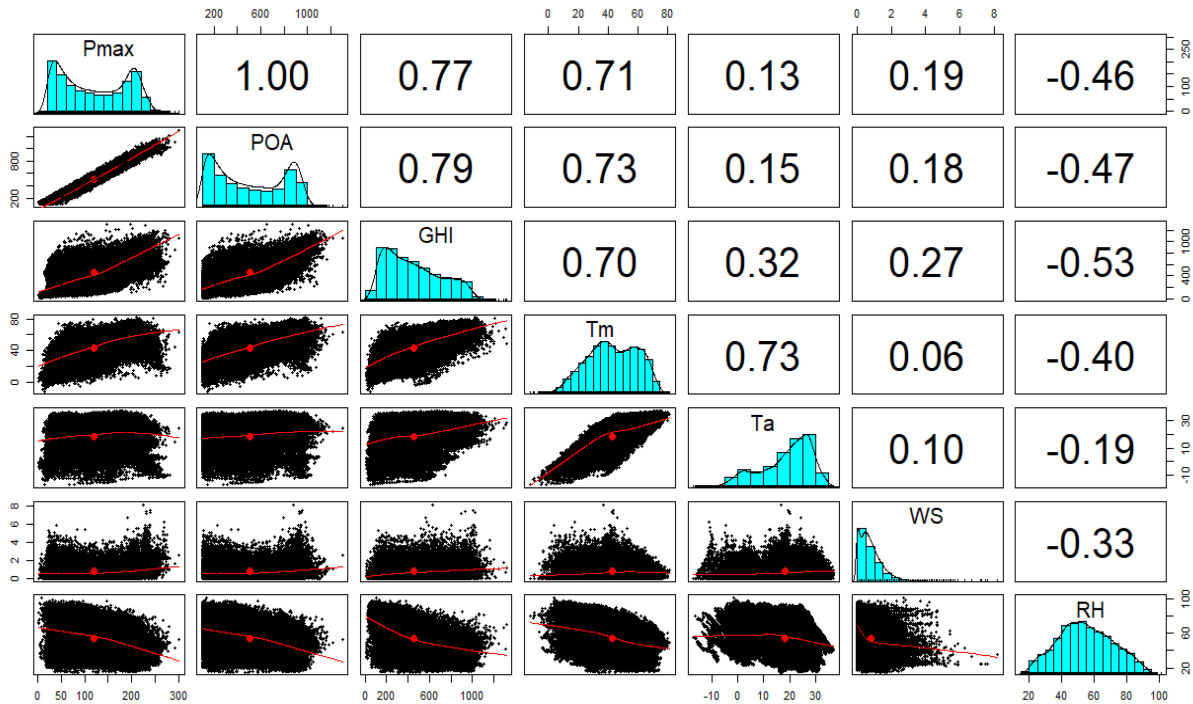

Fig. 3. Pairwise scatter plots and correlation coefficients between PV power output and environmental variables.

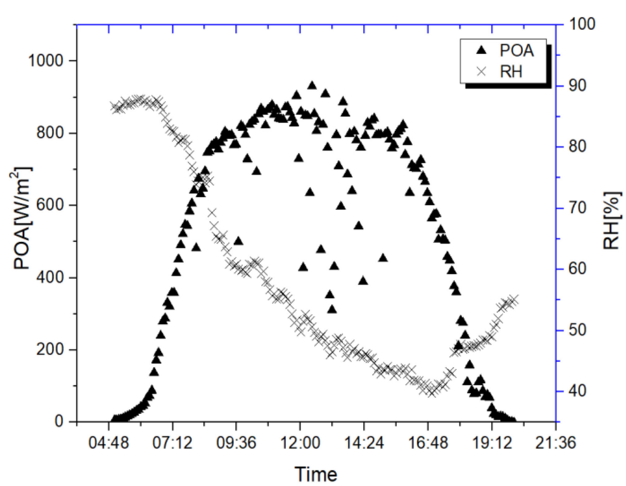

Fig. 4. Time series of POA and RH measured in a clear day.

The superior and independent variables can be described by

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + e_i. \tag{5}$$

The mean of the error terms is zero, and the variance follows a normal distribution with a constant variance among all the measurements. The coefficient β is estimated by least-squares regression, as follows:

$$\begin{aligned} \mathrm{L} &= \min\left[\sum_{i=1}^{n} {e_i}^2\right] = \min\left[\sum_{i=1}^{n} (y_i - \widehat{y_i})^2\right] \\ &= \left(y - X\hat{\beta}\right)'\left(y - X\hat{\beta}\right) \\ &= Y'Y - Y'X\hat{\beta} - \beta'X'Y + \hat{\beta}'X'X\hat{\beta} \end{aligned} \tag{6}$$

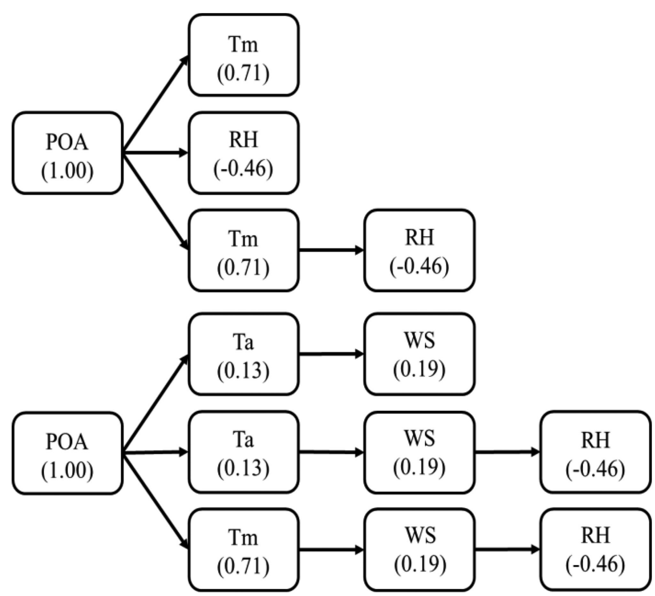

Fig. 5. Process flow for selecting predictive model variables through correlation analysis.

$$\frac{\partial L}{\partial \hat{\beta}} = -2X'Y' + 2X'X\hat{\beta} = 0 \tag{7}$$

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{8}$$

At this point, the row equation is solved by setting the partial differential of the sum of squared errors L with respect to the coefficient to zero, as shown in (8). This process results in an estimate of the regression coefficient β [33]. The coefficient of determination is generally used for linear regression. For

TABLE V
PREDICTION MODELS FOR PV POWER USING ENVIRONMENTAL VARIABLES

Model	Equations
P_1	$POA(\beta_1 + \beta_2 T_m)$
P_2	$POA[\beta_1 + POA(\beta_2 RH)]$
P_3	$POA[\beta_1 + \beta_2 T_m + POA(\beta_3 RH)]$
P_4	$POA[\beta_1 + POA(\beta_2 T_a + \beta_3 WS)]$
P_5	$POA[\beta_1 + POA(\beta_2 T_a + \beta_3 WS + \beta_4 RH)]$
P_6	$POA[\beta_1 + \beta_2 T_m + POA(\beta_3 WS + \beta_4 RH)]$

univariate simple linear regression models, i.e., with just one independent variable and one dependent variable, the coefficient of determination is the square of the Pearson correlation coefficient between these variables. However, for univariate multiple linear regression models, the coefficient of determination is defined as follows.

1) The ratio of the variance of the errors in a regression model to the total variance of the dependent variables.
2) The square of the correlation coefficient between the observed dependent variables and the regression model

$$e_i = y_i - \widehat{y_i} = (y_i - \bar{y}) - (\widehat{y_i} - \bar{y}) \tag{9}$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\widehat{y_i} - \bar{y})^2 + \sum_{i=1}^{n} e_i{}^2$$

$$= \mathrm{SSTO} = \mathrm{SSR} + \mathrm{SSE} \tag{10}$$

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SSTO}} = \frac{\sum_{i=1}^{n} (\widehat{y_i} - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}. \tag{11}$$

In (9)–(11), SSTO is the total sum of squares. The greater the variation of $\widehat{y_i}$, the greater is the SSTO. SSE denotes the sum of squared errors, and SSR is the regression sum of squares. The coefficient of determination is calculated from (11) [34].

IV. Model Validation

The absolute average deviation (AAD), RMSE, and mean absolute percentage error (MAPE) were calculated to evaluate the accuracy of the predicted model. The formulas for calculating these errors are given in (12)–(14). For MAPE, accuracy is typically expressed as a percentage, and the disadvantage in this form is that there is no upper bound for error rates if the prediction is too high. The RMSE analysis is proportional to the squares of the error and as such, is sensitive to outliers. The AAD uses the same scale as the data being evaluated and is commonly used for error forecasting in time-series analysis. To complement the shortcomings of each error analysis method, the model prediction accuracies estimated by the three methods were assessed jointly. In particular, the coefficients of determination, coefficients β_i, and errors for the regression models were calculated and compared. Table VI summarizes these metrics

TABLE VI
REGRESSION COEFFICIENT VALUES OF THE PREDICTION MODELS

	β_1	β_2	β_3	β_4
P_1	0.2592	-3.922e-04	-	-
P_2	0.2393	-3.445e-08	-	-
P_3	0.2569	-4.216e-04	1.096e-07	-
P_4	0.2456	-6.670e-07	3.750e-06	-
P_5	0.2432	-6.914e-07	3.749e-06	7.737e-08
P_6	0.2552	-4.173e-04	2.221e-06	1.042e-07

TABLE VII
COEFFICIENT OF DETERMINATION AND ERRORS OF THE PREDICTION MODELS

	R^2	AAD [W]	RMSE [%]	MAPE [%]
P_1	0.9980	4.232	5.174	6.198
P_2	0.9975	4.784	5.768	5.668
P_3	0.9980	4.176	5.138	6.030
P_4	0.9982	3.964	4.976	5.597
P_5	0.9982	3.936	4.957	5.468
P_6	0.9981	4.541	5.429	6.060

$$\mathrm{MAPE} = \sum_{i=1}^{n} |y_i - \widehat{y_i}/y_i| \text{x } 100\% \tag{12}$$

$$\mathrm{RMSE} = 100\% \text{ x} \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} (y_i - \widehat{y_i})^2} \Big/ \frac{1}{n}\sum\nolimits_{i=1}^{n} y_i \tag{13}$$

$$\mathrm{AAD} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|. \tag{14}$$

Tables VI and VII show the model errors for the predicted power generation and the coefficients for each model. Model P_1, which is defined by POA and T_m, has a lower AAD (4.23 W) and RMSE (5.17%) than model P_2, which uses POA and RH, but the RMSE is 5.77%. This result indicates that model P_2 will produce more prediction outliers than model P_1.

Model P_3, which includes POA, T_m, and RH, has similar errors (with AAD of 4.18 W, RMSE of 5.14%, and MAPE of 6.03%) as model P_1 . Model P_4, which uses POA, T_a, and WS, has lower errors (with AAD of 3.96 W, RMSE of 4.98%, and MAPE of 5.60%) than model P_3. Model P_6, which comprises all the environmental variables except T_a, has higher errors than model P_5. Model P_2 has the lowest coefficient of determination but the highest AAD and RMSE, whereas models P_4 and P_5

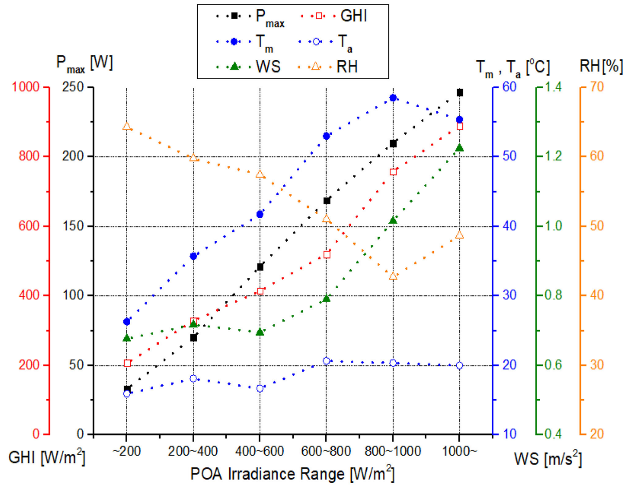

Fig. 6. Average of measured environmental variables and PV powers according to the range of POA irradiation.

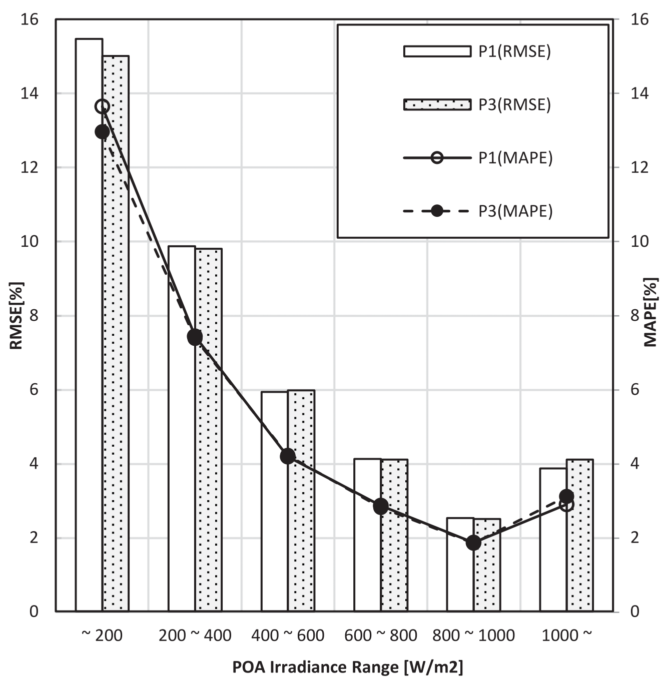

Fig. 7. Comparison of errors for models $\boldsymbol{P}_1$ and $\boldsymbol{P}_3$ by POA irradiance range.

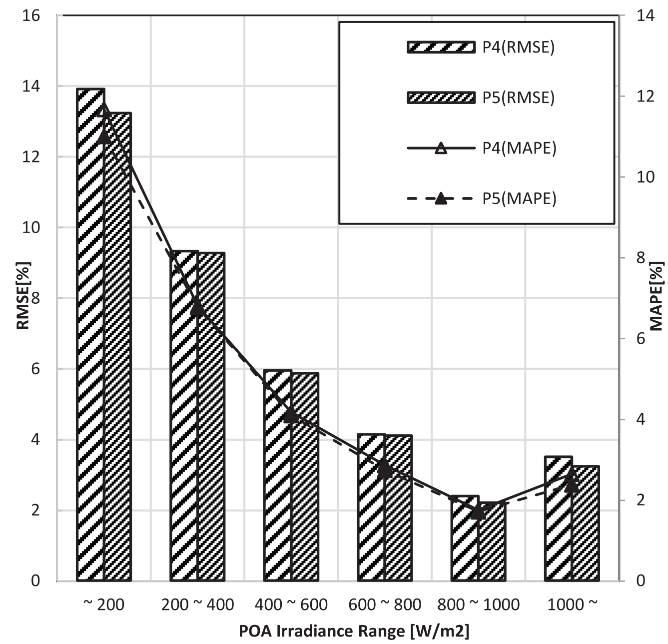

Fig. 8. Comparison of errors for models P_4 and P_5 by POA irradiance range.

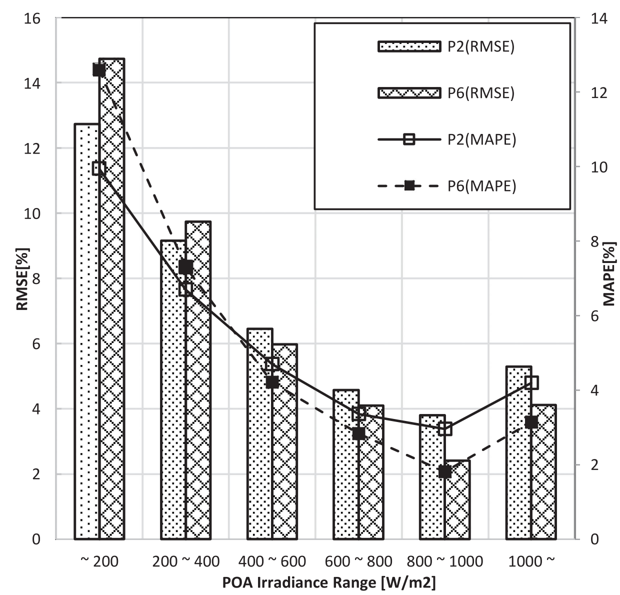

Fig. 9. Comparison of errors for models P_2 and P_6 by POA irradiance range.

have higher coefficients of determination. The model with the smallest AAD, RMSE, and MAPE over the entire monitored period was P_5. Overall, the error values and coefficients of determination indicate that model P_5 fits the measured values the most.

A. Analysis of Error Rates Over POA Irradiance Range

To verify the accuracy of the predictive model according to the measured environmental variables, the POA irradiance range was analyzed, as shown in Fig. 6. All the variables, except RH, are directly proportional to $P_{\max}$, which is consistent with the results of the correlation analysis in Section II. The values of T_a range between 16 and 20 °C over the POA irradiance range and have the smallest correlation with $P_{\max}$. Over the period of data collection, RH decreased as POA increased, but for POA values above 1000 W/m^2, RH increased. WS also rose sharply for POA values above 800 W/m^2.

In Figs. 7–9, the bars represent the RMSE and the lines represent the MAPE of the different models. The errors for models P_1 and P_3 (see Fig. 7) are similar over the POA irradiance range of $200-1000$ W/m^2 (note that the only difference between these models is the addition of RH in model P_3). On the other hand, model P_3 shows smaller errors for POA values less than 200 W/m^2. However, for POA greater than 1000 W/m^2, the errors increase. These results indicate that model P_3 is overfitted

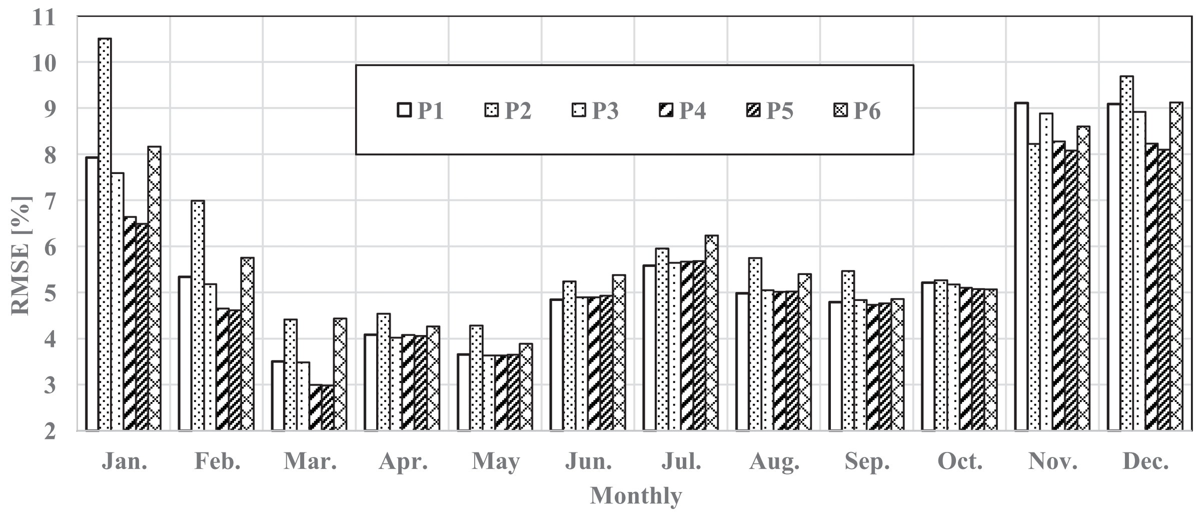

Fig. 10. RMSEs for monthly predicted PV power generation.

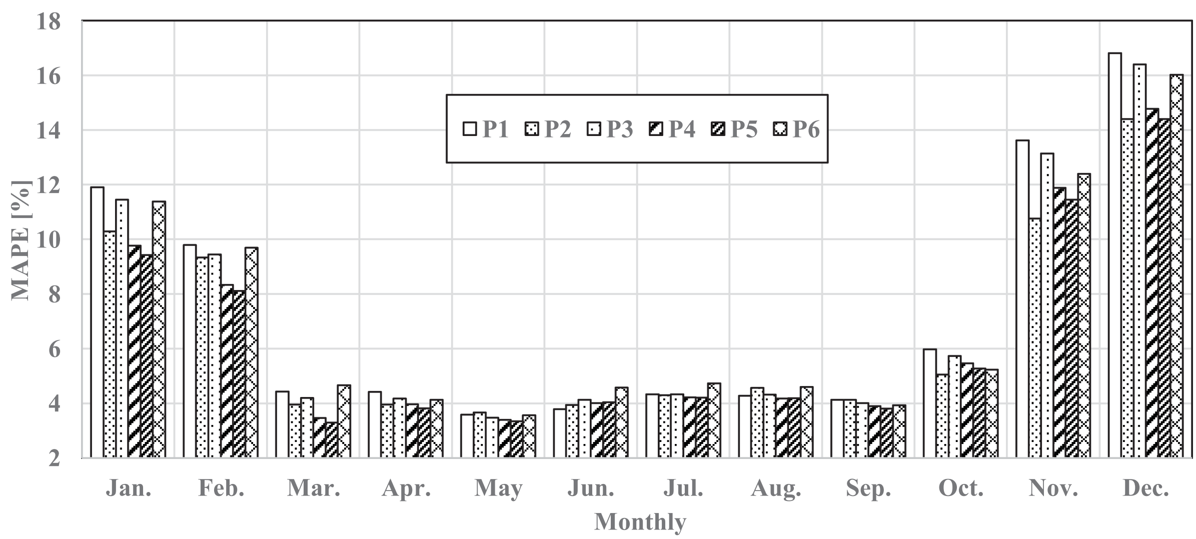

Fig. 11. MAPEs for monthly predicted PV power generation.

to increase incongruity of the model due to the low humidity over the range of 1000 W/m^2. Fig. 8 shows errors for model P_4 (which includes POA, T_a, and WS) and model P_5 (which adds RH to model P_4). It is notable that model P_5 has less errors than model P_4 over the entire range of irradiance. Fig. 9 shows the error graphs for model P_2 (which contains POA and RH) and model P_6 (which adds WS and RH to model P_1). Model P_2 has significantly low errors in the range below 200 W/m^2, but shows higher errors than model P_6 in the other ranges. This is because model P_2, which adds RH to POA, shows smaller error in the range below 200 W/m^2, due to the strong correlation between humidity and power under the low irradiance condition. On the other hand, P_6 has lower errors over the high irradiance range owing to the exclusion of humidity. From these results, it is important to note that the inclusion of all environmental variables with higher accuracies and correlations would be critical in predicting PV power.

B. Analysis of Error Rates by Month

Figs. 10 and 11 show the monthly error rates for each predictive model over a period of two years. These error estimates are considered in conjunction with the weather factors at the test site, as shown in Table IV. The RMSE and MAPE estimates show the highest error rates from November to February. Being the winter season, high humidity, low irradiation, and low temperature values were measured at the test site during this period. During the winter months, the model (P_5) that include POA, T_a, WS, and RH have relatively lower errors than models P_1, P_3, and P_6. Model P_2 has a lower MAPE but generally has a higher RMSE (the highest occurs in January). This shows that model P_2, with only humidity added, contains many outliers. The period from March to May show the lowest errors; this period has high irradiance and wind speed measurements. The error analysis shows that when POA and T_m are high, the model errors are lower, regardless of the values of T_a or WS, both of which have low correlations with $P_{\max}$. The correlation models that include T_a, WS, and RH have lower error values than the other models. Therefore, the combination of these variables in model P_5 may produce the lowest prediction error.

V. Conclusion

In this study, six predictive models for PV power generation were developed by analyzing the correlation between a set of weather variables (POA, T_m, $T_{\rm a}$, WS, and RH) and the power generation ($P_{\max}$) of the PV modules installed on a rooftop. The model equations composed of POA and T_m, which were highly correlated with $P_{\max}$, produced a large margin of error compared with those composed of less-correlated T_a, WS, and RH. This indicates an increase in prediction error caused by increasing number of variables of weather factors that have high correlation coefficients. Model P_5 (which included RH) generally produced a lower error value than model P_4 (that was composed of POA, T_a, and WS). The errors produced by model P_5 (RMSE of 4.957% and MAPE of 5.468%) were the lowest during the measurement period and over the entire range of irradiation. The results also indicated that prediction errors decreased in the winter months when RH was included in the model. It was also noted that the inclusion of RH produced low errors, regardless of how T_a, WS, or T_m were included in the model. This result, for the first time, indicates the importance of RH to the accuracy of the model predictions. In particular, the model composed of RH, T_a, and WS gave lower errors over the entire range of measurements than the model composed of RH and T_m, which is the model generally used in forecasting PV power in terms of the module temperature. From the results, it is very noticeable that humidity factor should be combined with other environmental parameters for the better accuracy of power prediction especially at low ambient temperature, low irradiance, and high humidity environments such as near lakes and marines.

Future studies will examine the correlation between the weather factors and power generation for large-scale PV plants by accounting for the effects of sea fog on floating photovoltaics and marine photovoltaics.

Acknowledgment

To the memory of Dr. S. Ryu.

References

[1] "Snapshot of global photovoltaic markets," Int. Energy Agency, Paris, France, Rep. IEA PVPS T1-33:2018, 2018.

[2] N. H. Reich, B. Mueller, A. Armbruster, W. G. Sark, K. Kiefer, and C. Reise, "Performance ratio revisited: Is pr > 90% realistic?" *Prog. Photovolt., Res. Appl.*, vol. 20, pp. 717–726, 2012.

[3] J. A. Duffie, and W. A. Bechman, *Solar Engineering of Thermal Processes*, 3rd ed. Hoboken, NJ, USA: Wiley, 2006

[4] D. Myers, "Evaluation of the performance of the PVUSA rating methodology applied to dual junction technology," Nat. Renew. Energy Lab., Golden, CO, USA, NREL/CP-550-45376, 2009.

[5] B. Kroposki, K. Emery, D. Myers, and L. Mrig, "A comparison of photovoltaic module performance evaluation methodologies for energy ratings," in *Proc. IEEE 1st World Conf. Photovolt. Energy Convers.*, 1994, pp. 858–862.

[6] R. W Taylor, "System and module rating: Advertised versus actual capability," *Sol. Cells*, vol. 18, pp. 335–344, 1986.

[7] *Standard Tables for Terrestrial Solar Spectral lrradiance at Air Mass 1.5 for a 37 Tilted Surface*, ASTM E 892, 1992.

[8] PVUSA Technical Specifications, 19271-EMT-3, Aug. 1991.

[9] *Standard Test Methods Electrical Performance of Nonconcentrator Terrestrial Photovoltaic Modules and Arrays Using Reference Cells*, ASTM E 1036, 1996.

[10] S. Nann and K. Emery, "A numerical analysis of PV-rating methods," in *Proc. Conf. Rec. 22nd IEEE Photovolt. Spec. Conf.*, Oct. 1991, pp. 789–795.

[11] E. Gianolii-Rossi and K. Krebs, "Energy rating of PV modules by outdoor response analysis," in *Proc. PV Sol. Energy Conf.*, Florence, Italy, 1988, pp. 509–514.

[12] B. K. Farmer, "PVUSA model technical specification for a turnkey photovoltaic power system," Univ. North Texas, Denton, TX, USA, Appendix C, p. c2, 1992.

[13] J. I. Rosell and M. Ibanez, "Modelling power output in photovoltaic modules for outdoor operating conditions," *Energy Convers. Manage.*, vol. 47, pp. 2424–2430, 2006.

[14] H. Yang, J. Burnett, and J. Ji, "Simple approach to cooling load component calculation through PV walls," *Energy Buildings*, vol. 31, pp. 285–290, 2000.

[15] C. R. Osterwald, "Translation of device performance measurements to reference conditions," *Sol. Cells*, vol. 18, pp. 269–279, 1986.

[16] M. Fuentes, G. Nofuentes, J. Aguilera, D. L. Talavera, and M. Castro, "Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in mediterranean climates," *Sol. Energy*, vol. 81, pp. 1396–1408, 2007.

[17] T. Huld *et al.*, "A power-rating model for crystalline silicon PV modules," *Sol. Energy Mater. Sol. Cells* vol. 95, pp. 3359–3369, 2011.

[18] D. L. King, J. A. Kratochvil, W. E. Boyson, and W. I. Bower, "Field experience with a new performance characterization procedure for photovoltaic arrays," in *Proc. 2nd World Conf. Exhib. Photovolt. Sol. Energy Convers.*, Vienna, Austria, 1998, pp. 1947–1952. .

[19] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Photovoltaic array performance model," Sandia Nat. Lab., Albuquerque, NM, USA, SAND2004-3535, 2004.

[20] C. Spearman, "The proof and measurement of association between two things," *Amer. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904.

[21] L. I-Kuei Lin "Concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, no. 1 , pp. 255–268, Mar., 1989.

[22] D. J. Best and D. E. Roberts, "Algorithm AS 89: The upper tail probabilities of Spearman's ρ," *J. Roy. Statist. Ser. C (Appl. Statist.)*, vol. 24, pp. 377–379, 1975. doi:10.2307/2347111.

[23] W. Revelle, Psych v1.8.4, 2018. [Online]. Available: https://www.rdocumentation.org/packages/psych/versions/1.8.4/topics/pairs.panels

[24] E.W. S. Weisstein, "Rank correlation, coefficient," 1999. [Online]. Available: http://mathworld.wolfram.com/SpearmanRankCorrelationCoefficient.html

[25] C. R. Landau, "Optimum tilt of solar panels," 2014. [Online]. Available: http://www.solarpaneltilt.com. Accessed on: May 5, 2015.

[26] C. M. Whitaker, H. J. Wenger, A. Iliceto, G. Chimento, and F. Paletta, "Effects of irradiance and other factors on PV temperature coefficients," in *Proc. Conf. Rec. 22nd IEEE Photovolt. Spec. Conf.*, 1991, pp. 608–613.

[27] S. Mekhilef, R. Saidur, and M. Kakalisarvestani, "Effect of dust, humidity and air velocity on efficiency of photovoltaic cells," *Renewable Sustain. Energy Rev.*, vol. 16, pp. 2920–2925, 2012.

[28] H. A Kazem and M. T. Chaichan, "Effect of humidity on photovoltaic performance based on experimental study," *Int. J. Appl. Eng. Res.*, vol. 10, no. 23, pp. 43572–43577, 2015.

[29] M. Alshakhs, "Challenges of solar PV in Saudi Arabia," in PH240, Stanford Univ., Stanford, CA, USA, 2013.

[30] K. John, M. K. Kaldellis, and K. A. Kavadias, "Temperature and wind speed impact on the efficiency of PV installations. Experience obtained from outdoor measurements in Greece," *Renewable Energy*, vol. 66, pp. 612–624, 2014.

[31] C. Schwingshackl *et al.*, "Wind effect on PV module temperature: Analysis of different techniques for an accurate estimation," *Energy Procedia*, vol. 40, pp. 77–86, 2013.

[32] W. K. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*. New York, NY, USA: Springer, 2007.

[33] W. H. Kruskal and J. M. Tanur, Eds., "Linear hypotheses," in *International Encyclopedia of Statistics*. New York, NY, USA: Free Press, 1978, pp. 523–541.

[34] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th ed. Boston, MA, USA: Cengage Learning, 2011, pp. 508–510.

Gyu Gwang Kim received the B.S. degree in electrical engineering from Konkuk University, Seoul, South Korea, in 2016. He is currently working toward the M.S. degree at Next Generation Photovoltaic Module and Power System Research Center, Konkuk University.

He is working with Chungbuk TechnoPark, Chungbuk, South Korea. His research interests include the predictive models for PV module performance and power system evaluation for marine photovoltaics and microgrid network.

Jin Ho Choi received the B.S. degree in electrical engineering from Konkuk University, Seoul, South Korea, in 2018. He is currently working toward the M.S. degree at Next Generation Photovoltaic Module and Power System Research Center, Konkuk University.

His research interests include power performance of bifacial PV module for floating photovoltaics and marine photovoltaics.

So Young Park received the B.S. degree in electrical engineering from Konkuk University, Seoul, South Korea, in 2018. She is currently working toward the M.S. degree at Konkuk University.

Her research interests include the issue of performance evaluation of MPVs and BIPVs.

Byeong Gwan Bhang received the B.S. degree in electrical engineering from Konkuk University, Seoul, South Korea, in 2017. He is currently working toward the M.S. degree at Konkuk University.

His research interests include utility of floating photovoltaics, marine photovoltaics, and their operations and maintenance using big data analysis.

Woo Jun Nam received the B.S. degree in information environmental engineering from Pukyong National University, Pusan, South Korea, in 2002, and the M.S. degree in electrical engineering from Konkuk University, Seoul, South Korea, in 2015. He is currently working toward the Ph.D. degree at Konkuk University.

He is currently working with PV Technical Center at CBTP (Chungbuk TechnoPark), Chungbuk, South Korea. His research interests include reliability of PV module and power station.

Hae Lim Cha received the B.S. and M.S. degrees in electrical engineering from Konkuk University, Seoul, South Korea, in 2015 and 2017, respectively. She is currently working toward the Ph.D. degree at Next Generation Photovoltaic Module and Power System Research Center, Konkuk university.

Her research interests include convergence system design of PVs and ESS. She is scheduled to work at KT (Korea Telecom) in 2018.

NeungSoo Park (M’03) received the B.S. and M.S. degrees in electrical engineering from Yonsei University, Seoul, South Korea, in 1991 and 1993, respectively, and the M.S. degree in computer engineering and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 2002.

He is currently a Professor with the Department of Computer Science and Engineering, Konkuk University, Seoul. He was a Senior Engineer with Samsung Electronics Corporation, South Korea. His research interests include parallel computing, computer architecture, embedded system, high-performance computing for signal processing, and multi-media systems and power analysis focusing on renewable energies using AI technology.

Hyung-Keun Ahn (M’92) received the B.S. and M.S. degrees in MOSFETs from Yonsei University, Seoul, South Korea, and the Ph.D. degree in HEMTs from the University of Pittsburgh, Pittsburgh, PA, USA, in 1993.

From 1986 to 1995, he was with the LG Semiconductor Corporation, where he was involved in silicon-based device design and process integration. In 1995, he joined the Department of Electrical Engineering, Konkuk University, Seoul, where he is currently a Professor and the Dean with the Next Generation Photovoltaic Module and Power System Research Center. From 2014 to 2016, he was also a Foreign Professor with the Department of Electrical and Control Engineering, Division of EE, Shandong University of Technology, Zibo, China. His research interests include reliability of both Si- and GaAs-based solar cells and PV modules for applications in micro roof-top to large scale of PVs connected with ESS, focusing on failure analysis, reliability, and maintenance, repair, and operation for durable power system. He was the first National Photovoltaic Research and Development Program Director of the Ministry of Knowledge Economy, South Korea, from 2009 to 2011. He is currently working on microgrid for net-zero energy house and city using ESS, bifacial rooftop, floating and marine PVs, geothermal, and fuel cell under the demand response control.

Dr. Ahn was a Committee Member of the National Science and Technology Commission, Energy Department, from 2005 to 2007.