Fault Detection for Photovoltaic Systems Using Multivariate Analysis With Electrical and Environmental Variables

B200041232 Teng Xuanzhi

Manuscript received July 13, 2020; revised August 7, 2020 and September 10, 2020; accepted October 16, 2020. Date of publication November 4, 2020; date of current version December 21, 2020. This work was supported in part by the New and Renewable Energy Technology Program of the Korea Institute of Energy, Technology, Evaluation, and Planning, and in part by the Ministry of Trade, Industry, and Energy, Republic of Korea, under Grant 20183010014260. *(Gyu Gwang Kim and Wonbin Lee are co-first authors.) (Corresponding author: Hyung-Keun Ahn.)* The authors are with the Next Generation Photovoltaic Module and Power System Research Center, Konkuk University, Seoul 05029, South Korea (e-mail: rbrhkd00@konkuk.ac.kr; dnjsqls6766@konkuk.ac.kr; shorev@konkuk.ac.kr; bbk0627@konkuk.ac.kr; hkahn@konkuk.ac.kr). Digital Object Identifier 10.1109/JPHOTOV.2020.3032974

Abstract: Fault detection and repair of the components of photovoltaic (PV) systems are essential to avoid economic losses and facility accidents, thereby ensuring reliable and safe systems. This article presents a method to detect faults in a PV system based on power ratio (PR), voltage ratio (VR), and current ratio (IR). The lower control limit (LCL) and upper control limit (UCL) of each ratio were defined using the data of a test site system under normal operating conditions. If PR exceeded the set range, the algorithm considered a fault. Subsequently, PR and IR were examined via the algorithm to diagnose faults in the system as series, parallel, or total faults. The results showed that PR exceeded the designated range between LCL (0.93) and UCL (1.02) by dropping to 0.91–0.68, 0.88–0.62, and 0.66–0.33 for series, total, and parallel faults, respectively. Moreover, VR exceeded the LCL (0.99) and UCL (1.01) by 0.95–0.69 and 0.91–0.62 for series and total faults, respectively, but not under parallel faults condition. IR did not change in series and total faults but exceeded the range of LCL (0.93) and UCL (1.05) by dropping to 0.66–0.33. Thus, faults in PV systems can be detected and diagnosed by analyzing quantitative output values.

Index Terms: Correlation coefficient, current ratio, electrical variables, fault detection algorithm, multivariate analysis, power ratio, regression analysis, voltage ratio, weather variables.

1. Introduction

THE significant developments in photovoltaic (PV) systems in China during the past three years have contributed to a global increase in the number of PV systems. In 2017, the top ten countries had an installed PV capacity of 344.5 GW [1]. However, owing to this increasing use of PV systems, the importance of system reliability and safety is also expanding [2]. Moreover, additional time and costs are incurred when failures or faults in the system are failed to be detected in time. Therefore, to maintain a high-quality system for extended periods, it is essential to calculate and analyze quantitative values to promptly identify the locations and times of faults and failures [3].

Furthermore, PV generation is influenced by various environmental factors, i.e., irradiance, ambient temperature (Ta), relative humidity (RH), and wind speed (WS)[4]. The relationship between these factors and the output of the PV system can be expressed using a correlation

coefficient. When the correlation coefficient is close to the absolute value of 1, it indicates a strong relationship between the variables, whereas a coefficient value close to 0 indicates a weak relationship between the two variables.

The maximum power output of the PV system can be achieved when incident irradiance enters the surface at a right angle [5]. Changes in surrounding environmental conditions, such as shading, are also among the main causes of power drops in PV systems [6]. When a PV system uses a bifacial PV module, the shading conditions of the rear side also need to be considered to avoid undesired power drops [7], [8]. When the global horizontal irradiance (GHI) and the plane of array (POA) irradiance were compared, POA showed a stronger correlation with the output power of the PV module than GHI, as shown in our previous work [9].

A PV system is greatly influenced by changes in the temperature. As the temperature of the PV module rises, the output power decreases. Generally, when the temperature of a crystalline silicon PV module increases by 1, the output power of the module decreases by 0.35%–0.4%[1210 12]. In a floating PV system, the operating temperature of the PV module is lower than that of a land-based system; therefore, a floating PV system can generate 10% more energy [13]. The correlation coefficient between the output power of the PV module and temperature is 0.71 for the module temperature and 0.13 for the ambient temperature. According to theoretical analyses, the coefficient between the temperature and output power should be negative; however, according to a correlation analysis, the temperature rises as irradiance rises, resulting in positive correlation coeffi- cients, also shown in our previous work[9].

Similarly, the RH affects PV systems in a similar fashion as dust accumulation affects the output power of the PV system. Moreover, water vapor particles in the air reduce the amount of irradiance, and light rays hitting these water droplets are scattered via refraction, reflection, or diffraction [14]. The correlation coefficient between the output power and RH is 0.46, which indicates that when irradiance at the installation site increases, the RH also decreases, resulting in increased output power of the system[15].

Finally, the correlation coefficient between the WS and the output power of the PV system is 0.19, which indicates a weak relationship. Nevertheless, it is still a positive value because wind cools the surface of the PV module, leading to an increase in the output power of the PV system [16]. In this study, a half-cut 36-cell PV module that produces one-sixth of the value of a 72-cell crystalline silicon PV module was used to design a PV system. The data of irradiance, module temperature, ambient temperature, RH, and WS from a test site were collected to model and estimate the output power, voltage, and current in the system. Calculations were performed to detect and diagnose faults in the PV system using the proposed algorithm based on output data under normal operating conditions. This algorithm can be particularly essential for ground, floating, and/or marine-based PV systems, where installation conditions are sensitive to sudden climatic changes.

2. FAULT DETECTION AND DIAGNOSIS OF PV SYSTEM

2.1. Performance Ratio (PR)

The PR is a general evaluation method for a PV generation system and calculates the ratio between the theoretical output and actual output as [17]

$$PR = Y_p/Y_r \tag{1}$$

The numerator and denominator of the right-hand-side term of (1) each represent the power rate of the PV system and the theoretical yield when incident irradiation hits the PV system, respectively. These are presented as follows:

$$Y_p = E_{P,d}/P_{A,s} \tag{2}$$

$$Y_r = H_{A,d}/G_{STC} \tag{3}$$

In (1), Yp represents the yield of the PV system, which is the rate of generated power with units in [kWh · *d*1 · (kW)1] or [h·*d*1]. *Ep,d* is the total cumulative energy generated by the PV generation system, and its unit is [kWh/d]. *PA,s* is the rated output of the PV system, and its unit is kilowatt [kW]. All variables capable of affecting the output of the PV system, such as the power loss, from the increasing resistance owing to multiple connections of modules in PV arrays or changes in output owing to changes in operating temperature are considered in the output value that is processed through the inverter, which is the final stage, and are included in *Yp.* The reference yield is expressed as *Yr,* and its unit is [kWh · *d*1 · (kW)1] or [h·*d*1]. *HA,d* is the total cumulative solar irradiation for a day, and its unit is [kWh/m2]. *G*STC is a day's worth of solar irradiation according to standard testing conditions (STC), and its unit is [d × 1 kW/m2]. Thus, *Yr* is the ratio of the actual cumulative solar irradiance to the STC solar irradiance. Since a PV system is installed outdoors, it is affected by various factors, and the temperature is one of the factors with the greatest impact. When the temperature rises by 1 ℃, the power decreases by a certain percentage, which is known as the temperature coefficient. If the power ratio is calculated without considering the temperature coefficient, errors may arise[18].

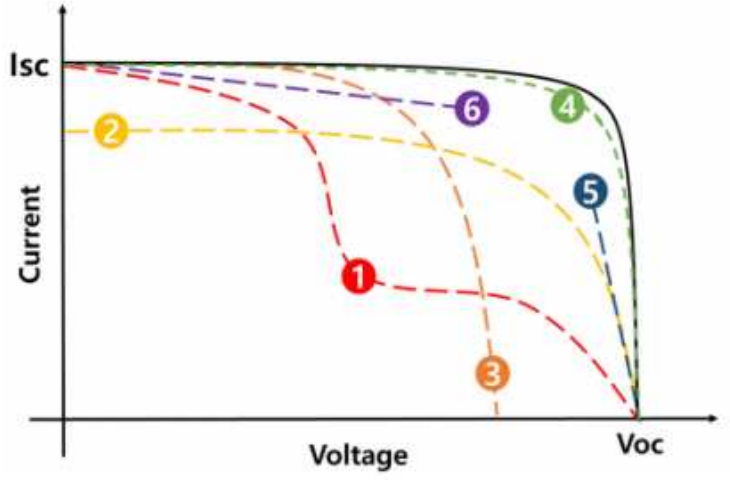

Fig. 1. *I–V* curves under different PV module fault conditions.

2.2. I–V Curve

The *I–V* curve of a PV module is represented by a graph of current and voltage. When the voltage of the PV module is 0, a short-circuit current (*I*sc) is generated, owing to the generation and collection of photogenerated carriers. In an ideal case where the power loss from the resistance is same as the ideal calculation, *I*sc and the light-induced current are identical. *I*sc is the maximum current available from the PV module, and the open-circuit voltage (*V*oc) is the maximum voltage available from the PV module. The fill factor (FF) is an element that determines the maximum power. The *I–V* curve of a PV system can be categorized into six types according to each fault condition. The *I–V* curve of PV module under the normal condition is calculated and visualized via a one-diode equation. The detection and diagnosis of faults in the PV module could be done by detecting different changes in *I–V* curve according to different types of faults. One method of using *I–V* curve to detect fault was studied in[19] by applying neurofuzzy fault detection method. Another method of using *I–V* curve to detect faults in the PV module was studied by Tingting and Xiaohong[**?**] where they used MATLAB and Simulink to make a hybrid model with variables, such as the number of PV modules, *V*oc, and *I*sc at maximum power point condition.

1) Mismatch: PV modules in a PV array are connected in either series or parallel. When a string, which refers to several PV modules connected in series, is connected to other strings in series and parallel, multiple dysfunctions can take place. One possible dysfunction is that the current produced from one of the strings could be lower than that of the other strings. This situation can occur when PV modules connected in strings are not capable of generating electricity in the normal way because the PV module has been somehow polluted or damaged. When such mismatching has occurred, the *I–V* curve of the system has the shape of curve 1 in Fig. 1.
2) Uniform Soiling: The mismatching of modules in strings or arrays usually occurs because of differences in generated electricity. Curve 2 of Fig. 1 is a case where the front surface of a PV system has been generally contaminated or plenty of time has passed since installation, resulting in decreasing generation.
3) Bypass Diode: When the surface of the PV module is soiled or shaded, PV cells in that particular region are unable to generate the same amount of electricity as normal cells. In such a fault condition, the voltage generated from other normal cells would be cast to the defected cells, resulting in hotspots or other malfunctions. A bypass diode is installed in the PV module to minimize the damage caused in such a situation. When the bypass diode is functioning, currents generated from normal cells flow through the bypass diode instead of the defective cells. When the bypass diode is in operation, the open-circuit voltage decreases, resulting in an *I–V* curve in the form of curves 1 and 3 in Fig. 1.

2.2.1. Mismatch:

PV modules in a PV array are connected in either series or parallel. When a string, which refers to several PV modules connected in series, is connected to other strings in series and parallel, multiple dysfunctions can take place. One possible dysfunction is that the current produced from one of the strings could be lower than that of the other strings. This situation can occur when PV modules connected in strings are not capable of generating electricity in the normal way because the PV module has been somehow polluted or damaged. When such mismatching has occurred, the *I–V* curve of the system has the shape of curve 1in Fig. 1.

2.2.2. Uniform Soiling:

The mismatching of modules in strings or arrays usually occurs because of differences in generated electricity. Curve 2 of Fig. 1 is a case where the front surface of a PV system has been generally contaminated or plenty of time has passed since installation, resulting in decreasing generation.

2.2.3. Bypass Diode:

When the surface of the PV module is soiled or shaded, PV cells in that particular region are unable to generate the same amount of electricity as normal cells. In such a fault condition, the voltage generated from other normal cells would be cast to the defected cells, resulting in hotspots or other malfunctions. A bypass diode is installed in the PV module to minimize the damage caused in such a situation. When the bypass diode is functioning, currents generated from normal cells flow through the bypass diode instead of the defective cells. When the bypass diode is in operation, the open-circuit voltage decreases, resulting in an *I–V* curve in the form of curves 1 and 3in Fig. 1.

2.2.4. Unknown Knee:

Curve 4 of Fig. 1 is an *I–V* curve of an unknown knee, which is the case where degradation occurs with no clear reason. Such a case requires significant time and general inspections throughout the entire PV system.

TABLE I
MATH SPACINGS USED BY LATEX

Size	Width	Cmd.	Used for	Example
small	1/6 em	\,	symbols	ab
medium	2/9 em	\:	binary operators	$a+b$
large	5/18 em	\;	relational operators	$a=b$
negative small	−1/6 em	\!	misc. uses	ab

2.2.5. Series Resistance:

There are various sources of series resistance in PV cells and modules. The contact resistance between the metal contact and silicon of a PV cell or mismatch between the PV modules can increase the series resistance. A major influence on the series resistance of a PV system is the reduction of the FF. However, if the series resistance is high enough, it could also result in a reduction of the short-circuit current. The series resistance does not affect the PV cell at the open-circuit voltage because the overall current flows through the PV cell, resulting in the series resistance of 0. By contrast, around the open-circuit voltage, the series resistance greatly affects the *I–V* curve. The *I–V* curve of a PV module as the series resistance increases could be drawn in a similar form as curve 5in Fig. 1. The series resistance can be approximated as the slope of the *I–V* curve at the open-circuit voltage point.

2.2.6. Design Factor:

The design factor is an analysis method that uses the operation time from the time of incident irradiation at the surface of the PV system to the time of the ac output generation of the inverter. This method is similar to the PR method except that the design factor is capable of detecting faults in each part of the PV system through the yield. This is represented by quantified values derived from analyzing each section of the PV system. The reference yield of solar irradiation (Yr), maximum PV array yield (Yam), temperature-corrected PV array yield (Yat), optimized PV array yield (Yao), PV array yield (Ya), and PV system yield (Yf) are defined as follows:

$$Y_r G_{a,meas}/G_{ref} \tag{4}$$

$$Y_{am} a_m \cdot Y_r \tag{5}$$

2.3. Exponentially Weighted Moving Average (EWMA)

EWMA is an effective method to detect small changes in the process mean with no concern for the size of the measured value. The moving average refers to a method of attaining more weight to new data, whereas the past data are given less weight exponentially, and the total sum of the weight is 1. T

By determining the upper and lower limits of the control range, abnormalities can easily be detected. However, since the control range greatly depends on the weight, determining the proper weight is crucial

he initial output data of the PV system, which were considered during the normal operation. Therefore, UCL and LCL would be defined as different values for different testing sites.

Three types of fault conditions in the PV system were designed: series, parallel, and total. These are shown in Fig. 15. The series and parallel faults were defined in Section III-C. The total fault is a case where both voltage and current of PV array have dropped due to complex reasons, such as open-/short-circuited problem and shading. In this study, the total fault condition has been emulated by decreasing the current and voltage of the array. Current is decreased by covering the surface of the PV module with mesh, and voltage is decreased by increasing the number of short-circuited PV modules. The series fault was emulated by short circuiting a faulty module and

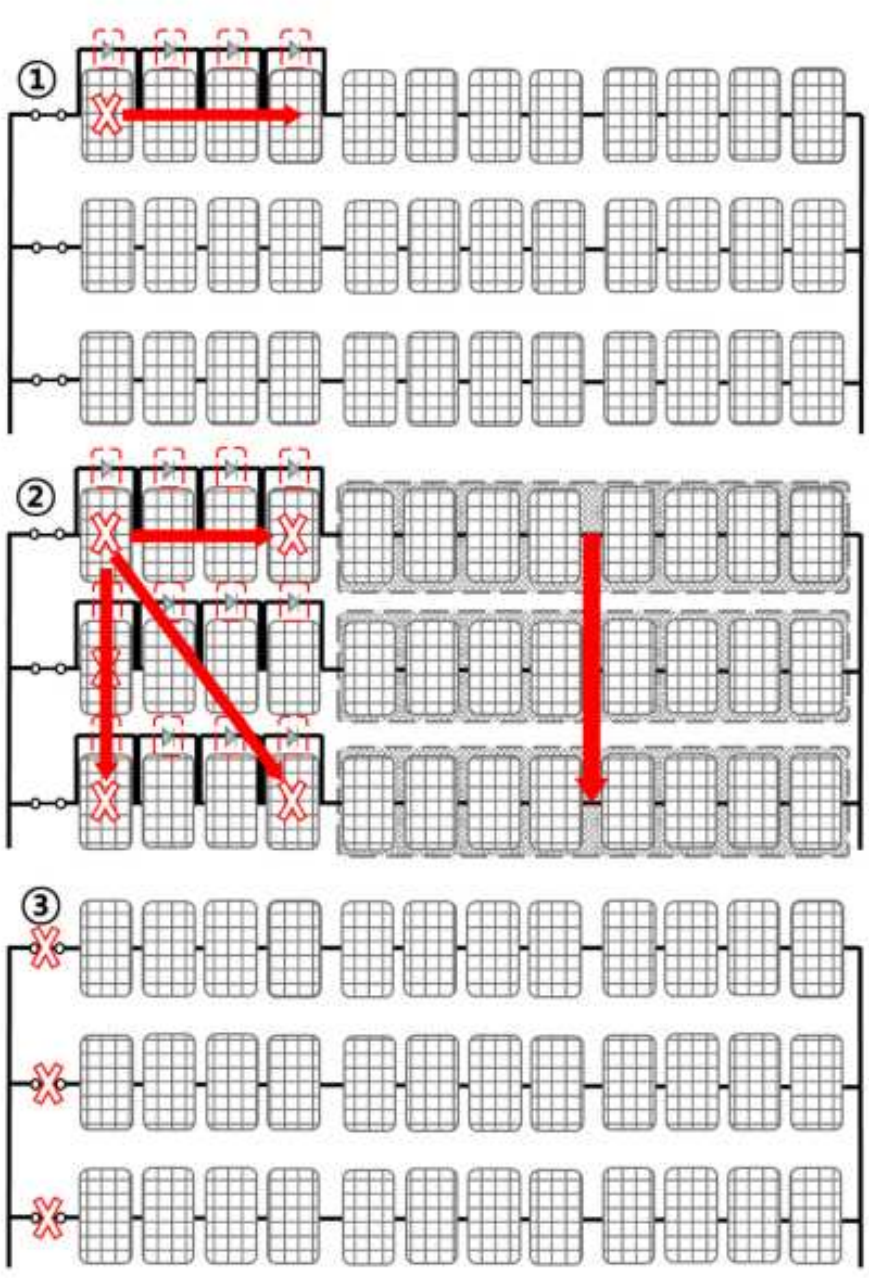

Fig. 2. Types of fault under test conditions (1: Series fault, 2: total fault, and 3: parallel fault).

then activating a bypass diode. The parallel fault was emulated by increasing the number of the open-circuited sting.

TABLE II
SPECIFICATIONS FOR PV MODULE

Range	$\Omega(m)$
$x < 0$	$\Omega(m) = \sum_{i=0}^{m} K^{-i}$
$x \geq 0$	$\Omega(m) = \sqrt{m}$

3. MATHEMATICAL MODELS OF PV SYSTEM

3.1. Correlation Analysis Between the PV System and Environmental Variables

Correlation analysis is a method of elucidating the interrelationship between two variables. Generally, correlation analysis can be divided into two main groups: Spearman and Pearson. In a Spearman analysis, the order of strength in the interrelationship between the two variables is analyzed . However, in a Pearson analysis, the correlation between the variables is analyzed based on the linear relationship between the two variables and expressed as a number between 1 and +1. When the two variables are in a proportional relationship, the coefficients are positive, where a number closer to 1 indicates a stronger relationship. By contrast, when two variables are in an inverse proportional relationship, the coefficients are negative

In this study, a Pearson correlation analysis was chosen to analyze the interrelationship between the output of the PV system (voltage, current, and power at the maximum power point) and the environmental variables (irradiance, module temperature *Tm*, *Ta,* WS, and RH)

4. TESTING SITE

An outdoor testing site was designed to verify the fault detection/diagnosis algorithm, as presented in Section III, and is shown in Fig. 14. The PV system was composed of 36 PV modules with the characteristics of 50 W (18 V and 2.77 A). A total of 12 modules are connected in series forming a string and three strings are connected in parallel to form an array. The specifications of the PV modules used in the test are listed in Table II. The geographical coordinates of the test site location were 36° 54 08.3 N and 127° 32 26.4 E. It was installed facing south, and the installation orientation was 30°. The data were collected for a week to determine the coefficients for the PV output prediction model. To enhance the accuracy of the model, data with less than 100 W/m2 of irradiation were excluded from the analysis.

The UCL and LCL of *PR*, *VR*, and *IR* were defined using the initial output data of the PV system, which were considered during the normal operation. Therefore, UCL and LCL would be defined as different values for different testing sites.

Three types of fault conditions in the PV system were designed: series, parallel, and total. These are shown in Fig. 15. The series and parallel faults were defined in Section III-C. The total fault is a case where both voltage and current of PV array have dropped due to complex reasons, such as open-/short-circuited problem and shading. In this study, the total fault condition has been emulated by decreasing the current and voltage of the array. Current is decreased by covering the surface of the PV module with mesh, and voltage is decreased by increasing the number of short-circuited PV modules. The series fault was emulated by short circuiting a faulty module and then activating a bypass diode. The parallel fault was emulated by increasing the number of the open-circuited sting.

Theorem 1 (Einstein-Podolsky-Rosenberg):

The back-propagation distance is determined by calculating the Fresnel zone plate (FZP) focal distance for the specific hologram geometry. For the specific geometry employed in this experiment, the FZP focal length is approximately the distance between the object and the recording medium.

The digital images of the holograms processed with the Fresnel propagation code generated the reconstructed images shown in Fig. **??**.

Holograms recorded in such a fashion can not be reconstructed in the conventional way with an optical reconstruction beam. In order to numerically reconstruct the holograms, the surface modulation was digitized with a Novascan atomic force microscope operated in tapping mode.

Lemma 1:
The back-propagation distance is determined by calculating the Fresnel zone plate (FZP) focal distance for the specific hologram geometry For the specific geometry employed in this experiment, the FZP focal length is approximately the distance between the object and the recording medium. The digital images of the holograms processed with the Fresnel propagation code generated the reconstructed images shown in Fig. **??**.

Holograms recorded in such a fashion can not be reconstructed in the conventional way with an optical reconstruction beam. In order to numerically reconstruct the holograms.

Proof:
The back-propagation distance is determined by calculating the Fresnel zone plate (FZP) focal distance for the specific hologram geometry. For the specific geometry employed in this experiment, the FZP focal length is approximately the distance between the object and the recording medium. The digital images of the holograms processed with the Fresnel propagation code generated the reconstructed images shown in Fig. **??**. ■

5. Conclusions

This study presented a method to predict the output of a PV system using environmental variables and subsequently detect and diagnose faults in the system. The normal running condition of a PV system was defined from a regression model by utilizing adapted environmental variables.
There are various causes of faults in PV systems; these causes can be categorized as series, parallel, and total faults that eventually lead to a decrease in the electrical output of the system. An output model was generated using the data of environmental variables collected from a test site. The accuracy of the *Vmp* prediction model was 0.81 MAPE (%) and 2.26 RMSE (%). The accuracy of the *Imp* prediction model was 2.17 MAPE (%) and 0.083 RMSE (%). A prediction model of *Pmp* was designed by multiplying the predicted value of *Vmp* and *Imp* with the smallest error rate, thus achieving the accuracies of 2.29 MAPE (%) and 18.67 RMSE (%). The modeled output data of the PV system were compared with the actual output data of a PV system collected from the same test site. *PR* was used as the criterion to detect faults in the PV systems. When a fault is caused in a parallel connection, the current in the circuit is affected; however, if the fault is caused in a series connection, the voltage in the circuit is affected. Based on these characteristics of the PV system, faults were diagnosed using *VR* and *IR*.
Using this research, it is possible to realize a secure power generation through the integrated management of PV systems. The proposed system can enable the owners of PV systems and companies in operations and maintenance to monitor the operating conditions of a system through quantitative data. Furthermore, faults in the PV system can be detected and diagnosed through basic output data of the PV system without requiring the installation of additional monitoring systems. The innovation of this work is that the proposed method is capable of detecting a fault and diagnosing the adapt maintenance method of the PV system, especially in cases of floating and marine PV system, which have a vast deviation in output power because of their sensitiveness to changes in environmental conditions. This algorithm can be particularly useful to monitor and maintain ground, floating, and marine-based PVs, wherein the installation conditions are sensitive to sudden climatic changes.

For future works, the presented method could be used for the diagnosis of PV power plants using general PV modules. In particular, this analysis method is planned to be applied to develop and diagnose faults in a testbed of PV power plant composed with a high-density power module, such as bifacial or shingled PV module, which is capable of generating higher power under low-temperature condition.

Acknowledgements

The authors wish to thank the anonymous reviewers for their valuable suggestions.

References

[1] "Snapshot of global photovoltaic markets," Int. Energy Agency, Paris,France, Rep. IEA PVPS T1-33:2018, 2018.
[2] Y. Mahmoud and E. F. El-Saadany, "Enhanced reconfiguration method for reducing mismatch losses in PV systems," *IEEE J. Photovolt.*, vol. 7, no. 6, pp. 1746–1754, Nov. 2017
[3] A. Walker, "PV O&M cost model and cost reduction," presented at Photovolt. Module Rel. Workshop, 2017. [Online]. Available: https://www.nrel. gov/docs/fy17osti/68023.pdf
[4] F. Touati, A. Massoud, J. A. Hamad, and S. A. Saeed, "Effects of environmental and climatic conditions on PV efficiency in Qatar," in *Proc. Int. Conf. Renewable Energies Power Qual.*, Mar. 2013, pp. 262–267.
[5] H. A. Kazem, T. T. N. Khatib, and A. Alwaeli, "Optimization of photovoltaic modules title angle for Oman," in *Proc. 7th IEEE Int. Power Eng. Optim. Conf.*, Jun. 2013, pp. 700–704.
[6] M. Catelani *et al.*, "Characterization of photovoltaic panels: The effect of dust," in *Proc. IEEE Int. Energy Conf. Exhib.*, 2012, pp. 45–50.
[7] B. G. Bhang *et al.*, "Power performance of bifacial c-Si PV module with different shading ratios," *IEEE J. Photovolt.*, vol. 9, no. 5, pp. 1413–1420, Sep. 2019
[8] H. Cha, B. G. Bhang, S. Park, J. Choi, and H. Ahn, "Power prediction of bifacial Si PV module with different reflection conditions on rooftop," *Appl. Sci.*, vol. 8, no. 10, pp. 1752–1771, Sep. 2018.
[9] G. G. Kim *et al.*, "Prediction model for PV performance with correlation analysis of environmental variables," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 832–841, May 2019.
[10] S. Dubey, J. N. Sarvaiya, and B. Seshadri, "Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world—A review," *Energy Procedia*, vol. 33, pp. 311–321, 2013.
[11] C. M. Whitaker *et al.*, "Effects of irradiance and other factors on PV temperature coefficients," in *Proc. Conf. Rec. 22nd IEEE Photovolt. Spec. Conf.*, 1991, pp. 608–613.
[12] B. Rupnik and O. Westbrook, "Ambient temperature correction of photovoltaic system performance data," in *Proc. IEEE 40th Photovolt. Spec. Conf.*, 2014, pp. 1973–1977.
[13] W. C. L. Kamuyu, J. R. Lim, C. S. Won, and H. K. Ahn, "Prediction model of PV module temperature for power performance of floating PVs," *Energies*, vol. 11, no. 2, Feb. 2018, Art. no. 447.
[14] S. Mekhilef, R. Saidur, and M. Kamalisarvestani, "Effect of dust, humidity and air velocity on efficiency of photovoltaic cells," *Renewable Sustain. Energy Rev.*, vol. 16, pp. 2920–2925, 2012.
[15] 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. 5, pp. 3979–3982, 2016.
[16] J. K. Kaldellis, M. Kapsali, 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, Jun. 2014
[17] "Photovoltaic system performance monitoring guidelines for measurement, data exchange and analysis," IEC 61724, Edition 1.0, 1998.
[18] N. H. Reich *et al.*, "Performance ratio revisited: Is PR ¿90% realistic?" *Prog. Photovolt., Res. Appl.*, vol. 20, pp. 717–726, 2012.
[19] L. Bonsignore, M. Davarifar, A. Rabhi, G. M. Tina, and A. Elhajjaji, "Neuro-fuzzy fault detection method for photovoltaic systems," *Energy Procedia*, vol. 62, pp. 431–441, Dec. 2014.