IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 1, 2019, accepted August 21, 2019, date of publication August 26, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937546

A Health Status-Based Performance Evaluation Method of Photovoltaic System

KUN DING [1], LI FENG [2], JINGWEI ZHANG [1], (Member, IEEE), XIANG CHEN [1], FUDONG CHEN [1], AND YUANLIANG LI [3]

[1]College of Mechanical and Electrical Engineering, Hohai University, Changzhou 213022, China
[2]Solar Computing Laboratory, Bielefeld University of Applied Sciences, 32427 Minden, Germany
[3]Changzhou Key Laboratory of Photovoltaic System Integration and Production Equipment Technology, Changzhou 213022, China

Corresponding author: Kun Ding (dingk@hhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51777059, in part by the Qing Lan Project, and in part by the Six Talent Peaks of Jiangsu Province under Grant GDZB-006.

ABSTRACT The large scale application of the photovoltaic (PV) systems is significantly beneficial to the mitigation of energy crisis. The quality and performance of PV systems directly influence the energy yield. It is necessary to establish more scientific and effective methods to maximize the energy yield. In this paper, the concept of health status is proposed to describe the performance of PV systems within a certain period of time. A health status based performance evaluation model is built by the Gaussian mixture models (GMM) and the empirical mode decomposition (EMD). Then, the health index (HI) of PV array is defined. Under the outdoor ambient conditions, the proposed model can sensitively detect the slight performance reduction caused by faults or partial shadings of the PV array in real-time. At last, the proposed method is verified by the simulations and experiments. Experimental results show that on the sunny day the average daily HI is 1.2 and the average performance ratio (PR) is 0.85, which both show the PV array is healthy. When one of the PV modules in the PV array is partially shaded, the PR is still approximate 0.9. However, the calculated HI is greater than the threshold and the fault is reported. The results indicate that the proposed method can sensitively identify pseudo health status that cannot be identified by the PR. The proposed evaluation method based on health status provides an alternative option to assess the performance of the PV systems. Combining with the PR, the comprehensive performance of the PV systems can be reflected more accurately.

INDEX TERMS Empirical mode decomposition, Gaussian mixture models, health status, photovoltaic system, performance evaluation.

I. INTRODUCTION

With the modern challenges that the fossil fuel resources continue dwindling, as a form of renewable energy, the solar energy provides the promising solution for modern energy supply. The photovoltaic (PV) system is one of the main application forms of solar energy. The global capacity of PV system continues increasing in recent decades. However, the performance of PV system is not only influenced by the current ambient conditions, but also degrades after long term outdoor operation. The performance degradation may be attributed to the aging or mismatch of PV modules, the low efficiency control strategies, the defects of design and installation [1]–[3]. Hence, the real-time monitoring and performance evaluation of the PV systems are necessary [4]–[6]. At present, many performance evaluation methods of PV systems have been presented. In the standard IEC 61724, the typical parameters, including the reference yield Y_r, array yield Y_A, final yield Y_f, capture losses L_c, system losses L_s, performance ratio (PR), mean PV array efficiency η_A, mean total efficiency η_{tot}, are introduced [7]. The detail guidelines and procedures are also described [7]. Among the above parameters, the reference yield and final yield are commonly used to quantify the equivalent hours of energy yield under the rated power of PV system, which already have been used to assess PV systems [8]–[15]. The normalized losses are also utilized to evaluate the loss degree of the PV system [6], [8]. Besides, the array or system efficiency is another basic index for assessing performance of PV array

The associate editor coordinating the review of this article and approving it for publication was Nagarajan Raghavan.

or system in a specific period of time [8], [10], [11], [14], [16]–[19]. Other researchers use the daily or monthly capacity factor (CF) to evaluate the actual energy yield relative to the energy under the rated power of PV system for the 24h operation [11], [13], [14]. As a regression method, the PVUSA rating is used to study the performance of PV system [9]. Nevertheless, the PR is used as the core criterion by the most researchers [6], [8]–[16], [20]–[24]. PR is an indicator calculated by the ratio between the final energy yield and the reference yield [7]. The comprehensive performance of the entire PV system can be expressed by the PR.

The above parameters only assess the performance of PV systems from the macro perspective. They are usually calculated by the monitoring data, including the in-plane irradiance, temperature of PV modules, voltage and current of PV array, voltage and current of the inverter, etc. The large amount of data cannot identify the detail faults or mismatch in a PV array. For instance, if few PV modules or a module in one string of the array are bypassed or affected by the dust in the air, the above parameters of the whole PV system are still in the normal range of the criteria. Eventually, due to unsuccessful detection, the energy continues losing in the PV system. Besides, the energy losses of a PV system are mainly attributed to the soiling, aging, partial shadows, dust of PV modules [1], [12], [14]. Some of these effects are recoverable, e.g. the partial shadows, soiling. The natural aging of PV modules or other system components, also result the slow degradation of system performance. However, these factors cannot be identified by the conventional performance indicators. In recent years, the line-line faults, line-ground faults and arc faults are focused due to the fact that they may cause catastrophic hazards [25]. In [25], different categories of faults in the PV system are classified. Corresponding advanced fault detection and diagnosis methods are surveyed and compared. In [26], the fault detection for the line-line and line-ground faults is integrated with maximum power point tracking (MPPT). In [27], a method for detecting the line-line fault, line-ground fault and partial shadings is proposed. The status of the PV array is classified as normal status, fault and partial shading. In [28], the compatibility of protection standards for line-line and line-ground faults are investigated by simulation and experimental results. In [29], the wavelet packet transform is used to extract features from the measured voltage and current of PV array for fault detection. The fault detection and diagnosis methods are also reviewed for the line-line and line-ground faults of PV array [30], [31].

In recent years, the artificial intelligence (AI) based methods are widely applied for the performance evaluation or fault detection of PV array [32]–[35]. In [32] the current-voltage (I-V) curves of PV array and the measured ambient meteorological data are used to train a deep residual network for fault detection and diagnosis. In [33] the voltage and current of the abnormal PV array or string are used as samples for training a random forest to build a real-time monitoring system of PV array. In [34] the kernel based extreme learning machine is used to extract key feature from measured I-V curves to detect faults. In [35], a long short-term memory network is applied to extract features from raw data of the PV array and to classify faults. However, for above methods, the sufficient fault data samples are necessary for training the deep learning networks. These fault data samples should be collected from simulation models or fault experiments of the PV array. Once the investigated PV array is changed, these fault data should be replaced and the new fault data samples are required.

In order to describe and quantify the detail performance of PV systems for further fault diagnosis, in this paper, the health status is introduced. The concept of health status is commonly used in prognostic and health management (PHM) for complex systems [36]. It utilizes the information, e.g. feature data or evidences sampled by the physical sensors, to discover the variation of system performance and the potential fault. Hence, the health status can be used to improve the system management and maintenance. It has attracted much attention in the fields of aerospace, electrical devices [37], [38]. PV system is a non-stationary complex system. Only using two statuses, i.e. normal or fault, cannot comprehensively describe the actual status of PV system. However, the health status is a promising concept to evaluate the overall performance of PV system under the certain ambient conditions.

In this paper, the indicator health index (HI) for PV systems is investigated and used to describe the real-time performance of PV system. Based on the Gaussian mixture model and empirical mode decomposition (EMD), a comprehensive performance evaluation model for calculating the HI is proposed. The performance reduction caused by the soiling, partial shading and faults, e.g. line-line short-circuit fault, line open-circuit fault are focused. Furthermore, the simulations and experiments are implemented to verify the feasibility and reliability of the proposed method. The proposed performance evaluation method can sensitively identify pseudo health status that cannot be identified by the performance ratio (PR). The proposed evaluation method based on health status provides an alternative option for researchers to assess the performance of PV systems. Combining with the PR, the comprehensive performance of PV systems can be reflected more accurately.

II. METHODOLOGY

A. SELECTION OF CHARACTERISTIC PARAMETERS FOR THE HEALTH STATUS

In this paper, the concept of health status is applied to analyze the real-time performance of PV systems. Referring to the definition of human health, the health of the PV system is defined as the deviation degree relative to the expected normal status under various outdoor conditions. The health status of PV system should reflect the overall status of the PV system and its subsystems.

Besides, the selection of the reasonable characteristic parameters is the key to evaluate the health status. The selected parameters influence the real-time diagnosis of the

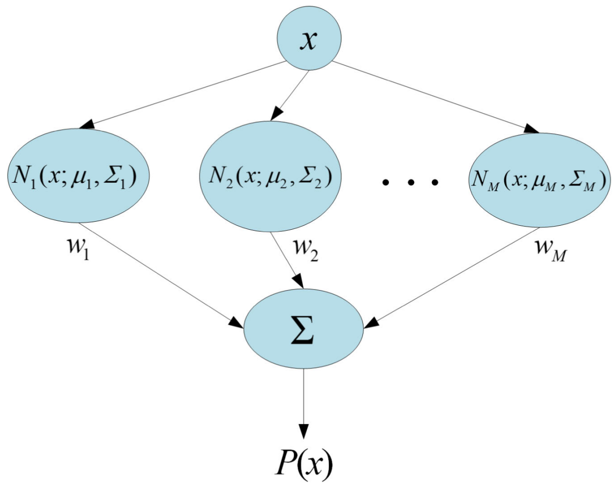


FIGURE 1. Structure of GMM.

operating status of the PV system. The energy generated from the PV systems is directly influenced by the ambient irradiance and temperature. The value of voltage and current of the PV array is the commonly used indicator for reflecting the output characteristics of PV array. Any minor variation can be expressed by the measured voltage and current of the PV array. Therefore, the in-plane irradiance and temperature of PV modules are selected as the basic ambient factors. The voltage and current of PV array are used for analyzing the health status of PV array.

B. BASIC THEORY OF GAUSSIAN MIXTURE MODEL

The PV system is a system with non-stationary random process. The model based on the voltage and current has high sensitivity and can quickly capture the characteristic of operation status. In this paper, the Gaussian mixture model (GMM) is applied to evaluate the performance of the PV system [39]–[41]. The GMM combines the advantages of nonparametric and parametric methods and it is a nonparametric model which considers the analytical features of closed mathematical models and the flexibility of nonparametric models. The modeling of GMM is easy, and its complexity is determined by the complexity of the studied problem. Furthermore, the GMM would not be easily affected by the size of the data sample. As long as the single Gaussian components are enough to participate in the mixture, the simple GMM can be used to approximate arbitrary complexity data distributions. The structure of the GMM is shown in Figure 1.

The GMM uses the weighted sum of the multiple Gaussian probability density functions to describe the complex distribution in the space of the probability vector. X denotes an n-dimensional random variable, i.e. X follows single Gaussian probability densities function if its probability density function is written as:

$$p(x) = \sum_{k=1}^{M} w_k p_k(x) = \sum_{k=1}^{M} w_k N(x; \mu_k, \Sigma_k) \tag{1}$$

where M is the number of the single Gaussian probability density functions. w_k is the weight of GMM, and satisfies $0 < w_k < 1$ and $\Sigma w_k = 1$. $N(x; \mu_k, \Sigma_k)$ represents the k-th single Gaussian probability density function, it can be expressed by:

$$N(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{n/2}|\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma^{-1}(x-\mu_k)} \tag{2}$$

where μ_k is the mean of the k-th single Gaussian probability density function, Σ_k is the covariance matrix of the k-th single Gaussian probability density function.

In the finite model parameter estimation, the maximum likelihood (ML) is applied for measuring a hybrid model. It is well-known that ML is expressed by:

$$\theta' = \arg\max_{\theta} \{\log p(x|\theta)\} \tag{3}$$

where $\theta = \left[w_1, w_2, \ldots, w_M, \mu_1, \mu_2, \ldots, \mu_M, \sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2\right]$,

$$\sum\nolimits_k = \sigma_k^2 \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix}$$

A common method for obtaining ML estimation of the mixture parameters is applying the expectation maximum (EM) algorithm [42]. The EM algorithm is based on the incomplete data X to acquire estimation values of model parameters from the training data sequence. Thus, the log-likelihood function of GMM about x_i is:

$$J(\theta) = \ln \prod_{i=1}^{M} p(x_i) = \sum_{i=1}^{M} \ln \left[w_k N\left(x; \mu_k, \sigma_k^2\right)\right] \tag{4}$$

With the Bayesian information criterion (BIC), the posterior probability is defined as the probability when the x is generated by the k-th Gaussian probability density function, which can be calculated by:

$$\beta_k(x) = p(k|x) = \frac{p(k)p(x|k)}{\sum\limits_{k=1}^{M} p(k)p(x|k)} = \frac{w_k N(x; \mu_k, \sigma_k^2)}{\sum\limits_{k=1}^{M} w_k N(x; \mu_k, \sigma_k^2)} \tag{5}$$

The derivative of equation (5) relative to μ_k and σ_k are:

$$\nabla_{\mu k} J(\theta) = \sum_{i=1}^{n} \beta_k(x_i) \left(\frac{x_i - \mu_k}{\sigma_k^2}\right) \tag{6}$$

$$\nabla_{\sigma k} J(\theta) = \sum_{i=1}^{n} \beta_k(x_i) \left(\frac{(x_i - \mu_k)^T (x_i - \mu_k)}{\sigma_k^3} - \frac{d}{\sigma_k}\right) \tag{7}$$

When the derivatives are zero, μ_k and σ_k are:

$$\mu_k = \frac{\sum\limits_{i=1}^{n} \beta_k(x_i)\, x_i}{\sum\limits_{i=1}^{n} \beta_k(x_i)} \tag{8}$$

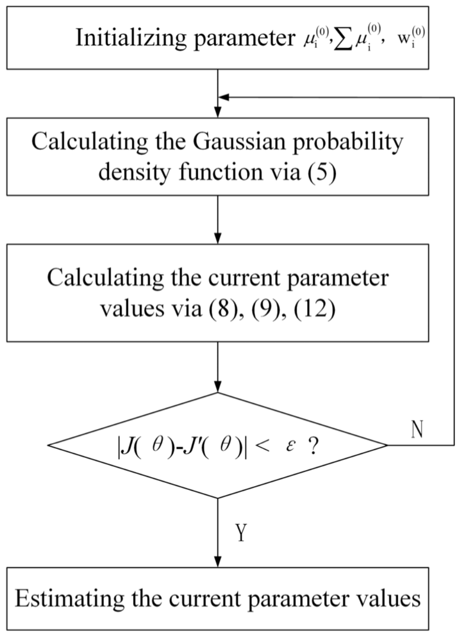


FIGURE 2. Flow chart of EM algorithm.

$$\sigma_k^2 = \frac{1}{d}\frac{\sum_{i=1}^{n}\beta_k(x_i)(x_i-\mu_k)^T(x_i-\mu_k)}{\sum_{i=1}^{n}\beta_k(x_i)} \tag{9}$$

Because the sum of w_k is 1, with the adopted Lagrange multiplier, a new objective function is defined as follows:

$$\begin{aligned}J_{new} &= J + \lambda\left(1-\sum_{k=1}^{M}w_k\right)\\ &= \sum_{i=1}^{n}\ln\left[\sum_{k=1}^{M}w_kN\left(x;\mu_k,\sigma_k^2\right)\right]+\lambda\left(1-\sum_{k=1}^{M}w_k\right)\end{aligned} \tag{10}$$

So the derivative of equation (10) relative to w_k is:

$$\begin{aligned}\frac{\partial J_{new}}{\partial w_k} &= \sum_{i=1}^{n}\frac{w_kN\left(x_i;\mu_k,\sigma_k^2\right)}{\sum_{k=1}^{M}w_kN\left(x_i;\mu_k,\sigma_k^2\right)}-\lambda\\ &= \frac{1}{w_k}\sum_{i=1}^{n}\beta_k(x_i)-\lambda\end{aligned} \tag{11}$$

When the equation (11) is zero, the final formula of w_k is:

$$w_k = \frac{1}{n}\sum_{i=1}^{n}\beta_k(x_i) \tag{12}$$

μ_k, σ_k, w_k are iteratively updated until $|J(\theta)-J'(\theta)|$ is less than ε ($\varepsilon < 10^{-5}$), where $J'(\theta)$ is the calculated result after parameters are updated. If the iteration converges, the algorithm will terminate. The detail flow chat of the iteration process is shown in Figure 2.

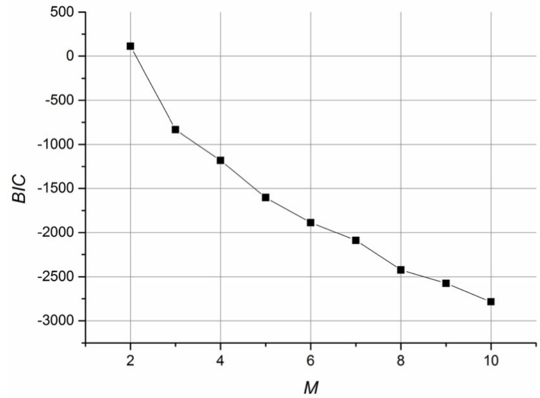


FIGURE 3. Relationship between *M* and *BIC*.

C. INITIALIZATION AND DETERMINATION OF HYBRID NUMBER OF GMM

The EM is a classic algorithm of the finite mixture model for maximum likelihood parameter estimation. However, it is sensitive to the initial values of the parameters, which significantly affect iteration speed of the EM. The poor initial value may lead the iteration converge to a local optimum. In this paper, the K-means algorithm is used to initialize the EM [43].

Besides, the number of the single Gaussian probability density functions M is another important parameter of GMM. In order to accurately fit the actual form of the data distribution, theoretically the greater value of M is a better choice. Nevertheless, it is limited by the size of the sampled data. Moreover, the insufficient mixed components may lead to the under-fitting and make the GMM difficultly describe the actual data distribution. Finally, it greatly reduces the accuracy rate of recognition. The sufficient mixed composition may lead to the over-fitting and increase the complexity of model. Hence, the reasonable selection of mixing degree to match the data distribution should be studied. In this paper, the BIC is used to determine the hybrid number M of GMM for the health status of PV system [44]. It is calculated by:

$$BIC = 2\times\max\{\log(p)(x|\theta)\} - M\times\log(n) \tag{13}$$

where $\max\{\log(p)(x|\theta)\}$ is the maximum likelihood estimation of GMM. The value of M corresponding to the maximum value of BIC is the optimal hybrid number of the GMM. The BIC is analysed in the range of 2 to 10, as shown in Figure 3. In this paper, when M is 2, the GMM is optimal to describe the most accurate data distribution.

After all the parameters of GMM are determined, the value of voltage and current for a certain operation status of PV system are standardized and used as an example to build the GMM. Corresponding complex multi-feature data of the standardized value of voltage and current are described by GMM and shown in Figure 4. Besides, the centroids of the results clustered by the cross entropy (CE) are also calculated and shown for comparison [45]. Obviously, the cluster results from K-means and CE are basically the same. Thus, in this

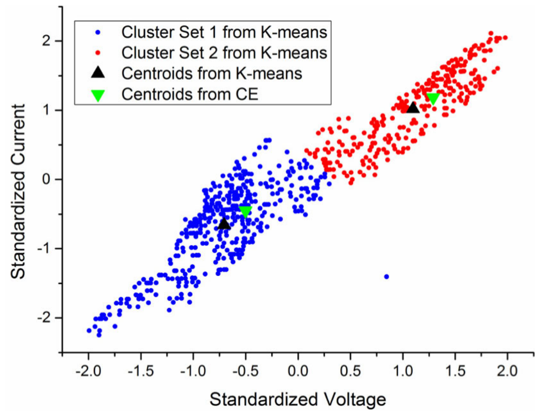


FIGURE 4. Complex multi-feature data described by GMM.

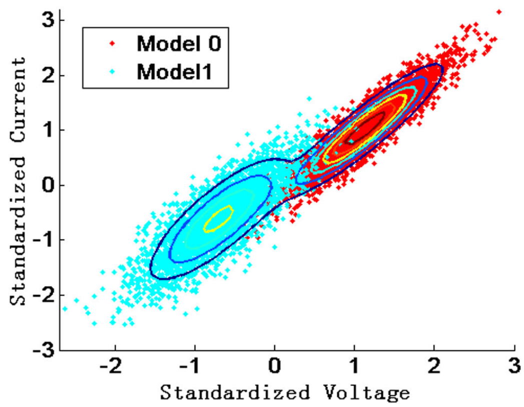


FIGURE 5. Contour map of GMM.

paper, the K-means is used to initialize the EM. The contour map of calculated GMM is shown in Figure 5. According to the actual characteristics of the voltage and current data, the built GMM can map the nonlinear relationship of health status for the PV system. Thus, the GMM based on multiple single weighted mixtures can be used for fitting the nonlinear data distribution. It is suitable to describe the complicated health status for the PV system.

D. PERFORMANCE EVALUATION METHOD BASED ON HEALTH STATUS

In this paper, the proposed real-time performance evaluation method based on the health status for the PV system is shown in Figure 6. The detail steps are as follows:

Step 1 (Obtaining Real-Time Data of the Current Status of PV System): The sampled data is acquired by the monitor system of the PV system, including the voltage and current of the each string of the PV array, the meteorological data (the coplanar irradiance, the temperature of back sheet of the PV modules).

Step 2 (Obtaining the Data of the Reference Status of PV System): Based on the measured coplanar irradiance and the back sheet temperature, the simulation model of PV system is built in the MATLAB/Simulink to calculate the electrical parameters of the reference status under the actual outdoor

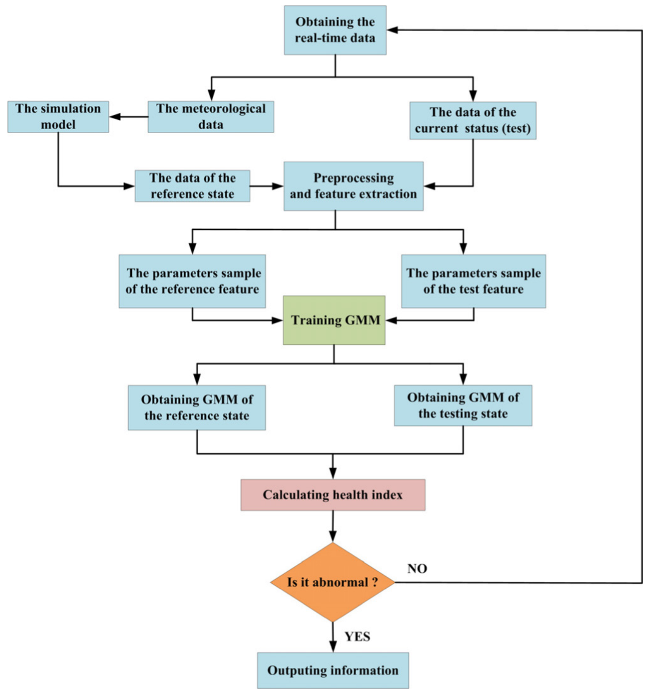


FIGURE 6. Flow chart of the real-time performance evaluation method for the PV system based on the health status.

condition, including the output voltage, current and power of each string in the PV array.

Step 3 (Data Processing and Feature Extracting): The sampled data and reference data of PV system are pre-processed to determine the reasonable data interval and to filter abnormal data points. Then, the data is extracted to obtain the reference and testing status samples by the EMD. The outdoor weather has the variability and instantaneity. In different time of the same day, the cloud coverage and the surface water evaporation are not exactly the same. These factors directly affect the absorbed solar radiation of the PV system. The output characteristics of PV system are changing with the ambient condition sensitively. Thus, in this paper, the original electrical data is extracted by the EMD at first. The EMD is a time domain analysis method that is widely used in the field of mechanical and electrical system for the fault diagnosis. It is suitable for analysing and processing non-stationary signals [46], [47]. In EMD, the composite signals are assumed to be the superposition of intrinsic mode functions (IMFs). The EMD algorithm is detailed as follows:

(1) At first, the maximal and minimal points are interpolated with the cubic spline curve to obtain the upper envelope E_1 and the lower envelope E_2 of the signal $x(t)$. The mean envelope m_1 is calculated as the average of the upper E_1 and lower envelopes E_2. Then, the IMF candidate h_1 is extracted from the original signal $x(t)$, as $h_1 = x(t) - m_1$.

(2) If the h_1 satisfies the IMF conditions, it is considered as IMF_1, and $IMF_1 = h_1$. If not, h_1 will replace the original signal $x(t)$, and the procedures (1) and (2) will be restarted to get a new the mean envelope m_{11}.

Next the new IMF candidate is computed with h_1 and m_{11}, $h_{11} = h_1 - m_{11}$. If the h_{11} is not matching the IMF conditions,the iteration will continue until the last h_{1k} satisfies the IMF conditions (k is iterations number), $h_{1k} = IMF_{1k}$. Where the IMF need to meet the following conditions:

1) The extreme points of the signal are equal to the zero crossing points or at most one difference form the zero crossing points.
2) At any time, the mean value between the upper and lower envelopes determined by the minimal and maximal values of the signal is zero.

(3) The residue signal R_1 is calculated by subtracting the first IMF from the original signal, $R_1 = x(t) - IMF_1$. The procedures (1) to (3) are repeated with n iterations until the residue signal R_n becomes a monotonic function from which no IMFs can be extracted.

$$\begin{cases} R_1 - IMF_2 = R_2 \\ \qquad \vdots \\ R_{n-1} - IMF_n = R_n \end{cases} \tag{14}$$

Finally, according to the fluctuations or trend of different characteristics scale (or frequency), the original signal $x(t)$ is decomposed to the n number of IMFs and the residue signal R_n:

$$x(t) = \sum_{i=1}^{n} IMF_i + R_n \tag{15}$$

In this paper, the whole voltage and current data are considered as the original signal $x(t)$ to be extracted with EMD. As shown in Figure 7 and 8, after the data of trend term are obtained, the interference factors are eliminated. Then the non-stationary time series is transformed to a stable time series.

Step 4 (Training GMM Model): Samples of the reference and testing status are used to obtain the model parameters for building the GMM of reference status and the GMM of testing status, respectively.

Step 5 (Calculating the Health Index (HI).): The mean value is an important parameter of the GMM. The position diversity represents the difference between different GMMs, as shown in Figure 9. If the system works normally, the data regions of testing and reference status are basically overlap. If the system is abnormal, a significant deviation between two regions may exist.

Thus, the HI is defined as the Euclidean distance between the GMM of the reference status and the GMM of the testing status:

$$HI = \sqrt{\sum_{i=1}^{M} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)} \tag{16}$$

where $\boldsymbol{\mu}_0$ is the mean vector of GMM of the reference status, $\boldsymbol{\mu}_1$ is the mean vector of GMM of the testing status.

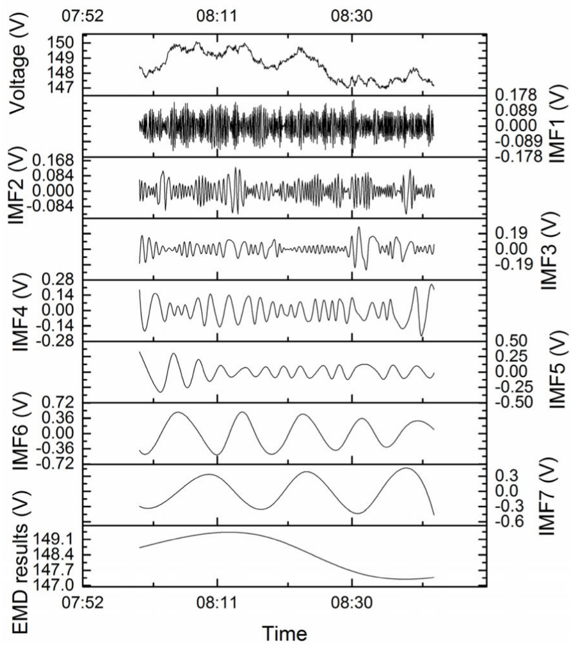


FIGURE 7. Results of EMD and the trend term.

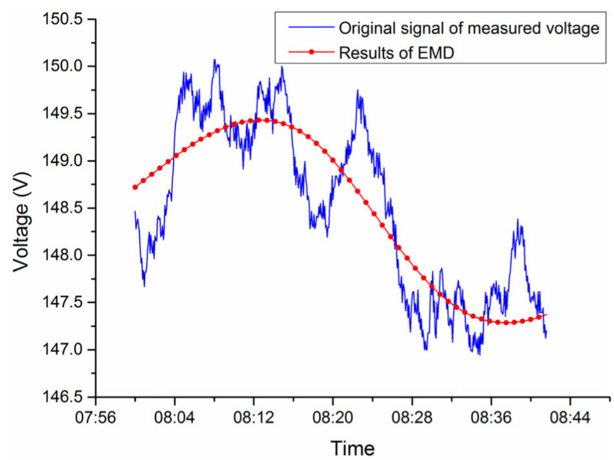


FIGURE 8. Extraction results.

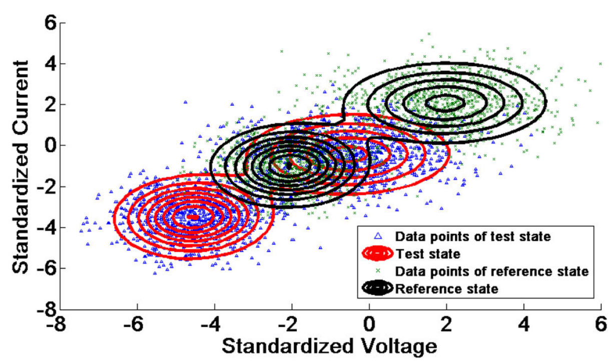


FIGURE 9. Calculation principle of health index.

Step 6 (Judging the Health Status of the PV System): A less value of HI indicates that the coincidence degree between the testing and the reference status is higher, i.e. the system is healthier, and the vice versa.

TABLE 1. Calculated HI under different dust conditions.

No.	Conditions	HI	PR
1	G=756W/m^2,T=30°C	11.13	0.70
2	G=456W/m^2,T=30°C	19.44	0.46
3	G=156W/m^2,T=30°C	63.11	0.17

III. SIMULATION AND ANALYSIS

In order to verify the accuracy and feasibility of the proposed method based on the health status, the simulation model of a 5×2 PV array is built in MATLAB/Simulink [48], as shown in Figure 10. Then, several commonly occurred faults of PV array are simulated and investigated. In this paper, the calculated HI is compared with the PR of PV array. Only considering the performance of the PV array without the inverter, the PR of PV array, i.e. DC PR, is expressed as [7], [49]:

$$PR = \frac{Y_A}{Y_r} = \frac{\tau_r \times \left(\Sigma_{day} P_A\right)/P_0}{\tau_r \times \left(\Sigma_{day} G_I\right)/G_{I,ref}} \tag{17}$$

where Y_A is the array yield, Y_r is the reference yield, τ_r is the sampling interval of the monitored data, P_A is the measured power of PV array, P_0 is the rated power of PV array, G_I is the measured in-plane irradiance, $G_{I,ref}$ is the reference in-plane irradiance of PV module.

A. SIMULATION ANALYSIS OF DIAGNOSIS OF SURFACE DUST OF PV MODULES

At first, the irradiance is respectively set to 756W/m^2, 456W/m^2 and 156W/m^2 to analyse the influence of different dust thickness for PV modules. The reference irradiance is assumed as 1056W/m^2, and the temperature of PV modules is 30°C. In this paper, the transient PR is compared with the proposed method. The results are shown in Table 1. Table 1 reveals that with increment of the surface dust thickness of PV modules, the equivalent irradiance is gradually reduced. The thicker dusts lead to less generated current of the PV array. Hence, the HI increases with the thickness of dusts. This result is consistent with the trend of PR under the same conditions.

B. SIMULATION ANALYSIS OF BYPASSED PV MODULES

Bypassed modules are PV modules that are bypassed by the wire or bypass diode. For simulating short circuit of PV modules, the bypassed modules in the simulation are replaced by diodes directly. The bypassed modules are assumed as PVM1, PVM2 in Figure 10 (a). The irradiance for other PV modules is set as 1056W/m^2 and the temperature of PV module is 30°C. The simulation results are shown in Table 2. With the increasing number of the bypassed PV modules, the HI increases significantly. The reason is that when the PV modules are bypassed, the output voltage of PV array drops dramatically and results to higher HI. It should be pointed out that, when a PV module is bypassed, the PR is 0.82 and

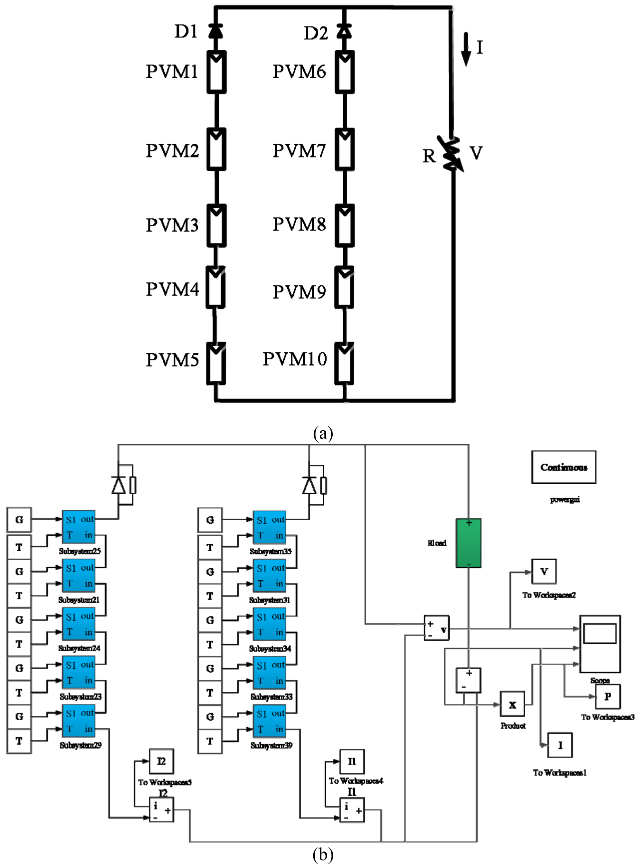


FIGURE 10. Structure and model of the 5×2 PV Array. (a) Structure of the 5×2 PV array. (b) Simulation model of the 5×2 PV array.

TABLE 2. Calculated HI when PV modules are bypassed.

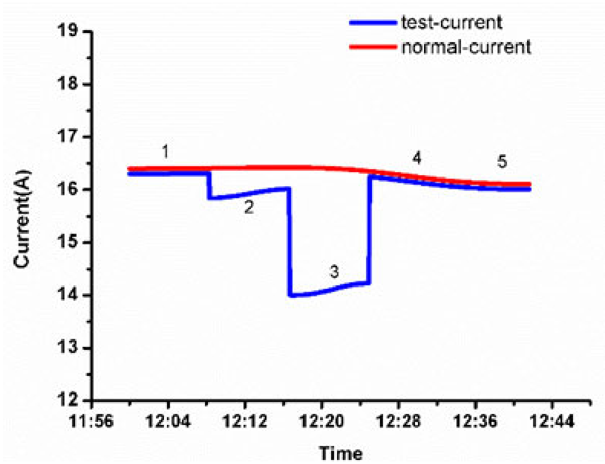
No.	Conditions	HI	PR
1	Non-bypassed PV array (ideal condition)	0	1
2	PVM1 is bypassed	55.82	0.82
3	PVM1 and PVM2 are bypassed	115.6	0.62

the system is mistaken to be considered as health. Under this circumstance, the proposed performance evaluation method based on health status can identify pseudo health status that is diagnosed as normal status by the PR.

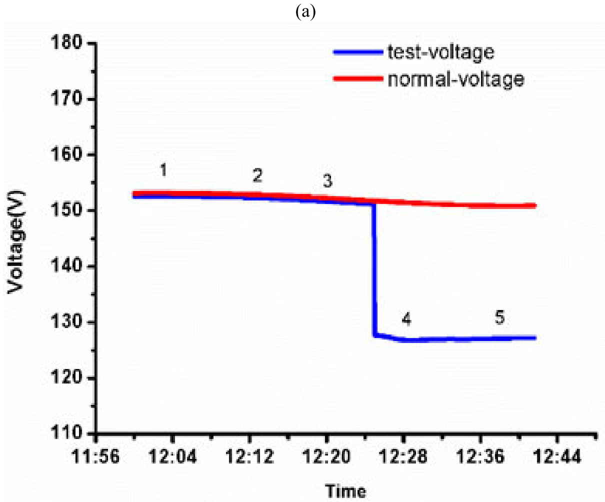
C. SIMULATION ANALYSIS OF PARTIALLY SHADED PV MODULES

The irradiance of PVM2 is set to vary from 200W/m^2 to 1000W/m^2 with 200W/m^2 interval, the irradiance of other PV modules are kept to 1000W/m^2, and the temperature of PV modules is 30°C. The output characteristics of PV array are shown in Figure 11.

In Table 3, the lowest HI corresponds to the result of the highest PR under G=1000W/m^2, T=30°C. The shaded



(a)



(b)

FIGURE 11. Output characteristics of PV array under shaded Condition. (a) Current curve of PV array. (b) Voltage curve of PV array.

TABLE 3. Calculated HI under different partial shading conditions.

No.	Conditions	HI	PR
1	G=1kW/m^2,T=30°C	1.217	0.8669
2	G=800W/m^2,T=30°C	1.683	0.8306
3	G=600W/m^2,T=30°C	4.759	0.7455
4	G=400W/m^2,T=30°C	48.61	0.7443
5	G=200W/m^2,T=30°C	47.68	0.7545

PVM2 is not bypassed under the equivalent irradiance G=800W/m^2 and G=600W/m^2. When the equivalent irradiance reduces to 400W/m^2, the voltage of PV array reduces significantly and the current of PV array recovers to normal. The reason is that the PVM2 is bypassed. In above two conditions, the PR is not changed obviously. Thus, the PR cannot identify the change of operation status of PV array in real-time. However, the HI can rapidly reflect the variations.

FIGURE 12. 10kWp distributed PV system for experiments.

TABLE 4. Specification of PV module TMS-PC05.

Parameters	Variable	Value
Maximum power	P_{MPP}	240W
Voltage at MPP	V_{MPP}	29.7V
Current at MPP	I_{MPP}	8.1A
Short circuit current	I_{SC}	8.62A
Open circuit voltage	V_{OC}	37.3V
Temperature coefficient of I_{SC}	α	0.047%/°C
Temperature coefficient of V_{OC}	β	-0.32%/ °C

IV. EXPERIMENTS AND APPLICATION

A. EXPERIMENTS UNDER NORMAL OPERATING CONDITION

A 10kWp distributed PV system on the roof is studied in this paper, as shown in Figure 12. It locates at north latitude 30° 125’ and east longitude 11° 6’. The tilted angle is 27°. The PV array is formed by 4 strings connected in parallel. Each string is formed by 10 PV modules connected in series. The specification of the PV module TSM-PC05 is shown in Table 4. Then, the PV array is connected to a 10kW three-phase grid-connected PV inverter.

The sampled data from 8:00 to 17:00 on March 16th, 2016 are analysed. The weather is sunny. The time interval for sampling is 5 seconds. The output characteristics of the PV system are shown in Figure 13. The evaluation result is shown in Figure 14. The average HI for the entire day is 1.2, which indicates the PV system is healthy on March 16th. The average PR is approximate 0.85 and also shows that the PV system operates normally. Hence, the evaluation result of HI is equivalent to PR under the normal operation status.

B. ABNORMAL OPERATION CONDITION

In this section, the proposed evaluation method under three abnormal conditions is studied, including the open circuit of one string, one shaded PV module and two shaded PV modules in one string. At first, the PV system operates under the

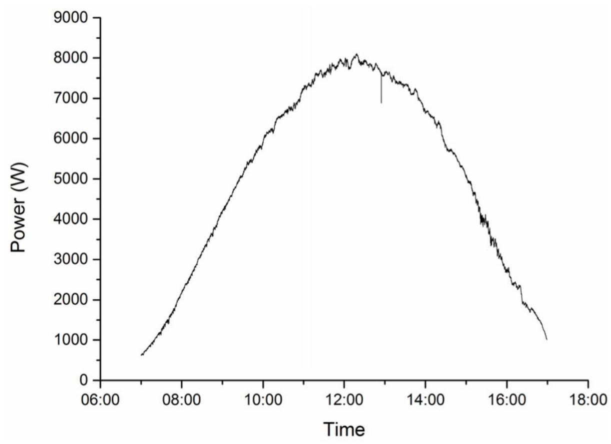


FIGURE 13. Measured power of the PV system on March 16th, 2016.

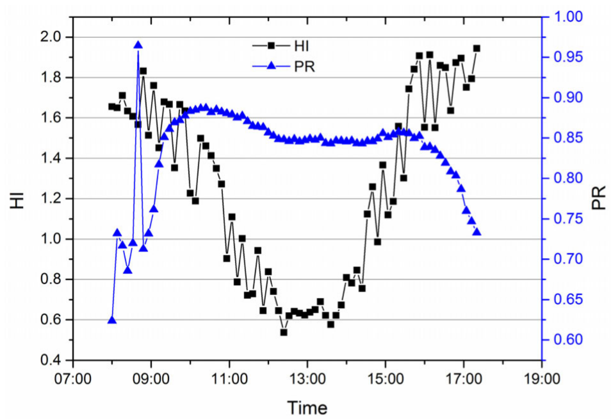


FIGURE 14. Calculated HI and PR on March 16th, 2016.

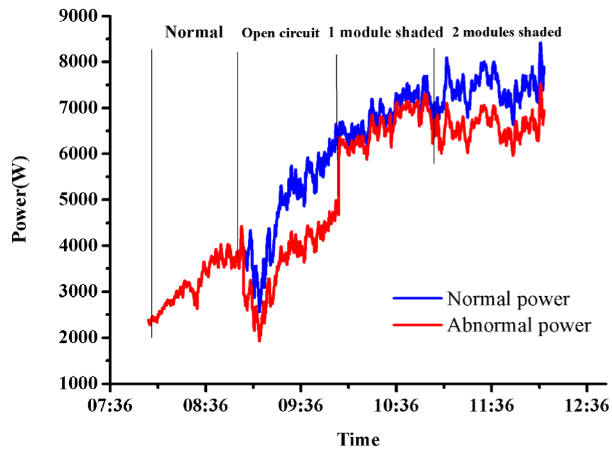


FIGURE 15. Power curve under abnormal conditions.

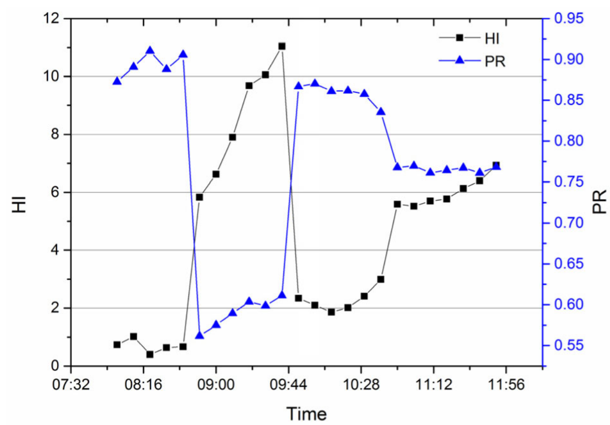


FIGURE 16. Diagnosis results under abnormal conditions.

normal status without any artificial faults. The average PR of PV system is 0.85 and indicates the PV system operates well enough. Then, the open circuit is artificially implemented by disconnecting one PV string from the PV array. Finally, the shaded PV modules (including artificially shaded one and two PV modules) are imitated to cover the PV module in one PV string by the black transparent polyvinyl chloride film and keep other PV strings operating normally. Figure 15 reveals the power curve of PV array from 8:00 to 12:00 on April 8th, 2016. In order to show the normal operating status of the PV array as a reference, the normal power of PV array is calculated by 4 multiplies the measured power of one normal PV string. In Figure 15, the power of PV array decreases dramatically for the open-circuit fault than other conditions. Besides, the loss of power for the condition of two shaded PV modules is greater than that of one shaded PV module.

In Figure 16, the HI agrees with the trend of PR and the actual status. Due to the fact that the power loss is the most obvious under the open circuit condition, the corresponding PR is the least and the calculated HI reaches the greatest. Comparing the diagnosis results under different shaded conditions, the HI under the condition of one shaded PV module is less than that under the condition of two shaded PV modules, i.e. the PV system performance of the former condition is healthier. Besides, when the PV system is under the condition of one shaded PV module, the PV system is evaluated as normal status by the PR, which is approximate 0.9. However, the HI is greater than the threshold value 2. Thus, though few PV modules have faults and cause slight performance variation of the whole PV system, the accuracy and sensitivity of the proposed model is better than that of PR.

It should be pointed out that the proposed evaluation method relies on the accuracy of the measured meteorological and electrical data of PV system. If the measured data is precise enough, the proposed evaluation method not only can evaluate the real-time performance in a certain period of time, but also it can be extended to long-term performance evaluation and fault diagnosis of PV systems.

V. CONCLUSION

In this paper, the concept of the health status of PV system is proposed to describe system performance. Then, the voltage and current are reasonably selected as the characteristic parameters of the health status. Furthermore, a real-time performance evaluation model based on health status is built with GMM and EMD. The estimation and selection of parameters of GMM are investigated. The established GMM can accurately map the complicated characteristic information of the PV system and describe its health status. In order to verify the effectiveness of the proposed real-time performance evaluation model, the simulation and experiments are implemented and analyzed. Experimental results show that on the sunny day the average daily HI is 1.2 and the average performance ratio (PR) is 0.85, which both show the PV array is healthy. When one of the PV modules in the PV array is partially shaded, the PR is still approximate 0.9. However, the calculated HI is greater than the threshold and the fault is reported. The experimental results indicate that the proposed performance evaluation model can not only achieve the same effect as PR under normal operating status of the PV system, but also accurately identify the pseudo health status that cannot be detected by the PR. The proposed evaluation method based on health status provides an alternative option to assess the performance of PV systems. Combining with the PR, the comprehensive performance of PV systems can be reflected more accurately.

REFERENCES

[1] A. Triki-Lahiani, A. B. Abdelghani, and I. Slama-Belkhodja, "Fault detection and monitoring systems for photovoltaic installations: A review," *Renew. Sustain. Energy Rev.*, vol. 82, pp. 2680–2692, Feb. 2018.

[2] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 1–17, Aug. 2018.

[3] A. Mohammedi, N. Mezzai, D. Rekioua, and T. Rekioua, "Impact of shadow on the performances of a domestic photovoltaic pumping system incorporating an MPPT control: A case study in Bejaia, North Algeria," *Energy Convers. Manage.*, vol. 84, pp. 20–29, Aug. 2014.

[4] M. Díez-Mediavilla, M. I. Dieste-Velasco, M. C. Rodríguez-Amigo, T. García-Calderón, and C. Alonso-Tristán, "Performance of grid-tied PV facilities: A case study based on real data," *Energy Convers. Manage.*, vol. 76, pp. 893–898, Dec. 2013.

[5] V. Komoni, I. Krasniqi, A. Lekaj, and I. Gashi, "Performance analysis of 3.9 kW grid connected photovoltaic systems in Kosova," in *Proc. 5th Int. Renew. Energy Congr.*, Hammamet, Tunisia, Mar. 2014, pp. 1–6.

[6] S. Bhattacharjee and S. Bhakta, "Analysis of system performance indices of PV generator in a cloudburst precinct," *Sustain. Energy Technol. Assessments*, vol. 4, pp. 62–71, Dec. 2013.

[7] *Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange and Analysis*, International Standard IEC, 61724, International Electrotechnical Commission, 1998.

[8] R. Dabou, F. Bouchafaa, A. H. Arab, A. Bouraiou, M. D. Draou, A. Neçaibia, and M. Mostefaoui, "Monitoring and performance analysis of grid connected photovoltaic under different climatic conditions in south Algeria," *Energy Convers. Manage.*, vol. 130, pp. 200–206, Dec. 2016.

[9] V. Sharma and S. S. Chandel, "Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 753–767, Nov. 2013.

[10] V. Sharma and S. S. Chandel, "Performance analysis of a 190 kWp grid interactive solar photovoltaic power plant in India," *Energy*, vol. 55, pp. 476–485, Jun. 2013.

[11] M. Farhoodnea, A. Mohamed, T. Khatib, and W. Elmenreich, "Performance evaluation and characterization of a 3-kWp grid-connected photovoltaic system based on tropical field experimental results: New results and comparative study," *Renew. Sustain. Energy Rev.*, vol. 42, pp. 1047–1054, Feb. 2015.

[12] E. Kymakis, S. Kalykakis, and T. M. Papazoglou, "Performance analysis of a grid connected photovoltaic park on the island of Crete," *Energy Convers. Manage.*, vol. 50, no. 3, pp. 433–438, Mar. 2009.

[13] A. M. Humada, M. Hojabri, H. M. Hamada, F. B. Samsuri, and M. N. Ahmed, "Performance evaluation of two PV technologies (c-Si and CIS) for building integrated photovoltaic based on tropical climate condition: A case study in Malaysia," *Energy Buildings*, vol. 119, pp. 233–241, May 2016.

[14] B. S. Kumar and K. Sudhakar, "Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India," *Energy Rep.*, vol. 1, pp. 184–192, Nov. 2015.

[15] A. Al-Otaibi, A. Al-Qattan, F. Fairouz, and A. Al-Mulla, "Performance evaluation of photovoltaic systems on Kuwaiti schools' rooftop," *Energy Convers. Manage.*, vol. 95, pp. 110–119, May 2015.

[16] T. Ozden, B. G. Akinoglu, and R. Turan, "Long term outdoor performances of three different on-grid PV arrays in central anatolia—An extended analysis," *Renew. Energy*, vol. 101, pp. 182–195, Feb. 2017.

[17] T. Khatib, K. Sopian, and H. A. Kazem, "Actual performance and characteristic of a grid connected photovoltaic power system in the tropics: A short term evaluation," *Energy Convers. Manage.*, vol. 71, pp. 115–119, Jul. 2013.

[18] A. M. Al-Sabounchi, S. A. Yalyali, and H. A. Al-Thani, "Design and performance evaluation of a photovoltaic grid-connected system in hot weather conditions," *Renew. Energy*, vol. 53, pp. 71–78, May 2013.

[19] R. Sharma and G. N. Tiwari, "Technical performance evaluation of stand-alone photovoltaic array for outdoor field conditions of New Delhi," *Appl. Energy*, vol. 92, pp. 644–652, Apr. 2012.

[20] A. M. Khalid, I. Mitra, W. Warmuth, and V. Schacht, "Performance ratio—Crucial parameter for grid connected PV plants," *Renew. Sustain. Energy Rev.*, vol. 65, pp. 1139–1158, Nov. 2016.

[21] A. K. Shukla, K. Sudhakar, and P. Baredar, "Simulation and performance analysis of 110 kW_p grid-connected photovoltaic system for residential building in India: A comparative analysis of various PV technology," *Energy Rep.*, vol. 2, pp. 82–88, Nov. 2016.

[22] G. A. Dávi, E. Caamaío-Martín, R. Rüther, and J. Solano, "Energy performance evaluation of a net plus-energy residential building with grid-connected photovoltaic system in Brazil," *Energy Buildings*, vol. 120, pp. 19–29, May 2016.

[23] IEA PVPS, "Photovoltaic module energy yield measurements: Existing approaches and best practice," Int. Energy Agency, Paris, France, Tech. Rep. IEA-PVPS T13-11, 2018.

[24] N. Ketjoy, C. Sirisamphanwong, and N. Khaosaad, "Performance evaluation of 10 kWp photovoltaic power generator under hot climatic condition," *Energy Procedia*, vol. 34, pp. 291–297, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187661021301000X

[25] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in PV systems," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 18–40, Aug. 2018.

[26] D. S. Pillai and N. Rajasekar, "An MPPT-based sensorless line–line and line–ground fault detection technique for PV systems," *IEEE Trans. Power Electron.*, vol. 34, no. 9, pp. 8646–8659, Sep. 2019.

[27] R. Hariharan, M. Chakkarapani, G. S. Ilango, and C. Nagamani, "A method to detect photovoltaic array faults and partial shading in PV systems," *IEEE J. Photovolt.*, vol. 6, no. 5, pp. 1278–1285, Sep. 2016.

[28] D. S. Pillai and R. Natarajan, "A compatibility analysis on NEC, IEC, and UL standards for protection against line–line and line–ground faults in PV arrays," *IEEE J. Photovolt.*, vol. 9, no. 3, pp. 864–871, May 2019.

[29] B. P. Kumar, G. S. Illango, M. J. B. Reddy, and N. Chilakapati, "Online fault detection and diagnosis in photovoltaic systems using wavelet packets," *IEEE J. Photovolt.*, vol. 8, no. 1, pp. 257–265, Jan. 2018.

[30] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A comparative evaluation of advanced fault detection approaches for PV systems," *IEEE J. Photovolt.*, vol. 9, no. 2, pp. 513–527, Mar. 2019.

[31] D. S. Pillai, J. P. Ram, N. Rajasekar, A. Mahmud, Y. Yang, and F. Blaabjerg, "Extended analysis on line-line and line-ground faults in PV arrays and a compatibility study on latest NEC protection standards," *Energy Convers. Manage.*, vol. 196, pp. 988–1001, Sep. 2019.

[32] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions," *Energy Convers. Manage.*, vol. 198, Oct. 2019, Art. no. 111793.

[33] Z. Chen, F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen, "Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents," *Energy Convers. Manage.*, vol. 178, pp. 250–264, Dec. 2018.

[34] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, "Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics," *Appl. Energy*, vol. 204, pp. 912–931, Oct. 2017.

[35] A. Y. Appiah, X. Zhang, B. B. K. Ayawli, and F. Kyeremeh, "Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis," *IEEE Access*, vol. 7, pp. 30089–30101, 2019.

[36] J. P. M. Smeulers, R. Zeelen, and A. Bos, "PROMIS methodology for prognostic health management," in *Proc. IEEE Autotestcon Syst. Readiness Technol. Conf.*, Valley Forge, PA, USA, Aug. 2001, pp. 559–568.

[37] H. Fang, H. Shi, Y. Xiong, and R. Li, "The component-level and system-level satellite power system health state evaluation method," in *Proc. Prognostics Syst. Health Manage. Conf.*, Zhangiiaijie, China, Aug. 2014, pp. 683–688.

[38] L. En-Wen and S. Bin, "Transformer health status evaluation model based on multi-feature factors," in *Proc. Int. Conf. Power Syst. Technol.*, Chengdu, China, Oct. 2014, pp. 1417–1422.

[39] K.-L. Lim, H. Wang, and X. Mou, "Learning Gaussian mixture model with a maximization-maximization algorithm for image classification," in *Proc. IEEE 12th Int. Conf. Control Automat. (ICCA)*, Kathmandu, Nepal, Jun. 2016, pp. 887–891.

[40] Y. Zamiri-Jafarian and S. Gazor, "Receiver design for diffusion-based molecular communication: Gaussian mixture modeling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[41] Q. Gemine, B. Cornélusse, M. Glavic, R. Fonteneau, and D. Ernst, "A Gaussian mixture approach to model stochastic processes in power systems," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Genoa, Italy, Jun. 2016, pp. 1–7.

[42] J. Gao, L. Zhou, and B. Du, "Parameter estimation of Gaussian mixture model and its application in multimode process monitoring," in *Proc. 12th World Congr. Intell. Control Automat. (WCICA)*, Guilin, China, Jun. 2016, pp. 2896–2901.

[43] V.-E. Neagoe and V. Chirila-Berbentea, "Improved Gaussian mixture model with expectation-maximization for clustering of remote sensing imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Beijing, China, Jul. 2016, pp. 3063–3065.

[44] M. Stephens, "Bayesian analysis of mixture models with an unknown number of components—An alternative to reversible jump methods," *Ann. Statist.*, vol. 28, no. 1, pp. 40–74, 2000.

[45] D. P. Kroese, R. Y. Rubinstein, and T. Taimre, "Application of the cross-entropy method to clustering and vector quantization," *J. Global Optim.*, vol. 37, no. 1, pp. 137–157, 2007.

[46] V. K. Mishra, V. Bajaj, A. L. Kumar, and G. K. Singh, "Analysis of ALS and normal EMG signals based on empirical mode decomposition," *IET Sci., Meas. Technol.*, vol. 10, no. 8, pp. 963–971, 2016.

[47] S. Tavildar and A. Ashrafi, "Application of multivariate empirical mode decomposition and canonical correlation analysis for EEG motion artifact removal," in *Proc. Conf. Adv. Signal Process. (CASP)*, Pune, India, Jun. 2016, pp. 150–154.

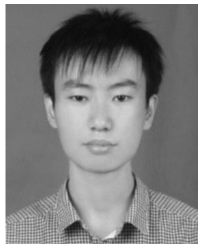
[48] K. Ding, X. Bian, H. Liu, and T. Peng, "A MATLAB-simulink-based PV module model and its application under conditions of nonuniform irradiance," *IEEE Trans. Energy Convers.*, vol. 27, no. 4, pp. 864–872, Dec. 2012.

[49] S. Rehman and I. El-Amin, "Performance evaluation of an off-grid photovoltaic system in Saudi Arabia," *Energy*, vol. 46, no. 1, pp. 451–458, 2012.

KUN DING received the M.S. degree in materials processing engineering and the Ph.D. degree in electrical engineering from Hohai University, Jiangsu, China, in 2000 and 2009, respectively, where he is currently a Professor with the College of Mechanical and Electrical Engineering. His research interests include modeling of PV module and PV system, PV system integration and optimization, PV system performance evaluation, mechatronics technology, and intelligent equipment technology.

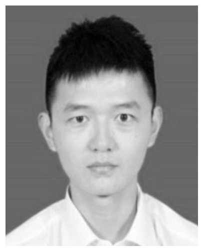
LI FENG received the M.S. degree in mechanical engineering from Hohai University, in 2016.

She is currently with the Bielefeld University of Applied Sciences, Minden, Germany. Her current research interests include multi-physics analysis, modeling, simulation, fault diagnosis, and performance evaluations of the photovoltaic systems.



JINGWEI ZHANG (S'14–M'18) received the M.S. degree in mechanical engineering from Hohai University, Jiangsu, China, in 2014, where he is currently pursuing the Ph.D. degree in power system and automation.

His current research interests include fault diagnosis of photovoltaic systems, multi-physics analysis of electromagnetic devices, and power electronics.



XIANG CHEN received the bachelor's degree in mechanical engineering from Hohai University, Jiangsu, China, in 2017, where he is currently pursuing the Ph.D. degree in hydraulic machinery.

His current research interests include photovoltaic module fault diagnosis technology and photovoltaic array health diagnosis.

FUDONG CHEN received the M.S. degree in mechanical engineering from Hohai University, in 2019.

His current research interests include fault diagnosis and health status analysis of the PV systems.

YUANLIANG LI received the M.S. degree in mechanical engineering from Hohai University, in 2019.

His current research interests include fault diagnosis and health status analysis of the PV systems.