PV module fault diagnosis based on micro-converters and day-ahead forecast

***Abstract*—The employment of solar micro-converter allows a more detailed monitoring of the PV output power at the single module level; thus, machine learning techniques are capable to track the peculiarities of modules in the PV plants such as regular shadings. In this way it is possible to compare in real-time the day-ahead forecast power with the actual one in order to better evaluate faults or anomalous trends which might have occurred in the PV plant. This paper presents a method for an effective fault diagnosis; this method is based on the day-ahead forecast of the output power from an existing PV module, linked to a micro-converter, and on the outcome of the neighbor PV modules. Finally, this paper proposes also the analysis of the most common error definitions with new mathematical formulations, by comparing their effectiveness and immediate comprehension, in view of increasing power forecasting accuracy and performing both real-time and offline analysis of PV modules performance and possible faults.**

***Index Terms*—Fault diagnosis, micro-inverter, PV system, day-ahead forecast.**

I. Introduction

THE study of the effective configuration of photovoltaic (PV) systems and their optimization has continuously attracted scientific and industrial research in the last decade.

Moreover, since Renewable Energy Sources (RES) are intermittent and variable, the availability of undisrupted operation is extremely important, especially in view of operation and maintenance (O&M), among many purposes.

Generally speaking, distributed module-converter layouts lead to a higher energy yield by diminishing the effect of mismatching and partial shading: recently, maximum power point tracking (MPPT) efficiency was increased at the module level by using DC-DC power optimizers [1] and sub-module PV systems such as DC-AC micro-inverters connecting a single PV module [2] to the electrical grid or in island mode [3].

However, the study of PV systems, along with their spread, has been facing also the problem of plants which were out of service due to single components failure [4], and the overall degradation of performance [5] may require a statistical analysis of the real-time data for supervision and monitoring [6].

Therefore, distributed electronics is also helpful, at module and sub-module level, in monitoring the PV production and for diagnostic purposes. In fact, the availability of micro-converters directly connected to several PV modules, usually with the same characteristics, would also allow an easy but detailed comparison among their performance and working status.

Furthermore, a suitable power forecast, performed in a prognostic way, can be employed for the identification and predictive maintenance of an equipment which will no longer fulfill its intended function. Such prediction is usually executed starting from the health state of the component (in particular the PV module) and taking into account its past history and future operation. Typically, these methods are largely classified as model-based, data-driven [7] and hybrid [8], respectively.

In particular, model-based methods adopt mathematical relationships of the degradation process in order to forecast degradation state [9]. As a matter of fact, data-driven methods are often used when an explicit model is not available, but there are enough historical data. These are statistical based models, learning trends from the amount of historical data [10].

Among data-driven models can be included: Autoregressive Moving Average techniques [11], Relevance Vector Machines [12], and Machine learning techniques, such as Artificial Neural Networks (ANNs), which were often used in the prognostic field [8], [12], [13]. In particular, feedforward ANNs have been used for the prediction of rotating machineries [14] and Lithium-ion batteries remaining useful life (RUL) [11], and Echo State Network (ESN) have been used for the Fuel Cells failure diagnostic [15]. Finally, hybrid approaches combine physics-based models of the degradation process with the use of historical data collected from degrading components [10].

Moreover, most of the available bibliographic sources for prognostics deal with different methods, mainly focusing on predicting and estimating the RUL of a specific system or component. However, the number of problems and possible faults of PV systems may depend also on different plant layouts and typology of installation, *i.e.* building integrated photovoltaics (BIPV) [1], [16] or large PV plants [17]–[20].

Given that the forecast horizon together with the temporal resolution (time sampling) strongly affect the prediction accuracy, forecasting models with prognostics purposes should take into account the different sources of uncertainty affecting predictions [21] such as:

- randomness in the equipment future degradation path, due to the intrinsic stochastic nature of the degradation process and the unknown future operation and environmental conditions;
- inaccuracy of the forecasting model;
- measurement noise or offsets;
- reliability of the monitoring system (data transmission).

This paper aims to present a method for PV system monitoring at the module-level, by adopting the day-ahead output power forecast of existing PV modules, connected to

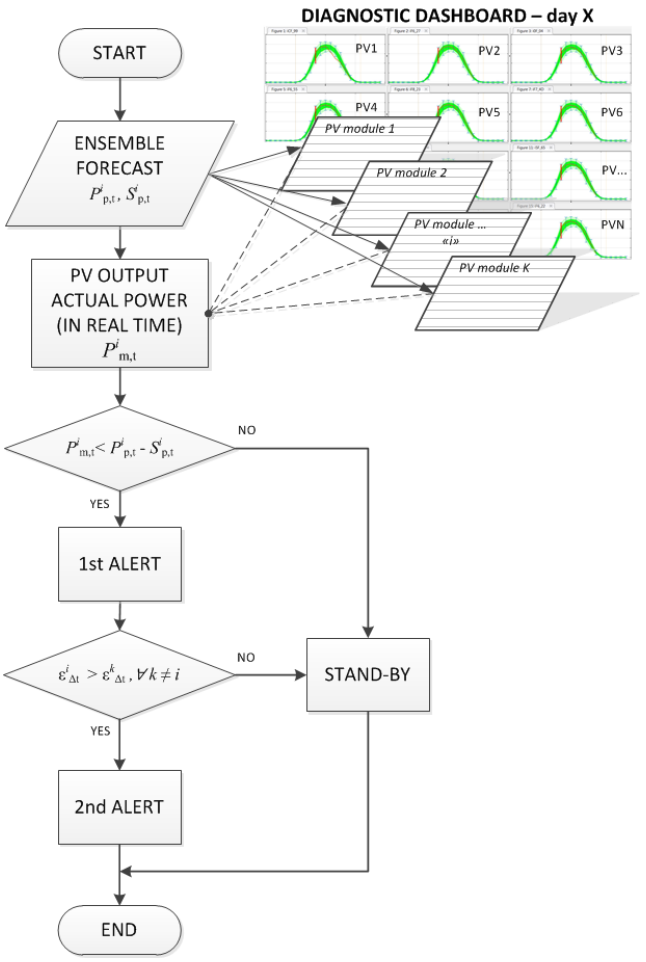

Fig. 1. Flowchart for the "faulty samples" detection and the Diagnostic Dashboard composition.

a micro-converter, for an effective fault diagnosis purpose. Additionally, this paper proposes the analysis of the most common error definitions employed in PV forecasting, by comparing their effectiveness with new mathematical formulations of performance indexes aimed to provide an immediate comprehension of the faulty status at the PV module level.

This paper is structured as follows: first of all, in Section II the Daily Diagnostic Dashboard is presented, where a new strategy for the real time data analysis coupled with the day ahead power forecast is showed with maintenance purposes. Then, in Section III the day-ahead PV output power forecasting method by means of ANN is presented. In Section IV a list of error definitions usually adopted in PV forecasting is presented and new ones are proposed in view of maintenance purposes; Section V presents the real case study here considered for validation. Finally, results of the proposed method are shown in Section VI and comments and conclusions are drawn in Section VII.

II. The Daily Diagnostic Dashboard

The methodology which has been applied for monitoring and diagnostic purposes entails two distinct steps concerning the PV module output:

1) real-time monitoring analysis;
2) offline analysis of the recorded data.

In particular, the first step of this diagnostic method leads to the "daily diagnostic dashboard" composition, which provides an immediate information about the status of the monitored PV modules by comparing in real-time the actual power with the forecast and the actual trends of the neighbor PV modules. Instead, the off-line analysis is performed at the end of the day providing useful indicators both in determining a faulty or anomalous trend and for the employment of reliable data in the continuous training of the forecasting model, as shown in Section VI.

A. The real-time monitoring analysis

The real-time monitoring analysis follows the flowchart reported in Fig. 1 and it is based, first, on the comparison between the measured power and its forecast, and secondly, between the measured power of similar modules, named as "neighbours". In particular, a different power prediction is computed for every PV module/converter: estimation of the power forecast is presented in detail in Section III.

The implemented procedure includes the following steps:

1) The predicted power $P^i_{p,t}$ in the $t-$th time sample is calculated for each $i-$th PV module/converter, with the related sample standard deviation $S^i_{p,t}$ (as defined in Section III, eq. 3).
2) The measured power $P^i_{m,t}$ of each $i-$th PV module/converter is recorded: this is the mean power measured in the t-th time sample (here a time resolution of 1 minute is considered, but for the sake of generality, a different sampling can be considered by averaging the measured values over the specific time interval).
3) The actual power $P^i_{m,t}$ is compared to the correspondent forecast $P^i_{p,t}$ for each $i-$th PV module/converter. When a module is not working properly, usually its measured power falls significantly below the forecast, and this may indicate anomalies. Therefore, if the following condition is verified for a single PV module/converter:
$$P^i_{m,t} < P^i_{p,t} - S^i_{p,t} \quad (1)$$
a "1^{st} level" alert is set on the status of the relevant module.
4) If the 1^{st} level alert condition is verified, the following check is performed between the given diagnostic indicator ε calculated in the time interval Δt of the $i-$th anomalous module and the $k-$th neighbours:
$$\varepsilon^i_{\Delta t} > \varepsilon^k_{\Delta t}, \forall k \neq i \quad (2)$$
where ε is any of the diagnostic indicators defined in Section IV and the here considered Δt is 15 minutes; if this last condition is verified a "2^{nd} level" alert is set on the status of the related micro-converter.

The status of all the PV modules is determined on the basis of the flowchart which has been explained, and a different color on the daily diagnostics dashboard is given according to the level of the alert. If the 1^{st} level alert is reached the background of the relevant micro-inverter graph on the

TABLE I
CONFUSION MATRIX OF THE POSSIBLE ALERTS COMBINATION

On-line analysis	Off-line analysis	
	3^{rd} level	4^{th} level
1^{st} level	No fault	Soft fault
2^{nd} level	Soft fault	Hard fault

dashboard is yellow colored, and the color is kept even if the anomalous trend is subsided. In the case of a faulty situation is ascribed to the PV module, a 2^{nd} level alert is given and the background of the related micro-inverter graph is red colored on the diagnostic dashboard.

B. The off-line analysis

The off-line analysis is performed at the end of the day on the basis of the diagnostics indicators presented in detail in Section IV. It follows these steps:

1) daily diagnostic indicators calculation for each module, as presented in Section IV;
2) mean values μ_i and standard deviation σ_i calculation of the diagnostic indicators for all the PV modules, excluding the i-th one, and setting of the related $\mu_i+\sigma_i$ and $\mu_i+3\sigma_i$ thresholds;
3) if the diagnostic indicators of the i-th module are falling below the $\mu_i+\sigma_i$ threshold, then the module/converter is considered "healthy";
4) if the diagnostic indicators of the i-th module/converter are falling between the $\mu_i+\sigma_i$ and $\mu_i+3\sigma_i$ thresholds, then the i-th module is given a 3^{rd} level alert;
5) if the diagnostic indicators of the i-th module are greater than the $\mu_i+3\sigma_i$ threshold, then the module will receive the 4^{th} level alert;
6) matching between the alerts given by the diagnostic indicators (3^{rd} and 4^{th} level alert) and the generated real time alerts (1^{st} and 2^{nd} level alert), according to the confusion table reported in Table I:
   - the combination of a 1^{st} level alert and 3^{rd} level alert determines a "No fault" condition for the given module/converter;
   - the combination of a 2^{nd} level alert and 4^{th} level alert determines the "Hard fault" condition for the given module, requiring specific maintenance;
   - the remaining combinations will result in a "Soft fault" condition: the module has no need for maintenance yet, however recorded data are not reliable for future uses in model training, as detailed later in Section VI-C.

The faults description is provided in Section V; an example of the real-time and off-line analyses will be presented later in Section VI.

It is worth mentioning that the selected thresholds have been defined and tested considering the behavior of the modules in the whole year 2017, with particular reference to those days presenting anomalies, as described later in Section V and VI.

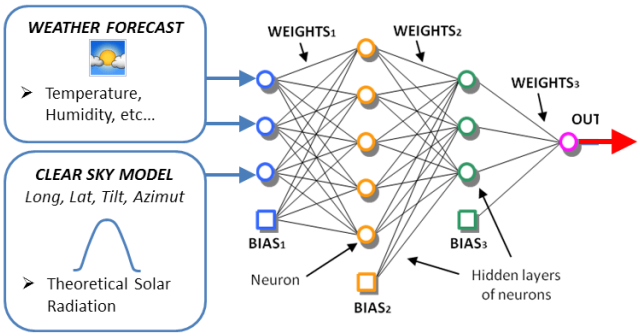

Fig. 2. PHANN forecasting method.

III. FORECASTING METHOD

The day-ahead forecast of the PV output power can be performed adopting several methods such as physical, statistical or hybrid ones. Although it can be achieved in many ways, hybrid methods proved their effectiveness because they merge in a unique method the upsides of the others [22].

Statistical methods, *i.e.* based on Artificial Neural Network (ANN), rely on the historical data measurements (weather conditions and produced power) for building up a suitable model of the plant. These have shown the main issue of committing higher percentage errors in specific hours of the day, in particular early in the morning and in the sunset hours [23].

Moreover, the so called Clear Sky Solar Radiation Model (CSRM) [24] is a physical model that computes the solar irradiance available without clouds for any tilted surface, whatever oriented, at a given location.

As shown in [23], when CSRM is included as an additional input of the ANN, it significantly reduces the error by defining the maximum hourly quantity of the solar irradiance and the expected daily hours of sunlight available for any specific location. This hybrid model, here adopted, is the Physical Hybrid Artificial Neural Network (PHANN).

A. Physical Hybrid Artificial Neural Network forecasting method

As shown in Fig. 2, PHANN forecasting method adopts ANN ability to learn from historical data the existing relationships among the weather forecast and the measured output power of the PV plant. Besides, in order to enhance the forecasting abilities, the deterministic solar irradiance under clear sky conditions (CSRM) is provided as an additional input.

Table II shows the complete list of the parameters, provided by the weather forecast service, which are employed as the input neurons together with the CSRM. In the end of this table also the DC output power, which is the forecast provided by the output neuron of the PHANN, is reported.

Learning phase is a very important step in ANN-based methods and many issues are related to it, as reported in [25], [26]. In fact, PHANN method needs is trained with the supervised learning. Supervised learning undergoes the

TABLE II
PARAMETERS EMPLOYED IN THE ANALYSIS

Symbol	Quantity	Unit of Measurement
d	day of the year	#
h	hour of the day	#
T_{DB}	dry-bulb air temperature	°C
G_{HI}	global horizontal solar irradiance	W/m^2
G_{POA}	global solar irradiance on the plane of the array	W/m^2
G^{cs}_{POA}	clear-sky global solar irradiance on the plane of the array	W/m^2
W_s	wind speed	m/s
W_d	wind direction	degrees
A_p	atmospheric pressure	hPa
R_f	rainfall	mm
C_c	cloud cover	%
C_t	cloud typology	Low/Mid/High
P_{out}	DC output power	W

matching between the historical input (weather parameters forecast and CSRM) to the actual output (measured DC output power). Available historical data (days of measurements) are divided into two sub-datasets, with different aims:

1) *training set*: used by ANN training procedure to learn relationships between input and output;
2) *validation set*: used by ANN training procedure to check if the model of the relationships is right.

The amount of data which have to be included in the different subsets have to pass through a sensitivity analysis, which has been previously performed by the authors, as detailed in [26].

Therefore, assuming a database of historical measurements continuously updated day by day, 90% of the available days are picked randomly to be used during training and the remaining 10% are for the validation set. This approach allows a continuous improvement of the prediction accuracy, as shown in [25].

Moreover, since the real power measured in the previous days is used to train the ANN, the forecasting model can be employed either for the whole PV plant or for each single module, as in the case described in this paper, where the use of micro-converters provides the availability of detailed power measurements from each single module.

B. Ensemble forecast

It has been proved in literature [27] that similar network models can produce slightly different results even if trained on the same dataset. This is due to the stochastic nature of ANN. Therefore, it was demonstrated that it is possible to reduce the forecasting error, by averaging the single output x_j of several parallel ANNs [28].

Thus, the above described process for PHANN training can be repeated on a number N_E of parallel networks (trials) in order to produce an average forecast $\overline{x}$ for each time sample, which is the mean value of all x_j samples.

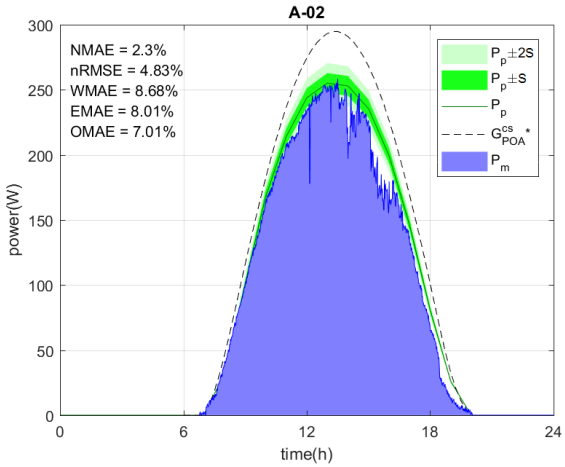

Fig. 3. Ensemble forecast $P_{p,t}$ of a single module output power with the relative standard deviation $S_{p,t}$ in a generic given day.

Moreover, due to the stochastic base of the PHANN, also the sample standard deviation $S^i_{p,t}$ of all the trials can be computed (in each time sample t and for each module i) as:

$$S^i_{p,t} = \sqrt{\frac{\sum_{j=1}^{N_E} (x_j - \overline{x})^2}{N_E - 1}} \tag{3}$$

Eq. 3 has the expression of the sample standard deviation, as the mean of the population of all the trials is not known *a priori*.

In this paper, the daily time series of $\overline{x}$ represents the PV output power profile P_p and it is referenced as the daily "ensemble forecast". In Figure 3, the violet area $P_{m,t}$ is the actual power of the PV module measured minutely by the micro-inverter "A-02"; on the left top corner of the picture, real time diagnostic indicators are reported. as detailed in the next section IV.

IV. DIAGNOSTIC INDICATORS

Because of the above mentioned intents, a steadily comparison between the actual energy and the expected one is extremely important in order to continuously check the system operation. Hence some statistical indicators, gathering the main features of the expected output power, provide an useful support for a reliable comparison with the actual power. As it is more likely to have a not null difference between the expected power and the actual one, a simple formulation of the error committed in the t-th sample of time is easy to find. Things become harder in finding a usable error definition which could be broadly used not only for diagnostic purposes but also which could give an effective assessment on the accuracy of the forecast at a glance. These indicators assess the forecasts accuracy providing a mathematical relation between the expected value and the actual value in the same sample of time.

The error e_t made in the t-th sample of time is the starting definition given as the difference between the values of the power measured $P_{m,t}$ and the forecast $P_{p,t}$ [29]:

$$e_t = P_{m,t} - P_{p,t} \quad \text{(W)} \tag{4}$$

From the expression of the absolute value $|e_t|$, most commonly used error definitions can be inferred [30] [31], such as:

- the well-known *Mean Absolute Percentage Error* ($MAPE$):

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{e_t}{P_{m,t}}\right| \cdot 100 \tag{5}$$

  where N is the number of considered time samples: in this analysis it is calculated for a single day;
- the *Normalized Mean Absolute Error* $NMAE$:

$$NMAE = \frac{\sum_{t=1}^{N}|e_t|}{N \cdot P_n} \cdot 100 \tag{6}$$

  where the percentage of the absolute error is referred to the rated power P_n of the PV module, instead of the hourly measured power $P_{m,t}$.
- the *Weighted Mean Absolute Error* $WMAE$ which is based on total energy production:

$$WMAE = \frac{\sum_{t=1}^{N}|e_t|}{\sum_{t=1}^{N}P_{m,t}} \cdot 100 \tag{7}$$

- the *Normalized Root Mean Square Error* $nRMSE$ is based on the maximum power output $max(P_{m,t})$:

$$nRMSE = \frac{\sqrt{\frac{\sum_{t=1}^{N}|e_t|^2}{N}}}{max(P_{m,t})} \cdot 100 \tag{8}$$

However, the daily evaluation indexes, expressed in (6), (7) and (8), could significantly differ, with $WMAE$ and $nRSME$ often showing values above 100%, thus being not able to provide a complete information "at a glimpse" on the accuracy of the prediction.

Starting from these assumptions, and in view of a more useful summary evaluation, additional performance indexes are proposed, aimed to provide a value between 0 and 100% of the forecast accuracy [32]. Therefore, the *Envelope-Weighted Mean Absolute Error*, $EMAE$ is defined as:

$$EMAE = \frac{\sum_{t=1}^{N}|e_t|}{\sum_{t=1}^{N}max(P_{m,t}, P_{p,t})} \cdot 100 \tag{9}$$

where the numerator is the same as $WMAE$, while the denominator is the sum of the maximum between the forecast and the measured power.

Additionally, a new diagnostics indicator is here introduced. It draws from the Performance Ratio $PR_{\%}$ coefficient as it is expressed in the IEC 61724 norm [33]:

$$PR_{\%} = \frac{\sum_{t=1}^{N}\frac{P_{m,t}}{P_n}}{\sum_{t=1}^{N}\frac{G_{POA,t}}{G_{STC}}} \cdot 100 = \tag{10}$$

$$= \frac{\sum_{t=1}^{N}P_{m,t}}{\sum_{t=1}^{N}G_{POA,t}} \cdot \frac{G_{STC}}{P_n} \cdot 100 \tag{11}$$

This indicator gives a good evaluation of the exploited energy from the PV system, by comparing it with the solar irradiance measured. Instead of the actual irradiance on the plane of array (G_{POA}) which is not always an available parameter, we adopted the following *Objective Mean Absolute Error* ($OMAE$):

$$OMAE = \frac{\sum_{t=1}^{N}|e_t|}{\sum_{t=1}^{N}G^{cs}_{POA,t}} \cdot \frac{G_{STC}}{P_n} \cdot 100 \tag{12}$$

where:

- $G^{cs}_{POA,t}$ is the solar irradiance on the plane of the array given by the clear sky solar irradiance model (CSRM) as it is described in [24];
- G_{STC} is the solar irradiance at standard test conditions (equals $1000W/m^2$).

From (12) it is possible therefore to rewrite the $OMAE$ highlighting the existing relationship with the former indicator $NMAE$:

$$OMAE = NMAE \cdot \frac{G_{STC}}{\sum_{t=1}^{N}G^{cs}_{POA,t}} \cdot N \tag{13}$$

As it can be seen in (13), the relationship between $NMAE$ and the new diagnostic indicator $OMAE$ is set. Finally, it is noticeable how $EMAE$ and $OMAE$ are both limited between 0% and 100%, providing an immediate indication on the magnitude of the daily diagnostic error, which is more likely related to a fault occurrence.

V. Case Study

In view of the validation of the proposed procedure, reported in the next Section VI, here we will use experimental data collected in the year 2017 at Solar Tech Lab, Milano, Italy. The whole PV plant is constituted by 21 silicon modules with different features; all the PV modules are facing South (0° Azimuth) and lay on fixed structures which are tilted at an angle of 30 degrees, (see Fig. 5). The maximum power point tracking is guaranteed by the micro DC-AC solar converters which are installed under each module, as described in [26]. The electrical parameters of the PV modules are collected remotely by the monitoring system; thus the maximum DC power values are recorded minutely.

For example, if we consider Fig. 4, it is possible to notice that the PV module connected to the micro-inverter "i85-87" had some problems (the DC output power from the PV module is constantly equal to 0 W). This faulty behavior can be either related to a fault which occurred in the data transmitter or to the DC-AC converter. On the other hand, regular partial shadings occurring in the morning and in the afternoon can be easily recognized in the different output of some PV modules.

VI. Examples and discussion

In the previous Section II, a dual-step diagnostic method was presented. Here we present and discuss the results obtained by applying the proposed approach on measured data from the experimental plant described in the previous section. In particular, an example of the "daily diagnostic dashboard"

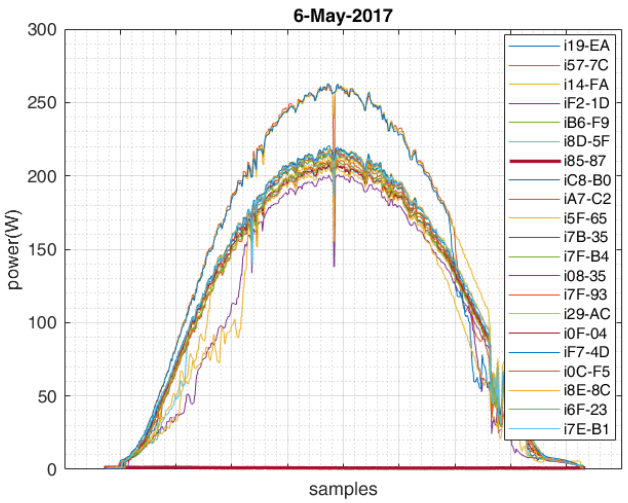

Fig. 4. DC output power of 21 PV modules, recorded at Solar Tech Lab during the 6th May 2017, with faulty samples for the converter i85-87.

Fig. 5. Solar Tech Lab picture on 13th April 2017 afternoon, with shadings of the railings on the PV modules.

is presented, showing real-time alerts for 15 specific micro-inverter PV-module status for a given day; then, the related off-line analysis is presented for the same modules. In addition, we will illustrate the acceptance criterion of the recorded sample in view of the forecasting model training, on the basis of the daily diagnostic indicators scored by each micro-inverter PV module set.

A. Real-time monitoring

An example of the Daily Diagnostic Dashboard for 13th April 2017 at 4:45PM is provided in Fig. 6. It can be noticed that PV modules connected to micro-inverters "C-03" and "A-05" are colored differently because of different alerts. In the first case, the actual power of the PV module provisionally differed from the forecast because of a temporarily shading, but returned to normal later. In the second case, the measured power constantly showed an anomalous lower trend compared both to the forecast and to the neighbor PV modules.

This analysis brought to the following four scenarios for each module:

1) *"healthy" or "good day"*: the online analysis gives no alerts related to anomalous trend (see, for instance, Fig. 3);
2) *regularly "partially shaded day"*, that is, either the shading has been properly forecasted or the neighbor PV modules show the same trend. For instance, as shown in Fig. 7, the partially shading in the morning is consolidated in the historical data of the PV module and the related forecast clearly shows this well-known trend. This condition still gives no alert for anomalies;
3) *unexpected partially shaded PV module*. In this case the power produced by the PV module diverges from the forecast and its trend is not consistent with the neighbours. For example, as shown in Fig. 8, the PV module connected to micro-inverter "C-03" has been intentionally partially shaded, starting from 11:45. As reported in Fig. 8, starting from that moment the measured power differed from the expected one sufficiently to generate a 1st level alert in the on-line analysis. In addition, neighbor modules did not behave in the same way and a 2nd level alert was generated. Finally, as the shading suddenly stopped at 12:15, the alert remained on the daily dashboard signaling the anomalous trend of the system. In this case, for this module, data were recorded as "soft fault", as described in Section II and in Table I.
4) *"faulty day"*. In this case the PV module output power has a completely anomalous trend: it differs both from the related forecast and from the neighbours' trends. For example, as shown in Fig. 9, the actual power of module "A-05" is significantly below the predicted power, and the computed diagnostic indicators are high: this means that a 2nd level alert is still on in the current time when the on-line analysis is performed.
This steep power loss is not associated to the regular ageing of the PV module; hence this sudden behavior was not predicted by the forecasting method. In fact, the regular ageing usually occurs along the years, therefore it can be predicted by the forecasting procedure described in Section III, since this is continuously trained on measured historical data for any specific module. In this case, as shown later, the complete diagnostic analysis will provide an "hard fault" alert. Indeed, after this alert, the PV module was inspected and many cells showed micro-cracks [34] on the surface, which brought the output power to decrease 20% within a few days.

B. Offline analysis

At the end of the day, given that in the on-line analysis the alerts referred to possible faults have been already set, the diagnostic indicators listed in Section IV are collected and analyzed off-line, as explained in II-B. The mean diagnostic indicators computed for the considered day are reported as an example in Table III: the micro-inverters scoring the highest daily values in all the indicators are "A-05" and "C-03".

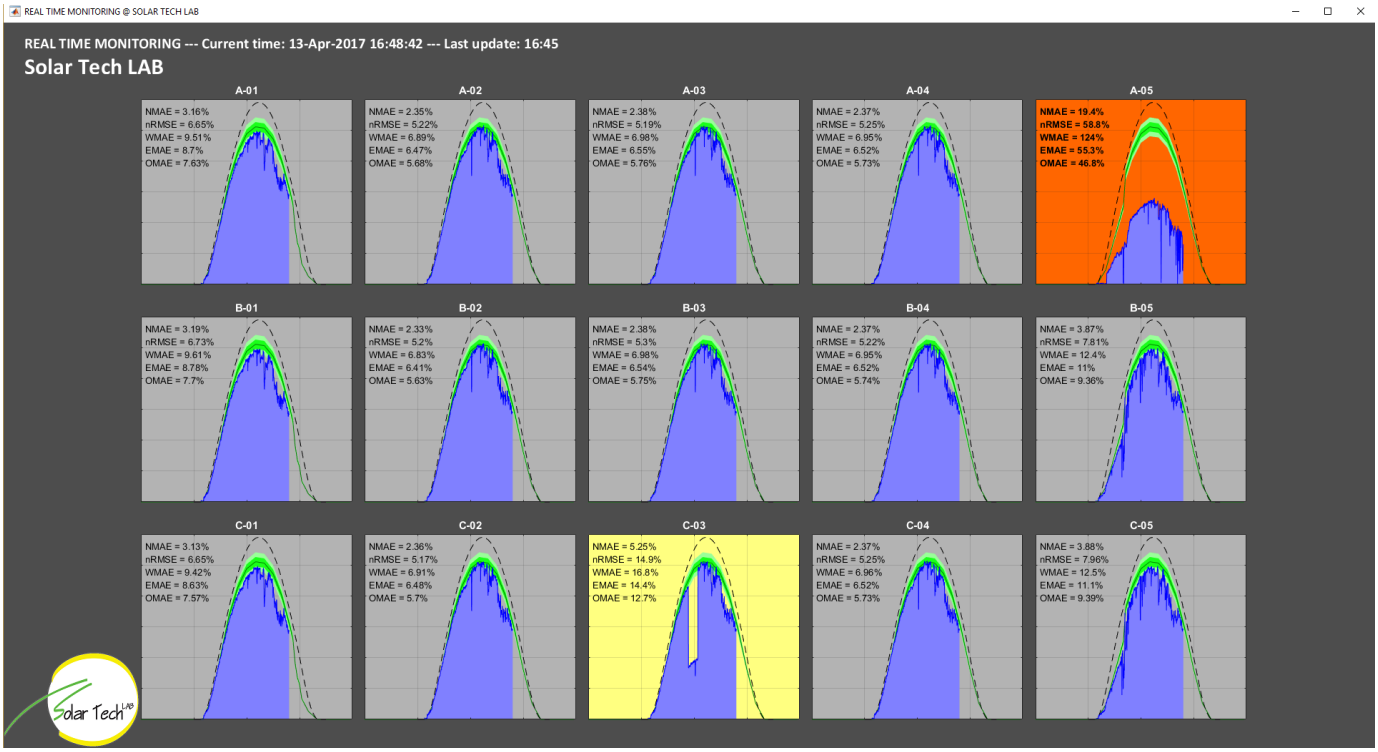

Fig. 6. Daily Diagnostic Dashboard for real-time monitoring of 15 modules (current status on 13th April 2017, afternoon).

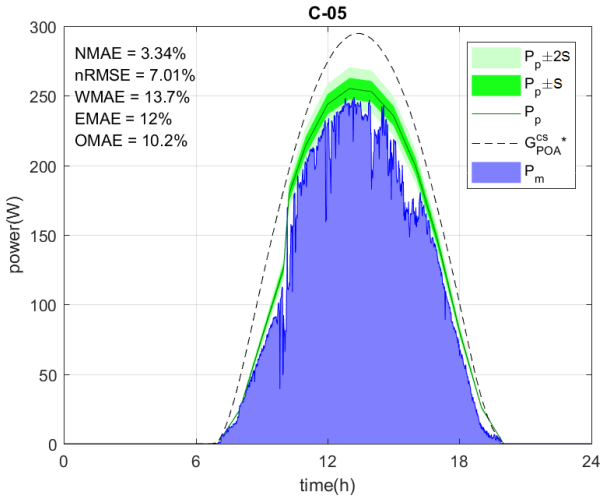

Fig. 7. Regularly "partially shaded" PV module in the morning.

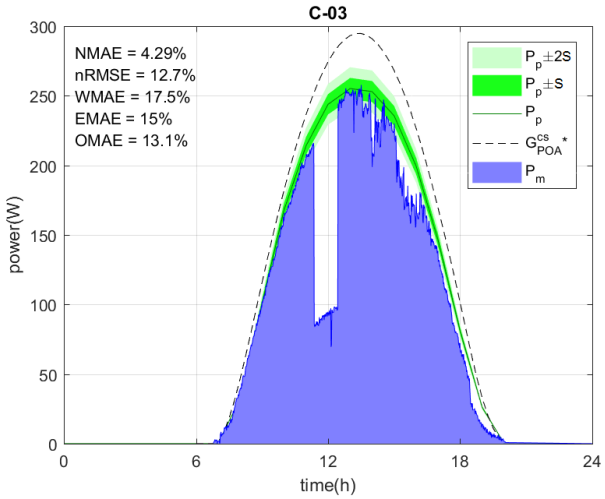

Fig. 8. Unexpected partially shaded PV module.

The highest indicator is $WMAE$ which exceeds 100%, while $NMAE$ is the lowest. Instead it is noticeable that $OMAE$ and $EMAE$ are greater than the regular ones (the indicators for the other micro-inverters are below 10% in all the cases), but they are not exceeding 100% maximum cap even if they are related to a probable faulty situation.

In particular, an example of $nRMSE$ for the off-line analysis is given in Fig. 10, highlighting the most critical anomalies with respect to the thresholds defined in Section II-B.

Therefore, on the basis of the off-line statistical analysis, it is now possible to point out the 3rd and 4th level alerts. By means of the alerts combination which has been previously described in II-B and Table I, the faulty day condition for the "A-05" inverter is determined as an "hard fault". Instead "C-03" inverter is marked as a "soft fault", which means that it is not in fault condition yet, but it is likely to have problems in the future if the same behavior is confirmed in the following days: thus, if no additional alerts are found later, the module can be considered healthy, while if a "soft fault" condition is found for consecutive days, the module will need to be carefully inspected by an operator in order to verify the cause of this systematic performance degradation. For this reason,

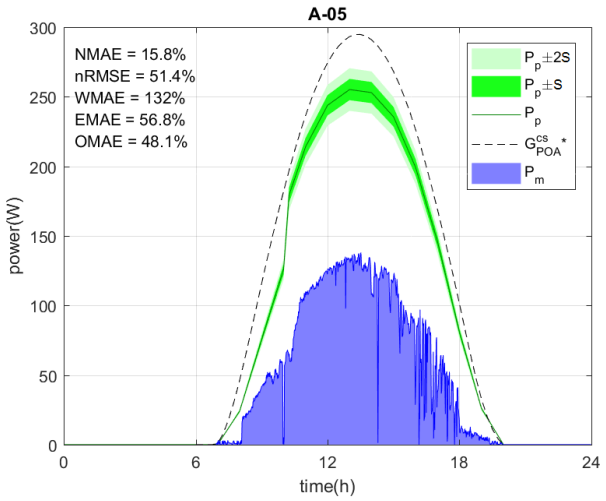

Fig. 9. "Faulty day" for the PV module connected to "A-05" micro-inverter.

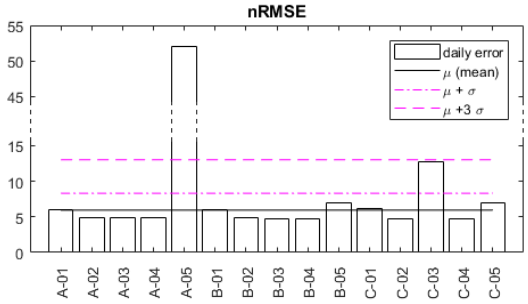

Fig. 10. $nRMSE$ off-line analysis with respect to thresholds.

TABLE III
MEAN DIAGNOSTICS INDICATORS RECORDED ON 13th APRIL 2017

module	$\overline{NMAE}$	$\overline{nRMSE}$	$\overline{WMAE}$	$\overline{EMAE}$	$\overline{OMAE}$
A-01	2.83	6.07	11.27	10.15	8.65
A-02	2.27	4.82	8.55	7.91	6.92
A-03	2.25	4.83	8.51	7.86	6.87
A-04	2.28	4.87	8.63	7.97	6.98
A-05	15.78	52.07	132.14	56.93	48.18
B-01	2.83	6.08	11.25	10.14	8.65
B-02	2.26	4.85	8.54	7.89	6.90
B-03	2.17	4.75	8.19	7.59	6.64
B-04	2.22	4.78	8.36	7.74	6.77
B-05	3.33	6.94	13.66	12.03	10.18
C-01	2.84	6.10	11.31	10.18	8.69
C-02	2.26	4.75	8.50	7.86	6.89
C-03	4.29	12.80	17.53	14.96	13.10
C-04	2.25	4.74	8.51	7.86	6.88
C-05	3.33	6.98	13.62	12.00	10.16

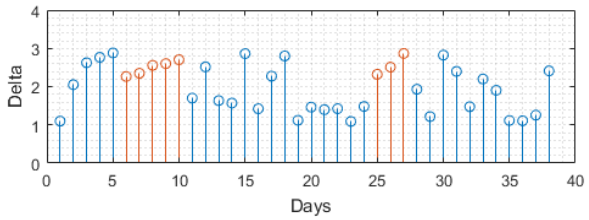

Fig. 11. Δ of the "faulty days" in 2017 with respect to $nRMSE$.

data measured from modules in "soft fault" condition will not be considered for the training of the model for the next days, as explained in the followings.

In order to define the method effectiveness based on experimental data, we have considered the measurements of one full year: in particular, in our analysis 307 days were available for the whole year 2017 (considering only valid measurements and removing incomplete data, related for instance to scheduled maintenance); among these, 38 days turned out to have "soft fault" alerts (including all the possible anomalies which could be referred to the PV system, from the module to the PV converter, before the connection to the grid).

To properly set the threshold for the off-line analysis, a Δ value can be considered to classify each diagnostic indicator for the i-th micro-converter, *e.g.* for the $nRMSE$ it can be defined as:

$$\Delta_i = \frac{nRMSE_i - \mu_i}{\sigma_i} \tag{14}$$

where μ_i and σ_i are the values defined in section II-B: as reported in Fig. 11, the values of Δ with respect to $nRMSE$ are usually defined in the range proposed in section II-B, when a "soft fault" was identified ($\sigma \div 3\sigma$). However, the alerts for these 38 days can be further analyzed *ex-post*, as proposed above, to evaluate how many false positive the method generated: in particular, the operator inspection confirmed that the "soft fault" alerts (marked in blue in Fig. 11) were mostly related to maintenance and other experimental activity conducted in the laboratory, which caused isolated performance degradation similar to the one shown in Fig. 8. On the other hand, the values highlighted in red in Fig. 11 represent the cases when the soft-fault alert was repeated on the same module for consecutive days: in this case an inspection was necessary to investigate the cause and, although the power reduction was not so relevant to activate a hard-fault alert, it was found that the module presented issues (hot spots) that caused a permanent performance reduction.

It is worth mentioning that the module performance can also be affected by the converter: indeed, each module is connected to a micro DC-AC solar converters, as described in [26]; in this case, it could be difficult to distinguish the real source of degradation. In fact, the proposed method is aimed to highlight possible faults based on the analysis of measured output power, but it is not able to detect a specific kind of fault or to directly identify the reasons. Therefore, as mentioned above, a careful operator check would be needed to determine the cause of any alert, but the proposed algorithm is able to locate where a fault condition happens and its degree, providing proper information to the system maintenance operators. In particular, with respect to a possible failure of micro-converters, these generally have a monitoring system of their status: thus, data will be completely missing or null in case of a hard fault, which is easy to detect; our analysis is based on measurements provided by healthy converters: if a converter has a failure, this will be clear from

TABLE IV
MEAN DIAGNOSTICS INDICATORS

	$\overline{NMAE}$	$\overline{nRMSE}$	$\overline{WMAE}$	$\overline{EMAE}$	$\overline{OMAE}$
C1	3.79	21.53	55.06	25.04	14.37
C2	8.41	727.96	1,893.1	58.13	29.89
C3	8.62	937.163	2,732.19	56.22	30.34

missing data.

C. Training data selection for diagnostic purposes

As reported in [26], the training approach and data selection are critical aspects for the effectiveness of forecasting, in particular when an online training is performed daily to continuously update the forecasting model.

For this reason, here we investigated the effect of including, in the training of the PHANN, the previously recorded measurements of inverter with alerts: this evaluation is performed by analyzing the resulting forecast accuracy.

In particular, we conducted an analysis based on forecasting faulty modules by considering two different training datasets, with previous "soft fault" condition either included or not; as mentioned in section VI-B, out of 307 days available, 38 days turned out to have "soft fault" alerts.

In Table IV the mean diagnostics indicators are reported with respect to the following configurations:

C1) this subset contains the power forecast for 269 days, based on a training dataset of 269 healthy days (data from healthy modules);
C2) this subset includes the forecast of 38 days with measured fault conditions: forecast is here based on a mixed training dataset composed by 231 healthy and 38 faulty days;
C3) this subset includes the forecast of 38 days with measured fault conditions: forecast is here based on a training dataset of 269 healthy days (the same training dataset as C1).

In all the three cases considered, the number of days considered in the training dataset was 269, for the sake of a fair comparison.

Moreover, Fig. 12 shows the daily $OMAE$ of the 38 faulty days forecasted with approaches C2 and C3. As reported in this figure, the forecast error is slightly lower when previous "faulty days" are included in the training set (C2), while the approach C3 is able to better highlight faults with respect to normal condition. Since our target is to increase the general forecasting accuracy and to emphasize fault conditions with respect to the predicted outcome of the plant, results suggest that it is better to exclude from training the previous days with detected "soft fault" condition, in order to have worse error indicators when fault conditions occur and to better detect faults.

VII. CONCLUSION

An effective diagnostic method for fault detection at PV-module level is here presented, based on monitored data by micro-inverter and day-ahead power forecasting. A suitable procedure has been implemented for real-time monitoring fault conditions by comparing neighbor PV modules and for performing a final offline evaluation at the end of the day, by means of properly defined diagnostic indicators. In particular, this method is aimed to provide a useful help in the fault detection at the PV-module level. The classification of faults and suggestion of possible measures are beyond the scope of this work and will be subject of further research.

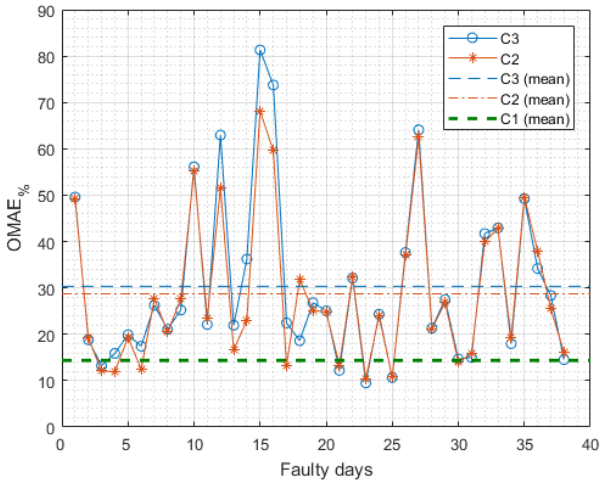

Fig. 12. $OMAE$ of the "faulty days" in 2017 with different training datasets.

The proposed method has been validated on 15 PV modules and the related micro-inverters in a laboratory facility, considering also one year of measurements for the definition of proper thresholds for the identification of different fault degrees. Results showed that the proposed approach is able to identify critical failures of PV modules, avoiding systematic errors like ageing or regular shadows, by means of a machine learning based forecasting, specific for each module. Moreover, the employment of specifically defined diagnostic indicators can provide an immediate comprehension of the fault status.

The proposed approach can be easily extended to optimizers in view of future implementations, by adding the feature of developing or the day-ahead forecast and the logic of the comparison among different PV module output on board.

REFERENCES

[1] R. W. Erickson and A. P. Rogers, "A microinverter for building-integrated photovoltaics," in *Applied Power Electronics Conference and Exposition, 2009. APEC 2009. Twenty-Fourth Annual IEEE*, pp. 911–917. IEEE, 2009.
[2] H.-J. Chiu, Y.-K. Lo, C.-Y. Yang, S.-J. Cheng, C.-M. Huang, C.-C. Chuang, M.-C. Kuo, Y.-M. Huang, Y.-B. Jean, and Y.-C. Huang, "A module-integrated isolated solar microinverter," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 2, pp. 781–788, 2013.
[3] C. T. Rodríguez, D. V. De La Fuente, G. Garcera, E. Figueres, and J. A. G. Moreno, "Reconfigurable control scheme for a pv microinverter working in both grid-connected and island modes," *IEEE transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1582–1595, 2013.
[4] A. Triki-Lahiani, A. B.-B. Abdelghani, and I. Slama-Belkhodja, "Fault detection and monitoring systems for photovoltaic installations: A review," *Renewable and Sustainable Energy Reviews*, 2017.

[5] J. Bastidas-Rodriguez, G. Petrone, C. Ramos-Paja, and G. Spagnuolo, "Photovoltaic modules diagnostic: An overview," in *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pp. 96–101. IEEE, 2013.

[6] S. Vergura, G. Acciani, V. Amoruso, G. E. Patrono, and F. Vacca, "Descriptive and inferential statistics for supervising and monitoring the operation of pv plants," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 11, pp. 4456–4464, 2009.

[7] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–a review on the statistical data driven approaches," *European journal of operational research*, vol. 213, no. 1, pp. 1–14, 2011.

[8] T. Brotherton, G. Jahns, J. Jacobs, and D. Wroblewski, "Prognosis of faults in gas turbine engines," in *Aerospace Conference Proceedings, 2000 IEEE*, vol. 6, pp. 163–171. IEEE, 2000.

[9] G. Petrone, G. Spagnuolo, R. Teodorescu, M. Veerachary, and M. Vitelli, "Reliability issues in photovoltaic power processing systems," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 7, pp. 2569–2580, 2008.

[10] M. Schwabacher, "A survey of data-driven prognostics," in *Infotech@ Aerospace*, p. 7002, 2005.

[11] B. Saha, K. Goebel, and J. Christophersen, "Comparison of prognostic algorithms for estimating remaining useful life of batteries," *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3-4, pp. 293–308, 2009.

[12] K. Goebel, B. Saha, and A. Saxena, "A comparison of three data-driven techniques for prognostics," in *62nd meeting of the society for machinery failure prevention technology (mfpt)*, pp. 119–131, 2008.

[13] L. Peel, "Data driven prognostics using a kalman filter ensemble of neural network models," in *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pp. 1–6. IEEE, 2008.

[14] A. K. Mahamad, S. Saon, and T. Hiyama, "Predicting remaining useful life of rotating machinery based artificial neural network," *Computers & Mathematics with Applications*, vol. 60, no. 4, pp. 1078–1087, 2010.

[15] S. Morando, S. Jemei, R. Gouriveau, N. Zerhouni, and D. Hissel, "Fuel cells prognostics using echo state network," in *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pp. 1632–1637. IEEE, 2013.

[16] S. Rosiek, J. Alonso-Montesinos, and F. Batlles, "Online 3-h forecasting of the power output from a bipv system using satellite observations and ann," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 261 – 272, 2018.

[17] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in pv arrays: types, detection, and mitigation techniques," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 982–997, 2015.

[18] M. K. Alam, F. H. Khan, J. Johnson, and J. Flicker, "Pv faults: Overview, modeling, prevention and detection techniques," in *Control and Modeling for Power Electronics (COMPEL), 2013 IEEE 14th Workshop on*, pp. 1–7. IEEE, 2013.

[19] M. Aghaei, F. Grimaccia, C. A. Gonano, and S. Leva, "Innovative automated control system for pv fields inspection and remote

[26] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "Comparison of training approaches for photovoltaic forecasts by means of machine learning," *Applied Sciences*, vol. 8, no. 2, p. 228, 2018.

control," *IEEE Transactions on Industrial Electronics*, vol. 62, DOI 10.1109/TIE.2015.2475235, no. 11, pp. 7287–7296, Nov. 2015.

[20] A. Dolara, G. C. Lazaroiu, S. Leva, G. Manzolini, and L. Votta, "Snail trails and cell microcrack impact on pv module maximum power and energy production," *IEEE Journal of Photovoltaics*, vol. 6, DOI 10.1109/JPHOTOV.2016.2576682, no. 5, pp. 1269–1277, Sep. 2016.

[21] P. Baraldi, F. Mangili, and E. Zio, "Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data," *Reliability Engineering & System Safety*, vol. 112, pp. 94–108, 2013.

[22] E. Ogliari, A. Dolara, G. Manzolini, and S. Leva, "Physical and hybrid methods comparison for the day ahead pv output power forecast," *Renewable Energy*, 2017.

[23] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "A physical hybrid artificial neural network for short term forecasting of pv plant power output," *Energies*, vol. 8, no. 2, pp. 1138–1153, 2015.

[24] F. Kasten and G. Czeplak, "Solar and terrestrial radiation dependent on the amount and type of cloud," *Solar energy*, vol. 24, no. 2, pp. 177–189, 1980.

[25] F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "Ann sizing procedure for the day-ahead output power forecast of a pv plant," *Applied Sciences*, vol. 7, no. 6, p. 622, 2017.

[27] M. Rana, I. Koprinska, and V. G. Agelidis, "Forecasting solar power generated by grid connected pv systems using ensembles of neural networks," in *2015 International Joint Conference on Neural Networks (IJCNN)*, DOI 10.1109/IJCNN.2015.7280574, pp. 1–8, Jul. 2015.

[28] M. Omar, A. Dolara, G. Magistrati, M. Mussetta, E. Ogliari, and F. Viola, "Day-ahead forecasting for photovoltaic power using artificial neural networks ensembles," in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, DOI 10.1109/ICRERA.2016.7884513, pp. 1152–1157, Nov. 2016.

[29] C. Monteiro, L. A. Fernandez-Jimenez, A. Ramirez-Rosado, I. J. ans Munoz-Jimenez, and P. M. Lara-Santillan, "Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques," *Mathematical Problems in Engineering*, vol. 2013, DOI 10.1155/2013/767284, p. 9, 2013.

[30] R. Ulbricht, U. Fischer, W. Lehner, and H. Donker, "First steps towards a systematical optimized strategy for solar energy supply forecasting," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2013)*, 2013.

[31] J. Kleissl, *Solar energy forecasting and resource assessment*. Academic Press, 2013.

[32] A. Nespoli, E. Ogliari, F. Grimaccia, S. Leva, and M. Mussetta, "Validation of ANN training approaches for day-ahead photovoltaic forecasts," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 4567–4572, Jul. 2018.

[33] IEC, "Photovoltaic system performance - part 1: Monitoring," *International Electrotechnical Commission 61724-1*, 2017.

[34] A. Dolara, S. Leva, G. Manzolini, and E. Ogliari, "Investigation on performance decay on photovoltaic modules: snail trails and cell microcracks," *IEEE journal of photovoltaics*, vol. 4, no. 5, pp. 1204–1211, 2014.