

1 Maximum Power Point Analysis for Partial Shading Detection and Identification 2 in Photovoltaic Systems

3
4 S. Fadhel^{1,2,4}, D. Diallo², C. Delpha³, A. Migan², I. Bahri², M. Trabelsi⁴, M.F. Mimouni⁴

5 ¹Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, 4023, Sousse, Tunisie

6 ²Université Paris- Saclay, CentraleSupélec, CNRS, Laboratoire Génie Electrique et Electronique de Paris, 91192,
7 Gif-sur-Yvette, France

8 ³Université Paris- Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 91192, Gif-sur-Yvette,
9 France

10 ⁴Ecole Nationale d'Ingénieurs de Monastir, Laboratoire Automatique, Systèmes Électriques et Environnement,
11 5035 Monastir, Tunisie

12

13 **Abstract**

14 Fault diagnosis of photovoltaic (PV) systems is a crucial task to guarantee security, increase
15 productivity, efficiency, and availability. In this regard, numerous diagnosis methods have been
16 developed. Methods requiring the interruption of power production are not adequate for economic
17 reasons. The development of large-scale PV plants and the objective of maintenance cost reduction
18 push toward the emergence of automatic on-line diagnosis methods that use available information. In
19 this study, we propose two data-driven methods for partial shading diagnosis using only the maximum
20 power point's information. It does not require the interruption of production, nor does it require any
21 additional equipment to obtain the I(V) curve. The analyses are conducted with principal component
22 analysis (PCA) and linear discriminant analysis (LDA) to detect and classify the faults. The
23 experimental dataset is collected from a 250 Wp PV module under four states of health (healthy, and
24 three severities of partial shading) for several meteorological conditions. The classification results
25 have a 100% success rate, and are robust to the variations of temperature and irradiance.

26 **Keywords:** Photovoltaic system, Fault diagnosis, Partial shading, Maximum Power Point (MPP),
27 Principal Component Analysis, Linear Discriminant Analysis

28

29

1. Introduction

Although photovoltaic (PV) cells/modules are the most reliable components in a PV installation [1] and considered as systems with low failure rates [2] requiring low maintenance [3], it is crucial to implement effective monitoring systems for PV modules to diagnose their state of health. Indeed, many PV faults can cause significant power losses, may lead to accelerated system ageing, reduced efficiency, and even safety hazards.

Due to uncertainties related to PV cell technology, manufacturing process and climatic conditions, correlating PV fault type to power loss is tricky. It is still an on-going area of research. In particular, the impact of climatic conditions is evaluated in [4], [5] by studying a large number of PV modules exposed to different climates. The annual degradation rate is proposed by numerous studies to quantify the impact of single or multiple faults. For example, it has been reported that crystalline silicon (c-Si) PV modules have degraded at an annual rate of approximately 7% due to discoloration and weakened solder bonds [6]. Another study reported that because of discoloration, the power loss could reach 10% per year in PV modules with concentrators [5]. According to [4], PV installations deployed after the 2000s are more susceptible to the occurrence of defects. However, the percentage of degradation would be significantly lower than in previous installations. In particular, for new c-Si PV modules, the biggest concern seems to be hot spots, followed by potential induced degradation (PID). Hot spots have an impact on PV module performance and could contribute to premature ageing if they are frequent [7]. It has also been reported that partial shading (PS) can lead to several early degradations, and is a significant cause of the appearance of hot spots. Therefore, we are interested in partial shading (PS) diagnosis.

Several PV faults have a visual signature such as delamination, burn, or discoloration marks. So, they can be simply detected through visual inspection. This technique is clearly irrelevant for large PV installations. Moreover maintenance policies are evolving towards Condition Based Maintenance

55 (CBM) to prevent failures and reduce maintenance costs. CBM requires continuous monitoring and
56 automatic fault diagnosis methods.

57 There are several methods for diagnosing PV faults, reviewed in the literature [8] to [11]. They can
58 be classified into two main families according to the measures used:

- 59 - The first group uses radiation-based techniques: Electroluminescence (EL) [12], [13], and
60 Infrared Thermography (IRT) under steady-state conditions [13], [14]. The main advantage of
61 these techniques is their ability to detect and locate the faults in real-time. EL is generally
62 performed to detect micro-cracks in PV cells and other faults within the cell material (contact
63 degradation...); however, it requires the system interruption. IRT can be performed at module
64 level, allowing a fast and effective inspection to detect different types of faults. This
65 technique can be tested without any operational interruption, but it requires mobile platforms
66 to diagnose many PV modules. For this reason, researches are ongoing to develop advanced
67 IRT techniques such as drone-mounted thermography [15]. Therefore, IRT could be more
68 suitable than the EL technique, especially for big PV plants.
- 69 - The second group uses electrical measures. It is composed of reflectometry techniques, I(V)
70 curve analysis, and maximum power point (MPP) coordinates analysis.

71 Reflectometry techniques are mainly used to detect catastrophic faults (ground faults and arc faults)
72 and locate open-circuit faults in the PV strings. These methods are effective. Nevertheless, they are
73 costly since they require a specific external signal function generator to stimulate the PV system [16]–
74 [20].

75 I(V) curves can also be used for fault diagnosis. However, not only do several defects have very
76 similar signatures, but the I(V) curve is also very sensitive to environmental conditions. Therefore
77 diagnosis methods based on the analysis of the entire I(V) curve's points need appropriate

78 preprocessing. This difficulty may be one of the reasons that justify the limited number of publications
79 using the full I(V) curve for fault diagnosis. From the literature review, we have only found two
80 studies dealing with the full I(V) curve analysis for diagnosis purposes [7], [23]. In [7], the analysis of
81 the first derivative of the ratio between the standard error and the PV voltage is used for the same
82 fault, the standard error results from the comparison between faulty and healthy I(V) data. Authors in
83 [23] have proposed the evaluation of the first and the second derivatives of the curve for shading
84 detection. An I(V) plotter has been used in this study to sweep the I(V) curve composed of 265 points.
85 We have proposed in two previous works a new method based on the full I(V) curve analysis to detect
86 and classify shading faults by using adequate variables [21], [22]. The experimental I(V)
87 characteristics are obtained thanks to a programmable voltage source, which provides the points from
88 the short-circuit current to the open-circuit voltage.

89 The other diagnosis techniques use the actual I(V) curve to extract several health indicators from
90 the shape of the curve. The fault diagnosis decision is made upon analysis of the residuals computed as
91 differences between measured indicators and estimated ones identified from the healthy I(V)
92 characteristics. This latter is generally obtained by considering the usual electrical equivalent circuit
93 model (single-diode or double-diode) [24]. At first, an optimization algorithm is applied to extract the
94 five optimal parameters, as proposed in [25]–[29]. Then, the non-linear equation $I = f(V)$ is solved
95 iteratively to reproduce the healthy behavior of the PV system.

96 In [30], the authors have defined five indicators (fill factor, equivalent series resistance...)
97 calculated from the shape of the outdoor measured on-line I(V) curve. Fault occurrence is detected by
98 comparing the actual values with their initial values measured during the commissioning phase of the
99 diagnosis system. Based on the one-diode model, authors in [31] have defined two reference indicators
100 to estimate the increment of the series resistance from the measures of the short-circuit current I_{SC} , PV
101 system temperature T_{PV} , maximum power point's voltage and current. The experimental indicators
102 analysis is performed with a variable load used to extract the full I(V) curve.

103 These methods suffer from several difficulties. First, they introduce additional time to sweep the
104 full I(V) characteristic, and the duration and the quality of acquisition depend on the used instrument.
105 Second, these approaches are confronted with the case of large-scale photovoltaic plants with limited
106 measurements. Furthermore, it is not practical to disconnect PV modules or strings to extract their I(V)
107 curves as it perturbs the system operation and leads to production losses. For instance, the power loss
108 due to the I (V) acquisition in [32] resulted in a power efficiency drop of 0.16%.

109 For this reason, it is not suitable for continuous monitoring. There is therefore, an interest in using
110 methods that do not require the interruption of production and that use only available data such as
111 voltage and current at the point of maximum power (MPP) commonly available in PV systems. This
112 approach is also cost-effective, as it does not require any additional measurement equipment.

113 The fault detection strategy developed in [33] is based on monitoring only the maximum power of
114 the PV system by analyzing the sample entropy. In [34], authors have proposed two diagnosis
115 attributes, which are the voltage ratio (VR) and power ratio (PR) computed from the division of the
116 theoretical and the measured values of the maximum power point's voltage and power, respectively.
117 More recently, to distinguish healthy from faulty states, three diagnosis indicators calculated from the
118 comparison between predicted and measured current, voltage, and power at MPP have been proposed
119 in [26]. The same electrical measurements are exploited in [35], [36], to define three residuals as fault
120 features. The data in these two studies are recorded from the same PV system whose behavior in the
121 healthy state is estimated based on the one-diode model. Residuals are analyzed with the one-class
122 support vector machine (1SVM) technique in [35], while the exponentially weighted moving average
123 (EWMA) approach is applied to the same residuals in [36], to detect and identify the faults. All the
124 previous analysis methods using the maximum power point's coordinates rely on physics-based
125 models. These methods are powerful in detecting many types of PV faults and can be applied to a
126 wide range of PV plants. However, they rely on an analytical model to reproduce the nominal PV
127 characteristics, and an optimization process is implemented for parameter extraction. Subsequently,

128 MPP's coordinates in a healthy state are estimated under real climatic conditions. When dealing with
129 large PV plants, the modeling stage can be quite challenging to reproduce the MPP's coordinates with
130 high accuracy. Besides, due to the sensitivity of PV systems to weather conditions, optimal thresholds
131 must be selected for effective and robust detection. Moreover, physics-based models need to be
132 updated to take into account the natural ageing and degradation of solar cells. Unfortunately, these
133 mechanisms are not yet well known in order to develop reliable models. Data-driven methods based
134 on historical data could be a valuable alternative for PV fault diagnosis.

135 To the best of our knowledge, fault detection based on data-driven approaches using only real data
136 at Maximum Power Point (MPP) data has not been well addressed; only one publication has been
137 found in the literature [3]. Authors have proposed graph-based semi-supervised learning for fault
138 diagnosis. The method is able to detect and identify line-to-line and open-circuit faults efficiently, but
139 it is not suitable for partial shading detection.

140 It should be remembered that partial shading (PS) is responsible for several early degradations, and
141 is an essential cause of the appearance of hot spots. That is why, in what follows, we focus on the
142 diagnosis of partial shading (PS) using a methodology based on maximum power point data.

143 The methodology is based on four steps:

- 144 - Modelling step in which the data is collected from an experimental PV module of 250 Wp,
- 145 - Preprocessing step in which the variables (voltage, current, and power at the maximum
146 power point) are selected, centered, reduced and processed through the logarithmic
147 function,
- 148 - Features extraction: in this step two statistical techniques, principal component analysis
149 (PCA) and linear discriminant analysis (LDA) are applied.

150 - Features analyses: in this step, PCA and LDA are used for partial shading faults detection
151 and classification. LDA is also used to determine discriminative laws for data separation.

152 We will evaluate the robustness of the method to environmental conditions: variable temperature
153 and irradiation.

154 The outline of the paper is as follows: section 2 describes how the database is collected and
155 extended. The detailed faults diagnosis methodology based on maximum power point (MPP)
156 coordinates are presented in section 3. A case study of an isolated PV module is presented to validate
157 the proposed methods. Section 4 discusses the application to large-scale PV plants and compares its
158 practical implementation with full I (V)-based diagnosis technique. Finally, section 5 concludes the
159 paper.

160

161 **2. Description of the dataset**

162

163 *2.1. Data acquisition within a 15-minute time window*

164 The experiments were carried out on March 21, 2019, using the experimental setup displayed
165 in Fig. 1. The main parameters of this module at standard test conditions (STC) are as follows
166 [22]: the maximum power $P_{mpp} = 250$ Wp, the maximum power current $I_{mpp} = 8.21$ A, the
167 maximum power voltage $V_{mpp} = 30.52$ V, the short-circuit current $I_{sc} = 8.64$ A, the open-circuit
168 voltage $V_{oc} = 37.67$ V. The temperature of the PV module T_{PV} is measured with a 4-wire Pt100
169 probe bonded on its back face.

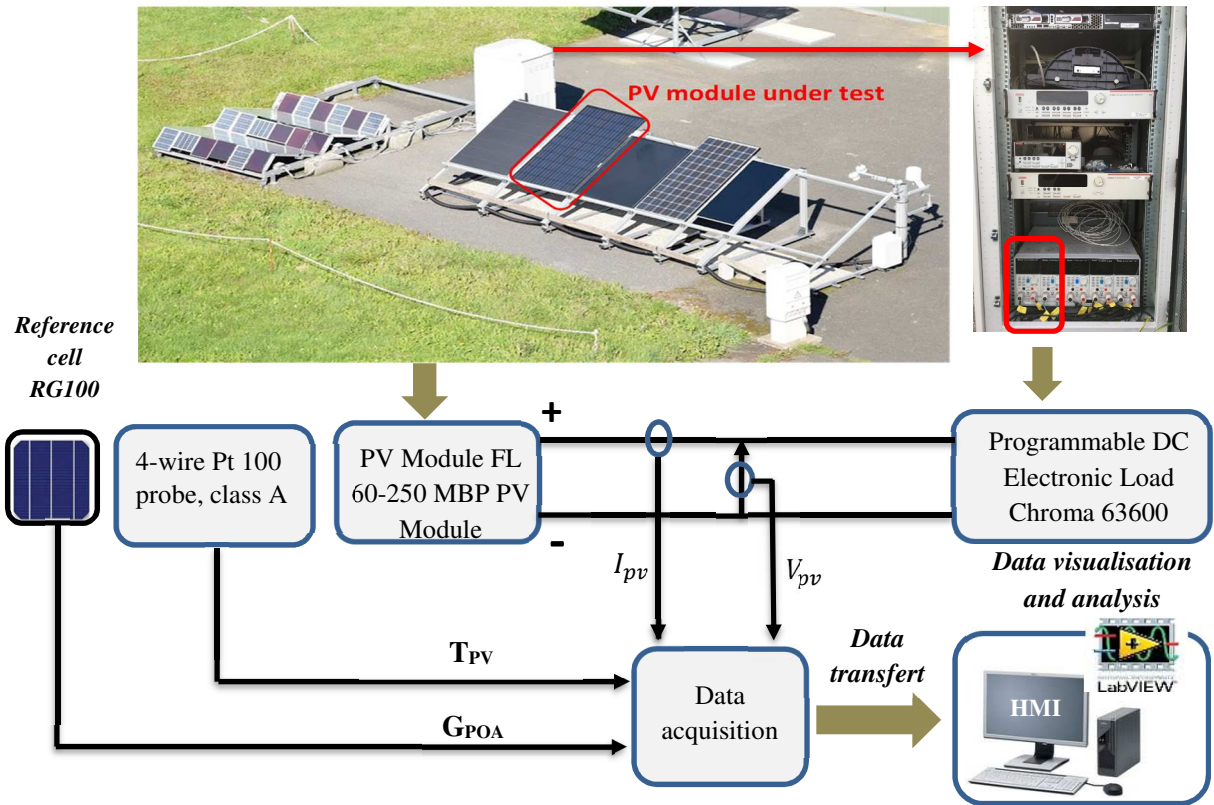


Fig. 1. Installed PV panel platform

170
171

172

173 A reference cell (SOLEMS RG100) is used to measure G_{POA} the plane of array irradiance.

174 The acquisition, display, and saving of each I (V) characteristic take one minute. Therefore, during 15

175 minutes, 15 experimental I(V) curves for each health status (healthy and the different shading profiles)

176 are recorded for several levels of temperature T_{PV} and irradiance G_{POA} . Each curve is composed of 101

177 points. As an example, for the environmental conditions T_{PV} and G_{POA} presented in Fig. 2, the

178 characteristics I(V) and P(V) under healthy status are displayed in Fig. 3. We can observe that due to

179 the temperature and irradiance fluctuations ($(\Delta G_{POA})_{max} = 25\%$ and $(\Delta T_{PV})_{max} = 13\%$), the

180 maximum power fluctuates from 149 Wp to 190 Wp.

181

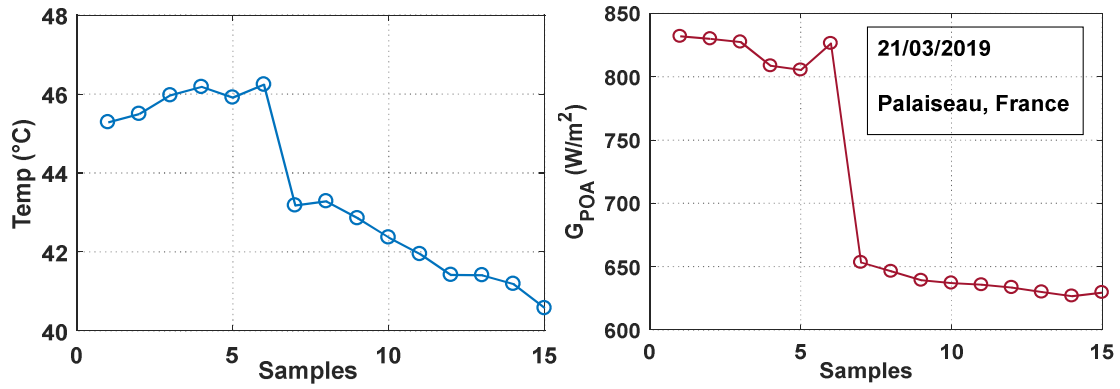


Fig. 2. Module temperature and irradiance variations for a healthy PV module

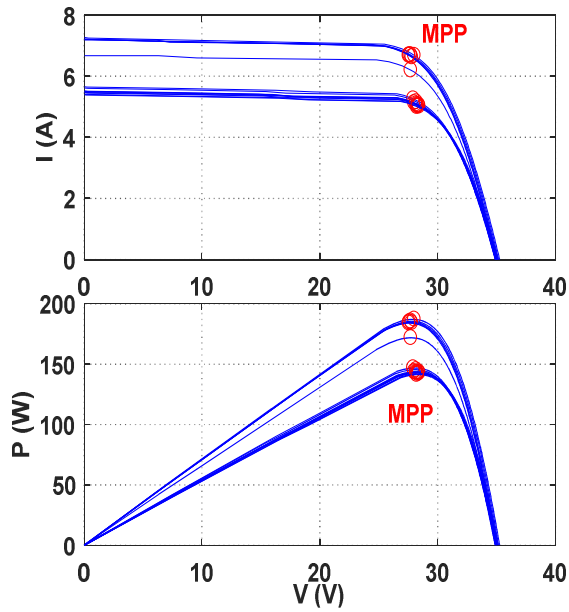


Fig. 3. Measured I(V) and P(V) curves for a healthy PV module

182

183 The shading profiles as in [22], and the corresponding experimental I(V) and P(V) curves are
 184 displayed in Fig. 4. One can notice the maximum power drop caused by the bypass diodes activation.

185 The solar irradiance and the module temperature measured during the partial shading experiments are
 186 plotted in Fig. 5. The maximum relative variations of irradiance and temperature are summarized in
 187 Table 1.

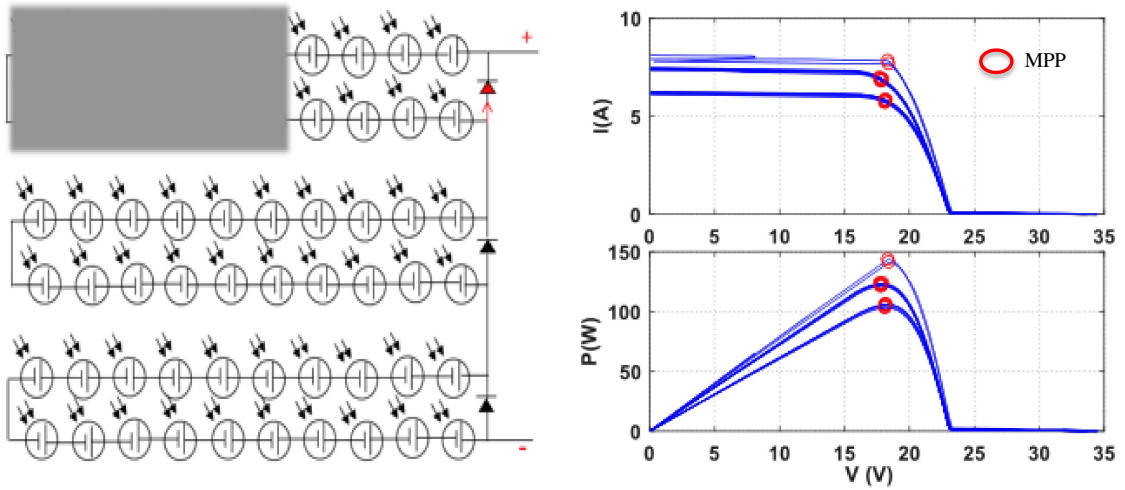
188

189 **Table 1.** Maximum relative variations of irradiance and temperature

| Partial Shading Condition | $\Delta(G_{POA})_{max}$ (%) | $\Delta(T_{PV})_{max}$ (%) |
|---------------------------|-----------------------------|----------------------------|
| 1 | 36 | 18.5 |
| 2 | 6 | 32 |
| 3 | 16 | 4.5 |

190

191

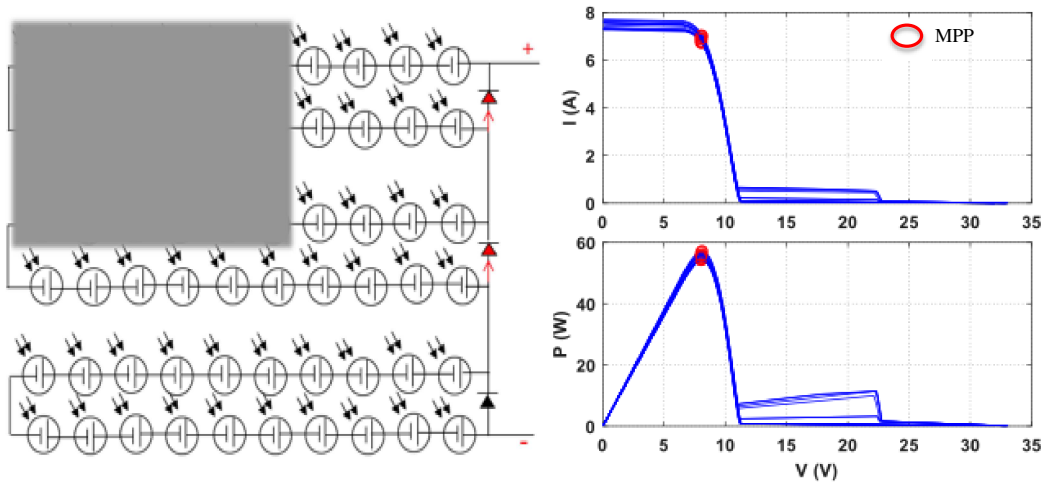


192

193

(a) Partial shading 1

194



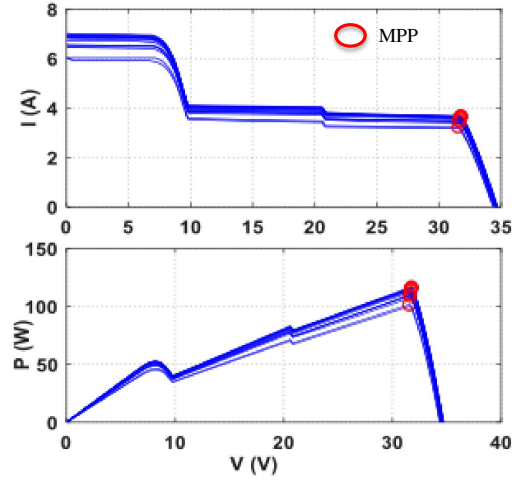
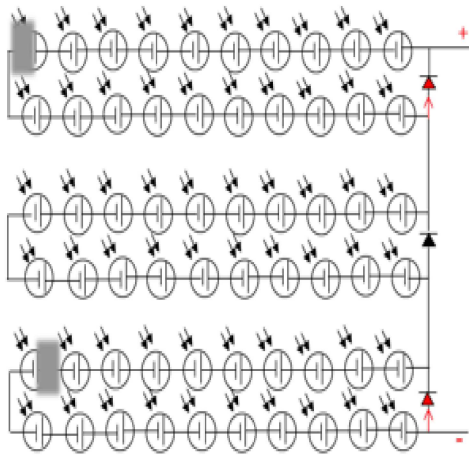
195

196

(b) Partial shading 2

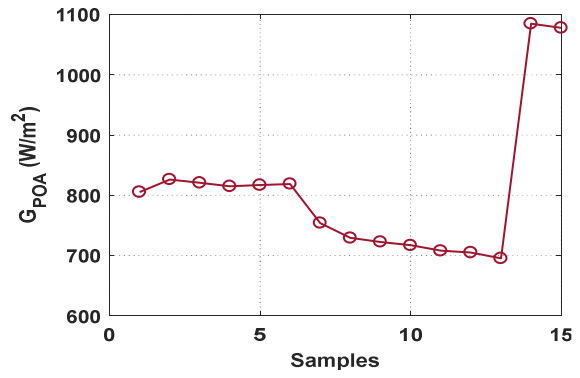
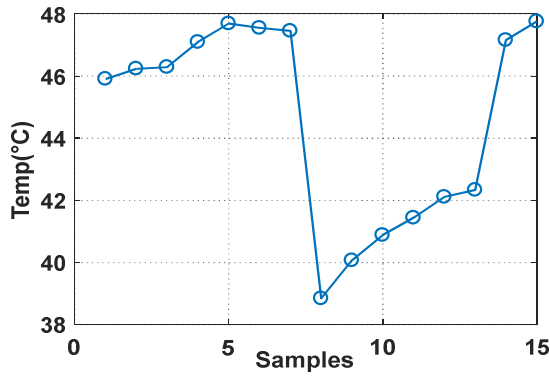
197

198

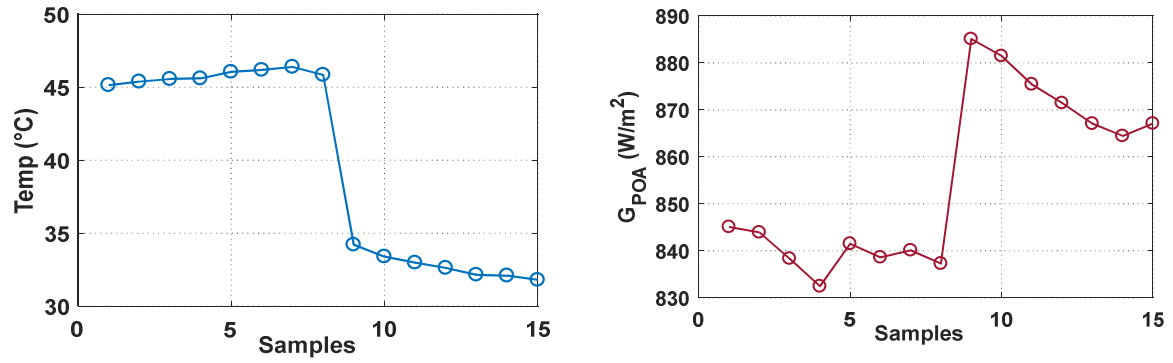


(c) Partial shading 3

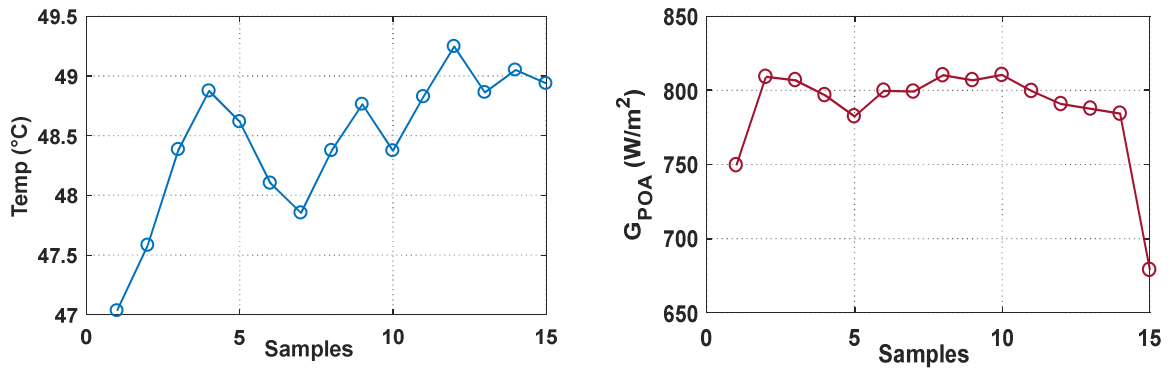
Fig. 4. Partial Shading profiles and measured I (V) and P (V) curves



(a) Partial Shading 1



(b) Partial Shading 2



(c) Partial Shading 3

199

200

Fig. 5. Module temperature and irradiance variations for the different partial shading conditions

201

202 2.2. Extension of the database

203

204

To increase the number of points in the database, we have generated new I(V) curves from the

205

experimental ones by varying the solar irradiance. In the following, we have doubled the number of

206

points; from the original 15 maximum power points for each health status, we have now 30 points.

207

Therefore, for the four modes (one healthy state and three partial shading ones), we have 30*4 MPP

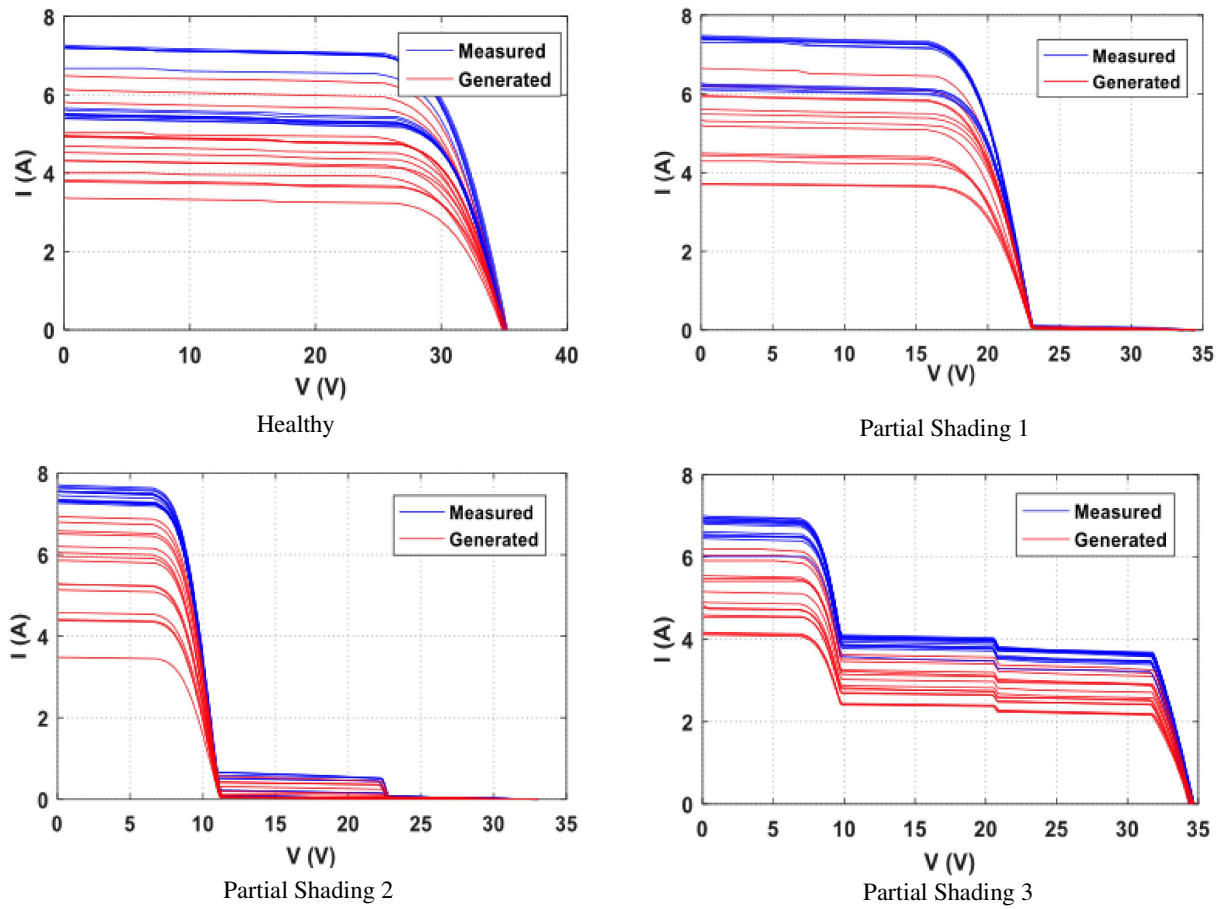
208

observations. Experimental and generated I(V) curves are plotted respectively with blue and red colors

209

in Fig. 6.

210 It must be noticed that MPP coordinates are almost always available in PV plants under operation,
211 thanks to the global maximum power point tracking (GMPPT) algorithm [37, 39].



212

213

214

Fig. 6. $I(V)$ curves of the extended database

215 We now have a database with a total of 120 maximum power points representative of four health
216 conditions for the PV panel. It should also be noted that these data are representative of the variability
217 inherent in real-world measurements in an outdoor environment. In the next section,

218

219 **3. Proposed methodology for fault diagnosis**

220

221 The proposed methodology for faults diagnosis can be described in four main steps, as presented in
222 the flowchart displayed in Fig. 7:

- 223 • *The modelling*: For this step, we have chosen a data-driven model based on the 120 maximum
224 power point coordinates extracted from the I (V) curves.
- 225 • *Data pre-processing*: This step is decisive for the performance level of the diagnostic method.
226 It consists in choosing the field of information, the descriptive variables, and in implementing
227 any technique allowing to keep the information useful and to eliminate nuisances. In our
228 study, we used data from the maximum power point in the time domain, to which we applied
229 the logarithmic function as in [22].
- 230 • *Features extraction and analysis*: The third and fourth steps aim to extract from the
231 maximum power points, the features (signatures) to be analyzed to detect, isolate, and
232 estimate the fault severity. One of the difficulties lies in choosing the most efficient tools for
233 the extraction and analysis of signatures, especially since in our case the data varies with the
234 environmental conditions. This sensitivity makes it difficult to distinguish the healthy state
235 from the faulty ones. Therefore, to ensure reliable fault detection, the most appropriate tool
236 should have the highest ability to separate the data from different health status. In this work,
237 we have selected PCA and LDA for features' extraction and analysis to perform the PS
238 detection and classification.

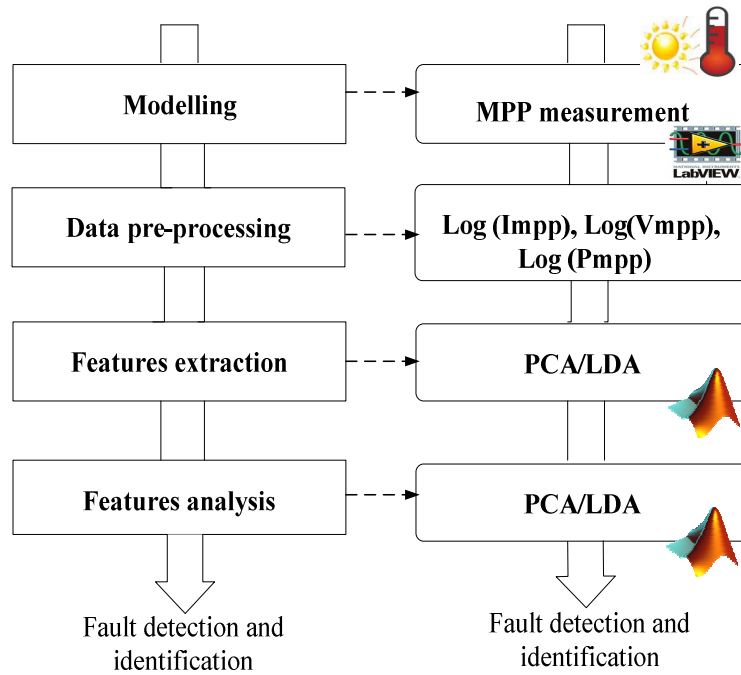


Fig. 7. Flowchart of the four-step diagnosis methodology

239
240

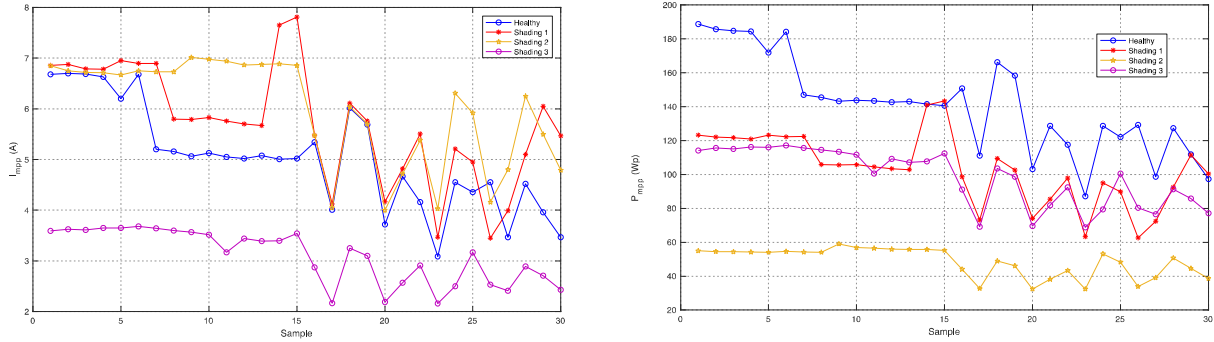
241

242 3.1.Review of normalization methods in the field of PV

243 When dealing with maximum power point (MPP) in outdoor conditions, it might not be easy to
 244 distinguish healthy status from shading ones. The maximum power point's coordinates change rapidly
 245 due to variations in environmental conditions, and it becomes then difficult to differentiate whether a
 246 fault or climatic changes cause the MPP variation. In other words, the faulty PV system may have
 247 similar MPP's coordinates as a healthy one. Besides, the fluctuation of MPPs for different fault types
 248 may be the same. This is confirmed by the evolutions of the maximum power points' coordinates
 249 plotted in Fig. 8. For the current at MPP, I_{mpp} , we can observe the overlapping of several points in
 250 healthy conditions, partial shading 1, and partial shading 2. For the power at the MPP, P_{mpp} , the
 251 overlapping is observed in healthy conditions, partial shading 1, and partial shading 3. These overlaps
 252 can be expected to be more pronounced when the data acquisition interval is greater. Therefore, to

253 ensure reliable fault detection, one should find appropriate processing techniques that highlight the
 254 data's separability.

255



256

257 **Fig. 8.** Overlapping MPP's coordinates over a wide range of temperature and irradiance

258 Several methods of normalization have been proposed in the literature to mitigate the effect of the
 259 environmental changes. Authors in [3] have defined two normalized parameters. The first one is the
 260 ratio of MPP's voltage by the reference open circuit voltage multiplied by N_{mod} , the number of series-
 261 connected PV modules. The second is the ratio of MPP's current by the reference short-circuit current
 262 I_{sc_ref} multiplied by N_{str} the number of parallel strings in the array. This technique is easy to
 263 integrate within a PV inverter whatever its topology. However, it requires the use of two additional
 264 modules identical to the PV modules under test. They are considered as modules of reference to
 265 measure instantaneous I_{sc_ref} and V_{oc_ref} with respect to solar irradiance and module temperature.
 266 This normalization is efficient for detecting and classifying line-to-line and open-circuit fault, but it
 267 cannot be used for partial shading. However, the main issue with this method is to guarantee that the
 268 modules under test, and the modules of reference are in the same conditions of temperature and
 269 irradiance. Another normalization method is proposed in [33]. The PV array current, and voltage, are
 270 divided by the STC short-circuit current, and the STC open-circuit voltage, respectively to calculate
 271 the normalized parameters, used to determine the entropy of the PV array output power. The method is

272 fast and effective for fault detection, however fault classification has not been considered. A per-unit
 273 standardization has been applied in two recent studies [29], [40]. This normalization requires accurate
 274 measurements of temperature and irradiance. It also assumes that the temperature coefficients of
 275 voltage and current are constant. Nevertheless, several experiments on the PV modules installed at the
 276 French observatory SIRTa have revealed that these coefficients vary depending on the meteorological
 277 data. Based on this normalization, the diagnosis method in [40] achieved high fault identification
 278 percentages for all the studied faults except for partial shading. In that case, the rate of correct
 279 identification is 53.66%, which is quite low.

280 Table 2 summarizes the normalization methods discussed above and the data size used to evaluate
 281 each diagnosis approach.

282 **Table 2.** Normalized expressions used for PV system diagnosis to mitigate the climatic conditions effect

| Variable | Normalized expression | Size of PV data | Reference |
|---------------------------------------|--|--|-----------|
| Normalized current and voltage at MPP | $I_{mppn} = \frac{I_{mpp}}{N_{str} \times I_{sc_ref}}$ | <u>Simulation test</u> : 5576 MPP data 1.43% used for training | [3] |
| | $V_{mppn} = \frac{V_{mpp}}{N_{mod} \times V_{oc_ref}}$ | <u>Experimental test</u> : 5000 MPP data 1% used for training | |
| Normalized current and voltage at MPP | $I_{mppn} = \frac{I_{mpp}}{I_{sc0}}$ $V_{mppn} = \frac{V_{mpp}}{V_{oc0}}$ | No training step | [33] |

| | | | |
|--|---|---|------|
| Per unit normalization of Current and voltage at MPP, ideality factor, open-circuit voltage and short-circuit current. | $I_{mppn} = \frac{I_{mpp}}{I_{mpp0}} \frac{G}{1000} [1 + \alpha \frac{I_{mpp0}}{I_{sc0}} (T - 25)]$ $V_{mppn} = \frac{V_{mpp}}{V_{mpp0}} [1 + \beta \frac{V_{mpp0}}{V_{oc0}} (T - 25)]$ $I_{scn} = \frac{I_{sc}}{I_{sc0}} \frac{G}{1000} [1 + \alpha (T - 25)]$ $V_{ocn} = \frac{V_{oc}}{V_{oc0}} [1 + \beta (T - 25)]$ | 123 I-V data 123/3 used for training | [40] |
|--|---|---|------|

| | | | |
|--|---|--|------|
| Per unit normalization of Current and voltage at MPP, ideality factor, open-circuit voltage and short-circuit current. | $n_n = \frac{n}{n_0} [1 + \alpha (T - 25)]$ | <u>Simulation test</u> : 4800 I-V data 75% used for training <u>Experimental test</u> : 2500 I-V data 75% used for training | [29] |
|--|---|--|------|

283

284 *3.2.Feature Extraction and Analysis with Principle Component Analysis*

285 In our case, before applying PCA, the three variables, I_{mpp} , V_{mpp} , and P_{mpp} are centred (zero
 286 mean), and reduced (unit variance). The challenge is also to deal with a limited number of samples
 287 (thirty points per health status). The robustness of our proposal to T_{PV} and G_{POA} variations will be
 288 evaluated.

289 *3.2.1. Principle*

290 The details of PCA are given in [22]. Briefly, PCA is a multivariate unsupervised technique that
 291 uses a linear combination of m original variables to create a low-dimensional set of new variables
 292 named principal components (PCs). The new variables are uncorrelated, and the new dataset is
 293 considered optimal when the variance of the information is maximal despite the reduction in the size

294 of the representation space. The original data is arranged in a matrix $X_{[N \times m]}$ where N represents the
 295 number of measurements/ observations for each of the m variables. The observations include data on
 296 healthy and faulty operating conditions. In the following, the PCA algorithm is applied to the matrix
 297 $X = \log[V_{mpp} \quad I_{mpp} \quad P_{mpp}]$.

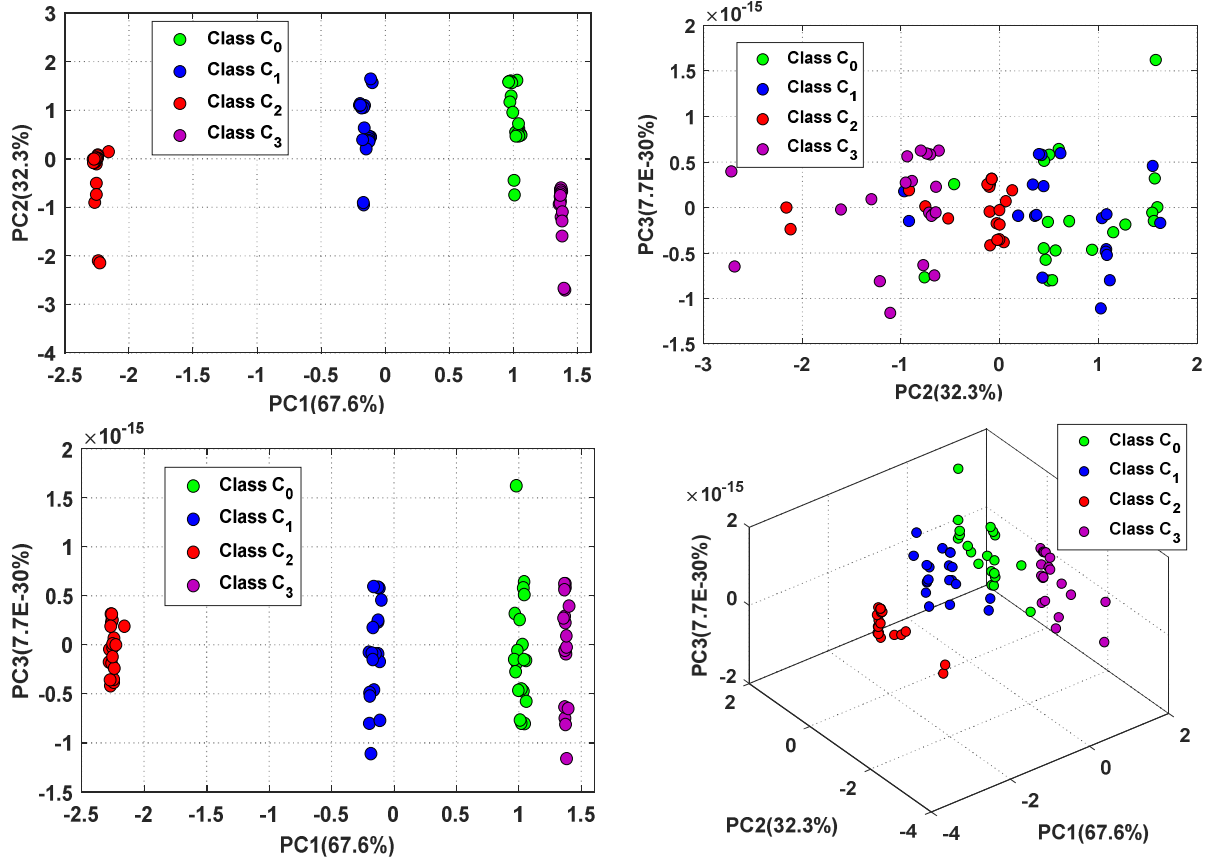
298 3.2.2. Results in the training stage with the PCA algorithm

299 In the training step, 80 observations representing 67% of the dataset is used. The PCA is
 300 applied to the matrix $X_{[80 \times 3]} = \log[V_{mpp} \quad I_{mpp} \quad P_{mpp}]$. We obtain three principal components PC_{*i*} (*i* =
 301 1,2,3), among which PC1 and PC2 preserve 99.9% of the initial total variance. The data can be
 302 projected in four different sub-spaces spanned by PC1-PC2, PC1-PC3, PC2-PC3, or PC1-PC2-PC3
 303 respectively. The projections of the training data are displayed in Fig.9. Except in the subspace
 304 spanned by PC2 and PC3 (there is less information in PC2, and PC3 compared to PC1), the four
 305 classes can be clearly distinguished in the other three subspaces: one healthy class C0 and three faulty
 306 ones noted C1, C2, and C3 (Table 3). The dispersion of the operating points found in the same group
 307 is due to the climatic conditions variations. Table 4 summarizes the percentages of classification, and
 308 shows that all the data samples are correctly classified. In this table, we also present the classification
 309 performance when using all the points of the I(V) curve in [22]. Despite the variations of T_{PV} and
 310 G_{POA} , and using only MPP's coordinates, the classification obtained in the present study is more
 311 accurate with 100% rate of classification.

312 **Table 3.** The four health status classes

| Test condition | Class |
|----------------|----------------|
| Healthy | C ₀ |
| Shading 1 | C ₁ |
| Shading 2 | C ₂ |
| Shading 3 | C ₃ |

313



314 **Fig. 9.** Projection of the training Maximum Power Points' coordinates in the PCA subspaces

315

316 **Table 4.** Confusion matrix for training dataset classification

| A priori Class | A posteriori class | | | | | | | |
|----------------------|----------------------|-------|----------------------|-------|----------------------|-------|----------------------|-------|
| | Class C ₀ | | Class C ₁ | | Class C ₂ | | Class C ₃ | |
| | MPP | I (V) | MPP | I (V) | MPP | I (V) | MPP | I (V) |
| Class C ₀ | 100 | 97.03 | 0 | 0 | 0 | 2.97 | 0 | 0 |
| Class C ₁ | 0 | 0 | 100 | 98.52 | 0 | 1.48 | 0 | 0 |
| Class C ₂ | 0 | 12.62 | 0 | 0 | 100 | 87.38 | 0 | 0 |
| Class C ₃ | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 100 |

317

318 3.2.3. Results in the validation stage with the PCA algorithm

319

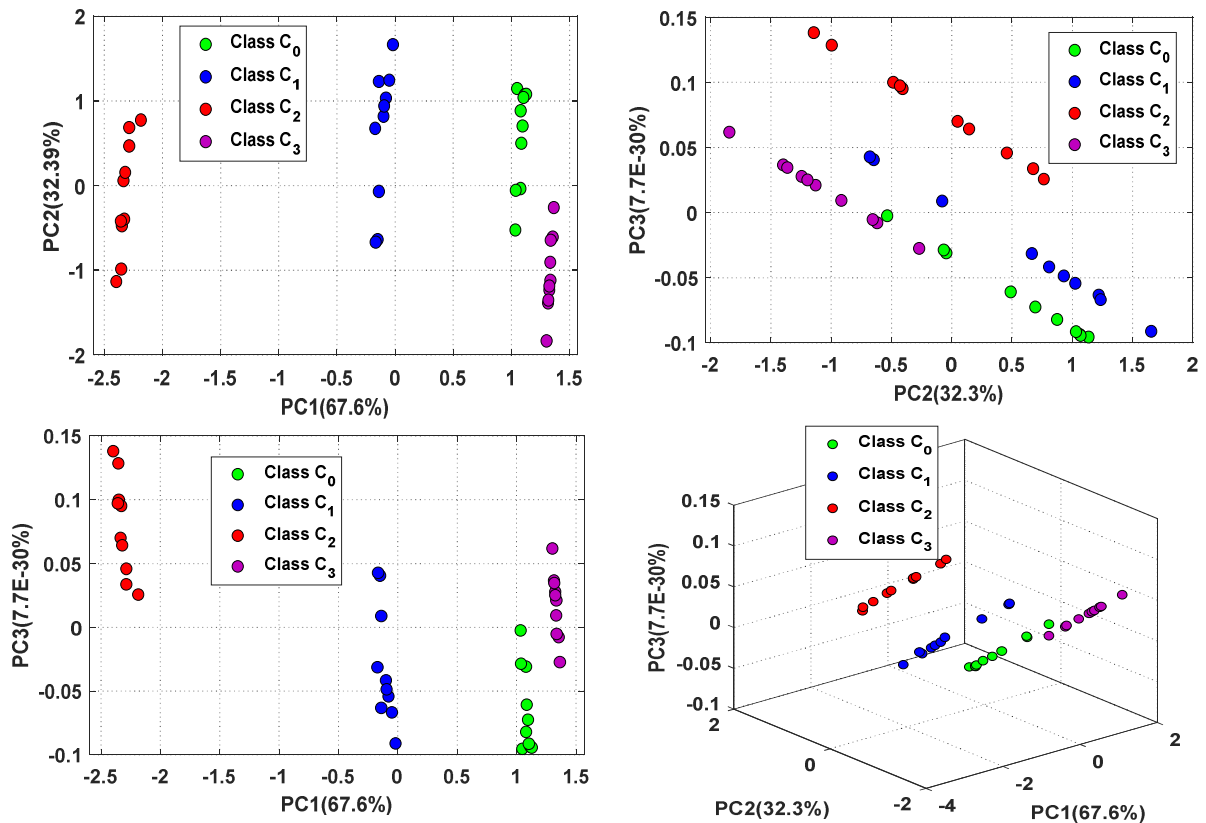
320 The remaining 33% of the database is used to test the effectiveness of the trained PCA model. The

321 test matrix is $X_{[40 \times 3]} = \log[V_{mpp} \quad I_{mpp} \quad P_{mpp}]$ where 10 test samples are used for each health status. The

322 projections of the data in the subspaces spanned by the principal components obtained in the training
 323 stage are displayed in Fig.10. The classes related to partial shading 1, 2, and 3 are clearly separated
 324 from the healthy one with a classification rate of 100%. These results are also compared with those
 325 obtained when using all the points of the curve I(V) (Table 5). They confirm that the analysis of the
 326 maximum power point's coordinates with the PCA is effective to detect and classify the partial
 327 shading (PS) accurately.

328 The justification for this high classification rate lies in the fact that partial shading more strongly
 329 affects the point of maximum power, which is at the inflexion point of the I(V) curve. Therefore, using
 330 only this information makes the classification more effective than using all the points of the I(V)
 331 curve.

332



333

334

Fig. 10. Projection of the test data in the PCA subspaces

335

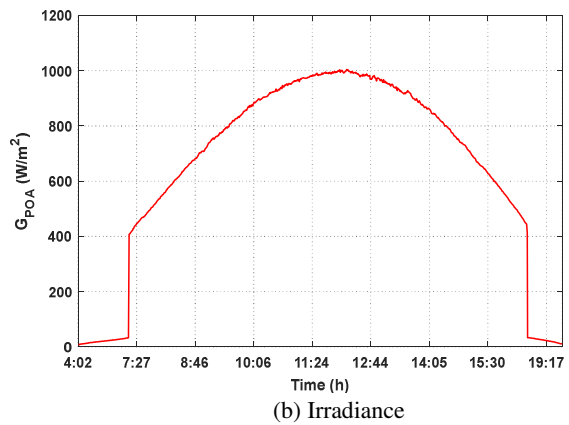
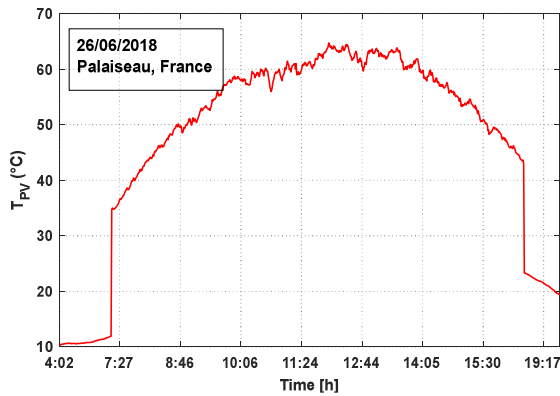
336 **Table 5.** Confusion matrix for test data classification

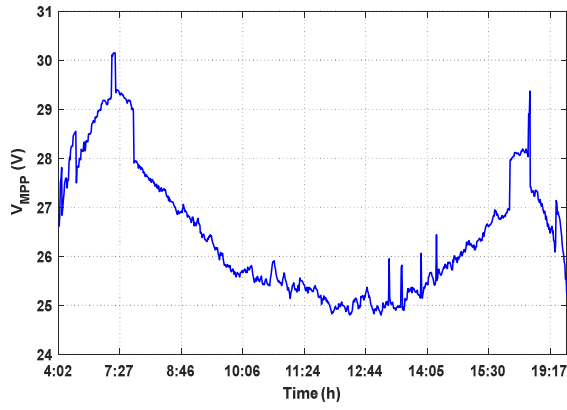
| A priori class | A posteriori class | | | | | | | |
|----------------------|----------------------|-------|----------------------|-------|----------------------|------|----------------------|------|
| | Class C ₀ | | Class C ₁ | | Class C ₂ | | Class C ₃ | |
| | MPP | I(V) | MPP | I(V) | MPP | I(V) | MPP | I(V) |
| Class C ₀ | 100 | 97.03 | 0 | 0 | 0 | 2.97 | 0 | 0 |
| Class C ₁ | 0 | 0 | 100 | 99.01 | 0 | 0.99 | 0 | 0 |
| Class C ₂ | 0 | 0 | 0 | 0 | 100 | 100 | 0 | 0 |
| Class C ₃ | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 100 |

337

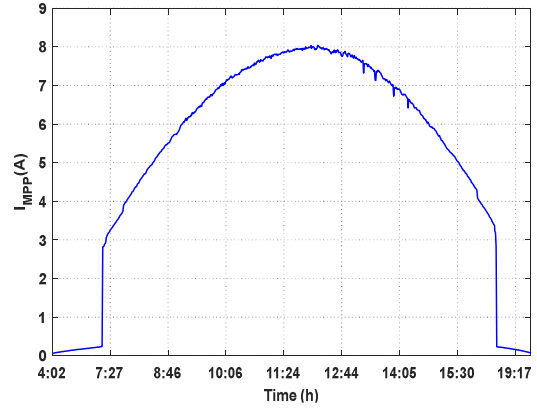
338 *3.2.4. Robustness of the classification*

339 This section aims to validate the robustness of the fault diagnosis method based on PCA and
 340 maximum power point (MPP) coordinates. To do so, previously used faulty MPP data (collected on
 341 March 21, 2019), are merged with MPP data collected during healthy operation on June 26, 2018. The
 342 corresponding daily profiles of temperature and solar irradiance, as well as the MPP voltage and
 343 current evolution under these conditions, are plotted in Fig.11. The robustness is evaluated using the
 344 data extracted at different times, and for moderate and high irradiation values. Table 6 presents the
 345 new samples used and the ranges of variation of T_{PV} and G_{POA} .





(c) Voltage at MPP



(d) Current at MPP

Fig. 11. Clear day profiles

346

347

348

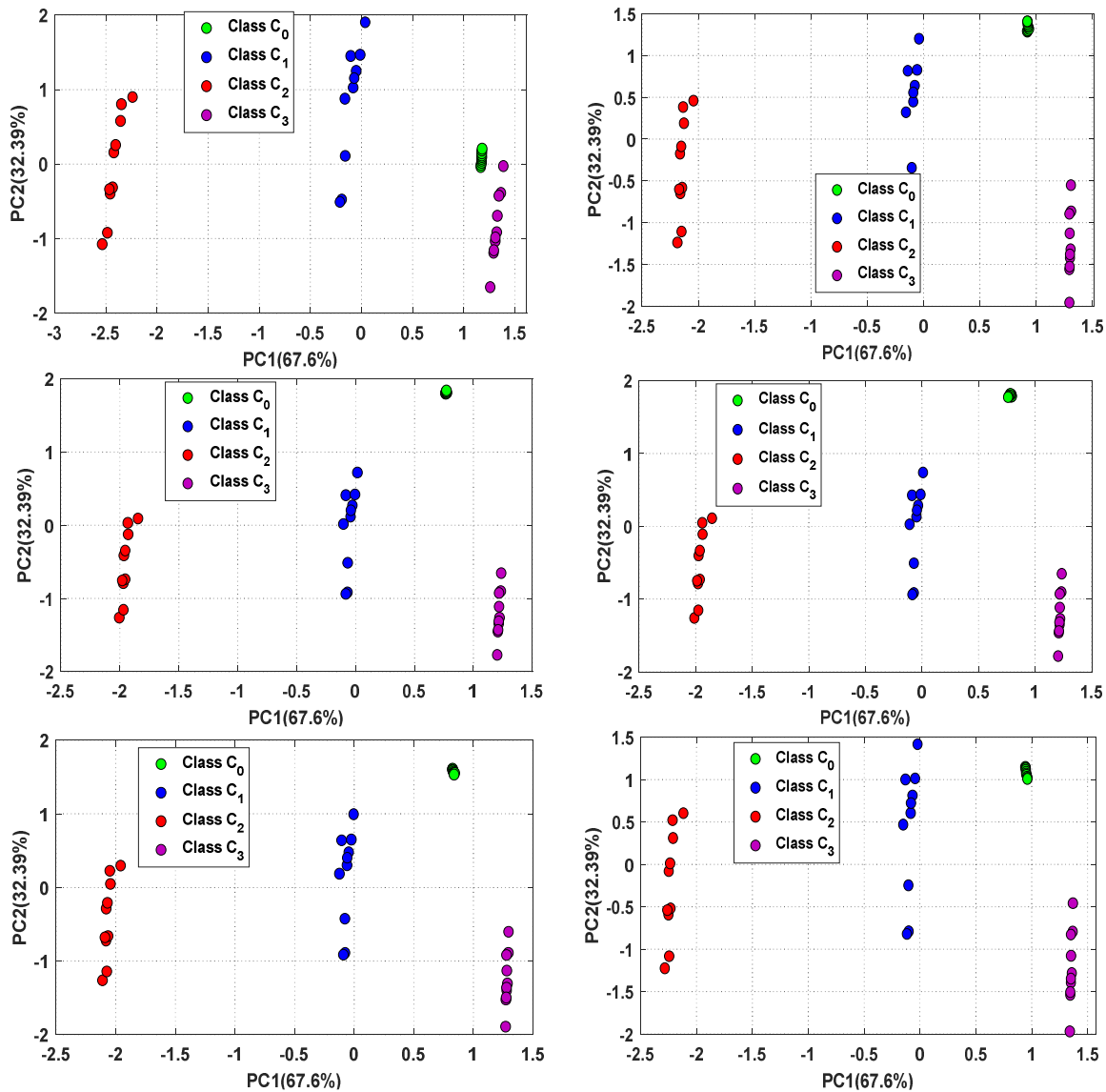
Table 6. Extracted samples for robustness evaluation

| Samples | Time (h) | T_{PV} (°C) | G_{POA} (W/m ²) |
|-------------|-----------------|-----------------|-------------------------------|
| [100 ; 109] | [7h26 ; 7h33] | [36.46 ; 37.9] | [445.53 ; 468.46] |
| [200 ; 209] | [8h44 ; 8h52] | [48.55 ; 49.81] | [680.71 ; 703.64] |
| [450 ; 459] | [12h03 ; 12h10] | [63.18 ; 64.19] | [992.10 ; 1004.50] |
| [500 ; 509] | [12h43 ; 12h50] | [59.68 ; 62.49] | [967.88 ; 982.80] |
| [621 ; 630] | [14h22 ; 14h29] | [56.93 ; 58.4] | [797.53 ; 815.39] |
| [700 ; 709] | [15h29 ; 15h37] | [48.96 ; 50.93] | [606.28 ; 628.28] |

349

350 The new matrix with the merged data $X_{[40 \times 3]} = \log[V_{mpp} \ I_{mpp} \ P_{mpp}]$ is projected in the PCA
 351 subspace spanned by the two principal components (PC1, PC2) obtained in the learning stage with the
 352 data collected on March 21, 2019. The results are presented in Fig.12. We observe that class C_0
 353 formed with the new healthy measurements is clearly separated from the other three classes of data
 354 corresponding to partial shading (PS) conditions. A successful classification of 100% is obtained.
 355 Moreover, we notice that all the new test data is positioned between classes C_2 and C_3 , as observed in
 356 the previous validation stage. These results are encouraging and prove the robustness of the method
 357 for PS detection and classification of PV modules.

358 When looking at the maximum power point's coordinates projected in the PCA subspace, one can
 359 observe that classes can be linearly separated. Therefore in the following, linear discriminant analysis
 360 (LDA) is used to retrieve the classification laws for fault identification.



361 **Fig. 12.** Projection of the merged dataset in PCA subspace (PC1, PC2)

362

363

364 *3.3. Features analysis and extraction with Linear Discriminant Analysis (LDA)*

365

366 *3.3.1. Formulation of the LDA algorithm*

367 The LDA algorithm used in [41] [42] [43] will be extended for application to photovoltaic system
368 diagnosis. LDA and PCA share a common characteristic, which is dimension reduction. Nevertheless,
369 theoretically, LDA is more powerful for data classification than PCA because LDA performs
370 dimensionality reduction while keeping maximum the discriminatory information among the classes.
371 For this reason, LDA is generally used to improve data classification when different data classes
372 overlap with the PCA algorithm.

373 Unlike the PCA, the LDA is a supervised analysis technique. It involves a learning phase during which
374 c a priori groups are defined with M descriptive variables and N observations or samples distributed
375 into a priori groups, as shown in Figure 13. LDA is both descriptive and decisive. In its descriptive
376 phase, it reduces the size of the data and separates the data according to their a priori defined groups.
377 Linear combinations of the variables in each group create $c - 1$ discriminant axes. Each axis is assigned
378 a percentage of the data variance, which is calculated from the corresponding eigenvalue. As in PCA,
379 the discriminant axes are ordered in descending order of eigenvalues. In its decisive phase, the LDA
380 generates $c - 1$ hyperplanes or separation laws that characterize the differences or similarities between
381 the groups allocated a priori. Thanks to these laws, it is possible to classify new samples of data
382 described by the same explanatory variables without a priori knowledge of their group.

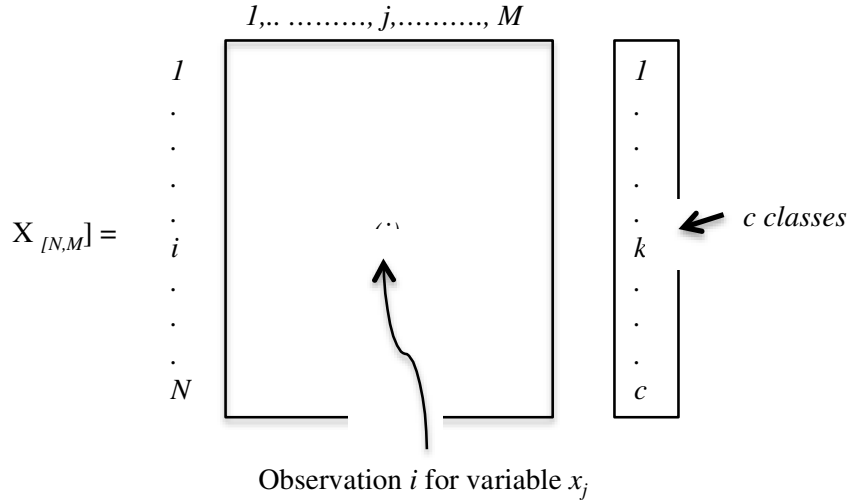


Fig. 13. Database organization for LDA algorithm

383

384

385 To perform best linear combinations, LDA aims to obtain a maximum interclass variance which
 386 highlights the differences between the classes, and a minimum intraclass variance [44]. The interclass
 387 and intraclass variances are respectively estimated by the between matrix variance S_B and the within
 388 matrix variance S_w .

389 Mathematically formulating, this analysis technique is based on maximizing the following objective
 390 function, also called Fisher criterion [45]:

391

$$J(u) = \frac{u^T S_B u}{u^T S_w u} \tag{1}$$

392

393 where u is the linear expression that best discriminates all classes of data.

394 Given N observations (x_1, x_2, \dots, x_N) of M centered reduced explicative variables, N_i , the total number
 395 of observations for each class C_i , $i = 1, 2, \dots, c$.

396 Let's consider μ_i as the mean value of observations belonging to C_i and μ as the mean value of the
 397 entire dataset given by:

$$\mu_i = \frac{1}{N_i} \sum_{x_n \in C_i} x_n \quad (2)$$

398

$$\mu = \frac{1}{N} \sum_{n=1}^N x_n \quad (3)$$

399 The between variance matrix can be expressed as:

$$S_B = \sum_{i=1}^c (\mu_i - \mu)^T (\mu_i - \mu) \quad (4)$$

400 The within variance matrix is calculated as:

$$S_w = \sum_{i=1}^c S_i \quad (5)$$

401 where :

402

$$S_i = \sum_{x_n \in C_i} (x_n - \mu_i)^T (x_n - \mu_i) \quad (6)$$

403

404 The eigenvector decomposition of $S_w^{-1}S_B$ gives the discriminant axes and their corresponding
 405 eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{c-1}$ that indicate the discriminant power of each axis. The eigenvectors
 406 associated with the largest eigenvalues define the discriminant axes that span the LDA space for
 407 classification. The original data centered reduced are finally projected onto this space to give the data
 408 class separability.

409 In the decisive phase, LDA uses three quantities to classify the new data samples in one of the defined
 410 a priori groups: posterior probability, prior probability, and cost.

411 The posterior probability that a new instance or observation I , described by the explicative variables, is
 412 classified into the class C_i is estimated by the Bayes formula:

$$\hat{P}(C_i / l) = \frac{P(l / C_i) \cdot P(C_i)}{P(l)} \quad (7)$$

413 where $P(C_i)$ represents the prior probability of class C_i .

414 With this probabilistic approach, the unknown observation l to be identified will be assigned to the
 415 class C_i with the highest posterior probability and minimum expected classification cost:

$$\hat{y} = \arg \min_{y=1, \dots, c} (P(C_1 / k)C(y / C_1) + \dots + P(C_i / k)C(y / C_i) + \dots + P(C_c / k)C(y / C_c)) \quad (8)$$

416 Where \hat{y} is the predicted class, c is the number of classes and $C(y / C_i)$ is the cost of classifying an

417 observation as y when its true class is C_i .

418 *3.3.2. Results of the training stage with LDA algorithm*

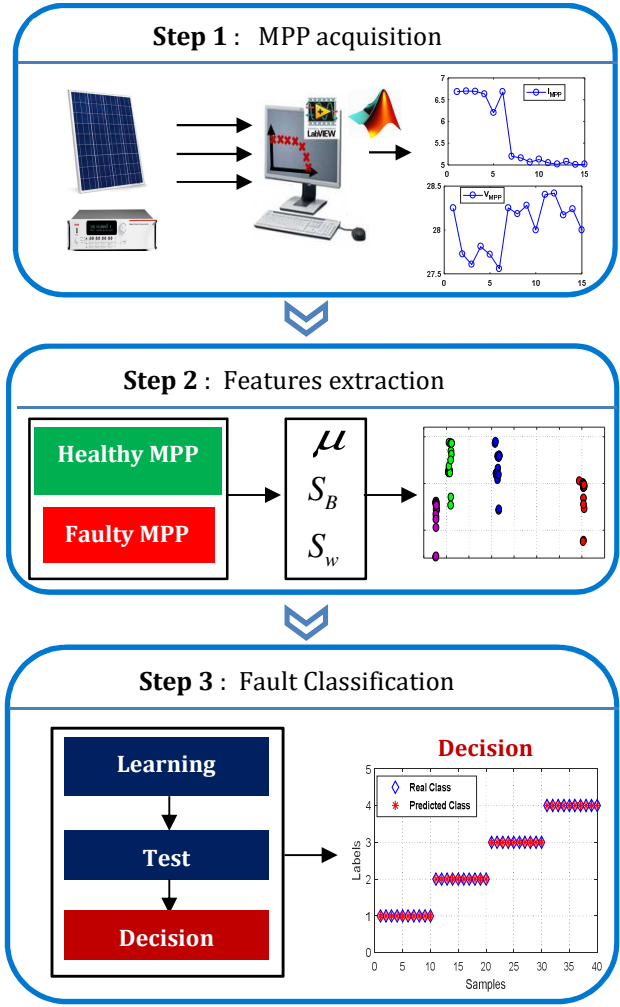


Fig. 14. Fault identification flowchart with LDA algorithm

419

420

421 The flowchart of the proposed algorithm is shown in Fig.14. The LDA is applied to extract the
 422 faults discriminating characteristics from maximum power points (MPPs) coordinates. Then, the labels
 423 of the data are assigned according to their class of membership (Table 7), and the classifier computes
 424 the linear separation laws allowing the classification of new observations. The four a priori classes for
 425 original labels are retrieved from the classification results with the PCA algorithm.

426

Table 7: Fault types and original labels

| Condition | Class | Label |
|-----------|-------|-------|
| Healthy | C_0 | 1 |

| | | |
|-------------------|-------|---|
| Partial Shading 1 | C_1 | 2 |
| Partial Shading 2 | C_2 | 3 |
| Partial Shading 3 | C_3 | 4 |

427

428 This training stage is carried out using 67% of the data as original labels. The same matrix

429 $X_{[80 \times 3]} = \log[V_{mpp} \quad I_{mpp} \quad P_{mpp}]$ built with data recorded on March 21, 2019 is used to extract the fault

430 features. Three discriminant axes, also called Fisher axes, are obtained. Fig.15 shows the projection of

431 the training dataset into the 2 – dimensional LDA subspace spanned by the first and the second

432 discriminant axes (DA1, DA2). These two axes represent 99.7%, and 0.18% of the data variance,

433 respectively. In this figure, we have also drawn the linear boundaries between the classes. The

434 magenta line is the boundary that separates the data for classes C_0 and C_3 . On the other hand, the blue

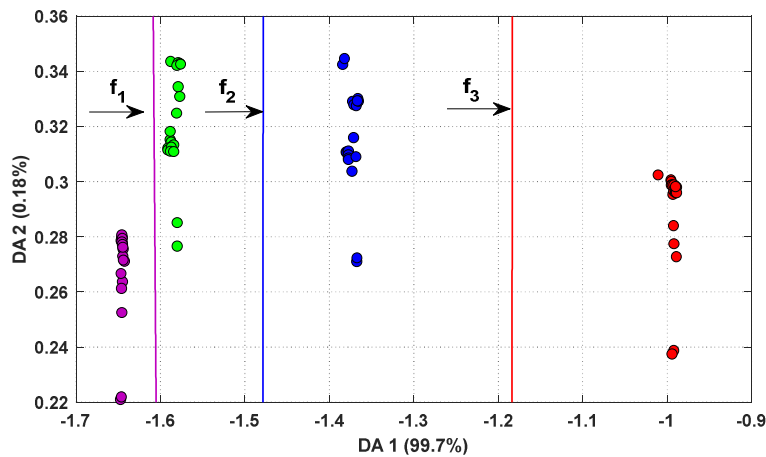
435 line separates the classes C_0 and C_1 , and the red line separates the observations of C_1 from those of C_2 .

436 Let us consider (x_1, x_2) as the coordinates of an instance l in the LDA subspace (DA1, DA2). The

437 three boundaries are described by three affine functions f_1, f_2 and f_3 where each function f_i is given by:

438
$$f_i(x_1, x_2) = L_i(1)x_1 + L_i(2)x_2 + K_i, i = 1, 2, 3 \quad (9)$$

439 The coefficients L_i and $K_i, i=1, 2, 3$, are given in Table 8.



440

441

Fig. 15. Projection of the training data in the subspace (DA1, DA2)

442

443 **Table 8:** Linear boundaries coefficients

| f_i | $L_i(1)$ | $L_i(2)$ | K_i |
|-------|--------------------|----------|--------------------|
| f_1 | $2.93 \cdot 10^3$ | 71 | $4.7 \cdot 10^3$ |
| f_2 | $-1.82 \cdot 10^4$ | 59 | $-2.15 \cdot 10^4$ |
| f_3 | $-1.01 \cdot 10^4$ | 17 | $-1.49 \cdot 10^4$ |

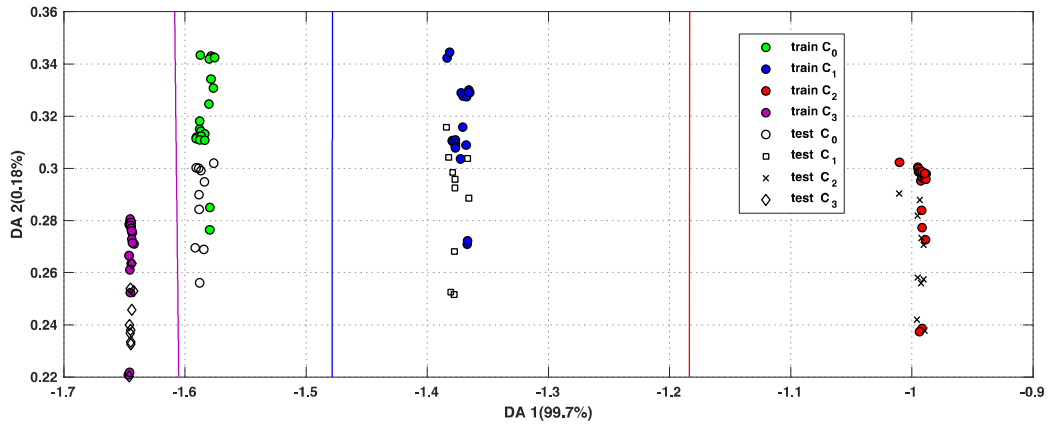
444

445 With the cross-validation method, the same success rate (100%) as for PCA is obtained. LDA also
 446 entirely separates the different classes even under climatic variations. In the following, the obtained
 447 separation laws will be used to classify new samples; it is the validation stage.

448

449 3.3.3. Results of the validation stage with LDA

450 The validation is done in two steps. In the first one, the calculated separation laws are used to
 451 classify the remaining MPP (test data) representing 33% of the database collected on March 21, 2019.
 452 The results, displayed in figure 16 show the successful classification of the test data.
 453 In the second step, the separation laws are evaluated with the healthy samples extracted from
 454 measurements on June 26, 2018, mixed with 33% of partial shading data measured on March 21,
 455 2019. The classification result is displayed in Fig.17. With the separation laws, all the test samples are
 456 correctly classified in their a priori groups; no misclassification errors were found. With the cross-
 457 validation method, discrimination's effectiveness between the measurements under variable
 458 experimental conditions is proved with a success rate of 100 % for all the classes (Table 9).

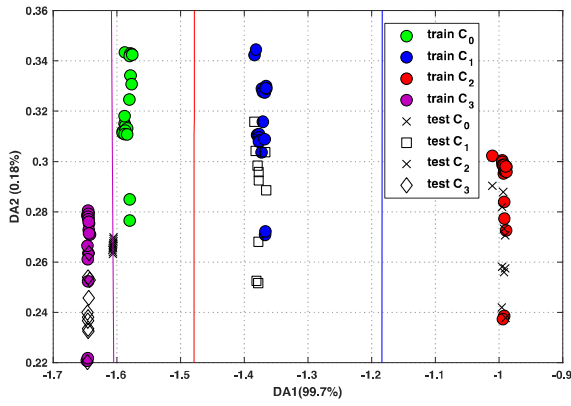


459

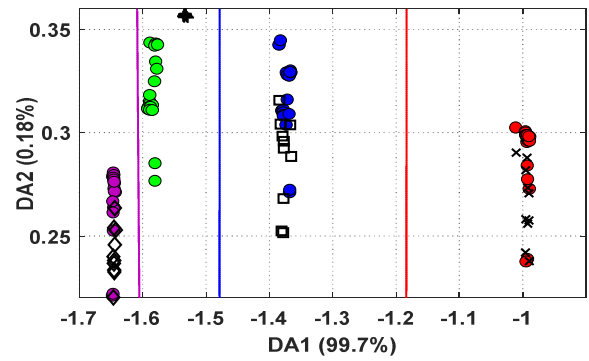
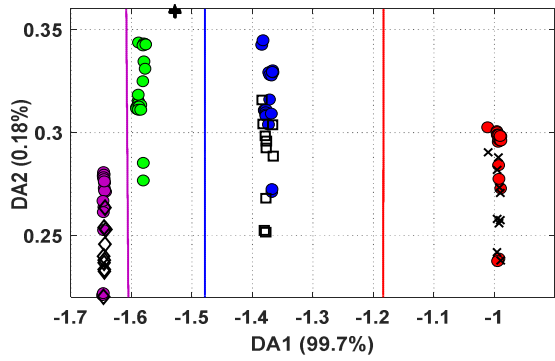
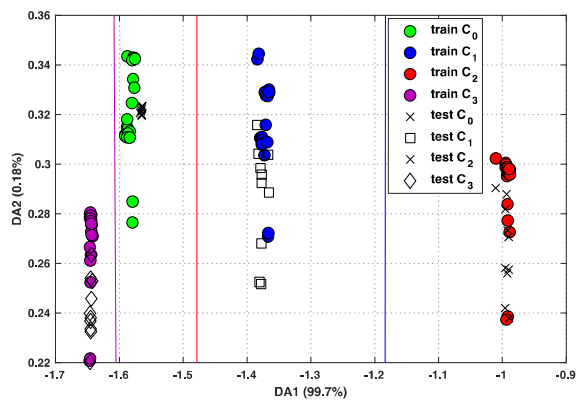
460

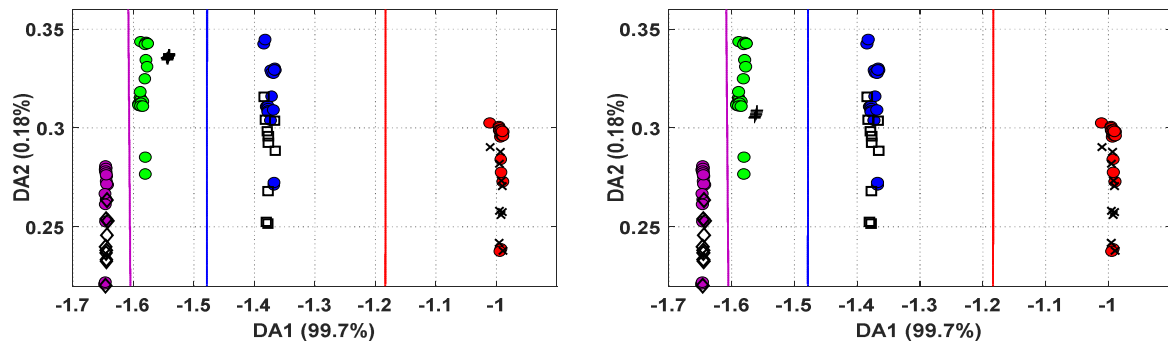
Fig. 16. Classification of test data in the discrimination space (DA1, DA2)

461



462





463

464

Fig. 17. Classification of mixed data in the discrimination space (DA1, DA2)

465

466

Table 9. Confusion matrix for test dataset with LDA

| Class a priori | Class a posteriori | | | |
|----------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | Class C ₀ (%) | Class C ₁ (%) | Class C ₂ (%) | Class C ₃ (%) |
| Class C ₀ | 100 | 0 | 0 | 0 |
| Class C ₁ | 0 | 100 | 0 | 0 |
| Class C ₂ | 0 | 0 | 100 | 0 |
| Class C ₃ | 0 | 0 | 0 | 100 |

467

468

469

470

471

472

473

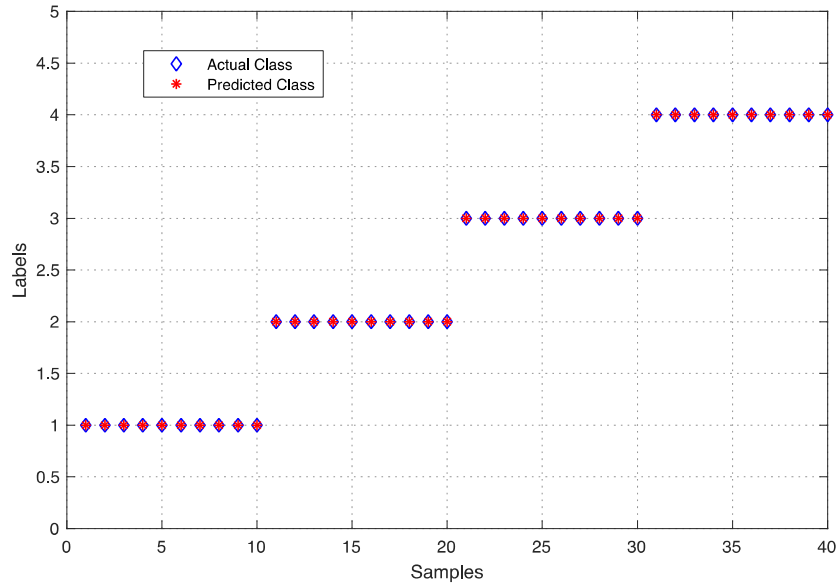
474

Even if for the studied cases the obtained identification is correct, it can be observed in Fig.17 that healthy state's data corresponding to samples 'test C₀', lie on the separation boundary between classes C₀ and C₃. However, as the probability of assignment to class C₀ is higher, these samples are predicted with label 1 and correctly identified as healthy data. Nevertheless, this ambiguity underlines the fact that in order to have good performance in classifying several types of defects with different levels of severity, it will be beneficial to have a rich database. It must contain several configurations (operating points and states of health) under different environmental conditions.

475

476

We can conclude that at the PV module level, the classification of measurements through LDA is successful as long as a large amount of labelled data is available for offline learning.



477
478 **Fig. 18.** Healthy and faulty class prediction based on LDA

479
480
481 **4. Discussion and comparison**
482

483 The classification results obtained through the analysis of the maximum power point's (MPP)
484 coordinates with PCA and LDA for partial shading diagnosis are very close. Table 10 summarizes the
485 main advantages and their limitations of the diagnosis methods proposed in this paper, the methods
486 developed in our previous work, especially for large PV plants.

487 Fault diagnosis based on the analysis of maximum power point's (MPP) coordinates appears to be
488 relevant for large PV plants. Indeed, this approach enables diagnosing a large number of PV modules
489 in real-time faster than using all the points on the I(V) curve, thus reducing power losses and costs.

490 From a field-application perspective, it would be of interest to evaluate these methods with other
491 faults.

492 MPP-based methods have up to now only been evaluated at the PV module level. It would be
493 interesting to evaluate these methods at the PV string and array levels.

494 Besides the ability to PV fault diagnosis, the proposed methods can also be used to quantify the
 495 power loss in real-time, thus allowing maintenance optimization. Moreover, they can be combined
 496 with reconfiguration algorithms to limit the shading effect. When several PV modules are connected in
 497 series, the shading fault detection and notably the isolation is decisive to balance the modules or at
 498 least mitigate the power production gaps between strings. Different reconfigurable architectures have
 499 been developed, taking into account the shading condition and have allowed significant energy savings
 500 [46], [47].

Table 10: Overview of the proposed methods for partial shading detection in PV systems

| Pre-processing | Advantages | Limitations for large scale PV plants |
|--|---|--|
| <ul style="list-style-type: none"> ▪ $\log\left(\frac{v}{\eta}\right)$ ▪ $\log(i)$ ▪ $\log\left(\frac{p}{\eta}\right)$ | <ul style="list-style-type: none"> - Adapted with online I (V) tracer - Sensitive to fault severity - Suitable for small-scale PV systems | <ul style="list-style-type: none"> - Less convenient - Time-consuming and functional disturbance - Power losses - Additional cost with local I (V) tracers - Sensitive to I (V) curve quality |
| <ul style="list-style-type: none"> ▪ $\log(I_{MPP})$ ▪ $\log(V_{MPP})$ ▪ $\log(P_{MPP})$ | <ul style="list-style-type: none"> - Earlier fault detection allowing an optimized maintenance - Cost-effective : no additional measurement of temperature or irradiance for data normalization - Saving memory storage - Sensitive to fault severity - Robust to climatic variations - Possible power losses quantification (LDA) | <ul style="list-style-type: none"> - Efficient and accurate fault identification depend on the learning data set (size and variety) (LDA) |

| Method | Variables |
|---|--|
| <ul style="list-style-type: none"> I (V) curve analysis – PCA [22] | <ul style="list-style-type: none"> Current i Voltage v Power p 303 observations for each variable |
| <ul style="list-style-type: none"> MPP analysis - PCA | <ul style="list-style-type: none"> Current I_{MPP} Voltage V_{MPP} Power P_{MPP} 30 observations for each variable |
| <ul style="list-style-type: none"> MPP analysis - LDA | 30 observations for each variable |

501

502

503 5. Conclusion

504

505 Fault detection and identification in PV systems are essential to expedite the maintenance and
 506 minimize the power losses. This study has presented a cost-effective and straightforward data-driven
 507 diagnosis approach to detect and identify partial shading using principal component analysis (PCA)
 508 and linear discriminant analysis (LDA). No prior information on the PV system and no other sensing
 509 circuit are needed. The proposed methods take advantage of available measurements in PV systems:
 510 voltage, current, and power at the maximum power point (MPP). The effectiveness of the proposed
 511 methods has been successfully proved at the PV module level. Based on collected data under real
 512 working conditions (variable irradiation and temperature), we have demonstrated that the methods can
 513 correctly detect and classify the partial shading fault with a success rate of 100%. The methods are
 514 also easy to integrate into a monitoring system for real-time diagnosis and are especially convenient
 515 for large-scale PV plants. The data required for the learning stage can benefit from the vast amount of
 516 data available in PV plants under operation for decades. Otherwise, the learning database could be

517 generated from simulation models. It would be of interest to evaluate the proposed methodology with
518 other faults (open-circuit, short-circuit, and contact degradation).

519

520 Acknowledgement

521 The authors gratefully recognize the financial support of the Ministry of Higher Education and
522 Scientific Research in Tunisia and Paris-Saclay University who provided the scholarship to the PhD
523 student.

524

525 References

- 526 [1] A. Golnas, "PV System Reliability: An Operator's Perspective," *IEEE J. Photovolt*, vol. 3, no. 1,
527 pp. 416–421, 2013, doi: 10.1109/JPHOTOV.2012.2215015.
- 528 [2] IEA, "Reliability Study of Grid Connected PV Systems," T7-08, 2002.
- 529 [3] Y. Zhao, R. Ball, J. Mosesian, J. de Palma, and B. Lehman, "Graph-Based Semi-supervised
530 Learning for Fault Detection and Classification in Solar Photovoltaic Arrays," *IEEE Trans.*
531 *Power Electron*, vol. 30, no. 5, pp. 2848–2858, 2015, doi: 10.1109/TPEL.2014.2364203.
- 532 [4] D. C. Jordan, T. J. Silverman, J. H. Wohlgmuth, S. R. Kurtz, and K. T. VanSant, "Photovoltaic
533 failure and degradation modes," *Prog. Photovolt: Res. Appl.*, pp. 318–326, 2017, doi:
534 10.1002/pip.2866.
- 535 [5] IEA, "Assessment of Photovoltaic Module Failures in the Field," 2017.
- 536 [6] X. Sun, R. V. K. Chavali, and M. A. Alam, "Real-time monitoring and diagnosis of photovoltaic
537 system degradation only using maximum power point—the Suns-Vmp method," *Prog Photovolt*
538 *Res Appl*, vol. 27, no. 1, pp. 55–66, 2019, doi: 10.1002/pip.3043.
- 539 [7] M. Bressan, Y. El Basri, A. G. Galeano, and C. Alonso, "A shadow fault detection method based
540 on the standard error analysis of I-V curves," *Renew. Energy*, vol. 99, pp. 1181–1190, 2016, doi:
541 10.1016/j.renene.2016.08.028.
- 542 [8] A. Y. Appiah, X. Zhang, K. A. Ben Beklisi, and F. Kyeremeh, "Review and performance
543 evaluation of photovoltaic array fault detection and diagnosis techniques," *Int. J. Photoenergy*,
544 vol. 2019, pp. 1–19, 2019, doi: <https://doi.org/10.1155/2019/6953530>.
- 545 [9] A. Triki-Lahiani, A. Bennani-Ben Abdelghani, and I. Slama-Belkhodja, "Fault detection and
546 monitoring systems for photovoltaic installations: A review," *Renew. Sustain. Energy Rev*, vol.
547 82, pp. 2680–2692, 2018, doi: <https://doi.org/10.1016/j.rser.2017.09.101>.
- 548 [10] D. S. Pillai, F. Blaabjerg, and N. Rajasekar, "A comparative Evaluation of Advanced Fault
549 Detection Approaches for PV Systems," *IEEE Journal of Photovoltaics*, vol. 9, no. 2, pp. 513–
550 527, 2019, doi: 10.1109/JPHOTOV.2019.2892189.
- 551 [11] S. Dali et al., "Monitoring, Diagnosis, and Power Forecasting for Photovoltaic Fields: A
552 Review," *International Journal of Photoenergy*, pp. 1–13, 2017, doi:
553 <https://doi.org/10.1155/2017/1356851>.
- 554 [12] F. Spertino et al., "A power and energy procedure in operating photovoltaic systems to quantify
555 the losses according to the causes," *Sol. Energy*, vol. 118, pp. 313–326, 2015, doi:
556 <https://doi.org/10.1016/j.solener.2015.05.033>.
- 557 [13] IEA, "Review on Infrared and Electroluminescence Imaging for PV Field Applications," T13-
558 10, 2018.

- 559 [14] M. Wolfgang, C. Hirschl, L. Neumaier, A. Kolger, and A. Plosch, "Identifikation von
560 fehlerhaften PV Modulen und Anlagen mit Thermografie," presented at the 29. Symposium
561 Photovoltaische Solarenergie 2014, Deutschland, 2014, pp. 1–11, doi: 10.13140/2.1.3636.2881.
- 562 [15] J. A. Tsanakas, L. D. Ha, and F. AL Shakarchi, "Advanced inspection of photovoltaic
563 installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery,"
564 *Renew. Energy*, vol. 102, pp. 224–233, 2017, doi: <https://doi.org/10.1016/j.renene.2016.10.046>.
- 565 [16] S. Roy, M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "An Irradiance-Independent, Robust
566 Ground-Fault Detection Scheme for PV Arrays Based on Spread Spectrum Time-Domain
567 Reflectometry (SSTDTR)," *IEEE Trans. Power Electron*, vol. 33, no. 8, pp. 7046–7057, 2018,
568 doi: 10.1109/TPEL.2017.2755592.
- 569 [17] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "PV ground-fault detection using spread
570 spectrum time domain reflectometry (SSTDTR)," presented at the IEEE Energy Conversion
571 Congress and Exposition (ECCE), USA, 2013, pp. 1015–1020, doi:
572 10.1109/ECCE.2014.6953848.
- 573 [18] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "PV arc-fault detection using spread spectrum
574 time domain reflectometry (SSTDTR)," presented at the IEEE Energy Conversion Congress and
575 Exposition (ECCE), USA, 2014, pp. 3294–3300, doi: 10.1109/ECCE.2014.6953848.
- 576 [19] T. Takashima, J. Yamaguchi, and M. Ishida, "Fault detection by signal response in PV module
577 strings," presented at the 33rd IEEE Photovoltaic Specialists Conference, USA, 2008, doi: Fault
578 detection by signal response in PV module strings.
- 579 [20] T. Takashima, J. Yamaguchi, K. Otani, T. Oozeki, K. Kato, and M. Ishida, "Experimental
580 studies of fault location in PV module strings," *Sol. Energy Mater. and Sol. Cells*, vol. 93, no. 6–
581 7, pp. 1079–1082, 2009, doi: <https://doi.org/10.1016/j.solmat.2008.11.060>.
- 582 [21] S. Fadhel, A. Migan, C. Delpha, D. Diallo, I. Bahri, M. Trabelsi, M.F. Mimouni, "Data-driven
583 Approach for Isolated PV Shading Fault Diagnosis Based on Experimental I-V curves Analysis,"
584 presented at the IEEE International Conference on Industrial Technology (ICIT), France, 2018,
585 pp. 927–932, doi: 10.1109/ICIT.2018.8352302.
- 586 [22] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, M.F. Mimouni, "PV shading
587 fault detection and classification based on I-V curve using principal component analysis:
588 Application to isolated PV system," *Sol. Energy*, vol. 179, pp. 1–10, 2019.
- 589 [23] Y. El Basri, M. Bressan, L. Seguir, H. Alawadhi, and C. Alonso, "A proposed graphical
590 electrical signatures supervision method to study PV module failures," *Sol. Energy*, vol. 116, pp.
591 247–256, 2015, doi: <https://doi.org/10.1016/j.solener.2015.02.048>.
- 592 [24] A. K. Tossa, Y. M. Soro, Y. Azoumah, and D. Yamegueu, "A new approach to estimate the
593 performance and energy productivity of photovoltaic modules in real operating conditions,"
594 *Sol. Energy*, vol. 110, pp. 543–560, 2014, doi: 10.1016/j.solener.2014.09.043.
- 595 [25] M. Merchaoui, A. Sakly, and M. F. Mimouni, "Particle swarm optimisation with adaptive
596 mutation strategy for photovoltaic solar cell/module parameter extraction," *Energy Conv. and
597 Manag.*, vol. 175, pp. 151–163, 2018, doi: <https://doi.org/10.1016/j.enconman.2018.08.081>.
- 598 [26] Y. Chaibi, M. Malvoni, A. Chouder, M. Boussetta, and M. Salhi, "Simple and efficient approach
599 to detect and diagnose electrical faults and partial shading in photovoltaic systems," *Energy
600 Conversion and Management*, vol. 196, pp. 330–343, 2019, doi:
601 <https://doi.org/10.1016/j.enconman.2019.05.086>.
- 602 [27] D. S. Pillai and N. Rajasekar, "Metaheuristic algorithms for PV parameter identification: A
603 comprehensive review with an application to threshold setting for fault detection in PV systems,"
604 *Renew. Sustain. Energy Rev.*, vol. 82, pp. 3503–3525, 2018, doi:
605 <https://doi.org/10.1016/j.rser.2017.10.107> Get.
- 606 [28] S. Kichou, S. Silvestre, L. Guglielminotti, L. Mora-Lopez, and E. Munoz-Ceron, "Comparison
607 of two PV array models for the simulation of PV systems using five different algorithms for the

- parameters identification,” *Renew. Energy*, vol. 99, pp. 270–279, 2016, doi: <https://doi.org/10.1016/j.renene.2016.07.002>.
- [29] Z. Chen, L. Wu, S. Cheng, P. Lin, W. Yue, and W. Lin, “Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and I-V characteristics,” *Applied Energy*, vol. 204, pp. 912–931, 2017, doi: <https://doi.org/10.1016/j.apenergy.2017.05.034>.
- [30] S. Spataru, D. Sera, T. Kerekes, and R. Teodorescu, “Diagnostic method for photovoltaic systems based on light I–V measurements,” *Sol. Energy*, vol. 119, pp. 29–44, 2015, doi: [10.1016/j.solener.2015.06.020](https://doi.org/10.1016/j.solener.2015.06.020).
- [31] J. D. Bastidas-Rodríguez, E. Franco, G. Petrone, C. A. Ramos-Paja, and G. Spagnuolo, “Model Based Degradation Analysis Of Photovoltaic Modules Through Series Resistance Estimation,” *IEEE Trans. Indus. Electr.*, vol. 62, no. 11, pp. 7256–7265, 2015, doi: [10.1109/TIE.2015.2459380](https://doi.org/10.1109/TIE.2015.2459380).
- [32] D. Wang, “A method for instantaneous measurement of PV I-V characteristics and its application for MPPT control,” presented at the 35th IEEE Photovoltaic Specialists Conference, USA, 2010, pp. 2904–2907, doi: [10.1109/PVSC.2010.5614505](https://doi.org/10.1109/PVSC.2010.5614505).
- [33] A. Khoshnami and I. Sadeghkhani, “Sample entropy-based fault detection for photovoltaic arrays,” *IET Renew. Power Gener.*, vol. 12, no. 16, pp. 1966–1976, 2018, doi: [10.1049/iet-rpg.2018.5220](https://doi.org/10.1049/iet-rpg.2018.5220).
- [34] M. Dhimish, V. Holmes, B. Mehrdadi, and M. Dales, “Multi-layer photovoltaic fault detection algorithm,” *IET High Voltage*, vol. 4, pp. 244–252, 2017, doi: [10.1049/hve.2017.0044](https://doi.org/10.1049/hve.2017.0044).
- [35] F. Harrou, A. Dairi, B. Taghezouit, and Y. Sun, “An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine,” *Sol. Energy*, vol. 179, pp. 48–58, 2019, doi: <https://doi.org/10.1016/j.solener.2018.12.045>.
- [36] E. Garoudja, F. Harrou, Y. Sun, K. Kara, A. Chouder, and S. Silvestre, “Statistical fault detection in photovoltaic systems,” *Sol. Energy*, vol. 150, pp. 485–499, 2017, doi: <https://doi.org/10.1016/j.solener.2017.04.043>.
- [37] L. Garcia-Gutierrez, M. Bressan, F. Jimenez, S. de-las Heras, and C. Alonso, “Design of a Global Maximum Power Point Tracking (GMPPT) for PV array based on precise PV shadow model,” presented at the 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, Oct. 2018, pp. 275–280, doi: [10.1109/ICRERA.2018.8566880](https://doi.org/10.1109/ICRERA.2018.8566880).
- [38] M. Aquib, S. Jain, and V. Agarwal, “A Time-Based Global Maximum Power Point Tracking Technique for PV System,” *IEEE Trans. Power Electr.*, vol. 35, no. 1, pp. 393–402, 2020, doi: [10.1109/TPEL.2019.2915774](https://doi.org/10.1109/TPEL.2019.2915774).
- [39] K. Kobayashi, I. Takano, and Y. Sawada, “A study of a two stage maximum power point tracking control of a photovoltaic system under partially shaded insolation conditions,” *Sol. Energy Mater. and Sol. Cells*, vol. 90, pp. 2975–2988, 2006, doi: <https://doi.org/10.1016/j.solmat.2006.06.050>.
- [40] H. Zhu, L. Lu, J. Yao, S. Dai, and Y. Hu, “Fault diagnosis approach for photovoltaic arrays based on unsupervised sample clustering and probabilistic neural network model,” *Sol. Energy*, vol. 176, pp. 395–405, 2018, doi: <https://doi.org/10.1016/j.solener.2018.10.054>.
- [41] M. Benbouzid, C. Delpha, Z. Khatir, S. Lefebvre, and D. Diallo, “Faults Detection and Diagnosis in a Static Converter,” in *Electrical Machine Diagnosis*, ISTE., 2011, pp. 272–319.
- [42] D. Y. Jung, S. M. Lee, H. Wang, J. H. Kim, and S. H. Lee, “Fault detection method with PCA and LDA and its application to induction motor,” *J. Cent. South Univ. Technol.*, vol. 17, pp. 1238–1242, 2010, doi: <https://doi.org/10.1007/s11771-010-0625-y>.
- [43] R. J. Martis, U. R. Acharya, and L. C. Min, “ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform,” *Biomedical Signal Processing and Control*, vol. 8, pp. 437–448, 2013, doi: <https://doi.org/10.1016/j.bspc.2013.01.005>.

- 658 [44] A. Quinquis, *Digital Signal Processing using MATLAB*, ISTE. André Quinquis, 2008.
659 [45] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York, 1979.
660 [46] M. A. Chaaban *et al.*, “Adaptive photovoltaic system,” presented at the 36th Annual Conference
661 on IEEE Industrial Electronics Society (IECON), USA, 2010, pp. 3192–3197, doi:
662 10.1109/IECON.2010.5675047.
663 [47] M. Alahmad, M. A. Chaaban, S. kit Lau, J. Shi, and J. Neal, “An adaptive utility interactive
664 photovoltaic system based on a flexible switch matrix to optimize performance in real-time,”
665 *Sol.Energy*, vol. 86, pp. 951–963, 2012, doi: <https://doi.org/10.1016/j.solener.2011.12.028>.
666