

## Article

# Hypothesis Tests-Based Analysis for Anomaly Detection in Photovoltaic Systems in the Absence of Environmental Parameters

Silvano Vergura 

Department of Electrical and Information Engineering, Polytechnic University of Bari, st. E. Orabona 4, I-70125 Bari, Italy; silvano.vergura@poliba.it; Tel.: +39-080-5963590; Fax: +39-080-5963410

Received: 28 January 2018; Accepted: 22 February 2018; Published: 25 February 2018

**Abstract:** This paper deals with the monitoring of the performance of a photovoltaic plant, without using the environmental parameters such as the solar radiation and the temperature. The main idea is to statistically compare the energy performances of the arrays constituting the PV plant. In fact, the environmental conditions affect equally all the arrays of a small-medium-size PV plant, because the extension of the plant is limited, so any comparison between the energy distributions of identical arrays is independent of the solar radiation and the cell temperature, making the proposed methodology very effective for PV plants not equipped with a weather station, as it often happens for the PV plants located in urban contexts and having a nominal peak power in the 3–50 kWp range, typically installed on the roof of a residential or industrial building. In this case, the costs of an advanced monitoring system based on the environmental data are not justified, consequently, the weather station is often also omitted. The proposed procedure guides the user through several inferential statistical tools that allow verifying whether the arrays have produced the same amount of energy or, alternatively, which is the worst array. The procedure is effective in detecting and locating abnormal operating conditions, before they become failures.

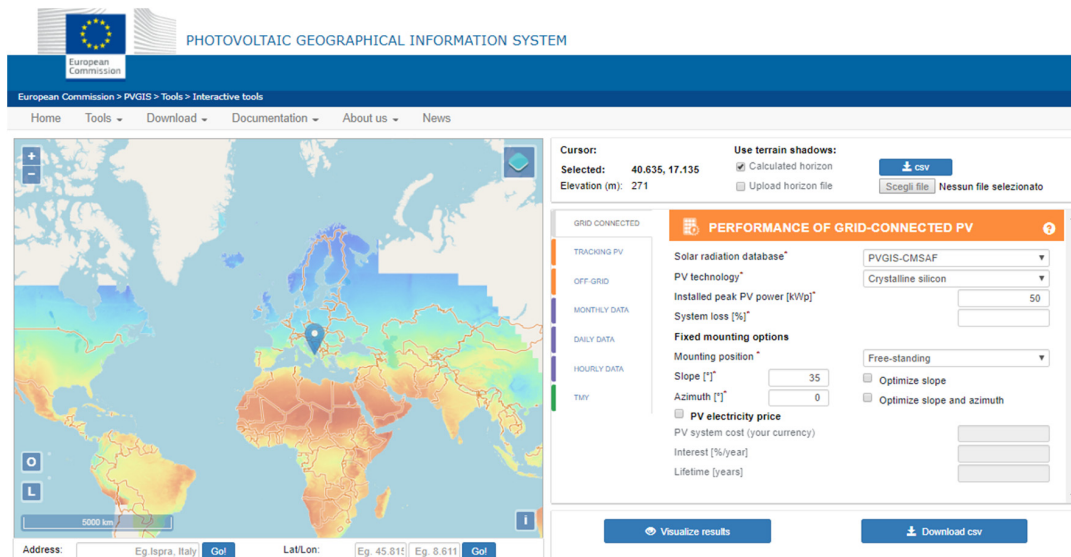
**Keywords:** ANOVA; Bartlett's test; Hartigan's dip test; Jarque-Bera's test; Kruskal-Wallis' test; Mood's Median test; residential buildings; Tukey's test; urban context

## 1. Introduction

The random variability of atmospheric phenomena affects the available irradiance intensity for photovoltaic (PV) generators. During the clear days an analytic expression for the solar irradiance can be defined, whereas this is not possible for cloudy days. The effects of the environmental conditions are studied in [1–5]. After the installation of a PV plant, a system for monitoring the energy performance in every environmental condition is needed. As the modules are main components of a PV plant, deep attention is focused on their state of health [6]. For this reason, techniques commonly used to verify the presence of typical defects in PV modules are based on infrared analysis [7,8], eventually supported by unmanned aerial vehicles [9], on luminescence imaging [10], or on their combination [11], while automatic procedures to extract information using thermograms are proposed in [12,13]. Nevertheless, these approaches regard single modules of the PV plants. When there is no information about the general operation of the PV plant, other techniques can be considered to prevent failures and to enhance the energy performance of the PV system, such as artificial neural networks [3,14], statistics [15–17], and checking the electrical variables [18–20]. More in detail, some of the PV fault detection algorithms are based on electrical circuit simulation of the PV generator [21,22], while other researchers use approaches based on the electrical signals [23,24]. Moreover, predictive model approaches for PV system power production based on the comparison between measured and modeled PV system outputs are discussed in [15,25–27]. Standard benchmarks [28], called *final PV system yield, reference*

*yield* and *Performance Ratio (PR)*, are currently used to assess the overall system performance in terms of energy production, solar resource, and system losses. These benchmarks have been recently used to review the energy performance of 993 residential PV systems in Belgium [29] and 6868 PV installations in France [30]. Unfortunately, these indices have two drawbacks: they only supply rough information about the performance of the overall PV plant and they do not allow any assessment of the behavior of single PV plant parts. Moreover, when important faults such as short circuits or islanding occur, the electrical variables and the produced energy have fast and not negligible variations, so they are easily detected. These events produce drastic changes and can be classified as high-intensity anomalies. On the other hand, low-intensity anomalies such as the ageing of the components or minimal partial shading produce minimal variations on the electrical variables and on the produced energy, so they are not easily detectable. Moreover, these minor anomalies can evolve into failures or faults, so their timely identification can avoid more serious failures and limit the occurrence of out of order states. With respect to the configuration defined in the design stage, any PV plant can be single-array or multi-array, being an array a set of connected PV modules, for which the electrical variables and the produced energy are measured. Moreover, PV plants with only two arrays are not common: the alternatives are between one-array PV plant—this is the case of small nominal peak power PV plant—and multiple-array PV plant for higher nominal peak power PV plant. The multiple-array solution is very common, because it has several advantages, thanks to the partition of the produced energy: lower current for each array (thus reduced section of the solar cables), high flexibility in the choice of the components (inverter, switches, electrical boxes, etc.), O&M services on each single array, avoiding situations where the whole plant is out of order, and so on. Moreover, the large PV plants, having a nominal peak power higher than 100 kWp are usually equipped with a weather station, able to measure and store the environmental parameters, which affect the energy production, i.e., the solar irradiance, temperature and wind. Frequently, the large PV plants with nominal peak power higher than 1 MWp are equipped with more than one weather station, because of the large occupied area, typically about 2 ha/MWp. Obviously, these last ones are solar farms and are located in extra-urban territory. Instead, the PV plants usually installed in a urban context have a nominal peak power that ranges between 3–50 kWp; the minimum value refers, for example, to a PV plant on the roof of a residential building, while the maximum value corresponds to a PV plant of a small company that locates it on the roof of its industrial building or in a free private area. These PV plants are usually multi-array and are not equipped with a weather station, because the costs of an advanced monitoring system are not negligible with respect to the initial investment as well as to the costs of a yearly O&M service, so these medium size PV plants are usually equipped with a simplified monitoring system, which stores the total produced energy, the electrical variables on the AC and DC side, having also the possibility to send alerts to the owner via SMS or email. This system does not perform any analysis of the produced energy, so it cannot detect any anomaly before it becomes a failure, and it can only alert when the failure is already happened. In these cases, valid support is provided by the PhotoVoltaic Geographical Information System (PVGIS) [31] of the European Commission Joint Research Centre (EC-JRC) that is based on the historical data of the solar irradiance. Figure 1 is a screenshot of the website. On the left hand-side, a colored map with the solar radiation is reported and the user can select the location of the PV plant, whereas, on the right hand-side, the user can insert the information on the typology of the PV plant (off-grid, grid-connected, tracking-based), the specifications of the PV plant (module technology, slope, rated power, etc.), and the required energy production data (monthly, daily, hourly). In this way, it is possible to estimate the productivity of the PV plant under investigation and to compare it with the real energy production. This can represent a preliminary check of the operation of the PV plant and will be used later, in the Sections 3 and 4. Nevertheless, it is extremely important to prevent a failure, detecting any anomaly in a timely way for two reasons. Firstly, when an anomaly is present, the energy production is already lower than the expected one and this implies an economic loss. Secondly, a timely action of the O&M service allows restoring the damaged parts of the PV systems with minimum costs and minimum time out of order, reducing either the Mean Time To

Repair (MTTR)—because the damage is limited—or the Mean Down Time (MDT), because the restore action is planned while the PV plant is still operating. This strategy, evidently, greatly increases the availability of the PV plant and its yearly energy performance.



**Figure 1.** PVGIS of the European Commission Joint Research Centre (EC-JRC).

With this in mind, this paper proposes a methodology to detect an anomaly in the operation of a PV system; this methodology can be easily applied to any multi-array PV plant, but it is particularly useful for PV plants not equipped with a weather station. This is often the situation of the PV plants in urban contexts, as previously explained. The proposed methodology, in fact, compares the energy distributions of the arrays with each other, on the basis of a statistical algorithm that does not consider the environmental parameters as inputs. This is possible because the area occupied by the PV modules of a PV plant in the urban context is limited and then the average environmental conditions can be considered to affect the identical arrays in the same way. The proposed procedure is completely based on several hypothesis tests and is a cheap and fast approach to monitor the energy performance of a PV system, because no additional hardware is required. The procedure also allows continuous monitoring because it is cumulative and new data can be added to the initial dataset, as they are acquired by the measurement system. The methodology is based on an algorithm, which suggests the user, step by step, the suitable statistical tool to use. The first one is the Hartigan's Dip Test (HDT) that is able to discriminate an unimodal distribution from a multimodal one. The verification of the unimodality can be also done on the basis of a relationship between the values of skewness and kurtosis [32,33]; nevertheless, in this paper only HDT will be used, because it is usually more sensitive than other methods. The check on the unimodality is very important to decide whether a parametric test can be used to compare the energy distributions of the arrays or not, because the parametric tests, being based on known distributions, are more performing than the nonparametric ones. Nevertheless, the parametric tests can be applied, only if specific assumptions are satisfied. A powerful parametric test to compare more than two statistical distributions is the well-known ANalysis Of VAriance (ANOVA) [34] that is based on three assumptions. The proposed algorithm suggests using the Jarque-Bera's test and the Bartlett's test to verify the assumptions. In the negative case, the procedure suggests to use the Kruskal-Wallis test or the Mood's median test, in absence or in presence of outliers in the dataset, respectively. As a last step, Tukey's test is run to do a multi-comparison one-by-one between the mean values of the distributions, in order to determine which estimates are significantly different.

A case-study is discussed in the paper. The algorithm is applied to a real operating PV plant and the methodology is run four times: the first one, based on the energy dataset of one month; the second

one, based on the energy dataset of three months; the third one, based on the energy dataset of six months; the last one, based on the energy dataset of the whole year. The paper is structured as follows: Section 2 describes the proposed algorithm, Section 3 describes the PV system under examination, Section 4 discusses the results, and the Conclusions end the paper.

## 2. Statistical Methodology

In this paper, it is considered that the PV plant is composed of  $A$  identical arrays, with  $A > 2$  for the reasons already explained in the Introduction. This constraint is mandatory for the proposed methodology, because it is based on the comparison among the energy distributions of the arrays constituting the PV system. Each array is usually equipped with a measurement system that measures the values of the produced energy in AC, other than the values of voltage and current of both the DC and AC sides of the inverter, with a fixed sampling time,  $\Delta t$ . At the generic time-instant  $t = q\Delta t$  of the  $j$ -th day, the  $q$ -th sample vector of the  $k$ -th array is defined as  $x_{j,k}(q) = [E_{j,k}(q) \ v_{j,k,DC}(q) \ i_{j,k,DC}(q) \ v_{j,k,AC}(q) \ i_{j,k,AC}(q)]$ , for  $k = 1, \dots, A$ ,  $j = 1, \dots, D$  (being  $D$  the number of the investigated days),  $q = 1, \dots, Q$ , where  $q = 1$  characterizes the first daily sample at the analysis time  $t = \Delta t$  and  $q = Q$  defines the last daily sample, acquired at the time  $t = Q \cdot \Delta t$ . For our aims, let us consider only the dataset constituted by the energy values  $E_{j,k}(q)$ . Thus, the proposed methodology can be applied to any PV plant, having a measurement system that measures at least the produced energy, no matter which are the other measured variables. The  $k$ -th array, at the end of the  $j$ -th day, has produced the energy  $E_{j,k} = \sum_{q=1}^Q E_{j,k}(q)$ , therefore the complete dataset of the energy produced by the PV plant in a fixed investigated period can be represented in a matrix form:

$$E = \begin{pmatrix} E_{1,1} & \cdots & E_{1,A} \\ \vdots & \ddots & \vdots \\ E_{D,1} & \cdots & E_{D,A} \end{pmatrix} \quad (1)$$

The columns of the matrix (1) are independent each other, because the values of each array are acquired by devoted acquisition units, so no inter-dependence exists among the values of the columns, which can be considered as separate statistical distributions. The flow chart in Figure 2 proposes the methodology to detect and locate any anomaly, before it becomes a fault.

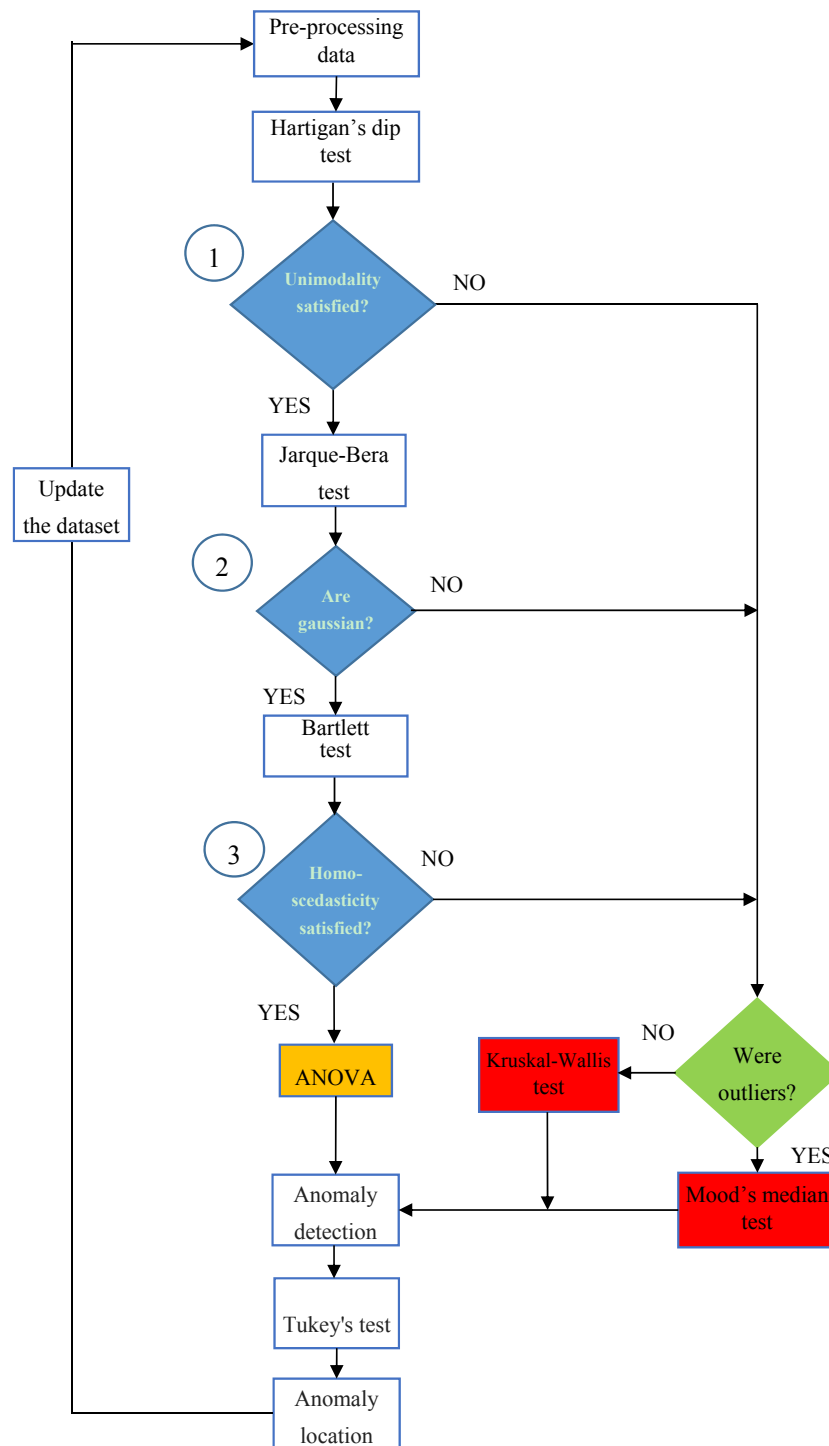
It is based on the mutual comparison among the energy distributions of the arrays; therefore the environmental data are not necessary. Obviously, this approach is valid only if the arrays are identical (same PV modules, same number of modules for each array, same slope, same tilt, same inverter, and so on); in fact, under this assumption, the energy produced by any array must be almost equal to the energy produced by any other array of the same PV plant, in each period as well as in the whole year (the changing environmental conditions affect the arrays in the same way, if they are installed next to each other without any specific obstacle).

Thus, the comparative and cumulative monitoring of the energy performance of identical arrays allows one to determine, within the uncertainty defined by the value of the significance level  $\alpha$ , if the arrays are producing the same energy or not. The first step of Figure 2 is the pre-processing of the energy dataset collected as previously explained, in order to check if outliers are present; the information about the presence or not of the outliers will be also useful later (green block). By default, an outlier is a value that is more than three scaled Median Absolute Deviations (MAD) away from the median. For a random dataset  $X = [X_1, X_2, \dots, X_D]$ , the scaled MAD is defined as:

$$MAD = F \times \text{median}(|X_j - \text{median}(X)|) \quad \text{for } j = 1, \dots, D \quad (2)$$

where  $F$  is the scaling factor and is approximately 1.4826 for a normal distribution.

After the data pre-processing, it is necessary to verify if the arrays have produced the same amount of energy. This goal can be pursued by using parametric tests or non-parametric tests. As the parametric tests are based on a known distribution of the dataset, they are more reliable than the non-parametric ones, which are, instead, distribution-free. For this reason, it is advisable to use always the parametric tests, provided that all the needed assumptions are satisfied.



**Figure 2.** Statistical methodology.

In particular, the parametric test known as ANOVA calculates the ratio between the variance *among* the arrays' distributions (divided by the freedom degree) and the variance *within* each array distribution (divided by the freedom degree). In other words, ANOVA evaluates whether the differences of the mean values of the different groups are statistically significant or not. For this aim, ANOVA calculates the following Fisher's statistic,  $F$ , [35]:

$$F = \frac{\frac{D}{A-1} \cdot \sum_{k=1}^A (\bar{x}_k - \bar{X})^2}{\frac{1}{A(D-1)} \cdot \sum_{k=1}^A \sum_{j=1}^D (x_{kj} - \bar{x}_k)^2} \quad (3)$$

where  $\bar{x}_k$  is the mean value of the  $k$ -th distribution,  $\bar{X}$  the global mean,  $x_{kj}$  the  $j$ -th occurrence of the  $k$ -th distribution. The cumulative distribution function  $F$  allows to determine a  $p$ -value, which has to be compared with the significance level  $\alpha$ , as later explained.

ANOVA is based on the null hypothesis  $H_0$  (Equation (4)) that the means of the distributions,  $\mu_k$ , are equal:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_A \quad (4)$$

versus the alternative hypothesis that the mean value of at least one distribution is different from the others. The output of the ANOVA test, as any other hypothesis test, is the  $p$ -value, which has to be compared with the pre-fixed significance value  $\alpha$ . Usually,  $\alpha = 0.05$ , so, if  $p\text{-value} < \alpha$  then the null hypothesis is rejected, considering acceptable to have a 5% probability of incorrectly rejecting the null hypothesis (this is known as type I error).

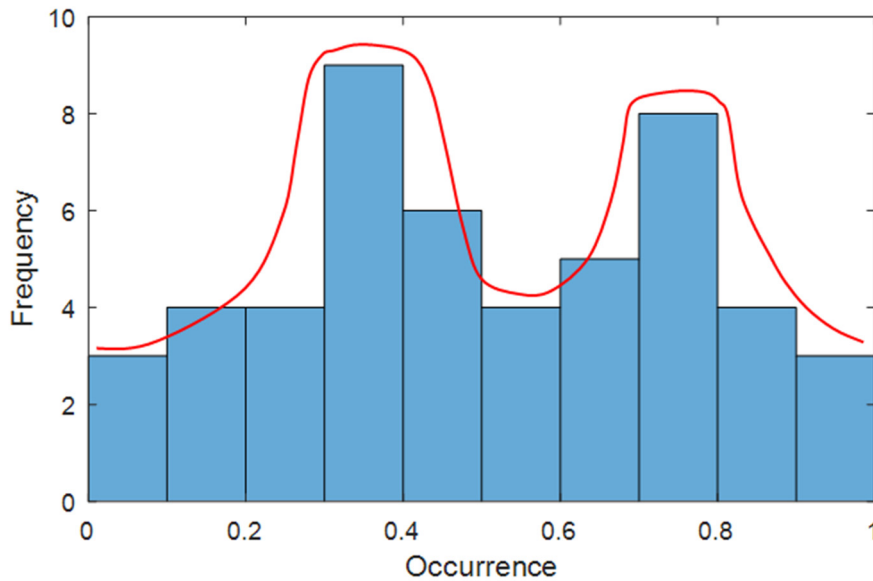
Smaller values of  $\alpha$  are not advisable to study the data of a medium-large PV plant, because the complexity of the whole system requires a larger uncertainty to be accepted. Nevertheless, ANOVA can be used only under the following assumptions:

- (a) all the observations are mutually independent;
- (b) all the distributions are normally distributed;
- (c) all the distributions have equal variance.

Finally, ANOVA can be applied also for limited violations of the assumptions (b) and (c), whereas the assumption (a) is always verified, if the measures come from independent local measurement units. So, before applying ANOVA test, several verifications are needed and they are represented by the three blue blocks of Figure 2. The first check (blue block 1) regards the unimodality of the dataset of each array, because a multimodality distribution, e.g., the bimodal distribution in Figure 3, is surely not Gaussian and violates the condition (b). Moreover, the daily-based energy distribution of an array of a well-working PV system is unimodal, because the daily solar radiation has the typical Gaussian waveform, which is unimodal; therefore, the multimodality of a daily-based energy distribution is a clear alert of a high-intensity anomaly. The Hartigan's Dip Test (HDT) is able to check the unimodality [36] and is based on the null hypothesis that the distribution is unimodal versus the alternative one that it is at least bi-modal. The HDT is a non-parametric test, so it is distribution-free. HDT return a  $p\text{-value}_{\text{HDT}}$ . By fixing the significance value  $\alpha = 0.05$ , if  $p\text{-value}_{\text{HDT}} < \alpha$  is satisfied, the null hypothesis of the unimodality is rejected, the distribution is surely not Gaussian, ANOVA cannot be applied and a nonparametric test has to be used.

In the general case of  $A$  arrays, with  $A > 2$ , the nonparametric test has to be chosen between Kruskal-Wallis test (K-W) [37,38] and Mood's Median test (MM), under the constraint of the green block; both K-W and MM do not require that the distributions are Gaussian, but only that the distributions are continuous.





**Figure 3.** Example of histogram of a bimodal distribution.

In the presence of outliers (detected, if present, in the first block), MM performs better than K-W, otherwise K-W is a good choice. Both K-W and MM are based on the null hypothesis that the median values of all the distributions are equal versus the alternative one that at least one distribution has a median value different from the others. As K-W as MM returns a  $p\text{-value}_{\text{K-W(MM)}}$  that has to be compared with the significance value  $\alpha = 0.05$ . If  $p\text{-value}_{\text{K-W(MM)}} < \alpha$  is satisfied, the null hypothesis is rejected and the arrays have not produced the same energy; otherwise, they have. Instead, if the unimodality is satisfied, other checks are needed, before deciding whether ANOVA can be applied. In fact, it is needed to verify the previous assumptions (b) and (c). Only if both of them are satisfied (blue blocks 2 and 3, respectively), ANOVA can be applied.

To check the condition (b), an effective statistical tool is the Jarque-Bera's Test (JBT). The JBT is distribution-free and based on independent random variable. It is a hypothesis test, whose null hypothesis is that the distribution is gaussian. Then, it calculates a statistical parameter, called JB, and returns a  $p\text{-value}_{\text{JBT}}$ . By fixing the significance value  $\alpha = 0.05$ , if  $p\text{-value}_{\text{JBT}} < \alpha$  is satisfied, the null hypothesis is rejected and the distribution is not gaussian, otherwise it is. It results:

$$\text{JB} = \frac{D}{6} \cdot \left[ \sigma_k^2 + \frac{k_u^2}{4} \right] \quad (5)$$

Being  $D$  the sample size,  $\sigma_k$  the skewness and  $k_u$  the Pearson's kurtosis less 3 (also known as excess kurtosis). The skewness is defined as:

$$\sigma_k = \frac{1}{D} \frac{\sum_{j=1}^D (x_j - \bar{x})^3}{(\hat{\sigma}^2)^{\frac{3}{2}}} \quad (6)$$

Being  $D$  the sample size,  $\bar{x} = \frac{1}{D} \sum_{j=1}^D x_j$  the mean value, and  $\hat{\sigma}^2 = \frac{1}{D} \sum_{j=1}^D (x_j - \bar{x})^2$  the variance. The skewness is the third standardized moment and measures the asymmetry of the data around the mean value. Only for  $\sigma_k = 0$  the distribution is symmetric; this is a necessary but not sufficient condition for a gaussian distribution. In fact, while the Gaussian distribution is surely symmetric, nevertheless there exist also symmetric but not gaussian distributions.

The excess kurtosis, instead, is defined as:

$$k_u = \frac{1}{D} \frac{\sum_{j=1}^D (x_j - \bar{x})^4}{(\hat{\sigma}^2)^2} - 3 \quad (7)$$

with the previous meaning of the parameters. The kurtosis is the fourth standardized moment and measures the tailedness of the distribution. Only for  $k_u = 0$ , the distribution is mesokurtic, which is the necessary but not sufficient condition for a Gaussian distribution. If the check of the blue block 2 is not passed, a non-parametric test (K-W or MM) has to be used, in accordance with the green block. Instead, if this verification is passed, it needs to test the assumption (c) of ANOVA, i.e., the homoscedasticity (blue block 3). This assumption can be verified by means of the Bartlett's Test (BT), which is again a hypothesis test that returns a  $p\text{-value}_{BT}$ . The BT is effective for Gaussian distributions; in fact, in the flow-chart of Figure 2 it is used only if the distributions are Gaussian. Also in this case, it is possible to fix the common significance value  $\alpha = 0.05$  and to compare it with the  $p\text{-value}_{BT}$ . If the inequality  $p\text{-value}_{BT} < \alpha$  is satisfied, the null hypothesis is rejected and the variances of the distributions of the arrays are different, then the condition (c) is violated, and ANOVA cannot be applied. In this case, it is necessary to use K-W or MM, in accordance with the green block. Otherwise, ANOVA can be applied and it return another  $p\text{-value}_{AN}$  that must be compared with the significance level  $\alpha = 0.05$ . If the inequality  $p\text{-value}_{AN} < \alpha = 0.05$  is satisfied, then the null hypothesis ( $H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_A$ ) is rejected and the conclusion is that the identical arrays have not produced the same amount of energy; so, a low-intensity anomaly is present and it is located in the array that has the mean value different from the other ones. To detect it, a multi-comparison analysis—one-to-one—between the distributions is done by means of the Tukey's Test (TT), which is a modified version of the well-known  $t$ -test and returns a  $p\text{-value}_{TT}$ , which states whether the means between two distributions are equal or not. For a small sample size (about 20 samples) the TT is reliable only for normal distribution, instead, for a larger sample size it is valid also for not normal distributions, because of the central limit theorem. Otherwise, no criticality is present and the dataset can be updated with new data to continue the monitoring of the PV plant. As the energy dataset increases, the monitoring becomes more accurate.

### 3. Description of the PV Plant under Investigation

The system under examination is a real operating 49.5 kWp grid-connected PV plant, installed in the south of Italy on the roof of the industrial building of a company. The PV plant has been designed and installed under the scheme of the feed-in tariff, financed by the Government. The 330 modules of the PV plant are equally divided in five arrays, each of them constituted by 66 PV modules. The nominal peak power of a single module is 150 W, and then the nominal peak power of a single array is 9.9 kWp. Each array is connected to the grid via a 10 kW inverter. The system faces the south and the slop is about 30°. By inserting these values in the previously mentioned PVGIS [31] of the EC-JRC, it results that the estimated yearly energy production is about 64,724 kWh, corresponding to about 1307 kWh/kWp per year. Moreover, the website provides also the estimated monthly energy production, which will be used in the next Section 4.1, Section 4.2, and Section 4.3. The PV plant is equipped with a datalogger that stores the data from the five arrays. The datalogger has a sample time of 2 s. After 10 min, the measured samples are equal to 30 (samples/min)  $\times$  10 (min) = 300 samples; an internal software calculates the average value of these 300 measured samples, whereas the energy produced in this time-slot of 10 min is calculated as  $P_{average} \cdot \frac{10}{60} [kWh]$ . This value is stored into the datalogger. So, the sampling time of the energy is 10 min, therefore there are 6 samples/hour, hence 144 samples/day, that are summed in the proposed procedure. Thus, the unique value of a day is not an average value, but a cumulative data that takes into account the variability of the environmental conditions happened during the whole day.

The measured variables are the power in AC, the energy in AC, and the voltage  $V_{dc}$  of each inverter; moreover, the number of the operating hours is stored. The default monitoring system of



the PV plant uses the power in AC and the voltage  $V_{dc}$ , the daily and total energy produced by each inverter, and the number of the operating hours. It is worth noting that the monitoring system is an internal software of the datalogger. As the operation of the monitoring system occupies the internal memory, for default the internal monitoring system does not utilize all the data available into the datalogger, in order not to occupy the internal memory quickly. This approach allows to monitor the PV plant for a longer time, but only the high-intensity anomalies can be detected. Instead, to detect even the low-intensity anomalies, it is necessary to use the methodology described in Figure 2 and all the data stored into the datalogger. Moreover, even if the measurement system of this PV plant does not measure all the variables mentioned in the Section 2 (the produced energy, other than the voltage and current in both the DC and AC side), nevertheless it acquires the produced energy that is the unique variable necessary for the proposed methodology; so it can be applied. The observation period refers to a full year during which the plant has shown some malfunctions, whereas in the previous years the PV plant has not shown any malfunctions, therefore the results of the previous years are not reported in the paper.

#### 4. Cumulative Statistical Analysis

The energy performance of the PV plant described in Section 3 has been studied by means of the statistical methodology proposed in Section 2. Statistical data analysis has been carried out in Matlab R2017 environment by using the standard routines of the Statistics toolbox and by implementing the flow chart of Figure 2. In particular, a Matlab routine that implements just the procedure of Figure 2 has been written and run for each analysis discussed later. As some tests (ANOVA, K-W, JBT) are implemented in the *Statistics Toolbox* of Matlab, these native-routines are recalled from the main routine, when necessary. As already explained, each array is equipped by a devoted measurement system, then the five distributions are mutually independent.

Four analyses are discussed, based on the dataset of the energy produced by each array:

- one-month analysis (January);
- three-months analysis (January–March);
- six-months analysis (January–June);
- one-year analysis (January–December).

The increase of the time window, updating the dataset as described in Figure 2, allows understanding how some characteristic benchmarks of the PV plants vary during the year, as new data are acquired. The following results will be reported for each analysis: the  $p$ -value<sub>HDT</sub> of each distribution to test the unimodality; the  $p$ -value<sub>JBT</sub> of each distribution to test whether each one of them is gaussian; the  $p$ -value<sub>BT</sub> to test the homoscedasticity among the distributions; the  $p$ -value<sub>AN</sub> to test whether all the distributions have the same mean value; the  $p$ -value<sub>K-W(MM)</sub> of the non-parametric test (when ANOVA cannot be applied) to check whether all the distributions have the same median value; the box plot of the ANOVA test or of the non-parametric test; the mean value of each distribution and its spread with respect to the global mean of the PV plant.

##### 4.1. One-Month Analysis (January)

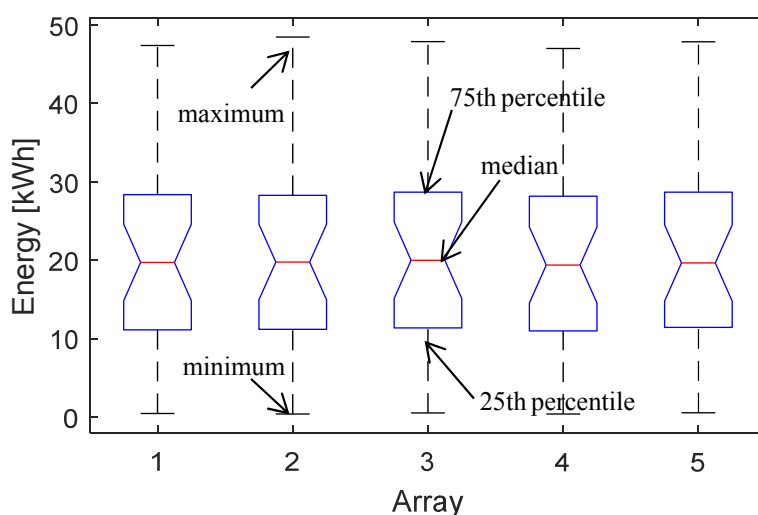
Table 1 reports the main numerical values of the parameters calculated by applying the procedure in Figure 2.

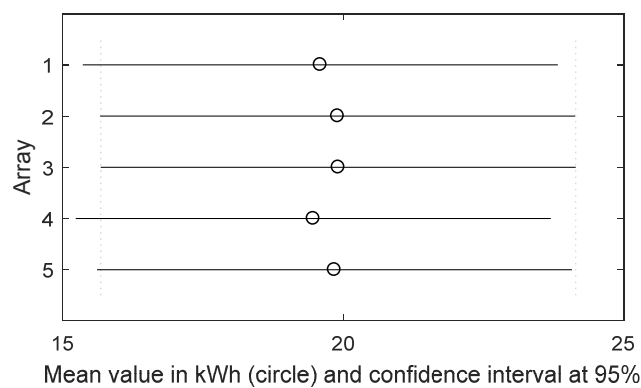
This dataset is constituted by 31 cumulative samples/array, each sample being the sum of 144 samples/day. The energy dataset of the first month does not contain outliers. The  $p$ -value<sub>HDT</sub>  $> \alpha = 0.05$  for each distribution, so all the distributions are unimodal. To apply ANOVA, conditions (b) and (c) have to be verified.

**Table 1.**  $p$ -Value of HDT, JBT, BT, ANOVA for the energy distribution of the arrays, and mean in kWh and spread with respect to the global mean for one-month analysis (January).

Array Number	1	2	3	4	5
$p\text{-value}_{\text{HDT}}$	0.628	0.364	0.674	0.658	0.670
JBT					
JB	0.699	0.676	0.674	0.718	0.638
$p\text{-value}_{\text{JBT}}$	0.500	0.500	0.500	0.500	0.500
$p\text{-value}_{\text{BT}}$			0.999		
$p\text{-value}_{\text{AN}}$			0.999		
Mean (kWh)	19.59	19.90	19.91	19.46	19.84
Global mean			19.74		
Spread %	−0.76	0.80	0.85	−1.40	0.51

Table 1 reports the JB values and the related  $p\text{-value}_{\text{JBT}}$ ; as  $p\text{-value}_{\text{JBT}} > \alpha = 0.05$ , all the distributions are Gaussian, so condition (b) of ANOVA is satisfied. Condition (c) about the homoscedasticity has to be verified by means of BT (see Figure 2). The  $p\text{-value}_{\text{BT}} = 0.999$  in Table 1 (again higher than  $\alpha = 0.05$ ) says that the homoscedasticity is verified, then all the variances are equal. Therefore, the main conditions of the flow chart in Figure 2 (blocks 1,2,3) are satisfied and ANOVA can be applied. The  $p\text{-value}_{\text{AN}} = 0.999$  in Table 1 says that the distributions have the same mean values, so all the arrays have produced the same energy in this month. Figure 4 is the box plot of ANOVA. For each box, the central red mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points. Figure 4 highlights that the five distributions have produced almost the same energy, both with respect to the median value (in red color) and to the first and third inter-quartiles; moreover, outliers are absent. Therefore, no anomaly is present in the PV plant. Particularly, from PVGIS [31], it results that the estimated average energy of the PV plant in January should be about 3173 kWh, corresponding to a daily average energy for each array of about  $3173/(31 \times 5) = 20.5$  kWh, that is almost equal to the global mean value 19.74 kWh of Table 1. Figure 5 diagrams the mean value and confidence interval at 95% of each distribution; the values are very similar each other, as it results also from Table 2 that reports the one-to-one comparisons of the mean values. In particular, the high  $p$ -values confirm that the differences are not significant.

**Figure 4.** Box plot of ANOVA test of the five distributions for the one-month analysis (January).



**Figure 5.** Mean value of the energy produced by each array for the one-month analysis (January).

**Table 2.** One-to-one comparison of the means for the one-month analysis (January).

Comparison between Samples		LowerBound	DifferenceEstimate	UpperBound	$p\text{-Value}_{TT}$
1	2	−8.77	−0.31	8.16	0.999
1	3	−8.78	−0.32	8.15	0.999
1	4	−8.34	0.12	8.59	0.999
1	5	−8.71	−0.25	8.21	0.999
2	3	−8.47	−0.01	8.45	1
2	4	−8.03	0.43	8.90	0.999
2	5	−8.41	0.06	8.52	1
3	4	−8.02	0.44	8.91	0.999
3	5	−8.40	0.06	8.53	1
4	5	−8.84	−0.38	8.09	0.999

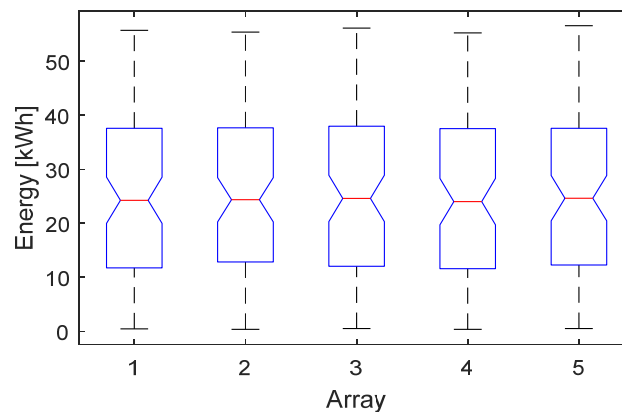
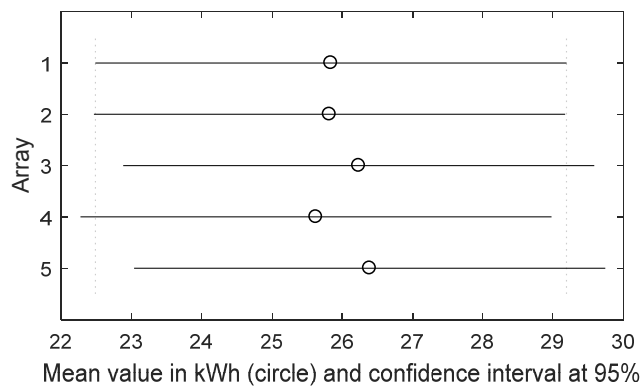
#### 4.2. Three-Months Analysis (January–March)

Table 3 reports the main numerical values of the parameters calculated by following the algorithm of Figure 2 for the energy dataset of three months, including the first month already considered in the previous analysis. This dataset is constituted by 90 cumulative samples/array, each sample being the sum of 144 samples/day. The  $p\text{-value}_{HDT} > \alpha = 0.05$  for each distribution, so all the distributions are still unimodal. The  $p\text{-value}_{JBT} > \alpha = 0.05$ , so all the distributions are gaussian and the condition (b) of ANOVA is satisfied. The homoscedasticity is also satisfied ( $p\text{-value}_{BT} > \alpha = 0.05$ ). Therefore, the main conditions of the flow chart in Figure 2 (blocks 1,2,3) are satisfied and ANOVA can be newly applied. The  $p\text{-value}_{AN} = 0.998$  affirms that the distributions have the same mean values, so all the arrays have produced the same energy also in these three months. Figure 6 is the box plot of ANOVA and it highlights that the five distributions have produced almost the same energy, both with respect to the median value (in red color) and to the first and third inter-quartiles; moreover, outliers are not present. Therefore, no anomaly is present in the PV plant in these three months. Particularly, from PVGIS [31], it results that the estimated average energy of the PV plant in the period January–June should be about 11,695 kWh, corresponding to a daily average energy for each array of about  $11,695/(90 \times 5) = 25.99$  kWh, that is almost equal to the global mean value 25.95 kWh of Table 2.

Figure 7 plots the mean value and confidence interval at 95% of each distribution; the values are very similar each other, as it results also from Table 4 that reports the one-to-one comparisons of the mean values. In particular, the high  $p$ -values confirm that the differences are not significant.

**Table 3.**  $p$ -Value of HDT, JBT, BT, ANOVA for the energy distribution of the arrays, and mean in kWh and spread with respect to the global mean for three-month analysis (January–March).

Array Number	1	2	3	4	5
$p\text{-value}_{\text{HDT}}$	0.776	0.892	0.818	0.856	0.830
$p\text{-value}_{\text{JB}}$	5.127	5.198	5.114	5.177	5.090
$p\text{-value}_{\text{JBT}}$	0.054	0.053	0.054	0.053	0.055
$p\text{-value}_{\text{BT}}$			0.999		
$p\text{-value}_{\text{AN}}$			0.998		
Mean (kWh)	25.84	25.82	26.23	25.62	26.39
Global mean			25.95		
Spread %	−0.54	−0.62	0.98	−1.36	1.58

**Figure 6.** Box plot of K-W test of the five distributions for the three-months analysis (January–March).**Figure 7.** Mean value of the energy produced by each array for the three-months analysis (January–March).**Table 4.** One-to-one comparison of the means for the three-months analysis (January–March).

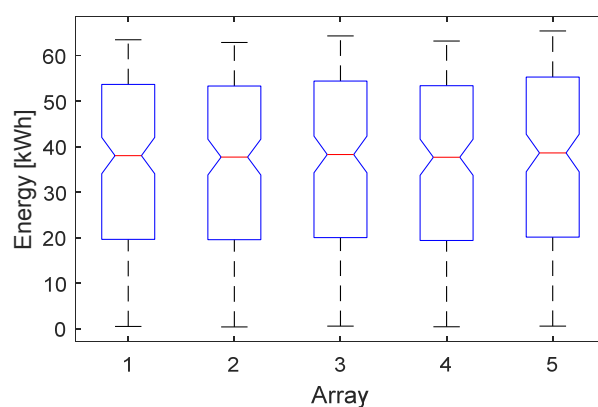
Comparison between Samples		LowerBound	DifferenceEstimate	UpperBound	$p\text{-Value}_{\text{TT}}$
1	2	−6.68	0.02	6.73	1
1	3	−7.10	−0.39	6.31	0.999
1	4	−6.49	0.21	6.92	0.999
1	5	−7.25	−0.55	6.15	0.999
2	3	−7.12	−0.41	6.29	0.999
2	4	−6.51	0.19	6.90	0.999
2	5	−7.27	−0.57	6.13	0.999
3	4	−6.09	0.60	7.31	0.999
3	5	−6.86	−0.15	6.55	0.999
4	5	−7.47	−0.76	5.94	0.997

### 4.3. Six-Months Analysis (January–June)

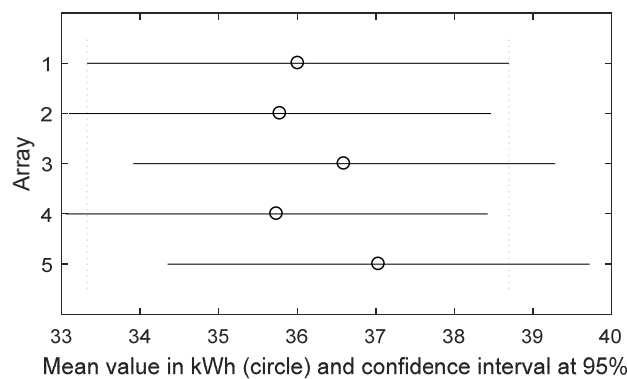
Table 5 displays the main numerical values of the parameters obtained by the application of the procedure in Figure 2, after updating the previous energy dataset (used for the January–March analysis), by adding the data of the successive three months. This dataset is constituted by 181 cumulative samples/array, each sample being the sum of 144 samples/day. The pre-processing of the new dataset has excluded the presence of outliers. The  $p\text{-value}_{\text{HDT}} > \alpha = 0.05$  for each distribution, so all the distributions are still unimodal. As  $p\text{-value}_{\text{JBT}} < \alpha = 0.05$  for each distribution, then the null hypothesis is rejected and the constraint of the block 2 (corresponding to the condition (b) of ANOVA) is not satisfied: the distributions are not Gaussian. Therefore, it has no sense to verify the homoscedasticity, because it is mandatory to use a nonparametric test. As no outlier is present, it is advisable to use K-W, as suggested by the green block. The  $p\text{-value}_{\text{K-W}} = 0.861$  affirms that the distributions have the same median values, so all the arrays have produced the same energy also in these six months, even if the distributions are no longer Gaussian. Figure 8 is the box plot of K-W and it highlights that the five distributions have produced almost the same energy, both with respect to the median value (in red color) and to the first and third inter-quartiles; moreover, it is confirmed that outliers are not present. Therefore, no anomaly is present in the PV plant in these six months. Particularly, from PVGIS [31], it results that the estimated average energy of the PV plant in the period January–June should be about 32,285 kWh, corresponding to a daily average energy for each array of about  $32,285/(181 \times 5) = 35.67$  kWh, that is almost equal to the global mean value 36.23 kWh of Table 3. Figure 9 illustrates the mean value and confidence interval at 95% of each distribution; the values are very similar each other, as it results also from Table 6 that reports the one-to-one comparisons of the mean values. In particular, the high  $p$ -values confirm that the differences are not significant.

**Table 5.**  $p$ -Value of HDT, JBT, BT, (KW) for the energy distribution of the arrays, and mean in kWh and spread with respect to the global mean for six-months analysis (January–June).

Array Number	1	2	3	4	5
$p\text{-value}_{\text{HDT}}$	0.794	0.722	0.782	0.808	0.842
<b>JBT</b>					
JB	13.95	14.02	13.92	13.94	13.85
$p\text{-value}_{\text{JBT}}$	0.007	0.007	0.007	0.007	0.007
$p\text{-value}_{\text{BT}}$					
$p\text{-value}_{\text{K-W}}$			0.861		
Mean (kWh)	36.01	35.78	36.60	35.74	37.04
Global mean			36.23		
Spread %	−0.60	−1.23	1.02	−1.35	2.23



**Figure 8.** Box plot of K-W test of the five distributions for the six-months analysis (January–June).



**Figure 9.** Mean value of the energy produced by each array for the three-months analysis (January–June).

**Table 6.** One-to-one comparison of the means for the three-months analysis (January–June).

Comparison between Samples		LowerBound	DifferenceEstimate	UpperBound	$p\text{-Value}_{TT}$
1	2	−5.14	0.22	5.60	0.999
1	3	−5.96	−0.58	4.78	0.999
1	4	−5.10	0.27	5.64	0.999
1	5	−6.40	−1.02	4.34	0.985
2	3	−6.18	−0.81	4.55	0.993
2	4	−5.33	0.04	5.41	1
2	5	−6.62	−1.25	4.11	0.968
3	4	−4.51	0.85	6.23	0.992
3	5	−5.81	−0.43	4.93	0.999
4	5	−6.67	−1.29	4.07	0.964

#### 4.4. One-Year Analysis (January–December)

Table 7 displays the main numerical values of the parameters obtained by the application of the procedure in Figure 2, after updating the previous energy dataset (used for the January–June analysis), by adding the data of the successive six months. This dataset is constituted by 365 cumulative samples/array, each sample being the sum of 144 samples/day. The pre-processing of the new dataset has excluded the presence of outliers. The  $p\text{-value}_{HDT} > \alpha = 0.05$  for each distribution, except for the distribution n. 4, for which  $p\text{-value}_{HDT}(4) = 0.006 < \alpha = 0.05$ ; therefore, the condition of the block 1 about the unimodality is not satisfied for all the distributions and ANOVA cannot be applied. Consequently, it is mandatory to use a nonparametric test. As no outlier is present, it is advisable to apply K-W, as suggested by the green block. As  $p\text{-value}_{K-W} = 0.009 < \alpha = 0.05$ , the null hypothesis is rejected, so the distributions have different median values. This implies that the arrays have not produced the same energy in the complete year, even if they had produced the same energy for the first six months. Figure 10 is the box plot of K-W and it highlights that the median value of the distribution n. 4 is significantly different from the others. It is also confirmed that outliers are not present. Particularly, from PVGIS [31], it results that the estimated average energy of the PV plant in the period January–December should be about 64,724 kWh, corresponding to a daily average energy for each array of about  $64,724 / (365 \times 5) = 35.46$  kWh, that is almost equal to the global mean value 35.13 kWh of Table 4. Therefore, high-intensity anomaly is not present, but a low-intensity anomaly is detected in the array n. 4, as confirmed also by the spreads of the mean values reported in Table 4. It can be observed that the array n. 4 produced 6.54% less than the average energy of the PV plant. Figure 11 shows the mean value and confidence interval at 95% of each distribution. It can be observed that the array n. 4 is very different from the other ones, as it results also from Table 8 that reports the

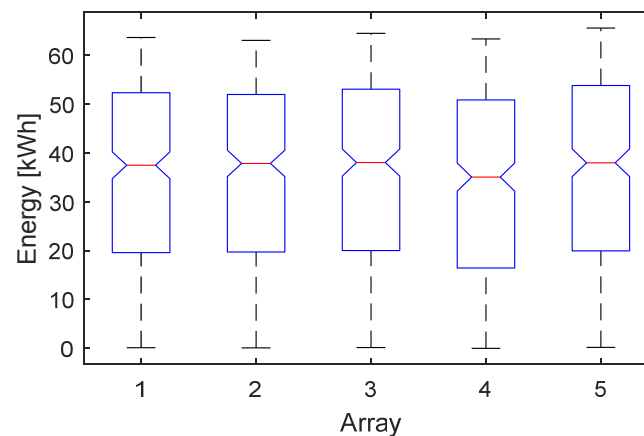


one-to-one comparisons of the mean values. In particular, the  $p$ -value  $0.043 < \alpha = 0.05$  rejects the hypothesis that the distribution 4 and 5 have the same mean value.

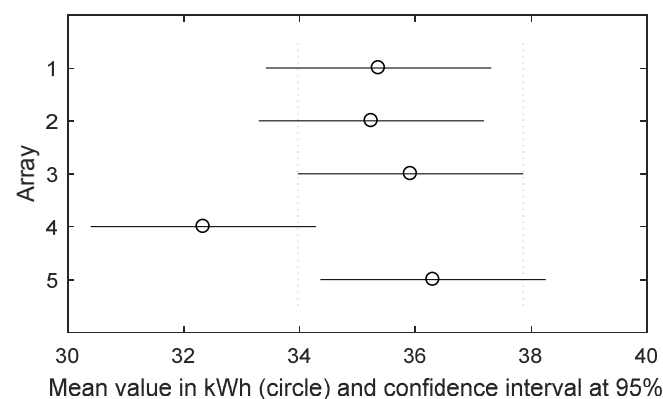
**Table 7.**  $p$ -Value of HDT and K-W for the energy distribution of the arrays, and mean in kWh and spread with respect to the global mean for one-year analysis.

	Array Number				
	1	2	3	4	5
$p$ -value <sub>HDT</sub>	0.820	0.846	0.898	0.006	0.636
$p$ -value <sub>BT</sub>	-	-	-	-	-
$p$ -value <sub>BT</sub>	-	-	-	-	-
$p$ -value <sub>K-W</sub>			0.009		
Mean (hWh)	35.37	35.24	35.92	32.83	36.31
Global mean			35.13		
Spread %	0.68	0.32	2.25	−6.54	3.35

Global mean	35.13				
Spread %	0.68	0.32	2.25	−6.54	3.35



**Figure 10.** Box plot of K-W test of the five distributions for the one-year analysis.



**Figure 11.** Mean value of the energy produced by each array for the one-year analysis.

**Table 8.** One-to-one comparison of the means for the one-year analysis.

Comparison between Samples		LowerBound	DifferenceEstimate	UpperBound	p-Value <sub>TT</sub>
1	2	−3.76	0.12	4.01	0.999
1	3	−4.44	−0.55	3.33	0.995
1	4	−0.86	3.02	6.91	0.210
1	5	−4.83	−0.94	2.95	0.964
2	3	−4.56	−0.67	3.21	0.989
2	4	−0.98	2.90	6.79	0.249
2	5	−4.95	−1.06	2.82	0.945
3	4	−0.31	3.57	7.47	0.088
3	5	−4.28	−0.38	3.50	0.998
4	5	−7.85	−3.96	−0.07	0.043

## 5. Conclusions

The paper proposes a statistical algorithm to monitor the energy performance of PV plants and detect anomalies. The procedure is cumulative and the algorithm can be iterated, as new data are acquired by the measurement system, in order to follow the most important benchmarks. The case study, referred to a real operating PV system, has shown the results of four cumulative analyses, starting from the dataset of only one month and finishing with a yearly-based dataset. As real operating PV systems are affected by atmospheric phenomena, their energy distributions are never perfectly Gaussian, so parametric tests should never be applied. Instead, as ANOVA can be applied for modest violations of its assumptions, the issue consists in evaluating the violation, in order to decide whether it is negligible or not. The proposed methodology, based on hypothesis tests, allows this evaluation. The first two analyses (based on the data of one and three months, respectively) have been carried out by means of the parametric test ANOVA, whereas the third and the fourth analyses have been based on the nonparametric test K-W, because the mandatory ANOVA's assumptions were not satisfied. Moreover, while the first three analyses have not evidenced any anomaly in the PV plant—in fact the energy distributions of the arrays were almost equal—instead the last analysis has shown a not negligible anomaly in the array n. 4. The proposed methodology does not allow identifying the origin of the anomaly, but only to detect and locate it. Finally, the proposed procedure is particularly effective in absence of environmental parameters, i.e., for monitoring PV plants not equipped with a weather station. In this case, the procedure allows extracting the main operating features of the PV plants without adding new hardware; thus, this approach is also cheap. Nevertheless, when a commercial PV plant has to be evaluated, it is mandatory to take into account the environmental parameters; so, if the PV plant is not equipped by a weather station, it is necessary to add this component and to use the monitoring methodologies based on the environmental data, even though this results more expensive.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Xiao, W.; Ozog, N.; Dundorf, W.G. Topology study of photovoltaic interface for maximum power point tracking. *IEEE Trans. Ind. Electron.* **2007**, *54*, 1696–1704. [[CrossRef](#)]
2. Mutoh, N.; Inoue, T. A control method to charge series-connected ultraelectric double-layer capacitors suitable for photovoltaic generation systems combining MPPT control method. *IEEE Trans. Ind. Electron.* **2007**, *54*, 374–383. [[CrossRef](#)]
3. Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. ANN Sizing Procedure for the Day-Ahead Output Power Forecast of a PV Plant. *Appl. Sci.* **2017**, *7*, 622. [[CrossRef](#)]
4. Dellino, G.; Laudadio, T.; Mari, R.; Mastronardi, N.; Meloni, C.; Vergura, S. Energy Production Forecasting in a PV Plant Using Transfer Function Models. In Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), Roma, Italy, 10–13 June 2015.

5. Vergura, S.; Pavan, A.M. On the photovoltaic explicit empirical model: Operations along the current-voltage curve. In Proceedings of the 2015 International Conference on Clean Electrical Power (ICCEP), Taormina, Italy, 16–18 June 2015.
6. Guerriero, P.; di Napoli, F.; Vallone, G.; D'Alessandro, V.; Daliento, S. Monitoring and diagnostics of PV plants by a wireless self-powered sensor for individual panels. *IEEE J. Photovolt.* **2015**, *6*, 286–294. [[CrossRef](#)]
7. Takashima, T.; Yamaguchi, J.; Otani, K.; Kato, K.; Ishida, M. Experimental Studies of Failure Detection Methods in PV module strings. In Proceedings of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion, Waikoloa, HI, USA, 7–12 May 2006; Volume 2, pp. 2227–2230.
8. Breitenstein, O.; Rakotoniaina, J.P.; Al Rifai, M.H. Quantitative evaluation of shunts in solar cells by lock-in thermography. *Prog. Photovolt. Res. Appl.* **2003**, *11*, 515–526. [[CrossRef](#)]
9. Grimaccia, F.; Leva, S.; Dolara, A.; Aghaei, M. Survey on PV Modules' Common Faults after an O&M Flight Extensive Campaign over Different Plants in Italy. *IEEE J. Photovolt.* **2017**, *7*, 810–816.
10. Johnston, S.; Guthrey, H.; Yan, F.; Zaunbrecher, K.; Al-Jassim, M.; Rakotoniaina, P.; Kaes, M. Correlating multicrystalline silicon defect types using photoluminescence, defect-band emission, and lock-in thermography imaging techniques. *IEEE J. Photovolt.* **2014**, *4*, 348–354. [[CrossRef](#)]
11. Peloso, M.P.; Meng, L.; Bhatia, C.S. Combined thermography and luminescence imaging to characterize the spatial performance of multicrystalline Si wafer solar cells. *IEEE J. Photovolt.* **2015**, *5*, 102–111. [[CrossRef](#)]
12. Vergura, S.; Marino, F. Quantitative and Computer Aided Thermography-based Diagnostics for PV Devices: Part I—Framework. *IEEE J. Photovolt.* **2017**, *7*, 822–827. [[CrossRef](#)]
13. Vergura, S.; Colaprico, M.; de Ruvo, M.F.; Marino, F. A Quantitative and Computer Aided Thermography-based Diagnostics for PV Devices: Part II—Platform and Results. *IEEE J. Photovolt.* **2017**, *7*, 237–243. [[CrossRef](#)]
14. Mekki, H.; Mellit, A.; Salhi, H. Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. *Simul. Model. Pract. Theory* **2016**, *67*, 1–13. [[CrossRef](#)]
15. Harrou, F.; Sun, Y.; Kara, K.; Chouder, A.; Silvestre, S.; Garoudja, E. Statistical fault detection in photovoltaic systems. *Sol. Energy* **2017**, *150*, 485–499.
16. Vergura, S.; Carpentieri, M. Statistics to detect low-intensity anomalies in PV systems. *Energies* **2018**, *11*, 30.
17. Ventura, C.; Tina, G.M. Development of models for on line diagnostic and energy assessment analysis of PV power plants: The study case of 1 MW Sicilian PV plant. *Energy Procedia* **2015**, *83*, 248–257. [[CrossRef](#)]
18. Silvestre, S.; Kichou, S.; Chouder, A.; Nofuentes, G.; Karatepe, E. Analysis of current and voltage indicators in grid connected PV (photovoltaic) systems working in faulty and partial shading conditions. *Energy* **2015**, *86*, 42–50. [[CrossRef](#)]
19. Vergura, S. A Complete and Simplified Datasheet-based Model of PV Cells in Variable Environmental Conditions for Circuit Simulation. *Energies* **2016**, *9*, 326. [[CrossRef](#)]
20. Vergura, S. Scalable Model of PV Cell in Variable Environment Condition based on the Manufacturer Datasheet for Circuit Simulation. In Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), Roma, Italy, 10–13 June 2015.
21. Chao, K.H.; Ho, S.H.; Wang, M.H. Modeling and fault diagnosis of a photovoltaic system. *Electr. Power Syst. Res.* **2008**, *78*, 97–105. [[CrossRef](#)]
22. Hamdaoui, M.; Rabhi, A.; Hajjaji, A.; Rahmoum, M.; Azizi, M. Monitoring and control of the performances for photovoltaic systems. In Proceedings of the International Renewable Energy Congress, Sousse, Tunisia, 5–7 November 2009.
23. Kim, I.-S. On-line fault detection algorithm of a photovoltaic system using wavelet transform. *Sol. Energy* **2016**, *226*, 137–145. [[CrossRef](#)]
24. Rabhia, A.; El Hajjajia, A.; Tinab, M.H.; Alia, G.M. Real time fault detection in photovoltaic systems. *Energy Procedia* **2017**, *11*, 914–923.
25. Plato, R.; Martel, J.; Woodruff, N.; Chau, T.Y. Online fault detection in PV systems. *IEEE Trans. Sustain. Energy* **2015**, *6*, 1200–1207. [[CrossRef](#)]
26. Ando, B.; Bagalio, A.; Pistorio, A. Sentinella: Smart monitoring of photovoltaic systems at panel level. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2188–2199. [[CrossRef](#)]
27. Harrou, F.; Sun, Y.; Taghezouit, B.; Saidi, A.; Hamlati, M.E. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renew. Energy* **2018**, *116*, 22–37. [[CrossRef](#)]

28. International Electrotechnical Commission (IEC). *Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange and Analysis*, International Standard 61724, 1st ed.; International Electrotechnical Commission (IEC): Geneva, Switzerland, 1998.
29. Leloux, J.; Narvarte, L.; Trebosc, D. Review of the performance of residential PV systems in Belgium. *Renew. Sustain. Energy Rev.* **2012**, *16*, 178–184. [[CrossRef](#)]
30. Leloux, J.; Narvarte, L.; Trebosc, D. Review of the performance of residential PV systems in France. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1369–1376. [[CrossRef](#)]
31. Available online: [http://re.jrc.ec.europa.eu/pvg\\_tools/en/tools.html](http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html) (accessed on 20 January 2018).
32. Rohatgi, V.K.; Székely, G.J. Sharp inequalities between skewness and kurtosis. *Stat. Probab. Lett.* **1989**, *8*, 297–299. [[CrossRef](#)]
33. Klaassen, C.A.J.; Mokveld, P.J.; van Es, B. Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Stat. Probab. Lett.* **2000**, *50*, 131–135. [[CrossRef](#)]
34. Hogg, R.V.; Ledolter, J. *Engineering Statistics*; MacMillan: Basingstoke, UK, 1987.
35. Roussas, G. *An Introduction to Probability and Statistical Inference*; Academic Press: Cambridge, MA, USA, 2015.
36. Hartigan, J.A.; Hartigan, P.M. The dip test of unimodality. *Ann. Stat.* **1985**, *13*, 70–84. [[CrossRef](#)]
37. Gibbons, J.D. *Nonparametric Statistical Inference*, 2nd ed.; M. Dekker: New York, NY, USA, 1985.
38. Hollander, M.; Wolfe, D.A. *Nonparametric Statistical Methods*; Wiley: Hoboken, NJ, USA, 1973.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).