Statistics to Detect Low-Intensity Anomalies in PV Systems

Silvano Vergura * [ID] **and Mario Carpentieri**

Department of Electrical and Information Engineering, Polytechnic University of Bari, 70125 Bari, Italy; mario.carpentieri@poliba.it

* Correspondence: silvano.vergura@poliba.it; Tel.: +39-080-5963-590; Fax: +39-080-5963-410

Abstract: The aim of this paper is the monitoring of the energy performance of Photovoltaic (PV) plants in order to detect the presence of low-intensity anomalies, before they become failures or faults. The approach is based on several statistical tools, which are applied iteratively as the data are acquired. At every loop, new data are added to the previous ones, and a proposed procedure is applied to the new dataset, therefore the analysis is carried out on cumulative data. In this way, it is possible to track some specific parameters and to monitor that identical arrays in the same operating conditions produce the same energy. The procedure is based on parametric (ANOVA) and non-parametric tests, and results effective in locating anomalies. Three cumulative case studies, based on a real operating PV plant, are analyzed.

Keywords: ANOVA; non-parametric test; unimodality; homoscedasticity; kurtosis; skewness

1. Introduction

A crucial problem for the management of PV plants is the strong dependence of the system response on many extrinsic factors, such as irradiance intensity, cloudiness and pollution, humidity, air velocity, environment temperature, and cell temperature. Several models that are able to evaluate the effects of such uncertainties have been proposed in [1–9], whereas the statistical methods have been used for the design of PV plants [10,11] and for the suitable choice of the electrical components [12,13]. When a PV plant is operating, a monitoring system to check the performance in all the environmental conditions is needed.

PV modules are the main components of a PV plant, so a deep attention has to be focused on their health state. For this aim, techniques that are commonly used to verify the presence of typical defects in PV modules are based on the infrared analysis [14,15], on the luminescence imaging [16], on their combination [17], while automatic procedures to extract information by the thermo-grams are proposed in [18,19]. Nevertheless, these approaches regard single modules of PV plants and they are useful when a defect has been roughly individualized. When there is no information about the general operation of the PV plant, other techniques have to be considered, as neural networks [20], check of the electrical variables [21], AutoRegressive Integrated Moving Average (ARIMA) models [22], statistics, and so on. For example, in [23], the authors proposed a decision algorithm based on both the descriptive and inferential statistics. The former one is used to characterize the data-set under investigation, on the basis of a descriptive model or a distribution family. The latter one is used for unknown dataset and consists of a data producing process with the aim to infer the characteristics of the population on the basis of a sub-set of the sample data, in order to predict anomalies, even though the amount of field measurements is limited. Nevertheless, when important failures/anomalies (short circuit, islanding, and so on) occur, the electrical variables and the associated energy have fast and not negligible variations, so they are instantaneously detected. These events can be classified as

high-intensity anomalies, because they produce a drastic change. Instead, the low-intensity anomalies (ageing of the components, minimal partial shading, corrosion and so on) produce a minimal variation on the values of the electrical variables and produced energy, so it is not trivial to detect them. Moreover, these anomalies can evolve in failures or faults, so a fast and correct identification can prevent them and limit the out of order. This paper proposes a cheap, easy, and fast statistical based on an algorithm that processes the data usually stored in the datalogger (Easy, Fronius, Bussolengo (VR), Italy) of the PV plants, therefore not requiring new hardware/components.

The proposed methodology evaluates a priori the presence of high-intensity anomaly in each array of the PV plant, then discriminates the statistical distributions between unimodal and multimodal, and finally applies several statistical tools to extract definitive conclusions about the similarity among the energy produced by the arrays, in order to detect and locate possible anomalies.

The paper presents the methodology in Section 2, describes the characteristics of a real PV plant under investigation in Section 3, and discusses the results in Section 4. Conclusions end the paper.

2. Statistics-Based Procedure

The random variability of the environmental conditions affects the irradiance intensity for the PV generators. During clear days an analytic expression for solar irradiance can be defined, whereas it is not possible for cloudy days, when, instead, the statistical approach results are very effective. In this paper, we consider that the PV plant is composed of k identical arrays with a devoted unit of measurement for each array. Each unit stores measurements of voltage, current and energy produced, whereas a central unit acquires the data coming from all the arrays. For our aims, let us consider the dataset constituted by the energy values. The descriptive statistical parameters (mean, median, and variance) are valuable tools to detect large failures (short circuits, open circuits, large shading, etc.) that provoke not negligible energy reductions. Instead, when a low-intensity anomaly appears in an array (light ageing, partial shading, and so on), the energy reduction is limited, and the previous parameters are not able to distinguish an anomaly from an acceptable tolerance. For this aim, here, we propose the methodology reported in Figure 1, based on the comparison among the energy performances of the arrays. This goal can be pursued by means of parametric or non-parametric tests. When the data distribution is not normal, a non-parametric test must be used. As the non-parametric tests make only mild assumptions, they are less powerful than the parametric ones for normally distributed data. So, it is preferable to use a parametric test each time that it is possible. Among the parametric tests, the Analysis of Variance (ANOVA) [24] is an effective tool for inferential purposes and returns reliable feedback, when comparing the variance inside each array distribution and among the arrays' distributions. In other words, ANOVA evaluates whether the differences of the mean values of the different groups are statistically significant or not. So, the null hypothesis H_0 is that the means μ_i are equal, i.e.,

$$H_0 : \ \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k \tag{1}$$

versus the alternative one that at least one distribution has mean different from the others.

Nevertheless, it can be applied only under three assumptions:

(a) equal variance for all the distributions;
(b) all the distributions are gaussian; and,
(c) all of the observations are independent each other.

In our case, the hypothesis (c) is always verified, because the k local measurement units are different; moreover, a modest violation for the assumptions (a) and (b) are allowed, because ANOVA is a robust test.

So, before applying ANOVA test, several verifications are needed. The first one regards the unimodality of the statistical distributions; in fact, a multimodality distribution (Figure 2) is surely not

normally distributed, thus violates the condition (b). The check on the unimodality can be done by means of skewness and kurtosis, as will be explained later. The skewness of a distribution is defined as:

$$\sigma_k = \frac{E(x - \mu)^3}{\sigma^3} \qquad (2)$$

being μ the mean of the data x, σ the standard deviation of x, and $E(z)$ the expected value of the quantity z. From a mathematical point of view, the skewness is a third standardized moment and it measures the asymmetry of the data around the mean. For

- $\sigma_k = 0$ the distribution is gaussian;
- $\sigma_k > 0$ the data are spread out more to the right of the mean than to the left; and,
- $\sigma_k < 0$ the data are spread out more to the left.

Pay attention that $\sigma_k = 0$ is a necessary but not sufficient condition for the symmetry; in fact, symmetric distributions with $\sigma_k \neq 0$ cannot exist, but asymmetric distributions with $\sigma_k = 0$ can.

The *kurtosis*, instead, is a measure of how outlier-prone a distribution is. Since the kurtosis can have several formulas, here, we consider the Pearson's kurtosis less 3, i.e.,

$$k_u = \frac{E(x - \mu)^4}{\sigma^4} - 3 \qquad (3)$$

From a mathematical point of view, the kurtosis is a fourth standardized moment and results that:

- $k_u = 0$ the distribution is gaussian;
- $k_u < 0$ the distribution is less outlier-prone than the gaussian one; and,
- $k_u > 0$ the distribution is more outlier-prone than the gaussian one.

Now, it results that skewness and kurtosis of a unimodal distribution have to satisfy the following constraint [25]:

$$U = \sigma_k^2 - k_u \leq \frac{6}{5} \qquad (4)$$

whether the distribution mode is equal to the distribution mean, and kurtosis is calculated as in (3).

Instead, a more inclusive constraint, valid whether the distribution mode is different from the distribution mean, is [26]:

$$U^* = \sigma_k^2 - k_u \leq \frac{186}{125} \qquad (5)$$

So, after the calculation of skewness and kurtosis, and of the mode and mean values of each distribution, if the constraint (4) (when mode = mean) or (5) (when mode $\neq$ mean) is not satisfied, it needs to apply the Kruskal-Wallis test (K-W) [27,28] or Mood's median test (MM), i.e., a non-parametric test, which requires only that the measurements come from a continuous distribution, even though it is not gaussian. Moreover, MM is preferred to K-W, when outliers are present.

Instead, K-W studies the variance based on the ranks of the data values and not the data values themselves. K-W is based on the same null hypothesis (1), i.e., that the data belong to k distributions having equal mean values. So, if p-value $< \alpha = 0.05$, the null hypothesis (1) can be refused, implying that the k arrays are not producing the same amount of energy, a low-intensity is present, and an alert is notified, highlighting which are the worst performing arrays. The parameter α is the well-known significance level.

Instead, if the constraint (4) (when mode = mean) or (5) (when mode $\neq$ mean) is satisfied, ANOVA could be applied under the conditions (a) and (b). Now, the check on the variances (condition (a)) can be made by means of the homoscedasticity's test (also known as homogeneity of variance), while the condition (b) can be verified by means of a normal probability test [29]. The normal probability plot returns the range of percentiles where the distribution is gaussian.
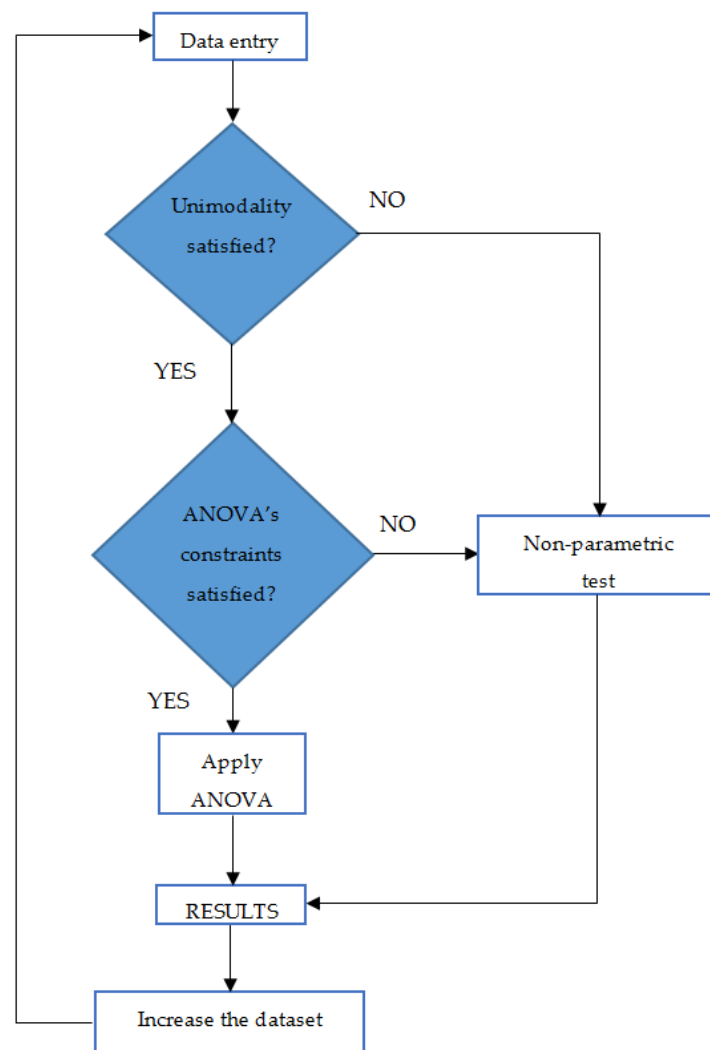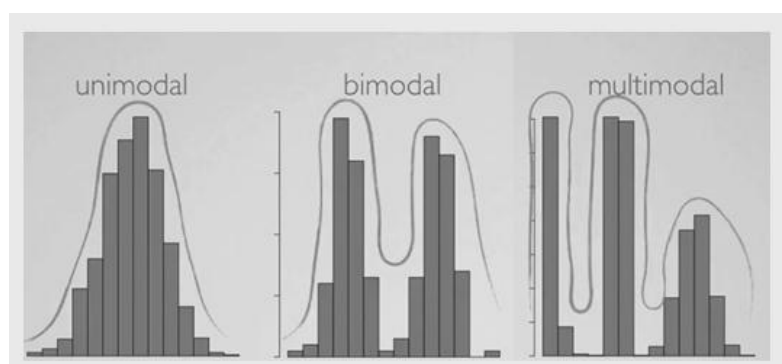
Figure 1. Proposed approach.

Figure 2. Different modality.

If the above assumptions are verified, then it is possible to apply ANOVA test to obtain the information related to the *p*-value, otherwise a non-parametric test (K-W or MM) has to be applied.

The final step is the analysis of all the calculated statistical parameters. A support is provided by the Pearson's distribution that we have used as approximation function, because we observed that the kurtosis is necessary to fit the data. In particular, the Pearson proposed is a Pearson Type IV

(p4PDF), which is an asymmetric version of the Student's *t* distribution. By considering the first four standardized moments mean μ, standard deviation (STD) σ, skewness σ_k, and kurtosis k_u, the p4PDF can be written as:

$$f(x|\mu,\sigma,\sigma_k,k_u) = \frac{k_u}{\sigma}\left[1 + \left(\frac{\hat{x}-\lambda}{a}\right)^2\right]^{-n} exp\left[-k tan^{-1}\left(\frac{\hat{x}-\lambda}{a}\right)\right] \ (n > 1/2) \tag{6}$$

where $\hat{x} = \dfrac{x-\mu}{\sigma}$, whereas:

$$
\begin{aligned}
\lambda &= \frac{a\kappa}{b} \\
\kappa &= \frac{2c_1(1-n)}{\sqrt{4c_0c_2 - c_1{}^2}} \\
a &= \sqrt{\frac{b^2 + (b-1)}{b^2 + \kappa^2}} \\
b &= 2n - 1 \\
n &= 1/2c_2
\end{aligned}
\tag{7}
$$

And:

$$
\begin{aligned}
\beta_1 &= \sigma_k^2 \\
c_0 &= \frac{4k_u - 3\beta_1}{D} \\
c_1 &= \frac{\sigma_k \cdot (k_u + 3)}{D} \\
c_2 &= \frac{\sigma_k(2k_u - 3\beta_1 - 6)}{D} \\
D &= 10k_u - 12\beta_1 - 18
\end{aligned}
\tag{8}
$$

In (7) n, κ, a, and λ are real-valued parameters, and $-\infty < x < \infty$. k_2 is a normalization constant that depends on n, κ, a, and it can be expressed by [30]:

$$k_2 = \frac{\left|\dfrac{\Gamma(n + (i\cdot\kappa/2))}{\Gamma(n)}\right|^2}{a \cdot \psi\left(n - \dfrac{1}{2}, \dfrac{1}{2}\right)} \tag{9}$$

As new data are acquired, the fault estimation and its location becomes more accurate. The proposed algorithm extracts statistical information from the produced energy by the arrays and performs a continuous supervision of the operation of the PV plant.

As in real cases, it is unlikely that the data distribution is perfectly gaussian and since ANOVA test is allowed also for a modest violation of condition (b), skewness and kurtosis, already defined in (2) and (3), can be used to quantify the divergence of the real distributions of the *k* arrays from a gaussian one and to select the most suitable path.

3. Case Study

The behavior of a 19.8 kWp grid-connected PV plant, located in the South of Italy, has been analyzed. The 132 panels of the plant are partitioned in $k = 6$ equal arrays. Each PV module (Sol 150 mono-crystalline, Solterra, Chiasso, Switzerland) has a nominal power of 150 Wp, so the peak power of each array is 3300 Wp. A 3000 W inverter (Sunny Boy 3000, SMA, Milano, Italy) is connected to each array. The system faces the south, and it is sloped at about 44°. The PV plant is on the roof of a private company building that is taller than any other obstacle around the same building. The inverter will be connected to the grid, only if the PV voltage of the array exceeds a prefixed threshold, which is checked by an internal regulation system of the inverter, and it is able to capture the MPP voltage. The PV plant has a data acquisition system, constituted by a datalogger that acquires the data from

the six inverters each 2 s. An internal software calculates the average value of the sampled data each 10 min and stores only this value into the database. The daily and cumulative values calculated and stored into the database are: (a) power and energy on the AC side of each inverter; (b) voltage V_{dc} on the DC side of each inverter; and, (c) total number of the working hours. The monitoring system can store up to 400 days. The investigation period refers to a full year, during which the plant has shown several anomalies.

4. Cumulative Statistical Analysis

To analyze the energy performance of the PV plant described in Section 3, a statistics-based algorithm introduced in Section 2 has been used. All of the analyses have been run in Matlab (R2017, MathWorks, Natick, MA, USA) environment by using standard routines and implementing the proposed algorithm.

Several analyses are presented, corresponding to several cycles of the algorithm of Figure 1, as the new data were acquired. So, for each new case study, new samples are added to the previous ones. It should be highlighted that the data have been filtered before inserting them in the proposed routine. The pre-processing is always needed, because some bad data or outliers could be present.

Thus, the incoming analyses represent a cumulative analysis for the statistical monitoring of the PV plant, in order to follow the time behavior of some benchmarks and to detect the low-intensity anomalies. Several analyses will be presented:

- monthly analysis (January);
- quarterly analysis (January–March); and,
- yearly analysis (January–December).

The following results will be reported:

- mean, median, variance and relative spreads of each array, in order to verify whether any large failure is present;
- skewness and kurtosis values, in order to evaluate the unimodality U or U^* of the k distributions, and also to quantify the mismatches with respect to a gaussian distribution; and,
- *p*-value, as explained in Section 2, having fixed $\alpha = 0.05$.

Observing Figure 1, the three analyses can be considered as the cumulative result of the flow-chart, starting from the data of one-month and adding, firstly, the new data of the successive two months (producing the quarterly analysis), and then adding the new data of the successive nine months (producing the yearly analysis). The procedure here described has been applied to the energy dataset of the same PV plant for two years. In this paper, we discuss only the application to the energy dataset of the second year, highlighting that in the first year no anomaly was revealed, the PV plant had produced the expected energy, the arrays had produced about the same energy.

4.1. Monthly Analysis (January)

Figure 3 reports the energy produced by each array while Table 1 reports some statistical parameters (mean, variance and medians of the energy produced by each array, their global means, and the spreads in per cent). It results that the spreads of the means are limited (the out of order of one PV module among the 22 of each array implies about 5% of energy reduction, thus no PV module is out of order), so large failures are not present. In order to evaluate the presence of a low-intensity anomaly, let us apply the flow-chart of Figure 1. To evaluate the unimodality, mode, skewness, and kurtosis are calculated (see Table 1), and are used in (5), because mean and mode are different for each array. For each array, the values of U^* (Table 1) are less than $186/125 = 1.488$, thus the constraint (5) is satisfied and each distribution is unimodal. Now, the ANOVA's constraints would have to be verified by means of homeschedasticity test and normal plot; nevertheless, it can be observed that the values of skewness and kurtosis are very different from zero for each array, so the distributions are surely not-gaussian,

violating the condition (b) of ANOVA test. Moreover, the spreads of the variances (contained in the range $-2.81 \div 3.46$) indicate a not negligible violation also of the condition (a). Therefore, ANOVA cannot be applied, so K-W is chosen, because outliers are not present (see Figure 3). Table 1 reports also the *p*-value of K-W (0.9999) which implies that:

- *p*-value > 0.05, so the null hypothesis H_0 in (1) cannot be refused; and,
- 1-*p*-value < 0.05, so the alternative hypothesis that at least one distribution has the mean different from the other ones has to be rejected.
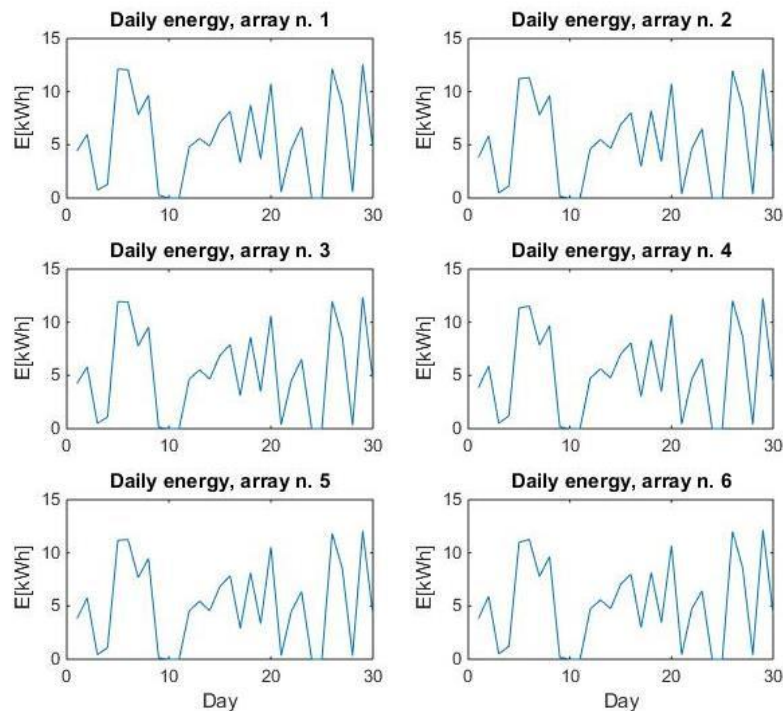

Figure 3. Energy, in kWh, produced by each array during one-month.

Table 1. Mean, median and variance of the daily energy (in kWh) produced by each array and relative spreads for one month. Skewness, kurtosis, and *p*-value.

	Array Number					
	1	**2**	**3**	**4**	**5**	**6**
Mean	5.40	5.17	5.25	5.24	5.11	5.19
Global mean			5.227			
Spread %	3.29	−1.04	0.50	0.23	−2.32	−0.66
Median	4.85	4.66	4.66	4.76	4.55	4.75
Global mean			4.707			
Spread %	2.96	−0.94	−0.94	1.23	−3.26	0.96
Variance	17.67	16.88	17.53	17.10	16.60	16.70
Global mean			17.080			
Spread %	3.46	−1.15	2.61	0.13	−2.81	−2.24
Mode	0.276	0.154	0.125	0.174	0.126	0.181
σ_k	0.135	0.107	0.134	0.100	0.103	0.102
k_u	−0.647	−0.675	−0.626	−0.667	−0.688	−0.644
U*	0.665	0.687	0.643	0.677	0.699	0.655
p-value (K-W)			0.9999			

4.2. Quarterly Analysis (January–March)

This analysis includes the data of the previous one (January). Table 2 reports the same parameters of Table 1, and it results that the spreads of the means are again less than 5% and lower than the previous case, so large failures are not present. Moreover, Equation (5) has to be applied also for this analysis and the U^* values, higher than the previous ones, are less than the limited value 186/125, thus all of the distributions are unimodal. The values of skewness and kurtosis say also that the distributions are not gaussian and flow-chart suggests applying a non-parametric test. The p-value = 0.9996, even if lower than the previous case, confirms the same considerations, i.e., the six arrays have analogous performances in the January–March period, and no anomaly, low-intensity or high-intensity, is present.

Table 2. Mean, median, and variance of the daily energy (in kWh) produced by each array and relative spreads for three months. Skewness, kurtosis and p-value.

| | Array Number | | | | | |
	1	2	3	4	5	6
Mean	8.390	8.218	8.199	8.306	8.098	8.315
Global mean			8.254			
Spread %	1.65	−0.44	−0.67	0.62	−1.89	0.73
Median	8.103	7.905	7.935	7.956	7.767	7.889
Global mean			7.926			
Spread %	2.23	−0.27	0.11	0.38	−2.00	−0.46
Variance	32.313	32.241	32.102	32.617	31.531	32.801
Global mean			32.268			
Spread %	0.14	−0.08	−0.51	1.08	−2.28	1.65
Mode	0.183	0.154	0.125	0.174	0.126	0.175
σ_k	0.211	0.221	0.206	0.220	0.219	0.236
k_u	−1.095	−1.082	−1.103	−1.081	−1.090	−1.065
U*	1.139	1.131	1.454	1.130	1.138	1.121
p-value (K-W)			0.9996			

4.3. Yearly Analysis (January–December)

The yearly analysis includes the previous data of January–March and returns complete information about the performance of the PV plant, because this analysis takes into account the yearly variability of the environmental conditions of the installation site of the PV plant.

Again, the spreads of the means that are reported in Table 3 are less than 5%, so large failures are not present. The corresponding values of mode and mean are different, so the unimodality has to be tested by (5), and all of the distributions satisfy the constraint $U^* < 186/125$. Nevertheless, it is important to point up that the sign of skewness is changed for all the six distributions; so, until March, the data was spread out more to the right of the mean than to the left, but, on the yearly basis, they are spread out more to the left of the mean and this happens for each array. The absolute values of the skewness are higher than the previous cases, so ANOVA cannot be applied. From K-W, p-value = 0.873, so:

- p-value > 0.05, so the null hypothesis H_0 in (1) cannot be refused; and,
- 1-p-value > 0.05, so neither the alternative hypothesis that at least one distribution has the mean different from the other ones can be rejected.

Moreover, not only 1-p-value = 0.127 is higher than α = 0.05, but it is higher than the double of α. In this case, it is concluded that there is a high probability that the yearly average values of the energy produced by the six arrays are different each other. In fact, an in depth view of the spreads of the mean values of Table 3 confirms the criticality of the array n. 5 (−1.79%), with respect the other arrays.

Table 3. Mean, median, and variance of the daily energy (in kWh) produced by each array and relative spreads values for 12 months. Skewness, kurtosis and *p*-value.

	Array Number					
	1	**2**	**3**	**4**	**5**	**6**
Mean	11.84	11.64	11.57	11.77	11.48	11.84
Global mean			11.69			
Spread %	1.25	−0.41	−1.01	0.66	−1.79	1.30
Median	12.60	12.26	12.42	12.38	12.08	12.31
Global mean			12.34			
Spread %	2.07	−0.63	0.61	0.32	−2.14	−0.22
Variance	37.63	37.80	37.01	38.41	36.97	39.34
Global mean			37.86			
Spread %	−0.61	−0.16	−2.24	1.45	−2.36	3.92
Mode	17.98	16.32	11.48	11.95	10.54	16.81
σ_k	−0.370	−0.354	−0.379	−0.353	−0.355	−0.333
k_u	−1.150	−1.168	−1.147	−1.168	−1.168	−1.185
U*	1.287	1.293	1.291	1.292	1.294	1.297
p-value (K-W)			0.873			

4.4. Discussion

The proposed procedure has highlighted that the energy performance of the six arrays are similar for the first three months, however this is not true for whole year, so the proposed cumulative analysis has highlighted a low-intensity anomaly. The procedure is not able to identify the origin of the anomaly, but it detects and locates it. In fact, an in-depth view on the spreads of the means reported the Table 3 highlighted the anomaly in the array n. 5 (spread of the mean equal to −1.79%). While the first two analyses have not revealed that the spreads were too critical (*p*-value was highest), the yearly based analysis has. Evidently, the anomaly was initially latent, and it has become detectable over time. This shows the effectiveness of the cumulative application of the proposed procedure. Again, array n. 1 is always the best performing, whereas the other ones have oscillating behavior. These considerations are also confirmed by the spreads of the median that assigns always the maximum positive value to the array n. 1 and the maximum negative value to the array n. 5. Figure 4 reports the probability density function computed considering a Pearson distribution in order to take into account all the four moments (mean, variance, skewness, and kurtosis) for the three investigated periods: 1 month, 3 months, and 12 months. Two main observations are possible. Firstly, all the curves in Figure 4a,b (monthly and quarterly, respectively) are spread out on the right, whereas all of the curves in Figure 4c (yearly) are spread out on the left. Therefore, although the environment conditions vary during the year, but equally affect the arrays of the PV plant, the distribution functions (see Figure 4a–c) of all of the arrays are very similar for each investigated period, highlighting that no fault or high-intensity anomaly has happened in that year. Secondly, the curves in Figure 4a,b are quite close each other, thus highlighting the similarity among the arrays from the energy point of view. Instead, the curves in Figure 4c are more spaced each other, highlighting that the energy performances of the arrays in the whole year are different, even if they were similar for the first three months. Then, low-intensity anomaly is surely present and the previous numerical analysis has also allowed for locating it. In conclusion, the diagram of Pearson distribution is useful to get preliminary and qualitative information on the several arrays if their distributions are unimodal, while the procedure of Figure 1 allows for identifying the modality of each distribution, quantifying some statistics for each array, and comparing each other the energy performance, such that possible low-intensity anomalies are detected and located.
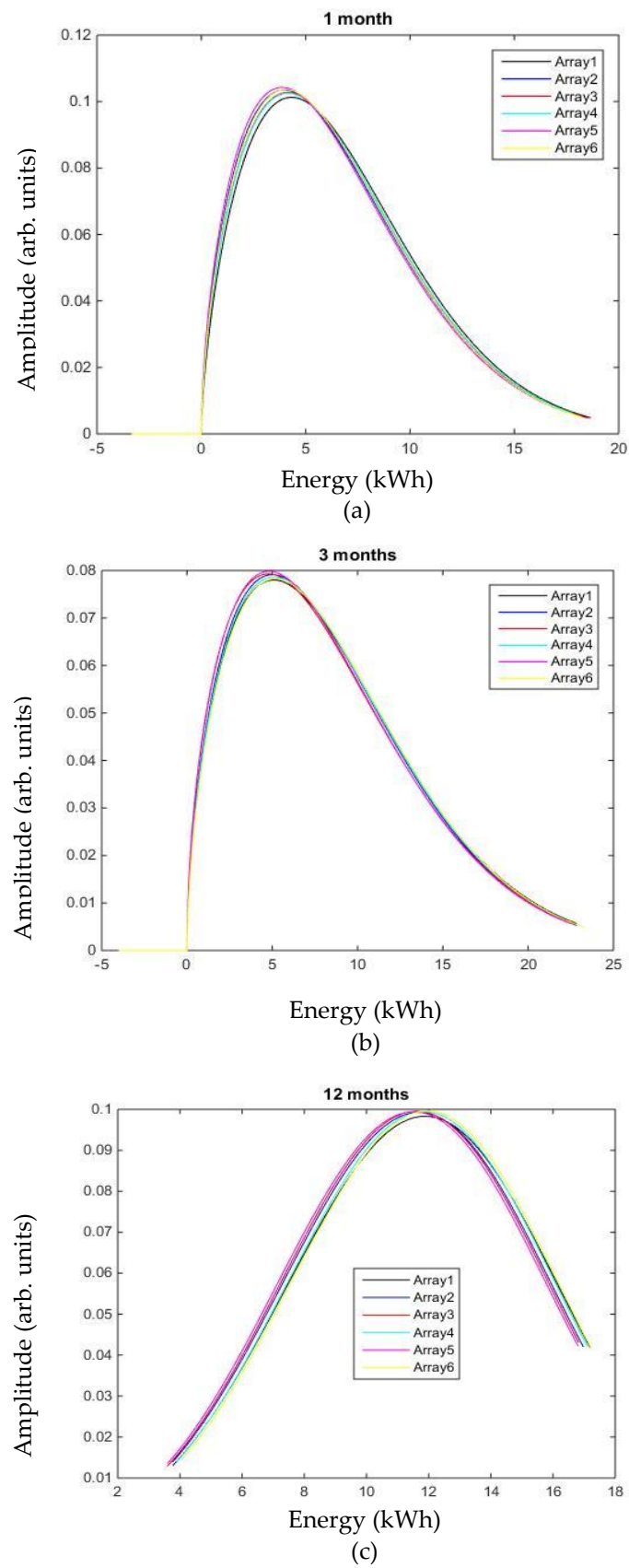
(a)

(b)

(c)

Figure 4. Pearson probability density functions for different periods: (**a**) monthly; (**b**) quarterly; (**c**) yearly.

5. Conclusions

The paper proposes a procedure for the statistical analysis of the energy performance of PV plants, in order to detect the presence of low-intensity anomalies. The procedure is cumulative and some statistical benchmarks are calculated as other measures are done. Experimental results on the yearly data acquired by the datalogger of a real PV plant are shown for three analyses, with the aim to explain point-to-point the iterative procedure for a complete year; nevertheless, the proposed approach is useful also for the real-time monitoring, if rigorous performance benchmarks are fixed. In this way, the trend of the benchmarks can be followed and the anomalies can be detected before they become faults.

If benchmarks exceed prefixed thresholds, alert messages can be sent, or a control procedure can be implemented. Obviously, the number of applications of cumulative analysis for detecting an anomaly depends on its severity. Real case study has shown the effectiveness of the proposed approach. We observed that the cumulative analysis of the whole period (12 months) gives information about the low-intensity anomaly of the array n. 5, as highlighted from the data of Table 3. Therefore, the method shows that a low-intensity anomaly can be identified, although it is not possible to understand the origin of the anomaly. Finally, by storing the statistical parameters of each run of the proposed algorithm, they can be compared year per year (or the same month/trimester of successive years), in order to track the statistical parameter of each single array.

Author Contributions: Silvano Vergura and Mario Carpentieri conceived and designed the experiments; Silvano Vergura performed the experiments; Silvano Vergura and Mario Carpentieri analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare 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. Won, J.-M.; Nam, K.-H.; Kwon, B.-H. Photovoltaic power conditioning system with line connection. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1048–1054.
3. Vergura, S. A Complete and Simplified Datasheet-based Model of PV Cells in Variable Environmental Conditions for Circuit Simulation. *Energies* **2016**, *9*, 326. [CrossRef]
4. Koizumi, H.; Mizuno, T.; Kaito, T.; Noda, Y.; Goshima, N.; Kawasaki, M.; Nagasaka, K.; Kurokawa, K. A novel micro controller for grid-connected photovoltaic systems. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1889–1897. [CrossRef]
5. Xiao, W.; Dundorf, W.G.; Palmer, P.R. Antoine Capel, Regulation of photovoltaic voltage. *IEEE Trans. Ind. Electron.* **2007**, *54*, 1365–1374. [CrossRef]
6. 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]
7. 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, Roma, Italy, 10–13 June 2015.
8. Xiao, W.; Lind, M.G.J.; Dunford, W.G. Antoine Capel, Real-time identification of optimal operating points in photovoltaic power systems. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1017–1026. [CrossRef]
9. 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, Taormina, Italy, 16–18 June 2015.
10. Boulanger, P.; Malbranche, P. Photovoltaic system performance statistical analysis. In Proceedings of the 3rd World Conference on Photovoltaic Energy Conversion, Osaka, Japan, 11–18 May 2003.
11. Takashima, T.; Koyanagi, T.; Otani, K.; Kato, K. Estimation system of in-plane irradiance by regression functions of in-plane/horizontal irradiance ratio vs. Time. In Proceedings of the Photovoltaic Specialists Conference, Lake Buena Vista, FL, USA, 3–7 January 2005.

12.　Paul, D.; Mukherjee, D.; Chaudhuri, S.B. Assessing solar PV behavior under varying environmental conditions—A statistical approach. In Proceedings of the 4th International Conference on Electrical and Computer Engineering ICECE, Dhaka, Bangladesh, 19–21 December 2006.

13.　McSharry, P.E. Assessing Photovoltaic Performance Using Local Linear Quantile Regression. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.146.2683&rep=rep1&type=pdf (accessed on 23 December 2017).

14.　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.

15.　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]

16.　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]

17.　Peloso, M.; 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]

18.　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]

19.　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]

20.　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]

21.　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]

22.　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, Roma, Italy, 10–13 June 2015.

23.　Vergura, S.; Acciani, G.; Amoruso, V.; Patrono, G.; Vacca, F. Descriptive and Inferential Statistics for Monitoring the Operation of PV Plants. *IEEE Trans. Ind. Electron.* **2009**, *56*, 4456–4464. [CrossRef]

24.　Hogg, R.V.; Ledolter, J. *Engineering Statistics*; MacMillan: Basingstoke, UK, 1987.

25.　Rohatgi, V.K.; Szekely, G.J. Sharp inequalities between skewness and kurtosis. *Stat. Probab. Lett.* **1989**, *8*, 297–299. [CrossRef]

26.　Klaassen, C.A.J.; Mokveld, P.J.; van Es, B. Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Stat. Prob. Lett.* **2000**, *50*, 131–135. [CrossRef]

27.　Gibbons, J.D. *Nonparametric Statistical Inference*, 2nd ed.; M. Dekker: New York, NY, USA, 1985.

28.　Hollander, M.; Wolfe, D.A. *Nonparametric Statistical Methods*; Wiley: Hoboken, NJ, USA, 1973.

29.　Chambers, J.M.; Cleveland, W.S.; Kleiner, B.; Tukey, B. *Graphical Methods for Data Analysis*; Wadsworth: Belmont, CA, USA, 1983.

30.　Heinrich, J. *A guide to the Pearson type IV distribution*; Technical Report; University of Pennsylvania: Philadelphia, PA, USA, 2004.