20th International Conference on Renewable Energies and Power Quality (ICREPQ'22)
Vigo (Spain), 27th to 29th July 2022
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.20, September 2022

Supervision of the energy performance of multi-arrays PV plants by means of bi-monthly violin plots

S. Vergura

Department of Electrical and information Engineering
Polytechnic University of Bari
st. E. Orabona, 4, 70125, Bari, ITALY
Phone:+0039 080 5963590, e-mail: silvano.vergura@poliba.it

Abstract. Small PV plants are frequently not equipped with a meteorological station, that is decisive to monitor the energy performance of a PV plant and to detect low-intensity anomalies, as premature ageing, before they become faults. The paper proposes a methodology for PV plant not equipped with a weather station and it is based on a violin plot, that is a statistical tool that integrate information from box plot and histogram. In absence of environment parameters, the monitoring is done, by comparing each other the energy performance of the arrays constituting the PV plant, that should produce the same environment condition affects all the arrays. The methodology is applied to a real 20kWp PV plant, studying the energy performance for different energy datasets.

Key words. kernel density estimation, photovoltaic plant, violin plot, box plot

1. Introduction

The energy produced by a photovoltaic (PV) plant depends mainly on the solar radiation and the air temperature, which affect the maximum power point. Since the PV module captures the solar radiation and produces the electrical energy, it is a very important component of a PV plant; so, its electrical behaviour is always directly or indirectly monitored [1]. Defected modules can be also detected by not destructive techniques [2-7] that can utilize unmanned aerial vehicles [8]. Other strategies can be used to monitor the energy performance of the entire PV plant; for example, some researchers propose methods based on statistics [9] or signal processing [10]. On the other hand, low-intensity anomalies produce small variations in the electrical variables, and their detection is complex. Small PV plants constituted by more than one array are said multi-arrays [9] and are usually equipped with a basic monitoring system that does not acquire the environment parameters. In this case, to evaluate the energy performance of the PV plant is not trivial and Photovoltaic Geographical Information System (PVGIS) [11] can be a useful tool for a rough, but insufficient, preliminary analysis (Fig. 1). In these cases, this paper proposes the use of a synthetic graphic representation, known as violin plot, which is based on several statistical tools and can monitor the operation of a multi-arrays PV system not equipped with a weather station. In fact, even if the measured data depend on the environmental conditions, the proposed supervision strategy does not depend on them, because it is based on the comparison among the energy values of the PV arrays.

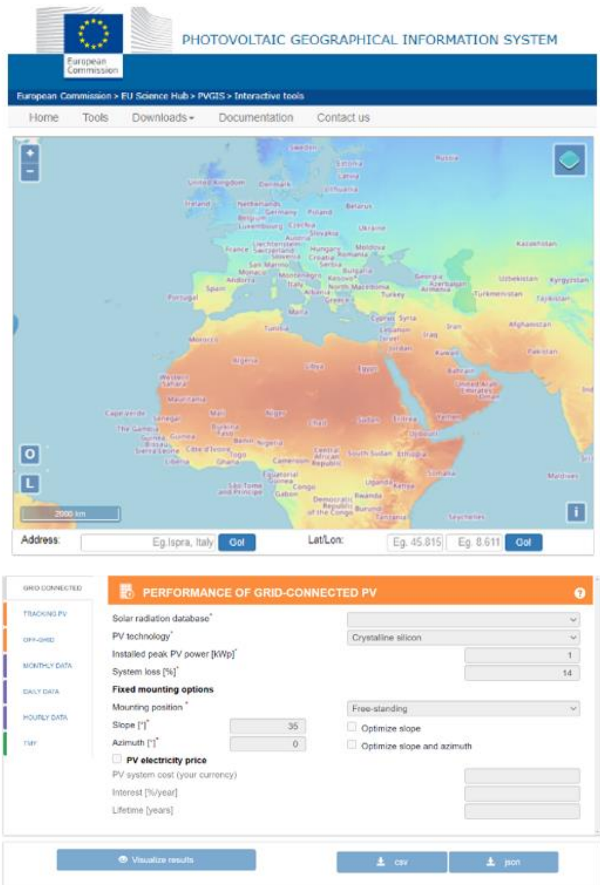

Fig. 1. Photovoltaic Geographical Information System (PVGIS).

2. Methodology

The violin graph [12] shows the distribution of the numerical data and allows comparing the distributions between multiple groups. Violin plot is the combination of a box plot and a Kernel Density Estimation (KDE). For a better understanding of the usefulness of a violin plot, box plot and KDE are separately studied. Fig. 2 reports some characteristic examples of violin plots, based on different distributions. For example, the fourth diagram is the violin plot of a bimodal distribution, as evidenced by the two peaks. The internal part is just the box plot. In fact, the thick black bar into each violin graph represents the interquartile range, i.e., from 25th to 75th percentiles, while the thin black line represents the rest of the distribution, and the white dot is the median. Extreme thin lines are outliers. Each example contains these values. Therefore, the violin graph contains information of a box plot inside, and further information on the edge.

In fact, the distance between the curve and the vertical axis provides information about the data points contributing to a certain parameter value, even if it is not the histogram, but the KDE. To better understand the difference, Fig. 3 reports a histogram and the associated superimposed KDE, which is a smoothed depending on the values of the data point of the histogram. In detail, KDE is a curve with a small area around each data point. The kernel function defines the shape and the width of the area, centered or not on the data point; the width is said bandwidth of the kernel. Therefore, the KDE is based on each data point but is characterized by a chosen kernel function, due to the shape and width. By changing the kernel function, different curves superimposed on histogram can be carried out, even if the histogram is unchanged (Fig. 3). The greater the bandwidth, the smoother the curve.

From a mathematical point of view, let the series $\{x_1, x_2, ..., x_n\}$ be n observations of a population X with an unknown probability distribution function f(x). Kernel estimate of the original f(x) associates each data point x_i a function $K(x,x_i,h)$ called Kernel Function (KF), such that [13]:

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) \qquad (1)$$

being h the bandwidth, and $K(x,x_i,h)$ a bounded function for all real x, i.e.:

$$0 \leq K(x, x_i, h) < \infty \qquad (2)$$

Some common KFs are the following:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \qquad (3)$$

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}} \cdot \left(1 - \frac{x^2}{5}\right) & for\ |x| < \sqrt{5} \\ 0 & otherwise \end{cases} \qquad (4)$$

being (3) the Gaussian KF and (4) the Epanechnikov KF.

For the case study in this paper, both of them were used, and the results were similar. So, only the diagrams based on Gaussian KF are reported.

If the arrays constituting the PV plants are identical for typology and installation conditions, they must produce the same energy in the same period, because partial random shadings offset each other. In fact, environmental condition affects equally the arrays during a medium period, being the arrays close. For this reason, by applying the violin plot to the energy datasets of all the arrays, the same results should appear, if no anomaly is present, i.e., all the violin plots must be almost identical. When this does not occur, a fault or an anomaly is already present in the PV plant. Consequently, by comparing the violin plots of the energies produced by the strings constituting the PV plant, it is possible to monitor the energy performance of the PV plant and to detect the presence of anomalies. The strategy will be applied to the energy datasets of a real PV plant, described in the next section.

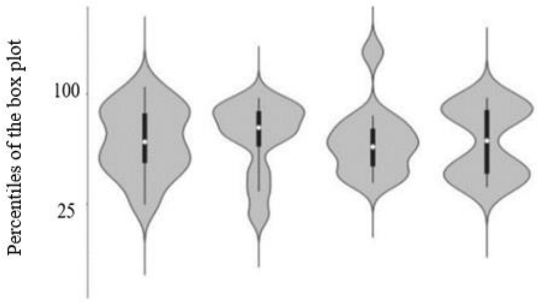

Fig. 2. Examples of violin plot: symmetric, skewed, with outlier, and bimodal.

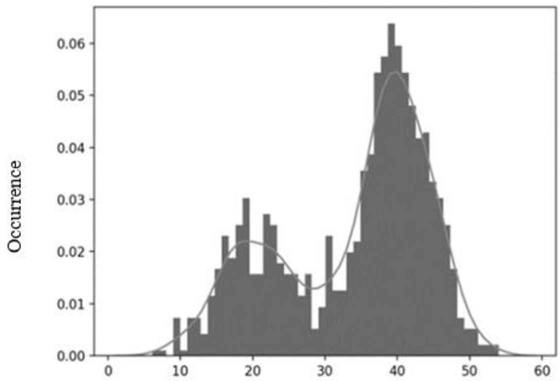

Fig. 3. Hystogram with a KDE.

3. Case study

The methodology was applied to the energy datasets a real 20kWp PV plant, constituted by six strings. It is in the south of Italy. Table I and II report the parameters from the manufacturer datasheets of the PV module and inverter, respectively. A complete description of the whole PV plant is available in [9]. For our scope, it is important to know that the datalogger stores six values for hour, therefore 144 samples for day and for array are available.

TABLE I. SPECIFICATIONS OF THE PV PLANT

Total number of modules	132
Type and manifacturer	Solterra 150W
Cells	monocrystalline
Max power [W]	150
Min. power [W]	140
Voltage at typical power [V]	34.7
Current at typical power [A]	4.32

Open circuit voltage [V]	43.2
Short circuit current [A]	4.7
NOCT [°C]	48° ± 1
Warranted Module efficiency (%)	11.80
Temperature coefficient voltage β	-153 mV/°C
Temperature coefficient current α	+0.90 mA/°C
Temperature coefficient power γ	-0.40 % /°C

The preliminary analysis based on the whole yearly energy dataset (Fig. 4) allows to understand if great criticalities are present. It results that the shapes are similar, and the median values (white dots) are almost identical. Therefore, great criticalities are not detected, even if the maximum values are different. Particularly, arrays #1, #3, and #6 show the maximum values, whereas the other strings have lower values. However, the differences are not excessive, and they are due to singular situations, not structural ones, as evidenced from the similarity between the shapes. To investigate in depth, a further analysis based on by-monthly energy datasets was carried out, and results are reported in Fig. 5.

TABLE II. SPECIFICATIONS OF THE INVERTER

MPP voltage range[V]	268 – 550
Max input voltage [V]	600
PV power [W]	4100
Max input current [A]	12.2
Nominal output power [W]	2600
Max. output power [W]	3000
Maximum efficiency	95%
Nominal mains voltage/frequency	230V / 50Hz
Total harmonic distortion	< 4.0%
Power factor	1

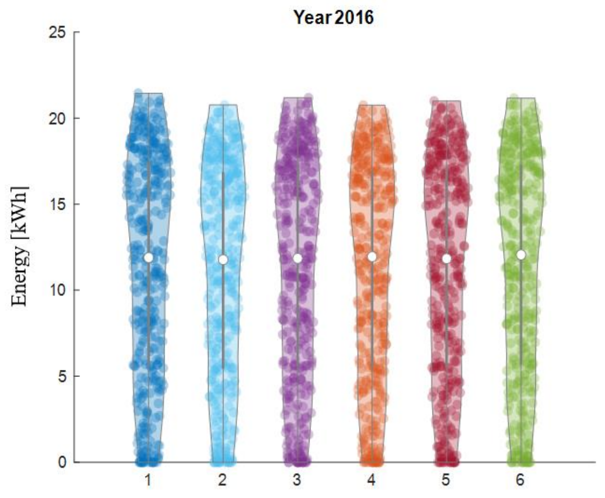

Fig. 4. Yearly violin graphs of the energy produced by the 6 PV arrays.

Each line of Fig. 5 reports the violin plots of the six arrays constituting the PV plant under test. These diagrams depend on the bi-monthly energy datasets; then, there are six lines, each of them representing the energy behaviors of the six arrays. Instead, each column describes the variable energy behavior of each array during the whole year. By observing the violin plots of each period, they result almost equal, highlighting the PV arrays always produce the same energy. This is a confirmation that no anomaly is present, considering that the strings, equally constituted from the component point of view, are under equal environment conditions. Moreover, by comparing the periods each other, you can observe the large variability of the violin plots during the year. In fact, the periods March-April and November-December show almost uniform distributions, highlighting that the values from zero to the maximum one are equally frequent. The first period January-February is characterized over of all by low values. The periods May-June and, over of all, July-August are characterized by very high values, highlighting that the maximum values of the produced energy are largely more frequent than the others. The shape of the diagrams in May-June remembers exactly the violin shape, while the violins in July-August are characterized by a bimodal distribution. By seeing the diagram of a fixed column, it is easy to observe the variability of the energy behavior from January to December, but the extreme distributions (January and December) are similar. Thus, the diagrams of a year can be usefully utilized as benchmark for the next year because they show a cyclic behavior. This hypothesis is confirmed for the energy datasets of the successive year of the PV plant under investigation. After this analysis, we can affirm that no difference in energy performance of the PV strings results in any analyzed period and all the PV arrays showed the same electrical behavior.

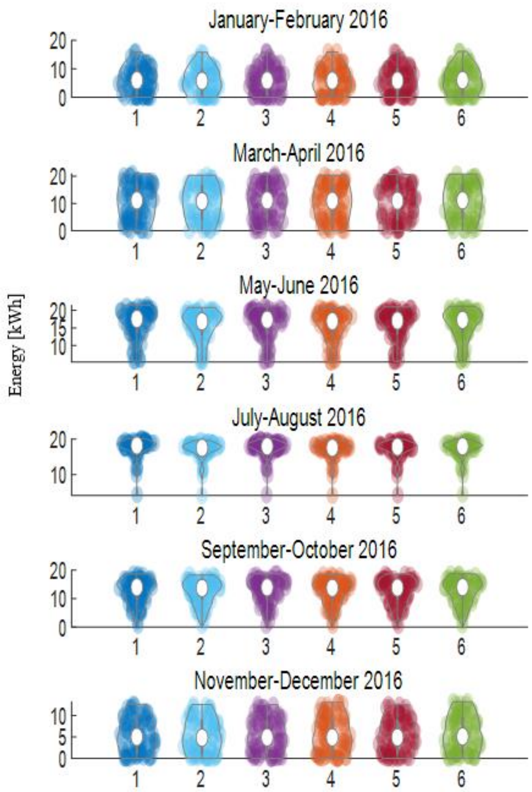

Fig. 5. Bi-monthly violin graphs of the energy produced by the 6 PV arrays. Each diagram represents energy in kWh.

The proposed approach allows to supervise the energy performance for two months, as shown, or for a wider period. Other statistical tools, as the simple box plots or the Bollinger bands [14], do not allow to get detaild information of the distributions, even if they are effective in defining the thresholds for a correct operation of the PV arrays. For analogy, methods usually applied to the electrical signals deriving from electrical systems [15-16] are not useful when the input data are unknown, as it occurs for PV plants not equipped by sensors to measure the environmental parameters.

4. Conclusion

The violin graph-based strategy to supervise the behaviour of multi-array PV plants is useful in absence of environmental data. Several kernel functions with different values of bandwidth will be applied to the energy datasets in order to evaluate the different results among them, and to check if any anomaly is detected by using only specific kernel function and not in the other cases.
The proposed methodology, based on both histograms and box plots, is effective in supervising the energy behaviour of multi-string PV plants, when the environmental data is not available. The procedure is applied to a real 20 kWp six-arrays PV plant for six bi-monthly periods. No anomaly was detected, comparing the diagrams of the six arrays within each period under investigation.
The proposed methodology was also applied by the authors to the daily energy datasets, but they are too limited to return reliable information. A simple cloud or temporary shadow on a string is detected as anomaly. For this reason, they diagrams are not reported in this paper, and the authors advise against using the proposed methodology on daily values.

References

[1] P. Guerriero, F. di Napoli, G. Vallone, V. D'Alessandro, S. Daliento, "Monitoring and diagnostics of PV plants by a wireless self-powered sensor for individual panels", IEEE J. Photovoltaics 2015, 6, 286–294.

[2] T. Takashima, J. Yamaguchi, K. Otani, K. Kato, M. Ishida, "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.

[3] O. Breitenstein, J.P. Rakotoniaina, M.H. Al Rifai, "Quantitative evaluation of shunts in solar cells by lock-in thermography", Prog. Photovolt. Res. Appl. 2003, 11, 515–526.

[4] S. Vergura, F. Marino, M. Carpentieri, "Processing Infrared Image of PV Modules for Defects Classification", International Conference on Renewable Energy Research and Applications, ICRERA 2015, 22–25/11/2015; Palermo, Italy, 2015, ISBN: 978-1-4799-9981-1.

[5] S. Johnston, H. Guthrey, F. Yan, K. Zaunbrecher, M. Al-Jassim, P. Rakotoniaina, M. Kaes, "Correlating multicrystalline silicon defect types using photoluminescence, defect-band emission, and lock-in thermography imaging techniques", IEEE Journal of Photovoltaics 2014, 4, 348–354.

[6] M.P. Peloso, L. Meng, C.S. Bhatia, "Combined thermography and luminescence imaging to characterize the spatial performance of multicrystalline Si wafer solar cells", IEEE Journal of Photovoltaics 2015, 5, 102–111.

[7] S. Vergura, M. Colaprico, M.F. de Ruvo, F. Marino, "A quantitative and computer aided thermography-based diagnostics for PV devices: Part II—platform and results", IEEE J. Photovoltaics 2017, 7, 237–243.

[8] S. Vergura, "Correct Settings of a Joint Unmanned Aerial Vehicle and Thermocamera System for the Detection of Faulty Photovoltaic Modules", IEEE Journal of Photovoltaics, Vol. 11, Issue 1, January 2021, pp 124-130.

[9] S. Vergura, "Hypothesis tests-based analysis for anomaly detection in photovoltaic systems in the absence of environmental parameters", Energies, 2018, 11, 485.

[10] I.-S. Kim, "Online fault detection algorithm of a photovoltaic system using wavelet transform", Solar Energy, 2016, 226, 137–145.

[11] http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html (accessed on March 19, 2022).

[12] J. M. Chambers. Graphical methods for data analysis. Chapman and Hall/CRC, 2017.

[13] B.W. Silverman, Density estimation for statistics and data analysis, Chapman and Hall, London, 1986.

[14] S. Vergura, "Bollinger bands based on exponential moving average for statistical monitoring of multi-array photovoltaic systems", Energies, 2020, 13, 3992.

[15] V. Puliafito, S. Vergura, M. Carpentieri, Fourier, Wavelet and Hilbert-Huang Transforms for Studying Electrical Users in the Time and Frequency Domain, Energies, 2017, 10, 188, pp. 1-14.

[16] R. Zivieri, S. Vergura, M. Carpentieri, Analytical and numerical solution to the nonlinear cubic Duffing equation: an application to electrical signal analysis of distribution lines, Applied Mathematical Modelling, Elsevier, Volume 40, Issues 21–22, November 2016, Pages 9152–9164.