Method to characterize variability of photovoltaics power output

Laurène PARENT
CEA-INES – France
laurene.parent@cea.fr

Delphine RIU
Univ. Grenoble Alpes, Grenoble INP, G2Elab – France
delphine.riu@g2elab.grenoble-inp.fr

Tuan Quoc TRAN
CEA-INES - France
quoctuan-tran@cea.fr

ABSTRACT

Effect of variability of renewables connected to electrical grids becomes a critical topic in the actual process of increasing the rate of green power. Some issues result from it, the grid operator has to identify them by carefully study the behaviour of these fluctuations. The output power was collected from four photovoltaics plants located in Martinique. A new methodology is developed to express simply the statistic behaviour of this power fluctuations. Two function are used to fit the density probability. A day classification is also performed. At the end this methodology allows to fit with a R^2 of 99% the real probability density. The result can be used in statistical study of the power grid with a good accuracy.

INTRODUCTION

The integration of renewable energy sources in the power system raises many concerns about the safety and reliability of the grid. Indeed grids with a high penetration of photovoltaics (PV) will be characterized by a lower inertia and short-circuit power than conventional grid. It will be greatly common for over-seas and tropical islands in the future. The lack of inertia makes the grid more vulnerable and sensitive to frequency fluctuations. A greater attention must be carried on the stability analysis of these grids in order to develop these structures in a confident way for users and operators. Thus, an accurate study of the variability by the operators is necessary. This work proposes a study of historical data of photovoltaic power plants to find an accurate model for the intermittency to design technical solution to face risks of frequency instability. This model can be implemented into a simulation tool to accurately simulate the stochastic behavior of the grid with variable power input and to investigate possible problems and weaknesses for system with high penetration of renewables.
To quantify the variability one of the most common metric is the Variability Index defined in [1]. It is given for the irradiance but for this study the output electrical power is available thanks to dedicated sensor investment. Thus to quantify their variation a more adapted index defined in [2] is used in this study in the same manner as for irradiance. A common way to study the output power variation consists in calculating the ramp variation which is the variation of power divided by the corresponding time duration as seen in [3].
Different methods are used to classify the variability of the data. Even though these methods are used to model the solar irradiance, they can be still adapted to the power output variability. In [1] four day types are classified according to the behavior of the irradiance from the variability index and the daily clear sky index. The clear-sky index is given as the ratio of the irradiance and the clear sky irradiance. These classes are defined as clear day, overcast day, mixed day and highly variable day. In [4] they used the daily probability of persistence and the clear sky index to classify days in ten classes. In [5] they use the clustering technics and obtain also ten classes. The work of this present paper use the same process but with the classification according to the variability of the power output.
Authors in [6] use the probability density and the cumulative distribution of the power variation to analyze the variability into the plant. This article shows that the probability density of the variation of the power is not a normal distribution. Thus a main is to identify the distribution, of the variation in order to develop reliability studies of grid stability against PV variation. In [7] such a distribution of the irradiance fluctuation is presented as an exponential function. [8] demonstrates also the non-Gaussian behavior of increments of wind and solar power. The probability density is similar to a q-exponential function. Authors present result for sampling time range from seconds to minutes.
First the data used in this paper will be presented. Then the methodology which used different aspect described below will be explained. In this part two probability density will be fully expressed, the classification process will also be showed. Then the main results will demonstrate a good accuracy for the methodology for those data and the conclusion will invite a reflection on the scope of this work.

DATA

Power output measures from photovoltaic (PV) power plants are collected in Martinique. The data were gathered in this French oversea department during almost two years,

Figure 1 Map of Martinique with the four PV power plants

in 2013 and 2014. A project with those data can be seen in [9] to deeper understand the context. Measures in each of the four power plants, called Prêcheur, Sainte-Marie, Ducos and Diamant, are done at the same time scale every one minute. The map in Figure 1 gives the approximate place and the nominal capacity of these PV plants. The power used here is the addition of each PV power output at the same time.

METHODOLOGY

Variables built from the data

The difference of this power between all adjacent times is then calculated. This value is divided by the total nominal capacity of the four PV plants. Thus, this value ΔP represents the percent rise and fall of PV power each time t for one minute time step.

Indices

In this study, two indices were chosen to classify the day, the energy index (EI) and the variability index (VI). The former index is a common metric shown in [2]. In this paper VI is the standard deviation of ΔP. The novel energy index is defined as the ratio between the energy over one day and the energy provided on an optimal day which is a virtual day composed with the maximal power sampled at each minute all over the years. This index helps to see how far a day of production is from the optimal day.

Days classification

Thereafter the days are classified according to those indices. To adjust the number of cluster to the data and to automate the computation a clustering algorithm is used. The K-means clustering is chosen according to [10] because this algorithm is very popular and is the simplest among partitioning algorithms. This clustering uses the square-Euclidean metric. A criterion, called Davies Bouldin, is added to establish the number of cluster automatically. According to [11] this index is well adapted for data with closed space cluster that is the case of our study. This process was chosen after numerous test and at the end showed to be the best way to obtain the clustering. A lot of criteria could be more adapted for other data. Tests have to be performed to choose the right algorithm and criterion.

Probability representation

The probability histogram of ΔP is shown in Figure 2. Probability that ΔP is between -0.21 and -0.19 is 0.12 %. Most ΔP are near 0, thus probability that ΔP is between -0.001 and -0.0095 is 45 %. Probability can also be represented by probability density. With this representation a function can be found to express with few parameters the statistic behaviour of ΔP. A probability density function is defined as positive, its total integral is

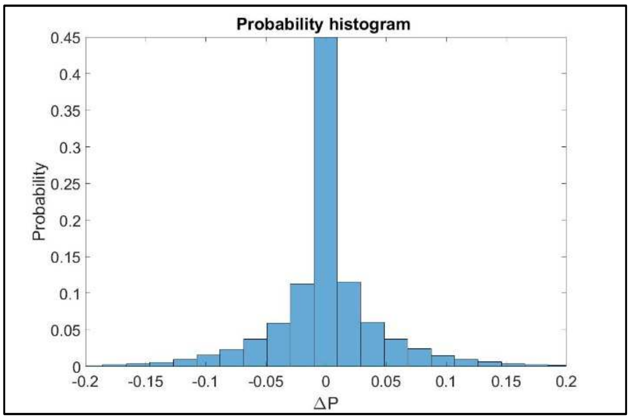

Figure 2 Probability histogram of ΔP

equal to 1, probability that a variable is between a and b is given by the integral of the probability density on [a, b]. The Kernel method [12] is used to have a continue density. ΔP probability density is built with a window of 0.001.

Probability density fitting

Some authors used some classical functions to fit with different probability densities. In [7] a probability density modelling the fluctuation of daily PV power is given by Equation 1.

Equation 1 $F_{\Delta P>u}(x) = be^{-b(x-u)}$

The parameter u denotes a threshold over which the variation of power is taken into account. The data are not considered as symmetric with such a distribution. This density will be called threshold-based density.
Another density law is presented in [8]. This density law is called the q-exponential function. In this work, the law is expanded to positive and negative values in order to be more representative of actual PV fluctuations. As the integral of density must be equal to one, the equation is changed slightly to Equation 2.

Equation 2 $F_q(x) = \frac{1}{2}\beta(2-q)(1-\beta(1-q)|x|)^{\frac{1}{1-q}}$

Q-exponential and threshold-based functions are fitted to the ΔP probability density. For both functions the maximum likelihood estimation allows to get the parameters, q and b for the q-exponential and b for the threshold based function. Then the R-squared coefficient is calculated between the real probability density and each function to evaluate the fitting. For the threshold based function, the threshold chosen is the one for whom the R-squared value is the higher.

RESULTS

First, probability density functions were fitted to ΔP probability density. Then a K-means algorithm classified the days according to the indices EI and VI. Finally the best fitting functions is found for each cluster.

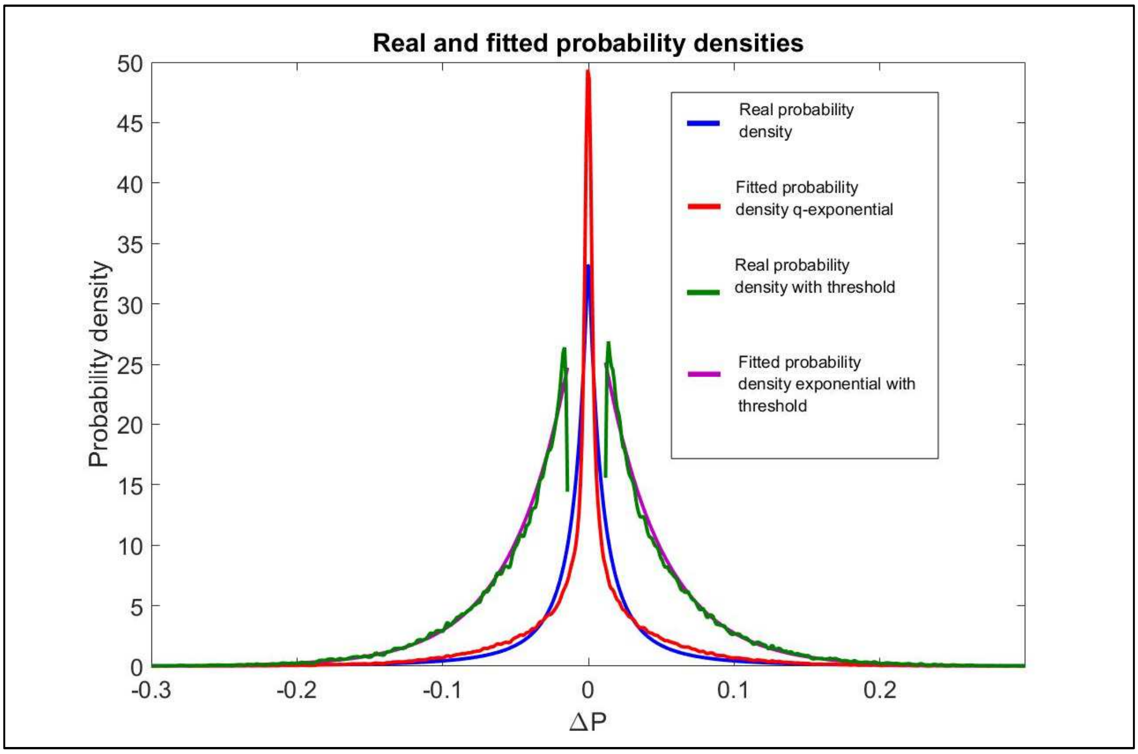

Figure 3 Real probability density and fitted one with the q-exponential and the threshold-based exponential

Fitting with one minute step time

ΔP probability density is fitted by the two functions described earlier, this can be seen on Figure 3. The R^2 values are shown on TABLE I. There is a poor R^2 value of q-exponential fitting whereas the other fitting is suitable.

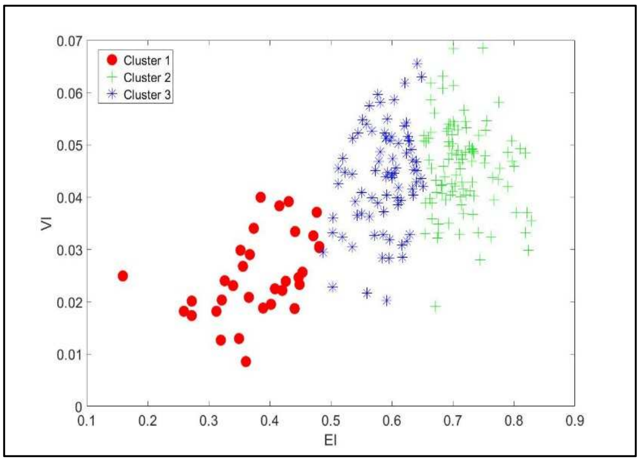

Figure 4 Individual days clustered in three groups according to EI and VI indices for one minute step time.

Classification

The K-means clustering algorithm gives three classes represented on the Figure 4. The clusters are strongly separated by the EI index. Clusters 2 and 3 have a similar VI while cluster 1 shows a smaller variability index.

Power output of three type of day from each cluster are shown on Figure 5. The day of cluster one has low power output and low variability. Both second and last cluster have high power output during their day but cluster 2 seems to have higher variability.

Fitting with cluster and one minute step time

Then for each cluster the fit is realized with both probability density function. The TABLE I sums up the result. With the q-exponential fitting the R^2 values are better for cluster 1 than the other cases. The cluster 2 and 3 have the best fit with the exponential–based function.

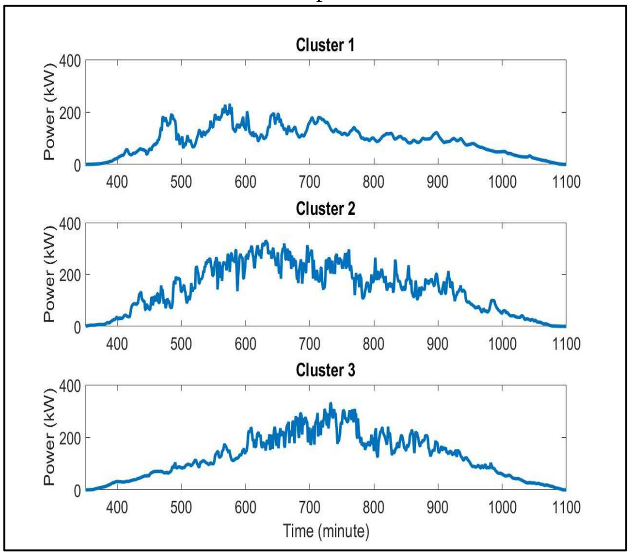

Figure 5 Example for the power output for three typical day for each cluster

TABLE I Comparison between the real probability density and the fitted probability density for three clusters for a step time of one minute

	R^2 values		
	q-exponential	**Threshold positive**	**Threshold negative**
All	90 %	98% u=1.18	98% u=1.43
Cluster 1	***99%***	97% u=1.22	97% u=1.44
Cluster 2	87%	***99%*** u=1.22	***99%*** u=1.12
Cluster 3	88%	***99%*** u=1.05	***99%*** u=1.29

CONCLUSION

With the use of power output from a PV plant this method allows to model the behaviour of the probability density with a good accuracy, with R-squared coefficient of 99 %. To implement this method various tools are used to automate the calculation. Thus the same method can be used for other data. Two functions found in the literature are well suited for the data.

At the end, from the probability density found, random variable can be generated. Those variable corresponding to statistical behaviour of a particular system can be implemented in the study of power grid.

Some questions are still under concern. Indeed, the statistical behaviour of a PV system depends of the size, capacity of the power plant and the time step of measure as noticed in [7]. If a relationship could be inferred between all those parameters and the parameters of the density probability found future system could be expanded with an idea of the future statistic behaviour. And some limitation on the grid could be prevented. Thus this is not for the moment possible.

REFERENCES

[1] J. S. Stein, C. W. Hansen, and M. J. Reno, 2012, "The variability index: A new and novel metric for quantifying irradiance and PV output variability", *World Renewable Energy Forum*, pp. 13–17.

[2] Murata Akinobu, Yamaguchi Hiroshi, and Otani Kenji, 2008, "A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area," *Electr. Eng. Jpn.*, vol. 166, no. 4, pp. 9–19.

[3] B. M. Mazumdar, M. Saquib, and A. K. Das, 2014, "An empirical model for ramp analysis of utility-scale solar PV power," *Sol. Energy*, vol. 107, pp. 44–49.

[4] B. O. Kang and K.-S. Tam, 2013, "A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data," *Sol. Energy*, vol. 94, pp. 102–118.

[5] S. Moreno-Tejera, M. A. Silva-Pérez, L. Ramírez-Santigosa, and I. Lillo-Bravo, 2017, "Classification of days according to DNI profiles using clustering techniques," *Sol. Energy*, vol. 146, pp. 319–333.

[6] X. Xu, Z. Yan, M. Shahidehpour, H. Wang, and S. Chen, 2018, "Power System Voltage Stability Evaluation Considering Renewable Energy With Correlated Variabilities," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3236–3245.

[7] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, 2011 ,"Power output fluctuations in large scale pv plants: One year observations with one second resolution and a derived analytic model," *Prog. Photovolt. Res. Appl.*, vol. 19, no. 2, pp. 218–227.

[8] M. Anvari *et al.*, 2016 "Short term fluctuations of wind and solar power systems," *New J. Phys.*, vol. 18, no. 6, p. 063027.

[9] T. Delaplagne *et al.*, "OPERATION OF THE HIGH TEMPERATURE NANICL2 BATTERIES STORAGE SYSTEM FOR MANAGEMENT OF PHOTOVOLTAIC PRODUCTION.".

[10] A. K. Jain, 2010 ,"Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666.

[11] A. Vij and P. Khandnor, 2016 ,"Validity of internal cluster indices," in *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, pp. 388–395.

[12] S. J. Sheather and M. C. Jones, 1991 ,"A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 53, no. 3, pp. 683–690.