Modeling of the nominal power of a PV generator under clear and cloudy sky conditions

José R Angulo[1,2], Brando X Calsi[1], Luis A Conde[1], Emilio Muñoz-Cerón[2], Rolf Grieseler[1], Jorge A Guerra[1], Juan de la Casa[2] and Jan A Töfflinger[1]

[1]Departamento de Ciencias, Sección de Física, Pontificia Universidad Católica del Perú, Av. Universitaria 1801 San Miguel Lima 32, Perú

[2] IDEA Research Group (Research and Development in Solar Energy), Centre for Advanced Studies in Energy and Environment (CEAEMA), Electronics and Automation Engineering Department, University of Jaén, Campus Las Lagunillas s/n 23071, Jaén, Spain

Abstract

The characterization of the nominal power under real operating conditions is crucial for the correct evaluation of a photovoltaic generator. Several earlier studies proposed different methods based on empirical models for the estimation of the nominal power. These methods require experimental data obtained during optimal days under clear sky conditions and are not suitable for days deviating from these optimal conditions and, thus, generating a significant amount of noise in the data. In this sense, we propose a non-parametric statistical approach to filter out this noise to reliably estimate the real nominal power in the latter conditions. The period of study was 107 days. These were divided in two categories, clear and partly cloudy sky conditions. The results show that our statistical method allows to obtain the same nominal power under partly cloudy conditions as under clear sky. This was possible by applying a kernel density estimation to filter the outliers and the noise under cloudy conditions.

Keywords: PV characterization, real PV power, testing in outdoor condition

1. Introduction

The costs of the installation of solar photovoltaic (PV) systems decreased up to 90 percent over the last years as the modules can easily be produced in a mass production process nowadays (Welter, 2019). Therefore, the production of electricity, using PV systems, has become cheaper than using gas-feed power plants and puts it at same level of black coal-feed power plants. The worldwide PV capacity installed at the end of 2018 was around 500 gigawatts (Watson and Schmela, 2018). The forecast expects this capacity to double within the next three years. In this current scenario, the knowledge of the optimal state of the photovoltaic system at standard test conditions, STC (irradiance = 1000 W/m^2; AM1.5 spectrum; cell temperature=25°C) is important for investors; since estimates of energetic projections are based on parameters extrapolated to STC (Markus Schweiger et al. 2017) which enable the analysis of possible failures in the PV generator (Talavera et al., 2019).

The nominal power (P_{M}^{*}) is one of the most relevant parameters of the PV generator (Kumar et al. 2017). It can be defined as equivalent to the maximum power point under STC. The meaning of this parameter is that, on the one hand, it plays a key role in estimating energy yield, which is necessary for commercial projects, and, on the other hand, represents value for purchase-sale transactions (Talavera et al. 2016).

At the module level this parameter is provided by the manufacturer's datasheet and certificate which is associated with quality assurance procedures. After installation, this given parameter value may differ from the actual P_{M}^{*} value; due to initial degradation after being exposed to light and temperature (Marcus Schweiger et al. 2017). Under outdoor conditions, the acquisition of the actual nominal power requires measurements of current-voltage (I-V) curves and extrapolation to STC in accordance the standard IEC 60891 (Reise et al. 2018).

At the PV system level and particularly for large PV plants, it is still an open question on how to reliably calculate P_{M}^{*} (de la Parra et al., 2017) taking into account the different system losses such as cabling, module mismatch, temperature, angular losses, soiling, among others. There are different methods described in the literature on how to obtain this value. As a first approach, one can take a few modules randomly and check the state at STC with a solar simulator. However, this value would not indicate or represent the real nominal power of the entire system. The more practical option in this case is to measure the P_{M}^{*} in-situ under outdoor conditions using the current and voltage (*I*-*V*) curve or by measuring the maximum power point (MPP). In principle, the MPP models consider the

doi:10.18086/swc.2019.14.01 Available at http://proceedings.ises.org

information available from the PV module datasheet. The MPP modeling has been applied more commonly than an I-V curve measurement because of the lack of commercial devices capable to measure the currents generated by the PV system (Muñoz et al. 2016).

To calculate the P_{M}^{*}, according to well-defined procedures by (Martínez-Moreno et al., 2012), it requires optimal testing condition or ideal environmental conditions to monitor irradiance and module temperature. This means sunny and clear sky conditions. This is established to avoid outliers in the operating conditions, which will introduce errors in the data processing. Notwithstanding, in several locations, where most of the time partly cloudy sky conditions are present, such optimal testing conditions may be seldomly given. When applying the methods proposed in the prior works, the data obtained under non-optimal conditions commonly cannot be considered due to the introduction of additional errors in the P_{M}^{*} estimation.

This paper aims to calculate the nominal power under partly cloudy day conditions, based on the approach given by (Martínez-Moreno et al., 2012). To achieve this, we introduce a statistical filtering procedure by applying a non-parametric approach in order to filter out noise and outliers.

2. Experimental details

The data in this paper was collected through outdoor measurements of a PV generator in operation located at Granada, Spain, which has been subject to prior other studies (Lomas et al., 2018; Muñoz-Cerón et al., 2018). The supposed nominal output power of this PV plant is 109.4 kW under STC according to the datasheet of the polycrystalline silicon modules. To assess the amount of solar irradiation (G) received and the module temperature (T_{m}), two calibrated panels of the same technology and in the same angle as the plant's modules were used. This plant was used as a test laboratory for research and outdoor monitoring by the IDEA group (Muñoz-Cerón et al., 2018).

Tab. 1: Configuration and electrical parameters of the photovoltaic generator at STC

Series connected modules per string	18
Strings in parallel	32
Current at the maximum power point (A)	257.6
Voltage at the maximum power point (V)	574.2
Power at maximum power point (kW)	109.4
Power temperature coefficient (%/°C)	-0.43

The MPP tracking efficiency of modern inverters is usually greater than 99%, thus, the DC power at the inverter entry (P_{DC}) can be suitably assumed as the power at the MPP. Hence, P_{DC} was measured with a calibrated wattmeter YOKOGAMA WT1600 (uncertainty of measurement less than 0.5%). Table 1 contains a short description of the PV generator. The measured values (G, T_{m}, P_{DC}) were recorded with a 60 seconds time step. To perform the P_{M}^{*} analysis, we used measurements from March 27 to September 30, 2018.

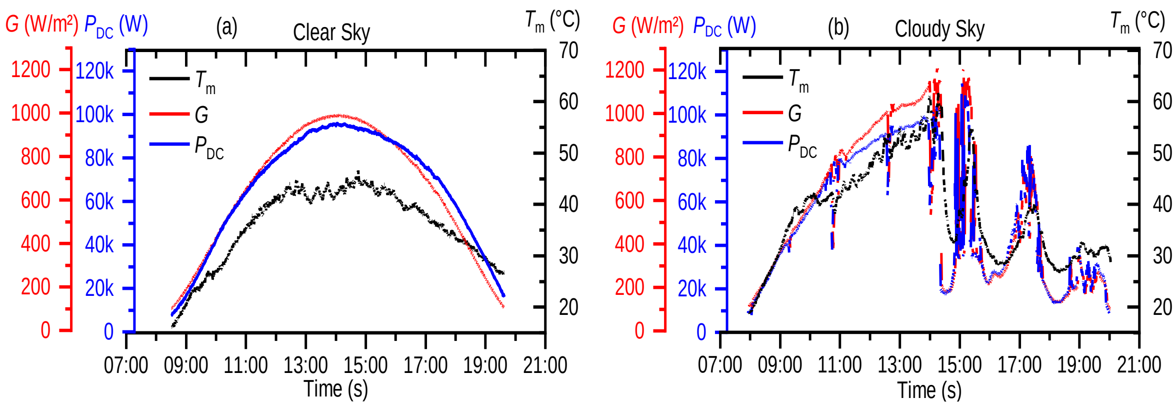

Fig. 1: Irradiance, module temperature and DC Power of two exemplary days under (a) clear sky and (b) cloudy sky conditions.

3. Method to calculate the nominal power

The objective of this work is to calculate the nominal power under cloudy sky conditions. To do this, it is required to divide the data $(G, T_{\mathrm{m}}, P_{\mathrm{DC}})$ into two categories. As a matter of example, the days of each category are represented in Fig. 1. The total number of days of data collected were 107. After applying a filtering procedure which distinguishes between both conditions, we obtained 37 days with clear sky and 90 days with cloudy sky conditions.

3.1. Nominal power estimation under clear sky conditions

The procedure suggested by (Martínez-Moreno et al., 2012) was followed and the P_{M}^{*} was calculated for each day in clear sky conditions.. Experimentally, this procedure requires sampling of data at least every minute during one day of a clear sky and it uses the Osterwald equation to calculate the nominal power, see equation (1). This empirical equation describes the relationship between the maximum power point (P_{DC}) with the module temperature (T_m) and the irradiance incident on the panel plane (G) for irradiances above 800 W/m^2.

$$P_{DC} = P_M^{*} \times \frac{G}{G^{*}}(1 + \gamma(T_{\mathrm{m}} - 25°\mathrm{C})) \qquad \text{(eq. 1)}$$

Here G^{*} is the irradiance under STC and γ is the temperature coefficient provided by the manufacturer module data sheet. The DC power is corrected to 25 °C ($P_{DC} \rightarrow P_{(G,T\rightarrow 25°\mathrm{C})}$), as described by equation 2.

$$P_{(G,T\rightarrow 25°\mathrm{C})} = \frac{P_{(G,T_{\mathrm{m}})}}{(1 + \gamma(T_{\mathrm{m}} - 25°\mathrm{C}))} \qquad \text{(eq. 2)}$$

Second, the set of points $(G, P_{(G,T\rightarrow 25°\mathrm{C})})$ are linear fitted with equation (3) to obtain P_{M}^{*} in an irradiance range of 800 – 1000 W/m^2. Fig. 2 (a) shows the linear fitting from which the nominal power is calculated for an exemplary clear sky day.

$$P_{(G,T\rightarrow 25°\mathrm{C})} = P_M^{*} \times \frac{G}{G^{*}} \qquad \text{(eq. 3)}$$

Fig. 2 (b), also depicts the P_{M}^{*} calculated in this way for the 37 days under clear sky conditions. The average value of P_{M}^{*}=104.13 kW and a dispersion of 1.5% is indicated. The observed dispersion is consistent with the one reported by (Martínez-Moreno et al., 2012). Furthermore, it is plausible to derive the P_{M}^{*} from a statistical point of view using the central limit theorem because there are more than 30 P_{M}^{*} values. The resulting P_{M}^{*} was taken as a reference for the following proposed method for calculating P_{M}^{*} with data of cloudy days.

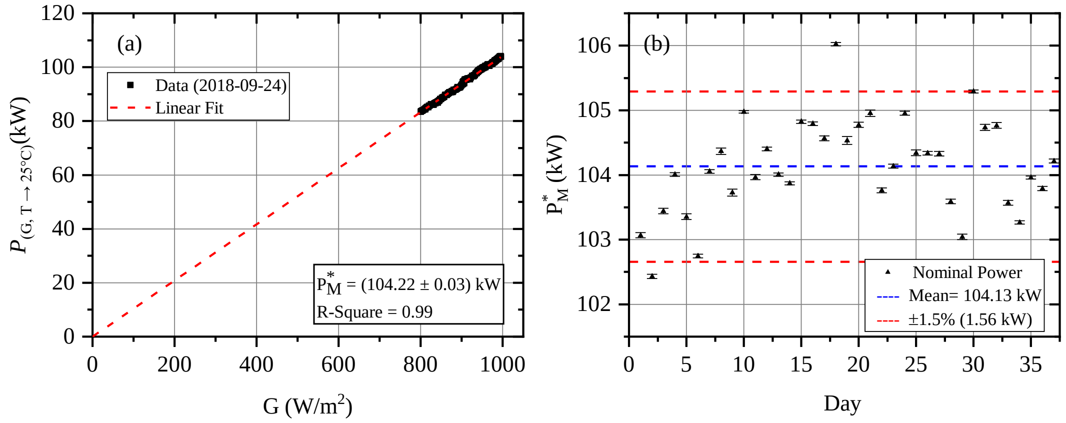

Fig. 2: (a) The linear fitting process for an exemplary day (2018-09-24) with clear sky conditions (b) the nominal power calculated for 37 individual days. These results present a mean of P_{M}^{*} (blue dotted line) and a 1.5% dispersion (red dotted line)

3.2. Nominal power estimation under cloudy sky conditions

The data obtained under cloudy sky conditions shows a considerable amount of noise and outliers, as seen in Fig. 3 (a). When the procedure outlined in section 3.1 is applied, the linear fitting of these data results in a nominal power that considerably deviates from that obtained in clear sky conditions. Therefore, to estimate the P_{M}^{*} in cloudy conditions, from the corresponding set of data, a subset of data needs to be filtered before applying the procedure described in section 3.1. This subset of data should represent the working conditions of the photovoltaic system during moments of clear sky conditions, such as can be seen in Fig. 1 (b) at certain moments even during a cloudy day.

To achieve this, first, the distribution of the measured data during cloudy days was studied. It was found that the data exhibits a non-parametric behavior since it presents registers of agglomerated data. The study of this type of non-parametric data can be done using a Kernel Density Estimation, KDE, (Trashchenkov et al., 2018). In order to visualize this behavior, exemplarily, the probability density or KDE for fixed irradiation of (990 ± 2) W/m^2 was calculated and is depicted in Fig. 3(b). This kind of data distribution has been previously reported in the modeling of photovoltaic generators (Trashchenkov et al., 2018). Furthermore, this KDE has several peaks, the one with the highest value, marked with an arrow, is assumed to best represent the PV generator as if under clear sky conditions. The other minor peaks are most likely artifacts generated by clouds that pass above the generator and/or irradiance sensors.

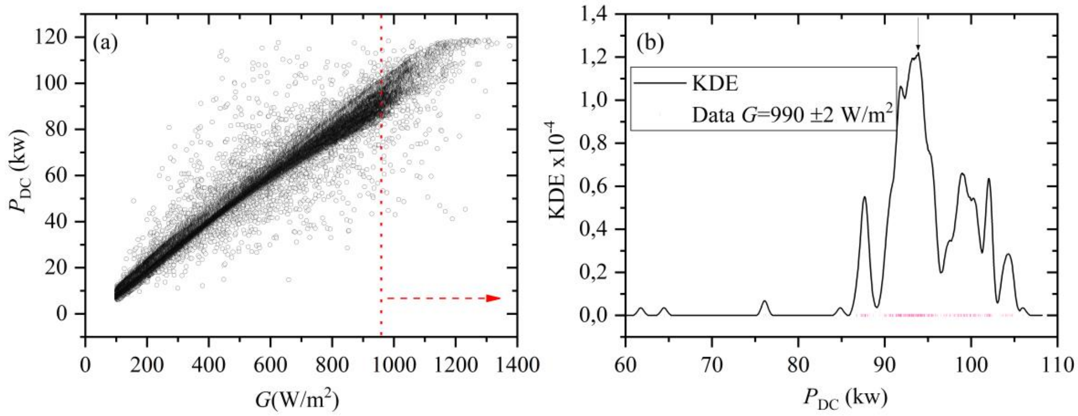

Fig. 3: (a) Data for 90 cloudy sky days. The red dotted line at 990 W/m^2 is an exemplary cross section of the data for KDE depicted in (b) to find the representative P_{DC} at the maximum indicated by the black arrow.

Generally, the KDE is a tool with the purpose to study the data distribution by generating the Probability Density Function (PDF) without any prior assumption of the data distribution. The KDE is defined as the convolution of multiple kernel functions (Qin et al., 2016), as described by the following eq. 4.

$$\hat{f}_{(p)} = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{p-p_i}{h}) \quad \text{(eq. 4)}$$

$\hat{f}_{(p)}$ represents the estimated probability density, n represents the sample size, h is the bandwidth of the estimation, d is the dimension of the data space, $K(z)$ represents the kernel function of z.

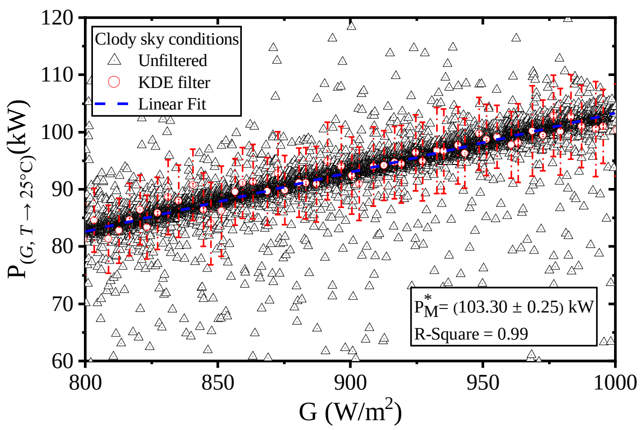

Fig.4: The corrected power at 25 ^{0}C after KDE filter applied in in difference to the unfiltered data of cloudy days condition

The type of kernel function has no significant effect on the estimation of KDE. Thus, for simplicity a Gaussian function can be considered for this type of data (Trashchenkov et al., 2018). In this work, a KDE in three dimensions is considered since there are three sets of input data. The latter avoids introducing further bias related to the calculation of the PDF via this method.

Given the three data collected (G, $T_{\mathrm{m}}, P_{\mathrm{DC}}$) at each instant of time. These were sorted by irradiance. Then, a set of measurements {$(G, T_{\mathrm{m}}, P_{\mathrm{DC}})_1$,…., $(G, T_{\mathrm{m}}, P_{\mathrm{DC}})_n$ } is considered, where n denotes the sample size in a defined interval of irradiance (ΔG = 4 W/m^2). This interval was chosen so that each sample contains at least 250 data points (Ren et al., 2014). For the 90 cloudy days, the sample size (n) was 300 points to generate the KDE.

After running the algorithm to calculate the KDE, a frequency distribution or the PDF is obtained. The maximum value $(G, T_{\mathrm{m}}, P_{\mathrm{DC}})^{Max}$ in the KDE for each set of measurements was defined as a representative value in clear sky conditions. The set of maximum values was then corrected in temperature by using equation 2.

Fig. 4 shows the corrected power values after applying KDE filtering in contrast to the unfiltered data of cloudy days condition. Notice how there are no more outliers present after the KDE filtering. In addition, the error bars represent the standard deviation of the data used for the estimation of the KDE since they will be considered in the calculation of the nominal power. Finally, to obtain the P_{M}^* in cloudy sky days, the values $P_{(G,T\to 25°\mathrm{C})}$ are fitted with equation 4, resulting in P_{M}^* = (103.3 ± 0.25) kW.

4. Results and discussion

The estimation of the nominal power according to the method proposed by (Martínez-Moreno et al., 2012) is strongly sensitive to the dispersion of data and outliers which deviate from the ideal working conditions of the PV generator. Their origin is attributed to cloudy day conditions and could be explained by:

(1) The temporal delay of the meteorological influence of the clouds on the measured parameters (Trashchenkov et al., 2018). For instance, a passing cloud can instantly affect G and P_{DC}, however T_{m} slowly adjusts to the new irradiation state and the measurement system will not record representative values under steady-state conditions.

(2) Passing clouds may introduce partial shading in the PV system and/or the irradiance sensor (Zhao et al., 2013) which would generate outliers in the measured data of power and irradiance. Particularly in relatively large power plants, such as the one under study, they may have a significant impact.

(3) The uncertainty in determining the operating temperature of the PV generator could have instantaneous temperature differences of up to 10 K (Muñoz Escribano et al., 2018).

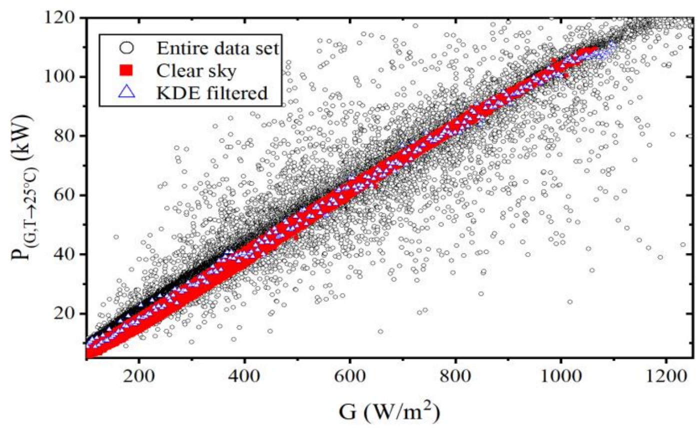

Fig. 5. The three data sets under study: all data collected (black circles), data only in clear sky conditions (red dots) and data on cloudy days after the filter by the blue KDE method (open triangles).

The identification of such anomalies are still controversially discussed since they are difficult to distinguish (Ding et al., 2018). However, from statistical point of view, the outliers can be defined as anomalous data that does not represent the steady-state operating conditions of the PV system. Hence, for the PV generator, the data point could be considered as an outlier value when the output power differs from data under clear sky conditions. The density of the measured data can show the regions where steady-state conditions occur. The exposure time of the photovoltaic generator to clear or partly cloudy conditions influences this data density.

The aim is to extract the clear sky conditions considering the data in cloudy conditions. To do this, statistical filters are found in the literature such as the 3-sigma rule, Hampel identifiers or boxplot rule outlier (Ding et al., 2018). These filters consider that the data has a normal distribution (Zhao et al., 2013). However, in section 3.2 Fig. 2 (b), The non-parametric behavior of the data was demonstrated in cloudy condition.

In this work, data filtering is based on finding the maximum KDE value that is considered the stable-state of the photovoltaic generator in clear sky conditions. Fig. 5 depicts the different data sets considered for this study. (1) The entire data set collected from March 27 to September 30, 2018, including clear and cloudy sky days (black circles), (2) the data set for only clear sky days (red squares) and (3) the data set obtained from only cloudy sky days after applying our proposed KDE filter (blue, open triangles).

To evaluate the calculation of nominal power in clear sky conditions, three approaches have been considered and compared:

(1) The entire data set of measurements in clear sky conditions was considered for a single fit. This means all the points in red in fig. 5.In this sense, in fig. 6(a), the irradiance had been divided into a small grid (ΔG=4 W/m^2) and it was treated under parametric statistics, $N(\mu, \sigma^2)$, with two parameters: the mean μ and variance σ^2 (σ is the standard deviation). The linear fit of the corrected power is shown in Fig. 6 (a). The result using equation 3 is P^*_{M} = (103.31 ± 0.04) kW.

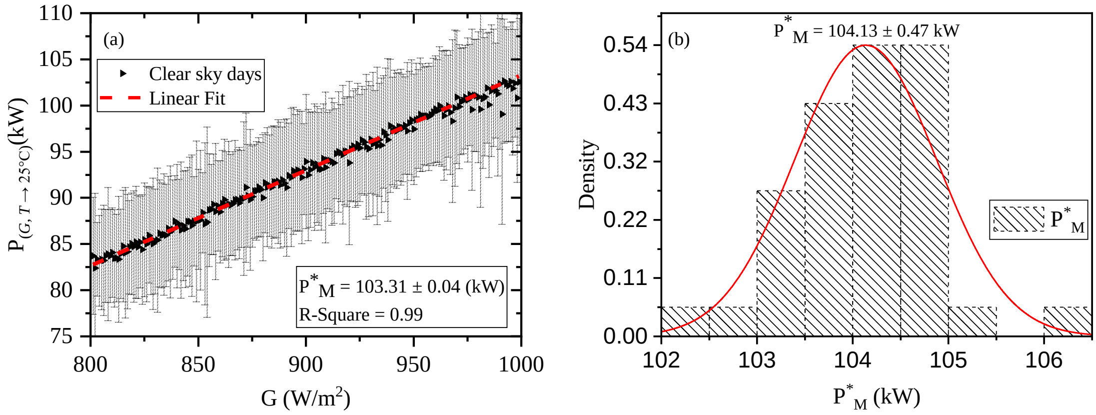

Fig. 6: (a) clear sky days with the corresponding mean and standard deviation. (b) Calculated distribution histogram collected from in the section 3.1

(2) In section 3.1, the nominal power for each day of clear sky was calculated. The calculated average nominal value and the dispersion were calculated resulting in P^*_{M} = (104.13 ± 1.56) kW. This indicates that if the nominal power is measured on any of the clear sky days, it will be in a 1.5% confidence range concerning the average value.

(3) Additionally, from the section 3.1, the corresponding histogram was calculated for all daily obtained nominal power values. In fig. 6 (b), a normal distribution approach with the nominal value and one sigma error is used. the P^*_{M} was calculated with the first sigma interval confidence. The result is P^*_{M} = (104.13 ± 0.47) kW.

As expected, the three approaches for clear sky conditions converge to similar P^*_{M} values. The aim is to contrast the P^*_{M} estimate in the entire measurement data that was divided into clear and partly cloudy days. In this sense, the approach (1) has been considered as a reference in favor of testing our estimation. Under partly cloudy conditions, the nominal power is calculated using the KDE filtering procedure, (103.3 ± 0.25) kW. This value is close to the values found in the three aforementioned cases. In this sense, the results on partly cloudy days show that the nominal power can be found within a range of acceptable error as under clear sky conditions.

5. Conclusions

To estimate the nominal power in a photovoltaic generator, the methodology proposed by Martinez et. al. was applied. Different approaches have been tested from a statistical point of view. To achieve this, the data of 37 days of clear sky and 90 days with cloudy sky conditions were collected.

For partly cloudy days, we have proposed a statistical method to filter out the noise and outlier data points. This method is based on the probability density function with a Kernel Density Estimation (KDE) of the data without modifying it. In this approach, the KDE filter allows us to find values within the range of the values that are considered to correspond to the operation of the system in moments of clear sky conditions during partially cloudy days.

After obtaining the filtered data in cloudy conditions, the P_{M}^{*} was calculated at (103.3 ± 0.25) kW. For clear sky conditions, the 37 days were considered in a single fit, getting a value of (103.31 ± 0.04) kW. These results clearly show that this method can be used well for the estimation of nominal power in either clear and cloudy sky conditions. Also, they are within the level of uncertainty of the measuring standard instrument (Yokogawa, 2001). Finally, this statistical tool could be applied in the daily diagnostic and monitoring of photovoltaic generators, for example by presenting alerts whenever the nominal power reaches values outside the ranges considered for optimal operation.

6. Acknowledgments

The authors would like to thank to IDEA research group for the measurements and for the helpful collaboration. This work was funded by the National Council of Science, Technology and Technological Innovation (CONCYTEC) under the PhD scholarship program, 236-2015-FONDECYT. Additional support has been provided by the Peruvian National Fund for Scientific and Technological Development (FONDECYT) and the World Bank through Contract N°045-2018-FONDECYT-BM-IADT-MU. Finally, J. A. Guerra, R. Grieseler and J. A. Töfflinger acknowledge the support provided by the Research Vicerectory of the Pontificia Universidad Católica del Perú, grant no. 2019-3-0041.

7. References

Ding, Hanxiang, Kun Ding, Jingwei Zhang, Yue Wang, Lie Gao, Yuanliang Li, Fudong Chen, Zhixiong Shao, and Wanbin Lai. 2018. Local Outlier Factor-Based Fault Detection and Evaluation of Photovoltaic System. *Solar Energy.* 164,139–48.

Kumar, Sandip Ravi, Francisco Gafaro, Andrew Daka, and Atul Raturi. 2017. Modelling and Analysis of Grid Integration for High Shares of Solar PV in Small Isolated Systems – A Case of Kiribati. *Renewable Energy.* 108, 89–97.

De la Parra, I., M. Muñoz, E. Lorenzo, M. García, J. Marcos, and F. Martínez-Moreno. 2017. PV Performance Modelling: A Review in the Light of Quality Assurance for Large PV Plants. *Renewable and Sustainable Energy Reviews.* 78, 780–97.

Lomas, J. C., E. Muñoz-Cerón, G. Nofuentes, and J. de la Casa. 2018. Sale of Profitable but Unaffordable PV Plants in Spain: Analysis of a Real Case. *Energy Policy.* 117.

Martínez-Moreno, F., E. Lorenzo, J. Muñoz, and R. Moretón. 2012. On the Testing of Large PV Arrays. *Progress in Photovoltaics: Research and Applications*. 20, 100–105.

Muñoz-Cerón, E., J. C. Lomas, J. Aguilera, and J. de la Casa. 2018. Influence of Operation and Maintenance Expenditures in the Feasibility of Photovoltaic Projects: The Case of a Tracking Pv Plant in Spain. *Energy Policy.* 121, 506–18.

Muñoz Escribano, Mikel, Miguel Garcia Solano, Inigo De La Parra Laita, Javier Marcos Alvarez, Luis Marroyo, and Eduardo Lorenzo Pigueiras. 2018. Module Temperature Dispersion within a Large PV Array: Observations at the Amareleja PV Plant. *IEEE Journal of Photovoltaics.* 8, 1725–31.

Muñoz, J. V., G. Nofuentes, M. Fuentes, J. de la Casa, and J. Aguilera. 2016. DC Energy Yield Prediction in Large Monocrystalline and Polycrystalline PV Plants: Time-Domain Integration of Osterwald's Model. *Energy.* 114, 951–60.

Qin, Xiaolin, Yonggang Li, Chao Shen, Zengqiang Zhang, and Xuyao Zeng. 2016. The Correlation Analysis of Clean Energy Output Based on Nonparametric Kernel Density Estimation Probability Models. *Atlantis*

Press. 133, 24–28.

Reise, Christian, Björn Müller, David Moser, Giorgio Belluardo, and Philip Ingehoven. 2018. *Uncertainties in PV System Yield Predictions and Assessments*.

Ren, Zhouyang, Wei Yan, Xia Zhao, Wenyuan Li, and Juan Yu. 2014. Chronological Probability Model of Photovoltaic Generation. *IEEE Transactions on Power Systems*. 29, 1077–88.

Schweiger, Marcus, Johanna Bonilla, Werner Hermann, Andreas Gerber, and Uwe Rau. 2017. Performance Stability of Photovoltaic Modules in Different Climates. *Progress in Photovoltaics: Research and Applications*. 20, 1114–29.

Schweiger, Markus, Werner Herrmann, Andreas Gerber, and Uwe Rau. 2017. Understanding the Energy Yield of Photovoltaic Modules in Different Climates by Linear Performance Loss Analysis of the Module Performance Ratio. *IET Renewable Power Generation*. 11, 558–65.

Talavera, D. L., E. Muñoz-cerón, J. P. Ferrer-rodríguez, and G. Nofuentes. 2016. Evolution of the Cost and Economic pro Fi Tability of Grid-Connected PV Investments in Spain : Long-Term Review According to the Different Regulatory Frameworks Approved. *Renewable and Sustainable Energy Reviews*. 66, 233–47.

Talavera, D. L., E. Muñoz-cerón, J. De La Casa, D. Lozano-Arjona, and M. Theristis. 2019. Complete Procedure for the Economic , Financial and with Self-Consumption. *Energies*. 12, 345.

Trashchenkov, Sergei, Sergio Pires Pimentel, Victor Astapov, Andres Annuk, and Enes Goncalves Marra. 2018. Kernel Density Estimation for Stochastic Modeling of PV Power Output. *7th International IEEE Conference on Renewable Energy Research and Applications, ICRERA*. 5, 1179–83.

Watson, James and Michel Schmela. 2018. Global Market Outlook. 1–80.

Welter, Philippe. 2019. PHOTON International. *March*, 1–81.

Yokogawa. 2001. "Digital Power Meter."

Zhao, Ye, Brad Lehman, Roy Ball, Jerry Mosesian, and Jean Francois De Palma. 2013. Outlier Detection Rules for Fault Detection in Solar Photovoltaic Arrays. *Conference Proceedings - IEEE Applied Power Electronics Conference and Exposition - APEC*. 2913–20.