Loughborough
University

This item was submitted to Loughborough's Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/

Outlier Identification in Outdoor Measurement Data - Effects of Different Strategies on the Performance Descriptors of Photovoltaic Modules

J. Zhu*, T.R. Betts, R. Gottschalg

Centre for Renewable Energy Systems Technology (CREST), Department of Electronic and Electrical Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom
*Corresponding author: Tel.: +44 1509 635313, Fax: +44 1509 635301, Email: J.Zhu@lboro.ac.uk

Abstract

Outdoor measurement campaigns of PV module performance are normally affected by a relatively large number of outliers. The aim of this paper is to develop a statistically sound approach of obtaining a dataset that allows one to analyse continuously monitored devices. This paper uses I_{SC} as a self-reference parameter to measure the incident irradiance on the module, which largely reduces the error due to spectral and angular effects. The outlier identification procedure is based on statistical distribution analysis of different performance descriptors and it assures 0.99 confidence level and the same skewness for the remaining data. This approach can be applied to whole datasets as well as for data in specific irradiance-temperature bins. The developed methodology will be used to analyze outdoor data from different devices at different locations with reduced uncertainty.

Introduction

Outdoor measurement campaigns of photovoltaic (PV) module performance are normally affected by a relatively large number of outliers. In order to analyse this data with low uncertainty one needs to carefully remove these points without skewing the resulting device statistics [1]. The aim of this paper is to develop a statistically sound approach of obtaining a dataset that allows one to analyse continuously monitored devices. The work makes only a limited number of approximations and is based entirely on the analysis of statistical distributions of performance descriptors investigated. This approach is applied to outdoor data measured at the Centre for Renewable Energy Systems Technology (CREST), Loughborough, which operates a well-instrumented outdoor system. Different test sites would not have the same instrumentation and the aim is to develop a general methodology that allows the extraction of data of similar quality on all sites without introducing artefacts due to inappropriate data limitations.

To analyze the power rating of PV devices, the module characteristics are sorted into different in-plane irradiance (G_i) bins. Data within each irradiance bin are corrected for temperature and irradiance using standard procedures described in IEC 60891 [2]. The spread of performance parameters in each bin is significantly affected by the meteorological conditions, such as temperature, irradiance, cloudiness, etc., as well as the equipment used for measuring. This inevitably causes bias as well as statistical errors. For all devices investigated in this work, the variation is much larger for low irradiance levels. This paper starts with describing the methodology of outlier identification in dataset. It then analyzes the parameter of short circuit current over in-plane irradiance (I_{SC}/G_i) as it will be demonstrated later that the I_{SC} is a good measure of G_i as seen by the module. This then allows the use of I_{SC} instead of G_i as a reference parameter to minimise uncertainty in further analysis, provided that an accurate calibration is carried out. Having established an appropriate effective irradiance on the module then allows re-evaluation of the G-T bins into I_{SC}-T bins. This gives further possibilities to identify outliers by investigating other descriptors within the obtained I_{SC}-T bins.

Methodology

Experimental data must be scrutinized for outliers before to draw meaningful conclusions from it. Typically, this is achieved by computing the mean and the standard deviation (std) using all the data and rejecting any that are over 3 std away from the mean (Normal distribution is assumed) [1]. However, as the number of data samples in the dataset increases, the probability of seeing extreme samples also increases. It is not considered good statistically to reject outliers without strong cause. Identifying outliers without cause typically results in underestimating the actual variability of the data. Therefore, the statistical rules for the outlier identification procedure employed in this paper are entirely based on the statistical distribution analysis. It should firstly guarantee data confidence level (CL) to be greater than 0.99 by selecting at least 99% of raw data, i.e. rejecting outliers not more than 1% of original measurements. Next, it should maintain the same skewness for the resulting data if the original data distribution is not symmetric by slightly adjusting the confidence interval window. This is the cornerstone of analyzing different descriptors.

Descriptor 1: Short Circuit Current over In-Plane Irradiance

As the I_{SC} exhibits a linear relationship with G_i, the ratio of I_{SC}/G_i is to be a performance descriptor and its distribution, as shown in Figure 1, needs to be studied. This distribution, i.e. probability density function (pdf), has a mean of 2.649 and std of 0.323 (units of $[mA]/[W/m^2]$). The skewness is 1.268, which indicates an asymmetry scale. It is shown that 81.2% of measurements are within the range of the mean +/- std. But considering data in the range of the mean +/- std only will not reflect the whole behaviour of this device. Data at relatively high or low I_{SC}/G_i levels might be obtained when the cloud is moving into or away from the sun and should not be ignored as it is often seen in the conditions of a partially cloudy day with fast moving clouds, as is often seen in the UK climate. Of course, this situation is site dependent which needs to be investigated separately. It is not difficult to see in Figure 1 that if filtering the data on the basis of the mean +/- 1, 2, or 3 std, the device statistical information will be skewed, which is to be avoided in the method developed in this paper.
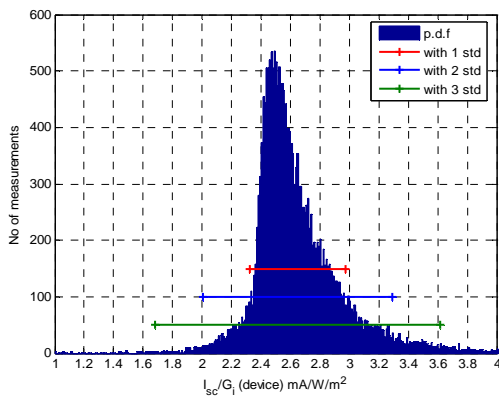

Figure 1: Distribution of I_{SC}/G_i of an unshaded c-Si module

Shading on Modules

One of the advantages of using the parameter of I_{SC}/G_i is that one can identify shading. This is demonstrated for a module deployed at the CREST outdoor facility, which experiences some shading due to a neighbouring structure. The structure is to the west of the rack on which the module installed. Thus it can cause a shadow on part of the module in late afternoon (depending on the solar elevation angle) and no shadow on the pyranometer or reference cell. This results in the pdf of I_{SC}/G_i shown in Figure 2 that low values of I_{SC}/G_i (below mean - std) are much more common than high values (above mean + std), (which are close to zero). In order to investigate the shading on modules, the pdfs of solar elevation and azimuth angle with I_{SC}/G_i below mean - std for the shaded module are generated

and are plotted in Figure 3 and Figure 4, respectively. Nearly 72% of the measured data occurred at solar evelation angles below 15°, indicating that these measurements were taken at either early morning or late afternoon. The asymmetric pdf of azimuth angle reveals that in addition to spectral mismatch effect, shading also occurs on the module and is indicated to occur in the afternoon.
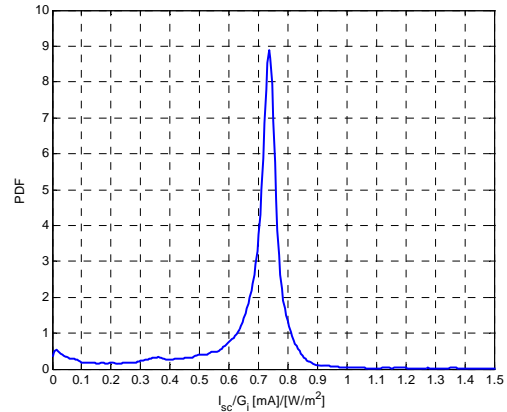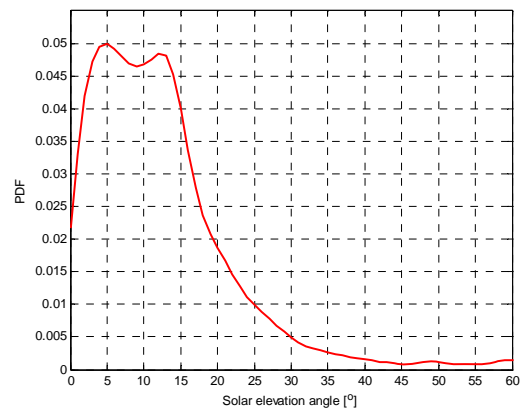

Figure 2: pdf of I_{SC}/G_i for the shaded module

Figure 3: pdf of solar elevation angle at low I_{SC}/G_i levels for the shaded module
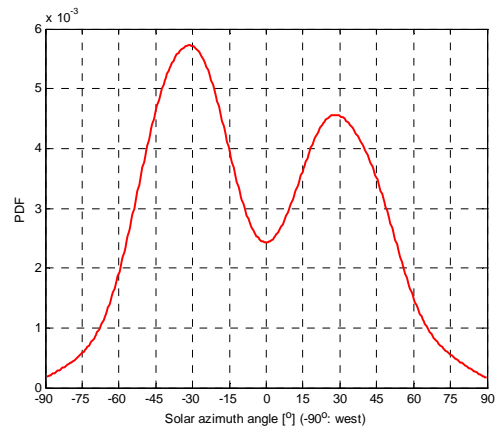

Figure 4: pdf of solar azimuth angle at low I_{SC}/G_i levels for the shaded module

Self-referenced by Short Circuit Current

In order to minimise the effects due to differences between the irradiance sensor and the module under measurements, such as slow response of pyranometers and difference in the spectral response and angle of incidence behaviour, it is proposed to use the self-reference method, i.e. using I_{SC} as reference parameter for irradiance instead of G_i. A calibration of I_{SC} to the effective irradiance is needed to establish the inter-relationship. In this study, outdoor data of an a-Si module at CREST is taken as an example.

Typically, I_{SC} behaves linearly with incident irradiance for all single junction devices, although the relative deviation is much larger at low irradiances than at high irradiances. Thus the I_{SC} is a good measure of the effective irradiance provided it is calibrated accurately. Calibration of I_{SC} to effective irradiance can be carried out at high irradiances, where typically the angle of incidence as well as incident spectrum is fairly constant and by definition very close to standard test conditions (STC). Outlier identification can be applied to a single irradiance bin using the same statistical rules. Figure 5 shows the pdf of I_{SC}/G_i at G_i in the range 950 – 1050 W/m^2, which will be used to determine the effective irradiance level at STC. Outliers have been removed based on the statistical rules described in the previous section and all measurements are corrected to the condition of 1000 W/m^2 and 25 °C using the irradiance and temperature coefficients derived from the measurements. The pdf has a mean of 0.247 and median of 0.247 [mA]/[W/m^2] with +/- 2.89% uncertainty (99% confidence level). Thus, the I_{SC} calbration can be established.
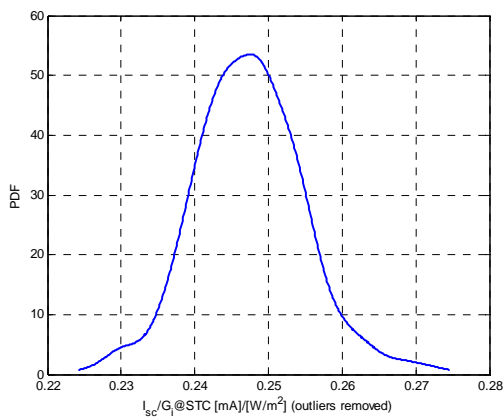
Figure 5: pdf of I_{SC}/G_i at G_i = 950 – 1050 W/m^2

Descriptor 2: Efficiency against Irradiance

The a-Si module efficiency exhibits a distinct relationship with irradiance, which can be used as a descriptor. Figure 6 plots the normalised module efficiencies calculated from the a-Si module measurements against measured in-plane irradiance (from thermopile pyranometer) and calculated effective irradiance (from I_{SC} self-reference), respectively. The red dots represent module efficiencies against effective irradiance, which show a clear trend of the module efficiency decreasing dramatically at low irradiances (below 100 W/m^2), reaching a maximum at medium levels (between 200 – 400 W/m^2), and then slightly decreasing at high irradiances. The blue dots, representing the efficiencies against measured irradiance, have a much larger spread. This is especially evident for low irradiance levels as the measured incident irradiance measured by pyranometers varies significantly from that seen by the semiconductor junction of the test sample.
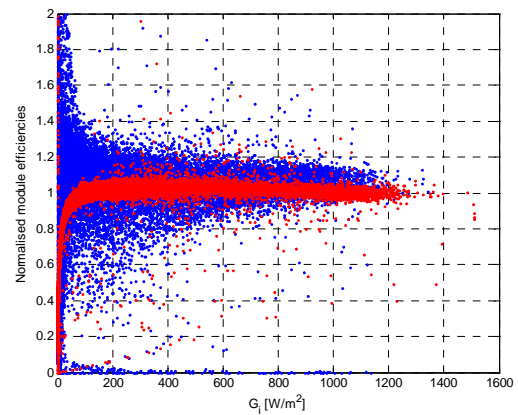
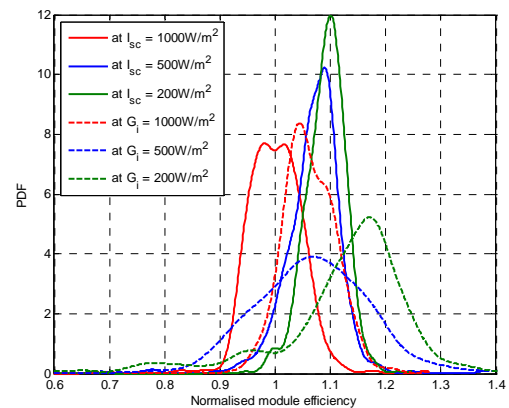**Figure 6: Normalised module efficiencies against irradiances**

Figure 7: pdfs of module efficiencies at different irradiance levels

Figure 7 plots pdfs of normalised module efficiencies at irradiance bins of 200, 500, 1000W/m^2 with +/-5% bin width. The solid lines represent pdfs obtained using I_{SC}, i.e. effective irradiances, while the dashed lines represent pdfs calculated by measured irradiance G_i. On one hand, as irradiance increases from 200 W/m^2 to 1000 W/m^2, the module efficiencies decrease

slightly due to series resistance losses. On the other hand, the pdfs calculated by G_i have much larger deviation than pdfs calculated by I_{SC}, and the skewness or shape of the distribution is different. Applying similar statistical rules as defined above, outliers can be identified at the specific irradiance bins one is interested in. These outliers are mainly due to measurement instrument failures, dirt or snow/ice/dew on pyranometers or modules. Taking the pdf at 200 W/m^2 for example, the mean and median with standard deviation are 1.09, 1.09 +/-3.16% for Efficiency-I_{sc} and 1.13, 1.15 +/- 9.78% for Efficiency-G_i, respectively.
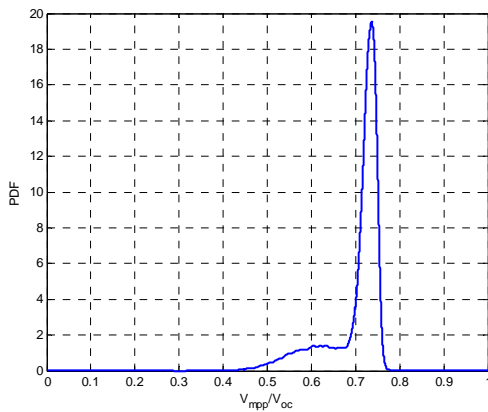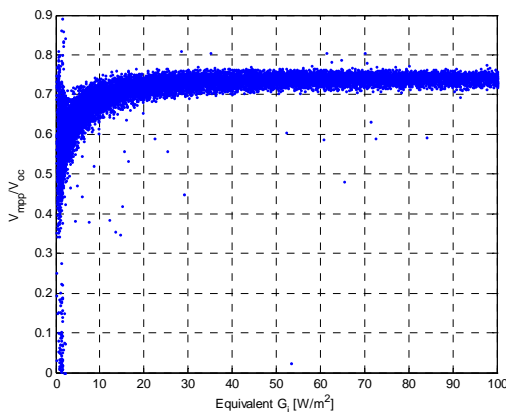

Figure 8: pdf of V_{MPP}/V_{OC} for an a-Si module

Figure 9: V_{MPP}/V_{OC} at G_i < 100 W/m^2

Descriptor 3: Voltage at Maximum Power Point over Open Circuit Voltage

The ratio of voltage at maximum power point to open circuit voltage (V_{MPP}/V_{OC}) provides some information about the shape of I-V curves and can be used as a further descriptor to identify outliers in measurements as it is a rather stable property as shown in Figure 8. This shows the pdf of V_{MPP}/V_{OC} for an a-Si module. Around 20% of measurements lie within the range of V_{MPP}/V_{OC} below mean - std. Looking closely at irradiances below 100 W/m^2, as shown in Figure 9, nearly all

the measurements of V_{MPP}/V_{OC} below 0.7 occur at irradiance less than 10 W/m^2. This is due to the shunt resistance limitation at low irradiances [3] and these data will not be considered in the further analysis of power rating or module degradation.

Cleaned Dataset

After filtering the outdoor measurements for different performance descriptors, a cleaned dataset of measurements is obtained, which preserves the original statistical information of the device behaviour investigated, and has greatly reduced the influence of outliers in further uncertainty analysis. This makes it possible to obtain high accuracy with the same uncertainty in analysis of energy rating of modules at different sites, or calculating PV module monthly degradation rates.

Conclusions

This paper uses I_{SC} as self-reference parameter to measure the incident irradiance on the module, which largely reduces the error due to spectral effect etc. Outlier identification procedures based on statistical distribution analysis assures confidence level of 0.99 and same skewness by rejecting outliers not more than 1% of original data and adjusting the confidence interval to match the original skewness. This can be applied either to the whole dataset or the specific irradiance-temperature bins of interest. This approach has been verified with outdoor data measured at CREST, Loughborough.

Acknowledgements

References

[1] D.S. Moore, G.P. McCabe, Introduction to the Practice of Statistics 5th Edition, 2006, W.H. Freeman and Company, USA.

[2] International Standard CEI IEC 60891, Procedures for Temperature and Irradiance Corrections to Measured I-V Characteristics Of Photovoltaic Devices.

[3] T.J. McMahon, T.S. Basso, and S.R. Rummel, Cell Shunt Resistance and Photovoltaic Module Performance, 25th IEEE Photovoltaic Specialists Conference Proceedings, l996, p.1291.