A CLUSTERING APPROACH FOR PV MODULE SAMPLING

Gianluca Corbellini[a], Sebastian Dittmann[a]
[a] University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Institute for Applied Sustainability to the Built Environment (ISAAC)
Campus Trevano, 6952 Canobbio, Switzerland
gianluca.corbellini@supsi.ch, +41 (0)58 666 62 34 , www.supsi.ch

ABSTRACT: To accomplish the quality of overall installation, it is fundamental to guarantee the quality of the batch of PV modules that are going to be installed, the plant will clearly benefit from high quality modules and low discrepancies between module-to-module performance in order to reduce the electrical mismatch effects.
For this purpose SUPSI developed a new procedure to select the modules to be tested in order to guarantee a statistical significance of the testing and cover possible subgroups of modules.
This tool has been designed and developed during the project 'Performance Plus - Tools for Enhanced Photovoltaic System Performance - Grant agreement no: 308991" inside the Framework Program 7 backed by the European Commission.

1 OVERVIEW OF THE PROBLEM

In the paper we propose a method to support laboratory testing, using statistical properties of batches of PV modules to support the lot acceptance procedures for medium and big size PV plants.
To accomplish the quality of overall installation, it is fundamental to guarantee the quality of the batch of PV modules that are going to be installed, the plant will clearly benefit from high quality modules and low discrepancies between module-to-module performance that will result in reduced electrical mismatch effects.
This could be applied to single installation of medium and large size, and for residential installers combining several installations together; the order of magnitude of module for batch to make it reasonable is in the order of hundreds.
As said before, for these testing it is crucial to find a good trade-off between the number of modules tested (that can let the testing be expensive) and the representativeness of results, in terms of statistical properties of the chosen sample.
The starting point of the testing is the flasher list that modules manufacturer usually provides to its customer, that list the electrical characteristic of modules as tested with the solar simulator installed at the end of the production line.
In SUPSI we tested several batch of modules for lot acceptance in our accredited laboratory "Swiss PV Module Test Center" (SPVMTC); based on lab experience we selected three main root causes significant differences between actual measurements and expected results:

- The solar simulator of the manufacturer is not well calibrated (we find a bias between our and manufacturer measures, bigger than the uncertainty of measures)
- The transportation of modules was not performed well, so some modules have cracked cells that could easily be observed with electroluminescence test
- Intentional distortion of measured values by the manufacturer

In order to reach the economic and reliability goals we developed a new procedure for PV batch testing which comprises of three steps:

1. Definition of number of modules to test depending on level of acceptance ("How many?")
2. Selection of the modules ("Which ones?")
3. Comparison of test results with flasher list provided by modules' manufacturer ("Is the lot acceptable?")

Task 1 is already covered by an existing software, later explained, while for tasks 2 and 3 new procedures and software have been implemented.

2 DEFINITION OF THE NUMBER OF MODULES

To establish the number of modules to be tested (task 1) RWTH Aachen developed APOS software [1], that starting from the flasher list (if provided, otherwise estimating statistical distribution of modules parameters) calculates the number of modules to be tested assuring the defined statistical significance of the test. So the important inputs that will impact on the number of modules are:

- Acceptable Quality Level (AQL), this is the maximal fraction of conforming modules which is in agreement with a "high quality" (of the lot). If the fraction of nonconforming modules of the lot is less or equal to the AQL, the lot is regarded as being of high quality.
- Rejectable Quality Level (RQL), this is the smallest fraction of non-conforming modules being regarded as "low quality". Thus, if the fraction of non-conforming modules of a lot is greater or equal to RQL, the lot is regarded as "low quality".

A screenshot of the software is represented in the fig.1.

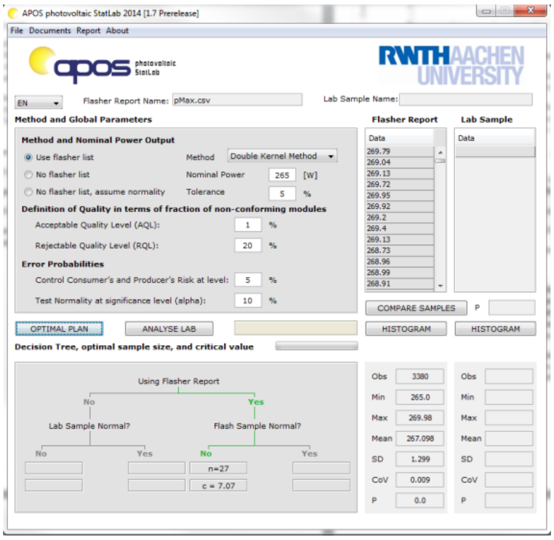


Figure 1: GUI of APOS software

3 LAYOUT SPECIFICATIONS

Once the task 1 is accomplished using APOS software, it is suggested by the normative [2] to randomly choose the modules from the batch; from a mathematical point of view it is correct, but effective only with a large number of modules; in real cases this can be risky with few modules as it often happens and so in SUPSI we developed a tool to have always a statically significant sample and avoid weird selections.

The PV modules as an electrical component are described by the well known I-V characteristic, unfortunately the manufacturer does not provide it for each module but the input we can use is the flasher list, that is typically structured with 5 values for each modules:

- V_{OC} [V]: the voltage at the open circuit point
- I_{SC} [A]: the current when the module is short circuited
- V_{MPP} [V]: the voltage in the MPP (maximum power point) of the module
- I_{MPP} [I]: the current in the MPP of the module
- P_{MAX} [W]: the power output in the MPP calculated as the product of current and voltage in the point of maximum output of the module

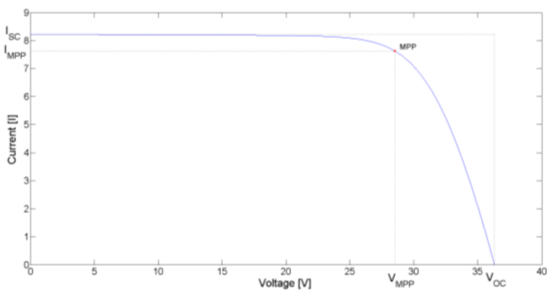


Figure 2: Typical I-V curve of a PV module with depicted V_{OC}, I_{SC}, V_{MPP} and I_{MPP}

These values (represented in fig.2) are measured using a solar simulator installed at the end of the production line of the manufacturer and will be compared with the results from laboratory testing.

One more indicator useful to describe the I-V characteristic is the Fill Factor (FF from now on) that represents the "squareness" of the curve and is defined as the ratio between the rectangles formed by (V_{MPP}, I_{MPP}) and (V_{OC}, I_{SC}):

$$FF = \frac{V_{MPP}\, I_{MPP}}{V_{OC}\, I_{SC}}$$

The idea behind this procedure is to identify a certain number of modules with different characteristic in order to cover possible discrepancies inside the batch:

- Modules tested with different solar simulators, or the same one re-calibrated
- Modules produced in different time periods
- Modules before and after improvements of production line

To cover all these aspects we considered a **clustering** of the batch of the modules in n clusters where n is the size of the sample as provided by APOS software.

The clustering considers the three dimensional space defined by V_{OC}, I_{SC} and FF; the maximum power, that is just the product of this three, has been neglected because the batch is already filtered into a tiny region (typically 2% or less wide) and so not significant as a variable; moreover being it just the product of the other three variables, it would add degree of freedom to the analysis.

Considering that these three variables have different units of measure and different order of magnitude, it is necessary to normalize these values before running the clustering; this normalization is performed using the average (μ) and the standard deviation (σ) of each variable:

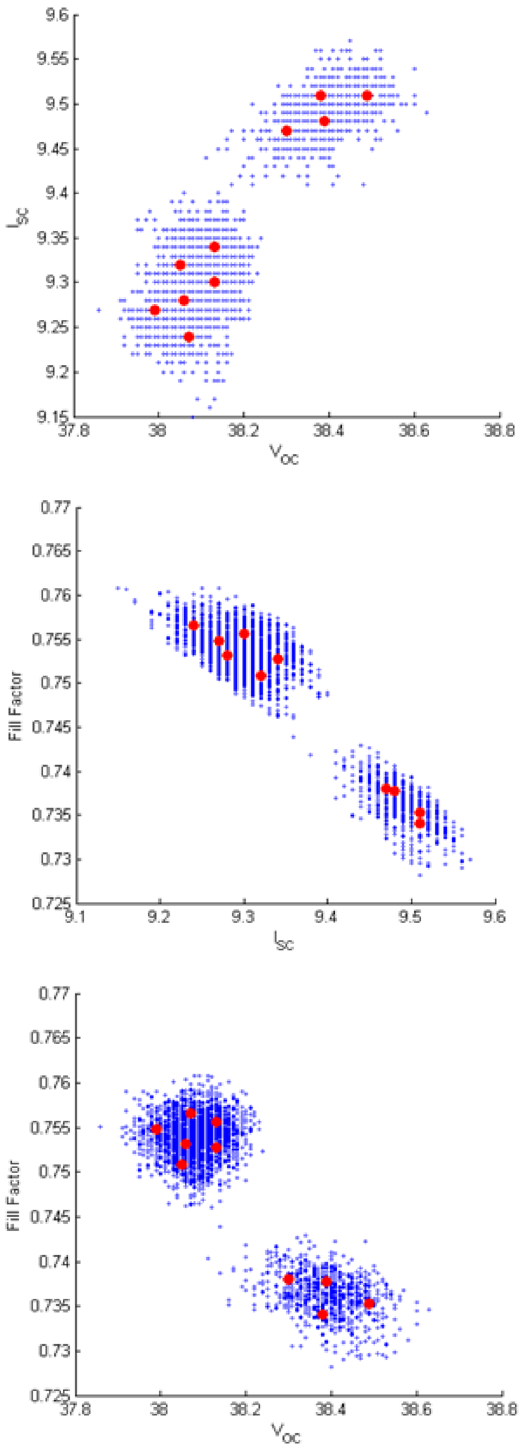
$$x^* = \frac{(x - \mu)}{\sigma}$$

Clustering algorithms are deeply studied for data mining purposes, in particular to solve this problem we adopted the **k-means** algorithm [2] that aims to partitions n observations (from 100 to 10000 in our case) into k clusters (10 to 40 typically) in which each observation belongs to the cluster with the nearest mean, serving as prototype for the cluster.

To solve the k-means problem, heuristic algorithm are commonly applied (while the problem is computationally NP-hard) and the resulting set of centroids depends on the initialization; to avoid non optimal solution we run a kind of **Monte Carlo k-means** (where the initialization is random) that consists in running k-means several times and then take the solution that minimize the sum of the distance of each point to the centroid assigned to it.

The software has additional features to study basic statistical properties of the batch, namely the minimum, the maximum, the median, the mean and the standard deviation for each of variables studied. In fig.3 it is shown an interesting case where the batch is the combination of two separate subset with significantly different characteristic, the 3-d representation is optimized to be aligned over principal components (computed from PCA) of the distribution.

An additional output shows the two dimension view of the parameters of the batch (in blue) and the selected modules for the testing (in red), examples in the fig. 3.
The tool has been developed in Matlab [3] and then compiled as a stand-alone application (fig. 4).



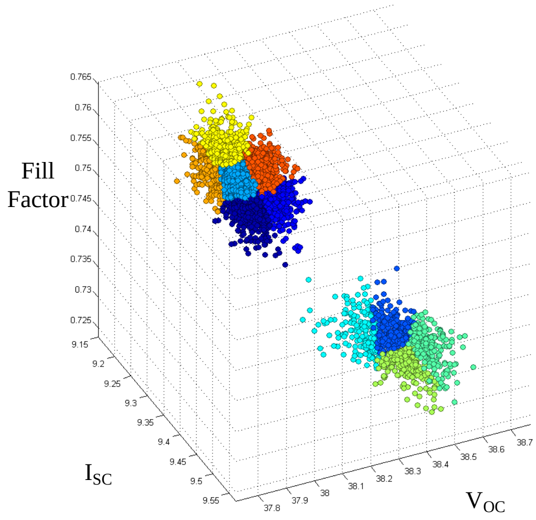


Figure 3: Two and three dimensional plot of the parameters, the tool is able to cover different sub-groups of modules as expected

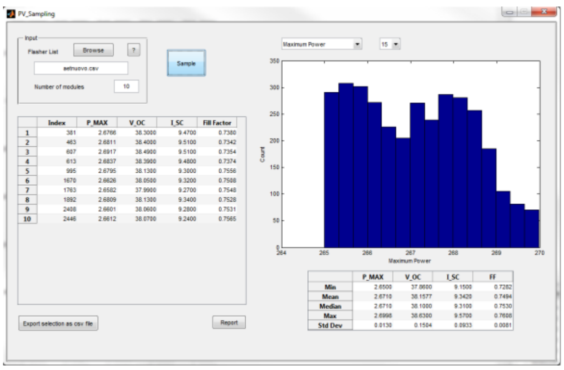


Figure 4: Screenshot of the tool, developed using Matlab and compiled as a stand-alone application

4 LOT ACCEPTANCE

Once modules have been selected, in laboratory I-V characteristic at standard conditions (G = 1000 W/m2, T = 25°C, Air Mass = 1.5) are performed and then the results of maximum power for each module will be used into the APOS software to test if the measurement are inside the range of acceptance.
To understand if the lot is acceptable or not, we have to apply the measurement uncertainties of the setup of our equipment and the one provided by the manufacturer to establish if the flasher list is reliable or not; for this reason the lower the uncertainties of the testing is, the higher the confidence in the testing will be.

5 CONCLUSIONS

The procedure and tool developed for PV module sampling will add a significant confidence in the lot acceptance procedure of PV module testing laboratories.
This has been developed inside a project focused on crystalline silicon modules but could be applied to other technologies as well.

REFERENCES

[1] IEA61215 normative - Crystalline silicon terrestrial photovoltaic (PV) modules - Design qualification and type approval

[2] - Christopher M. Bishop - Pattern Recognition and Machine Learning

[3] - www.mathworks.com/products/matlab/