21st European Photovoltaic Solar Energy Conference, 4-8 September 2006, Dresden, Germany

# STATISTICAL AND EXPERIMENTAL METHODS FOR ASSESSING THE POWER OUTPUT SPECIFICATION OF PV MODULES

W. Herrmann[1], J. Althaus[1], A. Steland[2], H. Zähle[2]
[1]TÜV Rheinland Immissionsschutz und Energiesysteme GmbH, D–51101 Cologne, GERMANY
Phone: +49 221 806 2272, e-mail: werner.herrmann@de.tuv.com
[2]RWTH Aachen, Institut für Statistik und Wirtschaftsmathematik
Wüllnerstr. 3, D-52056 Aachen, GERMANY
Phone: +49 241 80 94573, e-mail: steland@stochastik.rwth-aachen.de

ABSTRACT: End users and distributors of PV modules are often faced with the problem of how to verify the power output specification of a new shipment consisting of a larger number of PV modules. As a first step, claims can be based on the data sheet containing information on the nominal power output (related to STC) and the power output tolerance. The question then arises as to the evaluation of the shipment regarding fulfilment of the power output specifications. A control measurement of all modules is not realistic, so that in practice a certain sample size will be randomly selected and re-measured by a test institute. The issue is then whether the sample size is sufficient for a reliable inference and whether the test samples suitably represent the whole shipment. These questions often lead to disputes between module producers and consumers that not infrequently result in litigation. Accordingly, there is a great need for harmonised procedures suitable for evaluating whether the power output of a shipment lies within the manufacturer's specifications. On the other hand, inferring from random samples to the properties of larger populations is a major task of applied statistics and falls within the field of significance testing. The latter is based on statistical hypotheses and includes so-called error probabilities in reference to both the producer's risk and the consumer's risk. On the basis of these statistical methods we developed sampling plans and acceptance criteria for two practical cases: a) STC performance data of modules (as measured by the manufacturer) are supplied with the shipment, and b) no information at all is provided. Furthermore, we distinguish between the case of a normal distribution of the total production population and the case where no specific distribution is assumed. Our calculations resulted in minimum sample sizes and acceptance criteria that both correspond to a given statistical significance level. Including such decision rules in the conditions of acceptance of sales contracts may be a useful approach.

Keywords: PV module, performance, qualification and testing

## 1 INTRODUCTION

The power output specification of PV modules commonly refers to the maximum power point on the I-V characteristic that is measured under standard test conditions (STC). However, due to variations of the solar cell performance data and effects of the manufacturing process and the measurement system, the actual power output is subject to a certain scattering. The total production population of a given module type can accordingly be characterised as a parent population corresponding to a certain probability distribution of $P_{MAX}$. By sorting modules into power classes we can define different module types with a lower production tolerance. Together with the nominal value $\mu_0$ of the module type the production tolerance is stated in the data sheet and usually lies within the range ±3% to ±10%. Figure 1 shows an example of a $P_{MAX}$ frequency distribution. If a normal distribution is assumed, approximately 95% of the values should lie within the interval $\mu \pm 2 \cdot \sigma$.

We now wish to evaluate a population of N modules (shipment) as to the fulfilment of the power output specification. In practice, a certain number of n modules (sample size) are randomly selected and measured by a test laboratory. However, survey results show that the practices of test institutes for evaluating populations of modules with regard to fulfilment of the manufacturer's specifications are not harmonised at present and can vary greatly. For example, sample sizes of 1% or $\sqrt{N-1}$ are used which can lead to considerable differences in the number of test samples and therefore in testing costs. A detailed sampling method for the power rating of

photovoltaic modules has been developed by PowerMark in the US [1]. Here 7 modules are selected at random from a production batch or batches consisting of at least 100 modules produced on at least 5 different days. No information is provided on the statistical significance of this approach, however.
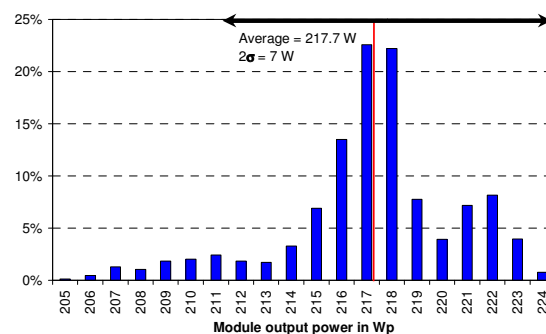


**Figure 1:** Example of a $P_{MAX}$ frequency distribution of 1741 modules with nominal power output 200 W.

On the other hand, PV module manufacturers do conduct individual performance measurements of PV modules at the end of the production line by means of pulsed solar simulators (flashers). The measurement results are commonly stored in databases that can be used for statistical analyses. More and more often these flasher reports accompany PV module shipments, at the customer's request. The reports afford consumers new ways to validate compliance with the data sheet specifications and directly check power ratings.

In this context we should mention TÜV Rheinland's procedure for the "power-controlled" certification of PV modules [2], which assures the end user of compliance of module power output with the power rating. Power-controlled certification involves the regular inspection of measurement equipment and measurement procedures in module production lines. In addition, one module per month is selected at random from production and shipped to TUV together with the flasher reports of all modules manufactured on the relevant days. A module type is deemed to fulfil power-controlled requirements if deviations between measurements by the producer and by the test lab fall within the stated production tolerance range.

Inferring from random samples to the properties of larger populations is a major task of applied statistics and falls in the field of significance testing. The latter is based on statistical hypotheses and includes so-called error probabilities with respect to both the manufacturer's risk and the consumer's risk. We have studied the utility of these approaches for the validation of PV module power ratings. Our work aimed at developing sampling methods for various realistic cases and defining clear-cut acceptance criteria guaranteeing a certain level of statistical relevance subject to agreement between the manufacturer and the consumer.

## 2 STATISTICAL DECISION RULES

### 2.1 Definition of decision rules for PV modules

A single module is called conformant (acceptable) if its power output lies above the threshold $\mu_0 - \varepsilon$, where $\mu_0$ is the nominal power output and $\varepsilon$ is the production tolerance. Otherwise the module is considered non-conformant. p denotes the fraction (percent) of non-conformant modules within the whole shipment. Intuitively, the whole shipment should be accepted if p lies "close" to 0 and rejected if p lies "far away" from 0.

The actual p is usually unknown and cannot be determined exactly, except in the case of a total re-measurement. Therefore the decision of accepting or rejecting the shipment can be based only on the control measurement, $x=(x_1,\ldots,x_n)$. In particular, we cannot use p in the decision-making. The solution is the construction of a suitable test function T which, based on the sample x, settles the acceptance or rejection of the shipment. If $T(x)$ falls within a suitable acceptance range, the shipment will be accepted, otherwise it will be rejected.

T is a random variable and hence the acceptance of the shipment is a random event. Therefore both the manufacturer and the consumer run the risk of deciding incorrectly. Incorrect decisions can be classified as follows:

Consumer risk: Based on the sample, the decision rule accepts the shipment, although it is of low quality.

Producer risk: Based on the sample, the decision rule rejects the shipment, although it is of high quality.

A decision rule should fulfil the following requirements for controlling both the consumer's and producer's risks.

(a) A high quality shipment should be accepted with probability not less than a given value α, which should be chosen as large as possible.

(b) A low quality shipment should be accepted with probability not exceeding a given value β, which should be chosen as small as possible.

The quality of a shipment is measured in terms of the proportion p of non-conformant PV modules. A shipment is of high quality, if the fraction of non-conformant modules is less than or equal to the acceptable quality level (AQL) $p_\alpha$. It is of low quality if p is greater than or equal to rejectable quality level (RQL) $p_\beta$ ($>p_\alpha$). Note that the proportion p is a property of the production process, and not a property of the decision rule. The AQL and RQL specify the required quality of the production process; both are typically small. The values α and β determine the error probabilities the contracting parties are willing to accept. The producer's error probability is 1-α, whereas the consumer's error probability equals β, thus determining the confidence level of the decision rule. For practical reasons the two error probabilities are related as β=1−α. Here we use α=0.9 and α=0.95.

Let A(p) denote the probability of the event that the test function T falls within the acceptance range provided that p is the actual proportion of non-conformant modules. A(p) is called the acceptance function. The requirements (a) and (b) can then be formalised as follows:

(A) $p \leq p_\alpha$ implies $A(p) \geq \alpha$
(B) $p \geq p_\beta$ implies $A(p) \leq \beta$.

Figure 2 illustrates an acceptance function satisfying the requirements (A) and (B). To guarantee the validity of (A) and (B) for a given test function, the acceptance range and the sample size have to be chosen properly.
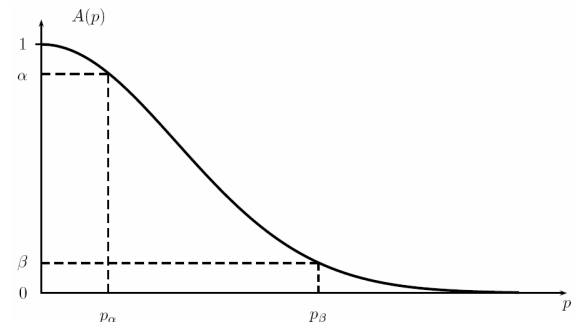


**Figure 2:** Acceptance probability of a decision rule.

### 2.2 Parameter variation program

The objective of our work was to select appropriate test functions for various scenarios, and to determine the required sample size and the corresponding acceptance range for a given confidence level and given AQL and RQL. The calculations were performed for a number of parameter sets that are listed in table 1.

| α | | $p_\alpha$ | β=1 − α | | $p_\beta$ |
|---|---|---|---|---|---|
| 0.9 | 0.95 | 0.01 to 0.05 | 0.1 | 0.05 | 0.03 to 0.1 |

**Table 1:** Investigated parameter ranges (decision rules)

The following scenarios for PV module shipments have been investigated:

I.  No flasher reports, arbitrary $P_{MAX}$ distribution
II.  No flasher reports, normal distribution of $P_{MAX}$
III.  Flasher reports, normal $P_{MAX}$ distribution
IV.  Flasher reports, arbitrary distribution of $P_{MAX}$

## 3 TEST FUNCTIONS AND MINIMUM SAMPLE SIZES FOR PARTICULAR DECISION RULES

All models assume that the nominal power output is specified in the manufacturer's data sheet by $\mu_0$ and that a single module is conformant if and only if its power output is at least $\mu_0–\varepsilon$, where $\varepsilon$ is the production tolerance; typically $\varepsilon = 0.05\ \mu_0$.

### 3.1 Scenario I

In this scenario, the consumer is assumed to have no information on the probability distribution of $P_{MAX}$. Apart from the sample size n and the number N of modules in the whole shipment, the number T of non-conformant sample modules is the only information the consumer can use in deciding on whether to accept or reject the shipment. T induces an appropriate decision rule: The shipment of PV modules will be accepted if and only if $T \le c$ for some integer c. The choice of the threshold c (i.e. of the acceptance range) and the sample size n is the crux for assuring conditions (A) and (B) in Section 2.1. It turns out that c should be chosen to be the smallest integer for which there is some integer n satisfying

$$G_{2(c+1)}(1-\beta) / (2p_\beta) \le n \le G_{2(c+1)}(1-\alpha) / (2p_\alpha);$$

in this case n provides the minimum sample size. Here $G_{2(c+1)}$ denotes the inverse of the distribution function of the $\chi^2$-distribution with $2(c+1)$ degrees of freedom. Note that our analysis is sufficiently accurate only for $N \ge 10n$.

### 3.2 Scenario II

We next assume the power outputs of the modules to be independent and identically normally distributed with unknown mean $\mu$ and unknown standard deviation $\sigma$. Intuitively, a necessary condition for accepting the shipment is that the unbiased estimate

$$\underline{x} = n^{-1}\sum_{i=1,\dots,n} x_i$$

of the true $\mu$ not fall short of $\mu_0–\varepsilon$, where $x_1,\dots,x_n$ denote the power outputs of the randomly chosen sample modules. Moreover, if the individual power outputs, and therefore $\underline{x}$, spread "much" around $\mu$, corresponding to "large" $\sigma$, then $\underline{x}$ should be "far" above $\mu_0–\varepsilon$. In fact, the size of the "safety gap" between $\underline{x}$ and $\mu_0–\varepsilon$ should be proportional to the standard deviation $\sigma/n^{1/2}$ of $\underline{x}$. Hence a shipment is acceptable if $\underline{x}$ exceeds $(\mu_0–\varepsilon) + c\cdot(\sigma/n^{1/2})$, where c is some suitable positive factor. As $\sigma$ is unknown, it will be replaced by its best estimator

$$\underline{S} = [(n-1)^{-1}\sum_{i=1,\dots,n} (x_i-\underline{x})^2]^{1/2}.$$

Hence

$$T = n^{1/2} (\underline{x}-(\mu_0–\varepsilon)) / \underline{S}$$

induces an appropriate decision rule: The shipment of PV modules will be accepted if and only if $T \ge c$. Again, the choice of the threshold c and the sample size n is the crux for assuring conditions (A) and (B) in Section 2.1. Nontrivial probability theoretical arguments yield

$$n = \lceil 4\Psi(\alpha)^2 \{1+[\Psi(p_\alpha)+\Psi(p_\beta)]^2 \} / [\Psi(p_\beta)-\Psi(p_\alpha)]^2 \rceil$$

$$c = - [\Psi(p_\alpha)+\Psi(p_\beta)] \ n^{1/2} / 2$$

where $\lceil x \rceil$ denotes the smallest integer m satisfying $x \le m$, and $\Psi$ is the inverse of the distribution function of the standard normal distribution.

### 3.3 Scenario III

In this scenario, the consumer obtains the flasher list from the producer with $P_{MAX}$–values $x_1{}',\dots,x_N{}'$ of all N modules in the shipment, such that $x_j{}' = \Pi_j + \Delta + r_j{}'$, where $\Pi_j$ is the real power output of the j-th module in the shipment, $r_j{}'$ refers to the measurement error and $\Delta$ represents a calibration error. The consumer re-measures the power outputs of n randomly chosen modules, and obtains a sample of $P_{MAX}$–values $x_1,\dots,x_n$ also subject to measurement error $r_i$ but not to a calibration error. That is, $x_i = \Pi_i + r_i$. We further assume the $x_j{}'$ (j=1,…N) as well as the $x_i$ (i=1,…,n) to be independent and $\Pi_j + r_j{}'$ and $\Pi_i + r_i = x_i$ to be $N(\mu,\sigma^2)$-distributed for all shipment indices j and all sample indices i. The crucial point is that both $x_j{}'$ and $x_i$ have the standard deviation $\sigma$. Hence $\sigma$ can be estimated by the $x_j{}'$–values from the flasher list. We again use the estimator $\underline{S}$ introduced in Section 3.2, which is now denoted by $\underline{S}'$ to emphasis the use of different input data. Since there is usually a large number of $x_j{}'$-values (compared with the number of $x_i$-values), $\underline{S}'$ can be regarded as the true standard deviation; note that $\underline{S}' \to \sigma$ as $N\to\infty$. In other words, we now have the scenario of Section 3.2 with the *known* standard deviation $\sigma (= \underline{S}')$. For the test function

$$T = n^{1/2} (\underline{x}-(\mu_0–\varepsilon)) / \underline{S}'$$

and the same decision rule as in Section 3.2 we obtain similar formulae for the acceptance threshold and the sample size:

$$n = \lceil 4\Psi(\alpha)^2 / [\Psi(p_\beta)-\Psi(p_\alpha)]^2 \rceil$$

$$c = - [\Psi(p_\alpha)+\Psi(p_\beta)] \ n^{1/2} / 2.$$

### 3.4 Scenario IV

We alter the assumptions of Section 3.3 by stating that the power outputs are not necessarily normally distributed. We allow for an arbitrary continuous $P_{MAX}$-distribution with finite second moment. For the same test function T and the same decision rule as in Section 3.3 we obtain

$$n = \lceil 4\Psi(\alpha)^2 / [F_N^{-1}(p_\beta)- F_N^{-1}(p_\alpha)]^2 \rceil$$

$$c = - [F_N^{-1}(p_\alpha)+ F_N^{-1}(p_\beta)] \ n^{1/2} / 2$$

where $F_N^{-1}$ is the inverse of the empirical distribution function of the random variables $y_j{}' = (x_j{}'-\underline{x}') / \underline{S}'$ :

$$F_N(y) = N^{-1}\sum_{j=1,\dots,N} \mathbf{1}_{(-\infty,y]} (y_j').$$

## 4 RESULTS

Figure 3 illustrates the influence of the decision rule on the sample size n (Scenario II):

- The sample size decreases with rising $p_\beta$.
- The sample size increases considerably if $p_\alpha$ is expanded.
- An improvement of the probability of acceptance for "High quality" ($\alpha$) from 0.9 to 0.95 can result in an increase of sample size by more than 100 modules.

A commonly used decision rule in statistical quality control is ($\alpha$=0.95, $p_\alpha$=0.01, $\beta$=0.05, $p_\beta$=0.03). Figure 4 shows the resulting optimal sample size for the different scenarios:

- The sample size is smallest if $P_{MAX}$ is normally distributed.
- The sample size for the empirical $P_{MAX}$ distribution of Fig. 2 (Scenario IV) differs considerably from Scenario III. This indicates that $P_{MAX}$ is not normally distributed.
- The availability of flasher reports reduces the sample size considerably. At $p_\beta$=0.03 the difference is >100 modules.
- The best case is a sample size of 55 modules (Scenario III).

Fig. 3 and 4 can also be used to infer the feasible quality claims given the sample size. Example: n=200 and ($\alpha$=0.95, $p_\alpha$=0.01, $\beta$=0.05) leads to $p_\beta$=0.05 for the blind case (Scenario I).
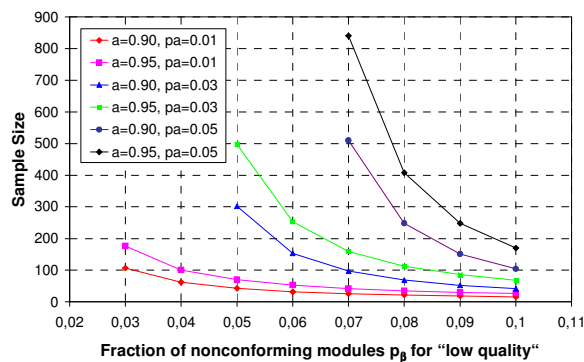


**Figure 3:** Sample size for different decision rules ($\alpha$, $p_\alpha$, $\beta$, $p_\beta$) calculated for Scenario II
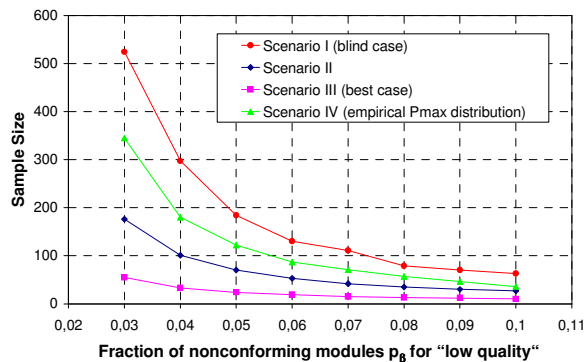


**Figure 4:** Comparison of sample sizes for different scenarios ($\alpha$=0.95, $p_\alpha$=0.01, $\beta$=0.05, $p_\beta$=variable)

## 5 CHECKING THE PROBABILITY DISTRIBUTION FOR NORMAL DISTRIBUTION

Because the optimal sample sizes for deciding whether to accept or reject a shipment are too optimistic if $P_{MAX}$ is not normally distributed, this assumption has to be checked. Many textbooks recommend a chi-square goodness-of-fit test, but then one must choose the number of classes by rule-of-thumb, which can be erroneous. In applied statistics a more common test is the Shapiro-Wilks test [3], which is implemented in many statistical software packages, e.g., SAS (PROC UNIVARIATE), R (shapiro.test), SPSS (EXAMINE MODULE), and Microsoft Excel (as an add-on).

This test checks the null hypothesis $H_0$ of normally distributed $P_{MAX}$ measurements with unspecified mean and variance. The alternative hypothesis specifies an arbitrary non-normal distribution for the measurements. The software usually reports the p-value of a statistical test, i.e., the largest significance level, which would result in a rejection of $H_0$ for the given data set. The null hypothesis is rejected if the p-value is less than the chosen significance level, e.g., 5%.

The Shapiro-Wilks test proceeds as follows: If the observations are normally distributed with mean $\mu$ and variance $\sigma^2$, the ordered values satisfy a linear regression model with slope equal to the variance $\sigma^2$, if the i-th regressor is chosen as $\Phi^{-1}((i-3/8)(n+1/4))$. The test now estimates the variance $\sigma^2$ by estimating the slope of the regression line, taking into account the correlation of the ordered measurements. That regression estimate is compared against the sample variance $s^2$ by taking the ratio. Test statistical values near 1 indicate that the distribution may be normal. The test statistics formula employs a matrix calculus and is therefore omitted here.

For the distribution of Fig. 1 the resulting p-value of the Shapiro-Wilks test is $10^{-16}$. The null hypothesis of normal distribution is therefore rejected.

## 6 CONCLUSIONS

Methods of statistical quality control can be applied in deciding whether to accept or reject a shipment consisting of a larger number of PV modules. However, the approach presupposes that a certain fraction of non-conformant modules is allowed, to be defined by a decision rule.

The optimal sample size depends on the decision rule subject to agreement between the PV module manufacturer and the customer. If flasher reports are available for inferring the $P_{MAX}$ distribution, the sample size can be reduced by up to 100 modules.

Because the sample size to accept or reject a shipment is strongly influenced by whether $P_{MAX}$ has a normal distribution, one should check this assumption on the basis of delivered flasher reports.

The definition of acceptance criteria regarding the fulfilment of power rating for small PV systems (<100 modules) requires different approaches. These questions will be considered in a follow-up study.

7 ACKNOWLEDGEMENT

REFERENCES

[1] Montgomery, D.C. *Introduction to Statistical Quality Control*. Wiley, New York (2005).
[2] Power Mark specification: Document PV-3.4 – Testing requirements for PV module power rating (2003)
[3] Shapiro, S. and Wilks, M. An analysis of variance for normality (complete samples). *Biometrika*, 52, 591-611 (1965).
[4] TÜV Rheinland specification 931/2.572.14 – *Power Controlled* certification of PV modules