# DEVELOPMENT OF A DATA ANALYSIS METHODOLOGY TO ASSESS PV SYSTEM PERFORMANCE

Odysseas Tsafarakis[1], Wilfried G.J.H.M. van Sark[2]
Utrecht University, Copernicus institute of Sustainable Development
1E: tsafod@gmail.com , 2E: w.g.j.h.m.vansark@uu.nl

Abstract
There are many possible operational problems that may occur in a PV system and the sooner they are detected and repaired, the lowest the energy loss will be. Most of the malfunctions are taking place in the inverter of a PV system [1]. The result of the majority of the inverter malfunctions is not a standard reduction on the output of the system but a variable energy loss [2]. These malfunctions, like bad Maximum Power Point Tracking MPPT or inverter outage, are very difficult to be detected using only the so-called Performance Ratio of the system, which is standard in PV performance analyses [1]. During this study, a new monitoring method was proposed in which energy yield is related to reference energy yield as scatterplot using the mean square error (MSE). Any unusual change of the linear regression implies the presence of a malfunction, and in fact our MSE method showed that inverter malfunctions and cases of partially shading can be detected, where other methods fail.

## 1. INTRODUCYION

The common impression is that the performance of a PV system is only depending on the solar radiation, while it also is depending on system (modules and inverter) characteristics and on their condition during their lifetime, in which operational malfunctions may be occurring. Clearly, the sooner they get identified and repaired, the lowest the energy loss will be [2]. Furthermore, the result of the majority of inverter's malfunctions is a variable energy loss; a fact which makes their detection very difficult [2].

As a solution to this problem, in the framework of the IEA PVPS Task 13, *"performance and reliability of PV systems"* a malfunction detection and identification method, was proposed, named as the "Stamp Collection" [3], where its principle is to plot one parameter of the PV system versus another and study their (linear) relation. Unexpected deviations of the slope of the linear regression can be observed and pinpoint the presence of a malfunction [3].

## 2. PROBLEM DEFINITION

In this study the effectiveness of this method, in the aspects of the malfunction detection and application in small residential PV systems, without complicated monitoring tools, was examined. For this reason the main plot of the "Stamp Collection" was examined, the:

*"System Yield ($Y_f$) vs Reference Yield ($Y_r$)"*

Furthermore, from a literature review and a survey with participating PV owners and installation companies revealed that the most common malfunctions are causing variable energy loss and that they are linked to the operation of the inverter [1].

However, malfunctions that are causing variable energy loss and that are linked to the operation of the inverter are not detectable through the study of Performance Ratio neither with the linear regression of the plot $Y_f$ vs $Y_r$.

## 3. PROPOSED SOLUTION

For the solution of this major problem, the use of the statistical parameter Mean Square Error (MSE) [4] in the plot: "System Yield *vs* Reference Yield" was proposed and examined in this study.

With the use of this parameter in the available monitoring data, inverter malfunctions and low partially shading were identified, while the change in the hourly Performance Ratio and the linear regression of the plot was insignificant.

## 4. THEORETICAL BACKGROUND

### System Yield & Reference Yield

System yield is defined as the AC energy of the entire PV plant, which is delivered to the grid per kilowatt peak of installed PV array [5]:

$$System\ Yield\ (Y_f) = \frac{E_{AC}}{W_{peak}}\left[{}^{Wh}/_{W}\right]$$

Reference yield is defined as the total irradiation in plane-of-array per the irradiation under standard test conditions ($1000 W/m^2$) [5]:

$$Reference\ Yield\ (Y_r) = \frac{H_{poa}}{G_{STC}}\left[\frac{{}^{Wh}/_{m^2}}{1000\ {}^{W}/_{m^2}}\right]$$

### Mean Square Error (MSE)

The mean squared error (MSE) of linear regression (LR) is one option to calculate the difference between values disguised by the LR and the right values of the measure that was predicted. [4]

$$MSE = \frac{1}{n}\sum(Y_i - \hat{Y}_i)$$

Where Ŷ are the predicted from the LR values and Y the true values (see Figure 1).

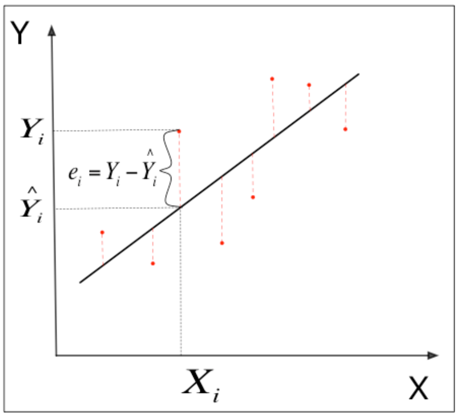
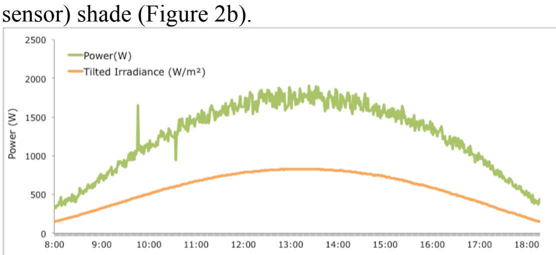



Figure 1. Illustration of mean square error method

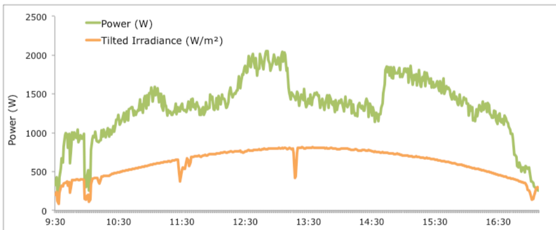
## 5. RESULTS

**Story 1-Shading Fault** [6]

The system is affected by partial shade. In Figure 2 the power of the system and the tilted irradiance are illustrated, during a clear day (Figure 2a) and a day under the influence of (undetected from the solar radiation sensor) shade (Figure 2b).




**a**




**b**

Figure 2. PV system power and irradiance for a clear day without shade (a), and with shade (b).

In Figure 3, the $Y_f$ *vs* $Y_r$ plots of these days are compared and the values of PR and LR are presented in Table I.

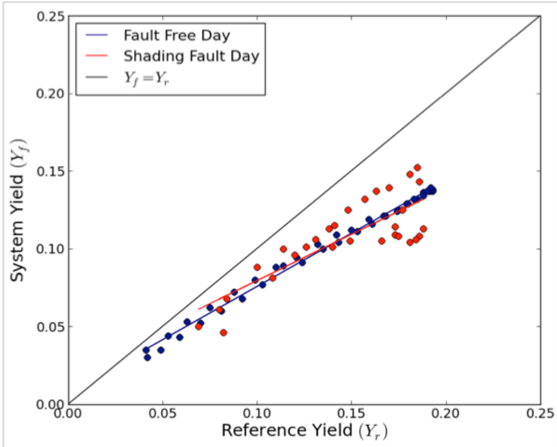



Figure 3. System Yield *vs* Reference Yield for the cases of Figure 2.

Table I. Performance ratio and linear regression slope for the cases of Figure 2

| | PR | LR |
|---|---|---|
| Fault free day | 73.7% | 0.68/34° |
| Shading Fault day | 73.1% | 0.59/31° |

While the error is noticeable through visual inspection of Figure 3, it is not detected through the study of the values in Table I. However, the ratio of the MSE of the two days clearly detects the influence of the shade since:

$$\frac{MSE\ Shading\ Fault}{MSE\ Fault\ free} = 26.75$$

**Story 2-Power Derating**

The examined system consists of two identical inverters, connected with identical PV systems, one next to the other.

During days with smooth irradiation inverter power derating causes a noticeable difference in output between these two systems.

In Figure 4 the power vs time plot of the inverters is illustrated, while in Figure 5 the *"$Y_f$ vs $Y_r$"* "Stamp" is shown.

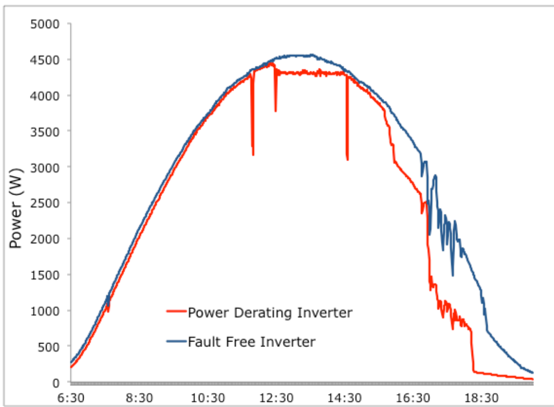



Figure 4. PV system power for fault free inverter and inverter with malfunction.

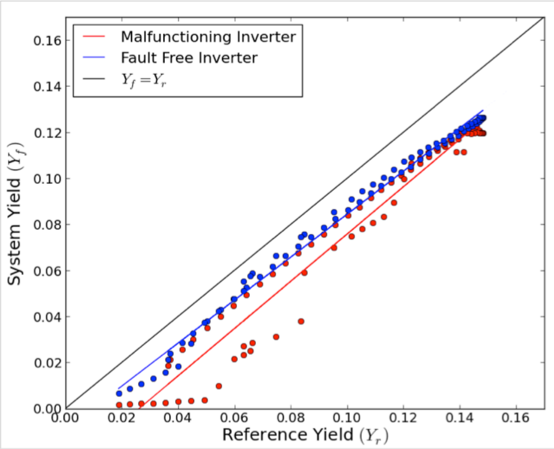



Figure 5. System Yield *vs* Reference Yield for the cases of Figure 4.

The slopes of their LR and their PR are compared in Table II.

Table II. Performance ratio and linear regression slope for the cases of Figure 2

| | PR | LR |
|---|---|---|
| Fault free System | 84.4% | 0.93/43° |
| Power Derating system | 75.8% | 1.03/46° |

Similar to story 1 the influence of the malfunction is mostly pinpointed through the comparizon of the MSE of the systems, since:

$$\frac{MSE\ Power\ Derating}{MSE\ Fault\ free} = 7$$

## 6. CONCLUSION

A statistical tool, such as MSE, as a parameter in PV monitoring can help in the detection of inverter malfunctions (story two) or cases of partially unexpected shadow (story one), which are not often detectable through the already used methods, since their influence on the system varies, according to the available solar radiation [2].

## 7. REFERENCES


[1] Tsafarakis Odysseas, Wilfried van Sark. *"Development of a data analysis methodology to assess PV system performance"*. Utrecht University; 2014.

[2] Drews A., de Keizer A.C., Beyer H.G., Lorenz E., Betcke J., van Sark W.G.J.H.M., Heydenreich W., Wiemken E., Stettler S., Toggweiler P., Bofinger S., Schneider M., Heilscher G., Heinemann D.. *"Monitoring and remote failure detection of grid-connected PV systems based on satellite observations."* Solar Energy. 2006;81(4):548-564.

[3] A.Woyte, M.Richter, D.Moser, N.Reich, M.Green, S.Mau, H.Georg Beyer. *"Analytical Monitoring of Photovoltaic Systems Good Practices for Monitoring and Performance Analysis."*. International Energy Agency. T13-03:2013.

[4] Lehmann E. L., Casella G. *"Theory of Point Estimation."* New York: Springer.

[5] N.H. Reich, B. Mueller, A. Armbruster, W.G.J.H.M. van Sark, K. Kiefer, Ch. Reise. *"Performance Ratio revisited: Are PR > 90% realistic?"* Progress in Photovoltaics. 2012;20:717-726.

[6] Silvestre S, Chouder A, Karatepe E. *"Automatic fault detection in grid connected PV systems."* Solar Energy. 2013 August;94:119-127.