Universiteit Utrecht

Faculty of Geosciences
Department Energy and Recourses
Copernicus Institute of Sustainable Development

PV system performance evaluation by clustering production data to normal and non-normal operation

Odysseas Tsafarakis[1], Kostas Sinapis[2], Wilfried G.J.H.M. van Sark[3]

[1,3]Utrecht University, Copernicus institute of Sustainable Development, Princetonlaan 8a, 3584 CB, Utrecht, The Netherlands
[2]Solar Energy Application Centre, HTC 21, 5656 AE, Eindhoven, The Netherlands
[1]E: O.Tsafarakis@uu.com, [2]E: sinapis@ecn.nl, [3]E: w.g.j.h.m.vansark@uu.nl

Motivation

The most common assessment method[1] of a photovoltaic (PV) system is by comparing its energy production to reference data (irradiance or neighboring PV system). At normal operation, these sets of data tend to show a linear relationship. Deviations from this linearity are mainly due to malfunctions occurring in the studied PV system or data input anomalies and they have to be detected to study separately for the detection of any occurred malfunction.

Research Target

To deliver an **algorithm** that:

- Automatically clusters the points of a scatterplot which tends to follow linear regression to:
  1. *Inliers – points that comply with the linearity.*
  2. *Outliers – points that do not.*

Reference Yield

$$Y_R = \frac{GTI}{1000}\left[\frac{Whr/m^2}{W/m^2}\right]$$

Error

$$\varepsilon = Y_f - Y_r$$

System Yield

$$Y_f = \frac{E_{AC}}{kWp}\left[\frac{Whr}{W}\right]$$

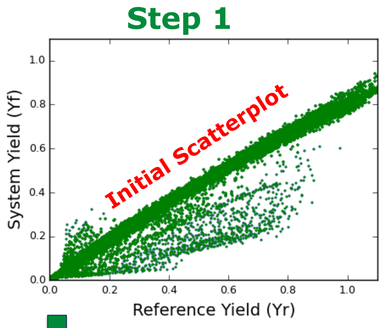
Application of the algorithm:

It is applied to the scatterplot for two main purposes:

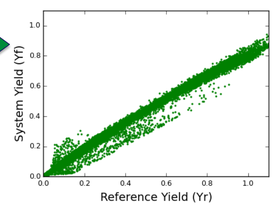
a) to detect and separate measurements where the PV system is functioning properly from the measurements that show that the photovoltaic (PV) system is malfunctioning.
b) to detect and exclude any data input anomalies, mainly due to use as reference data global horizontal irradiance converted to tilted irradiance from solar models.

Methodology

Step 1

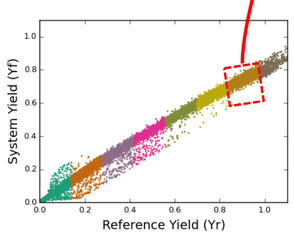


Ran.Sac.[2] filters any extreme outliers

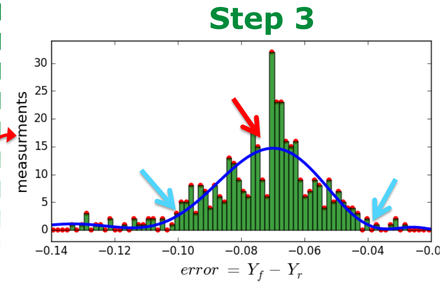


Step 2

The Sample is clustered based on the reference data level. Each cluster is studied separately in the next phase



Step 3



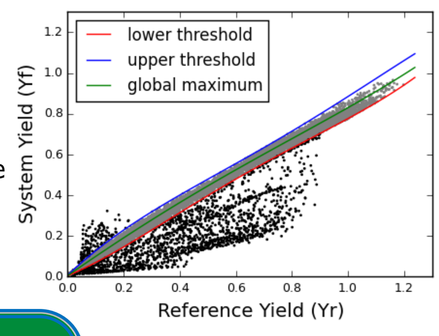
For each cluster the histogram of the errors and its polynomial fit are created in order to obtain the following three extrema:

1. **Global maximum** as the most common error
2. **Two local minima** (left & right of maximum) as the thresholds, where points within them considered as inliers

For the three different groups of extrema, their polynomial fits are calculated and solved with respect to Y_f:

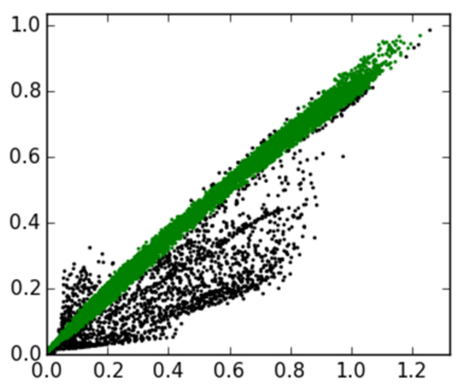
- $Y_{f,lower} = \varepsilon_{lower}^{Poly}(Y_R) + Y_R$
- $Y_{f,upper} = \varepsilon_{upper}^{Poly}(Y_R) + Y_R$
- $Y_{f,max} = \varepsilon_{max}^{Poly}(Y_R) + Y_R$

Step 4



Every measurement between the upper and lower threshold is characterized as inlier and outside as outlier

Final result



Examples

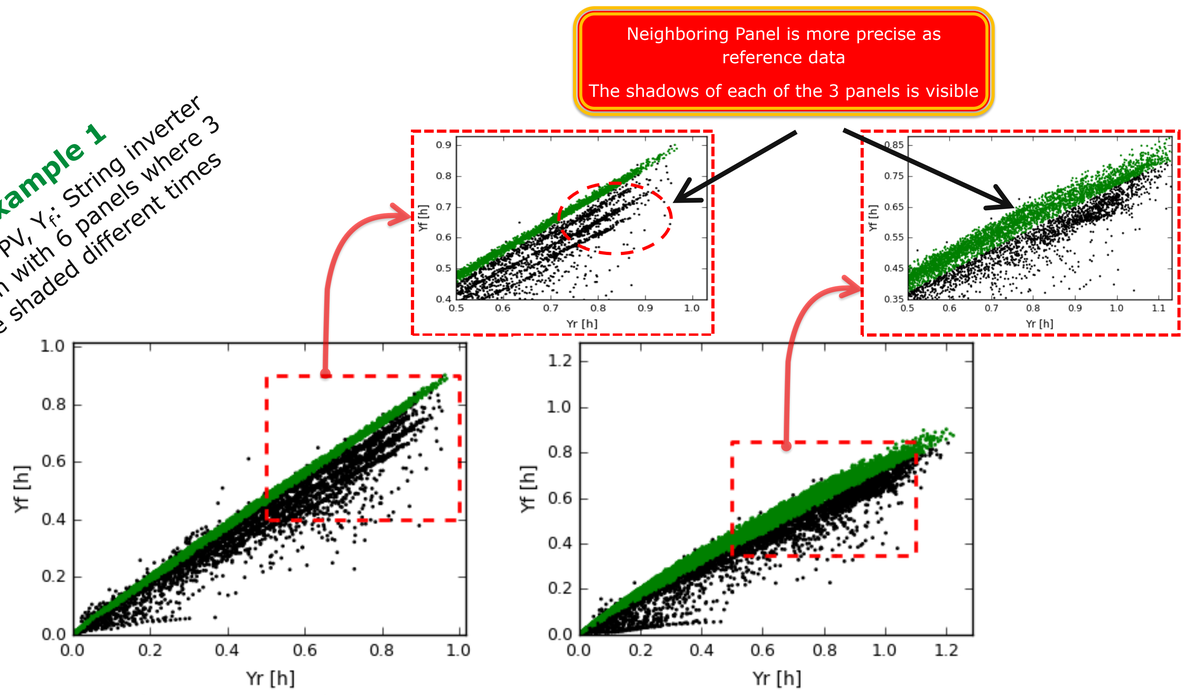
Purpose 1

Detect and separate measurements where the PV system is functioning properly from the ones that the PV system is malfunctioning.

Example 1
Studied PV, Y_f: String inverter system with 6 panels where 3 are shaded different times

Neighboring Panel is more precise as reference data

The shadows of each of the 3 panels is visible

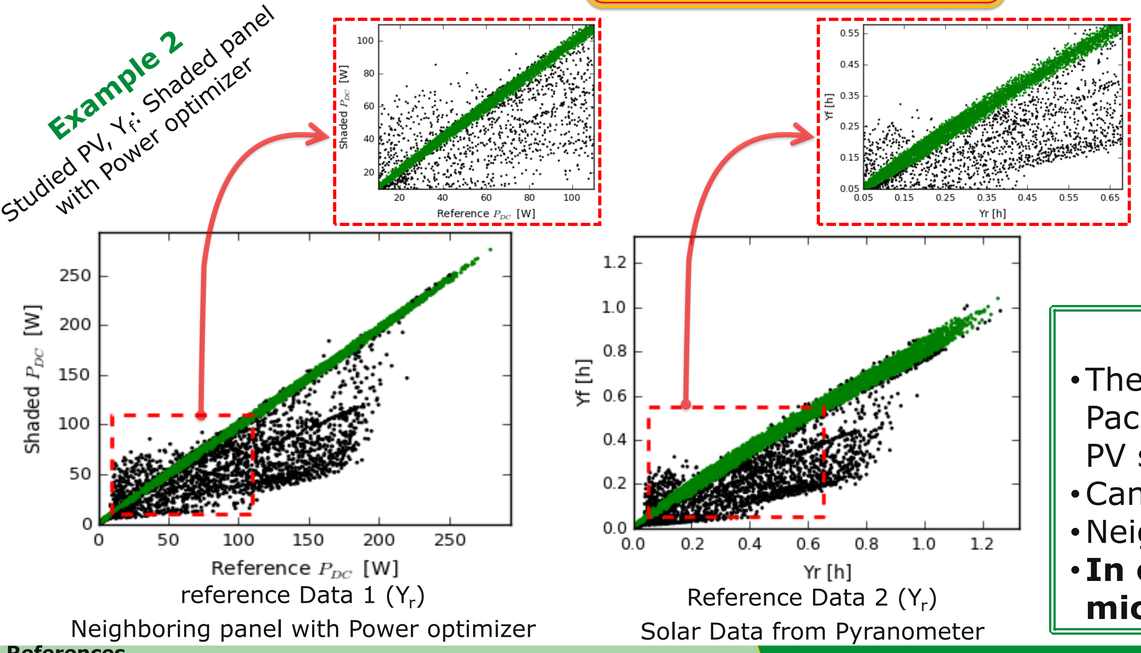


reference Data 1 (Y_r)
Neighboring panel with Power optimizer

Reference Data 2 (Y_r)
Solar Data from Pyranometer

Neighboring Panel is more precise and linear, since its production is also reduced in higher temperatures

Example 2
Studied PV, Y_f: Shaded panel with Power optimizer



reference Data 1 (Y_r)
Neighboring panel with Power optimizer

Reference Data 2 (Y_r)
Solar Data from Pyranometer

Purpose 2

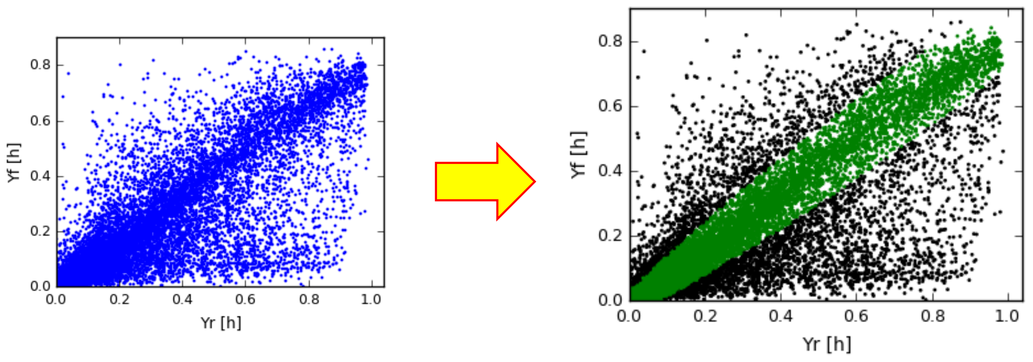
Detect and exclude any data input anomalies, manly due to use of GTI and solar models as reference data.

Examples

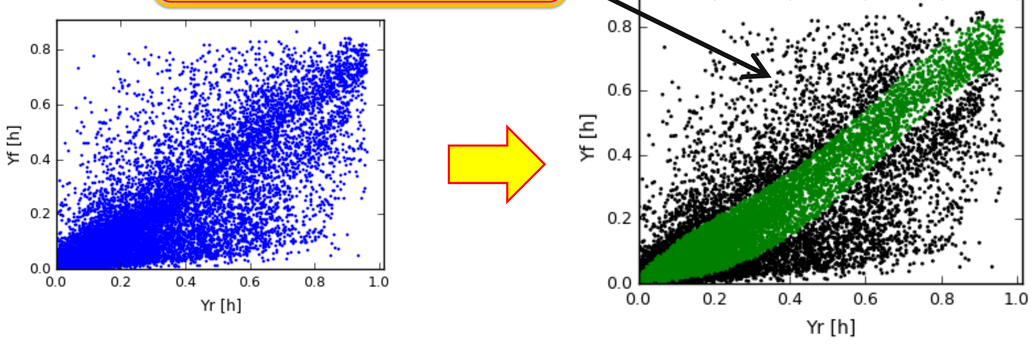
Studied PV, Y_f: PV systems on rooftops in Utrecht area

Reference Data, Y_r: Global Horizontal irradiance, obtained by local weather station and converted to tilted irradiance through the HDKR model[3]

Causes of outliers: a)Geospatial reasons – Station is far from PV Systems
b)Errors of transition models
c)Shadow



The majority of measurements tend to be linear, although a curve is formed wich is still detected by the algorithm



Conclusions

- The proposed method is offering a low cost monitoring solution, since it needs only a simple Pac data logger, to small residential PV systems, since it requires only P_{AC} of two neighboring PV systems, with similar tilt and orientation.
- Can be used as well for the comparison of any type of datasets which tend to be linear
- Neighboring panel or PV system offers more precise monitoring than a pyranometer
- **In current form it is perfect for malfunction detection and performance analysis of micro level power electronics (MLPE) systems**

References
[1] IEC 61724. Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange and Analysis, 10th ed.; International Electrotechnical Commission: Geneva, Switzerland, 1998.
[2] M. a Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981
[3] Hay, J.E. Calculating solar radiation for inclined surfaces: Practical approaches. Renew. Energy 1993, 3, 373–380.