

Statistical analysis of the performance and simulation of a two-axis tracking PV system

O. Perpiñan*

*Grupo de Sistemas Fotovoltaicos, IES-UPM, UPM,
Ciudad Universitaria s/n, 28040 Madrid, Spain[†]*

Abstract

The energy produced by a photovoltaic system over a given period can be estimated from the incident radiation at the site where the Grid Connected PV System (GCPVS) is located, assuming knowledge of certain basic features of the system under study. Due to the inherently stochastic nature of solar radiation, the question “How much energy will a GCPVS produce at this location over the next few years?” involves an exercise of prediction inevitably subjected to a degree of uncertainty. Moreover, during the life cycle of the GCPVS, another question arises: “Is the system working correctly?”. This paper proposes and examines several methods to cope with these questions. The daily performance of a PV system is simulated. This simulation and the interannual variability of both radiation and productivity are statistically analyzed. From the results several regression adjustments are obtained. This analysis is shown to be useful both for productivity prediction and performance checking exercises. Finally, a statistical analysis of the performance of a GCPVS is carried out as a detection method of malfunctioning parts of the system.

Keywords: two-axis tracking, solar radiation, performance calculation, statistical analysis

*Solar Energy, 83:11(2074–2085), doi:10.1016/j.solener.2009.08.008

[†]Electronic address: oscar.perpinan@upm.es

List of Symbols

\overline{D}	Mean difference between the reference and the simulation daily productivities
$D_d^{(i)}$	Daily difference between the productivity of the i-th group and the productivity of the reference
D_{dy}	Difference between the reference and the simulation daily productivities in the day d of the year y
\overline{D}_i	Mean of the daily difference between the productivity of the i-th group and the productivity of the reference
d_i	Distance of the individual productivity of the i-th group to the mean productivity of the system
E_{ac}	Energy produced by a grid connected PV system
$\overline{G_d(0)}$	Interannual average of the daily global horizontal irradiation
$G_{d,y}(0)$	Daily global horizontal irradiation in the day d of the year y
$\overline{G_m(0)}$	Interannual average of the monthly means of the daily global horizontal irradiation
$G_{m,y}(0)$	Monthly mean of the daily global horizontal irradiation of the month m of the year y
$\overline{G_y(0)}$	Interannual average of the yearly means of the daily global horizontal irradiation
$G_y(0)$	Yearly mean of the daily global horizontal irradiation of the year y
$\overline{G_{efd}(I)}$	Interannual average of the daily effective irradiation
$G_{efd,y}(I)$	Daily effective irradiation in the day d of the year y
$\overline{G_{efm}(I)}$	Interannual average of the monthly means of the daily effective irradiation
$G_{efm,y}(I)$	Monthly mean of the daily effective irradiation of the month m of the year y
$\overline{G_{efy}(I)}$	Interannual average of the yearly means of the daily effective irradiation
$G_{efy}(I)$	Yearly mean of the daily effective irradiation of the year y
P_g^*	Nominal power of the PV generator
\overline{rD}	Relative mean difference between the reference and the simulation daily productivities
ρ	Correlation coefficient between the reference and the simulation daily productivities
$RMSD$	Root mean squared difference between the reference and the simulation daily productivities
$rRMSD$	Relative value of the root mean squared difference between the reference and the simulation daily productivities
σ_D	Standard deviation of the difference between the reference and the simulation daily productivities

σ_{Di}	Standard deviation of the daily difference between the productivity of the i-th group and the productivity of the reference
σ_{G0d}	Standard deviation of the daily global horizontal irradiation
σ_{G0m}	Standard deviation of the monthly means of the daily global horizontal irradiation
$\overline{\sigma_{G0y}}$	Standard deviation of the interannual average of the yearly means of the daily global horizontal irradiation
σ_{G0y}	Standard deviation of the yearly means of the daily global horizontal irradiation
σ_{Gefd}	Standard deviation of the daily effective irradiation
σ_{Gefm}	Standard deviation of the monthly means of the daily effective irradiation
$\overline{\sigma_{Gefy}}$	Standard deviation of the interannual average of the yearly means of the daily effective irradiation
σ_{Gefy}	Standard deviation of the yearly means of the daily effective irradiation
σ_i^*	Difference between the standard deviation of the daily productivities of the i-th group and the standard deviation of the reference
σ_{ref}	Standard deviation of the daily productivity of the reference
σ_{sim}	Standard deviation of the simulation of daily productivity
σ_{Yfd}	Standard deviation of the daily productivity
σ_{Yfm}	Standard deviation of the monthly means of the daily productivity
$\overline{\sigma_{Yfy}}$	Standard deviation of the interannual average of the yearly means of the daily productivity
σ_{Yfy}	Standard deviation of the yearly means of the daily productivity
$\overline{Y_{fd}}$	Interannual average of the daily productivity
$\tilde{Y}_{fd,y}$	Median of the daily final productivities of a group of PV systems in the day d of the year y
$Y_{fd,y}^{(ref)}$	Daily final productivity of the reference PV system in the day d of the year y
$Y_{fd,y}^{(sim)}$	Simulation of the daily final productivity in the day d of the year y
$Y_f^{(i)}$	Productivity of the i-th group of a PV system
$\overline{Y_{fm}}$	Interannual average of the monthly means of the daily productivity
$Y_{fm,y}$	Monthly mean of the daily final productivity of the month m of the year y
$Y_{f,ref}$	Final productivity of the reference grid connected PV system
$\overline{Y_{fy}}$	Interannual average of the yearly means of the daily productivity

I. INTRODUCTION

The energy produced by a photovoltaic system over a given period can be estimated from historical values of the incident radiation at the site where the Grid Connected PV System (GCPVS) is located, assuming knowledge of certain basic features of the system under study. Due to the inherently stochastic nature of solar radiation, the question “How much energy will a GCPVS produce at this location over the next few years?” involves an exercise of prediction inevitably subjected to a degree of uncertainty. Moreover, during the life cycle of the GCPVS, another question arises: “Is the system working correctly?”.

This paper proposes and examines several methods to cope with these questions:

- The daily performance of a PV system is simulated. This simulation is statistically analyzed and validated. The results could be used both for productivity prediction during the life cycle of the system and for checking its correct operation.
- The interannual variability of both radiation and productivity is statistically analyzed. From the results several regression adjustments are to be obtained. Again, these regression lines are useful for both productivity prediction and performance checking exercises.
- Finally, a statistical analysis of the performance of a GCPVS is carried out as a detection method of malfunctioning parts of the system.

These methods will take use of the performance data of a 6.02 MWp two-axis tracking GCPVS located in Carmona (Sevilla, España; latitude =37.4°N); 225 trackers organised in 57 groups: 54 of them with 4 trackers and 3 of them with 3 trackers. Each of the trackers supports a PV generator composed of 132 Isofotón IS-207 modules which feeds a 25 kW inverter located inside the column of the tracker. In this study the analyzed performance data represent 18 groups of 4 trackers. Therefore, the nominal power of the generator of each group for our analysis is $P_g^* = 109.3 \text{ kWp}$. The daily energy production data, E_{ac} , of each group is registered by an energy meter. For the analysis, the final productivity is to be used: $Y_{f,ref} = E_{ac} / P_g^*$.

The performance of tracking PV systems has already been studied by other authors, although none of them have undertaken a statistical analysis. Several references are to be considered: Gordon and Wenger [GW91] analyze the relation between shadows, area occupation, tracker and plant geometry, limitation of tracking angle and electrical configuration of the generator; Panico et al. [PGWS91] proposes the backtracking technique for horizontal axis trackers; Macagnan and Lorenzo [Mac93] include a chapter devoted to geometrical considerations of tracking systems, the energy produced by each tracking technology and the analysis of mutual shadows; Lorenzo et al. [LPEA02] examine the geometry of shadows in an azimuthal tracking system, applying the results to the design of a PV plant; Narvarte and Lorenzo [NL08] study the tracking and shading geometry for single vertical axis, single horizontal axis and two axes, and present simulation results regarding energy production and ground cover. Detailed information about PV tracking, a large set of equations describing the movement of several types of trackers, and a method for mutual shadows calculation can be found in [Per08].

II. MODELLING OF A TWO-AXIS TRACKING PV SYSTEM: PROCEDURE AND VALIDATION

A. Calculation procedure

The source data for the performance simulation is a set of daily values of global horizontal irradiation. This database has been obtained from the meteorological station of Carmona-Tomejil (?), 20 km far away from the PV plant.

Each element of this database, $G_{d,y}(0)$, corresponds to the daily global horizontal irradiation in the day d of the year y . From each element, the daily global effective irradiation on the generator is calculated following the first six steps of Table I. The daily effective irradiation of the day d of the year y is represented with $G_{efd,y}(I)$. The final productivity of a typical system is calculated for each cell of the matrix with the last four steps of Table I. The simulation of productivity of the day d of the year y is represented with $Y_{fd,y}^{(sim)}$.

The available performance data spans from July 2007 to June 2008. The real performance data of the groups which are working correctly is the reference for the validation of the simulation procedure. This reference can be constructed filtering the data of the faulty groups (outliers in a statistical sense). Since the median is a measure of center regarded as robust against the outliers [Nav08], the daily median of the performance of the 18 groups is calculated and used as the reference ($Y_{f,dy}^{(ref)} = \tilde{Y}_{f,dy}$, where \tilde{X} stands for the median of X).

B. Validation

The comparison between the simulation, $Y_{fd,y}^{(sim)}$, and the productivity data, $Y_{fd,y}^{(ref)}$, for the period from July 2007 to June 2008, both for daily values and for monthly means of daily values, is shown in the figure 1.

The comparison of daily values shows fluctuations which become weaker with the monthly means. Moreover, the higher the global irradiation, the better the correspondence between model and data, as a consequence of the correlations between global and diffuse irradiation. Some other aspects must be remarked to understand the deviation between model and data:

- The distance between the meteorological station and the plant location imposes a difference between the radiation being measured and the radiation on the plant.
- The inverter has been modeled with three generic coefficients. For a better adjustment these coefficients should be obtained from a real performance curve fitting and considering the effect of voltage. These curves were not available when the initial estimation was carried out.
- The module parameter dispersion losses and the Joule effect losses in wiring has been modeled with constant coefficients estimated *a priori*.

The daily difference between the reference and the simulation data is $D_{dy} = Y_{fd,y}^{(sim)} - Y_{fd,y}^{(ref)}$. The mean and the standard deviation of the difference is denoted with \bar{D} and

Table I: Calculation procedure for the estimation of energy produced by a PV system from daily global horizontal irradiation data

Step	Method
Decomposition of daily global horizontal Irradiation	Correlation between diffuse fraction of horizontal radiation and clearness index proposed by Collares-Pereira and Rabl [CPR79]
Estimation of instantaneous irradiance	Ratio of global irradiance to daily global irradiation proposed by Collares-Pereira and Rabl [CPR79]
Estimation of irradiance on inclined surface	The direct irradiance is calculated with geometrical equations. The estimation of the diffuse component makes use of the anisotropic model proposed by Hay and McKay [HM85]
Albedo irradiance	Isotropic diffuse irradiance with reflection factor equal to 0,2
Effects of dirt and angle of incidence	Equations proposed by Martin and Ruiz [MR01]. A low constant dirtiness degree has been supposed (2%)
Shading effects	Only the direct irradiance is supposed to be affected by shadows. Besides, this effect is modelled as a linear relation with the shadow factor, SF_{xx} (where xx depends on the direction of the source of shadows), following the method detailed in [Per08]
$G_{efs} = D_{ef} + R_{ef} + B_{ef} \cdot (1 - SF_{xx})$	
Ambient Temperature	The ambient temperature has been modeled with the constant $T_a = 25^\circ C$.
Parameters of the PV generator	$dV_{oc}/dT_c = 0,475 \frac{\%}{^\circ C}$ $T_{ONC} = 47^\circ C$
Efficiency of the Inverter	The characteristic coefficients of the inverters [JSS92] are: $k_0^o = 0.01$, $k_1^o = 0.025$, $k_2^o = 0.05$.
Other losses	Losses due to wiring and electrical protections, dispersion of module parameters and algorithm of MPP tracking have been modeled with constant coefficients.

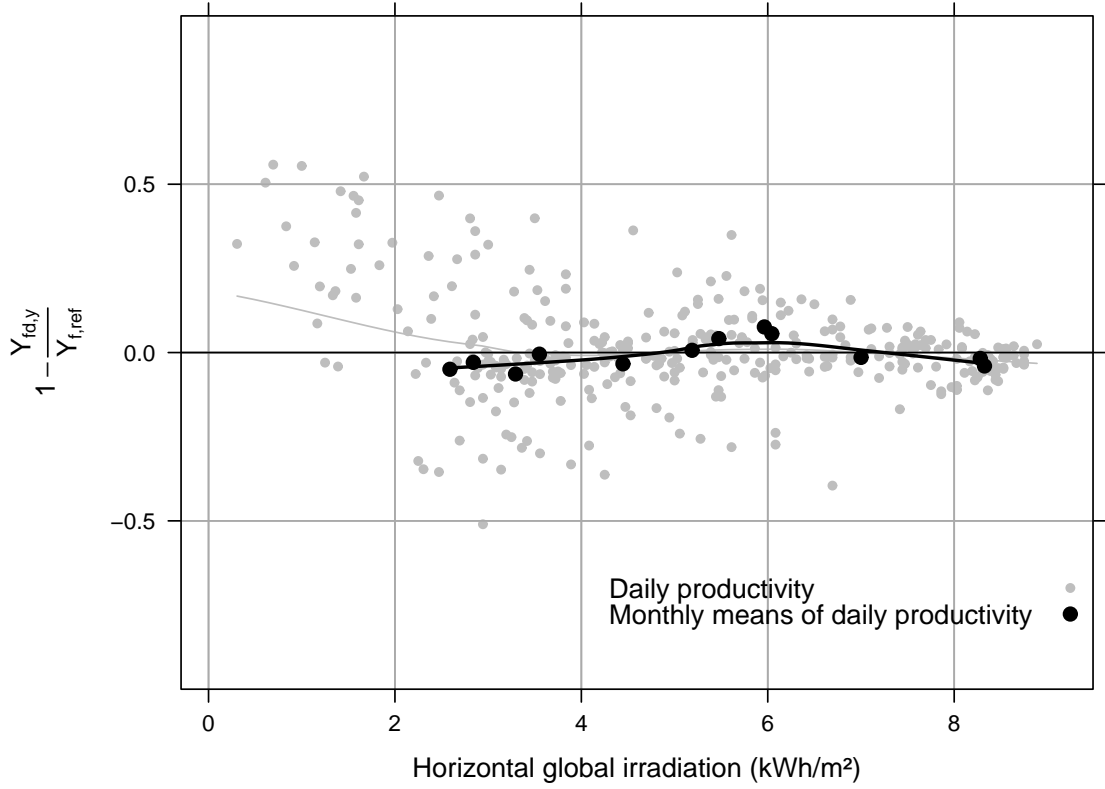


Figure 1: Relative difference between the productivity reference and the simulation for daily values and monthly means of daily productivity.

σ_D , respectively. The correspondence between the reference and the simulation can be summarized with the use of some statistics [SJM⁺09, Tay00]:

- The correlation coefficient, ρ , which gives a measure of linear relation between the reference and the model.
- The root mean squared difference, $RMSD$, which measures the size of the discrepancies between the reference and the model. It can be calculated with [Pee01]:

$$RMSD^2 = \sigma_D^2 + \bar{D}^2$$

- The standard deviation of the simulation, σ_{sim} , and the reference, σ_{ref} .

For the \bar{D} and $RMSD$ statistics, their relative value is preferred, calculated with the mean of the productivity reference: $rRMSD = RMSD / \overline{Y_f^{(ref)}}$ and $r\bar{D} = \bar{D} / \overline{Y_f^{(ref)}}$. The statistics for the whole period for daily values are summarized in Table II, while its monthly evolution is presented in the figure 2.

The set of statistics for daily values (Table II) shows a good behaviour of the simulation. The simulation gets a variance similar to the reference, a $rRMSD$ of 10% and a good correlation coefficient. The statistics for monthly means of daily values (Table III)

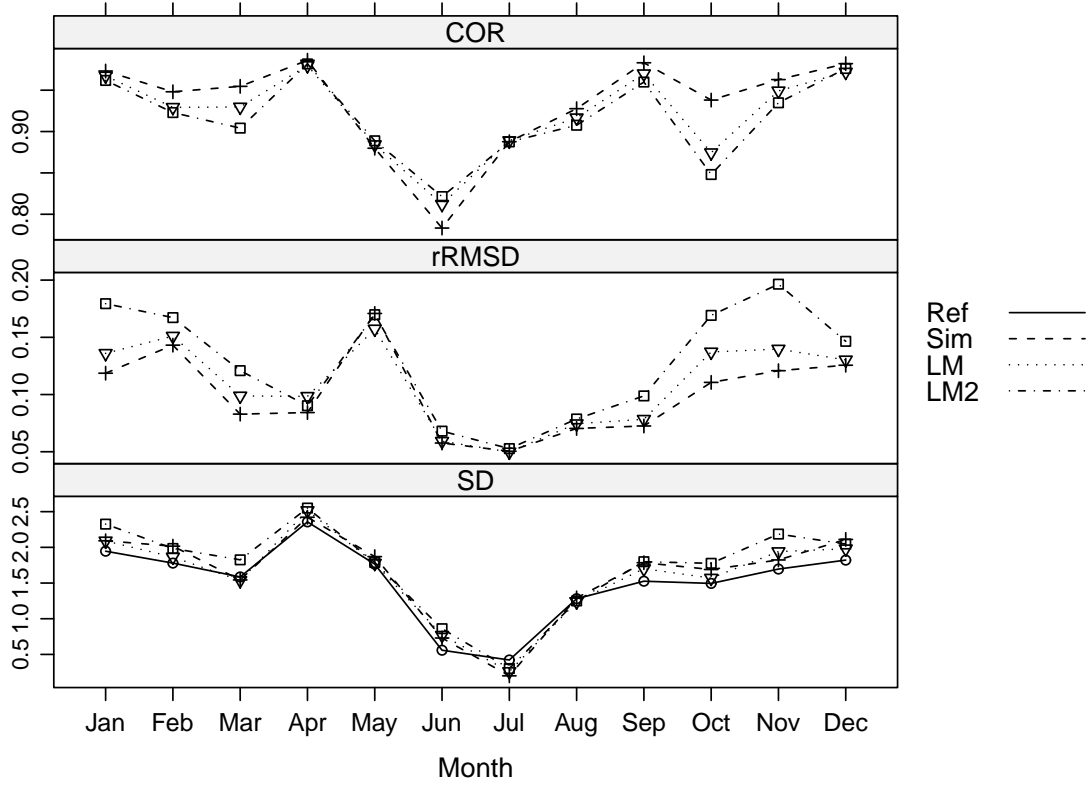


Figure 2: Monthly evolution of statistics of the difference between the productivity data (Ref) and simulation (Sim) and regression adjustments (LM: $Y_{fd,y}(I) \propto G_{d,y}(0)$ and LM2: $Y_{fd,y}/G_{d,y}(0) \propto G_{d,y}(0)$) for daily values.

evidence an improvement in the whole set of statistics with a $rRMSD$ below 5% . The yearly simulation (Table IV) deviate less than 1% from the productivity data .

The monthly evolution of these statistics (Figure 2) shows that the simulation reproduce correctly the trend of the variance of observation data.

The trend of the $rRMSD$ resembles the variance of the productivity data. A plausible explanation is that higher variances of productivity are related with low values of the clearness index, which is the region where the correlations between diffuse and global radiation produces worse results and consequently the $RMSD$ is higher.

The trend of the correlation coefficient also tracks the tendency of the variance of data. Surprisingly the correlation coefficient is worse for the months with best clearness index values. An explanation about this fact will be provided with the help of the Figure 4.

Table II: Statistics of the correspondence between the reference and simulation for daily values

Method	σ	\bar{D}	$\bar{D}/Y_f^{(ref)}$	RMSD	rRMSD	ρ
Reference	2.09					
Simulation	2.19	-0.02	-0.00	0.58	0.10	0.97
$Y_{fd,y} \propto G_{d,y}(0)$	2.16	-0.03	-0.00	0.62	0.10	0.96
$Y_{fd,y}/G_{d,y}(0) \propto G_{d,y}(0)$	2.26	-0.07	-0.01	0.72	0.12	0.95

Table III: Statistics of the correspondence between the reference and simulation for monthly means of daily values

Method	σ	\bar{D}	$\bar{D}/Y_f^{(ref)}$	RMSD	rRMSD	ρ
Reference	1.42					
Simulation	1.44	-0.03	0.00	0.26	0.04	0.98
Linear Regression	1.42	-0.06	-0.01	0.28	0.05	0.98
Rule of Three	1.44	0.06	0.01	0.26	0.04	0.98

III. STATISTICAL ANALYSIS OF THE VARIABILITY OF THE IRRADIATION AND THE PRODUCTIVITY

A. Calculation procedure

The source data for the analysis of the variability is again a set of daily values of global horizontal irradiation from the meteorological station of Carmona-Tomejil [Jun08]. For this analysis, the time period spans from September 2001 to September 2008.

This database can be arranged in a matrix of 365 rows and 8 columns (one per year), where each element $G_{d,y}(0)$ corresponds to the daily global horizontal irradiation in the day d of the year y . Since some values were not available, part of this matrix is empty. From each element, the daily global effective irradiation on the generator is calculated following the first six steps of Table I. The daily effective irradiation of the day d of the year y is represented with $G_{efd,y}(I)$. The final productivity of a typical system is calculated for each cell of the matrix with the last four steps of Table I. The productivity of the day d of the year y is represented with $Y_{fd,y}$.

From each matrix it is possible to extract a new matrix of monthly means of daily values. This new matrix consists of 12 rows and 8 columns. Each element of this matrix is represented with $G_{m,y}(0)$, $G_{efm,y}(I)$ and $Y_{fm,y}$, respectively. Besides, a vector of 8 yearly means of daily values can be constructed from each original matrix, where $G_y(0)$, $G_{efy}(I)$

Table IV: Statistics of the correspondence between the reference and simulation for yearly means of daily values

Method	Y_f	\bar{D}	$\bar{D}/Y_f^{(ref)}$
Reference	2 169.16		
Simulation	2 178.06	8.90	0.004
Linear Regression	2 171.62	2.46	0.001
Rule of Three	2 132.31	-36.85	-0.02

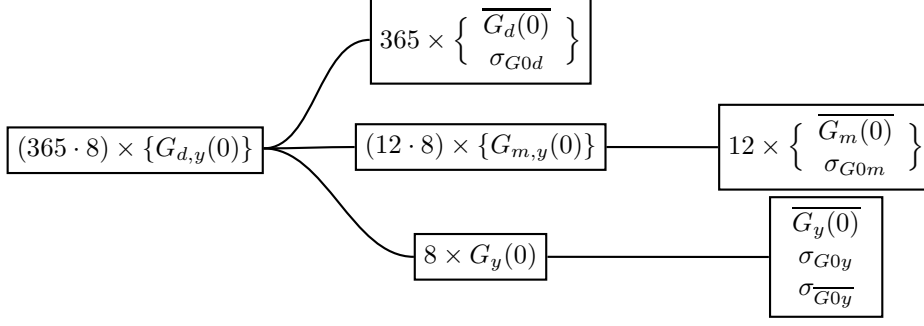


Figure 3: Sequence of transformations from daily values of global horizontal irradiation to monthly and yearly means of radiation values.

and Y_{fy} stand for each element of the three vectors.

From these matrices and vectors, three different classes of time averages can be calculated:

- The interannual average of the daily values: each of the 365 elements of this vector is represented with $\overline{G_d(0)}$, $\overline{G_{efd}(I)}$, $\overline{Y_{fd}}$ and calculated with $\overline{X_d} = \frac{1}{8} \sum_{y=1}^8 X_{d,y}$. A vector of standard deviation values is connected to each of these vectors, with its elements represented with σ_{G0d} , σ_{Gefd} and σ_{Yfd} .
- The interannual average of the monthly means: each of the 12 elements of this vector is represented with $\overline{G_m(0)}$, $\overline{G_{efm}(I)}$, $\overline{Y_{fm}}$ and calculated with $\overline{X_m} = \frac{1}{8} \sum_{y=1}^8 X_{m,y}$. A vector of standard deviation values is connected to each of this vector, with its elements represented with σ_{G0m} , σ_{Gefm} and σ_{Yfm} .
- The interannual average of the yearly means, represented with $\overline{G_y(0)}$, $\overline{G_{efy}(I)}$, $\overline{Y_{fy}}$. The standard deviation associated to this average is represented with σ_{G0y} , σ_{Gefy} and σ_{Yfy} .

Moreover, the standard deviation of the mean (SDOM) is of interest. The three vectors $G_y(0)$, $G_{efy}(I)$ and Y_{fy} , which belong to a limited time period, can be supposed to be a sample of the population of the yearly productivities during the N years of the life cycle of the system. It is then possible to estimate the mean and standard deviation of the population with the sample mean and standard deviation. Therefore, the mean productivity during the N years of the life cycle of the system can be quantified with the mean $\overline{Y_{fy}}$ and the SDOM, $\sigma_{\overline{Y_{fy}}}$, where $\sigma_{\overline{Y_{fy}}} = \sigma_{Y_{fy}} / \sqrt{N}$. The SDOM will be represented with $\sigma_{\overline{G0y}}$, $\sigma_{\overline{Gefy}}$, $\sigma_{\overline{Y_{fy}}}$.

This sequence of transformations is summarised in Figure 3 for the global horizontal irradiation.

B. Interannual variability of the global irradiation

The interannual variability or relative uncertainty is calculated with: $\delta_X = \sigma_X / \overline{X}$. For the set of data of this station, the daily ($\sigma_{G0d} / \overline{G_d(0)}$), monthly ($\sigma_{G0m} / \overline{G_m(0)}$) and annual

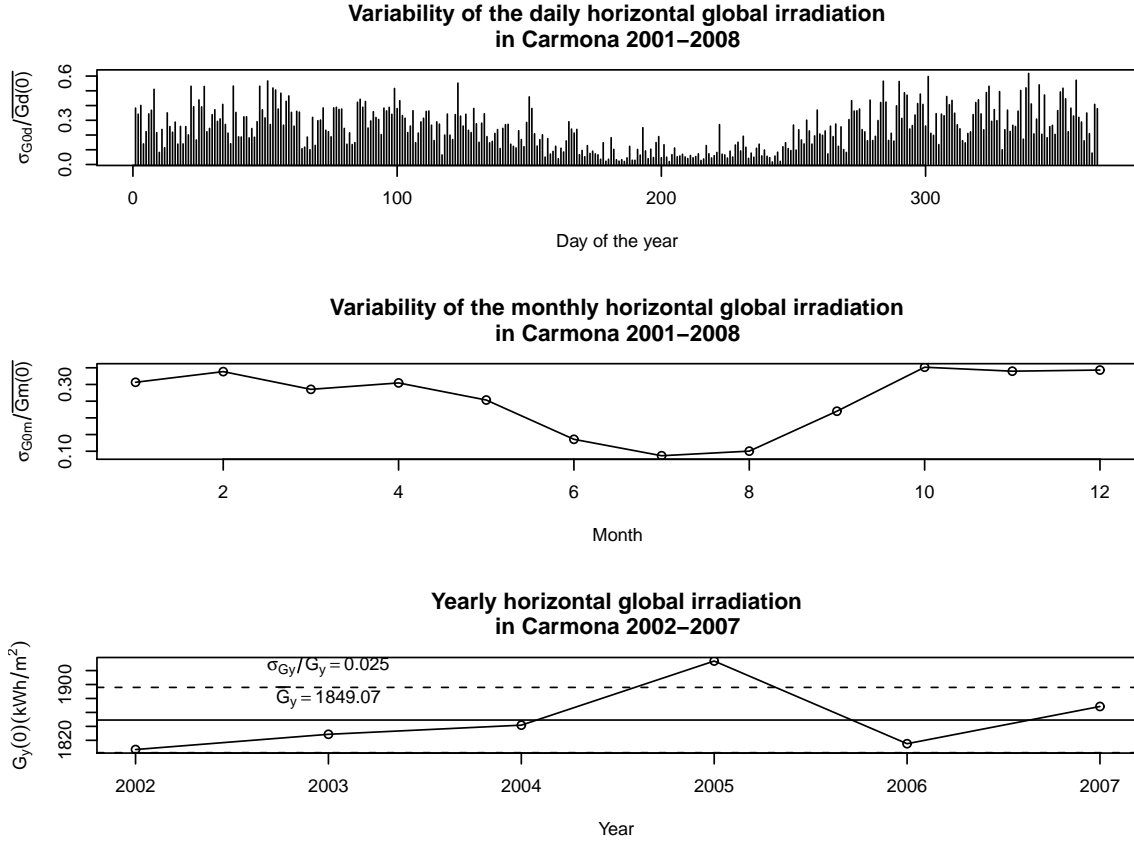


Figure 4: Variability of daily, monthly and yearly horizontal global irradiation during the period 2001-2008.

$(\sigma_{G_{0y}}/\overline{G_y(0)})$ variability of the global horizontal irradiation during the period 2001-2008 is shown in Figure 4:

The daily variability of the global horizontal irradiation ranges from 0.6 in winter to 0.1 in summer. The monthly values range from 0.35 in winter to 0.1 in summer. The yearly variability for the period 2002-2007 (2001 and 2008 are excluded since they are incomplete years) is 2.5%. These results are similar to those published by PVGIS ([SHD⁺07]).

Since the effective irradiation is a function of the horizontal irradiation, its variability can be derived following the theory of error propagation [Nav08]:

$$\delta_{G_{eff}} = \frac{\overline{G(0)}}{G_{ef}(I)} \cdot \left| \frac{dG_{ef}(I)}{dG(0)} \right| \cdot \delta_{G(0)}$$

In Figure 5 the ratio between *daily* effective incident and horizontal irradiation versus horizontal irradiation is displayed for each month of the year. It is interesting to observe that the ratio $G_{efd,y}/G_{d,y}(0)$ and the slope of the regression line $G_{efd,y}(I)/G_{d,y}(0) = m \cdot G_{d,y}(0) + n$ are higher for winter than for summer. Here, from Figure 2 we remember that the correlation coefficient in the validation results was worse for the months with best clearness index values. This may be explained if the figure 4 is connected to the fact that the correlation coefficient only quantifies the *linear* relation between model and data. In months with moderate clearness index values the productivity of a PV system spans

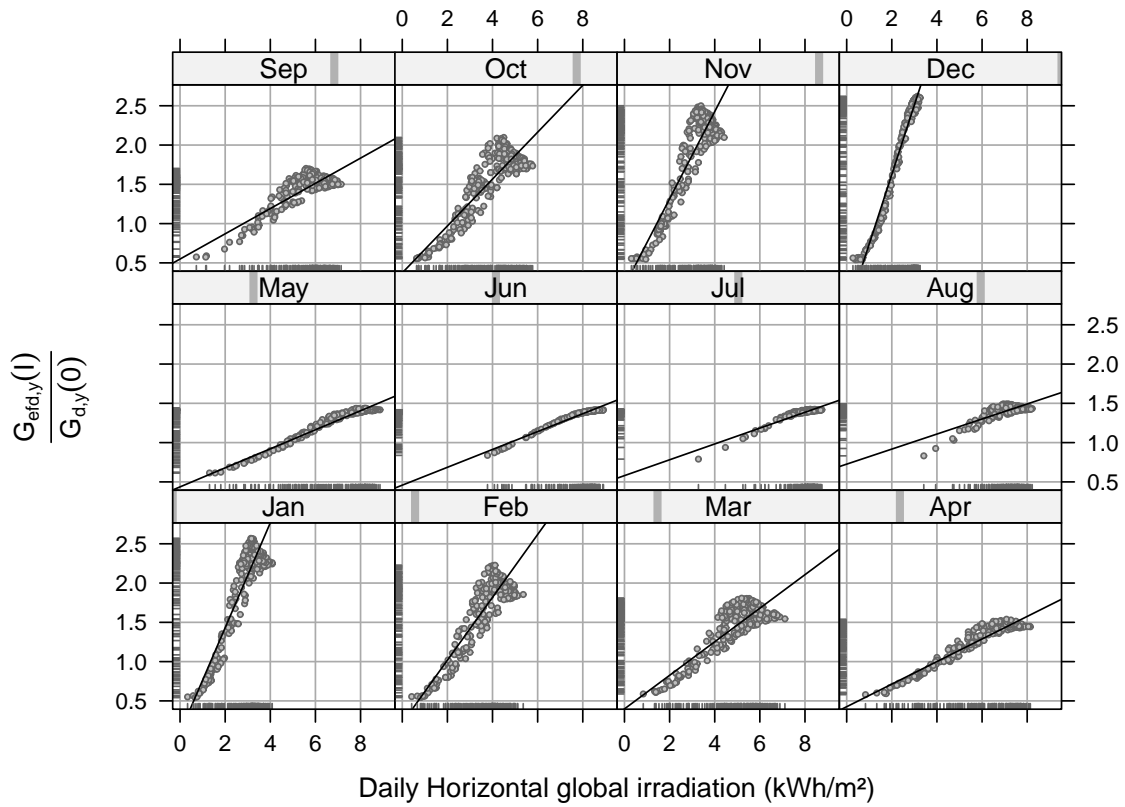


Figure 5: Relation $G_{efd,y}(I)/G_{d,y}(0) = m \cdot G_{d,y}(0) + n$ between daily horizontal and effective radiation for each month during the period 2001-2008

over a large range. However, on months with high clearness index values, the cloud of values is concentrated in a smaller region where a linear relation could be more difficult to find.

The global effective irradiation is a nonlinear function of the horizontal irradiation, so the derivative included in the last equation depends on the value of the horizontal irradiation. For this location and for a two-axis tracker, the absolute value of this derivative is greater than 1 during the whole set period. Therefore, the variability of the effective irradiation is higher than the variability of horizontal irradiation. The ratio between the variability of horizontal and effective incident irradiation versus the mean of the daily horizontal irradiation is shown in Figure 6. This ratio is approximately linear in a range comprised between 1 and 2.5, decreasing with increasing horizontal irradiation.

In Figure 7 the ratio between *monthly* effective incident and horizontal irradiation versus horizontal irradiation is displayed for each month of the year. Once again the ratio $G_{efm,y}/G_{m,y}(0)$ and the slope of the line $G_{efm,y}(I)/G_{m,y}(0) = m \cdot G_{m,y}(0) + n$ is larger for winter than for summer.

It is worth to note that, although we are analysing estimated effective irradiation, the described relations are not due to the calculation procedure. This same analysis has been applied to radiation data from the HELIOS-IES meteorological station (<http://helios.ies-def.upm.es/>). In this database both daily global horizontal irradiation and global irradiation incident on a fixed plane are available. The nonlinear relation be-

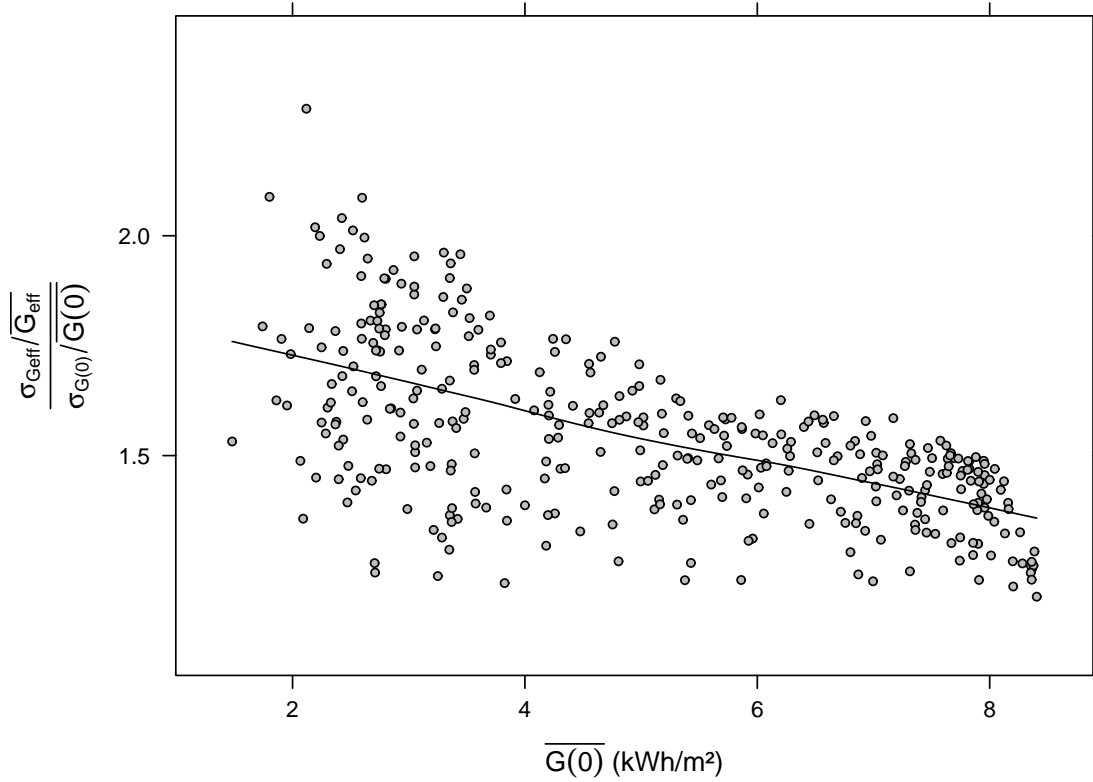


Figure 6: Ratio of variability of daily horizontal and effective global irradiation during the period 2001-2008

tween horizontal and incident irradiation is found again. The derivatives for a fixed system and for a two-axis tracking system are different, but their absolute values are greater than unity in both cases for almost the whole set of data. Consequently, the variability of the incident irradiation is higher than the variability of the horizontal irradiation.

C. Interannual variability of the productivity of a two-axis tracking PV system

This analysis can also be applied to the final productivity. The relation between daily final productivity and effective irradiation can be assumed as quasi-linear ([PLC07]), so the derivative of the equation of variability is approximately constant (in a range of values from 0.7 to 0.8). Although this fact reduces the amplification of the variability due to the transformation from horizontal to effective irradiation, the variability of the final productivity is higher than the variability of the horizontal irradiation.

When a performance prediction for the life cycle of the system is needed, the average of yearly means ($\overline{Y_{fy}}$) is of common usage. In this case, the uncertainty due to the variability of radiation can be quantified with the SDOM, $\sigma_{\overline{Y_{fy}}}$. However, if the productivity is to be checked each year, then the standard deviation $\sigma_{Y_{fy}}$ has to be used.

When a daily performance analysis is carried out, the vector of averages of daily val-

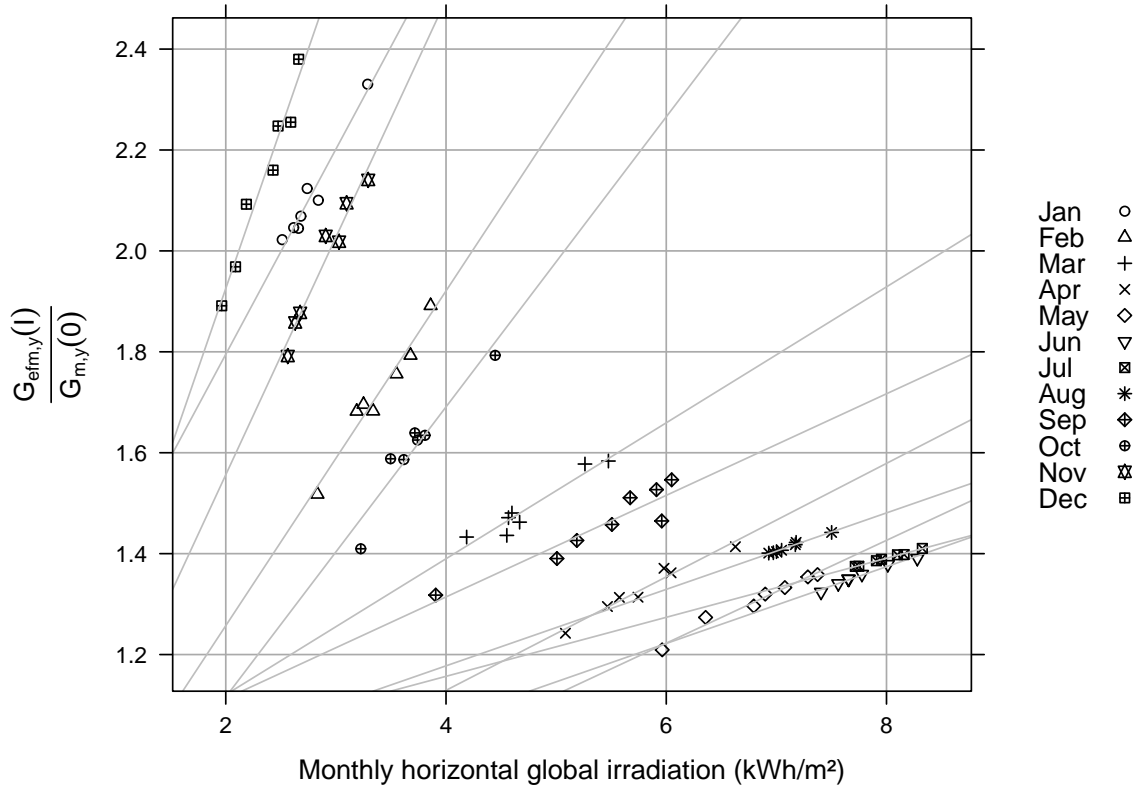


Figure 7: Relation $G_{efm,y}(I)/G_{m,y}(0) = m \cdot G_{m,y}(0) + n$ between monthly horizontal and effective radiation during the period 2001-2008

ues ($\overline{Y_{fd}}$ and $\sigma_{\overline{Y_{fd}}}$) is the appropriate data. This information is seldom reported and instead the vector of averages of monthly values ($\overline{Y_{fm}}$) is supplied. Due to the variability of radiation and productivity, this vector has to be corrected with the current daily global horizontal radiation. Although it is a customary practice, it is not advisable to check the productivity of a system by means of a rule of three between current horizontal irradiation and the ratio of estimated productivity and mean horizontal irradiation. As it will be detailed in the next section, it is preferable to use the 12 regression lines $Y_{fd,y}/G_{d,y}(0) = m \cdot G_{d,y}(0) + n$, equivalent to those shown in the Section 5.

It shall be remarked that the uncertainty due to the variability of the radiation is only one of the sources of uncertainty in a productivity prediction. There are several additional sources of uncertainty (for example, the radiation measurement errors) that should be included in the calculation of the total standard deviation [Lor03].

D. Productivity calculation with regression adjustments

The approach described in section II does necessarily make use of a computer, which every day has to simulate the performance of the system from the daily global horizontal irradiation. In order to reduce the dependence from the computer, it is sensible to benefit from the variability analysis. In Figure 5 the regression lines $G_{efd,y}(I)/G_{d,y}(0) = m \cdot$

Table V: Coefficients of the regression adjustments $Y_{fd,y}/G_{d,y}(0) \propto G_{d,y}(0)$ and $Y_{fd,y} \propto G_{d,y}(0)$ for daily values, and $Y_{fm,y} \propto G_{m,y}(0)$ for monthly means of daily values.

Month	Daily values				Monthly means	
	$Y_{fd,y}/G_{d,y}(0) \propto G_{d,y}(0)$		$Y_{fd,y} \propto G_{d,y}(0)$		$Y_{fm,y} \propto G_{m,y}(0)$	
	n	m	n	m	n	m
1	0.15	0.47	-1.97	2.27	-2.20	2.35
2	0.23	0.28	-1.83	1.82	-2.39	1.98
3	0.37	0.14	-1.81	1.48	-1.86	1.49
4	0.39	0.09	-1.92	1.32	-2.19	1.36
5	0.40	0.08	-2.17	1.29	-2.64	1.35
6	0.45	0.07	-2.85	1.37	-2.67	1.34
7	0.54	0.06	-2.38	1.31	-1.81	1.24
8	0.64	0.05	-1.83	1.29	-1.90	1.30
9	0.49	0.10	-1.63	1.38	-1.47	1.35
10	0.34	0.20	-1.57	1.62	-2.65	1.91
11	0.22	0.39	-1.77	2.08	-2.53	2.34
12	-0.05	0.63	-2.05	2.48	-2.25	2.56

$G_{d,y}(0) + n$ are shown for each month. The correspondent regression lines for $Y_{fd,y}$ can be extracted from the same calculation exercise. Therefore, with the current value of global horizontal irradiation, these equations provide the estimation of productivity of the system.

As previously explained, the relation between effective incident and horizontal irradiation is non-linear. However, it is interesting to study the results obtained with a regression between $Y_{fd,y}$ and $G_{d,y}(0)$, instead of $Y_{fd,y}/G_{d,y}(0)$ and $G_{d,y}(0)$. The Table V contains the coefficients of the 12 regression lines for each approach. This table also includes the coefficients of the regression of monthly means $Y_{fm,y} \propto G_{m,y}(0)$. It is interesting to note the similarity between the slopes of the regression lines of daily values ($Y_{fd,y} \propto G_{d,y}(0)$) and monthly means ($Y_{fm,y} \propto G_{m,y}(0)$).

For monthly and yearly values, the designer of the system usually provides the estimation of productivity related to the reference of global horizontal irradiation. In order to check the performance of the system, it is of common usage a rule of three to correct the estimated productivity with the current monthly or yearly means of horizontal irradiation. This approach has also been analyzed in this comparison.

The set of statistics for daily values (Table II) shows a good behaviour of the two regression adjustments. The variance is similar to the reference, the $rRMSD$ is around 10% and the correlation coefficient is acceptable. The statistics for monthly means of daily values (Table III) evidence again an improvement in the whole set of statistics with a $rRMSD$ around 5%, both for the regression adjustment and the rule of three. The yearly calculation (Table IV) deviate less than 1% from the productivity data when using the regression adjustment and around 2% when using the rule of three.

The monthly evolution of these statistics (Figure 2) shows that the regression lines correctly track the trend of the variance of observation data.

IV. STATISTICAL ANALYSIS OF THE PERFORMANCE OF A TWO-AXIS TRACKING PV SYSTEM

Since the 18 groups of the 6 MWp PV system are theoretically identical, their performance along the time should be the same. Due to their practical differences –power tolerance, dispersion losses, dust–, the individual performance of each group will deviate from the average behaviour. However, when a group is performing correctly, these deviations are constrained inside a range and should not be regarded as sign of malfunctioning.

If these common deviations are assumed as a random process, a statistical analysis of the performance of the whole set of groups can identify a faulty group as the one that departs significantly from the mean behaviour. This identification can be carried out by graphical analysis or with a numerical criterion. Moreover, it is possible to analyse only the performance during the current day, or the trend of the groups during a time period.

A. Current day approach

In statistical theory, there are several methods of assessing whether one piece of experimental data from a set of observations is likely to be an outlier. It is worth to mention the graphical method with boxplots based in the interquartile range [Cle93], the Chauvenet’s criterion [Tay97], the Pierce’s criterion [Ros03] or even an application of the Chebyshev theorem [AFC05]. Both the Chauvenet’s and Pierce’s criterion assume the measurements are governed by a gaussian distribution. However, the Chebyshev theorem is not dependent upon any knowing how the data is distributed. In our case, the premise of a gaussian distribution is reasonable, and consequently Chauvenet’s and Pierce’s criterion are applicable.

In Chauvenet’s method, there is an arbitrary assumption that a measurement may be rejected if the probability of obtaining the deviation from the mean for that value is less than the inverse of twice the number of measurements. The theoretical development of Peirce’s criterion does not make any such assumption. In addition, Chauvenet’s criterion makes no distinction between the case of one or several suspicious data values, whereas Peirce’s criterion can be applied in the case of several suspicious data values using a table included in the previous reference. However, due to its ease of presentation, the application of the Chauvenet’s criterion to our PV system is described here.

1. For the current day, the mean, \overline{Y}_f , and the standard deviation, σ_{Y_f} , of the productivity of the set of the 18 groups is calculated.
2. With the daily mean and standard deviation, the distance of the individual performance of each group, $Y_f^{(i)}$, to the mean performance of the set is calculated with:

$$d_i = \frac{Y_f^{(i)} - \overline{Y}_f}{\sigma_{Y_f}}$$

3. In this method, it is assumed that a measurement may be rejected if the probability of obtaining the deviation from the mean for that value is less than the inverse of

twice the number of measurements. For $N = 18$, this probability is 0.0277. In a cumulative normal distribution (z-score table) this probability corresponds to $d_i = -1.92$ (only negative values of d_i are of interest). Therefore, a group can be detected as faulty if its relative individual performance is:

$$\frac{Y_f^{(i)}}{\overline{Y_f}} < 1 - 1.92 \cdot \frac{\sigma_{Y_f}}{\overline{Y_f}}$$

In the time period from July 2007 to June 2008, the Chauvenet's criterion produces 249 identifications (3,8%) from a total of 6570 observations (18 groups, 365 days). During this period the group failures were not systematically registered. Therefore, it was not possible to quantify the number of false positives and missing negatives of this criterion. However, a random application of this procedure correctly detected faulty groups.

The ratio $\sigma_{Y_f}/\overline{Y_f}$ during this period ranged from 1.1% in summer to 5.8% in winter. Therefore, this method can detect a faulty group of this system if its individual performance deviation from the group is higher than 2.12% in summer and 11.1% in winter.

B. Time period approach

In order to take in consideration the trend of past days, we construct a daily reference ($Y_{f,d}^{(ref)}$) and compare the daily performance of each group ($Y_{f,d}^{(i)}$) with this reference during a time period of N_d days preceding the current day. The real performance data of those groups which are working correctly is a useful reference which can be constructed filtering the data of those groups detected as faulty. Again, the daily median of the performance of the 18 groups is calculated and used as the reference ($Y_{f,d}^{(ref)} = \tilde{Y}_{f,d}$).

The daily difference between the reference and the individual data is $D_d^{(i)} = Y_{f,d}^{(i)} - Y_{f,d}^{(ref)}$ (thus, a negative value of $D_d^{(i)}$ means that the i th group produced below the reference during the day d). The mean and the standard deviation of the difference for the i th group during a period of N_d days is denoted with \overline{D}_i and σ_{D_i} .

For this analysis, the statistics to be calculated for each group are: the correlation coefficient, ρ_i ; the root mean square difference, $RMSD_i$; the standard deviation of individual performances, σ_i , and reference, σ_{ref} ; the difference of standard deviations defined as $\sigma_i^* = \sigma_i - \sigma_{ref}$.

When several comparisons are to be carried out (18 in this analysis), this set of statistics can be summarised in graphical tools taking advantage of the mathematical relations between them [JKS⁺09, Tay00]. In our study we will work with the "target" diagram, which plots together the root mean square difference, the average difference and the standard deviation of the difference. Besides, this diagram includes the sign of the difference of standard deviations.

The Figure 8 includes a target diagram where the reference and individual performances are compared during four different time periods ($N_d \in \{5; 10; 20; 30\}$ in this example). The mean difference serves as the y-axis, and the standard deviation of the difference as x-axis. The distance between the origin and a point of the diagram is the $RMSD$ of that point. Therefore, the circles are the set of points with the same $RMSD$. In

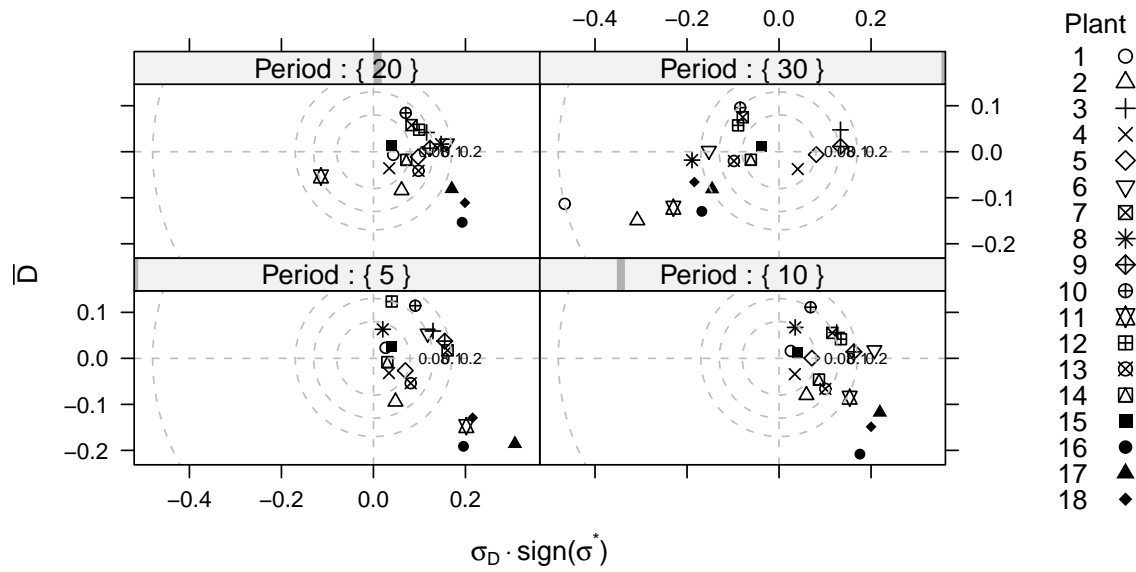


Figure 8: Target diagram for the productivity reference and the individual performances during different time periods. The mean difference serves as the y-axis, and the standard deviation of the difference as x-axis. The circles are the set of points with the same RMSD. Four circles are drawn: the first quartile, the median and the third quartile of the RMSDs of the group and another circle equivalent to the largest RMSD value.

this diagram four circles are drawn: the first quartile, the median and the third quartile of the RMSDs of the group and another circle equivalent to the largest RMSD value.

Since σ_{Di} is always positive, the negative region of the Cartesian coordinate space, $X < 0$, may be utilized if this standard deviation is multiplied by the sign of difference of standard deviations (σ_i^*). Thus, the resulting target diagram provides information about whether the standard deviation is larger ($X > 0$) or smaller ($X < 0$) than the reference standard deviation, in addition to a positive ($Y > 0$) or negative ($Y < 0$) mean difference.

In the figure 8, it is clear that the groups no. 11, 16, 17, and 18 were producing below the group during the 5 days previous to the day in analysis. Their \bar{D} is negative and their RMSD is above the 3rd quartile of the RMSDs of the group. This result partly agrees with the Chauvenet's criterion, which produces 1 detection for the groups no. 11, 16, and 17, but none for group no. 18. During the 10 days period, the groups no. 16, 17, and 18 are clearly performing below the group, while the group no. 11 has improved its result. The Chauvenet's criterion produces 1 detection for this group during this 10 days period, and 3, 1 and 2 detections for groups no. 16, 17 and 18, respectively. This same analysis is applicable to the rest of groups and periods.

V. CONCLUSION

The statistical comparison between the data and the simulation of the performance of a PV system has shown acceptable results in a daily, monthly and yearly basis. Moreover, this simulation can be condensed by means of regression lines. The statistical comparison between the performance data and the regression lines has also proved to be acceptable. Therefore, both the simulation and the regression adjustments can cope with the question about the energy to be produced by a GCPVS. The interannual variability of the horizontal irradiation, and the amplification of this variability in the productivity results, imposes limits to this prediction exercise.

When this calculated performance is to be used as a reference for checking the operation of a PV system, the uncertainty associated to simulation and to the variability of the radiation restricts its applicability. As it has been shown in Table II, the $rRMSD$ for daily values is about 10%. Therefore, in general it is not possible to reckon a group as faulty if its daily performance is above the 80% of the simulation reference. However, this $rRMSD$ is below 5% during the summer months. During this period, it could be possible to observe defects below the 90% of the simulation values. This minimum value of $rRMSD$ is similar to the value obtained in a monthly basis. Hence, a weekly or even a monthly analysis can detect problems which are hidden in a daily checking [DdKB⁺07]. As previously shown, those days with low irradiation levels do not contribute positively to this routine of detection, so the period should be enlarged if those days are found. It is important to note that the degree of the problem to be detected is directly related with the period of analysis. It is possible to mark as faulty a group deeply defective in a short period of time, but for a small deviation it is necessary to devote more time.

On the other way, it has been shown that it is possible to analyze and decide about a faulty group of a PV system with statistical methods without irradiation information. The current day approach takes advantage of statistical methods for the detection of outliers. The limits of these methods are related to the standard deviation of the individual performances, which is higher during winter months (about 5%) than during summer months (around 1%). The time period approach plots several statistics with summary diagrams in order to study the evolution of each groups in the context of the whole system during one or several time periods. In this manner it is possible to identify if the problem is located in a particular day or period or if it is maintained over time. These patterns can help to discriminate the cause of the problem, as shown in [DdKB⁺07].

Acknowledgments

I thank Prof. Eduardo Lorenzo for his fruitful discussions.

The simulation and statistical analysis were carried out with the R-project free software ([R D08]). The statistical graphics were generated with the package lattice ([Sar09]) of the R-project.

[AFC05] AMIDAN, B. G., FERRYMAN, T. A., and COOLEY, S. K.: *Data outlier detection using the chebyshev theorem*. Aerospace Conference IEEE, pages 3814–3819, March 2005.

- [Cle93] CLEVELAND, W. S.: *Visualizing Data*. Hobart Press, Summit, New Jersey, U.S.A., 1993.
- [CPR79] COLLARES-PEREIRA, M. and RABL, ARI: *The average distribution of solar radiation: correlations between diffuse and hemispherical and between daily and hourly insolation values*. *Solar Energy*, 22:155–164, 1979.
- [DdKB⁺07] DREWS, A., KEIZER, A. C. DE, BEYER, H. G., LORENZ, E., BETCKE, J., SARK, W. G. J. H. M. VAN, HEYDENREICH, W., WIEMKEN, E., STETTLER, S., TOGGWEILER, P., BOFINGER, S., SCHNEIDER, M., HIELSCHER, G., and HEINEMANN, D.: *Monitoring and remote failure detection of grid-connected pv systems based on satellite observations*. *Solar Energy*, 81:548–564, 2007.
- [GW91] GORDON, J. M. and WENGER, H. J.: *Central-station solar photovoltaic systems: field layout, tracker, and array geometry sensitivity studies*. *Solar Energy*, 46(4):211–217, 1991.
- [HM85] HAY, J. E. and MCKAY, D. C.: *Estimating Solar Irradiance on Inclined Surfaces: A Review and assessment of Methodologies*. *Int. J. Solar Energy*, (3):pp. 203, 1985.
- [JKS⁺09] JOLLIFF, J., KINDLE, J. C., SHULMAN, I., PENTA, B., FRIEDRICH, M. A. M., HELBER, R., and ARNONE, R. A.: *Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment*. *Journal of Marine Systems*, 76:64–82, 2009.
- [JSS92] JANTSCH, M., SCHMIDT, H., and SCHMID, J.: *Results on the concerted action on power conditioning and control*. In *11th European photovoltaic Solar Energy Conference*, pages 1589–1592, 1992.
- [Jun08] JUNTA DE ANDALUCIA: *Estación Carmona-Tomejil*, 2008. <http://www.juntadeandalucia.es/agriculturaypesca/estacionesAgroclimaticas/>, [last read November 2008].
- [Lor03] LORENZO, E.: *Energy Collected and Delivered by PV Modules*. John Wiley, 2003, ISBN 0471491969.
- [LPEA02] LORENZO, E., PÉREZ, M., EZPELETA, A., and ACEDO, J.: *Design of tracking photovoltaic systems with a single vertical axis*. *Progress in Photovoltaics: Research and Applications*, 10(8):533–543, 2002.
- [Mac93] MACAGNAN, M. H.: *Caracterización de la radiación solar para aplicaciones fotovoltaicas en el caso de Madrid*. PhD thesis, Instituto de Energía Solar, UPM, 1993.
- [MR01] MARTIN, N. and RUÍZ, J. M.: *Calculation of the PV modules angular losses under field conditions by means of an analytical model*. *Solar Energy Materials & Solar Cells*, 70:25–38, 2001.
- [Nav08] NAVIDI, W.: *Statistics for Engineers and Scientists*. McGraw-Hill, 2008.
- [NL08] NARVARTE, L. and LORENZO, E.: *Tracking and ground cover ratio*. *Progress in Photovoltaics: Research and Applications*, 16(8):703–714, 2008.
- [Pee01] PEEBLES, P. Z.: *Probability Random variables, and random signal principles*. McGraw-Hill Education, 2001.
- [Per08] PERPIÑÁN, O.: *Grandes Centrales Fotovoltaicas: producción, seguimiento y ciclo de vida*. PhD thesis, UNED, 2008. <http://e-spacio.uned.es/fez/view.php?pid=bibliuned:20080>.
- [PGWS91] PANICO, D., GARVISON, P., WENGER, H. J., and SHUGAR, D.: *Backtracking: a novel strategy for tracking PV systems*. In *IEEE Photovoltaic Specialists Conference*, pages 668–673, 1991.
- [PLC07] PERPIÑÁN, O., LORENZO, E., and CASTRO, M. A.: *On the calculation of energy produced by a PV grid-connected system*. *Progress in Photovoltaics: Research and Applications*, 15(3):265–274, 2007.
- [R D08] R DEVELOPMENT CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. <http://www.R-project.org>, ISBN 3-900051-07-0.
- [Ros03] ROSS, S. M.: *Peirce's criterion for the elimination of suspect experimental data*. *Journal of Engineering Technology*, 2003.

- [Sar09] SARKAR, DEEPAYAN: *lattice: Lattice Graphics*, 2009. <http://CRAN.R-project.org/package=lattice>, R package version 0.17-20.
- [SHD⁺07] SÚRI, M., HULD, T., DUNLOP, E. D., ALBUISSON, M., LEFEVRE, M., and WALD, L.: *Uncertainties in photovoltaic electricity yield prediction from fluctuation of solar radiation*. In *22nd European Photovoltaic Solar Energy Conference*, September 2007.
- [SJM⁺09] STOW, C. A., JOLLIFF, J., MCGILLICUDDY, D. J., DONEY, S. C., ALLEN, J. I., FRIEDRICHS, M. A. M., ROSE, K. A., and WALLHEAD, P.: *Skill assessment for coupled biological/physical models of marine systems*. *Journal of Marine Systems*, 76:4–15, 2009.
- [Tay97] TAYLOR, J. R.: *An Introduction to Error Analysis*. University Science Books, 2nd. edition, 1997.
- [Tay00] TAYLOR, K. E.: *Summarizing multiple aspects of model performance in a single diagram*. Technical report, Program for Climate Model Diagnosis and Intercomparison, 2000. <http://www-pcmdi.llnl.gov/publications/pdf/55.pdf>.