An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation

Yves-Marie Saint-Drenan, S. Bofinger, R. Fritz, S. Vogt, G.H. Good, J. Dobschinski

An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation

Yves-Marie Saint-Drenan[a,*], Stefan Bofinger[a], Rafael Fritz[a], Stephan Vogt[a], Garrett H. Good[a], Jan Dobschinski[a]

[a]*Fraunhofer Institute for Wind Energy and Energy System Technology (IWES), 34119 Kassel, Germany*

Preprint submitted to Solar Energy in October 2015

Abstract

The aim of this work is to develop an algorithm that can utilize historical PV power measurements to establish the parameters of a physical model for power production. The chosen approach consists in evaluating the parameters of a PV model that maximize the likelihood that simulations match with power measurements. The proposed method offers advantages beyond the standard approaches used for the simulation or prediction of PV power production, as it makes maximum use of the information typically available on a PV plant (plant description and measurement history). Furthermore, an interpretation and control of the algorithm output is made possible. The performance of the proposed approach has been evaluated and analysed using measurements from two PV plants. It is shown that the proposed approach may identify the orientation angles of a PV module to within an accuracy of less than 2° in optimal cases. Situations were also found with a difference between the estimated and actual angles of 5°, for which the estimated parameters lead to better simulation/forecast accuracy than the actual ones as they balance the systematic error of the chosen PV-model.

Keywords: Photovoltaic, Simulation, Forecast, Characterization

1. Introduction

It has become commonplace for photovoltaic forecast suppliers or academic groups to need to generate forecasts for PV plants for which little information aside from historical power measurements is available. The two common approaches in this case are the physically motivated approach and the statistical approach.

The physically motivated approach maximizes the use of the information available for the plant (Drews et al., 2006; Kidwelly, 2006). Optimally, PV plant information includes orientation angles of the PV modules along with the module and inverter specifications. In this case, a calculation of the PV power from meteorological data using available models from the literature is possible. Systematic differences between the

*Corresponding author
Email address: yves-marie.saint-drenan@iwes.fraunhofer.de (Yves-Marie Saint-Drenan)

simulated/forecasted and measured power can however be frequently observed. This error typically results from differences between the information used and the actual characteristics of the PV plants (approximate module orientation, deviation from manufacturer specifications, etc...). A manual correction of the plant information used is always possible, but may be time-consuming. The physical approach applies to ideal conditions, but it is unfortunately often the case that its implementation is impossible due to missing PV plant information needed for the calculation of the power from meteorological data.

The alternate approach is the statistical one, for which PV plant information is not a necessary prerequisite. In this approach, the best possible use is made of historical measurements. Artificial neural networks have become a standard practice to this end, and numerous works can be found in the literature on such methods (de Rocha Vaz, 2014; Dolara et al., 2015; Espinar et al., 2010). Though a statistical approach avoids the problems faced by the physical approach, other issues are present themselves. A neural network or any other statistical method learns dependencies between input and output data using a training dataset. For this purpose, it is important to exclude data affected by measurement errors or plant outages from the training dataset that would hinder the training phase of the statistical method used. Though obvious measurement errors can be easily detected and excluded from the training dataset, other errors like downed power lines or module shading may be more difficult to identify. The performance of the statistical approach is thus strongly dependent on the quality of the training dataset, which is sometimes difficult to guarantee. In the case of a deficient training set, it is not possible to check or fit the statistical coefficients to make manual fixes as were possible with the physical approach. Lastly, no use of the available plant parameters can be made with the statistical approach.

Both the physical and statistical approaches thus have advantages and drawbacks, and the optimal approach may depend on the quality of the available dataset. Still, neither approach is ideal as both ignore some part of the available information: historical measurements are not used in the physical approach and plant parameters is overlooked in the statistical approach.

Time series of power measurements implicitly contain a wealth of information on a PV plant. A visual inspection of this data may for example easily reveal whether the PV modules are oriented to the east or the west. This shows that it may be possible to derive (or train) parameters of a physical model from historical measurements. Such a hybrid approach would offer many advantages. First, the simulation model could integrate physical models available from the literature. Then, information contained in historical measurements would be fully exploited. Finally, it would be possible to control and modify trained parameters, which would have a physical sense. Regarding the last point, information available on a PV plant (module orientation, inverter or module specifications) could be explicitly used for the validation or modification of the assessed parameters.

The goal of the work presented in this paper is to develop a hybrid approach, in which parameters of a PV model are estimated from historical PV power measurements and meteorological data. The focus of this paper is consequently put on the choice of the appropriate configuration parameters of a PV model (module azimuth and tilt angle, power curve, etc...), rather than on minimizing the forecast/simulation error. A minimization of the error by e.g. removing systematic errors from the meteorological input data and/or by means of model output statistical methods may be conducted once the parameters of the considered plant are known, but this step is not addressed here.

Characterizing a PV plant requires estimating the parameters of a PV plant model that lead to the best match between measurements and simulation. A preliminary step is to choose a PV model, which is the focus of section section 2. The approach used for assessing the parameters is then described in section 3. In section 4, parameters of two plants are evaluated with the proposed approach; and compared with the known plant characteristics. Advantages and limitations of the parameter estimation algorithm introduced in this paper are then finally discussed in conclusion.

2. PV plant model selection

The aim of the proposed approach is to derive parameters of a PV model from historical measurements, facilitating the simulation/forecast of the power production of a PV plant from meteorological data with the best accuracy. This goal has two objectives at odds with one another. On one hand, the best simulation accuracy is obtained by using complex models requiring detailed information on the configuration of a PV plant. Though power measurements implicitly contain a lot of information on a PV plant, it is clear that it is not possible to ascertain each detailed characteristics of a PV plant from this data. The choice of an overly complex model would thus make the parameter estimation impossible. On the other hand, it can be expected that while the parameters estimation of a very simple model would be much easier, the choice of an overly simple model could limit the simulation accuracy due to its inherent uncertainty. Regarding the choice of the set of equations for the simulation of the PV power production from amongst the different models available in the literature, a compromise is thus required between minimizing the amount of information on the PV plant needed by the model and maximizing the model accuracy.

To find the optimal model, the choice of the PV model considers different PV plant characteristics and their respective effects on the power output. First, all processes occurring in a PV plant, whose consideration with the chosen approach is unrealistic, were neglected (e.g. local shading, the effect of wind on the module temperature, voltage-dependency of the inverter efficiency. . .). The characteristics of a PV plant to which the output power is most sensitive were then identified. These are the two module orientation angles, the set of parameters describing the optical losses of the module glazing, the electrical characteristics of the power module and the power curve of the PV inverter. This information is important for the choice of the set of models describing the different parts of a PV plant. Indeed, in order to decrease the modelling error, accurate models should be preferred to describe the effects these key characteristics on the output power.

In contrast, simpler models can be chosen for other parameters whose effect on the power is lower. Based on these considerations, a set of models to simulate the output power from meteorological data has been selected from amongst those available in the literature. The resulting calculation steps are described in the following paragraphs.

With the effect of local shading being neglected, the plane of array (POA) irradiation can be estimated from the global irradiation and the sun position using a set of models commonly used for this purpose (Iqbal, 1983; Quaschning, 1999). Here, the separation and transposition models proposed by Skartveit et al. (1998) and Perez et al. (1993) are each respectively used for estimating the plane of array irradiation from the global horizontal irradiation. The module azimuth and tilt angle as well as the ground albedo are the PV plant information required for this first step. To limit the number of model parameters and considering its limited effect on the output power, the ground-albedo is assumed to be constant and equal to 20%.

To estimate the POA irradiation effectively contributing to the photovoltaic effect (effective irradiation), optical losses occurring within the module glazing have to be considered. The formulation of Martin and Ruiz (2001) for calculating the angular losses has been chosen from the models existing in the literature (Souka and Safwat, 1966; Standard et al., 1977; King et al., 1997) as it offers the best compromise between simplicity and physicality. Indeed, Martin and Ruiz propose an analytical model based on theoretical and experimental results that only requires two parameters (the angular loss coefficient and a fitting coefficient for the diffuse and reflected irradiation) for the determination of the angular losses of the direct, diffuse and reflected irradiations. As the output PV power is little sensitive to the fitting coefficient of the Martin and Ruiz (2001) model for the diffuse and reflected irradiation, it is assumed constant and set to a value representative for crystalline modules (0.07).

The influence of the variations of the solar spectrum on the power production of PV cells is neglected so that the output of the PV power modules can be directly evaluated with the effective irradiation and the module temperature.

The calculation of the module temperature can be nontrivial, as it is affected by local conditions (wind cooling the module back-side, thermal inertia of the building, etc…). However, a detailed modelling of the module temperature requires information on a PV plant that cannot be considered in the proposed approach. As a result, the expression proposed by Ross (1976) has been chosen, where the difference between the module and air temperature is assumed to be proportional to the POA irradiation.

Further models may be chosen for the remaining calculation steps that would result in a relatively large set of additional parameters describing the respective influences of the PV module characteristics, DC-losses, inverter efficiency, and so on, on the output power. A general consideration of the remaining simulation steps however shows that a unique value of the produced power corresponds to each value of the effective irradiation and module temperature. For the present application, individually modelling each component of the plant is not necessary, since only their cumulative effect is needed for the power calculation. Accordingly, a pragmatic simplification was made, using a look-up table (LUT) describing the combined behaviour of the PV module, cable losses, inverter efficiency, etc... rather than simulating each effect individually Additionally, with the assumption that the difference between module and air temperature is proportional to the POA irradiation (Ross, 1976), it can be shown that the explicit simulation of the module temperature can be avoided. Indeed, under this assumption, a single power value corresponds to any pair of effective irradiation and air temperature. It was thus decided to use a look-up-table giving the output PV power for all values of the effective irradiation and air temperature.

An advantage of the look-up-table is that parameters which are difficult to assess are implicitly considered (effective capacity, soiling loss, mismatch losses, etc...). Furthermore, eventual modelling weaknesses are avoided since it is not necessary to choose a mathematical model describing a relationship between the input and output data. The use of an LUT may therefore not necessarily lead to a reduction of the model accuracy. Finally, the use of an LUT instead of a set of additional parameters actually considerably simplified the estimation of the model parameters from historical measurements (see following section).

In total, the chosen PV plant model uses three parameters (the module azimuth and tilt angles and the angular loss coefficient) and an LUT describing the "total power curve" of the PV plant. A flow chart illustrating the PV model is given in Figure 1. The input meteorological data are the global horizontal irradiation and the air temperature (upper row) and the PV plant parameters are the module orientation angles, the angular losses parameter and the LUT (left column).

One last issue remains to be addressed regarding the physical model. The power output of an increasing number of PV plant is capped when the power exceeds a certain level (i.e. 70% of the peak capacity). This limitation on the power is commonly referred to as inverter clipping. Though inverter clipping is not explicitly discussed in this section, it is implicitly considered in the look-up table. Indeed, all irradiation and temperature values leading to power values larger than the limit under normal conditions are associated with the clipping limit. The effect of the power limitation is therefore contained in the look-up table and no specific measure is required to consider the effect of inverter clipping.

3. Determination of the simulation parameters of a PV plant

With a PV model chosen, it remains to discover how the set of parameters best describing a PV plant can be evaluated from power measurements. The basic idea is to identify the set of parameters with which power simulated from meteorological data best matches with the measurements. Two issues need however be clarified prior to the parameter search (section 3.3). Firstly, it is unclear what meteorological data are the best suited for the determination of the configuration parameters (section 3.1). At the same time, given the presence of a look-up-table in the model parameters and that measurements can be affected by issues such as power line failures, it is unclear what cost function is suited to the present problem (section 3.2).

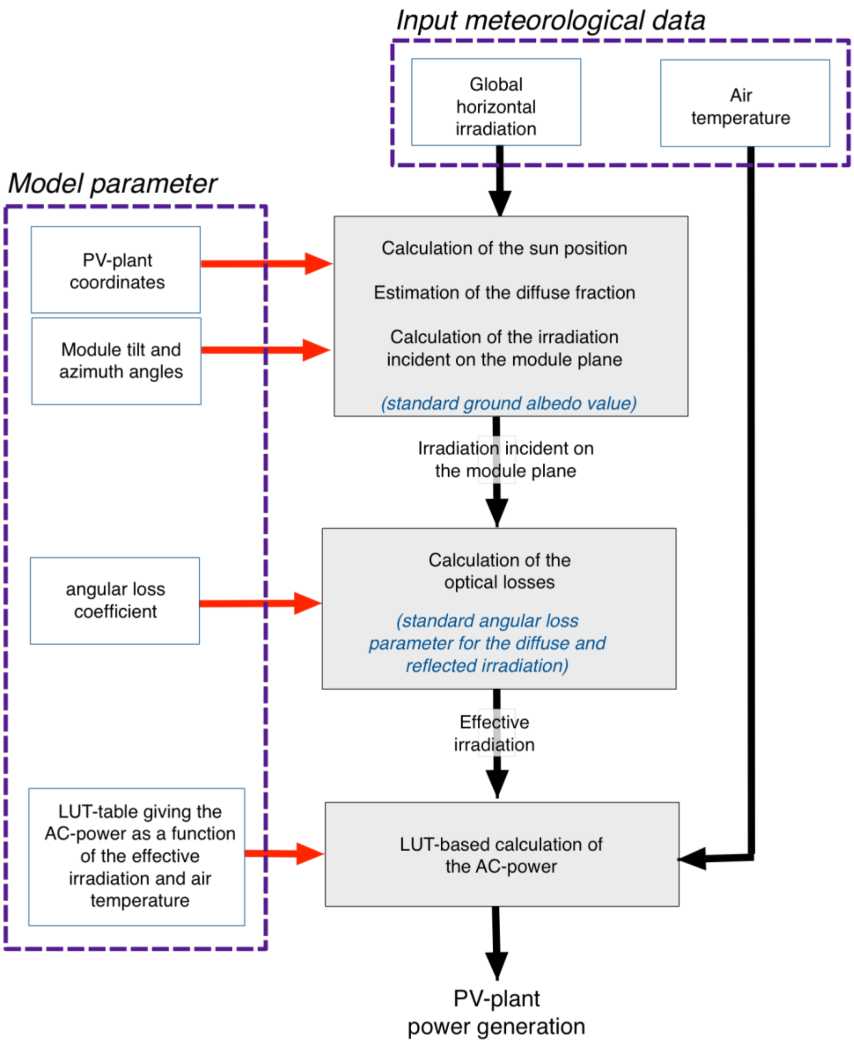

Figure 1: Flow chart of the simplified PV plant model

3.1. Choice of meteorological input data for the model parameter evaluation

Apart from its configuration (parameters of the PV model), the power production of a PV plant depends on the solar irradiation and the air temperature. The first step in the determination of the simulation parameters of a PV plant thus consists in collecting meteorological data for each point in time a measurement is available.

Should the present approach be needed to calculate PV power forecast, it may appear at first glance natural to use irradiation and temperature forecast to estimate the configuration parameters of the considered plant. Deviations between forecasted and actual weather conditions when the power was measured may however result in noise that limits the performance of the parameter estimation. It is thus preferable to use the most accurate meteorological information available as an input for the parameter estimation.

Two situations can occur at this preliminary step. In the optimal situation, irradiation and temperature are measured parallel to the AC power production. In this first situation, the meteorological information needed for the power calculation is available directly from the set of measurements. A more common situation is that only the power generation is measured and another source of meteorological data must consequently be used. In this latter case, it is possible to extract for example irradiation from satellite-based products and temperature from NWP analysis for the desired location and time period.

3.2. Choice of the cost function

Once meteorological data is available for each value of the power measurement, it remains to identify with which set of parameters the PV power simulated from meteorological data best fits the measurements. This is a common optimization problem that can be solved by choosing a cost function to quantify the simulation error and by searching for its global minimum over the parameter space.

At first glance, it may seem natural to choose a common measure of the simulation error such as the RMSE or MAE as the cost function. In practice, the implementation of this approach is difficult due to the existence of a look-up table in the parameter set. Indeed, each value contained by the LUT needs to be estimated by the optimization, such that the parameter space is too large for the optimization. Another approach (or problem formulation) is thus necessary to solve the issue caused by the LUT.

The use of a look-up table in the simplified model has been motivated by the fact that, with the assumed simplifications, a single value of the output PV power corresponds to any pair of air temperature and effective irradiation values. This characteristic of the chosen PV model can also be exploited to evaluate the optimal module orientation angles and optical loss coefficient (the LUT is not considered in a first time). Indeed, these three parameters can be expected to have the following effects:

- If the module orientation angles and the optical loss coefficient are optimally chosen, little dispersion should be observable amongst measurements corresponding to similar values of the simulated effective irradiation and temperature (e.g. left-side plot in Figure 2).
- In contrast, a sub-optimal set of parameters should result in a higher dispersion among measurements corresponding to similar values of the simulated effective irradiation and temperature (e.g. right-side plot in Figure 2).

Based on the considerations above, it should be possible to search for the three parameters (module tilt and azimuth angles and angular loss coefficient) by minimizing the dispersion of the measurements for any values of the effective irradiation and temperature. The advantage of this approach is that the shape of the power curve (quantified by the LUT) is not necessary for the optimization, which only focuses on maximizing the density of points on this unknown power curve. As a result, the parameter space is reduced to the three dimensions formed by the module tilt and azimuth angles and the angular loss coefficient. To implement this idea, a cost function must still to be chosen that quantifies the dispersion of the measurements for any values of the effective irradiation and air temperature.

For a given set X_{param} of the three model parameters (module azimuth angle, module tilt angle and angular loss coefficient), the effective irradiation can be calculated from the global horizontal irradiation. Three time series are thus available as input data for the cost function: power measurements, air temperature and effective irradiation.

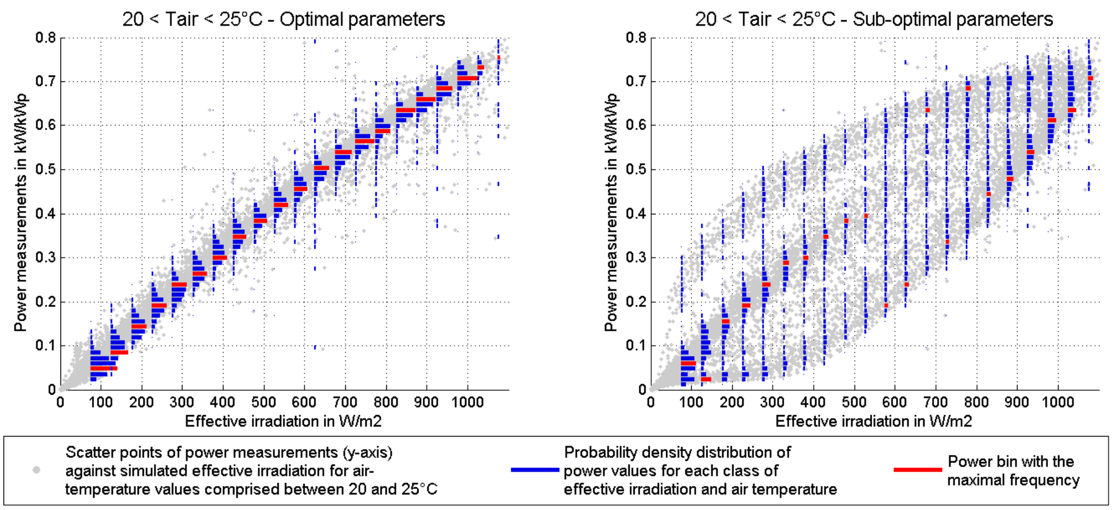

Figure 2: Illustration of the approach used for estimating the performances of a given set of parameters.

The dispersion of the data is first evaluated by calculating the joint probability distribution of the three considered variables. For this purpose the number of occurrences of the three considered quantities within different bins is counted. Bin widths of $0.01kW/kW_p$, $20W/m^2$ and 2°C have been used for the power, irradiation and air temperature, respectively. This first step is illustrated in Figure 2 for air temperature values ranging between 20 and 25°C and for a set of optimal and sub-optimal parameters (left and right picture respectively). Scatter points represent the adequacy between power measurements (ordinate) and simulated effective irradiation (abscissa). The number of values present in different power bins for each class of effective irradiation is represented by a horizontal bar. This operation is conducted for each class of air temperature. The meaning of the bar colours is discussed later.

The joint probability distribution concerns the distribution of the dataset in the entire space covered by the data. Since only the frequency of the occurrence of power values in the vicinity of the (unknown) power curve is needed, it remains to extract this information from the joint probability distribution.

When the set of parameters is optimal, it can reasonably be expected that the frequency of measurements will be higher for power bins corresponding to the power curve than for those elsewhere. In this case, the required information can thus be assessed for any value of the temperature and irradiation by selecting the bin with the maximal frequency from amongst all bins of power measurement. These bins are marked in red in the example given on the left picture of Figure 2. A final summation over all temperature and irradiation bins should give approximately the percentage of measurements lying in the vicinity of the unknown power curve (sum of frequency corresponding to all red bars in the left picture of Figure 2).

When the set of parameters is sub-optimal, the assessment described above becomes meaningless since it can no longer be expected that the power bin with the maximum frequency corresponds to the power curve. This situation is illustrated by the right picture in Figure 2. However, in this case, the previous calculation should lead to a lower value that when optimal parameters are used. Indeed, based on the example given in Figure 2, red bars are higher in the left (optimal parameter set) that in the right picture (sub-optimal parameter set). In that sense, this approach can still be used to evaluate the performances of a set of coefficients.

Finally, the cost function used for the estimation of the module azimuth angle, the module tilt angle and the angular loss coefficient is thus:

$$f^{cost}\left(X_{param}\right) = \sum_{j,k}\left[\max_{i}\left(p(PW_{Meas} = PW_i, G_{eff}(G_{hor}, X_{param}) = G_j, T_{air} = T_k)\right)\right] \tag{1}$$

Where:

- X_{param} is the set of considered parameters (module azimuth angle, module tilt angle and optical loss coefficient),
- $PW_{Meas}, G_{hor}, T_{air}$ are the power measurements, the global horizontal irradiation and air temperature data, respectively,
- $G_{eff}(G_{hor}, X_{param})$ is the effective irradiation calculated from G_{hor} with the parameter set X_{param},
- PW_i, G_j, T_k are the i^{th}, j^{th} and k^{th} bins of the power, irradiation and air temperature, respectively,
- $p(X = X_i, Y = Y_j, Z = Z_k)$ is the probability that X, Y and Y are equal to X_i, Y_j and Z_k (joint probability distribution), and,
- $f^{cost}\left(X_{param}\right)$ is the cost function for the set of parameters $_{Xparam}$.

3.3. Determination of the configuration parameters

Using time series of power measurements and the corresponding irradiation and temperature data, the first three parameters can be evaluated by finding the set of parameters X_{param} that maximizes the cost function introduced in the previous section (1):

$$X^{Opt}_{Param} = argmax\left(f^{cost}\left(X_{Param}\right)\right) \tag{2}$$

This optimization is made in a three-dimensional space formed by the module tilt angle, the module azimuth angle and the optical loss coefficient, so that the estimation of the first three parameters is relatively fast.

Once the three parameters that maximises the cost function are found, the next step of the algorithm consists in evaluating the look-up table that corresponds to the power curve of the PV plant. For this purpose, the simulated effective irradiation, air temperature and power measurements are used and the evaluation is made in two steps.

In the first step of the evaluation, for any value of the effective irradiation and air temperature, the power curve value is evaluated as the most frequent value of the power measurement (bin with the largest number of power measurements). These are represented by the red bars in Figure 2. The most frequent (or modal) value is preferred over e.g. the average value because it was judged to be more stable given the problems potentially affecting the measurements (line outage, measurement errors, etc...).

At this stage, only values of the effective irradiation and air temperature covered by the measurement dataset can be evaluated in the look-up-table. This would not be a problem if the measurement dataset were sufficiently large such that all possible values of air temperature and effective irradiation were covered. However, it cannot be excluded that a simulation could require a value from the look-up table that could not be assessed with the available measurements. An estimation of the values undefined in the look-up-table was thus necessary, which is the second step of the evaluation.

For the purpose of the estimation of the undefined LUT values, a linear dependency between the output power and the air temperature for each value of the effective irradiation is assumed. With this assumption, at each value of the effective irradiation, the two coefficients describing the linear dependency between output AC-power and air temperature are estimated with the available data and the undetermined values of the look-up table are filled by extrapolating the data with this linear relationship.

4. Sample applications

4.1. Test PV plants

Two PV plants have been chosen to illustrate the performances of the parameter estimation algorithm presented in this paper. These plants have been selected from amongst numerous plants for which the algorithm has been implemented, with the intention of demonstrating not only the performances obtained but also of showing the limitation of the proposed approach. Power measurements used in the two chosen examples are thus affected by local shading and measurement errors. Measurement errors have been intentionally left in the dataset for assessing how the proposed approach copes with such issues. This is discussed later in the validation of the results.

The measurements of two plants provided by the Technical University of Bern have been chosen to illustrate the operation and evaluate the performance of the proposed algorithm. A short description of the two PV plants used is given in Table 4.2.

	Stade de Suisse Wankdorf (DA1)	EBL Liestal
Latitude	46°57'51"	47°29'16"
Longitude	7°27'55"	7°46'59"
Year of installation	2005	1992
Azimuth and tilt angles [1]	-63° E / 20.5°	0° S / 30°
Peak power	127.575kW$_p$	18.510 kW$_p$
Reference of module	Kyocera KC-167GH-2	Kyocera LA361H51
Number of modules	729	363
Reference of inverter	Sputnik SolarMax125	Sputnik SolarMax20
Period used for the parameter estimation	01.01.2008 – 31.12.2008	01.01.2009 – 31.12.2009

Table 1: Description of the characteristics of the EBL Liestal PV plant.

For these two plants, one year of five-minute measurements of the output AC-power, global horizontal irradiation, POA irradiation, air temperature and module temperature are available.

For both plants the algorithm has first been run with local meteorological measurements and then using remote-data. Remote data include irradiation calculated from satellite images with the helioclim-3v4 method (Espinar et al., 2012) and air temperature taken from Cosmo-DE analysis (Schulz and Schättler, 2014). Where remote data have been used fifteen-minute average power measurements have been evaluated to match the time resolution of the satellite data. The original time resolution of five minutes has been used when the algorithm is run with local meteorological measurements.

4.2. Validation of the estimated parameters

For the sake of brevity, the optimization conducted for the estimation of the parameters is not detailed in this paper. Alternatively, reports generated by the algorithm are given in the appendix for each algorithm run conducted. These reports provide an overview of all end- and intermediary results, which are important for assessing the quality of the parameter estimation. Only the final results of the algorithm are discussed in this section.

Module orientations and optical loss coefficients found with the algorithm are given in Table 4.2 and scatter plots of the normalized measured power (y-axis) as a function of the effective irradiation (abscissa) are displayed in **??**.

Meteorological data used for the parameter estimation	Estimated parameters	**Stade de Suisse Wankdorf (DA1)**	**EBL Liestal**
Local irradiation and temperature measurements	Azimuth angle	-68°E	-3°E
	Tilt angle	22°	30°
	Optical loss coefficient	0.14	0.25
Satellite-derived irradiation and Cosmo-DE analysis temperature	Azimuth angle	-68°E	-1°E
	Tilt angle	22°	30°
	Optical loss coefficient	0.06	0.085

Table 2: Illustration of the approach used for estimating the performances of a given set of parameters.

Table 3: Results of the parameter estimation.

The module orientation angles estimated with the algorithm can be directly compared with the values available from the plant description. A validation of the angular loss coefficient and power curve look-up table is by contrast not possible, as the actual values of these parameters are not available from the plant information. An indirect validation of these parameters is therefore realized by verifying that the estimated power curve matches the dependencies between power measurements, effective irradiation and air temperature.

Validation of the estimated module orientation angles

The module orientation angles evaluated by the algorithm and those provided by the plant operator can be found in and respectively.

There is a good agreement between module orientation angles found with local meteorological measurements and remote data. For the Liestal plant the module orientation found with the local and remote data are (-3°E; 30°) and (-1°E; 30°) respectively, while the same module orientation has been found with the two datasets for the Wankdorf PV plant (-68°E; 22°).

The module tilt angle found at Liestal corresponds exactly to that provided by the plant operator (30°). In contrast, estimated azimuth angles correspond to a slightly eastward orientation (-3°E and -1°E) while a southern orientation is indicated in the plant description. An aerial view of the plant taken from Google Earth (left picture in Figure 3) reveals that the plant is indeed slightly oriented to the east, such that the results of the algorithm are plausible.

In the left picture of Figure 3, it can also be observed that a part of the PV plant is shaded in the morning. The production deficit resulting from the shading may explain the higher dispersion of the scatter points in the two right plots from Figure 5 for an effective irradiation between 0 and $600 W/m^2$.

A larger discrepancy is found between module orientations given by the plant operator (-63°E; 20.5°) and those found by the algorithm (-68°E; 22°) in the second example (Wankdorf Stade de Suisse). The difference in tilt angle is relatively small (overestimation of +1.5°) but the larger azimuth angle difference of 5° is not negligible. A control of the module orientation with Google Earth (right picture in Figure 3)

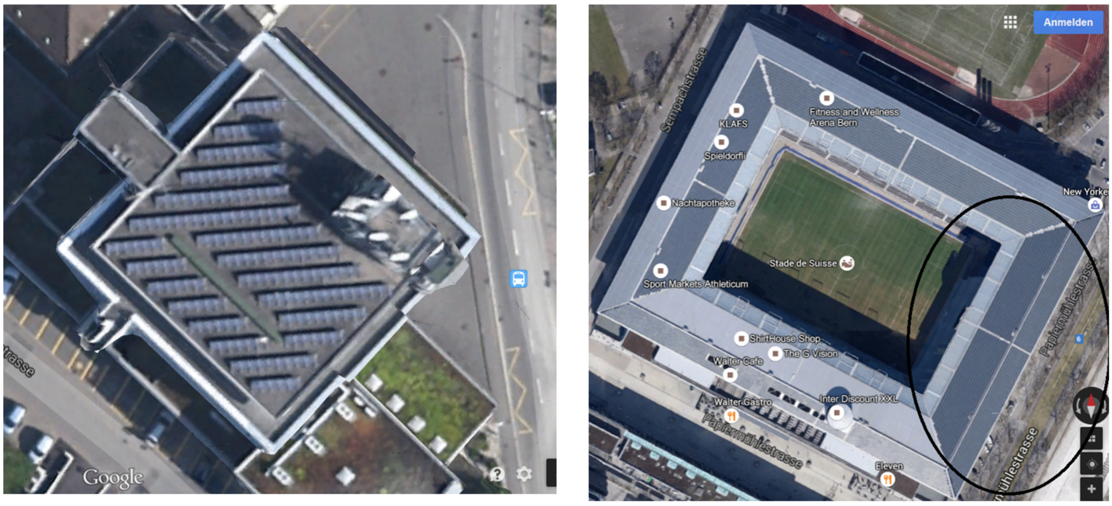

Figure 3: Aerial view of the Liestal left picture) and Wankdorf Stade de Suisse PV plants (right picture) Source: Google earth

confirmed that the azimuth angle provided by the plant operator (-63°E) is correct. The module orientation estimated by the algorithm therefore seems to deviate from its actual value for this plant.

Numerous intermediate measurements available from the Wankdorf PV plant (POA irradiation, air and module temperature, DC and AC power) allowed for the validation of the different steps of the PV power calculation in order to understand the reason for this difference. An analysis of these intermediate results revealed that this difference in the azimuth angle results from the assumption made for modelling the module temperature. Indeed, the Ross model was chosen, in which the difference between the module and air temperature is assumed to be proportional to the POA irradiation. This implies that a single module temperature corresponds to each value pair (POA irradiation and air temperature). The analysis of the intermediary measurements showed that the characteristics of the module temperature do not fully satisfy this simplifying assumption.

To highlight the behaviour of the module temperature responsible for the deviation of the estimated azimuth angle from its actual value, differences between measured module and air temperature are displayed as a function of the measured POA irradiation in Figure 4. Since a dependence of the scatter points with the time of the day was identified, scatter points have been coloured according to the solar azimuth angle.

It can be observed in Figure 4 that for a given POA irradiation the difference between the module and air temperature is lower in the morning than in the afternoon. For example, for a POA irradiation of $400 W/m^2$, a temperature difference of 10°C is observed at a solar azimuth of 120°, while it increases to 20°C as the solar azimuth is 240°. Under the same external conditions (air temperature and POA irradiation), a difference depending on the solar azimuth reaching up to 10°C can thus be observed, which is inconsistent with the simplifying assumption made.

The observed dependency of the module temperature on the solar azimuth (or the time of the day) can be

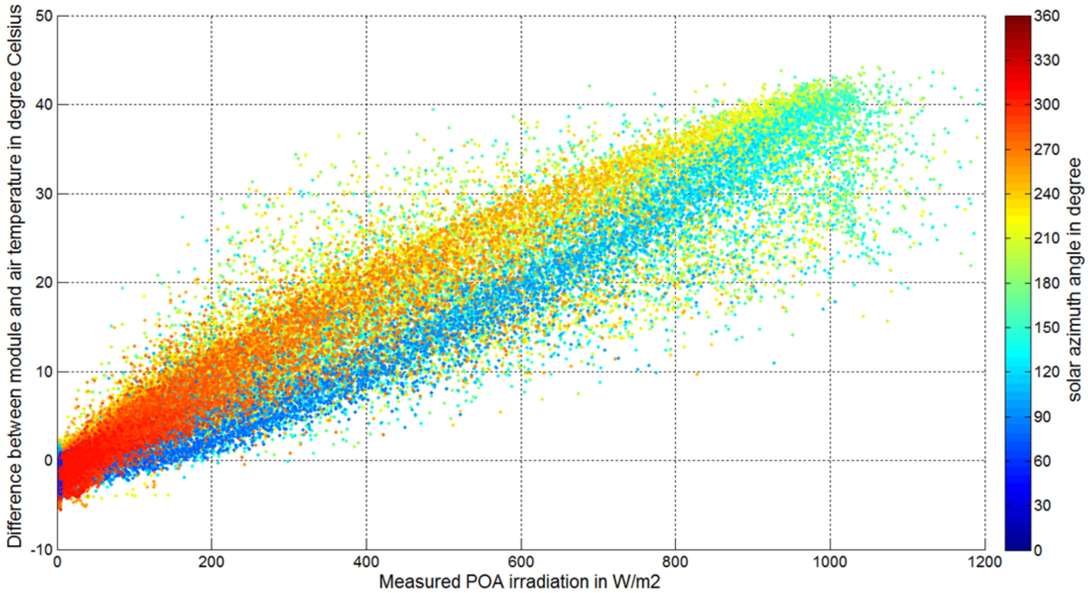

Figure 4: Dependence of the difference between the module and air temperature (ordinate) on the POA irradiation (abscissa) and the solar azimuth angle (colour of the scatter points) for the Wankdorf PV plant.

easily explained by the fact that the PV modules are directly integrated on the roof of the Wankdorf stadium. The air behind the module is heated by the incoming irradiation in the course of the day, which heats the backside of the PV module so that the module temperature exhibits a dynamic behaviour influenced by the thermal inertia of the building. As the consideration of the thermal inertia of the module was not foreseen in the chosen model, the parameter estimation algorithm has balanced the resulting modelling error by overestimating the module azimuth angle.

As previously mentioned, an explicit consideration of effects such as those illustrated in Figure 4 have been intentionally omitted in the chosen PV model (they would have required information of excessive detail on a PV plant). It is thus clear that a modelling error may occur for plants like Wankdorf where the validity of the simplifying assumption is limited. Given that the proposed algorithm estimates model parameters by maximizing the probability that a simulation matches the measurements, it is not surprising that a set of parameters different from the actual ones is found at Wankdorf. In a way, it can be considered that the difference between the estimated and actual parameters compensates for the weaknesses of the simplified PV model for this plant.

Validation of the angular loss coefficients and power curve LUT

As already mentioned, a direct validation of the angular loss coefficient and power curve LUT is not possible, as their actual values are not available from the description of the PV plant. Therefore, an indirect validation has been conducted, where it was verified that these parameters describe well the dependence between the effective irradiation, air temperature and power measurements.

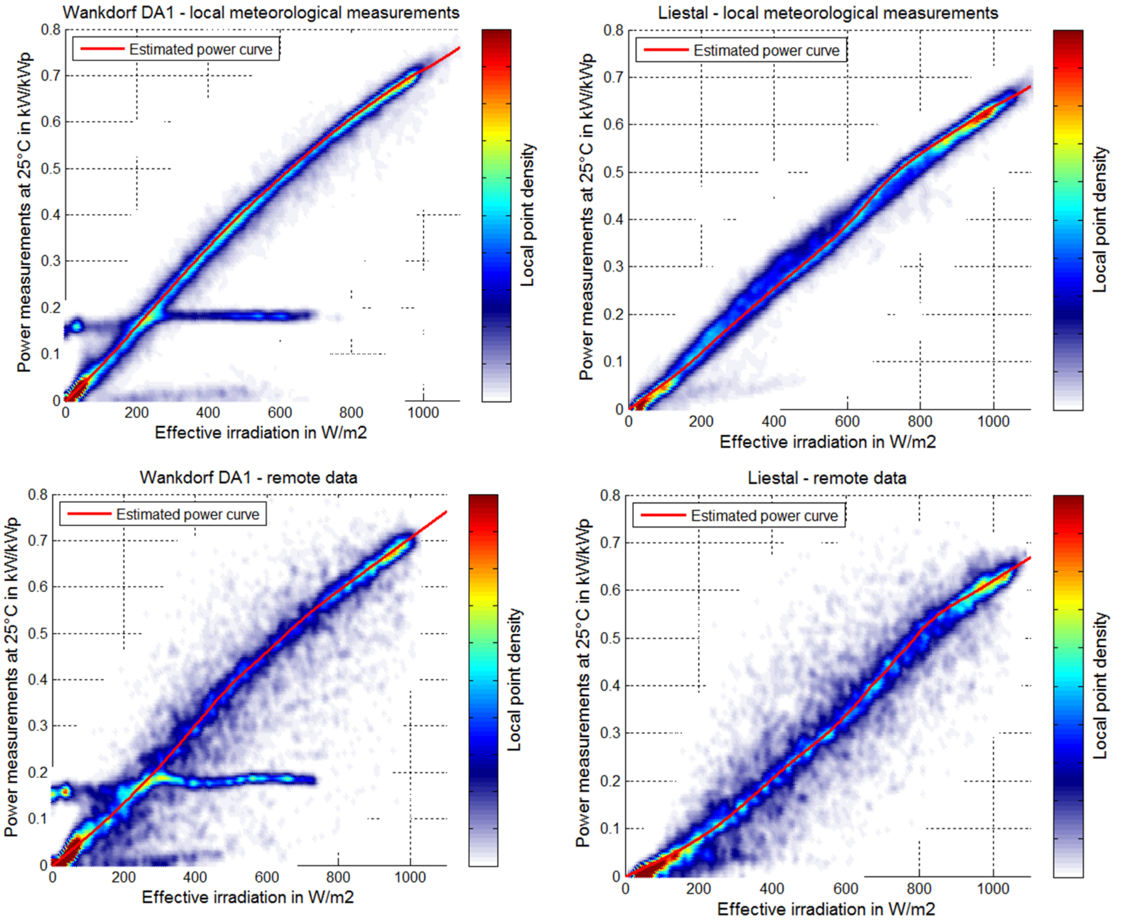

Figure 5: Scatter plots of the normalized measured power (ordinate) as a function of the effective irradiation (abscissa) and the estimated power curve (red line) – the colour of the scatter points represent the local point density.

In **??**, scatter plots of the normalized measured power as a function of the effective irradiation can be found for the four algorithm runs. Power measurements were corrected for their dependency on the air temperature to facilitate the visualisation of the data. For this purpose, the linear dependency assessed during the construction of the look-up table (see previous section) has been used to evaluate AC-power values corresponding to an air temperature of 25°C. The colour of the points represents the local density of the scatter points. A light blue to blue point occurs rarely, while a red point is very frequent. The power curves corresponding to an air temperature of 25°C, evaluated by the parameter estimation algorithm, are displayed in each scatter plot by a red curve.

The scatter plot corresponding to the estimation of the model parameters of the Wankdorf plant using local meteorological measurements is displayed in the upper left picture in Figure 5. A line of scatter points with a high density (light blue to red dots) starting from the origin ($0W/m^2$; $0kW/kW_p$) and ending at ($1000W/m^2$;$0.7kW/kW_p$) can be observed in this figure, which corresponds to the power curve of the PV plant. A horizontal line of scatter points with a high density can also be observed in this scatter plot. These points result from a time period where values delivered by the data logger were constant (intentionally not excluded from the dataset). It is interesting to note that these points did not seem to affect the algorithm, which was expected, given the chosen cost function. The red line represents power curve values obtained from the algorithm. The red line matches very well to the line of scatter points of high density. In this first example, the module orientation and optical loss coefficient seem to be correctly estimated, as the scatter plot contains a continuous line with a high density of points. The LUT evaluated by the algorithm seems also very plausible because the estimated power curve matches well for regions of high scatter point density. In this first example, the results of the algorithm are very plausible and the algorithm performance appears to be insensitive to measurement errors. The same conclusion as that previously described can be drawn from the observation of the scatter plot corresponding to the estimation of the parameters of the Wankdorf plant using remote data (lower left picture in **??**). More noise than in the previous plot can however be observed here, which results from the uncertainty of the satellite-derived irradiation and the Cosmo-DE temperature. The power curve obtained with remote data is very similar to that resulting from local meteorological measurements. Lower values can however still be observed for values of the effective irradiation ranging from 0 to $400W/m^2$. This difference may be explained by the difference in the angular loss coefficient between the two runs, or by a bias in the satellite-derived irradiation. Despite these minor differences, it is interesting to note that similar results are obtained with the algorithm when local measurements or remote meteorological data are used.

The dispersion of scatter points corresponding to the parameter estimation of the Liestal PV plants from local meteorological measurements (upper right picture in Figure 5) is larger than that observed at Wankdorf (upper left picture in Figure 5). A visual inspection of the intermediate measurements available showed that a module shading occurring in the morning at low solar elevation (already observed in Figure 3) is responsible for this spread. The red line matches well with scatter points of high density. However, it is very likely that these points are affected by the shading and that the estimated power curve is lower than the actual one. The same effects can be observed in the scatter plot corresponding to the parameter estimation of the Liestal PV plants from remote data (lower right picture). In these last two examples, it is interesting to note that correct module orientation angles were found despite the effect of the local shading on the measurements.

5. Discussion and Conclusion

An algorithm has been developed that derives the parameters of a physical model from historical PV power measurements. For this purpose, a simple PV model fulfilling the requirements of the intended application has been chosen (**??**) and a parameter-estimation method dealing with usual issues occurring in a PV plant (e.g. line outage, measurement errors) has been proposed in section 3.

The operation and performance of the algorithm have been illustrated for two PV plants in section 4. Outputs of the algorithm were found to be plausible and in good agreement with the information available on the PV plants. It was found that the parameters estimated with the algorithm may deviate from their actual values when, due to modelling error, they result in better simulation results. In this sense, the output of the algorithm should be seen as a set of parameters that lead to the best simulation and not necessarily as the actual characteristics of the PV plant. Nevertheless, a physical interpretation of the algorithm output is possible albeit with some precaution.

With the chosen cost function, the algorithm was shown to be little sensitive to outliers resulting from measurement errors or power line outages, which constitutes an advantage in comparison to statistical methods. The performance of the proposed method were found to be limited when PV module are shaded. In that case, for the considered examples, the module orientation was correctly assessed but the power curve was underestimated for power values affected by the shading. An explicit consideration of this issue could improve the proposed approach in the future.

The algorithm has been tested with several hundred PV plants. These have shown that at least six months of power measurements are necessary for an accurate estimation of the module tilt angle. When less than six month measurement is available and should the module orientation angles be available from plant information, it is possible to only assess the optical loss coefficient and the power curve of the plant by setting orientation angles to their known value. With regards to this, the proposed method is much more flexible than traditional statistical or physical approaches. It has also been shown that its performance is limited in some situations. For example, it often occurs that a power production time series corresponds to the aggregated production of modules with different orientations. The algorithm performs poorly in such cases, since it is based on the assumption that only a single orientation exists for a PV plant. A simulation error was also observed to result from the assumption that soiling losses are constant with time.

The parameter assessment algorithm described in this paper is German patent pending (Saint-Drenan and Bofinger, 2012).

Acknowledgements

The authors thank the Technical University of Bern for their valuable measurements, which allowed validating and understanding the limitation of the method presented in this paper. The authors are indebted to Lucien Wald and Philippe Blanc for their help in understanding the Helioclim data. We would also like to thank the Transvalor team, which is in care of the SoDa Service that makes the access to the HelioClim databases efficient and user-friendly.

References

de Rocha Vaz, A. G. C., 2014. Photovoltaic forecasting with artificial neural network. Ph.D. thesis, University of Lisbon.

Dolara, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliari, E., 2015. A physical hybrid artificial neural network for short term forecasting of PV plant power output. Energies 8 (2), 1138–1153.

Drews, A., Lorenz, E., Betcke, J., Keizer, A., van Sark, W., Beyer, H. G., Heydenreich, W., Wiemken, E., Stettler, S., Toggweiler, P., Bofinger, S., Schneider, M., Heilscher, G., Heinemann, D., 06 2006. Remote performance check and automated failure identification for grid-connected pv systems – results and experiences from the test phase within the pvsat-2 project.

Espinar, B., Aznarte, J. L., Girard, R., Moussa, a. M., Kariniotakis, G., 2010. Photovoltaic Forecasting: A state of the art. 5th European PV-Hybrid and Mini-Gird Conference 33, 250–255.
URL http://hal.inria.fr/docs/00/77/14/65/PDF/Espinar-Tarragona2010.pdf

Espinar, B., Blanc, P., Wald, L., Gschwind, B., Ménard, L., Wey, E., Thomas, C., Saboret, L., 2012. HelioClim-3: a near-real time and long-term surface solar irradiance database. In: COST WIRE workshop on "Remote Sensing Measurements for Renewable Energy". p. 4 pp.

Iqbal, M., 1983. An Introduction to Solar Radiation. Academic Press.
URL https://books.google.fr/books?id=BjCqswEACAAJ

Kidwelly, P. (Ed.), 7 2006. PVSAT-2: Results of Field Test of the Satellite-Based PV System Performance Check. Vol. 4 of 5. The organization, The name of the publisher, The address of the publisher, an optional note.

King, D. L., Kratochvil, J. A., Boyson, W. E., 1997. Temperature coefficients for PV modules and arrays: Measurement methods, difficulties, and results. In: Conference Record of the IEEE Photovoltaic Specialists Conference. pp. 1183–1186.

Martin, N., Ruiz, J. M., 2001. Calculation of the PV modules angular losses under field conditions by means of an analytical model. Solar Energy Materials and Solar Cells 70 (1), 25–38.

Perez, R., Seals, R., Michalsky, J., 1993. All-weather model for sky luminance distribution-Preliminary configuration and validation. Solar Energy 50 (3), 235–245.

Quaschning, V., 1999. Regenerative Energiesysteme, Technologie, Berechnung, Simulation, 2nd Edition. Hanser, von Volker Quaschning.

Ross, R. G., 1976. Interface Design Considerations For Terrestrial Solar Cell Modules. In: Proceedings of the 12th IEEE Photovoltaic Specialists Conference. pp. 801–806.
URL https://www2.jpl.nasa.gov/adv{_}tech/photovol/ppr{_}75-80/InterfaceDesConsid{_}PVSC76.pdf

Saint-Drenan, Y.-M., Bofinger, S., Aug. 2012. Verfahren zur bestimmung von parametern einer photovoltaikanlage. DE Patent DE102012214329A1.

Schulz, J., Schättler, U., 2014. Kurze Beschreibung des Lokal-Modells Europa COSMO-EU (LME) und seiner Datenbanken auf dem Datenserver des DWD. DWD.
URL https://books.google.fr/books?id=ku8JvwEACAAJ

Skartveit, A., Olseth, J. A., Tuft, M. E., 1998. An hourly diffuse fraction model with correction for variability and surface albedo. Solar Energy 63 (3), 173–183.

Souka, A. F., Safwat, H. H., 1966. Determination of the optimum orientations for the double-exposure, flat-plate collector and its reflectors. Solar Energy 10 (4), 170–174.

Standard, A., et al., 1977. Methods of testing to determine the thermal performance of solar collectors. American Society of Heating, 93–77.

Panel (1):

The location of the PV plant is displayed on a map by a square whose colour corresponds to the maximum value of the cost function evaluated by the algorithm at this location. The abscissa is the longitude and the ordinate the latitude.

If the exact location of the PV plant is known, this map is trivial. Should the exact location of the plant not be known but rather for example only the postal code, an estimation of the coordinates of the PV plant is represented in this map. This is achieved by selecting all points where meteorological information is available in a given area (for example all pixels of the satellite) and assessing the pixel with the highest value of the cost function, which should serve as an approximation of the location of the PV plant.

Panels (2),(3):

An overview of the search of the maximum value of the cost function in the space formed by the three unknown parameters is given in these two plots. With the cost function having been assessed for all values in the three-dimensional space formed by the unknown parameters, the result of the optimization is a four-dimensional array that requires simplification for a visualisation of the results.

In panel (2), the maximum value of the cost function obtained for each value of the module orientation is displayed in colour as a function of the azimuth angle (abscissa) and tilt angle (ordinate). A blue square represents a small value of the cost function and a red pixel a high value of the cost function (no colour scale is given). The module orientation corresponding to the maximum value of the cost function is displayed by a white cross and the module orientation provided by the meta-information is represented by a white diamond.

In panel (3) the maximum value of the cost function obtained for each value of the angular loss coefficient (ordinate) is represented as a function of the angular loss coefficient (abscissa). The red cross represents the optimal angular loss coefficient.

Panels (4):

In this table, the available meta-information on the PV plant and the results of the parameter estimation are summarized. Common statistical measures of the simulation error obtained with the estimated parameters and the used meteorological data are also indicated.

Panels (5),(6):

In panel (5), a scatter plot of the power measurements corrected for the temperature effect at $25°C$ (ordinate) as a function of the effective irradiation (abscissa) is displayed. The colour of the scatter points represents the local scatter point density. A blue point corresponds to a point with a low local density and a red point to a high local density. The power curve estimated at $25°C$ is superimposed using a black dashed line.

In panel (6), the effect of the air temperature (x-axis) on the PV power (ordinate) is illustrated for four values of the effective irradiation. The different values of the effective irradiation are recognizable by the colour of the scatter points (very light blue, light blue, blue and green points), which correspond to effective

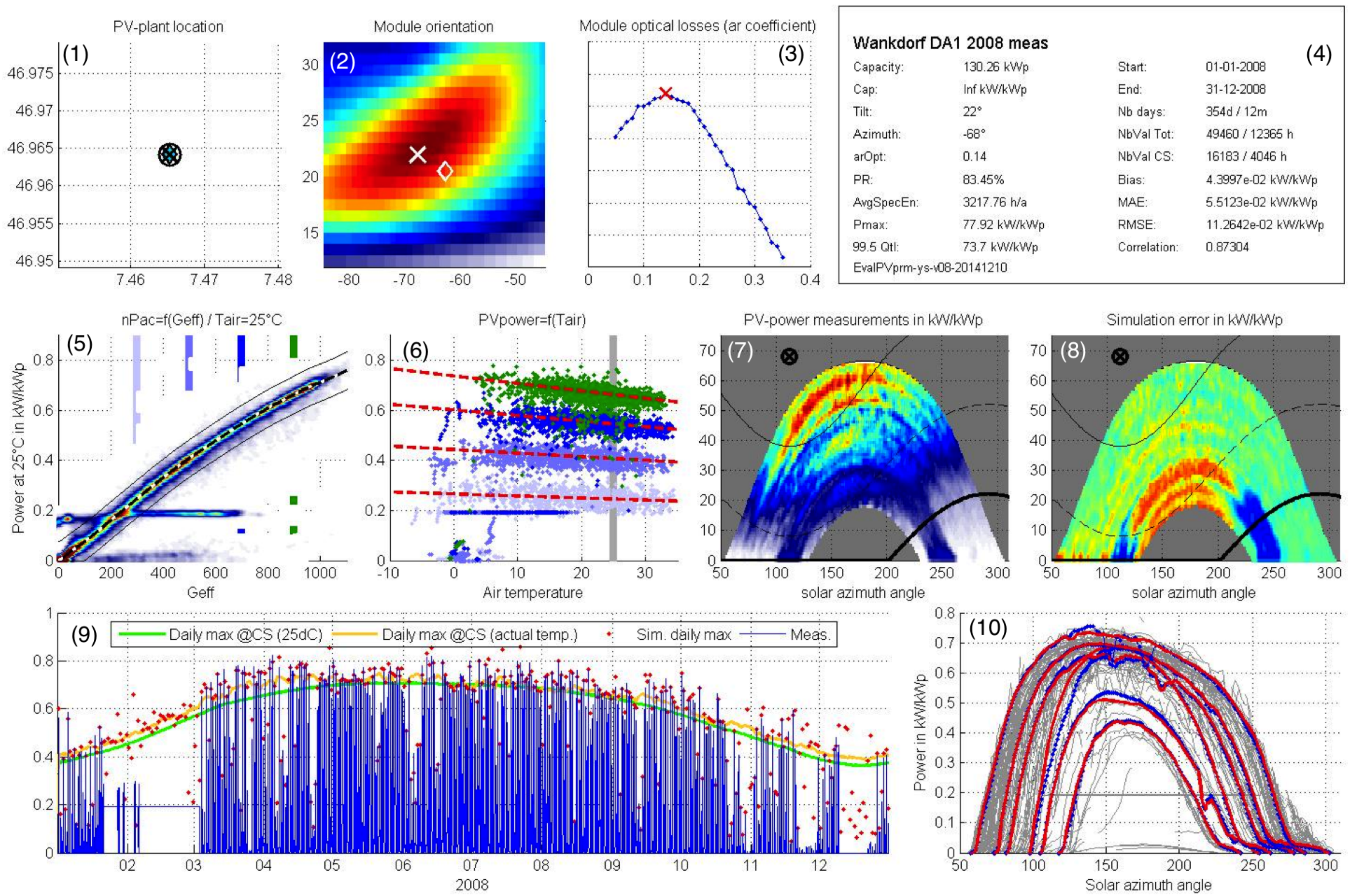

Figure .6: Report generated for the estimation of the parameters of the PV plant Wankdorf DA1 using local meteorological measurements.

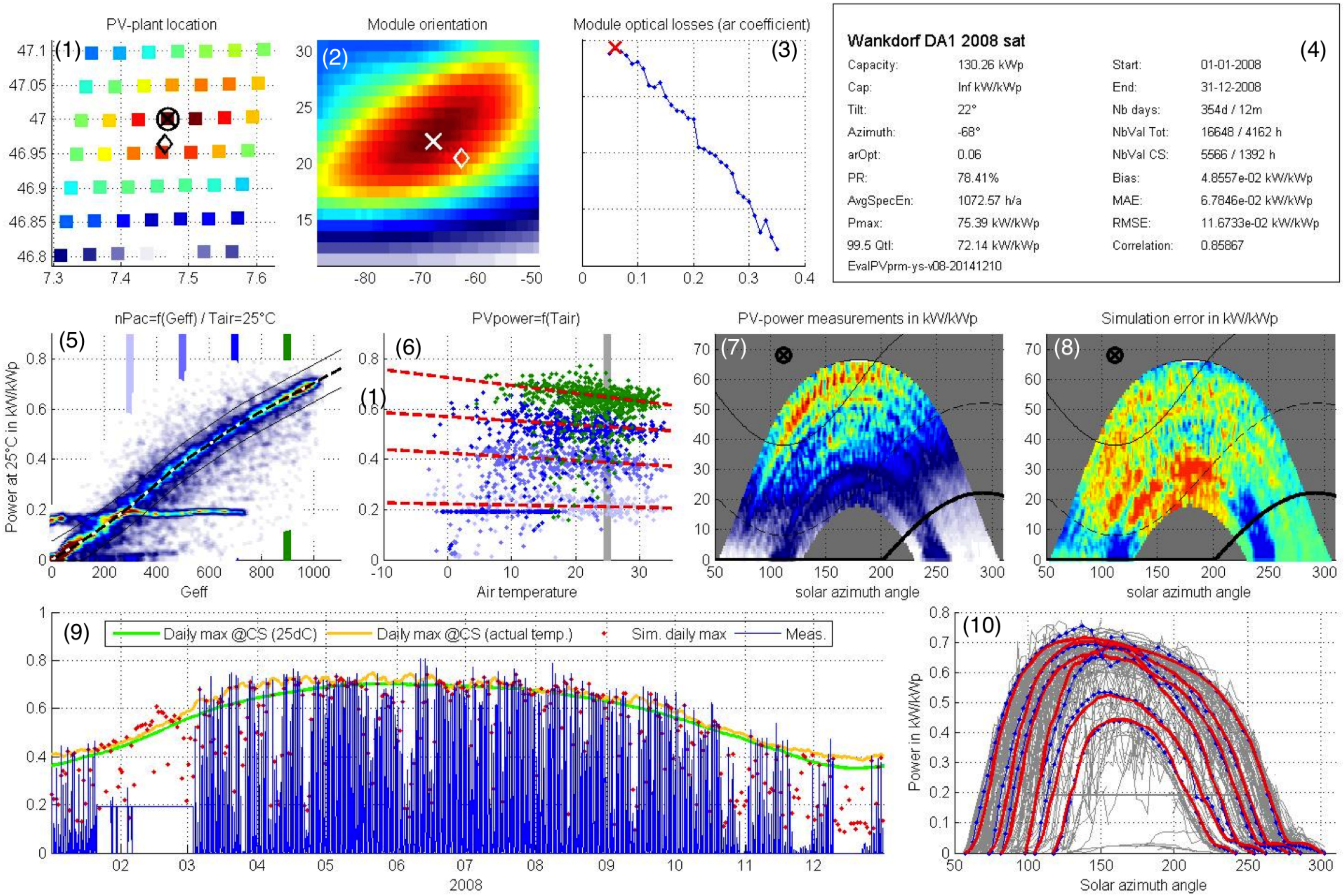

Figure .7: Report generated for the estimation of the parameters of the PV plant Wankdorf DA1 using satellite derived irradiation and analysis temperature.

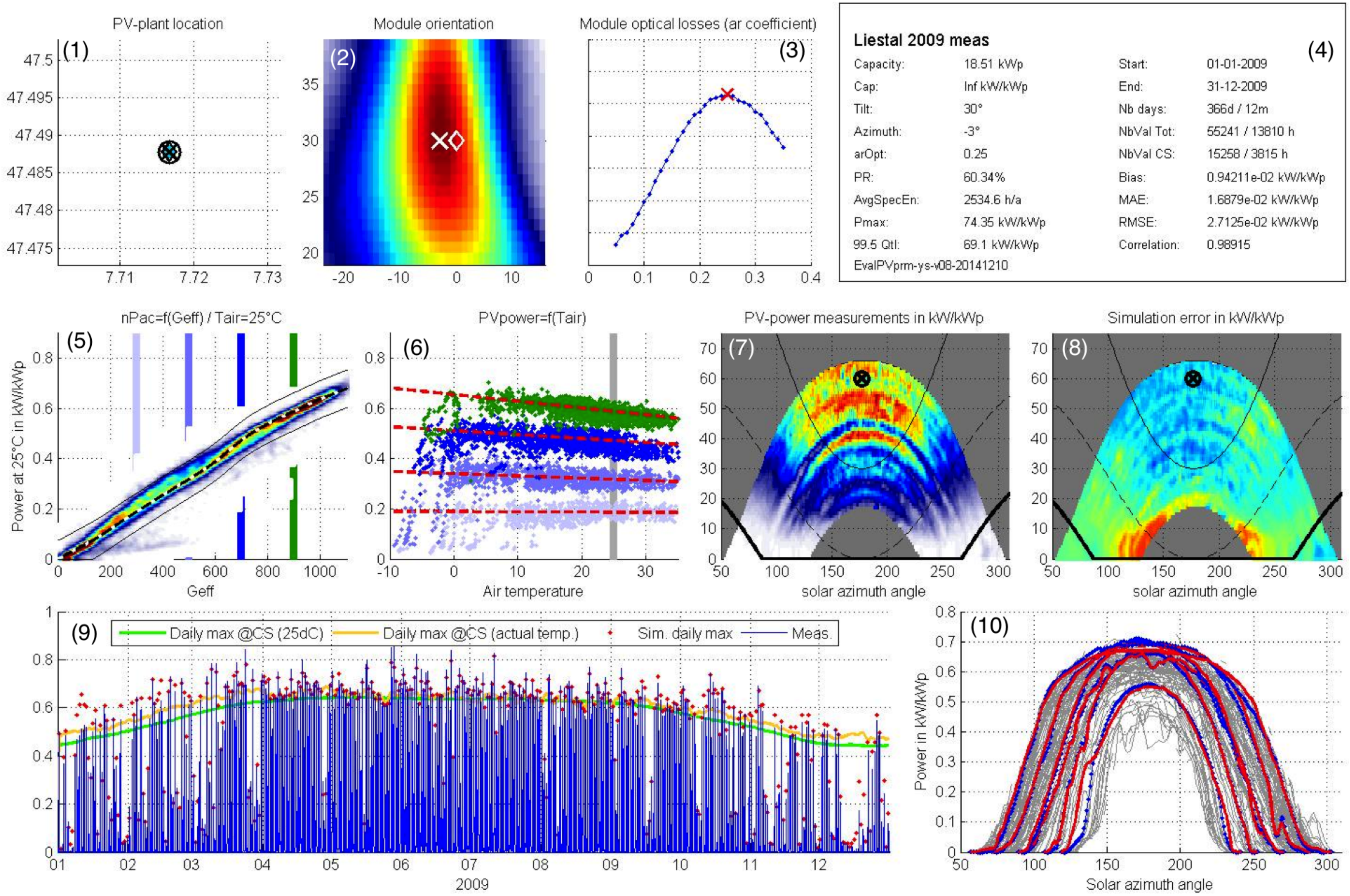

Figure .8: Report generated for the estimation of the parameters of the Liestal PV plant using satellite derived irradiation and analysis temperature.

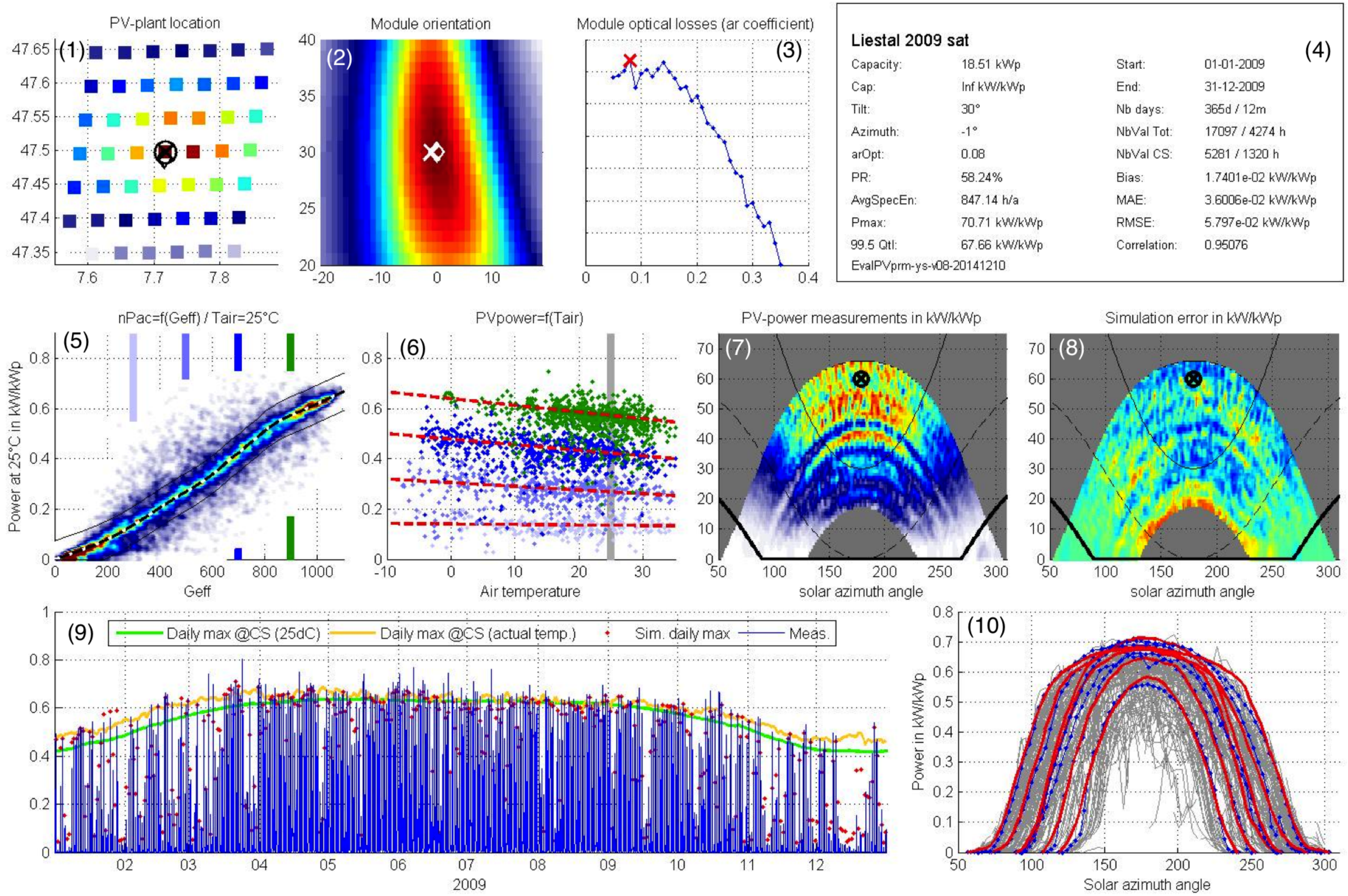

Figure .9: Report generated for the estimation of the parameters of the Liestal PV plant using satellite derived irradiation and analysis temperature.

irradiation values of 300, 500, 700 and 900 W/m^2, respectively. The red dashed lines are the values of the LUT corresponding to the different effective irradiations.

Panels (7):

The measured power values (colour of the scatter points) are displayed as a function of the solar azimuth (abscissa) and elevation angles (ordinate). White-to-blue points correspond to small power values (0 to 0.15 kW/kWp), while red points represent large power values (0.6 to 0.8 kW/kWp). Isolines of the incidence angles resulting from the estimated module orientation are shown (90, 60, 30 and 0°). Under clear-sky conditions, with the maximum power being reached at small incidence angles, a correspondence should be observable between the scatter points and the isolines of the incidence angles. Thus, the comparison of the two allows for verification of the estimated module orientation.

Panels (8):

The differences between the measurements and the power calculated with the estimated parameters (colour of the scatter points) are displayed as a function of the solar azimuth (abscissa) and elevation angles (ordinate). A light green point corresponds to a simulation error close to zero while blue points (red points) represent a simulated power 0.025 kW/kWp smaller (larger) than the measured power.

As in the previous plot, sun positions corresponding to incidence angles of 90, 60, 30 and 0° are displayed by three black lines and a black circle, respectively. This representation can be useful for identifying local shading effects on the power measurements.

Panels (9):

Time series of the simulation and measurements are compared for the entire training period. The measurements are shown by the blue line. To improve the readability of this graphic all simulated values were not displayed, but instead only the daily maximum of the simulated power. Additionally, the maximum daily simulated value that would have been reached under a clear-sky situation is represented by the yellow line. These two values of the simulation allow for the quick verification of the yearly shape of the measurements being well described by the simulation. These various outputs allow for verification that the seasonal variation of the PV power is described well by the estimated parameters.

Panels (10):

With the focus of panel (9) being on the yearly behaviour of the power data, the daily behaviour is represented in panel (10). For a better visibility, only clear-sky days are displayed here. Power measurements are displayed as a function of the solar azimuth instead of as a function of time, in order to avoid the effect of the yearly variation of the solar noon.

The power measurements are displayed for all selected clear-sky days by a light grey line. To avoid clutter, simulated power values are only displayed for 5 days chosen arbitrarily from amongst the set of clear-sky days. For these example days the measurements are displayed by a bold black line and the simulation by a red line.