Validation and self-shading enhancement for SoL: A photovoltaic estimation model

Carolina Nicolás-Martín*, Panagiotis Eleftheriadis, David Santos-Martín

University Carlos III Madrid, Spain

ARTICLE INFO

ABSTRACT

The estimation of electrical power generation in photovoltaic (PV) grid-connected systems based on meteorological data is a nontrivial, highly useful task, for instance to achieve accurate energy assessment. Widely used PV generation simulators are PV Systems (PVsyst), System Advisor Model (SAM) and PVLib. These simulators are characterized by presenting numerous features and providing complete results, however the PV estimation model SoL is an example of a new approach to PV generation estimation. SoL is characterized by its simplicity and computational efficiency. The objective of this paper is validating the recently published SoL model using real data from two PV locations for several years and facilities and comparing the results with those of three other PV simulators, namely PVsyst (in Spain), SAM (in Denver) and PVLib (both). It has been found that SoL estimates power production accurately for both locations and its estimations are more precise than those given by PVsyst, SAM and PVLib. It proves to be more computationally efficient than PVsyst, it can work with higher resolutions than SAM and PVsyst and requires fewer inputs than PVLib, SAM or PVsyst. Finally, a self-shading model is proposed as an enhancement for the SoL model. The number of inputs required is minimal, and it is an approximate yet efficient model. The estimation when using the self-shading enhancement is even more accurate than the previous estimation for SoL in locations where self-shading is evident. SoL proves to be an appropriate model for power estimation, and its results are enhanced when using the self-shading model proposed in this paper.

1. Introduction

The estimation of electrical power generation in photovoltaic (PV) grid-connected systems based on meteorological data is a nontrivial, highly useful task that can be applied, for example, for accurate energy assessment. A PV power simulator must encompass the synthesis and processing of meteorological and geographical data, as well as characteristics of the PV modules and inverter, to provide accurate power estimation. This power estimation is performed through the conversion of irradiance into the tilted plane (Perez et al., 1992; Hay and Davies, 1980; Gueymard, 2009), and from it into output power from the PV array (Jordan and Kurtz, 2013) and inverter (King et al., 2007; Chen et al., 2013).

Characteristics that enhance the value of a PV generation simulator are as follows: user friendliness, computational efficiency, accuracy of results, accessibility, and so on. Currently, there are numerous simulation tools for estimating PV power production (Smith and Reiter, 1984 provides a detailed overview of photovoltaic system performance models). Rodden et al. (2011) compares the performance of four different PV modelling tools. Some of the most relevant state-of-the-art simulation tools that can be used for PV estimation are as follows: The System Advisor Model (SAM), PV systems (PVSyst), and PVLib (Gurupira and Rix, 2017 and Stein et al., 2016).

These simulation tools are characterized by presenting numerous features and providing complete results. To provide such complete outputs, several inputs and extensive know-how about the program are needed. As a different approach to PV simulation, less complicated models have been developed (Zhou et al., 2007; Villalva et al., 2009 or Al-Amoudi and Zhang, 2000 among others). These models are characterized by simplicity, computational efficiency, and a reduced number of inputs.

The objective of this paper is to validate one of the simplest and more cost-effective models for PV estimation, namely the SoL model, (Santos-Martin and Lemon, 2015). The validation will be conducted by analyzing the error in power estimation using SoL and then comparing this error to the error rates of other state-of-the-art PV simulators. Partial shading in the modules has proven to have a significant impact on power production (Deline, 2009). The SoL model does not account

for partial shading between modules. Therefore, in addition to the model validation, a self-shading (SS) submodel is proposed as a new feature for SoL model for improvement of the results.

2. Description of the models used for validation

The SoL model is presented in Section 2.1. The two softwares used to compare the results of the SoL model, PVsyst and PVLib respectively, are introduced in Sections 2.2 and 2.3. Their structure, inputs, and main differences are explained. SAM, another PV software used for efficiency comparison in this paper, is also briefly explained in Section 2.4.

2.1. SoL model

The SoL model, explained in Santos-Martin and Lemon (2015), is a PV generation model for grid integration analysis in distribution networks. It is a simple model that does not need graphic design or specific requirements of the PV farm, and it focuses on simulating power generation given a certain number of inputs. The main goal of the model is fully describing the conversion of irradiance into electric power.

The system works using nine submodels with equations from state-of-the-art models (sometimes updated or reformulated to fit the needs of the global model). The structure and data flow of these nine submodels are represented in Fig. 1. These equations are generally algebraic relations with extremely low computational cost. This makes SoL a very convenient model for managing distribution networks, because of its functional nature (it can be used numerous times for automatic/stochastic calculations), however it is also suited for individual PV plant

Table 1
Required inputs for the SoL Model.

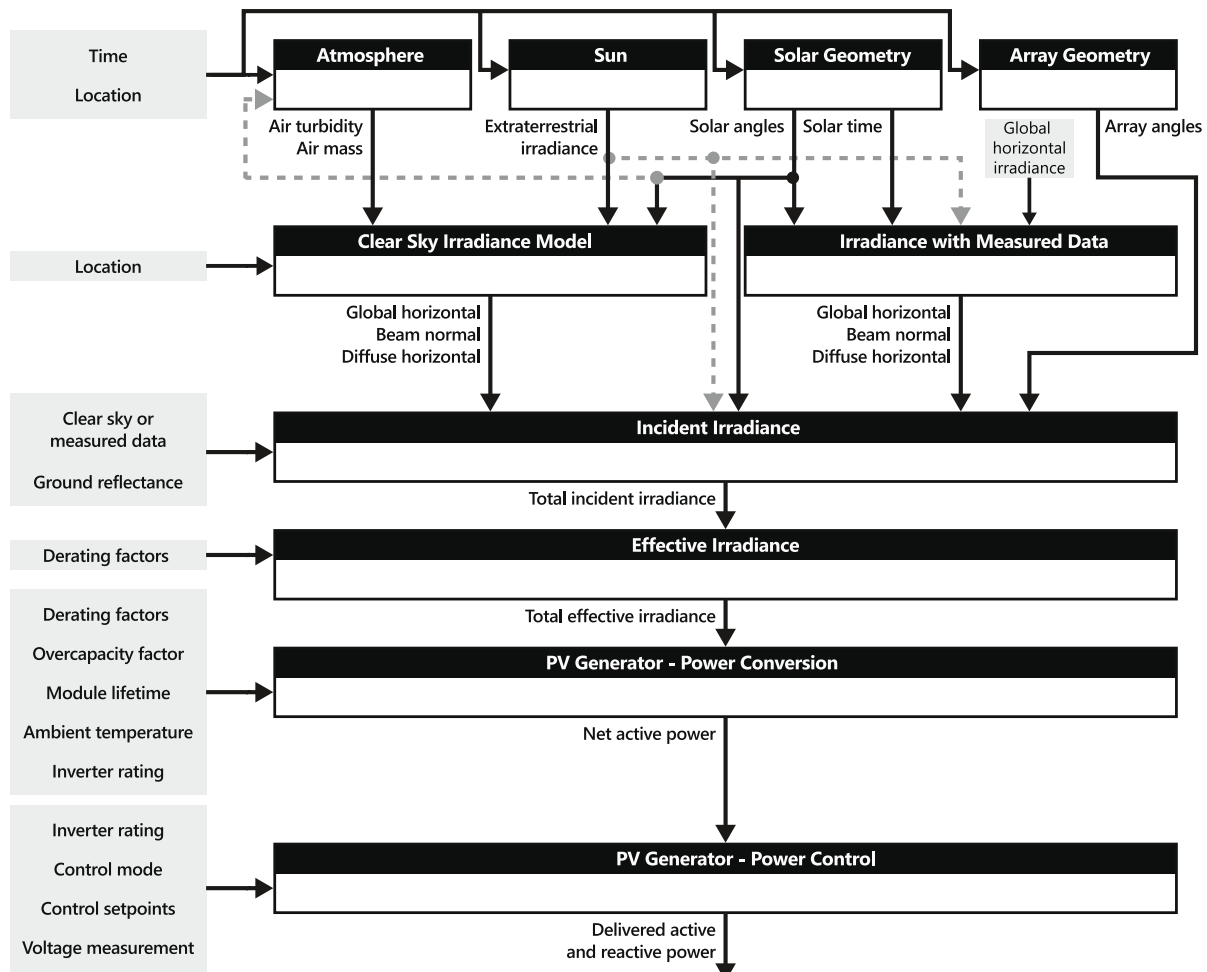| Input	Format	Units/ Range
Time	Structure with fields:	
	• Year	Positive integer
	• Month	Positive integer (1–12)
	• Day	Positive integer (1–31)
	• Hour	Positive integer (0–24)
	• Minute	Positive integer (0–60)
	• Second	Positive integer (0–60)
	• UTC offset	Integer
	• Day of the year	Positive integer
	• Daylight savings	1 daylight savings, 0 elsewhere
	Each field is a vector of equal length, except the UTC offset	
Location	Structure with fields:	
	• Latitude	degrees
	• Longitude	degrees
	• Elevation	meters
Rating of the inverter	Numerical value	Watts

Fig. 1. Flowchart of the SoL model showing submodels and key variables. Adapted from .Santos-Martin and Lemon (2015).

Table 2
Optional inputs for the SoL Model and its default values.

Input	Format	Units	Default value
Horizontal irradiance	Vector with values for each instant of time	Wm^{-2}	Clear sky irradiance values calculated by the Ineichen model
Ambient temperature	Vector with values for each instant of time	˚C	Constant 20 ˚C
Oversizing ratio	Numerical value	–	1.05
Module age	Numerical value	years	0
Power factor	Numerical value	–	1
Clearness index	Vector with values for each instant of time	-	Global horizontal irradiance extraterrestrial irradiance ratio, with value 0 when the sun elevation is very low.
Ground albedo	Numerical value	–	0.2
Azimuth offset	Numerical value	–	0
Linke turbidity pollution factor	Numerical value	–	0
Tilt angle	Numerical value	*degree*	Function of latitude
Effective irradiance factors	Structure with fields:		
	• Soiling	–	0.98
	• Shading	–	1
Derating factors	Structure with fields:		
	• Mismatch	–	0.98
	• Wiring	–	0.98
	• Connections	–	0.995
	• Light induced degradation	–	Function of module age, initial light induced degradation (0.985) and yearly light induced degradation (0.005)
	• Nameplate rating	–	0.99

analysis. Because of this functional nature, the SoL model is easily implemented as a Matlab function and can be easily operated by experienced or non-experienced users. For a list of well-defined inputs, some required (Table 1) and some optional (Table 2), there is a unique output, namely, the estimated power generation of the solar panels.

One of the key aspects of the SoL model is the reduced amount of inputs required. No specific information about the PV modules is needed due to the great effort of statistical summaries included in the several proposed submodels. Other parameters, are all considered with default values carefully chosen according to literature or regression analysis, Santos-Martin and Lemon (2015). Additionally, those default values can be easily replaced by real information regarding the PV plant (as optional inputs in Table 2), if it is available, thus enhancing the accuracy of results.

The resolution of the output data depends on that of the input data. The SoL model is designed to operate with an ideal resolution of 10 min. There is no internal database associated with the model; therefore, no meteorological inputs (temperature, horizontal irradiance) can be selected within the model given a certain location. The SoL model is designed so that the user is able to include their own local meteorological inputs. Local values for horizontal irradiance are strongly suggested, as the default clear sky model is not a common scenario.

The SoL model is, for now, limited to PV configurations with mono and polycrystalline modules (not suited for bifacial modules, DC-DC optimizers, HJT or CdTe). It is also not suited for PV systems using trackers. This is a significant limitation (for the time present) for the SoL model, since the current market is advancing towards trackers and bifacial PV modules according to the 2018–2019 International Technology Roadmap for Photovoltaic ITRPV (2019).

2.2. PVsyst

PVsyst is a commercial software program that focuses on modeling, sizing, designing, simulating, and analyzing PV systems. The inputs that the user is required to insert are numerous. The procedure for obtaining

the desirable results is complicated, and the steps must be followed with precision. In contrast, the software offers extensive reports and a variety of outputs. Because of its complexity, it can be highly detailed and tailored to the needs of each project. It is developed to meet the requirements of experienced users (engineers, architects) due to the requisite definition of the system (specific inverter, PV modules type and layout, etc.). All the above make it complete in terms of addressing all the needs of professionals. Its maximum data resolution is 1 h.

The steps needed to create and evaluate the results of a grid connected project in PVsyst are summarized in Fig. 2. The different inputs and outputs, as well the various options and settings, are schematically represented. PVsyst can provide default values for some inputs.

Information regarding PVsyst software and the physical models it is built on can be found from PVsyst Team (2019).

2.3. PVLib

PVLib is a set of well documented functions for simulating the performance of PV energy systems (see Sandia National Laboratories, 2019). A series of performance functions are developed and documented so that the user can benefit from them. PVLib is developed both in Phython and Matlab.

PVLib, unlike PVsyst or SAM, is not a software as such or a closed model like SoL. It is an open source toolbox. This means its functions are completely free, available and benefit from the input of various PV professionals. This makes it a very complete and adjustable tool containing functions for: time and location data treatment, irradiance and atmospheric functions, irradiance analysis and translation functions, PV system functions (including some parameter estimation functions). It also has some numerical utilities and very powerful databases.

Due to the adjustability of PVLib there are numerous ways to perform a PV grid connected power estimation analysis. The minimum number of inputs required for all the modules to work is shown in Fig. 3a. Module and inverter parameters can be found in the Sandia database, nevertheless, if the module cannot be found every parameter should be introduced manually or estimated. PVLib provides default
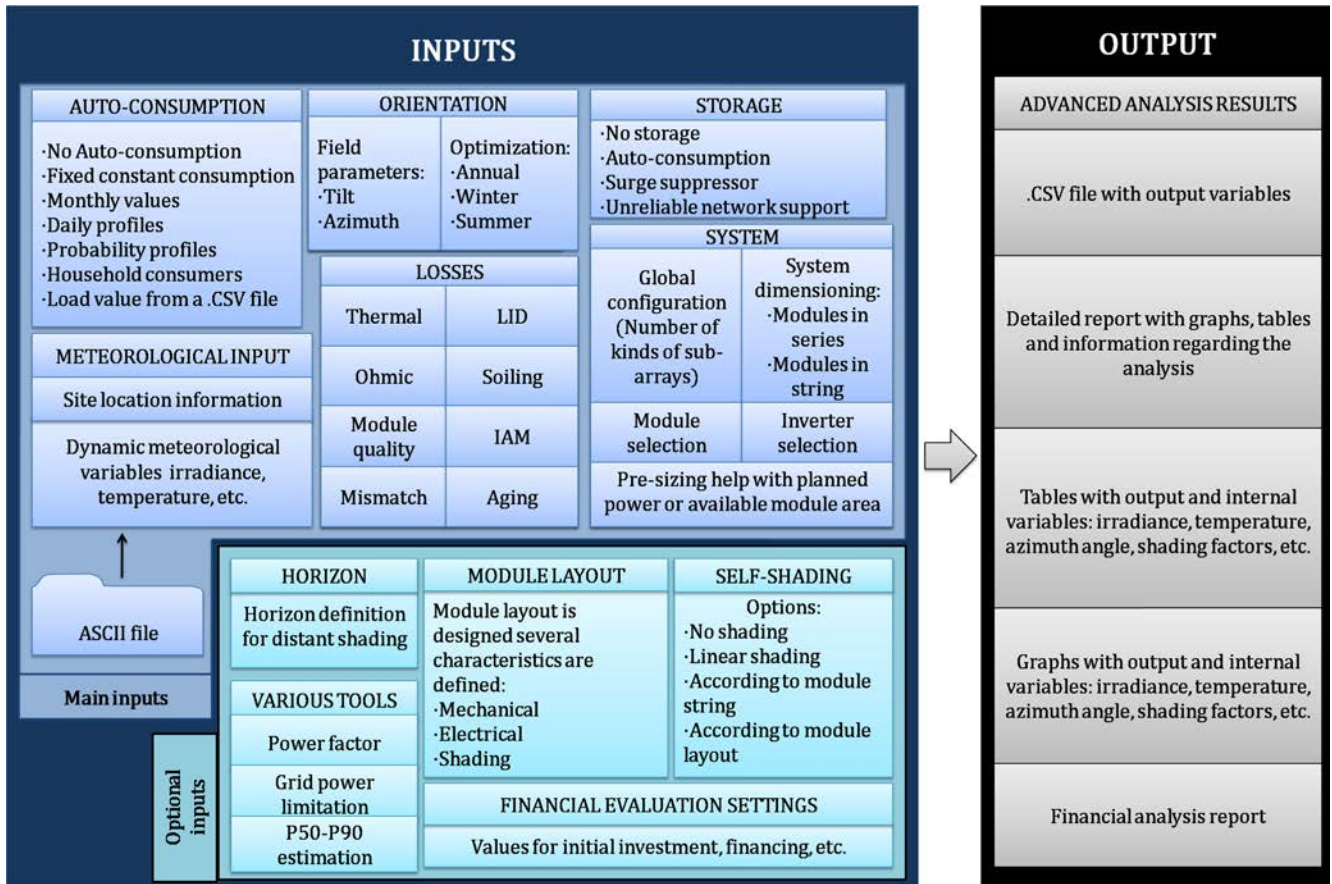
Fig. 2. PVsyst schematic use scheme.

values for other intermediate variables it uses (these are not specified for clarity purposes). The functions used with PVLib during the power estimation analysis in this paper are presented in Fig. 3b. This set of functions and inputs has been decided by combining the available information, minimizing the required number of inputs and adapting the available functions to the analysis to be performed.

2.4. SAM

SAM is a free performance and financial model for the renewable energy industry developed by the National Renewable Energy Laboratory (NREL). It is a powerful and complete tool, involving several models and databases. PV systems can be found among its performance models. It can provide power estimation based on meteorological and the PV system data for a maximum data resolution of 1 h. Information regarding the software can be found from National Renewable Energy Laboratory (2019).

3. SoL model validation

The validation of the SoL model is performed by comparing the output inverter power estimation from the model to that power measured from the inverter in real PV facilities. Two geographically diverse locations are selected for the study: A parcel in a PV farm in Tarragona, Spain is described in Section 3.1 and three NREL facilities in Denver, USA, are employed in Section 3.2.

The methodology of the model validation is described here. The power output, both measured and from the model, consists of data vectors with a given time resolution. The accuracy of the model results is measured according to several error indicators. Power error is only considered during power production hours, because night values give a 0% error. All the power errors are normalized by the maximum difference in power readings for the considered set of data. Since the minimum power reading will always be 0, it is equivalent to say that the power errors are normalized by the maximum power recorded. The error indicators are described below:

1. **Annual energy production signed error**: The annual energy production is computed based on the sum of power production over the given period of time. The annual percentage energy production error, δ_{model}, is computed according to Eq. (1), where E_{model} and E_{real} are the total annual energy generation values for the model and the real plant, respectively:

$$\delta_{model} = \frac{E_{model} - E_{real}}{E_{real}}. \tag{1}$$

2. **Power error indicators**: Root mean square error (RMSE), mean absolute deviation (MAD) and bias of each set of power data are considered. All these indicators are normalized, making the comparison between PV plants easier. The tracking signal (TS), which is the ratio between the bias and MAD value, is also considered to
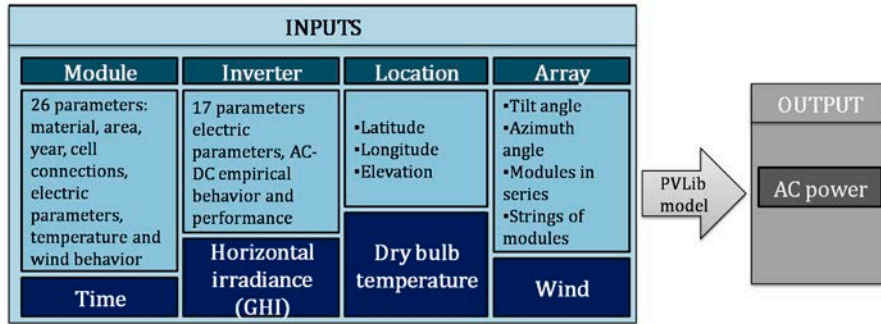
(a) Input-output diagram for PVLib
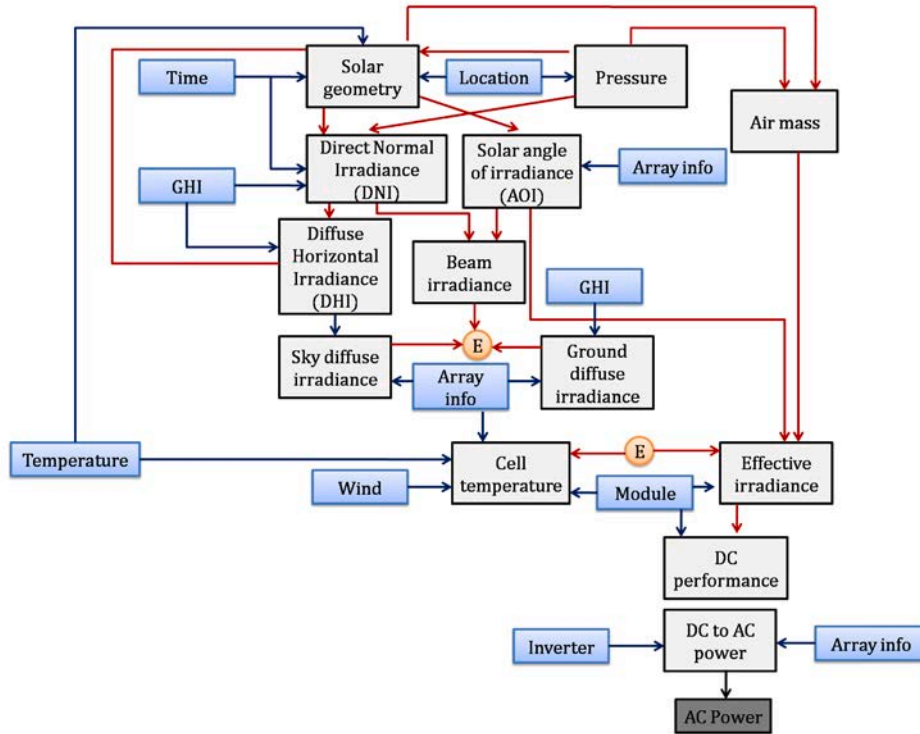
(b) Functions flowchart for PVLib. Inputs are represented in blue. Functions in white and output in grey. Internal variables are represented as orange circles.

Fig. 3. Inputs and functions flowchart for PVLib PV grid connected system estimation.
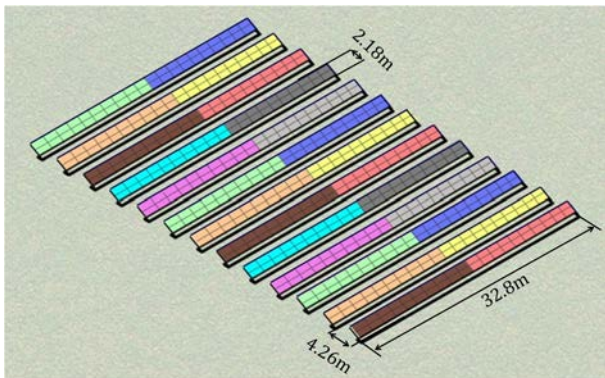

Fig. 4. Blueprint for the 21 parcel in Els Valentins PV farm.

Table 3
Energy percentage error for SoL model and PVLib for years 2015, 2016 and 2017.

	Year		
	2015	**2016**	**2017**
SoL model	3.32%	3.74%	3.49%
PVLib	7.23%	8.26%	8.42%

ensure that the bias in estimation is not excessive.

3. **Power error percentiles**: The 0th, 1th, 5th, 25th, 50th, 75th, 90th, 95th, 99th, and 100th percentiles of the normalized absolute power errors are calculated.

4. **Error heatplots**: To understand the distribution of errors, in

Table 4
Normalized power measurement error indicators for SoL and PVLib for years 2015, 2016, 2017. 10-min resolution data.

	RMSE		MAD		BIAS		TS	
	SoL model	PVLib	SoL model	PVLib	SoL model	PVLib	SoL model	PVLib
2015	3.59	4.52	2.43	3.36	1.39	3.03	0.57	0.9
2016	3.51	4.57	2.38	3.55	1.48	3.28	0.62	0.92
2017	3.94	5.11	2.56	3.86	1.51	3.64	0.59	0.94

(a) Heatmap for SoL MAD values for the year 2015

(b) Heatmap for SoL BIAS values for the year 2015

(c) Heatmap for PVLib MAD values for the year 2015

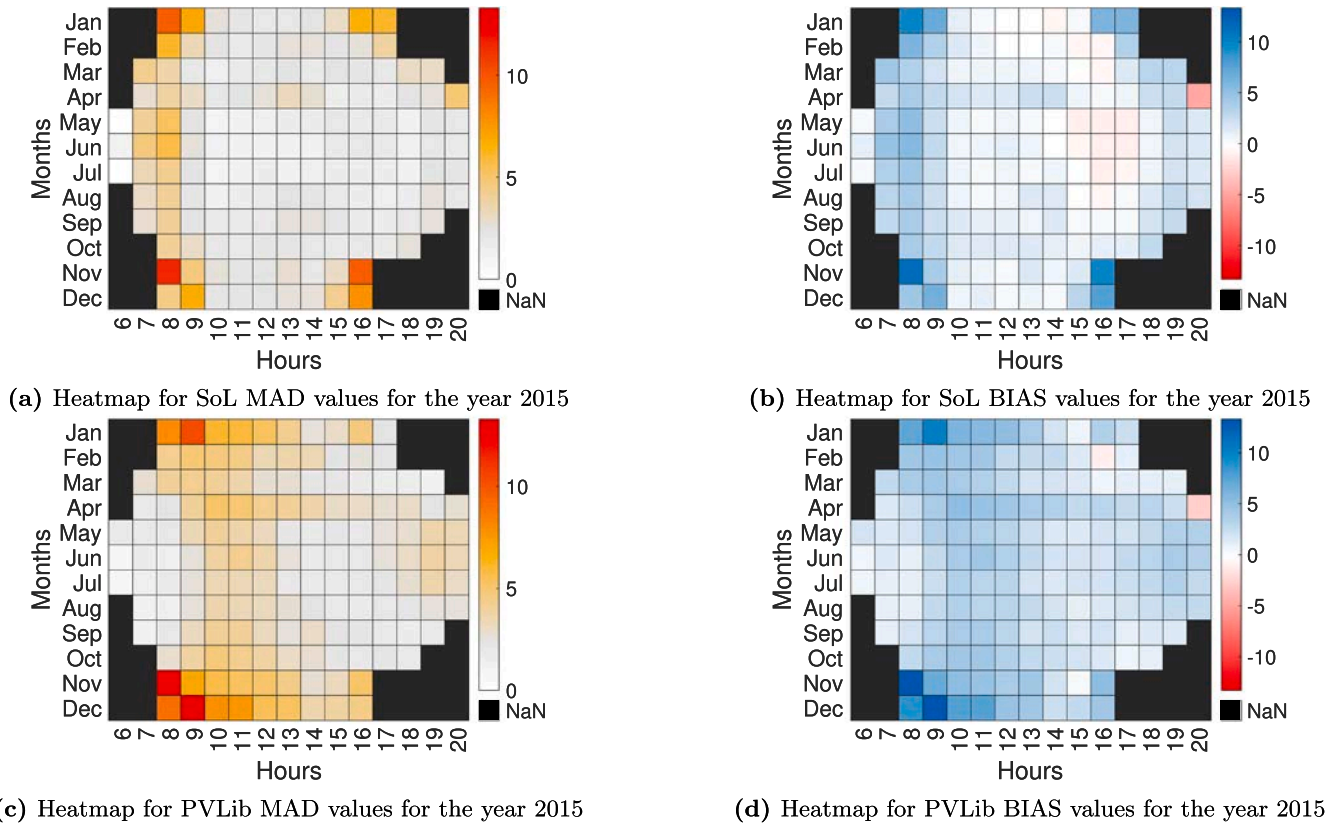**(d)** Heatmap for PVLib BIAS values for the year 2015

Fig. 5. Heatmaps of hourly normalized MAD and BIAS errors for Els Valentins.10 min resolution data.

Table 5
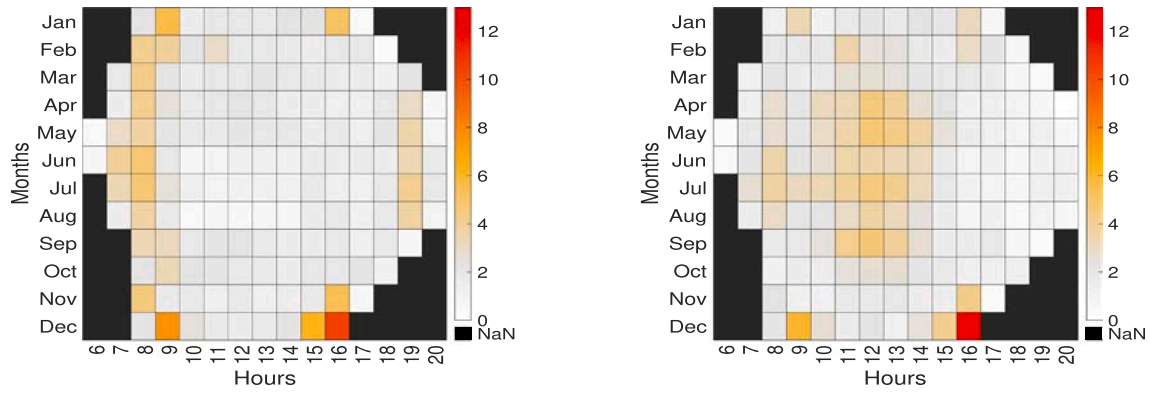Energy percentage error for SoL model and PVsyst for years 2015, 2016 and 2017.

	Year		
	2015	2016	2017
SoL model	3.25%	3.69%	3.54%
PVsyst	3.77%	3.91%	3.83%

addition to the magnitude and sign, hourly and seasonal error heatplots were considered. For each month, the MAD and bias normalized values for a certain power-producing hour are computed and displayed in a color map.

To complete the SoL model validation, the mentioned error indicators are compared with those of other models. Data which is not physically coherent, such as abnormal power production due to inverter shutdowns, incorrect reading of the sensor, or shutdowns due to snow, are excluded from the study.

Table 6
Normalized power measurement error indicators for SoL and PVsyst for years 2015, 2016, 2017. 1 h resolution data.

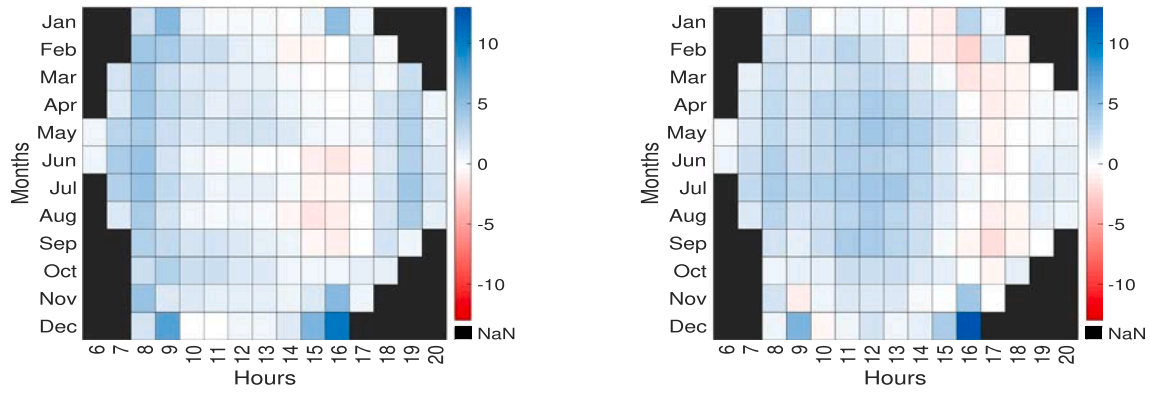|	RMSE		MAD		BIAS		TS	
	SoL model	PVsyst	SoL model	PVsyst	SoL model	PVsyst	SoL model	PVsyst
2015	2.98	4.94	2.13	3.45	1.32	1.53	0.62	0.44
2016	3.02	3.31	2.18	2.31	1.49	1.57	0.68	0.68
2017	4.32	4.55	2.53	2.68	1.58	1.61	0.63	0.64

(a) Heatmap for MAD values for the year 2016 for SoL (left) and PVsyst (right).

(b) Heatmap for BIAS values for the year 2016 for SoL (left) and PVsyst (right).

Fig. 6. Heatmaps of hourly normalized MAD and BIAS errors for Els Valentins. 1 h resolution data.

Fig. 7. Blueprint for the NREL facilities.

3.1. First location: Els Valentins, Spain

The SoL model has been used to estimate the power production of the 21st parcel of the Els Valentins PV farm located in Tarragona, Spain. The years 2015, 2016, and 2017 have been considered. To validate the model, its estimated power is compared with that read by the inverter at the plant (real power). The results are compared to the power estimated by PVLib (10-min data resolution) and PVsyst (1-h data resolution) in Sections 3.1.1 and 3.1.2 respectively. This is done by using the same inputs as those chosen for the SoL model (plus any other additional ones required by PVLib or PVsyst to function).

The PV farm is located in Partida de Els Anolls–Ulldecona

Table 7
Input data for SoL model for the three plants in Denver, USA.

	Model of employment	Facility		
		Visitor Parking	RSF2	STFS
Tilt angle	SoL, PVLib	8°	10°	10°
Azimuth angle (from North)	SoL, PVLib	165°	165°	164°
Location Latitude	SoL, PVLib	39.74°	39.74°	39.74°
Longitude		−105.18°	−105.18°	−105.17°
Elevation		1829 m	1829 m	1829 m
Inverter rating (kW)	SoL	500	500	75
Oversizing ratio	SoL	1.05	0.82	1.25
Module age	SoL	0	0	1 (2011) 2 (2012)
Module	PVLib	Sun Power SPR-315-E-WHT	Sun Power SPR-315-E-WHT	Evergreen Solar ES-190-RL
Inverter	PVLib	SM America's SC250U 480 V (2 inverters)	SM America's SC250U 480 V (2 inverters)	Sacton Technology Corporation PVS-75 480 V
Modules in series	PVLib	8	8	15
Strings of modules	PVLib	208	162	33

Table 8
SoL model and PVLib energy percentage errors for the Denver PV plants.

	PV plant			
	Visitor Parking	**RSF2**	**STFS**	
			2011	**2012**
SoL	−0.48%	1.27%	−0.06%	−0.1%
PVLib	5.51%	5.24%	7.08%	7.4%

(Tarragona, Spain). Its installed total capacity is 2.8 MW, which is distributed in 100-kW parcels. The arrangement of the solar panels, as well as the parcel and PV modules' dimensions, is described in Fig. 4.

The SoL model validation requires a series of data regarding the plant. These data are to be used as input for SoL, PVLib and PVsyst. All the data variables and parameters that have been used in the model validation are contained in Table A.16; its source, type, resolution, value, units of measurement, and model in which it is used, as well as any relevant comment related to how the data has been treated.

Monthly values for all the dynamic variables used in the study for the average of 2015, 2016, and 2017 are contained in Table B.17 as follows: horizontal irradiance, power measured in the inverter, and temperature.

For the power data analysis and comparison, a certain percentage of the data (7% for 2015, 5% for 2016, and 6% for 2017) has been excluded due to incoherent physical behavior as explained at the beginning of this section.

3.1.1. Validation of SoL against the real PV plant readings and PVLib comparative

The SoL model power estimation results are compared with the real power data provided by the Supervisory Control and Data Acquisition (SCADA) system in the plant. The SoL model uses the data described in Section 3.1 as input. The resolution of the input data, output power data from the SoL model and real power measures from the plant is adapted to 10 min (the ideal resolution of SoL model).

In addition, the same estimation is conducted with PVLib. The data resolution is also 10 min. The approach described in Fig. 3 is used. It should be noted that both the Siliken modules and inverter could not be found in the database. Therefore, their associated parameters were introduced manually or assumed to be the same as the module or inverter in the database with the closest characteristics. Wind data was also not available for Els Valentins. Since wind is a necessary input for PVLib wind was estimated to be a constant 0 m/s in order to eliminate the weight of that variable in the model. A wind versus annual energy sensibility test performed estimates a ~1% annual energy increase for every m/s for constant winds between 0 and 10 m/s.

Initially, the signed energy percentage error for the SoL model and PVLib is considered. The annual signed energy percentage errors for 2015, 2016, and 2017 are in Table 3. The estimation error for SoL is

positive and always inferior to 4% (see Table 3). This means that the SoL model slightly overestimates energy production for every year. PVLib overestimates energy production more than SoL does.

At this point, the power error must be considered. Several mean deviation indicators are shown to represent the characteristics of the power estimation error. RMSE, MAD, bias and TS for the power errors for each year studied are in Table 4. The power estimation absolute error percentiles (normalized) for 2015, 2016, and 2017 are found in Table C.20.

As shown in Table 4, MAD values are inferior to 3% for SoL for all three years. This is an acceptable error for instant power estimation. A higher value for the RMSE for 2017 indicates the existence of larger errors in this year, as RMSE gives a higher weight to higher deviations. The bias is positive for all the years. This indicates, in a similar way to the positive energy percentage errors in Table 3, that the SoL model overestimates power more frequently that it underestimates it. Nevertheless, the TS value indicates that the bias in the estimation is acceptable. Both absolute errors and bias are higher for PVLib.

To understand the distribution of the error data, and find possible sources for these errors, heat-plots with the MAD and bias values (both normalized) for the different hours in the day for each month in the year for 2015 (an arbitrarily chosen year) are represented in Fig. 5. Only one year is displayed, given that the heat-plot pattern is repeated during 2016 and 2017. By observing these heatplots two main conclusions can be drawn:

1. As a general repeated pattern for SoL during every year the biggest MAD values (see Fig. 5a) are focused on the first and final hours in winter. These few, low-power-producing hours accumulate the most significant errors in estimation. These errors are probably due to SS between modules, as the SoL model does not consider this effect and neither does PVLib.

2. By studying the bias heat-plots for SoL, it can be concluded that the biggest absolute errors are positive errors (supporting the SS hypothesis). The bias plots also show that the SoL model overestimates power for the first and last hours of every month (although greater errors are evident in winter months), while errors are not biased in any specific pattern for the remaining hours of the day. This is consistent for all the years studied.

3. By observing the bias heatplots for PVLib, it can be appreciated that, while the overestimation due to SS is also evident, there exist a clear and general overestimation independently of season or producing hours. This is thought to be due to the fact that PVLib does not incorporate a detailed electric losses (missmatch, connections, and so on) model, while SoL contains one with default values.

The power plots for the SoL model and PVLib outputs and the real power data from the Els Valentins plant for clear and cloudy days for each of the four seasons are represented in Fig. D.18. The error phenomenon explained in point 1 of the heatplot discussion can be clearly appreciated in Fig. D.18a. The SoL model clearly overestimates power
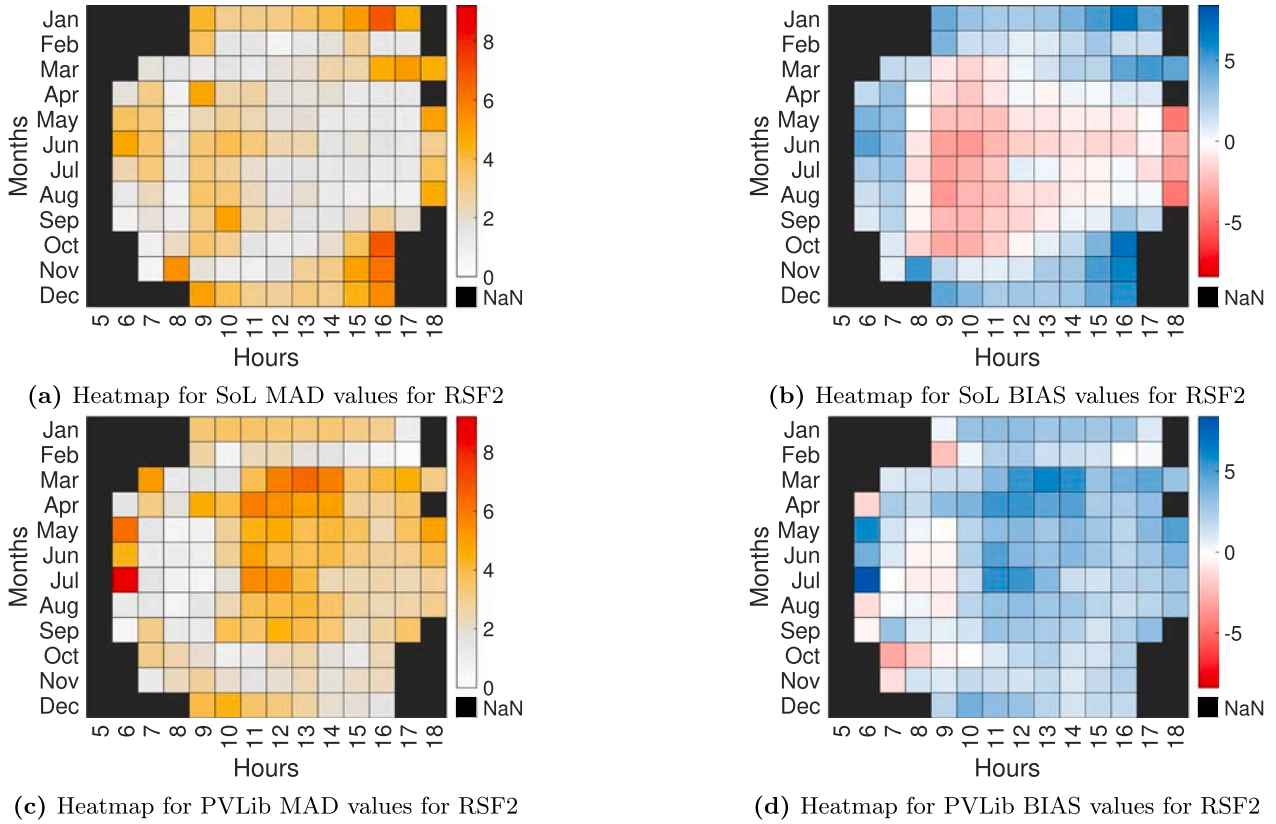
Table 9
Normalized power measurement error indicators for SoL model and PVLib for the Denver PV plants 1 h resolution data.

			RMSE		MAD		BIAS		TS	
			SoL	**PVLib**	**SoL**	**PVLib**	**SoL**	**PVLib**	**SoL**	**PVLib**
	Visitor Parking		4.02	4.45	3	3.4	−0.2	2.28	−0.07	0.67
	RSF2		3.77	4.42	2.53	3	0.54	2.25	0.22	0.75
STFS		**2011**	5.34	5.93	3.85	3.99	−0.03	3.5	−0.01	0.88
		2012	6.51	6.52	4.1	4.1	−0.05	3.54	−0.01	0.88

(a) Heatmap for SoL MAD values for RSF2

(b) Heatmap for SoL BIAS values for RSF2

(c) Heatmap for PVLib MAD values for RSF2

(d) Heatmap for PVLib BIAS values for RSF2

Fig. 8. Heatmaps of hourly MAD and BIAS normalized power errors for SoL and PVLib for Denver. 1 h resolution data.

Table 10
Comparison between SoL and SAM models normalized power errors. 1 h resolution data.

	Visitor Parking		RSF2		STFS			
	SoL	SAM	SoL	SAM	SoL		SAM	
					2011	2012	2011	2012
Annual energy error (%)	−0.48	−1.18	1.27	−1.36	−0.06	−2.2	−0.1	−2.73
Normalized power RMSE (%)	4.02	4.25	3.77	5.1	5.92		5	
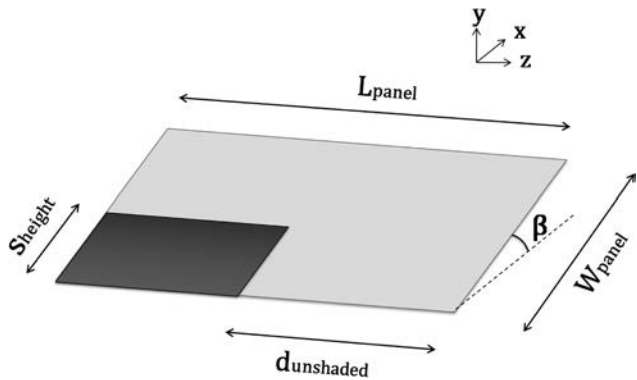
Fig. 9. Shadow dimensions in a PV panel.
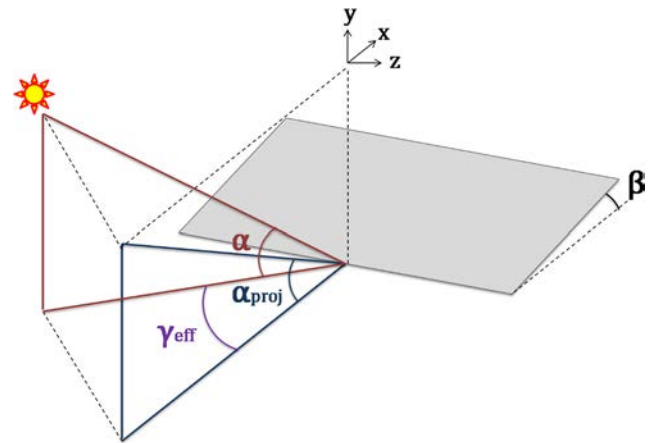
Fig. 10. Graphical representation of the projection of the elevation angle onto the plane perpendicular to the panel, α_{proj}.

production in the first and final hours in winter (see Fig. 5a), while for the rest of the seasons, the power estimation is significantly more accurate.

It can be concluded that while overestimation in the hours where SS is evident is the same for SoL and PVLib, SoL estimates power more accurately for the rest of the hours. The reason for this is that SoL considers electric losses and that inaccuracies for PVLib could be due to

the lack of some of the module and inverter input parameters. The assumption of a constant wind of 0 m/s is not considered to be a source of extra inaccuracies, because stronger wind, and refrigeration would
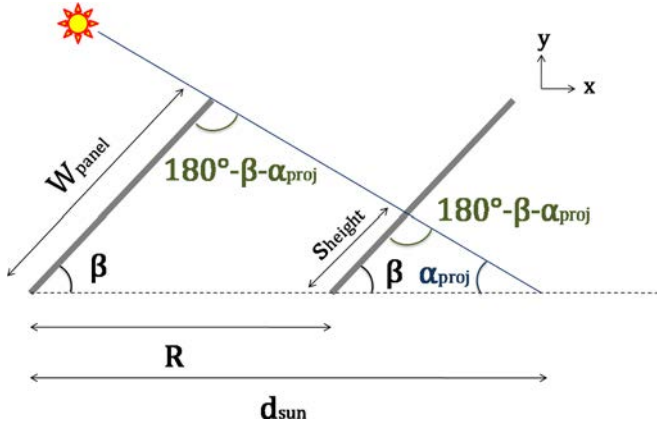
Fig. 11. Graphical representation of the height of the shadow of the panel, S_{height}.

imply even higher overestimations for PVLib.

3.1.2. SoL model validation compared with PVsyst

The SoL model power estimation output data are compared with the output data of the commercial software PVsyst. The highest data resolution allowed by PVsyst is 1 h. Therefore, the resolution of the input data for PVsyst, SoL model output, real power provided by the SCADA system and PVsyst for this study are also averaged for 1 h.

The annual signed energy percentage errors for the SoL model and PVsyst against real power data for 2015, 2016, and 2017 is listed in Table 5.

PVsyst presents a higher energy error for all three years studied. The energy error is positive, meaning that PVsyst also overestimates energy production.

Power error is now considered. The same error indicators used in Section 3.1.1 are employed. RMSE, MAD, bias, and TS values for 2015, 2016, and 2017 are found in Table 6. Note that, in Table 4, the errors are calculated with an estimation resolution of 10 min, in contrast to 1-h resolution in Table 6. The absolute power error percentiles for both models for all the years are indicated in Table C.21.

Absolute error indicators, such as the MAD or RMSE in Table 6 are also larger for PVsyst than for SoL model for each of the years studied.

PVsyst results also present a higher bias for each year, which is coherent with it presenting higher energy percentage error.

The heat-plots for MAD values for the different hours in the day for each month of 2016 for PVsyst and SoL model are represented in Fig. 6a. The error pattern is repeated for SoL and PVsyst for the three different years, and 2016 is chosen as an example. The equivalent heat-plots for bias values are represented in Fig. 6b. The error distribution varies between models as follows:

– PVsyst also concentrates the biggest inaccuracies during winter months, but only in the final hours in the day. In addition, it presents significantly large errors in high production during the summer months. The bias of the error also follows a different pattern than that of the SoL model. The biggest absolute errors also present a positive bias, but more negative bias values are registered than emerge in the SoL model. Most of these negative bias errors are concentrated in the last hours of the summer months, and present higher absolute error values than those in SoL.
  PVsyst does provide a SS option based on the module layout. However, when using this option, the results were more inaccurate than when the shading option was not used. PVsyst strongly underestimated power during shading hours. This is thought to be due to the 1-h data resolution, which is clearly insufficient for power estimation; this is discussed further in Section 4.1.3.

The error study between the two models shows that the SoL model presents to some extent more accurate results for both instant power and energy production for all the years studied. In addition to the error indicators discussed, examples of power estimation for clear and clouded days in each of the seasons are found in Fig. D.19. It can be confirmed visually that the SoL model power estimation curve is more accurate than that of PVsyst, especially on clear sky days.

In addition to presenting less accurate results for 1-h resolution data than SoL does, PVsyst power estimation is limited to a 1-h resolution, while the SoL model can work with higher and lower data resolutions. This limitation also affects accuracy and interpretation of the results as it can clearly be visualized by comparing Fig. D.18f to Fig. D.19f, as they both present results for the same day of the same year.
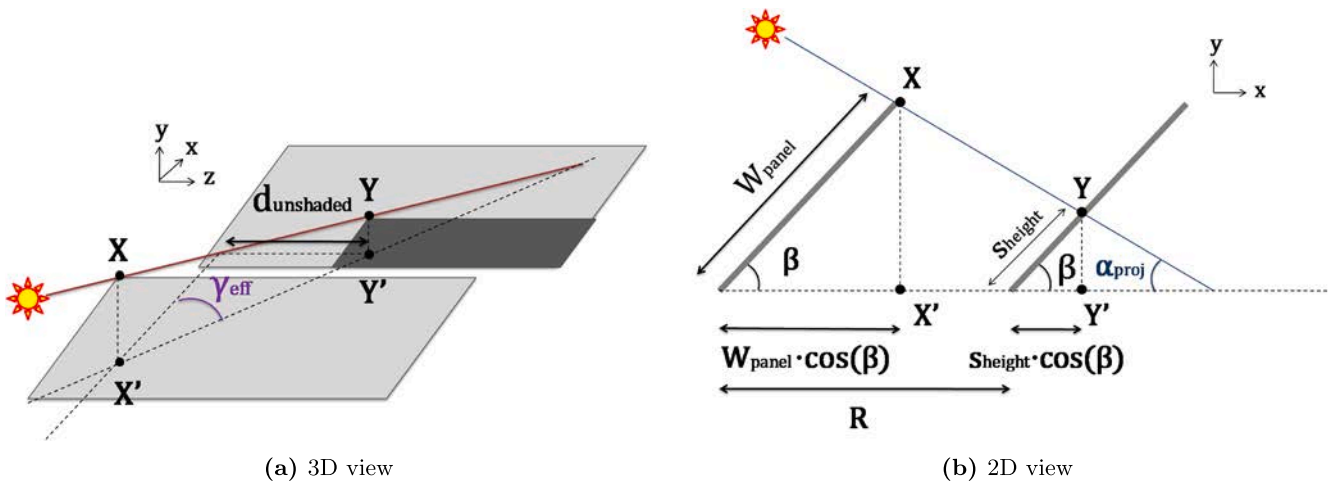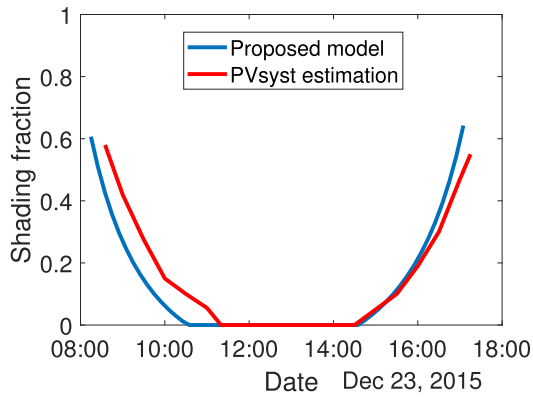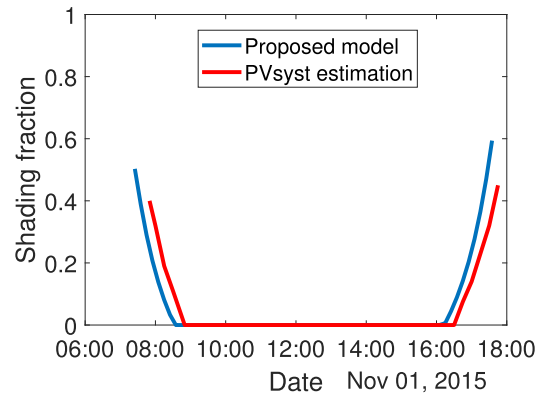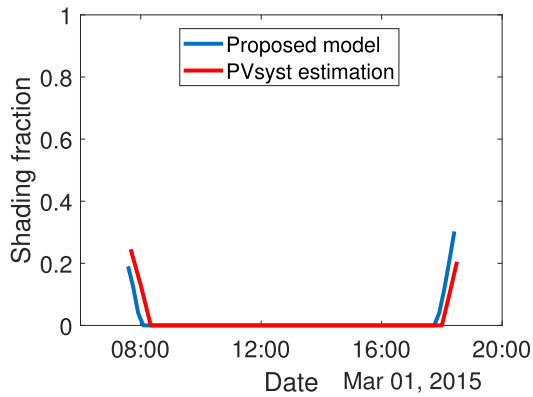

(a) 3D view **(b)** 2D view

Fig. 12. Graphical representation of the unshaded length of the panel, $d_{unshaded}$.
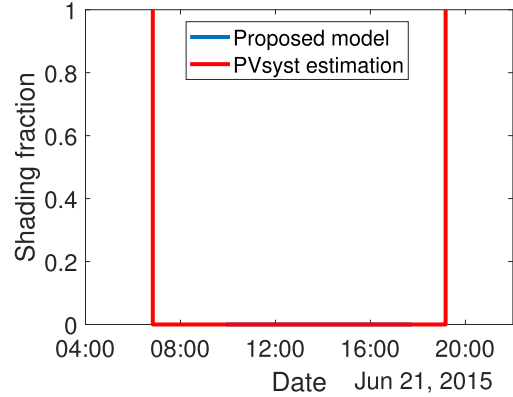
(a) Shading fractions for a winter day

(b) Shading fractions for an autumn day

(c) Shading fractions for a spring day

(d) Shading fractions for a summer day

Fig. 13. Shading fraction estimations from the proposed SS model and PVsyst SS feature using Els Valentins data.
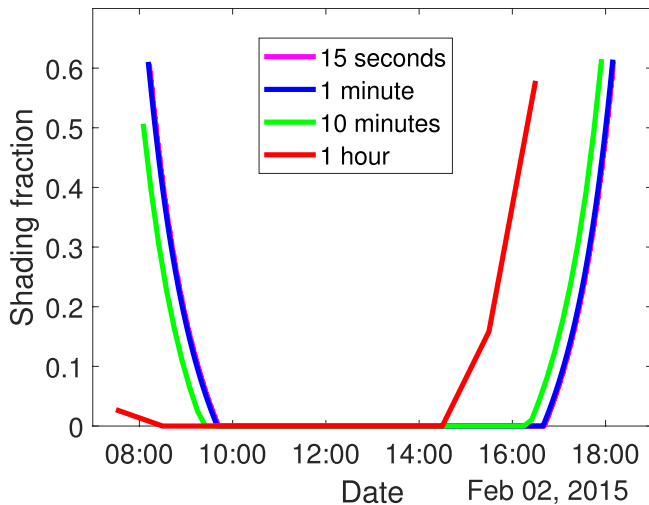

Fig. 14. Influence of data resolution in SS estimation. Els valentins data.

Fig. 15. Example of a module layout.

Table 11
Energy percentage error for SoL model for years 2015, 2016 and 2017 with and without SS feature. 10 min resolution data.

	Year		
	2015	**2016**	**2017**
SoL model	3.32%	3.74%	3.49%
SoL model with SS	2.33%	2.83%	2.48%

3.2. Second location: Denver, USA

The SoL model has been used to estimate the power produced by three different NREL PV facilities in Denver, Colorado, USA. All three of the research facilities are extremely close to each other. One is located on the visitor's parking lot roof at NREL and is referred to as Visitor Parking. The second is located at the second Research Support Facility and is referred to as RSF2. The last one is the smallest one and it is located in the Science and Technology Facility Study (it will be referred to as STFS). The model validation is conducted for 2012 for all three facilities, as well as for 2011 in STFS. Information from the facilities has been extracted from Freeman et al. (2013).

The measured power in each facility is compared with the output of the SoL model and will be also compared to the estimation provided by PVLib following a procedure analogous to Section 3.1.1.

A satellite view of the facilities is presented in Fig. 7. Information regarding the inputs from each plant and the model employing them (SoL, PVLib or both) is summoned in Table 7.

Time, global horizontal irradiance, and ambient temperature are the only dynamic variables used as input for SoL. PVLib additionally

Table 12

Power measurement error indicators for years 2015, 2016, 2017 with and without SS feature for 10-min resolution data

	RMSE		MAD		BIAS		TS	
	SoL model	SoL model SS	SoL model	SoL model SS	SoL model	SoL model SS	SoL model	SoL model SS
2015	3.59	3.08	2.43	2.22	1.39	0.98	0.57	0.44
2016	3.51	2.87	2.38	2.18	1.48	1.13	0.62	0.52
2017	3.94	3.26	2.56	2.31	1.51	1.08	0.59	0.47

Fig. 16. Influence of the SS feature.

Table 13

SS model variables.

	Variable	Description	Range/Units
Input	R	Distance between panels in the array	m
	n_{column}	Number of columns of modules in a panel	integer
	n_{row}	Number of rows of modules in a panel	integer
	L_{module}	Length of the PV module	m
	h_{module}	Height of the PV module	m
Input[†]	P_n	Net active power	W
	α_s	Solar elevation angle	(0°,90°)
	γ_s	Solar azimuth angle	(0°,360°) clockwise from N
	γ	Azimuth angle of the PV from North	(0°,360°)
	β	Tilt of the PV array from horizontal	(0°,90°)
	G_{bi}	Beam (direct) incident irradiance	W m^{-2}
	G_{ri}	Reflected incident irradiance	W m^{-2}
	G_{di}	Diffuse incident irradiance	W m^{-2}
Output	P_{SS}	Net active power considering SS	W
	s_{height}	Height of the shaded fraction of a PV panel	(0, $L_{module} \cdot n_{row}$) in m
	$d_{unshaded}$	Length of the unshaded fraction of a PV panel	(0, $h_{module} \cdot n_{column}$) in m
	d_{sun}	Distance from the origin of a PV panel to the intersection point between the ray of sun and the ground	m

[†] Input from another submodel notes.

3.2.1. Validation of the SoL model against the real readings of the PV plant and PVLib comparative

The measured power for the three PV facilities is compared with the output of the SoL model. The SoL model uses the data described in Section 3.2. The resolution of the meteorological data and power output
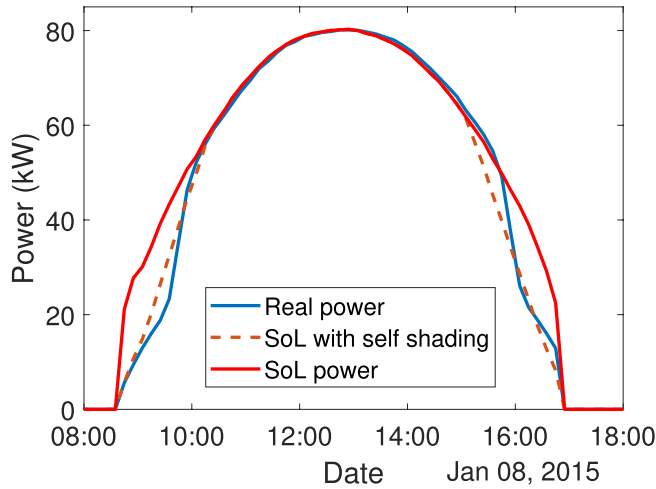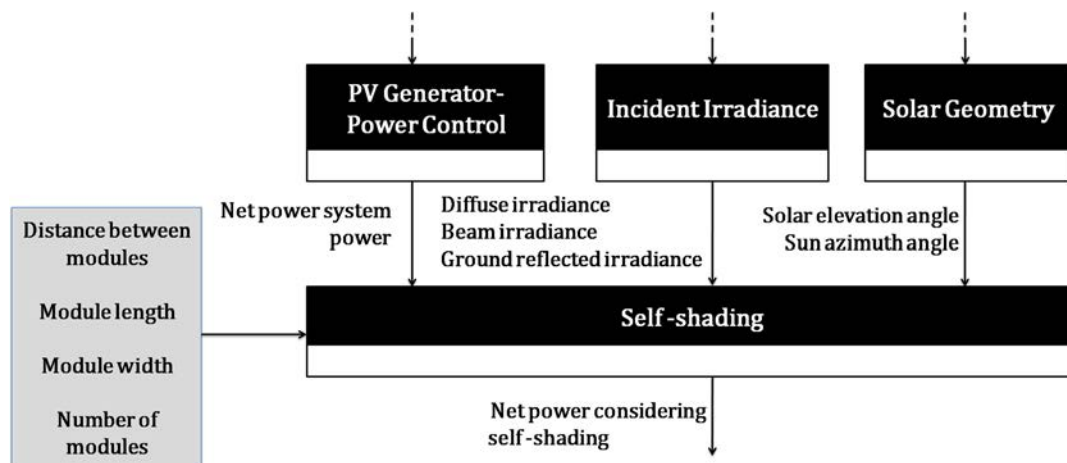
requires wind data. Hourly meteorological data for the site were provided by the Solar Radiation Research Laboratory. The meteorological data are the same for all three facilities, as they are situated in close proximity to each other. The hourly resolution of the data was not altered and it is used for simulating power output with the SoL model and PVLib. The monthly values for the meteorological data used for both PVLib and SoL, as well as the measured power in each of the facilities, is shown for 2011 in Table B.18 and 2012 in Table B.19.

For the power data analysis and comparison, a certain percentage of the data (14% for Visitor Parking, 7% and 6% for STFS in 2011 and 2012, respectively and 10% for RSF2) has been excluded due to incoherent physical behaviour (as explained in Section 3). Reasons for this significant amount of excluded data are snow covering the panels during winter months and accidental inverter shutdowns.

Fig. 17. Addition of the SS model into SoL's structure.

Table 14
SS model equations.

Id	Equation	References
(16)	$d_{sun} = \begin{cases} h_{module} \cdot n_{column} \cdot \dfrac{\sin\left(180° - \beta - \arctan\left(\frac{\tan\alpha_S}{\cos(\gamma_S - \gamma)}\right)\right)}{\sin\left(\arctan\left(\frac{\tan\alpha_S}{\cos(\gamma_S - \gamma)}\right)\right)} & \text{if } |\gamma_S - \gamma| \leqslant 90° \\ NaN & \text{otherwise} \end{cases}$	Adapted from Braun and Mitchell (1983) and Duffie and Beckman (1980)
(17)	$s_{height} = \begin{cases} h_{module} \cdot n_{column} - R \cdot \dfrac{\sin\left(\arctan\left(\frac{\tan\alpha_S}{\cos(\gamma_S - \gamma)}\right)\right)}{\sin\left(180° - \beta - \arctan\left(\frac{\tan\alpha_S}{\cos(\gamma_S - \gamma)}\right)\right)} & d_{sun} > R \\ NaN & \text{otherwise} \end{cases}$	Adapted from Braun and Mitchell (1983) and Appelbaum and Bany (1979)
(18)	$d_{unshaded} = \tan(\gamma_S - \gamma) \cdot (R + (s_{height} - h_{module} \cdot n_{column}) \cdot \cos\beta)$	Adapted from Braun and Mitchell (1983) and Appelbaum and Bany (1979)
(19)	$P_{SS} = \begin{cases} P_n \cdot \dfrac{G_{di} + G_{ri}}{G_{di} + G_{ri} + G_{bi}} & \text{if } s_{height} > h_{module} \text{ and } d_{unshaded} > L_{module} \\ P_n \cdot \dfrac{G_{di} + G_{ri} + \left(1 - \frac{s_{height}}{h_{module}}\right) \cdot G_{bi}}{G_{di} + G_{ri} + G_{bi}} & \text{if } 0 < s_{height} \leqslant h_{module} \text{ and } d_{unshaded} > L_{module} \\ P_n & \text{otherwise} \end{cases}$	Adapted from Deline et al. (2013)

is of 1 h. The same procedure is repeated also for PVLib.

Initially, the signed percentage error for the SoL model is considered. The annual signed energy percentage error for all the facilities and years studied is contained in Table 8.

It can be observed that SoL estimates the annual energy production very accurately for the three plants considered (less than 1.5% error for all of them). It slightly underestimates production for STFS and Visitor Parking, while it overestimates the annual production for RSF2. The behaviour is different for PVLib as it overestimates energy production for every facility.

The power error is now considered. RMSE, MAD, and bias error indicators and the TS are found in Table 9. STFS holds the largest RMSE and MAD values (absolute error indicators). Although slightly larger than those described in Section 3.1.1 in Els Valentins, the absolute error indicators for SoL for the rest of the facilities are under 5%. The bias values are coherent with the annual production results of Table 8. The TS indicates that bias in estimation is perfectly acceptable. While absolute error indicators for PVLib are similar to those of SoL they are always biased. The normalized SoL model and PVLib error percentiles for all three facilities are presented in Table C.22.

To understand the distribution of errors, the MAD and bias heatplots for RSF2 are presented in Fig. 8. The error distribution is consistent for all three facilities; therefore, for simplicity, only RSF2 is shown. The error distribution for SoL varies significantly from that shown in Section 3.1.1 for Els Valentins. The conclusions to be drawn are the following:

1. Winter months show higher absolute error values than any other month, accumulating the highest errors in the last hours of such months. However, although the biggest absolute errors are concentrated on the last hours, the Denver facilities accumulate bigger inaccuracies in peak production hours than those in Els Valentins. In this case, no shading is expected, as the PV modules have a low tilt angle due to being installed on the roof.
2. A clear seasonal bias is evident for SoL when observing Fig. 8b. In winter months, the power is over estimated by SoL, while in the central and last hours of summer the months, the power is lower for the SoL model's output than it is in reality.
3. Like in Section 3.1.1, there is a clear general overestimation of power for PVLib due to the lack of consideration of electrical losses.

The seasonal bias phenomenon for SoL and the general power overestimation for PVLib can be visualized in Fig. D.20 which represents the real, SoL and PVLib power output plots for clear and cloudy days for each season for RSF2.

3.2.2. Comparison between the SoL model and SAM power errors

Freeman et al. (2013) published a validation report for the PV simulator SAM using the same facilities employed for SoL validation in Section 3.2.1.

The annual energy errors and RMSE normalized values for the three facilities studies for SoL and SAM are in Table 10.

The annual energy percentage errors are always smaller for the SoL model, which estimates the annual power production more accurately. It is specially remarkable how, for RSF2, SoL overestimates the energy production, while SAM underestimates it. For the remaining cases, the sign of the error is the same for both models. The SoL model presents lower RMSE values for the Visitor Parking and RSF2, while SAM presents a better RMSE value for STFS.

It should be noted that, aside from globally providing more accurate results for the facilities studied, SoL is a simpler model than SAM is. SoL requires no information regarding the type or layout and dimensions of the modules or type of inverter, while SAM does. In addition, SoL results were expected to be suboptimal because its optimal resolution (10 min) could not be used.

4. The SS enhancement feature for SoL

Recurrent small inaccuracies were found for SoL model power estimation during the first and last hours for the winter months for the Els Valentins study. They arose because SoL does not account for SS power losses. SoL does include an optional shading loss factor input; however, the entire power output is affected by this factor. SS could be considered in terms of annual energy production by this shading factor, but instant power errors remain.

To account for the instant power losses due to SS and improve the overall accuracy of SoL, a simple SS model is presented in this section. The proposed SS model consists of 2 submodels. The first (Section 4.1) estimates the geometry of the shaded portion of the module array as a function of known solar angles and the basic dimensions of the PV array and it is adapted from Braun and Mitchell (1983). The theoretical

formulation of this submodel is explained in Section 4.1.1 and validated against the Els Valentins data in Section 4.1.2. The second submodel (Section 4.2) estimates a power derating factor given the shadow dimension in the array and the connections between modules. The theoretical formulation of this submodel is explained in Section 4.2.1 and validated against the Els Valentins data in Section 4.2.2. The introduction of the SS model as a feature for SoL is explained in Section 4.3.

4.1. Shadow dimension submodel

The shadow dimension submodel proposed here provides the dimensions of the shaded portion of the panels of an array as an output. The submodel can be used for a series of n_{panel} solar panels of constant length L_{panel} and width W_{panel}, separated by a constant distance R and with a fixed tilt angle β, located on a flat field surface. The shadow formed on a panel because of its adjacent one is assumed to be perfectly rectangular, as a simplification. The shadow dimension height is s_{height}, and the unshaded length of the panel is $d_{unshaded}$ (see Fig. 9).

The purpose of the submodel is providing a series of general and simple expressions to calculate the dimensions of the shadow as a function of array information, comprising n_{panel}, L_{panel}, W_{panel}, R, β, the solar elevation angle α_s, and the effective azimuth angle γ_{eff}, which is calculated according to Eq. (2), where γ_s is the sun azimuth angle and γ is the panel surface azimuth angle:

$$\gamma_{eff} = \gamma_s - \gamma. \tag{2}$$

In Section 4.1.1 the theoretical formulation of the submodel, based on geometrical relations, is explained. In Section 4.1.2 this submodel is used for the data from Els Valentins and validated against the PVsyst shadow fraction feature.

4.1.1. Theoretical formulation

First, the projection of the solar elevation angle, α_s, onto the plane perpendicular to the ground and parallel to the normal of the panel surface is calculated. The projection plane corresponds to plane xy in Fig. 10:

The projected elevation angle, denoted as α_{proj} can be calculated according to Eq. (3), adapted from Braun and Mitchell (1983), using trigonometric functions.

$$\alpha_{proj} = \arctan\left(\frac{\tan\alpha}{\cos\gamma_{eff}}\right). \tag{3}$$

The trajectory of a ray of sun seen through plane xy and the height of the shadow one panel projects onto the adjacent one, s_{height} is represented in Fig. 11.

The distance d_{sun} can easily be calculated through the sine theorem, as shown in Eq. (4), adapted from Duffie and Beckman (1980):

$$d_{sun} = W_{panel} \cdot \frac{\sin(180° - \beta - \alpha_{proj})}{\sin(\alpha_{proj})}. \tag{4}$$

Once the distance d_{sun} is computed, the shadow height, s_{height}, can be calculated using the same theorem. This is shown in Eq. (5), adapted from Appelbaum and Bany (1979):

$$s_{height} = \left(d_{sun} - R\right) \cdot \frac{\sin\alpha_{proj}}{\sin(180° - \beta - \alpha_{proj})}. \tag{5}$$

The length of the unshaded part of the panel, $d_{unshaded}$, delimited by the ray of sun acting on the highest leftmost point on the adjacent panel is represented in Fig. 12.

The expression for calculating $d_{unshaded}$, Eq. (6), is adapted from Appelbaum and Bany (1979). The projection of the distance $\overline{X'Y'}$ onto the xy plane (Fig. 12a) is computed, and then, by simple trigonometry, $d_{unshaded}$ is obtained (Fig. 12b, Fig. 13):

$$d_{unshaded} = \tan\gamma_{eff} \cdot (R + (s_{height} - W_{panel}) \cdot \cos\beta). \tag{6}$$

In order for Eqs. (5) and (6) to be physically coherent, they can only be used provided all the conditions in (7)–(9) are met:

$$|\gamma_{eff}| \leqslant 90°, \tag{7}$$

$$\alpha > 0, \tag{8}$$

$$d_{sun} > R. \tag{9}$$

Analogously, the dimensions of the shadow should be bounded to the maximum dimensions of the panel. This is translated as Eqs. (10) and (11):

$$\text{if } s_{height} > W_{panel} \text{ then } s_{height} = W_{panel} \tag{10}$$

$$\text{if } d_{unshaded} > L_{panel} \text{ then } d_{unshaded} = L_{panel} \tag{11}$$

The fraction of a panel that is shaded, χ_{panel}, can be easily calculated by dividing the shaded area by the total area of the panel (see Fig. 9), as shown in Eq. (12):

$$\chi_{panel} = \frac{s_{height} \cdot (L_{panel} - d_{unshaded})}{W_{panel} \cdot L_{panel}}. \tag{12}$$

The fraction of the total area of the array of panels that is shaded is computed assuming that every panel is equally shaded, except for the first panel, which is not shaded at all because it has no other panel in front of it. The expression for this total fraction, χ_{total}, is shown in Eq. (13):

$$\chi_{total} = \chi_{panel} \cdot \frac{n_{panel} - 1}{n_{panel}}. \tag{13}$$

4.1.2. Validation of the shadow dimension submodel with the Els Valentins data

The theoretical model proposed in Section 4.1.1 was used with the data in the Els Valentins study. The basic array geometrical parameters were provided by the PV farm staff, and the geometrical angles required for the model were obtained from SoL's Solar Geometry model.

Due to the unavailability of empirical shadow dimension data, the model has been validated against a shadow dimension estimation feature in PVsyst. This is presented in PVsyst as an independent feature. Results are provided for various resolutions (higher than the maximum 1-h resolution for power estimation), which makes it a viable validation source (influence of data resolution for SS analysis is addressed in Section 4.1.3). This shadow dimension feature provides a graph of the shaded percentage of the modules due to SS for a given time resolution. Both the shadow fraction estimation for the 21st Els Valentins parcel, using the PVsyst feature and the model presented here, are shown in Fig. 9. This shading fraction estimation is shown for four different days of the year (each corresponding to a different season) to appreciate the behavior of the model in every possible SS scenario. The data resolution used is of 10 min (optimal for SoL) and available in the PVsyst SS feature.

The shadow dimension model used for the SS feature of SoL and the PVsyst SS feature provide very similar results.

Table 15
Comparative analysis of the PV tools used.

			SoL	PVLib	PVsyst
Licence			Free	Open source	License required.
Data resolution			No limit (optimal 10-min).	No limit.	Maximum: 1-h resolution.
Inputs			Very reduced. No specific module, array or inverter information required.	Reduced. Module, array and inverter information required. If module or inverter is not on database, high number of inputs required.	Numerous.
Accuracy	Els Valentins	Power	Average RMSE: 3.68	Average RMSE: 4.73	Average RMSE: 4.27
		Energy	Average annual error: 3.51%	Average annual error: 7.97%	Average annual error: 3.84%
	NREL	Power	Average RMSE: 4.91	Average RMSE: 5.33	–
		Energy	Average annual error: 0.47%	Average annual error: 6.31%	–
Self-shading			Contains a simple, yet effective SS option.	No SS option.	Contains a SS option. It has proven not to be effective due to data resolution limitations.
Adjustability			Not very adjustable as it is fixed model (for now). Some optional inputs can be selected.	Highly adjustable at various levels of development.	Moderately adjustable. Numerous optional inputs and settings.
PV systems knowledge required			Very reduced. Output is provided automatically. Anyone with the required inputs can use it.	Moderate knowledge required, as it is very adjustable and does not provide the final output automatically. However, it is well documented.	Extensive knowledge regarding PV systems and the software is needed.
Popularity			Not popular. Has recently been developed.	Very popular in the academic and open-source community.	Highly used in industry.

4.1.3. Influence of data resolution in SS estimation

SS is a phenomenon that occurs during a reduced period of time. Therefore, data resolution is thought to be a key aspect when considering its effect on power estimation. A shading fraction estimation for a given day using different time resolutions is represented in Fig. 14.

The loss of accuracy regarding SS estimation when lowering the data resolution is visually noticeable in Fig. 14. Differences for 15-s versus 1-min resolution are unnoticeable. Lowering the data resolution to 10 min leads to slight loss of symmetry in the data, a small variation in the maximum shaded fraction values, and a slight data offset. However, when the resolution is lowered to 1 h, data symmetry is completely lost, and the shading fraction results are unacceptably different from reality. In the day considered as an example, the derating effect of SS in the first hours will be totally overlooked if a 1-h data resolution is chosen, and the SS effect will be overestimated in the last hours of the day.This is why the power estimation considering using the SS provided by PVsyst proved to be less accurate than the one without that feature.

4.2. Power derating factor submodel

The design of an accurate power derating factor submodel due to SS is a highly complex process. An accurate model for SS requires the following:

- A large list of input variables with information regarding electrical connections of the panels, modules and solar cells, electrical parameters of the modules (voltage–current-power curves), precise geometrical information of the panel array, etc.
- A different submodel for each array configuration, as different connections imply different behaviors toward SS.
- An additional submodel representing the effect of SS regarding diffuse and ground irradiance (note that the submodel presented in Section 4.1 only accounts for how beam irradiance is blocked). The influence of SS regarding diffuse and ground irradiance is minimal compared with its effect on beam irradiance; nevertheless, if an accurate SS model is required, it should also be considered.

All the above requirements make the addition of an accurate SS model incompatible with the simple and effective nature of the SoL model. Therefore, the power derating factor submodel presented here is a simplified and approximate model. This will be adequate for SoL for several reasons, which are as follows:

- Although a general, simplified model for SS may not be suitable for every array configuration, SoL would implement it as an optional feature. This implies the SS feature will always represent an enhancement and never a source of inaccuracies.
- Reducing the number of inputs for the SS model will not compromise SoL's simplicity significantly.
- The SS phenomenon is only relevant during the first and last low-producing hours of certain months. The overall cost of inaccuracies is low in the context of an annual time span. Therefore, the investment of resources in developing a precise model for SS would not be optimal.

4.2.1. Theoretical formulation

The submodel presented here is designed for panels formed by a number n_{row} of rows of modules (2 in Fig. 15) and a number n_{column} of columns of modules (3 in Fig. 15) connected between them.

Because the power loss is bigger when a module is completely shaded than when it is partially shaded (effect of bypass diodes), two alternative derating factor expressions are used depending on whether the bottom row of modules is completely shaded (Eq. (14)) or not (Eq. (15)). These equations are adapted from Deline et al. (2013). The proportional factor to G_{bi} in Eq. (15), in a continuous way, accounts for

the effect of bypass diodes. A continuous approach for this phenomenon allows the model to be used for any number of bypass diode configurations (thereby reducing the number of required inputs of the model):

$$\zeta_{derating} = \frac{G_{di} + G_{ri}}{G_{di} + G_{ri} + G_{bi}},$$

(14)

$$\zeta_{derating} = \frac{G_{di} + G_{ri} + \left(1 - \frac{s_{height}}{h_{module}}\right) \cdot G_{bi}}{G_{di} + G_{ri} + G_{bi}}.$$

(15)

In Eqs. (14) and (15), G_{di} is the diffuse irradiance at that moment of time, G_{ri} is the ground reflected irradiance, and G_{bi} is the beam irradiance. Eqs. (14) and (15) are valid provided $d_{unshaded}$ is sufficiently low ($d_{unshaded} < L_{module}$).

4.2.2. Validation of the SS model with Els Valentins' power data

This section provides an error analysis analogous to that in Section 3.1.1 for the power estimation for SoL once the power derating factor, $\zeta_{derating}$, has been considered for the data in Els Valentins for 2015, 2016, and 2017. The error results are compared with those for SoL without $\zeta_{derating}$. The resolution of the data is 10 min.

The annual energy percentage errors are compared in Table 11. The power error indicators are compared in Table 12.

There is an improvement regarding every error indicator considered. This validates the SS model proposed.

The measured power, SoL model estimation, and SoL model SS enhanced estimation for a winter day is represented in Fig. 16. It can be visually appreciated how, after introducing the SS enhancement, the overestimation during the first and last hours of the day disappears.

The discrete versus continuous approach to SS modelling explained in Section 4.2.1 regarding Eq. (15) can be very clearly appreciated in Fig. 16. Els Valentins' modules have three bypass diodes. During hours when SS is a noticeable phenomenon, the real power curve shows three discrete steps, while the SS SoL model is continuous. However, the error caused by this continuous approximation is extremely small considering the power production and the difference with the SoL estimation without SS. The level of simplicity obtained by using a simplified model highly outweighs its reduced inaccuracies.

4.3. SS as part of the SoL model

The proposed SS model is introduced into the SoL model structure. The new SS submodel would be introduced into the original flowchart (see Fig. 1) as proposed in Fig. 17.

Tables 13 and 14 contain the variables and equations for the SS submodel in SoL according to the structure presented in Santos-Martin and Lemon (2015).

5. Conclusions

Two locations have been studied, Els Valentins and the NREL facilities. The SoL model provides fairly good results for both. However,

the nature of the inaccuracies varied between locations as described below:

1. The inaccuracies in Els Valentins are caused by self-shading between the modules. This leads to the development of a new enhancement (and optional use) feature for SoL regarding SS losses. The results for SoL with the SS enhancement for Els Valentins are extremely positive. This validates the SS enhancement model for SoL for the data available. However it is a simple model that can be enhanced in the future. Additionally, its performance will be fully confirmed when it is validated against more data.
2. For the facilities in Denver, a seasonal bias in power estimation is observed. SoL overestimates power during winter months while it underestimates it for the summer months. Because of the module layout, there is no SS phenomenon.

SoL model's performance has been compared to that of other PV tools for the facilities and years studied sharing the same input data. They are compared in terms of accuracy, user-friendliness, performance, and so on in Table 15. It can be noted that PVLib's results could be improved by implementing a model to account for electrical (mismatch, connections, etc.) losses. Furthermore, SoL has also been compared to SAM for the NREL facilities. SoL provided better RMSE results for 2 out of the 3 facilities (3.9 average RMSE for SoL compared to average 4.7 RMSE for SAM for those two facilities). It also provided better energy estimation values for all the facilities, with an average 0.48% energy error compared to 1.87% of SAM. SoL is also simpler and can work with higher resolutions than SAM.

This validation and comparative evaluation provides a clear conclusion. SoL is an appropriate model for energy and power estimation or as part of a forecast tool for PV systems with mono or polycrystalline modules and fixed geometries. Its results are enhanced when using the SS model proposed in this paper, for panel array layouts where SS is a relevant phenomenon. It is not meant to replace other tools like PVsyst or SAM for PV power plants, as these can work with a higher amount of features, and SoL cannot, for now, meet that level of detail and complexity. However, if the user is searching for the simplest PV model or wishes to obtain a very accurate estimation for power production with a very reduced number of inputs (for example checking the hypothetical profitability of a PV system without knowing exact information about modules or inverters) SoL will probably be one of the most adequate models to use. Furthermore, it is very well suited for stochastic analysis where multiple plants or configurations want to be studied.

Appendix A. Els Valentins data table

Table A.16.

Table A.16
Data for Els Valentins PV plant

Magnitude	Type	Origin	Type/resolution or value	Units	Model of employment	Comments
Time	Input	Els Valentins	Dynamic variable. Resolution: 15 s	Time units	SoL, PVSyst, PVLib	15 s resolution changed to 10 min and 1 h resolution for SoL model and PVSyst data analysis by selecting the middle point of the time span considered.
Horizontal irradiance	Input	Els Valentins SCADA system. Pyranometer (INGENIEURBÜRO Si-420TC-T-K).	Dynamic variable. Resolution: 15 s	Wm^{-2}	SoL, PVsyst, PVLib	15 s resolution changed to 10 min and 1 h resolution for SoL model and PVsyst data analysis respectively by performing static mean to the data.
Ambient temperature	Input	Red de Estaciones Metereologicas Automaticas in Uldecona, Catalunya	Dynamic variable. Resolution: 30 min	°C	SoL, PVsyst, PVLib	30 min resolution changed to 10 min resolution by interpolating the data. Also changed to 1 h resolution by performing static mean to the data
Power	Output	Els Valentins SCADA system. Inverter reading	Dynamic variable. Resolution: 15 s	kW	–	15 s resolution changed rightarrow 10 min and 1 h resolution for SoL model and PVsyst data analysis respectively by performing static mean to the data.
Inverter rating	Input	PV farm data	Parameter. Value: 115	kW	SoL, PVSyst, PVLib	The inverter's plate AC power rating is 100 kW, but 115 kW was chosen as larger values for measured power were recorded.
Oversizing ratio	Input	PV farm data	Parameter. Value: 1	–	SoL	–
Location	Input	PV farm data	Data structure with fields: Latitude: 40.62 Longitude: 0.355 Elevation: 217	° ° m	SoL, PVSyst, PVLib	–
Tilt angle	Input	PV farm data	Parameter. Value: 31.5	°	SoL, PVSyst, PVLib	SoL provides the option to calculate it automatically. That estimated value is 32.8°
Azimuth angle	Input	PV farm data	Parameter. Value: 180	°	SoL, PVSyst, PVLib	–
Module age	Input	PV farm data	Parameter. Value for 2015: 6.83 Value for 2016: 7.83 Value for 2017: 8.83	years	SoL, PVSyst	PVsyst only allows integers as module age inputs, therefore the values have been rounded to the closest integer for PVsyst.
Losses	Input	PV farm data	Soiling: 2% Shading: 0% Mismatch: 2% Wiring: 2% Connections: 0.5% Initial Light Induced Degradation: 1.5%	–	SoL, PVSyst, PVLib*	In PVLib only soiling losses are considered.
PV Module	Input	PV farm data	Siliken SLK60P6L 224 Wp	–	PVSyst, PVLib	For PVLib it was not available in database, the closest module: Solar Fun SF220-30-P220 is selected. Known parameters of the module are included manually.
Inverter	Input	PV farm data	Siliken Se100i	–	PVsyst, PVLib	For PVLib it was not available in database, the closest inverter: CPS SC100KT-O-CA-480 is selected. Known parameters of the inverter are included manually.
Modules in series	Input	PV farm data	Parameter. Value: 20	–	PVsyst, PVLib	–
Strings of modules	Input	PV farm data	Parameter. Value: 26	–	PVsyst, PVLib	–

Appendix B. Averaged monthly data tables

Tables B.17, B.18, B.19.

Table B.17
Averaged monthly data for Els Valentins for years 2015 to 2017.

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Horizontal Irradiance (Wm^{-2})	91	132	174	254	274	306	286	252	189	135	109	86
Temperature (°C)	8.6	9.8	12	13.8	18.1	22.3	24.9	24.1	20.5	17.1	12.2	8.8
Averaged monthly power (kW)	12.9	16.4	21.5	23.1	24	24.7	22.5	20.7	20.8	19.3	18.6	13.2

Table B.18
Averaged monthly data for Denver for the year 2011.

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Horizontal Irradiance (Wm^{-2})	110	156	205	233	232	297	288	257	215	173	125	104
Temperature (°C)	3.7	3.8	7.2	8.6	11	20.3	23.7	24.5	18.2	12.4	6.5	2.5
Power for STFS (kW)	25.6	32.5	33.5	32.4	29.7	33.3	32.7	33.2	33.2	30	27.1	23.7

Table B.19
Averaged monthly data for Denver for the year 2012.

	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
Horizontal Irradiance (Wm^{-2})	107	154	222	258	271	292	271	246	213	149	114	94
Temperature (°C)	4.7	1.6	10.2	11.6	14.9	23.4	23.8	22.8	18.3	10	7.8	3.2
Power for STFS (kW)	9.2	10.7	18.8	19.5	20.5	21	19.8	18.9	17.4	12.3	10.1	7.8
Power for RSF2 (kW)	103.2	125.9	168.3	166.8	158.4	161.5	145.7	144.7	150.2	129.3	112.7	100.9
Power for Parking (kW)	124.8	162.2	203.2	211.9	193.5	199.5	180.7	196.4	168.3	154.3	125	115.9

Appendix C. Percentile error tables

Tables C.20, C.21, C.22.

Table C.20
Normalized power percentile errors (in percentage) for SoL and PVLib for years 2015, 2016 and 2017. 10-min resolution data.

		Percentile									
		0	1	5	25	50	75	90	95	99	100
2015	SoL	0	0.02	0.15	0.87	1.83	2.96	5.14	6.93	13.27	88.64
	PVLib	0	0.07	0.37	1.51	2.64	4.31	6.55	8.59	15.44	86.8
2016	SoL	0	0.01	0.15	0.88	1.83	2.97	4.76	6.26	13.3	51
	PVLib	0	0.08	0.41	0.1.65	2.92	4.74	6.64	8.43	15.13	28.2
2017	SoL	0	0.02	0.14	0.83	1.84	3.07	5.42	7.24	16.15	64.68
	PVLib	0	0.11	0.49	1.77	3.12	4.92	7.31	9.68	17.44	64.57

Table C.21
Normalized power percentile errors (in percentage) for SoL and PVsyst for years 2015, 2016 and 2017. 1 h resolution data.

						Percentile					
		0	1	5	25	50	75	90	95	99	100
2015	**SoL**	0	0.03	0.14	0.81	1.67	2.67	4.6	6.12	10.15	27.94
	PVsyst	0	0.04	0.17	0.93	2.57	4.84	7.3	9.29	19.14	33.18
2016	**SoL**	0	0.03	0.17	0.91	1.75	2.76	4.49	5.64	9.75	46.73
	PVsyst	0	0.02	0.13	0.66	1.62	3.32	4.95	6.5	10.54	43.63
2017	**SoL**	0	0.04	0.15	0.8	1.78	2.93	5.37	7.26	14.57	75.68
	PVsyst	0	0.02	0.13	0.76	1.86	3.41	5.61	7.55	13.06	77.44

Table C.22
Normalized power percentile errors (in percentage) for SoL and PVLib for the Denver PV plants.

							Percentiles					
			0	1	5	25	50	75	90	95	99	100
Visitor Parking		**SoL**	0	0.04	0.17	0.84	2.24	4.6	6.54	7.96	11.08	22.7
		PVLib	0	0.05	0.17	1.23	3.37	6.07	7.91	8.88	13.01	21.22
RSF2		**SoL**	0	0.03	0.15	0.81	1.78	3.48	5.33	7.07	11.4	46.98
		PVLib	0	0.03	0.16	0.83	2.18	4.56	6.32	7.3	13.78	50.29
SRFS	**2011**	**SoL**	0	0.04	0.24	1.47	3.21	5.28	7.48	9.14	16.8	72.1
		PVLib	0	0.04	0.14	0.83	1.94	3.98	6.29	8.46	16.66	78.4
	2012	**SoL**	0	0.04	0.28	1.41	3.24	5.42	7.68	9.35	22.57	89.12
		PVLib	0	0	0.2	1.48	3.37	5.19	7.13	8.83	17.12	96.18

Appendix D. Daily power plots
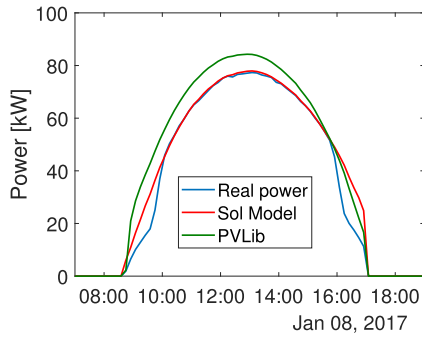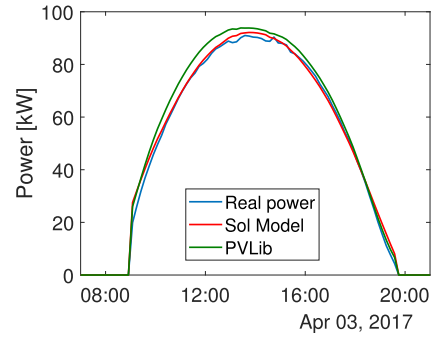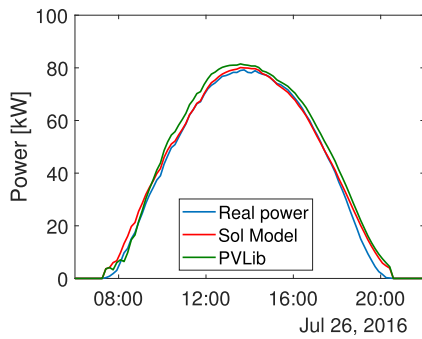
(a) Clear sky winter day

(b) Clear sky spring day

(c) Clear sky summer day
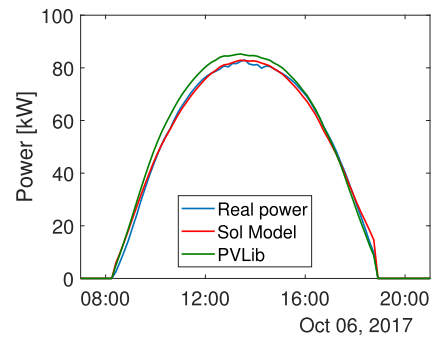
(d) Clear sky autumn day
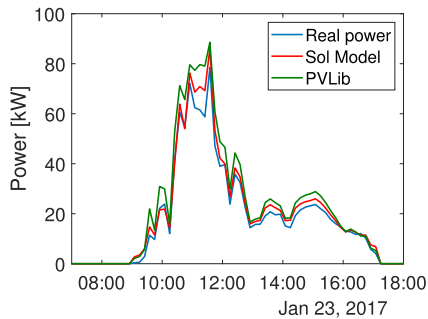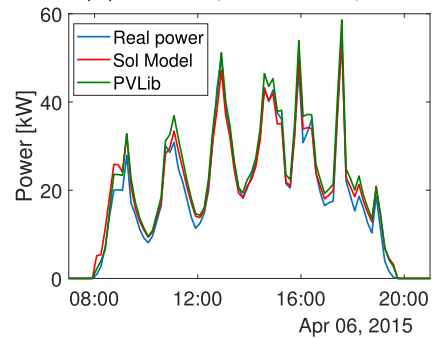
(e) Cloudy winter day
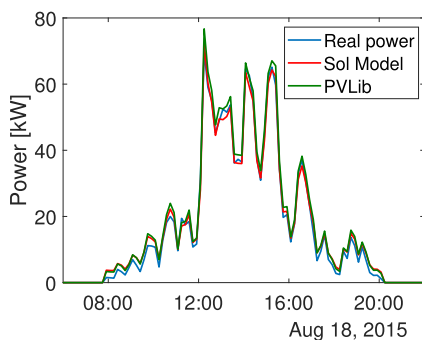
(f) Cloudy spring day
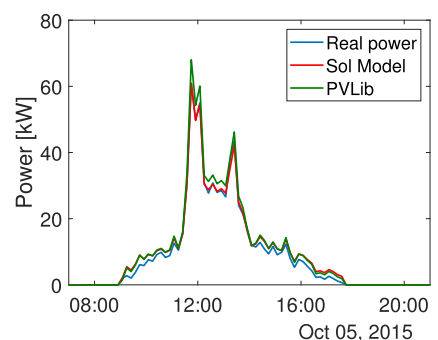
(g) Cloudy summer day

(h) Cloudy autumn day

Fig. D.18. Clear sky and cloudy day plots for the power estimated by SoL, PVLib and the real readings of the Els Valentins plant. 10 min resolution data.
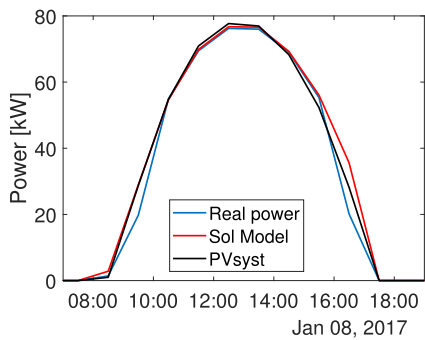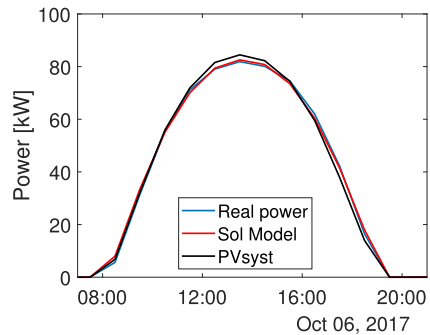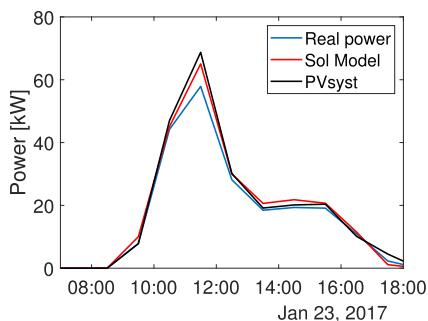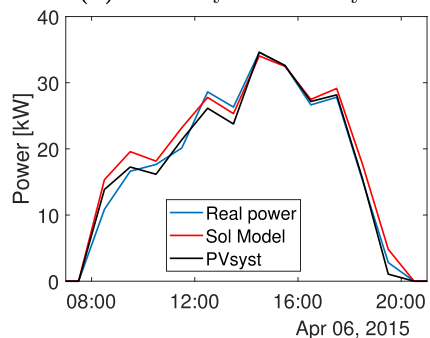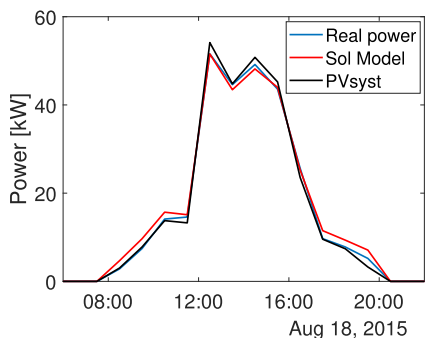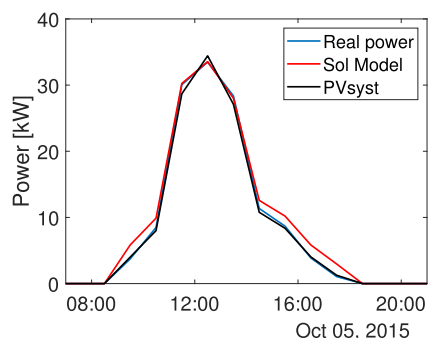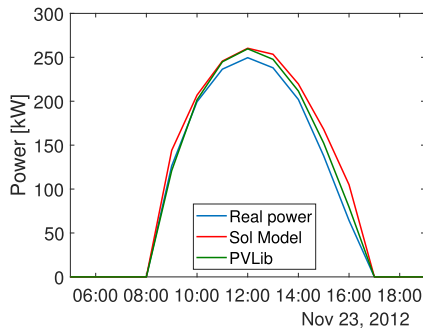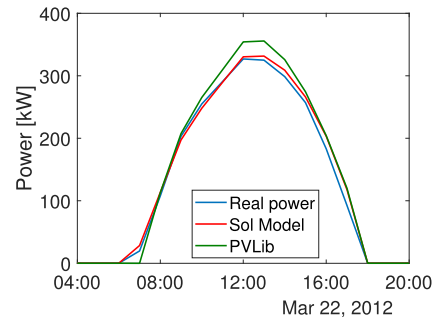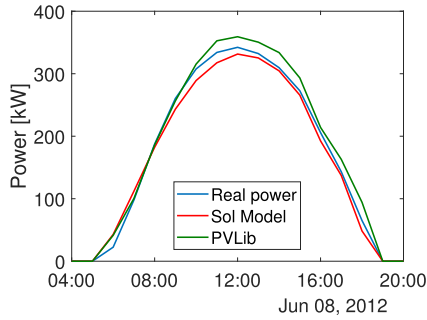
(a) Clear sky winter day
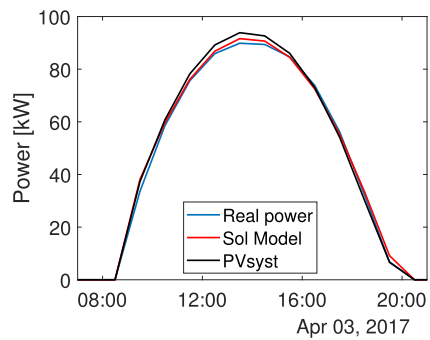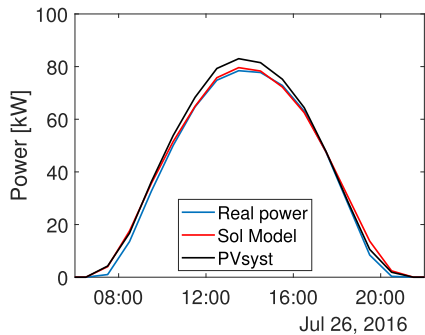
(b) Clear sky spring day

(c) Clear sky summer day

(d) Clear sky autumn day

(e) Cloudy winter day

(f) Cloudy spring day

(g) Cloudy summer day

(h) Cloudy autumn day

Fig. D.19. Clear sky and cloudy day plots for the power estimated by SoL, PVsyst and the real readings for the Els Valentins plant. 1 h resolution data.
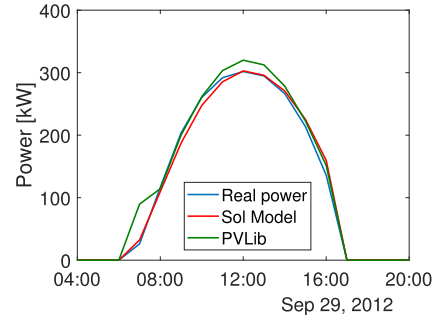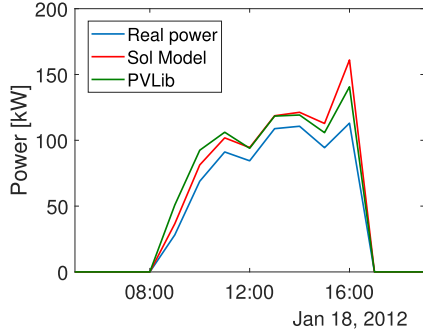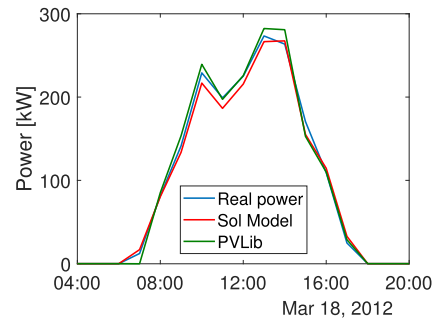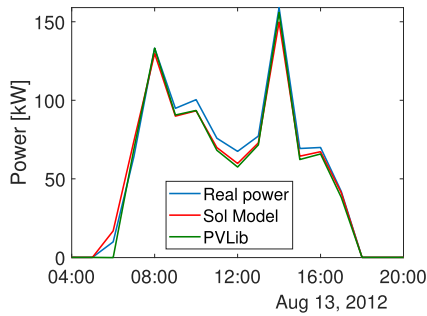
(a) Clear sky winter day

(b) Clear sky spring day

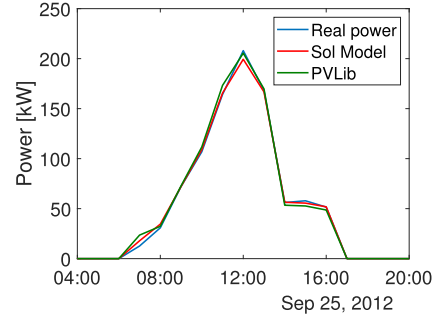**(c)** Clear sky summer day

(d) Clear sky autumn day

(e) Cloudy winter day

(f) Cloudy spring day

(g) Cloudy summer day

(h) Cloudy autumn day

Fig. D.20. Clear sky and cloudy day plots for the power estimated by SoL, PVLib and the real readings for the Denver plant. 1 h resolution data.

References

Al-Amoudi, A., Zhang, L., 2000. Application of radial basis function networks for solar-array modelling and maximum power-point prediction. IEE Proceedings: Generation, Transmission and Distribution 147 (5), 310–316, cited By 141. URL <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034272428> & doi=10.1049%2fip-gtd%3a20000605&partnerID=40&md5=0975a526dd2862066bd4d561c50ae727.

Appelbaum, J., Bany, J., 1979. Shadow effect of adjacent solar collectors in large scale systems. Sol. Energy 23 (6), 497–507.

Braun, J., Mitchell, J., 1983. Solar geometry for fixed and tracking surfaces. Sol. Energy 31 (5), 439–444.

Chen, S., Li, P., Brady, D., Lehman, B., 2013. Determining the optimum grid-connected photovoltaic inverter size. Sol. Energy 87, 96–116.

Deline, C., 07 2009. Partially shaded operation of a grid-tied pv system. pp. 001268–001273.

Deline, C., Dobos, A., Janzou, S., Meydbray, J., Donovan, M., 2013. A simplified model of uniform shading in large photovoltaic arrays. Sol. Energy 96, 274–282.

Duffie, J.A., Beckman, W.A., 1980. Solar Engineering of Thermal Processes. Wiley, New York.

Freeman, J., Whitmore, J., Kaffine, L., Blair, N., Dobos, A.P., 2013. System advisor model: Flat plate photovoltaic performance modeling validation report. Tech. rep., NREL.

Gueymard, C.A., 2009. Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications. Sol. Energy 83 (3), 432–444.

Gurupira, T., Rix, A., 01 2017. PV Simulation software comparisons: PVsyst, NREL SAM and PVlib.

Hay, J.E., Davies, J.A., 1980. Calculation of the solar radiation incident on an inclined surface. In: Proc. of First Canadian Solar Radiation Data Workshop, Ministry of Supply and Services Canada. Vol. 59. pp. 59–72.

(ITRPV, 2019. International Technology Roadmap for Photovoltaic webpage. Accessed: March 2020. URL https://itrpv.vdma.org/.

Jordan, D.C., Kurtz, S.R., 2013. Photovoltaic degradation rates - an analytical review. Prog. Photovolt.: Res. Appl. 21 (1), 12–29.

King, D.L., Gonzalez, S., Galbraith, G.M., Boyson, W.E., 2007. Performance model for grid-connected photovoltaic inverters. Tech. Rep. SAND2007-5036, Sandia National Laboratories, USA. <http://energy.sandia.gov/wp/wp-content/gallery/uploads/Performance-Model-for-Grid-Connected-Photovoltaic-Inverters1.pdf>.

National Renewable Energy Laboratory, 2019. SAM webpage. Accessed: March 2019.. https://sam.nrel.gov/.

Perez, R., Ineichen, P., Seal, R., Maxwell, E., Zalenka, A., 1992. Dynamic global-to-direct irradiance conversion models. ASHRAE Trans. 98 (1), 354–369.

PVsyst Team, 2019. PVsyst webpage. Accessed: March 2019.. https://www.pvsyst.com/.

Rodden, P., Frearson, L., Lee, G.R., 2011. An assessment of photovoltaic modelling software using real world performance data.

Sandia National Laboratories, 2019. PVLib webpage. Accessed: August 2019.. https://pvpmc.sandia.gov/applications/pv_lib-toolbox/.

Santos-Martin, D., Lemon, S., 2015. SoL – A PV generation model for grid integration analysis in distribution networks. Sol. Energy.

Smith, J.H., Reiter, L., 1984. Summary of photovoltaic system performance models.

Stein, J.S., Holmgren, W.F., Forbess, J., Hansen, C.W., 2016. Pvlib: Open source photovoltaic performance modeling functions for matlab and python. In: 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), pp. 3425–3430.

Villalva, M.G., Gazoli, J.R., Filho, E.R., May 2009. Comprehensive approach to modeling and simulation of photovoltaic arrays. IEEE Trans. Power Electron. 24 (5), 1198–1208.

Zhou, W., Yang, H., Fang, Z., 2007. A novel model for photovoltaic array performance prediction. Appl. Energy 84, 1187–1198.