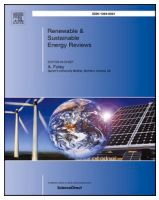
Renewable and Sustainable Energy Reviews 170 (2022) 112968

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser



Performance analysis and modelling of a 50 MW grid-connected photovoltaic plant in Spain after 12 years of operation

Enrique Fuster-Palop [a,b], Carlos Vargas-Salgado [a], Juan Carlos Ferri-Revert [c], Jorge Payá [a,*]

[a] Instituto Universitario de Investigación en Ingeniería Energética, Universitat Politècnica de València, Camí de Vera s/n, Valencia, Spain
[b] ImpactE, C/ Joan Verdeguer nº 16, Valencia, 46024, Spain
[c] Entoria Energy, Avenida Maisonnave, 41, 7 C D, 03003, Alicante, Spain

ARTICLE INFO

Keywords:
Photovoltaic
PV utility-Scale monitoring
Performance ratio
Low irradiance losses
Multiple linear regression
Random forest

ABSTRACT

This study aims to estimate the performance and losses of a 50 MW photovoltaic (PV) utility-scale after 12 years of operation. The PV plant has monocrystalline and polycrystalline silicon modules and is located in the central region of Spain with an annual insolation of 1976 kWh/m^2. Monitoring data over the entire year 2020 has been analyzed and filtered to assess the performance results following the IEC 61724 standard guidelines. The annual average reference yield, final yield, performance ratio and capacity utilization factor are of 5.44 h/d, 4.28 h/d, 79.24%, and 19.77%, respectively. Besides the experimental analysis, this work improves the estimation of the daily performance ratio, especially in days with low insolation. Two different modelling approaches have been assessed and compared. In first place, a physical model has been adopted, based on the most common losses, and including an exponential expression to account for low irradiance losses. In second place, statistical models have been used, with either multiple linear regressions or random forest algorithms. In contrast with other published models which require many inputs, the best accuracy has been reached with the random forest model using only the ambient temperature and solar irradiance as predictors, obtaining a RMSE of 1% for the PR and for the energy production.

1. Introduction

Photovoltaic (PV) energy systems are a key technology to increase the share of renewables in the energy mix, especially in countries with a high solar resource. In the last decade, the rapid cost reduction of up to 82% [1], together with the favorable decarbonization policies [2], has increased exponentially the global PV installed capacity from a total of 72 GW in 2011 to 707.5 GW in 2020 [3].

Literature on the operation of large photovoltaic plants is rather recent. Most of these plants are located in hot, desert, arid or semi-arid climates, such as 5 MW_p in Sivagangai (India) [4], 9.36 MW_p in Gujarat (India) [5], 10.13 MW_p in Soroti City (Uganda) [6], 11.15 MW_p in Shagaya PV plant (Kuwait) [7], 15 MW_p in Nouakchott (Mauritania) [8], 20.05 MW_p in southwestern Algeria [9], and 23.92 MW_p in El Bayad (Algeria) [10]. In most of the cases, these analysis were performed after only 1–5 years of operation, which provide limited insights on the long-term performance. To cover this gap, this paper investigates the PV production after 12 years of operation, for the largest PV power plant (50 MW) for which the performance is reported in literature.

The PV energy production potential estimation is essential to provide more accuracy in the design and monitoring stages of new PV utility-scales and to guarantee their integration to the power grid [9], and a proper performance and reliability throughout their life-cycle [11]. For this purpose, commercial modelling softwares are generally employed, with a reliability which depends on the accuracy of the irradiance and electrical submodels [12]. The latter includes parameters such as the power losses at different stages of the facility, namely, the performance ratio (PR).

In addition to the study of the performance of a power plant, this paper also investigates the modelling of the PR as one of the main points to estimate the AC energy yield (E_{AC}) in PV systems using irradiance time-series. These models are widely spread in technical specification manuals [13], open-source libraries *pvlib* [14], research literature [15], and commercial software [16]. Generally, the main inputs are the in-plane global irradiance (I_{POA}), the nominal capacity, along with the PR. The latter is introduced as a product of the different installation losses, which are strongly dependent on the technology, the system

* Corresponding author. Instituto de Ingeniería Energética, Universitat Politècnica de València, Camino de Vera s/n Edificio 8E semisótano frente acceso J, Valencia, 46022, Spain.
E-mail address: jorge.paya@iie.upv.es (J. Payá).

https://doi.org/10.1016/j.rser.2022.112968
Received 23 May 2022; Received in revised form 24 September 2022; Accepted 4 October 2022
Available online 17 October 2022

Nomenclature

Parameters used in equations

a, b, c	Exponential fit coefficients
A	Area of PV modules
CUF	Capacity Utilization Factor
E_{AC}	AC PV energy production
I_{POA}	Global irradiance in the plane of array
IRR	Internal rate of return
Isc	Short-circuit current
G_{STC}	Reference solar irradiance at standard conditions
IQR	Interquartile range
LCOE	Levelized cost of electricity
MAE	Mean absolute error
mtry	Number of predictors selected at each split of the regression tree
N_{mod}	Number of PV modules from a PV array
NOCT	Normal operating cell temperature
NPV	Net present value
nRMSE	Normalized root mean squared error
ntree	Number of trees of the random forest algorithm
$P_{PV,rated}$	Rated installed PV power of the system at standard test conditions
$P_{inv,rated}$	Rated installed power of the inverter
PR	Performance ratio
PR′	Performance ratio without considering low irradiance losses
$PR_{measured}$	Performance ration obtained from measurements
Pmax	Peak power
R^2	Coefficient of determination
RMSE	Root mean squared error
T_a	Ambient temperature
T_{cell}	Cell temperature
THD	Total harmonic distorsion
VIF	Variance inflation factor
Voc	Open circuit voltage
$\overline{y_i}$	Mean measured value
$\widehat{y_i}$	Predicted value
y_i	Measured value
Y_F	Final yield
Y_R	Reference yield
Y_{R1}	Reference yield from weather station 1
Y_{R2}	Reference yield from weather station 2
β_0, β_1, β_2, β_3	MLR coefficients
γ	Maximum power temperature coefficient
$\eta_{wiring,\ DC}$	DC wiring losses
η_{deg}	Degradation losses
η_{inv}	Inverter efficiency
η_{LIL}	Low irradiance losses efficiency
$\eta_{mismatch}$	Mismatch losses
$\eta_{PV,stc}$	PV efficiency under STC
η_{soil}	Soiling losses
η_{temp}	Temperature losses

Abbreviations

AC	Alternating current
ANN	Artificial neural network
ANOVA	Analysis of variance
a-Si	Amorphous silicon
DC	Direct current
IEC	International Electrotechnical Commission
LIL	Low Irradiance Losses
mc-Si	Mono-crystalline silicon solar cell
ML	Machine learning
MLR	Multiple linear regression
MPPT	Maximum power point tracker
pc-Si	Poly-crystalline silicon solar cell
O&M	Operations and maintenance
PV	Photovoltaic
RD	Royal decree
RF	Random forest
STC	Standard test conditions
SCADA	Supervisory control and data acquisition
SVM	Support vector machine
WS	Weather station

Units

A	Ampere
c€	Euro cent
d	Day
GW	Nominal Gigawatt
GWh	Gigawatt hour
h	Hour
Hz	Hertz
kV	Kilovolt
kVA	Kilovolt-ampere
kW	Nominal kilowatt
kWh	Kilowatt hour
kW_p	Kilowatt peak
m	Linear metre
m^2	Square metre
nm	Nanometre
MW	Nominal Megawatt
MWh	Megawatt hour
MW_p	Megawatt peak
M€	Millions of Euros
°C	Degree Celsius
V	Volt
W	Nominal Watt
W_p	Watt peak
μV	Microvolt
Ω	Ohm

design, and the climatic conditions [17]. There is abundant literature regarding losses in which affect on the PR [18], especially when facilities operate far from the standard test conditions (STC) [19]. However, there are only a few publications which quantify the impact of low irradiance losses (LIL) [20] on the PR by adding a correction factor. Irradiances below 200–400 W/m^2 cause a non-negligible drop in efficiency of the modules [21], leading to an overprediction in the E_{AC} results when operating in such range of irradiances [22].

Generally, the modelling of LIL is addressed with logarithmic expressions to estimate either directly the PV production or the module efficiency [23] together with several empirical models with scarce peer-reviewed comparisons with other previous model proposals [24]. Another common approach to face the non-linearity of these low irradiances is by defining two or more empirical expressions by irradiance ranges [25]. The simplest model which provides a LIL correction is found in the in-house program PR-FACT [26]. Nevertheless, the employed continuous efficiency curve is not publicly reported and cannot be implemented by other researchers. More complex low irradiance models have been developed [27], but they require detailed electrical characteristics of the solar cell, which hinder the replicability when compared to simpler models. To cover this gap, this paper develops a replicable method to estimate the LIL, and hereby increase the

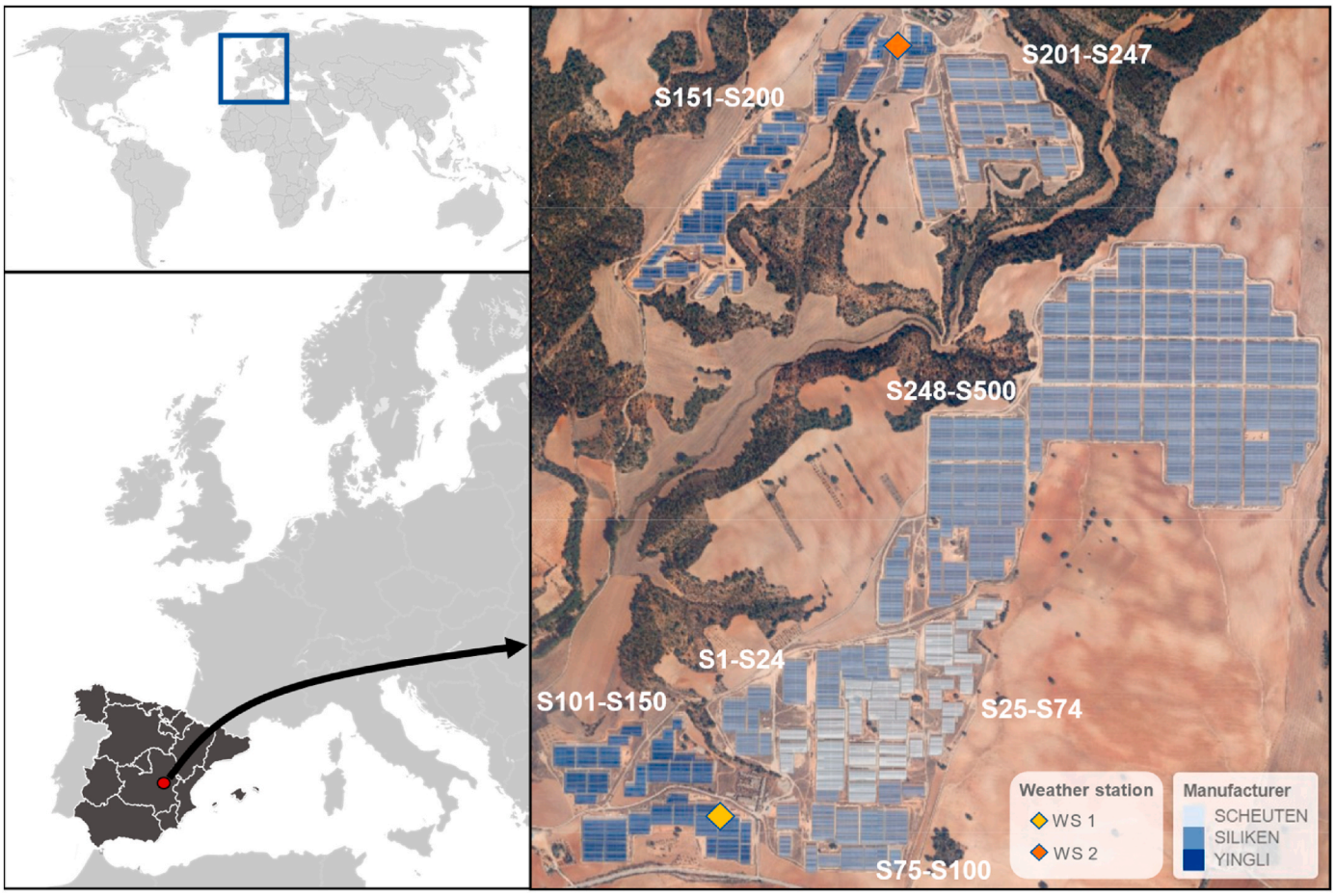


Fig. 1. Aerial view of the PV plant (500 sectors: S1–S500).

accuracy of the global PR, using the irradiance exclusively as input.

Besides physically based models, statistical and Machine Learning (ML) models have been proposed in recent years to estimate the PV production [28]. However, literature in this field is scarce. An artificial neural network (ANN) was applied to predict the PR of the PV modules with a root mean squared error (RMSE) below 0.02 [29]. The PR was calculated by means of a physical expression dependent on the temperature and irradiance. S. Bandong et al. [30] developed a Support Vector Regression (SVR) and Multiple Linear Regression (MLR) using 26 climatic variables as predictors, obtaining a RMSE of 1.5% compared with measured data. Behzad Hashemi et al. [31] reduced the number of inputs to 5, obtaining a RMSE of 0.06 with Long Short-Term Memory networks and 8 years of recorded data from a 1.44 kW_p facility. The complete replicability of these models is nevertheless limited due to the large number of climatic variables that must be measured over a long period of time.

The present work explores the capability of simpler ML models to predict the global PR with only two climatic variables: I_{POA} and the ambient temperature (T_a). Two regression models have been employed: a MLR, which is the simplest algorithm, and Random Forest (RF), which is computationally simpler than ANN and well-suited for predicting stochastic PV generation reducing bias and variance [32]. The authors have not found any published research on its application to estimate the global PR, which is a useful alternative for prediction when there is not

Fig. 2. Photograph of the PV plant.

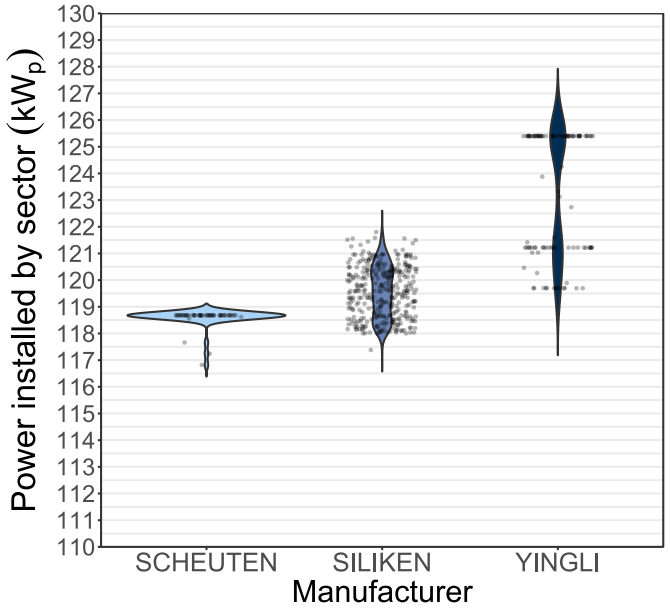


Fig. 3. Installed peak power grouped by manufacturer.

enough information available on the facility to develop a physical model.

To sum up, given the previous literature review, this article presents the following novelties:

- Provide relevant experimental data regarding the PV performance of a large PV system (50 MW) after 12 years of operation under Mediterranean climatic conditions.
- The development and assessment of a method to estimate the daily LIL, based on an empirical exponential expression and using the I_{POA} as input.
- The development and assessment of a MLR and a RF model to estimate the global PR of the utility-scale using only as inputs T_a and I_{POA}.

2. Materials

The grid-connected PV utility-scale of the present work is located in the east of Olmedilla de Alarcón, Spain (39.6155°N, 2.0905°W). The plant was commissioned in October 2008 with a nominal power of 50 MW, a peak power of 60.103 MW_p and a total land occupation of 175.3 ha. According to the Köppen climate classification, the climate of the power plant is classified as Csa (hot-summer Mediterranean), with daily average temperatures that vary between 0 °C and 31 °C, and a horizontal irradiation of up to around 1050 W/m^2, according to the measured data of this study.

The solar PV power plant (Figs. 1 and 2) consists of 500 independent sectors, each with an inverter of 100 kW and an array of different PV modules whose total peak power varies per sector from 116.5 kW_p to 127.5 kW_p. The peak power distribution is shown in Fig. 3 for the different manufacturers.

The PV modules have a fixed 30° tilt angle and are oriented towards the south. There are three different module manufacturers with mc-Si and pc-Si. The characteristics of the PV modules are summarized in Table 1. Each sector contains several models of the same manufacturer with different rated powers.

All sectors use the same inverter model INGETEAM INGECON SUN100 with a nominal power of 100 kW and an efficiency of 96%. The rest of the electrical parameters of the inverter are shown in Table 2. The energy output of the inverters is expanded to a medium voltage level of 20 kV by means of 500 transformers of 100 kVA. The voltage is finally

Table 1
Characteristics of the PV modules.

Sectors	Manufacturer and model	Technology	Area	Peak power (Pmax)	Efficiency	Peak power voltage	Peak power current	Open-circuit voltage	Short-circuit current	Temperature coefficient of Pmax	Temperature coefficient of Isc	Temperature coefficient of Voc	Maximum system voltage	Number of cells	NOCT	Distance between PV rows
(–)	(–)	(–)	(m^2)	(W_p)	(%)	(V)	(A)	(V)	(A)	(%/°C)	(%/°C)	(%/°C)	(V)	(–)	(°C)	(m)
1–24	Siliken	pc-Si	1.62	199	12.26	28.72	6.91	36.28	7.68	−0.43	0.62	−0.36	600	60	46	2.6
75–100	SLK60P6 L			203	12.50	28.86	7.03	36.36	7.76							
201–500				205	12.63	28.90	7.09	36.40	7.80							
				209	12.87	28.98	7.21	36.48	7.88							
				212	13.06	29.04	7.30	36.54	7.94							
				215	13.24	29.10	7.39	36.60	8.00							
				218	13.43	29.16	7.48	36.66	8.06							
				221	13.61	29.22	7.56	36.72	8.12							
				224	13.80	29.28	7.65	36.78	8.18							
				227	13.98	29.38	7.72	36.84	8.25							
				230	14.17	29.50	7.79	36.90	8.32							
25–74	Scheuten MULTISOL 200-P5	mc-Si	1.72	233	14.35	29.50	7.89	36.90	8.41	−0.47	0.50	−0.33	1000	96	45	2.8
				210	12.21	45.90	4.58	57.60	5.02							
				215	12.50	46.15	4.66	57.80	5.08							
				230	13.37	46.90	4.90	58.40	5.24							
101–200	Yingli YL210	pc-Si	1.70	210	12.40	26.60	7.90	33.60	8.45	−0.45	0.6p	−0,37	1000	54	46	2.68
				220	12.95	26.80	8.04	33.70	8.57							

Table 2
Summary of the characteristics of the inverters.

Parameter	Value	Units
Maximum input voltage	900	V
Maximum input current	286	A
MPPT voltage range	405–750	V
Number of inputs	4	-
Number of maximum power trackers	1	-
Nominal output power	100	kW
Nominal operating voltage	3 × 220–3x400	V
Nominal frequency range	50/60	Hz
Maximum output current	340	A
European efficiency	96	%
Power factor	1	-
THD	<3	%

Table 3
Summary of the pyranometer specifications.

Feature	Pyranometer		Temperature sensor	
	Value	Units	Value	Units
Manufacturer	Delta Ohm	-	E+E Elektronik	-
Model	LP PYRA 02	-	EE21	-
Sensitivity	10	μV/ Wm^{-2}	10	mV/ °C
Measuring range	0 ÷ 2000	W/m^2	−40 ÷ 60	°C
Operating temperature range	−0,5	°C	−40 ÷ 60	°C
Impedance	33 ÷ 45	Ω	-	-
Spectral range	283 ÷ 2800	nm	-	-
Type of sensor	-	-	Pt100 (tolerance class A, DIN EN 60751)	-
Accuracy	-	°C	0.2 ÷ 0.7	°C

increased in a substation up to 132 kW before its injection into the grid.

For the present study, the hourly E_{AC} has been measured by the monitoring system of the inverters. The inverter measurements are transferred to the supervisory control and data acquisition (SCADA) system, which is installed in a high-performance workstation, through an industrial RS232/RS485 to Ethernet converter and an IP-based network. The PV plant also has two weather stations (WS) located at its northern (39.6348°N, 2.0867°W) and southern (39.6151°N, 2.0938°W) ends (Fig. 1). The WS are equipped with a T_a sensor and a pyranometer which measures the I_{POA}. Both systems provide measurements every 5 min, and their main specifications are summarized in Table 3. The instrumentation of each WS is connected to an independent programmable logic controller system Omron CJ1M-CPU11 to condition the signals and send them to the SCADA system. The recorded data is stored and used for real-time monitoring, alarm management, signal processing, report generation, as well as the integration of the SCADA to the web. The measurement equipment is calibrated annually by the Spanish Centre for Energy, Environmental and Technological Research.

3. Methods

This section describes the methodology to analyze the measured data and to estimate the PR and E_{AC}.

The methodology is summarized in Fig. 4. The aim is to perform an analysis of the 50 MW PV power plant and to propose a novel method based on climatic data that improves the PR estimations and helps reach a more accurate estimation of the E_{AC}.

The first step in the methodology (section 3.1) is to carry out an exploratory data analysis of the collected data. The E_{AC} data of each sector and the climatic data have been initially filtered to remove potential outliers. Afterwards, the main performance parameters of the utility-scale have been calculated, and an exploratory data analysis of these results has been performed. Additionally, the results are compared with other power plants in similar climatic regions.

As a second step, two approaches have helped to model the PR with climatic data: the first method is a physical model, considering the

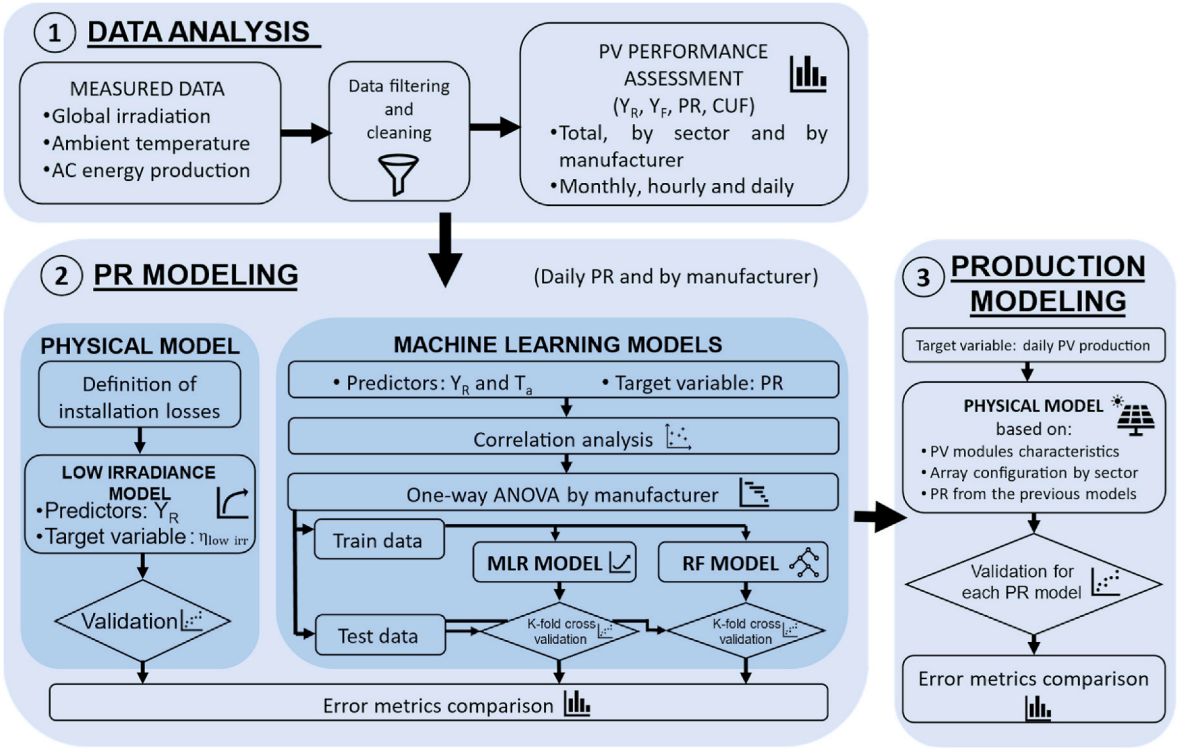


Fig. 4. Workflow of the methodology.

Table 4
Filtering criteria applied for the hourly measured data.

Min	Parameter	Max
20 W/m^2	I_{POA}	1500 W/m^2
−30 °C	T_a	50 °C
0	E_{AC}	1.02 · $P_{inv,\ rated}$

Table 5
Description of the PV performance parameters.

Parameter	Description	Expression	Units
Reference Yield (Y_R)	The maximum theoretical solar energy available in a specific location is defined as the ratio between the total daily in-plane insolation ($I_{POA,d}$) and the reference solar irradiance at standard conditions (G_{STC} = 1 kW/m^2).	$Y_R = \frac{G_{POA,d}}{G_{STC}}$	h/d
Final Yield (Y_F)	The ratio between the E_{AC} of the system during a certain period, in this case daily ($E_{AC,d}$), and the PV rated installed power of the system at standard test conditions ($P_{PV,rated}$).	$Y_F = \frac{E_{AC,d}}{P_{PV,rated}}$	h/d
Performance Ratio (PR)	The ratio between the Y_F and the Y_R. It can be understood as an efficiency parameter that measures the energy losses between actual output of the plant with its irradiation input. PR allows comparing performance results between different PV systems regardless the geographical location and the installed peak power.	$PR = 100 \cdot \frac{Y_F}{Y_R}$	%
Capacity Utilization Factor (CUF)	Relationship between the Y_F of the plant and the maximum possible energy production, defined by its installed capacity in a given period.	$CUF = 100 \cdot \frac{E_{AC,d}}{P_{PV,rated} \cdot 24}$	%

product of several factors that describe the energy losses in different stages of the facility. The LIL have been modelled to improve the estimated PR accuracy for low irradiances. Several regression and ML models have been assessed. This requires examining the correlations between the predictors (climatic data) and the predicted variable (PR) and obtaining one different model per manufacturer, justified by a one-way ANOVA test. Two different models have been studied: a MLR model and a RF model with their respective k-fold cross-validations with measurements.

The last step consists of predicting the total E_{AC} of the power plant through a physical model that considers the PR previously calculated from the different models, the module characteristics, and the array configuration of each sector. Finally, the production results have been compared with the measurements for the different PR models.

3.1. Data pre-processing

The performance data has been analyzed on an hourly basis for year 2020. The core of the experimental data is the E_{AC} from the inverter, the T_a, and the I_{POA} for the tilt and azimuth angle of the PV arrays (30° S). To provide stable measurements and a consistent analysis without measurement errors, the data was initially filtered according to the guidelines the standard IEC 61724, as in similar PV utility-scale analysis [22]. The hourly measurements were filtered according to the ranges indicated in Table 4.

In Table 4, $P_{inv,\ rated}$ (upper filter threshold for the E_{AC}) can be understood as the nominal power of the inverter (in this case, $P_{inv,\ rated}$ =

Table 6
Description of the losses included in the PR physical model.

Parameter		Description	Value	Reference
η_{soil}	Soiling losses	Optical losses due to dust and particles accumulated on the PV modules' surface over time. The latter fluctuate mainly with the frequency of the rainfalls, the maintenance schedule, and dust type. For this facility with a 30° tilt, 2% of soiling losses has been assumed.	0.98	[38–40]
η_{deg}	Degradation losses	Power decay over time in the output of the PV modules due to different causes of deterioration: cell cracks, corrosion, discoloration, glass breakage, etc [41]. Values extracted from the manufacturer data sheet.	0.916 0.890 0.920	Manufacturer Siliken Scheuten Yingli
η_{temp}	Temperature losses	Drop in the PV module efficiency due to the increase of the cell operating temperature. The calculation of the hourly values is further described in this section.	$\eta_{temp_i}(T_{cell})$	[15,37] and manufacturer
η_{LIL}	Low irradiance losses	This parameter gathers the nonlinear efficiency drop of the PV modules for low irradiance values. The calculation methodology is further explained in this section and the value depends on the I_{POA} or Y_R.	$\eta_{LIL}(Y_R)$	Manufacturer
η_{inv}	Inverter efficiency	A constant value for inverter efficiency is provided by the manufacturer mentioned in section 2.	0.96	Manufacturer
$\eta_{mismatch}$	Mismatch losses	Losses due to the interconnection of solar modules of cells with different electrical properties. Following published literature, a typical loss of 2% has been assumed.	0.98	[38,42]
$\eta_{wiring,DC}$	DC wiring losses	Direct current losses caused by the ohmnic resistance of the wiring that interconnects the PV strings with the inverter. Due to the great spatial extension of the power plant a 2% of losses has been assumed [38].	0.98	[38,43]

100 kW).

With respect to the data filtering, some authors employ a higher threshold for the minimum I_{POA} (200 W/m^2) [33]. Nevertheless, to keep as much available data as possible in this work, a less restrictive threshold of 20 W/m^2 has been considered according to the recommendation of part one of the standard IEC 61724 [34]. Keeping the irradiance values in the range 20–200 W/m^2 has enabled the development of a specific characterization of the system for low irradiance values.

The typical reporting periods to assess the performance of PV utility-scales are equal or longer than one day (e.g. annually, monthly, or daily) [35]. The hourly measurements were aggregated daily to avoid potential underestimations of module performances as reported in bibliography [36].

3.2. PV performance parameters

The performance of the present PV utility-scale has been evaluated following the IEC 61724 standard guidelines. Their definitions and expressions of the main performance parameters are described in Table 5.

3.3. Energy production model

The hourly E_{AC} ($E_{AC,h}$) of the entire utility-scale has been calculated using Eq. (1). The daily production ($E_{AC,d}$) can be obtained aggregating the hourly production as shown in Eq. (2).

$$E_{AC,h} = \sum_{i=1}^{500} \sum_{j=1}^{8} PR_i \cdot \eta_{PV,stc_{i,j}} \cdot A_i \cdot N_{mod_{i,j}} \cdot G_{POA} \quad (1)$$

$$E_{AC,d} = \sum_{h=1}^{24} E_{AC,h} \quad (2)$$

Where:

- $A_{i,j}$ is the area of a PV panel of sector i in the array j, provided by the manufacturer in Table 1.
- $\eta_{PV,stc_{i,j}}$ is the PV efficiency under STC of sector i in the array j, as provided by the manufacturer in Table 1.
- $N_{mod_{i,j}}$ is the total number of modules of the sector i in the array j.
- G_{POA} is the hourly measured in-plane global irradiance.
- PR_i is the performance ratio of the sector i and has been obtained either with a physical quantification of the plant losses (section 3.4) or with a statistical analysis (section 3.5).

3.4. PR physical model

The physical definition of the PR is based on the determination of the losses which occur in every energy transmission or conversion stage from I_{POA} to the E_{AC} of the inverters. The losses of the transformation stage have not been included since the measured data is before the grid injection. There are no shadow losses in the PV plant.

According to several authors [37], the PR of each sector i can be defined as the product of the losses indicated in Eq. (3). The term PR_i' (base PR model) refers to the PR before introducing the LIL.

$$PR_i' = \eta_{soil} \cdot \eta_{deg,i} \cdot \eta_{temp,i} \cdot \eta_{inv} \cdot \eta_{mismatch} \cdot \eta_{wiring,DC} \quad (3)$$

Where η is the efficiency of each stage, as indicated in Table 6. Whenever the efficiency data is not available, the values have been obtained from similar facilities in literature.

The temperature losses efficiency (η_{temp}) is obtained with Eq. (5) through the temperature coefficient of the PV modules (γ), defined in Table 1, and the PV cell temperature (T_{cell}). The latter can be estimated with Eq. (4) using the hourly measured T_a and I_{POA}, as well as the nominal operating cell temperature (NOCT in Table 1), which is defined as the cell temperature obtained with T_{amb} 20 °C and a solar irradiance of 1 kW/m^2. This approach is widely employed in literature and provide conservative loss values compared to other cell temperature models [44].

$$T_{cell,i} = T_a + G_{POA} \frac{NOCT_i - 20}{800} \quad (4)$$

$$\eta_{temp_i} = (1 - \gamma_i \cdot (T_{cell_i} - 25)) \quad (5)$$

Since only a single complete year with measurements is available, the absence of a cyclical component in the time series limits the use of year-on-year and statistical methods [45], which present robust results with time series of several years. As an alternative, instead of directly using the degradation losses supplied by the manufacturers (Table 6), the degradation losses were calculated with the daily PV production balance of Eq. (1), and breaking down the PR between $\eta_{deg,i}$ and another factor with the rest of the losses contemplated in Table 6. The $\eta_{deg,i}$ daily values were then averaged for the entire year, resulting in values of 0.9022 for Siliken, 0.8711 for Scheuten, and 0.8934 for Yingli. These coefficients represent the total loss due to degradation after 12 years of operation.

The efficiencies of Table 6 are typically employed to quantify the PR [18]. In the present work a new coefficient has been added, the LIL (η_{LIL}), to account for the drastic drop of the module PV production at low irradiances (below 200 W/m^2) [20]. Additionally, this coefficient includes the drop of the inverter efficiency when the power input is low, at low irradiance values. This helps to compensate the fact that a constant inverter efficiency had been assumed in the factor η_{inv}.

η_{LIL} has been calculated in Eq. (6) as the ratio between the PR obtained from the measurements ($PR_{measured}$) and PR', which does not consider the impact of the LIL.

$$\eta_{LIL} = \frac{PR_{measured}}{PR'} \quad (6)$$

Different correlations have been developed to relate the LIL with Y_R. The best fitting has been achieved with the expressions indicated in (7) and (8).

$$\eta_{LIL} = 1 - \exp(b \cdot Y_R) \quad (7)$$

$$\eta_{LIL} = 1 - a \cdot Y_R^c \cdot \exp(b \cdot Y_R) \quad (8)$$

Finally, the PR employed in the production model includes η_{LIL}, as shown in Eq. (9).

$$PR = PR' \cdot \eta_{LIL} \quad (9)$$

In order to evaluate the accuracy of the developed correlations, the daily PR obtained from the measurements has been compared with the calculated PR using the error metrics described in section 3.6.

3.5. PR statistical and Machine Learning models

A different approach to estimate the PR is by means of statistical models (e.g. MLR) and ML models (e.g. RF). Each of the developed models employs exclusively climatic data (I_{POA} and T_a) as predictors. The data corresponds to year 2020 which is representative for the behavior at half-life of the facility.

Given the variety of equipment, a different fit is proposed for each manufacturer to provide accurate predictions of the daily PR of the PV utility-scale. A one-way ANOVA test has been employed to determine which level of aggregation is more appropriate to define the statistical models. In other words, the one-way ANOVA tests helps to determine if a single global model is better for all sectors, in comparison to a different model for each of the module manufacturers. The null hypothesis is that the manufacturer groups are equal, whereas the alternative hypothesis is that at least one of the distributions is significantly different from the

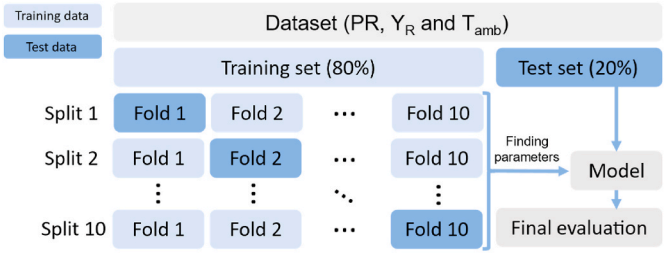


Fig. 5. Validation methodology of the MLR and RF models.

others [46].

The confidence of the results rely on the degree the one-way ANOVA assumptions are met [47]. A significance value (type-I error) of 5% has been assumed for all the hypothesis tests. The choice is based on S. Vergura [48], who indicated that medium-large PV plants present a larger uncertainty due to their high complexity.

Once the PR modelling by manufacturers was justified with the one-way ANOVA test, both MLR and RF models were trained and then tested.

As shown in Eq. (10), the MLR model assumes a linear relationship between the predicted variable (PR) and the predictors (Y_{R1} from WS1, Y_{R2} from WS2, and T_a). The *stats* R package has been applied. The regression parameters for each manufacturer i ($\beta_{0,i}$, $\beta_{1,i}$, $\beta_{2,i}$, $\beta_{3,i}$) are estimated by ordinary least squares [49].

$$PR_i = \beta_{0,i} + \beta_{1,i} \cdot Y_{R1} + \beta_{2,i} \frac{Y_{R2}}{Y_{R1}} + \beta_{3,i} \cdot T_a \tag{10}$$

The confidence of the regression parameters depends on the degree of compliance of the MLR assumptions [50], which are evaluated through their respective hypothesis test in section 4.2. The multi-collinearity is quantified employing the variance inflation factor (VIF) indicated in Eq. (11).

$$VIF_i = \frac{1}{1 - R_{i,j}^2} \tag{11}$$

Where $R_{i,j}$ is the correlation coefficient of the i predictor on the remaining explanatory variables. VIF values greater than 4 arise multi-collinearity problems [51].

Parallel to the MLR model, a RF model has been developed. The RF algorithm is a non-linear ML model [52] that potentially explains the PR with a better accuracy for the range of low irradiances. Since the PR is a continuous variable, the suggested RF model is constituted by regression trees.

The RF model was trained using the *caret* R package [53]. In the RF algorithm, several hyperparameters need to be defined by the user. The two most relevant optimization parameters are the number of predictors at each split (mtry) and the number of trees to grow for aggregation (ntree) [54].

The mtry value is calculated by default by the algorithm as the rounded down result of the square root of the total number of predictor

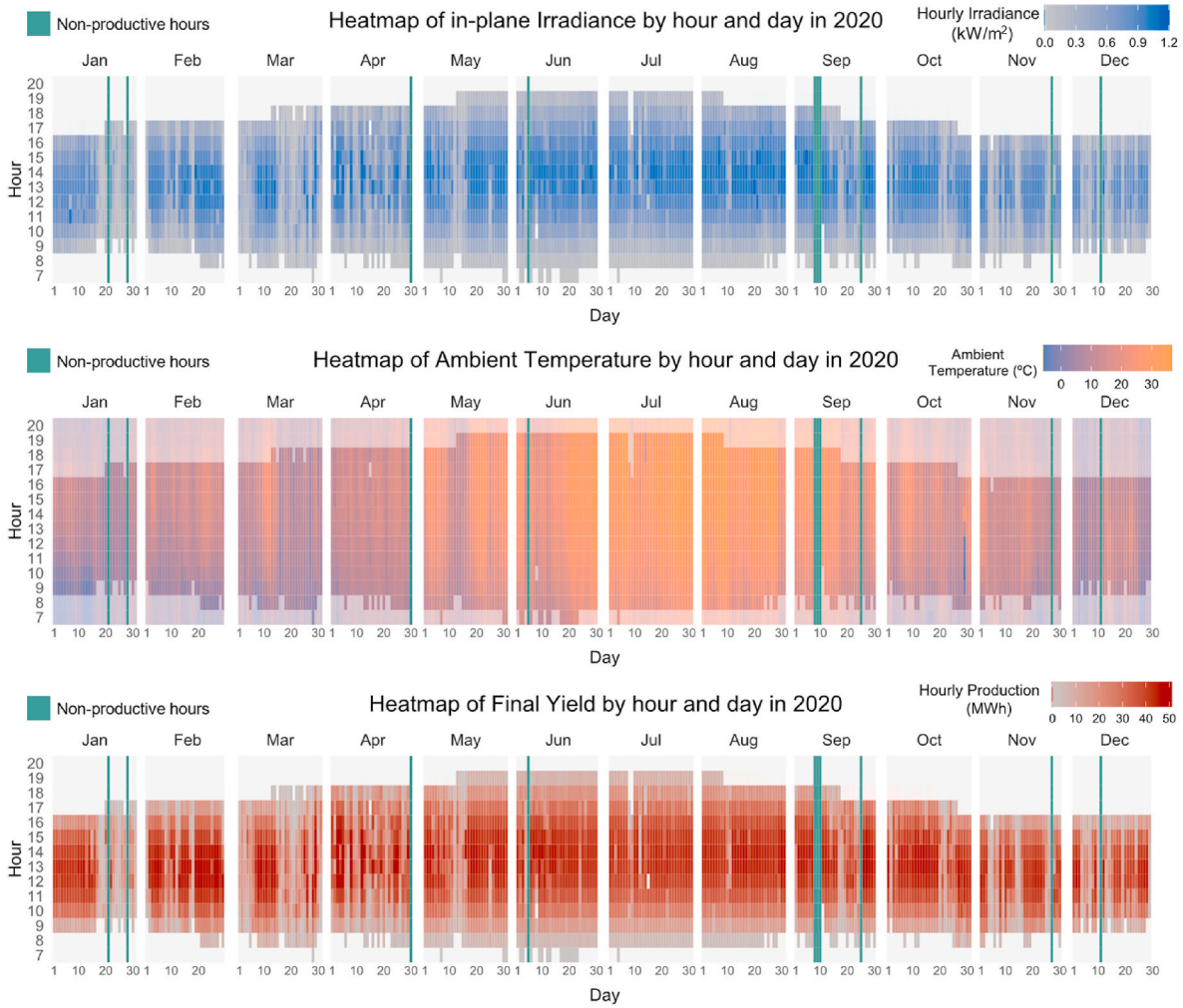


Fig. 6. Heatmaps of hourly measured I_{POA}, T_a and E_{AC} in 2020.

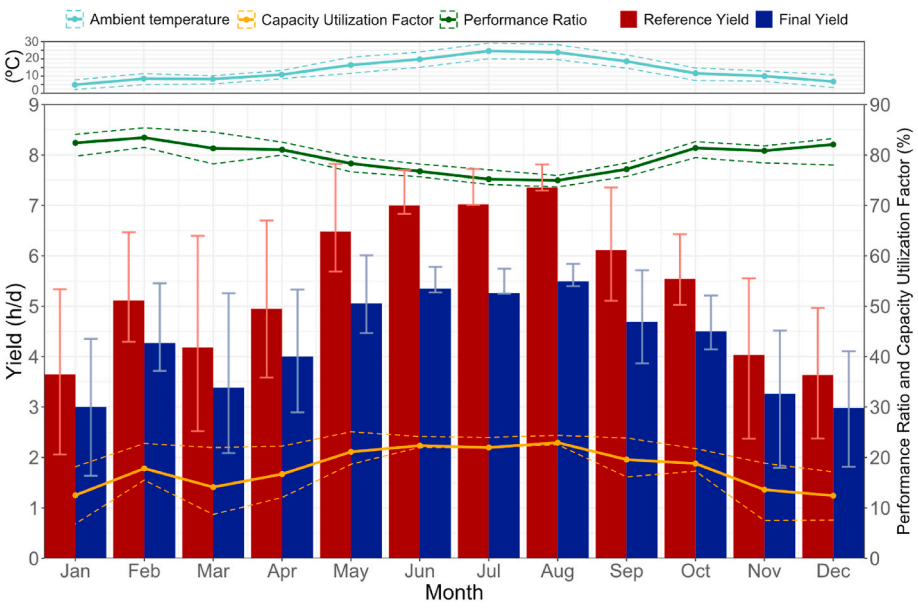


Fig. 7. Monthly average of the T_a and PV performance parameters (Y_R, Y_F, PR and CUF) in 2020. The interquartile ranges are represented by error bars and dashed lines.

variables. In this case, since there are three predictors, the mtry value is 2. The ntree value was fixed once the increase of ntree improved the RMSE in less than 1%. The previous criteria is commonly used by researchers [55] and employed in RF models applied to PV applications [56].

To assess the accuracy and robustness of the MLR and RF models, the validation was performed as depicted in Fig. 5. The pre-processed dataset obtained in section 3.1 was randomly split using an 80:20 ratio to create a train dataset and test dataset. This partition is done to perform an external validation of the models with unseen data from the train set [57]. Then, a k-fold cross validation was conducted with the train set formed by the 80% of the original dataset to obtain the optimized regressors coefficients of the MLR model and build the RF model. For this work a k value of 10 was considered, which is commonly used in literature [58]. Finally, external model validations with the remaining 20% of the original dataset were employed.

3.6. Model deviation

The accuracy of the three models for the daily PR and the derived PV production has been compared. For the MLR and RF models, the trained model performance is evaluated on the test set using as error metrics the root mean squared error (RMSE), the normalized root mean squared error (nRMSE), the mean absolute error (MAE) and the coefficient of determination (R^2) as defined in equations (12)–(15) [59]:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{12}$$

$$\mathrm{nRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\overline{y_i}} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{14}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{15}$$

Where y_i, $\hat{y}_i$, $\overline{y_i}$ are the measured, predicted and the mean measured values, respectively, and N the number of samples of the dataset.

4. Results

Following the methodology described in Fig. 4, this section presents the performance analysis of the PV utility-scale, the results and validation of the PR models, and finally in the impact on the PV production.

4.1. Performance results of the PV facility

The raw measured data consists of 8727 hourly measurements of three variables: I_{POA}, T_a and E_{AC} (in 500 sectors), during 364 days of 2020. After applying the data filtering explained in section 3.1, the resulting dataset was reduced by 59.33% of the original hours. The minimum irradiation threshold was the main effective filter since 49.74% of the original data was removed due to nighttime hours, and 8.30% during the sunrise and sunset hours. Additionally, there were 10 days (1.29% of the raw data) when the PV production stopped (non-productive hours in Fig. 6. Stops on individual days are mainly due to inverter failures caused by high temperatures, blown fuses and powered

Table 7
Statistical results of the global daily performance parameters in 2020.

Parameter	Units	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Standard deviation	Skewness[a]
Y_R	h/d	0.59	4.10	5.90	5.44	7.30	8.03	2.09	−0.68
Y_F	h/d	0.43	3.31	4.74	4.28	5.63	6.33	1.59	−0.85
PR	%	69.11%	75.93%	79.15%	79.24%	82.20%	93.44%	4.08%	0.23
CUF	%	1.81%	13.81%	19.77%	17.88%	23.52%	26.43%	6.64%	−0.85

[a] Dimensionless Value.

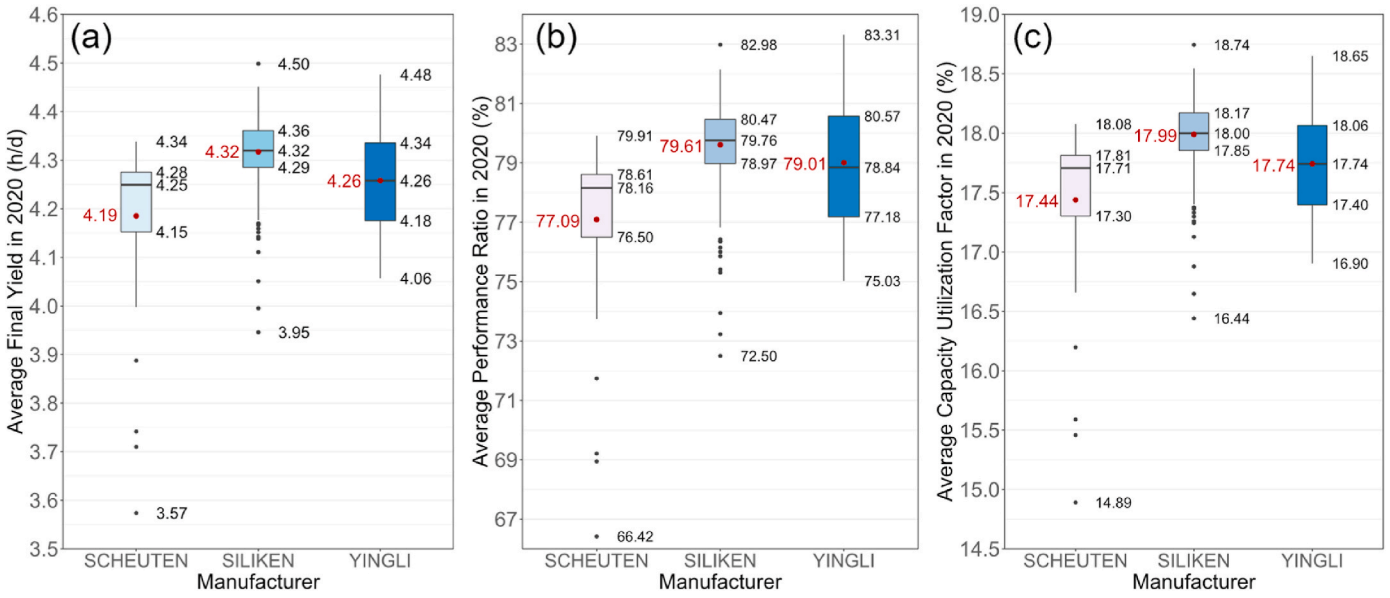


Fig. 8. Boxplots of the average Y_F (a), PR (b) and CUF (c) of each sector and manufacturer.

surge protection devices, which need staff intervention before starting-up again. Fig. 6 shows the hourly values of the total measured E_{AC} of the utility-scale, I_{POA}, and T_a after the data filtering.

The annual I_{POA} was 1975.52 kWh/m^2 and the E_{AC} of the complete utility-scale in 2020 was 91.32 GWh, with a monthly average of 7.61 GWh.

At the top of Fig. 7 the average T_a for each month is shown. The average daily T_a during the year is 16.61 °C, fluctuating from a minimum average of 7.43 °C in January, and a maximum average of 28.28 °C in July.

The monthly variation of the PV performance parameters is also presented in Fig. 7 and Table 7. According to section 3.2, the Y_R reached its maximum in August and its minimum in December with a value of 7.35 h and of 3.63 h of irradiance equivalent to 1 kW/m^2, respectively.

Regarding the Y_F, the yearly average was 5.44 h/d. The minimum was 2.98 h/d in December and the maximum 5.49 h/d in August. However, the maximum daily values, up to 6.33 h/d (in March), were registered during the Spring season, where the lower T_a compared to summer provided a higher efficiency. There are also noticeable differences in the yearly average Y_F for the different module manufacturers. The best performance is obtained by Siliken with 4.32 h/d, followed by Yingli with 4.26 and, lastly, Scheuten with 4.19 h/d. The latter also provides the lowest performances within some sectors, reaching lower values than 4 h/d, as shown in Fig. 8.

The PR ranges between in 74.97% in August and 83.46% in February, with an annual average of 79.24%. The lowest PRr are

Table 8
Comparison of the PV performance in different PV facilities with hot-summer Mediterranean climate (Csa).

Location	Commisioning Year	Monitoring period	Module type	Peak power (kW_p)	Y_F (h/d) min/mean/max	PR (%) min/mean/max	CUF (%) min/mean/max	Ref.
Olmedilla de Alarcón, Spain	2008	2020	mc-Si/pc-Si	60,103.4	0.43/4.28/6.33	69.11/79.24/93.44	1.81/17.83/26.43	Present study
Ar Ramtha, Jordan	-	2017–2018	pc-Si	5000	2.80/4.60/5.80	74.00/80.00/90.00	11.00/18.10/23.00	[62]
Albacete, Spain	2010	2010–2013	pc-Si	4600	2.20/-/4.37	59.46/-/78.84	9.16/-/18.22	[63]
Mugla, Turkey	-	2008	pc-Si	2730	2.53/4.77/6.65	-/72/-	-	[64]
Albacete, Spain	2008	2012–2016	pc-Si	2700	4.71/-/7.92	78.93/-/84.86	19.64/-/33.02	[63]
Şahinkaya, Turkey	2016	2017	pc-Si	2130.7	2.32/4.53/6.28	73.92/81.15/91.78	9.65/18.86/26.16	[37]
Albacete, Spain	2007	2012–2016	pc-Si	1400	7.92/-/2.2	83.56/-/85.60	20.14/-/29.87	[63]
Albacete, Spain	2007	2010–2013	mc-Si	1300	3.10/-/5.69	80.36/-/85.66	12.91/-/18.66	[63]
Albacete, Spain	2008	2010–2013	mc-Si	1300	3.50/-/6.69	64.91/-/68.83	14.65/-/27.29	[63]
Monteroni, Italy	2011	2012–2015	mc-Si	960	1.70/3.80/6.20	75.00/84.4/94.00	6.90/15.60/25.60	[11]
Ciudad Real, Spain	2013	2013–2016	pc-Si	370	4.29/-/4.63	80.39/-/81.39	17.86/-/19.30	[63]
Sitia, Crete, Greece	2002	2007	pc-Si	171.36	1.96/3.66/5.07	58.00/67.36/73.00	-/15.26/-	[15]
Manisa, Turkey	2018	1 year	mc-Si	30	1.53/4.16/6.09	81.22/83.61/86.15	6.38/17.35/25.39	[65]
Bouzareaha, Algeria	2004	2016–2018	pc-Si	9.5	-/3.37/-	-/70.00/-	-	[66]
Tangiers, Morocco	-	2015	pc-Si	5	1.96/4.45/6.42	58.00/79.00/98.00	6.55/14.84/21.42	[67]
Chania, Crete, Greece	-	2010–2012	a-Si/mc-Si	2.18	1.83/-/6.55	80.40/-/95.40	-	[68]
Tangiers, Morocco	-	2016	pc-Si	2	3.38/4.72/5.90	71.23/77.24/84.00	10.83/11.76/12.78	[69]
Los Angeles, United States	-	-	-	0	-/4.22/-	-/72.10/-	-	[70]
Casablanca, Morocco	-	-	-	0	-/4.29/-	-/71.90/-	-	[70]

Table 9
Parameters of the economic analysis.

Variable	Value	Units	Reference
Total investment cost	384	M€	Present study
Averaged yearly E_{AC} measured for the life cycle	91,967	MWh	Present study
Fixed electricity price (2008–2014)	22.976	c€/kWh	[71]
Averaged electricity market price (2014-Present)	6.186	c€/kWh	[75]
Specific remuneration for the operation (2014-Present)	31.754	c€/kWh	[76]
Specific remuneration for return on the investment (2014-Present)	244,850	€/MW·year	[76]
Average degradation rate	Siliken: 0.816 Scheuten: 1.074 Yingli: 0.888	%/year	Present study
O&M cost	11.6	€/kW_p·year	[77]
Annual Spanish inflation rate (averaged between 2008 and 2020)	1.062	%	[78]
Annual discount rate	7.090	%	[79]
Life cycle of the facility	25	years	Present study

The NPV, IRR and payback period are 93.02 M€, 9.19%, and 17.61 years, and the LCOE is 0.359 €/kWh.

obtained in summer due to the higher temperatures. There is a clear correlation with the temperature. The global PR was above 75% for 84.2% of the days with measurements, proving that the system has been working correctly in global terms. There are significant fluctuations in the PR when comparing the different manufacturers: the Siliken and Yingli (pc-Si) sectors provide an annual average of 79.61% and 79.01%, respectively, while the Scheuten (mc-Si) sectors yield 77.09%.

The PR of the mc-Si sectors is on average around 2% lower than the pc-Si sectors. Considering the similarities among manufacturer characteristics in the STC efficiencies, the performance difference is mainly caused by a greater drop in efficiency when the temperature increases. This issue can be observed in Table 1, since the temperature coefficient in the mc-Si modules is higher than for the pc-Si modules. The same phenomenon was also found in similar climate conditions, both in northern Algeria [60] and in Morocco [61].

The monthly average CUF ranges between 12.44% in December and 22.93% in August, with an annual average of 17.88%. The CUF dispersion decreases during the months with more sunny hours and stable weather. The annual average CUF among module manufacturers differs slightly: 17.99% for Siliken, 17.44% for Scheuten and 17.74% for Yingli, following a similar distribution scheme as the PR.

According to the assumptions described in section 3.4, the estimated degradation losses are generally 1–2% higher than the rates provided by the manufacturer datasheets. After 12 years of operation, the average degradation loss for Siliken modules is 9.79%, 12.89% for Scheuten, and 10.66% for Yingli, and their respective averaged yearly degradation rates are 0.816%/year, 1.074%/year, and 0.888%/year. Consequently, the averaged modules efficiency at STC drops to 12.13% for Siliken, 10.85% for Scheuten, and 11.02% for Yingli. The highest degradation is suffered by the mc-Si technology.

As indicated in Table 8, the performance of the present utility-scale is comparable with other PV power plants reported in scientific literature under Csa Mediterranean climate. However, due to the long operating period and consequent degradation losses, the average Y_F, PR, and CUF are slightly lower than in the other plants, where the performance was measured a few years after their commissioning. Among the registered PV utility-scales the CUF is generally greater for bigger installed powers, and mc-Si technologies on average provide better PR results.

Besides the energy performance assessment, an economic analysis has also been conducted. The remuneration of the facility has depended on two different Spanish legislative frameworks during its operation. The first period, from its commissioning in 2008 until July 2014, followed RD 661/2007 [71], with a fixed price. The second period, which is still in force, follows RD 413/2014 [72]. The remuneration calculations for this period are described in detail for other plants in literature [73].

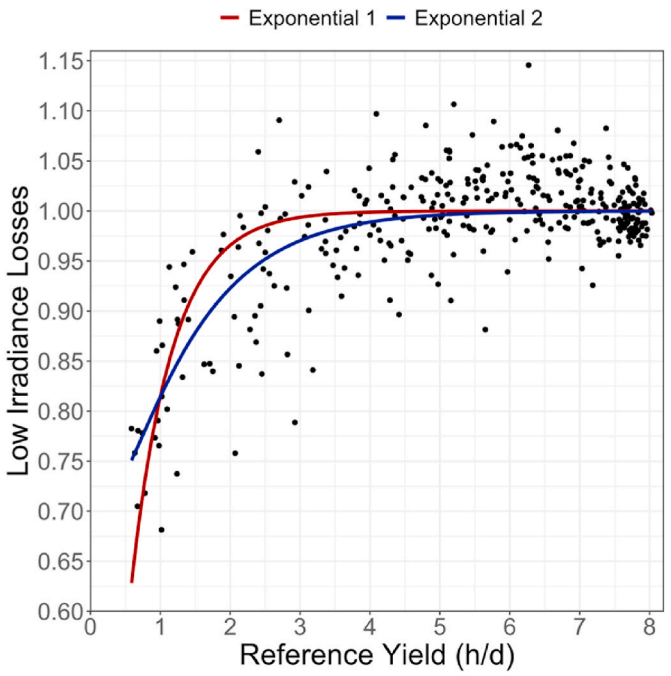


Fig. 9. Relationship between LIL (η_{LIL}) and the Y_R.

The net present value (NPV), internal rate of return (IRR), payback period and the levelized cost of electricity (LCOE), were estimated according to N. Bansal et al. [74], considering the initial investment cost, the O&M costs, the cashflows generated by the energy selling, the annual degradation rate of the modules, the inflation rate and the discount rate summarized in Table 9.

4.2. PR modelling results

Fig. 9 shows the deviations between the modelled daily PR considering the physical losses and the measured daily PR, as calculated with Eq. (6). While for regular days with daily Y_F higher than 3 h/d the deviations fluctuate around 1, which means that no significant corrections are required, there is a clear drop for daily Y_R values below 2 h/d.

These deviations were modelled with two nonlinear exponential fits, whose coefficients and error metrics are given in Table 10. The employment of an exponential fit allows reducing selectively these differences only for low Y_R values. A linear regression fit would tend to overestimate the LIL. The exponential 1 fit, despite its simplicity, tends to excessively reduce the PRs with low Y_R values s. The exponential fit 2 presents a more moderate fit and reduces the error compared with the exponential 1 for Y_R values around 2 h/d. The exponential 2 was consequently selected for the comparison with the PR models.

Adding the correction of the exponential fit 2 in Eq. (9) clearly improves the results, as may be inferred by comparing Fig. 10a and b. There are significant overpredictions with the base PR model (up to 15% of relative error) which are mitigated when η_{LIL} is introduced. With the exponential fit 2, the nRMSE decreases by 48.22%.

The compliance with the one-way ANOVA assumptions has been verified prior to its application. To meet the normality of the annual average PR, the inferior outliers below the limit defined by Tukey (Q1-1.5·IQR = 0.757) were filtered applying the same method as in other PR analyses [46]. As a result, 22 facilities were omitted and the remaining

Table 10
LIL exponential model coefficients and error metrics.

Model	Expression	a	b	c	RMSE	MAE	R^2
Exponential 1	$\eta_{LIL} = 1 - \exp(b \cdot Y_R)$	-	−1.688	-	0.047	0.033	0.514
Exponential 2	$\eta_{LIL} = 1 - a \cdot Y_R^c \cdot \exp(b \cdot Y_R)$	0.539	−1.067	0.273	0.043	0.032	0.585

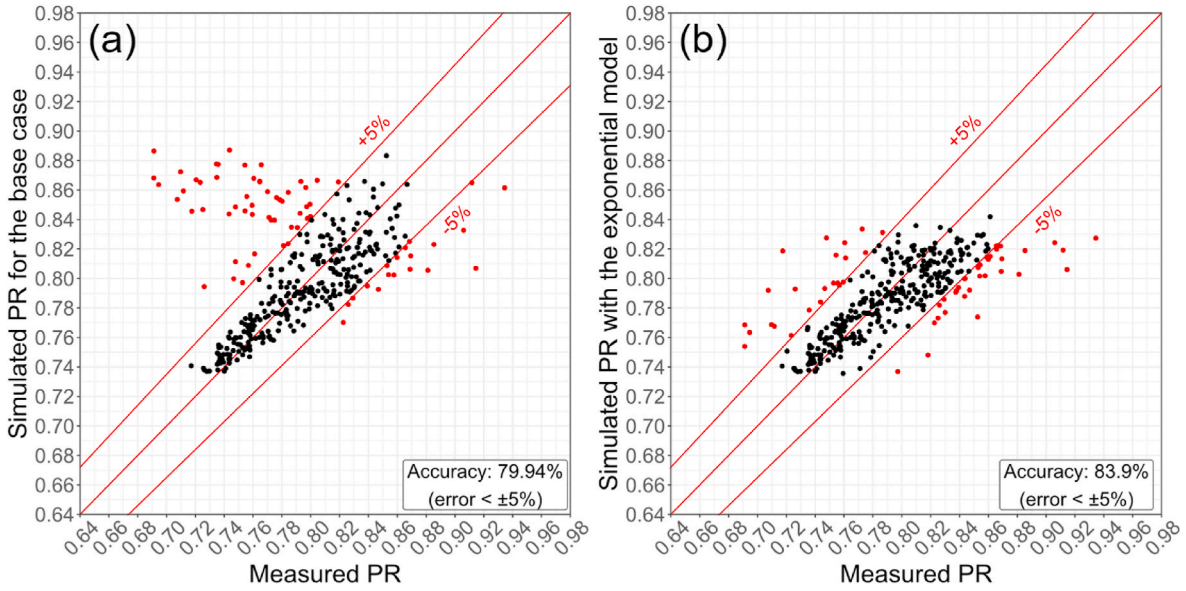


Fig. 10. PR results validated with the base (a) and exponential (b) models.

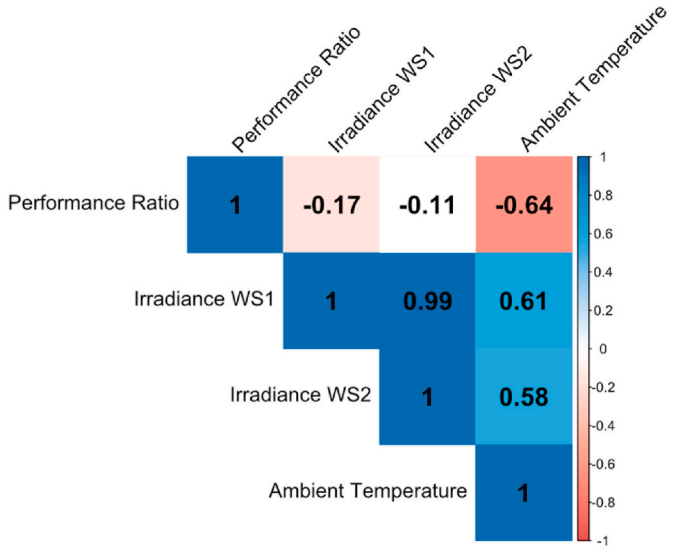


Fig. 11. Correlation matrix of the PR and the climatic variables as predictors.

PR followed a normal distribution with a mean of 0.795 and a standard deviation 0.014, as verified with the Kolmogorov-Smirnov test with a p-value of 0.239.

The normality of each PR distribution was tested with the Kolmogorov-Smirnov test for the Siliken, Scheuten and Yingli modules, providing p-values of 0.724, 0.516 and 0.0806, respectively. All the p-values are consequently higher than the type-I error threshold (0.05). A p-value of 0.239 was obtained. However, the homoscedasticity among the three PR samples was not met applying the Barlett's test. To reduce the heterogeneity of variances, the Welch's correction factor [80] was included in the one-way ANOVA test [81], which provided a p-value of 2.7e-14. Thus, the null hypothesis was rejected, which led to develop three independent PR statistical models.

For the MLR and RF model, the predictors were selected considering the global Pearson correlation coefficients between the climatic data and the global PR of the plant (Fig. 11) and the multicollinearity among predictors measured with the VIF. There is a low negative correlation with the measured irradiance from the two WS and a moderate correlation with T_a. This reveals that the higher PR is generally reached in cold days. The VIF values between T_a and the irradiances are below 1.6, presenting reduced multicollinearity. However, there is high multicollinearity between Y_{R1} and Y_{R2} with a VIF value of 50.25. The selected predictors are T_a, Y_{R1}, and the ratio Y_{R2}/Y_{R1} to consider weather fluctuations and have both irradiance predictors uncorrelated. This ratio presents a correlation coefficient with PR and T_a of 0.51 and −0.11,

Table 11
Coefficients and p-values of the MLR models to estimate the PR of the three manufacturers.

Manufaturer	MLR coefficients				p-values			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
Siliken	0.403	0.003	0.435	−0.004	<2e-16	1e-04	<2e-16	<2e-16
Scheuten	0.687	0.008	0.111	−0.005	<2e-16	2–14	<2e-16	1e-04
Yingli	0.933	0.004	−0.114	−0.004	<2e-16	3e-07	<2e-16	8e-06

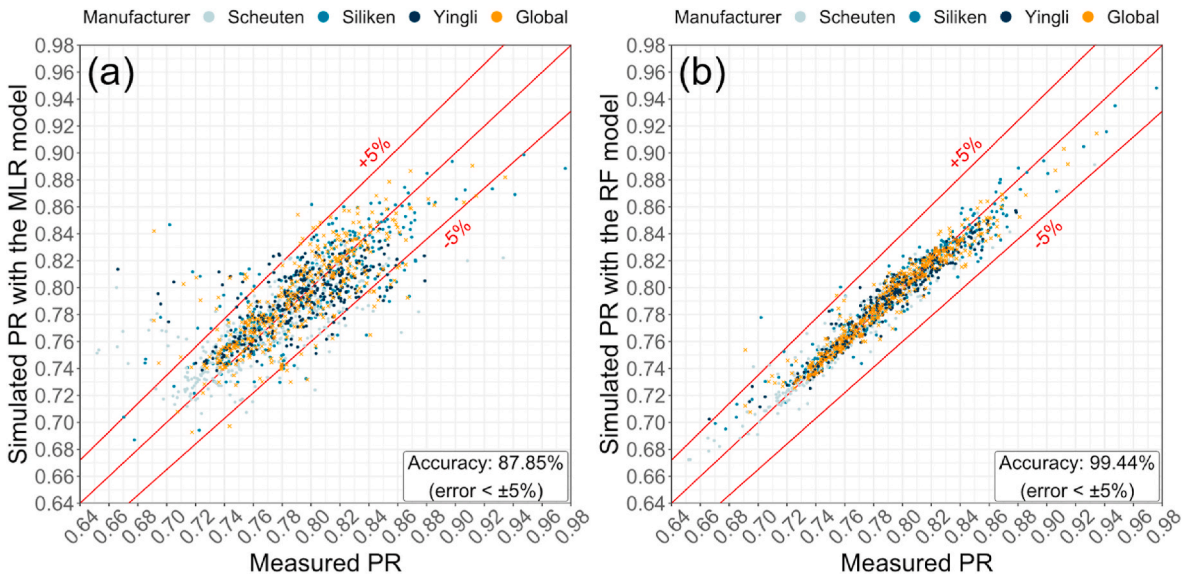


Fig. 12. Validation of the PR results obtained with the MLR (a) and RF (b) models compared with the measured PR.

Table 12
Error metrics of the PR for the MLR and RF models.

Manufacturer	MLR model				RF model			
	RMSE	nRMSE	MAE	R^2	RMSE	nRMSE	MAE	R^2
Siliken	0.026	0.033	0.019	0.689	0.024	0.030	0.017	0.741
Scheuten	0.028	0.037	0.020	0.547	0.023	0.029	0.016	0.701
Yingli	0.025	0.031	0.018	0.497	0.019	0.024	0.014	0.683

respectively.

The MLR expression results from the linear combination of the variables shown in Eq. (10), and the fitted coefficients gathered in Table 11. The intercept (β_0) value is the most influential coefficient for every manufacturer, followed by the fluctuations of the measured irradiances between both WSs. The negative coefficients of β_3 explain the reduction of the PR for increasing T_a values. The regression fit provides p-values below 0.05 for all the manufacturers (see Table 11). This supports the null hypothesis that the independent variables do not affect significantly the dependent variable.

Another conclusion is that including nonlinear combinations of the predictors have any power in explaining the PR, as verified with the Ramsey's RESET test (p-value of 0.5671). The residuals of each regression fit follow a normal distribution according to the Kolmorov-Smirnov test with an averaged p-value of 0.0998. The residuals are uncorrelated, given the averaged Durbin-Watson D statistic of 1.839 (p-value of 0.198) [82].

Fig. 12a compares the measured and the predicted PR of the three MLR models. There is a higher overprediction for lower PR values, as happened with the base PR model and this is not explained with linear relationships. The nRMSE represents for the three regressions around 3%. Fig. 12a, shows the global PR obtained by weighting the estimations of each regression with the number of sectors associated for each manufacturer, providing an accuracy of 87.85%. The accuracy represents the number of estimations with a relative error lower than 5%. The global nRMSE is 0.0324, which is close to the value obtained for the Siliken sectors which are the most frequent sectors. The global nRMSE improves by 41.54% and 15.29% compared with the base model and the exponential model, respectively. However, there is a trend to overestimate the lower PR values due to the linearity of the model similar to the base model.

For the RF regression models, the hyperparameters were first tuned to provide the lower RMSE. The Siliken sectors do not require more than 50 trees, and the other manufacturers require up to 100 trees to provide stability in RMSE. The nRMSE was below 3%. This is a major improvement compared to the MLR, as shown in Fig. 12b, where the number of outliers has been reduced, especially for low PR values. The greater deviations are found in the extreme PR values. Nevertheless, the global accuracy rises up to 99.44%. Weighting all the sectors the global PR yields a nRMSE of 0.013, shown in Table 13. This represents a reduction of 77.04% with respect to the base model and is similar or lower than the SVR model found in literature [30].

Table 13
General error metrics of the PR and E_{AC} models.

Model	PR				E_{AC}					
	RMSE (−)	nRMSE (−)	MAE (−)	R^2 (−)	RMSE (MWh)	nRMSE (−)	MAE (MWh)	R^2 (−)	Annual error (%)	Annual error ($Y_R < 3$ h/d) (%)
Base case	0.044	0.055	0.028	0.144	8.268	0.032	6.334	0.993	1.814	7.280
Exponential 1	0.034	0.043	0.023	0.432	7.973	0.031	5.968	0.993	−1.141	−1.006
Exponential 2	0.029	0.037	0.021	0.517	7.924	0.031	5.870	0.994	−0.706	−0.474
MLR	0.026	0.032	0.018	0.617	6.301	0.024	4.978	0.996	−0.116	−0.787
RF	0.010	0.013	0.007	0.945	2.626	0.010	1.973	0.999	0.099	2e-04

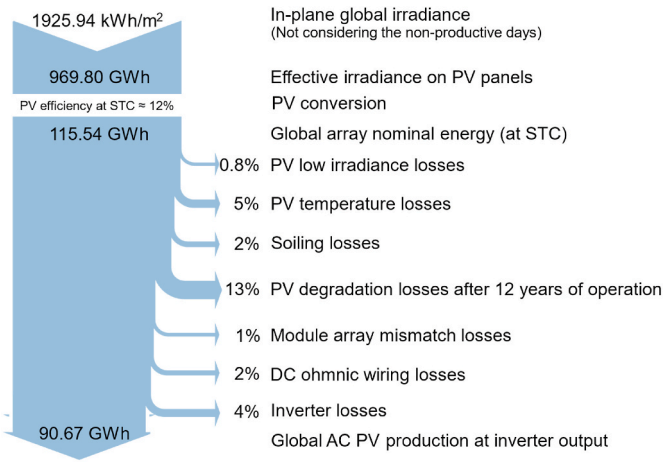


Fig. 13. Sankey diagram of the annual losses in the PV utility-scale according to the PR physical model.

The manufacturer's error metrics of both MLR and RF models are shown in Table 12.

4.3. PV production modelling results

The global E_{AC} of the PV power plant has been obtained for the base case, with the theoretical PR values, and compared with the production obtained with the three PR models (the exponential fit, the MLR and the RF model).

All the energy losses estimated in the energy balance are quantified in Fig. 13 by means of a Sankey diagram. The annual in-plane global irradiance does not consider the non-productive days since they have been filtered. LIL represent 0.78% of the annual array nominal energy at STC and the degradation losses have the biggest weight due to the long operating time of the utility-scale. The estimated annual E_{AC} with the physical model differs by −0.71% compared with the measurements. The annual production would rise 2.61% (up to 93.03 GWh) if the 10 non-productive days were considered.

The validation results for each model are shown in Fig. 14. The base model systematically overpredicts the production when the daily

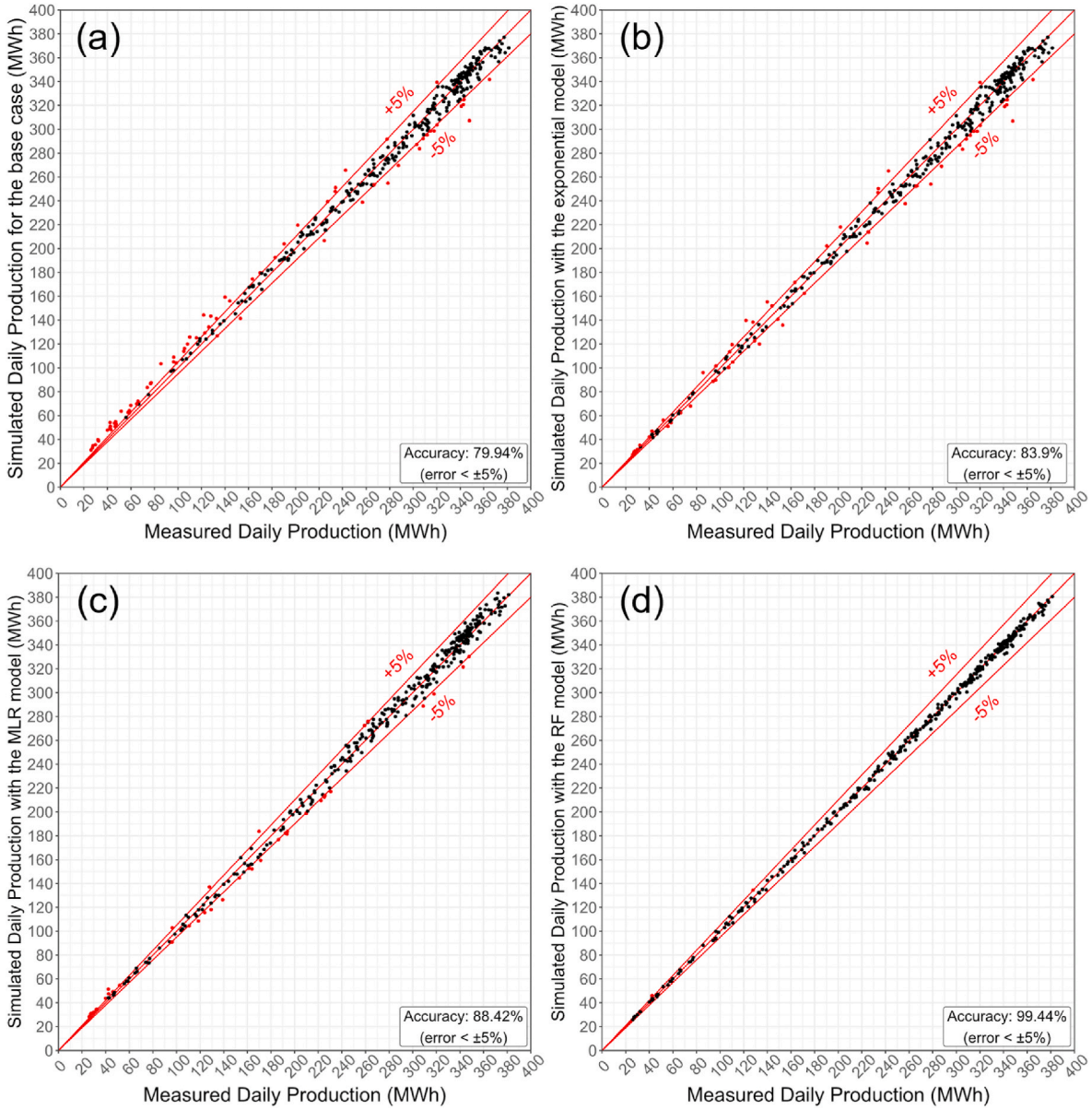


Fig. 14. Validation of the global PV production obtained considering the PR for the base (a), exponential (b), MLR (c) and RF (d) models.

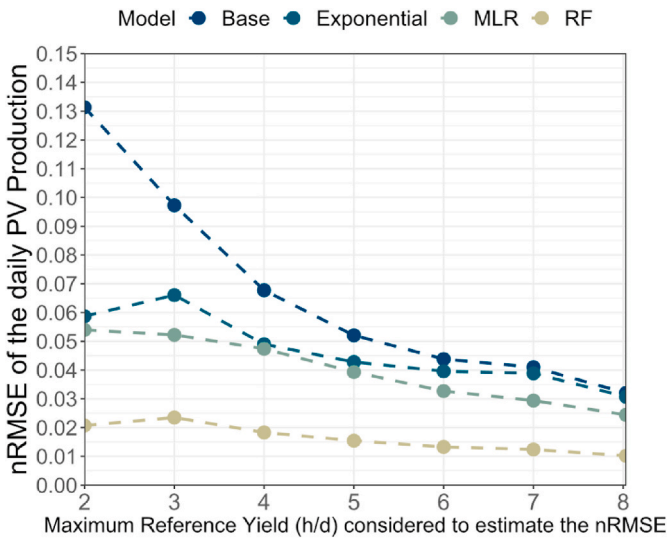


Fig. 15. nRMSE of the estimated daily production of the PV plant for days with irradiances equal or lower than the irradiance shown on the x-axis.

irradiance is low; however, the three PR models significantly reduce the number of outliers for low daily irradiances. Compared with the base case, the total RMSE is reduced up to 68.24% with the RF, while the exponential and the MLR models provide moderate improvements in RMSE of 4.16% and 23.79%, respectively. The relative error of the estimated annual production was reduced up to 0.09% with the RF model. Filtering only the days of the year in which the Y_R is below 3 h/d, an improvement in the annual production error of close to 7% is observed for the three previous models, and the RF provides the best results. Table 13 provides the general error metrics in the estimation of the production.

Fig. 15 represents the performance of each model, by means of the nRMSE of the daily E_{AC}. The nRMSE is clearly bigger in low irradiance days. Daily irradiation measurements lower than 2 kWh present a nRMSE of 13.14%, which is significantly higher than the nRMSE of 3.21% which is obtained for the full irradiance range. By incorporating the PR the LIL factor, the nRSME is reduced in the full range up to 5.87%. This value drops to 2.44% with the MLR model and 1.01% with the RF model. These three models reduce the nRMSE for low irradiances by more than half compared with the base model. Nevertheless, the RF model provides the lowest fluctuations and confident production predictions for the complete range of daily irradiances.

5. Conclusions

The present work involves the analysis of a 50 MW PV utility-scale plant in Olmedilla de Alarcón (Spain) after 12 years of operation under Mediterranean climatic conditions. The experimental campaign consists of a monitoring period of one year with measurements of climatic data and E_{AC} from the inverters. Using this data, the main PV performance parameters have been obtained.

The annual average and the minimum and maximum monthly average registered for the Y_R, Y_F, PR and CUF respectively are: 5.44 h/d, 4.28 h/d, 79.24%, and 19.77%. These results provided a clear seasonality, with lower system efficiencies during the summer due to the high temperatures. The performance is slightly lower than other PV power plants in the Mediterranean, although with more years of operation. Nevertheless, the PR is over 80% for almost 42% of the measured days, proving a correct performance. Furthermore, the pc-Si sectors provided PR values around 2% greater than the mc-Si sectors, mainly due to the higher PV temperature and degradation losses of the mc-Si sectors. The estimated degradation losses of the modules are approximately 2–3% lower than according to the manufacturer data. The degradation losses yield the greatest weight in the energy balance, representing a global energy loss of 13% of the global energy at STC.

After the performance analysis, a more in-depth study has been performed to reduce the outliers in the predictions in low irradiance days. A physical model was developed, as the product of the different losses of the PV system, including the LIL through irradiance measurements and an exponential fit. The results improved the nRMSE by 1.9% compared with the conventional model, increasing the R^2 from 0.144 to 0.553 for low irradiances.

A second approach has been applied using two statistical methods using only T_a and Y_R as predictors. The RF model has provided the best performance with a nRMSE of 1.27%. These results indicate a better performance than SVR models found in literature, which require significantly more predictors. In contrast, the MLR model has reduced the nRMSE by 2.30% with an accuracy of 87.85%.

The inclusion of the improvements in the PR and in the PV daily production model has provided improvements in nRMSE of 0.11%, 0.76%, and 2.19% for the exponential, MLR, and the RF models, respectively. These improvements are significantly greater for the low irradiance days, providing reductions in the nRMSE up to 7.27%, 7.75% and 11.07% for the exponential, MLR and RF models, respectively. In any case, both statistical models provided a better PR accuracy than the physical model, and are recommended to forecast the PR whenever measured data is available. Moreover, they constitute an alternative to model and predict the PR when there is scarce technical data of the plant.

As future work, the degradation of the PV modules will be studied in more detail by analyzing the performance after more years of operation.

Credit author statement

Enrique Fuster-Palop: Methodology, Software, Formal analysis, Resources, Validation, Investigation, Writing - Original Draft, Visualization, **Carlos Vargas-Salgado**: Conceptualization, Writing - Review & Editing, **Juan Carlos Ferri-Revert**: Resources, Investigation, **Jorge Payá**: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Valencia, May 23, 2022.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors gratefully acknowledge the operation & maintenance staff of the PV Power Plant in Olmedilla de Alarcón for providing the measured data of the solar PV power plant.

References

[1] David Feldman RM, Ramasamy Vignesh, Fu Ran, Ramdas Ashwin, Jal Desai, "U.S. Solar photovoltaic system and energy storage cost benchmark. 2020. p. 2021.
[2] Mitrašinović AM. Photovoltaics advancements for transition from renewable to clean energy. Energy 2021;237:121510. https://doi.org/10.1016/J.ENERGY.2021.121510. Dec.
[3] IRENA. Renewable Capacity Statistics 2021:2021.
[4] Sundaram S, Babu JSC. Performance evaluation and validation of 5 MWp grid connected solar photovoltaic plant in South India. Energy Convers Manag Aug. 2015;100:429–39. https://doi.org/10.1016/J.ENCONMAN.2015.04.069.
[5] Bansal N, Jaiswal SP, Singh G. Long term performance assessment and loss analysis of 9 MW grid tied PV plant in India," Mater. Today Proc., Feb. 2022. https://doi.org/10.1016/J.MATPR.2022.01.263.

[6] Oloya IT, Gutu TJL, Adaramola MS. Techno-economic assessment of 10 MW centralised grid-tied solar photovoltaic system in Uganda. Case Stud Therm Eng 2021;25:100928. https://doi.org/10.1016/J.CSITE.2021.100928. Jun.

[7] AL-Rasheedi M, Gueymard CA, Al-Khayat M, Ismail A, Lee JA, Al-Duaj H. Performance evaluation of a utility-scale dual-technology photovoltaic power plant at the Shagaya Renewable Energy Park in Kuwait. Renew Sustain Energy Rev 2020; 133:110139. https://doi.org/10.1016/J.RSER.2020.110139. Nov.

[8] Elhadj Sidi CEB, Ndiaye ML, El Bah M, Mbodji A, Ndiaye A, Ndiaye PA. Performance analysis of the first large-scale (15 MWp) grid-connected photovoltaic plant in Mauritania. Energy Convers Manag Jul. 2016;119:411–21. https://doi.org/10.1016/J.ENCONMAN.2016.04.070.

[9] Bentouba S, Bourouis M, Zioui N, Pirashanthan A, Velauthapillai D. Performance assessment of a 20 MW photovoltaic power plant in a hot climate using real data and simulation tools. Energy Rep Nov. 2021;7:7297–314. https://doi.org/10.1016/J.EGYR.2021.10.082.

[10] Dahmoun MEH, Bekkouche B, Sudhakar K, Guezgouz M, Chenafi A, Chaouch A. Performance evaluation and analysis of grid-tied large scale PV plant in Algeria. Energy Sustain. Dev. Apr. 2021;61:181–95. https://doi.org/10.1016/J.ESD.2021.02.004.

[11] Malvoni M, Leggieri A, Maggiotto G, Congedo PM, De Giorgi MG. Long term performance, losses and efficiency analysis of a 960 kWP photovoltaic system in the Mediterranean climate. Energy Convers Manag Aug. 2017;145:169–81. https://doi.org/10.1016/J.ENCONMAN.2017.04.075.

[12] Copper JK, Sproul AB, Jarnason S. Photovoltaic (PV) performance modelling in the absence of onsite measured plane of array irradiance (POA) and module temperature. Renew Energy Feb. 2016;86:760–9. https://doi.org/10.1016/J.RENENE.2015.09.005.

[13] IDAE. "Pliego de Condiciones técnicas de Instalaciones conectadas a red. 2011.

[14] Holmgren WF, Hansen CW, Mikofski MA. Pvlib python: a python package for modeling solar energy systems. J. Open Source Softw. Sep. 2018;3(29):884. https://doi.org/10.21105/JOSS.00884.

[15] Kymakis E, Kalykakis S, Papazoglou TM. Performance analysis of a grid connected photovoltaic park on the island of Crete. Energy Convers Manag Mar. 2009;50(3): 433–8. https://doi.org/10.1016/J.ENCONMAN.2008.12.009.

[16] Gilman P, Dobos A, Diorio N, Freeman J, Janzou S, Ryberg D. SAM photovoltaic model technical reference update. 2016.

[17] Ozden T. A countrywide analysis of 27 solar power plants installed at different climates. Sci. Reports 2022;12(1):1–11. https://doi.org/10.1038/s41598-021-04551-7. 2022 121.

[18] Khalid AM, Mitra I, Warmuth W, Schacht V. Performance ratio – crucial parameter for grid connected PV plants. Renew Sustain Energy Rev Nov. 2016;65:1139–58. https://doi.org/10.1016/J.RSER.2016.07.066.

[19] Quesada B, Sánchez C, Cañada J, Royo R, Payá J. Experimental results and simulation with TRNSYS of a 7.2kWp grid-connected photovoltaic system. Appl Energy May 2011;88(5):1772–83. https://doi.org/10.1016/j.apenergy.2010.12.011.

[20] Mavromatakis F, Vignola F, Marion B. Low irradiance losses of photovoltaic modules. Sol Energy Nov. 2017;157:496–506. https://doi.org/10.1016/J.SOLENER.2017.08.062.

[21] Soler-Castillo Y, Rimada JC, Hernández L, Martínez-Criado G. Modelling of the efficiency of the photovoltaic modules: grid-connected plants to the Cuban national electrical system. Sol Energy Jul. 2021;223:150–7. https://doi.org/10.1016/J.SOLENER.2021.05.052.

[22] Lindig S, Louwen A, Moser D, Topic M. Outdoor PV system monitoring—input data quality, data imputation and filtering approaches. Energies 2020;13. https://doi.org/10.3390/EN13195099. Page 5099, vol. 13, no. 19, p. 5099, Sep. 2020.

[23] Huld T, Gottschalg R, Beyer HG, Topič M. Mapping the performance of PV modules, effects of module type and data averaging. Sol Energy Feb. 2010;84(2): 324–38. https://doi.org/10.1016/J.SOLENER.2009.12.002.

[24] de la Parra I, Muñoz M, Lorenzo E, García M, Marcos J, Martínez-Moreno F. PV performance modelling: a review in the light of quality assurance for large PV plants. Renew Sustain Energy Rev Oct. 2017;78:780–97. https://doi.org/10.1016/J.RSER.2017.04.080.

[25] Roumpakias E, Stamatelos A. Comparative performance analysis of grid-connected photovoltaic system by use of existing performance models. Energy Convers Manag Oct. 2017;150:14–25. https://doi.org/10.1016/J.ENCONMAN.2017.08.001.

[26] Busch L, Schäfer T, Song W, Mack M, Egler M. PV system energy yield calculation program PR-FACT," 28th. Eur. Photovolt. Sol. Energy Conf. Exhib. Nov. 2013: 3699–708. https://doi.org/10.4229/28THEUPVSEC2013-5CO.6.3.

[27] Parretta A, Sarno A, Vicari LRM. Effects of solar irradiation conditions on the outdoor performance of photovoltaic modules. Opt Commun 1998;153(1–3): 153–63. https://doi.org/10.1016/S0030-4018(98)00192-8.

[28] Trigo-Gonzalez M, et al. Development and comparison of PV production estimation models for mc-Si technologies in Chile and Spain. J Clean Prod 2021;281:125360. https://doi.org/10.1016/j.jclepro.2020.125360. Jan.

[29] Tossa AK, et al. Artificial intelligence technique for estimating PV modules performance ratio under outdoor operating conditions. J Renew Sustain Energy Oct. 2018;10(5):053505. https://doi.org/10.1063/1.5042217.

[30] Bandong S, Leksono E, Purwarianti A, Joelianto E. Performance ratio estimation and prediction of solar power plants using machine learning to improve energy reliability. SAVE Proc 2019:36–41. https://doi.org/10.1109/ICA.2019.8916687. 2019 6th Int. Conf. Instrumentation, Control. Autom. ICA, . [Accessed July 2019].

[31] Hashemi B, Taheri S, Cretu AM, Pouresmaeil E. Systematic photovoltaic system power losses calculation and modeling using computational intelligence techniques. Appl Energy 2021;284. https://doi.org/10.1016/J.APENERGY.2020.116396. Feb.

[32] Ahmad MW, Mourshed M, Rezgui Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy Dec. 2018;164:465–74. https://doi.org/10.1016/J.ENERGY.2018.08.207.

[33] Kunaifi K, Reinders A, Lindig S, Jaeger M, Moser D. Operational performance and degradation of PV systems consisting of six technologies in three climates. Appl Sci 2020;10. https://doi.org/10.3390/APP10165412. Page 5412, vol. 10, no. 16, p. 5412, Aug. 2020.

[34] International Electrotechnical Committee. IEC 61724 photovoltaic system performance monitoring - guidelines for measurement, data exchange and analysis. 1998 [Online]. Available: https://global.ihs.com/doc_detail.cfm?document_name=IEC 61724& item_s_key=00290218. [Accessed 27 February 2022]. Accessed.

[35] Dhimish M. Thermal impact on the performance ratio of photovoltaic systems: a case study of 8000 photovoltaic installations. Case Stud Therm Eng 2020;21: 100693. https://doi.org/10.1016/J.CSITE.2020.100693. Oct.

[36] Ransome S, Funtan P. Why hourly averaged measurement data is insufficient to model PV system performance accurately. In: 20th European photovoltaic solar energy conference; 2005. p. 2752–5.

[37] Cubukcu M, Gumus H. Performance analysis of a grid-connected photovoltaic plant in eastern Turkey. Sustain Energy Technol Assessments 2020;39:100724. https://doi.org/10.1016/J.SETA.2020.100724. Jun.

[38] Roberts JJ, Mendiburu Zevallos AA, Cassula AM. Assessment of photovoltaic performance models for system simulation. Renew Sustain Energy Rev May 2017; 72:1104–23. https://doi.org/10.1016/J.RSER.2016.10.022.

[39] Pérez NS, Alonso-Montesinos J, Batlles FJ. Estimation of soiling losses from an experimental photovoltaic plant using artificial intelligence techniques. Appl Sci 2021;11. https://doi.org/10.3390/APP11041516. Page 1516, vol. 11, no. 4, p. 1516, Feb. 2021.

[40] Polo J, et al. Characterization of PV soiling losses in urban mediterranean environment. 2019. https://doi.org/10.18086/swc.2019.15.03.

[41] Kim J, Rabelo M, Padi SP, Yousuf H, Cho EC, Yi J. A review of the degradation of photovoltaic modules for life expectancy. Energies 2021;14. https://doi.org/10.3390/EN14144278. Page 4278, vol. 14, no. 14, p. 4278, Jul. 2021.

[42] Shiva Kumar B, Sudhakar K. Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India. Energy Rep Nov. 2015;1:184–92. https://doi.org/10.1016/J.EGYR.2015.10.001.

[43] Marion B, et al. Performance parameters for grid-connected PV systems. Conf Rec IEEE Photovolt Spec Conf 2005:1601–6. https://doi.org/10.1109/PVSC.2005.1488451.

[44] Santiago I, Trillo-Montero D, Moreno-Garcia IM, Pallarés-López V, Luna-Rodríguez JJ. Modeling of photovoltaic cell temperature losses: a review and a practice case in South Spain. Renew Sustain Energy Rev Jul. 2018;90:70–89. https://doi.org/10.1016/J.RSER.2018.03.054.

[45] Malvoni M, Kumar NM, Chopra SS, Hatziargyriou N. Performance and degradation assessment of large-scale grid-connected solar photovoltaic power plant in tropical semi-arid environment of India. Sol Energy Jun. 2020;203:101–13. https://doi.org/10.1016/J.SOLENER.2020.04.011.

[46] Taylor J, Leloux J, Hall LMH, Everard AM, Briggs J, Buckley A. Performance of distributed PV in the UK: a statistical analysis of over 7000 systems. Conf. 31st Eur. Photovolt. Sol. Energy Conf. Exhib. 2015. https://doi.org/10.13140/RG.2.1.2019.6568.

[47] Seltman HJ. Experimental design and analysis. 2018.

[48] Vergura S. A statistical tool to detect and locate abnormal operating conditions in photovoltaic systems. Sustain Times 2018;10. https://doi.org/10.3390/SU10030608. Page 608, vol. 10, no. 3, p. 608, Feb. 2018.

[49] Kabacoff RI. R in Action SECOND EDITION Data analysis and graphics with R. 2015.

[50] Flatt C, Jacobs RL. Principle assumptions of regression analysis: testing, techniques, and statistical reporting of imperfect data sets. vol. 21, no. 4, pp. 484–502, https://doi.org/10.1177/1523422319869915; 2019. 10.1177/1523422319869915.

[51] Fávero LP, Belfiore P. Simple and multiple regression models. Data Sci. Bus. Decis. Mak. Jan. 2019:443–538. https://doi.org/10.1016/B978-0-12-811216-8.00013-6.

[52] Breiman L. Random forests. Mach Learn 2001;451. https://doi.org/10.1023/A:1010933404324. 45, no. 1, pp. 5–32, Oct. 2001.

[53] Kuhn M. Building predictive models in R using the caret package. J Stat Software Nov. 2008;28(5):1–26. https://doi.org/10.18637/JSS.V028.I05.

[54] Probst P, Wright MN, Boulesteix AL. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. May 2019;9(3): e1301. https://doi.org/10.1002/WIDM.1301.

[55] Belgiu M, Drăgu L. Random forest in remote sensing: a review of applications and future directions. ISPRS J Photogrammetry Remote Sens Apr. 2016;114:24–31. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011.

[56] Assouline D, Mohajeri N, Scartezzini JL. Large-scale rooftop solar photovoltaic technical potential estimation using Random Forests. Appl Energy May 2018;217: 189–211. https://doi.org/10.1016/J.APENERGY.2018.02.118.

[57] Mathai N, Chen Y, Kirchmair J. Validation strategies for target prediction methods. Briefings Bioinf May 2020;21(3):791–802. https://doi.org/10.1093/BIB/BBZ026.

[58] Kuhn M, Johnson K. Applied predictive modeling. Appl. Predict. Model. Jan. 2013: 1–600. https://doi.org/10.1007/978-1-4614-6849-3.

[59] Bounoua Z, Ouazzani Chahidi L, Mechaqrane A. Estimation of daily global solar radiation using empirical and machine-learning methods: a case study of five Moroccan locations. Sustain. Mater. Technol. 2021;28:e00261. https://doi.org/10.1016/J.SUSMAT.2021.E00261. Jul.

[60] Guenounou A, Malek A, Aillerie M. Comparative performance of PV panels of different technologies over one year of exposure: application to a coastal

Mediterranean region of Algeria. Energy Convers Manag Apr. 2016;114:356–63. https://doi.org/10.1016/J.ENCONMAN.2016.02.044.
[61] Elamim A, Hartiti B, Haibaoui A, Lfakir A, Thevenin P. Analysis and comparison of different PV technologies for determining the optimal PV panels- A case study in Mohammedia , Morocco. IOSR J Electr Electron Eng Jan. 2017;12:37–45. https://doi.org/10.9790/1676-1201013745. 01.
[62] Alshare A, Tashtoush B, Altarazi S, El-Khalil H. Energy and economic analysis of a 5 MW photovoltaic system in northern Jordan. Case Stud Therm Eng 2020;21:100722. https://doi.org/10.1016/J.CSITE.2020.100722. Oct.
[63] Martín-Martínez S, Cañas-Carretón M, Honrubia-Escribano A, Gómez-Lázaro E. Performance evaluation of large solar photovoltaic power plants in Spain. Energy Convers Manag Mar. 2019;183:515–28. https://doi.org/10.1016/J.ENCONMAN.2018.12.116.
[64] Eke R, Demircan H. Performance analysis of a multi crystalline Si photovoltaic module under Mugla climatic conditions in Turkey. Energy Convers Manag Jan. 2013;65:580–6. https://doi.org/10.1016/J.ENCONMAN.2012.09.007.
[65] Murat Ates A, Singh H. Rooftop solar Photovoltaic (PV) plant – one year measured performance and simulations. J King Saud Univ Sci May 2021;33(3):101361. https://doi.org/10.1016/J.JKSUS.2021.101361.
[66] Bouacha S, et al. Performance analysis of the first photovoltaic grid-connected system in Algeria. Energy Sustain. Dev. Aug. 2020;57:1–11. https://doi.org/10.1016/J.ESD.2020.04.002.
[67] Attari K, Elyaakoubi A, Asselman A. Performance analysis and investigation of a grid-connected photovoltaic installation in Morocco. Energy Rep Nov. 2016;2:261–6. https://doi.org/10.1016/J.EGYR.2016.10.004.
[68] Savvakis N, Tsoutsos T. Performance assessment of a thin film photovoltaic system under actual Mediterranean climate conditions in the island of Crete. Energy 2015;90:1435–1455, Oct. https://doi.org/10.1016/J.ENERGY.2015.06.098.
[69] Herbazi R, et al. Performance evaluation and analysis of polycrystalline photovoltaic plant located in Northern Morocco. https://doi.org/10.1080/01430750.2019.1694985; 2019. 10.1080/01430750.2019.1694985.
[70] A. A. Merrouni, A. I. Amrani, and A. Mezrhab, "Electricity production from large scale PV plants: benchmarking the potential of Morocco against California, US," Energy Procedia, vol. 119, pp. 346–355, Jul. 2017, doi: 10.1016/J.EGYPRO.2017.07.118.
[71] Government of Spain. Ministry of Industry Tourism and Trade, "Royal Decree 661/2007," BOE, 2007. [Online]. Available: https://www.boe.es/eli/es/rd/2007/05/25/661. [Accessed 14 September 2022]. Accessed.
[72] Government of Spain. Ministry of industry tourism and trade. Royal Decree 413/2014," BOE, 2014. [Online]. Available: https://www.boe.es/buscar/doc.php?id=BOE-A-2014-6123. Accessed: 18-Sep-2022.
[73] Blanco-Díez P, Díez-Mediavilla M, Alonso-Tristán C. Review of the legislative framework for the remuneration of photovoltaic production in Spain: a case study. Sustain Times 2020;12. https://doi.org/10.3390/SU12031214. Page 1214, vol. 12, no. 3, p. 1214, Feb. 2020.
[74] Bansal N, Pany P, Singh G. Visual degradation and performance evaluation of utility scale solar photovoltaic power plant in hot and dry climate in western India. Case Stud Therm Eng 2021;26:101010. https://doi.org/10.1016/J.CSITE.2021.101010. Aug.
[75] Comisión Nacional de los Mercados y la Competencia (CNMC). Estadísticas. Precios del Mercado de Producción de Energía Eléctrica [Online]. Available: https://www.cnmc.es/estadisticas?hidtipo=12750. [Accessed 18 September 2022]. Accessed.
[76] Comisión Nacional de los Mercados y la Competencia (CNMC). Evaluación del régimen retributivo específico correspondiente al periodo 2014-2020. Expediente núm. INF/DE/037/21 2021;1. https://www.cnmc.es/expedientes/infde03721. [Accessed 12 September 2022]. 8-8.
[77] Walker A, et al. Model of operation-and-maintenance costs for photovoltaic systems. 2020.
[78] inflation.eu. Historic inflation Spain – historic CPI inflation Spain [Online]. Available: https://www.inflation.eu/en/inflation-rates/spain/historic-inflation/cpi-inflation-spain.aspx [Accessed: 18-Sep-2022.
[79] Ministry for the ecological transition and the demographic challenge, "orden TED/171/2020. BOE, 2021. [Online]. Available: https://www.boe.es/buscar/act.php?id=BOE-A-2020-2838. [Accessed 12 September 2022]. Accessed.
[80] Wilcox RR. ONE-WAY ANOVA. Appl. Contemp. Stat. Tech. Jan. 2003:285–328. https://doi.org/10.1016/B978-012751541-0/50030-4.
[81] Liu H. Comparing Welch's ANOVA, a kruskal-wallis test and traditional ANOVA in case of heterogeneity of variance. Theses Diss.; 2015. https://doi.org/10.25772/BWFP-YE95.
[82] Schreiber-Gregory DN, Jackson HM, Bader K. Logistic and linear regression assumptions : violation recognition and control. 2018.