Photovoltaic power electricity generation nowcasting combining sky camera images and learning supervised algorithms in the Southern Spain

Mauricio Trigo-González [a], Marcelo Cortés-Carmona [a], Aitor Marzo [b,*], Joaquín Alonso-Montesinos [c,d], Mercedes Martínez-Durbán [e], Gabriel López [f], Carlos Portillo [a], Francisco Javier Batlles [c,d]

[a] Centro de Desarrollo Energético Antofagasta (CDEA), Universidad de Antofagasta, Avda. Angamos 601, 1270300, Antofagasta, Chile
[b] Department of Optics, University of Granada, 18071, Granada, Spain
[c] Department of Chemistry and Physics, University of Almería, 04120, Almería, Spain
[d] CIESOL, Joint Centre of the University of Almería-CIEMAT, 04120, Almería, Spain
[e] Department of Informatics, University of Almería, 04120, Almería, Spain
[f] Dpto. Ingeniería Eléctrica y Térmica, de diseño y Proyectos, Universidad de Huelva, 21007, Huelva, Spain

ARTICLE INFO

ABSTRACT

The alternation between cloudy and clear skies alters the photovoltaic production. This makes it necessary to anticipate these disturbances hours in advance for the correct operation of the electricity distribution plants and networks. In this paper, two short-term forecasting models (3 h) are developed to forecast the photovoltaic production in an integrated plant in the CIESOL building of the University of Almería. The methodology used is based on sky camera images and Artificial Intelligence techniques. Two models have been developed and compared applying artificial neural network (ANN) and support vector machine (SVM) techniques. The global irradiance predicted using sky camera images is used as an input variable in both models. In addition, the operational status of the plants has been included as an input parameter through the performance ratio. The results have shown that the errors made by ANN and SVM are very similar. For all sky conditions, the uncertainty of the production forecast differs by less than 2% from the uncertainty of the solar resource, which is the main source of error in the production models developed.

1. Introduction

The development of societies is linked to energy consumption, which is necessary to meet their development and human needs. Fossil fuels and other conventional energy sources are limited, so their prices are subject to availability and the market, and they contribute to greenhouse gas emissions and thus to global warming [1]. Global warming is already a reality and a very serious issue, as reflected in the latest report of the Intergovernmental Panel on Climate Change (IPCC), which warns that emissions must be halved by 2030 if the temperature increase is to be limited to 1.5 °C in the face of the evidence presented [2].

Solar photovoltaic (PV) is the fastest growing power generation technology since 2002, with an average annual increase of 48%. The latest IRENA report shows that the world's installed solar capacity was 716 GW in 2020, 99% of which corresponds to PV technology. Currently the number of off-grid installations has grown considerably to 4,314 MW

[3]. The IEA in its 2020 Task 1 report indicates that of the 138 GW installed in Europe, almost 55% are residential and commercial rooftop installations [4]. This shows that the price of photovoltaic electricity is competitive with the price of conventional electricity.

One of the factors to be taken into account when installing a photovoltaic power plant or supplying electricity to a house is the distance between the place of electricity generation and the place of consumption. The greater the distance, the higher the price of energy due to the increased cost of the power line. In other words, the ideal solution is for the production to be generated as close as possible to the place of consumption. Distributed electricity grids and self-consumption systems are based on these principles. The basic principle of a self-consumption system is to supply electricity 24 h a day, regardless of the day-night cycle or inclement weather.

Nowadays, self-consumption emerges as an option in the context of rising electricity market prices and as an ecological alternative to the

---

consumption of energy from conventional sources. The main components of a self-consumption system are the photovoltaic modules, the batteries, the charge regulator system and the current converter or solar inverter [5]. The PV modules transform the incident solar radiation into direct current. The batteries store the unconsumed energy. The charge regulator allows access to the stored energy at times when production is insufficient. And finally, the current converter, called photovoltaic inverter, transforms the direct current generated by the PV modules into 220V alternating current. One of the advantages of self-consumption is the elimination of the costs associated with transporting the electricity supply and intermediaries. In recent times, self-consumption of electricity has evolved from the residential sector of single-family homes to small distribution networks in small housing developments in and, nowadays, to large plants in buildings and industrial warehouses of small and medium-sized companies [3].

An important factor to consider in self-consumption systems is the possibility of connection to the electricity grid [6]. If the system is not connected, it is self-sufficient and is called self-consumption without surpluses. This means that the system only produces what is consumed. The most important element of this type of system is the battery. In contrast to self-consumption without surpluses, self-consumption with surpluses injects the excess energy into the electricity grid, i.e., the user sells the excess energy produced to the electricity company. This type of self-consumption depends on the political strategy of each country and the rates imposed by the electricity company. In general, the company sets a purchase price, which is usually lower than the production price and variable depending on the market and fixes the cost of transport from the point of generation to the point of connection to the grid. An intermediate solution is self-consumption with a net balance, where the user can inject or take from the grid depending on the existence of a deficit or surplus of energy for consumption. In this case, both the selling price of surplus energy and the purchase of the energy deficit are set by the electricity company and can vary throughout the day depending on market prices. Given these uncertainties, it is difficult to predict the profitability of the self-consumption system with net balance. Regardless of the type of self-consumption system available, it must ensure the stability of the electricity supply.

A solar photovoltaic plant can be affected not only by inclement weather [9], but also by other factors related to the structure of the plant itself, such as losses due to orientation, wiring, inverters, excessive heating of the panels, degradation and soiling, as the most important factors [7–9]. There is a term that tries to collect all losses and unify them into a single factor known as the performance ratio (PR). The performance ratio is a global loss factor that considers all the losses of a photovoltaic solar power plant. For this reason, the PR is often described as a production index for quantifying the PV-plant quality, varying between 0 and 1, with low values representing many losses and high values representing few losses in the PV electricity generation [10]. In arid climates, production losses of up to 25% after three months of operation have been quantified [11].

To avoid the problem of transient phenomena, it is necessary to know in advance the amount of radiation that will reach the solar field. However, it is a difficult task to accurately predict the timing, duration and quantification of a significant drop in the solar resource. Long-term forecasts, such as those made 24 h in advance, do not allow detailed estimation of global irradiance variations over short time intervals on the order of a minute. Short-term solar irradiance forecasting, also known as nowcasting, is a current topic of research and provides predictions on a time horizon of the order of one to 3 h and with a resolution of the order of a minute.

Artificial Intelligence techniques based on machine learning have been used with remarkable success to estimate the solar resource. Machine learning aims to develop techniques that allow computers to learn. To this end, there are two options: techniques with unsupervised training and with supervised training. The first ones include self-organizing networks that group similar input vectors or patterns without the use of training data to specify the behaviour of a typical component of the group to which the vector or pattern belongs. Artificial Neural Networks (ANNs) are one of the most extensive and widely used techniques as a supervised training network corresponding to non-linear models based on machine learning. Different researchers have used ANNs to estimate solar radiation [12–15], as well as to model cooling systems [16]. López et al. [17] used ANNs to provide 1-day-ahead forecasts of global solar radiation under different sky conditions. In this paper, it was shown that adding time series inputs, in the form of common weather variables, does not improve the current ANN model. Gomes et al. [18] forecasted the solar resource in a 4-h time horizon, using meteorological variables as inputs to the ANN.

Support Vector Machines (SVMs) has proved to be one of the best classifiers for a wide range of situations and is therefore considered one of the benchmarks in the field of statistical learning and machine learning. Its applications in the photovoltaic area have been numerous, many of them also in production prediction of photovoltaic plants [19–30]. Among them, the work published by Shi J. et al. [30] in 2012, where they developed a one-day-ahead PV power production forecasting model for a single station, based on weather forecast data, real historical power production data and the SVM principle, stands out. After applying it to a PV station in China (with a capacity of 20 kW), the results showed that the proposed forecasting model for grid-connected PV systems was effective and promising. The proposed model showed errors of 2.10 MW RMSE and 8.64% MRE. Subsequently, Eseye et al. [28] proposed a hybrid short-term forecasting model combining wavelet transform, particle swarm optimisation and SVMs (Hybrid WT-PSO-SVM) in 2018 for short-term (day-ahead) generation power forecasting of a real microgrid PV system. The model showed superior performance with respect to seven other forecasting strategies with which it was compared, MAPE and NMAE had 4.22% and 0.4% average values, respectively. More recently in 2019, Catalina et al. [31] compared different machine learning techniques in combination with satellite data for the development of short-term PV production forecasting models (nowcasting) with good results.

The METEOSAT satellite, in its different generations, is one of the most widely used satellites for weather forecasting, representing a breakthrough in the study of solar resource assessment and forecasting. Escrig et al. in 2013 and Alonso-Montesinos et al. in 2016 developed a methodology to detect and classify cloud height (low, medium and high) [32,33]. Escrig [32] uses a meteorological model and Alonso-Montesinos [33] Bayesian classifiers. The METEOSAT Second Generation (MSG) satellite, also for has been used for nowcasting the type of cloud present on the observer's horizon [34]. Alonso-Montesinos et al. estimated and nowcast solar radiation with a 3-h time horizon using an adaptation of the Heliosat 2 model to the MSG satellite. The problem with the satellite lies in the overlapping of clouds. In the case of low, medium and high clouds, the satellite does not recognize the low clouds, because its spatial view is from the top to the bottom of the atmosphere [35,36].

In recently, sky cameras are being used for meteorological applications. Sky cameras are devices that provide a hemispheric view of the sky, with a temporal resolution that can be less than a minute. In contrast to the satellite, the field of view is from the lowest to the highest layers of the atmosphere. That is, in the case of overlapping clouds, the first to be seen by the cameras are the low clouds. The investigations that can be done with the camera are very wide and diverse, from cloud detection to aerosol characterization. The most significant applications are: aerosol optical depth analysis [37,38], cloud identification and classification [39–41], solar resource assessment [42] and the estimation of the three components of solar radiation (global, direct and diffuse), for which digital image levels have been used [35].

The aim of this paper is to develop a methodology based on sky cameras images and Artificial Intelligence techniques to forecast in the short term (1–180 min) the energy production of a photovoltaic plant. For this, the performance of two models based on artificial neural

networks (ANN) and support vector machines (SVM) will be developed and compared using the database of the photovoltaic plant located in the CIESOL building in Almería, Spain. In addition to the solar resource predicted with the sky camera system, the operational state of the plant will be considered as an input parameter, which will contribute to improving the performance of the prediction models. These models will be a fundamental support tool in the operation tasks of PV plants and electrical distribution networks. The models will allow the accurate prediction of production ramps due to cloud transit, giving operators time to mitigate such changes in electricity production, which can worsen the quality of supply. Therefore, these models will mean economic savings for companies dedicated to the exploitation of photovoltaic plants.

2. Location, instrumentation, and data

2.1. Location

Both the photovoltaic plant and the meteorological station are sited in the CIESOL building. This building is located in the northern part of the campus of the University of Almería, in the southeast of the Iberian Peninsula on the Mediterranean coast. The building has a south-east orientation and is 700 m from the coast. The solar field is integrated into the roof of the CIESOL building on a 30° sloping surface. The photovoltaic plant covers the electrical needs of the building and operates on a self-consumption basis with net balance.

Almería has an annual number of sunshine hours in the order of 3,000 and the global daily average daily radiation is 5 kWh m^{-2}. The annual average maximum and minimum temperatures are of the order of 35 °C and 25 °C, respectively, and the annual rainfall is of the order of 228 mm. Its climate is Warm steppe hot summers (BSh), according to the Köppen-Geiger climate classification [43].

2.2. Instrumentation

Local values of global, direct and diffuse irradiance, air temperature and relative humidity are recorded at the weather station. The global and diffuse horizontal irradiances are measured with two pyranometers model CMP 11. According to the ISO 9060, CMP 11 are classified as spectrally flat Class A radiometer and has an expected daily uncertainty of 2%. The direct normal irradiance is measured with a pyrheliometer model CHP1. According to the ISO 9060, the CHP1 is a first-class classification pyrheliometer has an expected daily uncertainty of 1%. All the radiometers are of the Kipp&Zonen manufacture and are installed on a Kipp&Zonen 2AP Sun Tracker All technical information on the instruments can be found at [44]. Diffuse horizontal and direct normal irradiances are use for the quality control of global horizontal irradiance measurements according the quality control criteria defined in Ref. [45].

Air temperature is measured with a thermocouple HMP45A/D from Vaisala. The uncertainty value [46] is less than ± 0.4 °C [47].

A TSI 880 model all-sky camera complements meteorological measurements with images of sky conditions. The all-sky camera has a hemispherical view and a rotational shadowband (specifically a TSI 880 model). The images are saved in JPEG (joint photographic expert group) format with a resolution of 352x288 pixels. Each pixel can have a value between 0 and 255 of digital levels. All images were recorded in 1-min periods for solar elevations higher than 5° [48].

Both weather variables and sky images are recorded every minute and stored on a data server.

2.3. PV plant characteristics

The CIESOL photovoltaic plant has an installed power of 9 kWp. The photovoltaic modules used are ATERSA model A-222P. These modules are made up of 60 polycrystalline cells and have a power of 222 Wp each. The installation consists of 42 panels arranged in three series of 14

modules. They face south and are tilted at an angle of 22° to the horizontal plane. The installation also has three CICLO model 3,000 DC/AC inverters. Table 1 shows the electrical characteristics of the installation. Fig. 1 shows a diagram of the same installation operating in a self-consumption regime with net balance. The CIESOL PV plant has been in continuous operation since 2010 and measurements from the plant have been recorded since then.

2.4. Database

For the generation of the database, the measurements recorded in the photovoltaic plant and the meteorological station located in the CIESOL building of the University of Almería, Spain, have been used. The database comprises a year of measurements between May 2013 and May 2014.

On one hand, the recorded parameters were global horizontal irradiance (GHI), ambient temperature (T), and generated photovoltaic power (P). The measurements were performed every minute and their average values were recorded in the database in 15 min intervals.

On the other hand, in order to increase the information contained in the database, other parameters such as those related to the relative position of the sun were calculated: solar zenith angle (θ_z), solar azimuth angle (θ_{az}) and extraterrestrial irradiance (I_0); and the performance ratio (PR) of the PV plant. These parameters are described below.

The solar zenith angle, θ_z, is the angular distance between the zenith of the observer and the position of the sun on the celestial vault. The θ_z is obtained from the following expression [49]:

$$cos(\theta_z) = sen(\delta) * sen(\varnothing) + cos(\delta) * cos(\varnothing) * \cos(\omega) \tag{1}$$

where δ is the solar declination, $\varnothing$ the latitude of the location and ω is the hour angle.

Extraterrestrial solar irradiance (I_0) is defined as the power density of solar radiation received on a fictitious surface of unit area parallel to the Earth's surface and located at the outer boundary of the Earth's atmosphere. It is calculated as follows [49]:

$$I_0 = E_0 \, I_{sc} \cos \theta_z \tag{2}$$

Where E_0 is the Earth orbit eccentricity factor correction and I_{sc} is the solar constant (1361.1 W m^{-2}).

The Performance Ratio is the relationship between the energy effectively produced (used), with respect to the energy that would be produced if the system were working under standard conditions (STC). Therefore, it can serve as an indicator of production losses due to faults, soiling, efficiency and other factors, that is, it is related to the operating status of the PV plant. The PR is defined in the IEC EN 61724 standard. It

Table 1
Electrical characteristics of the PV modules.

| Parameter	Values
Model	A-222P
Power	2.22 kW
Current I_{SC}	8.17 A
Current I_{MPP}	7.57 A
Voltage V_{MPP}	29.32 V
Voltage V_{OC}	36.42 V
$T_K(P_{MPP})$	−0.46%/C
$T_K (V_{OC})$	−0.35%/C
$T_K (I_{SC})$	+0.05%/C
Efficiency (η)	13.63%
Number of PV panels	42
PV panel inclination	22
Azimuth angle (α)	0 S
Number of inverters	3
PV plant inverter capability	2.5 kW
Efficiency of PV plant inverters	96%
PV plant power	9.3 kW$_P$
Installation year	2009
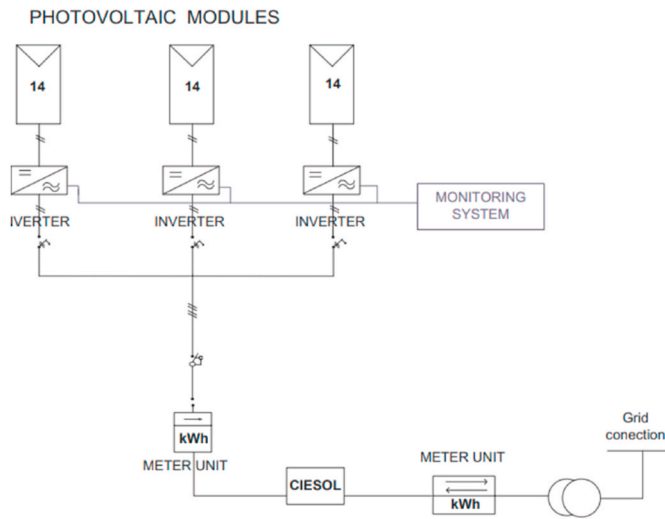
Fig. 1. CIESOL PV system diagram.

is calculated from the following expression [10]:

$$PR = \frac{\frac{E}{P_{STC}}}{\frac{H}{G_{STC}}} \qquad (3)$$

where E is the generated energy in Wh, P_{STC} is the installed capacity in Wp, H is the in-plane irradiance and G_{STC} is the irradiance under standard testing conditions (STC), i.e., 1000 W m^{-2}.

3. Methodology

The main objective of this paper is to forecast the production of a PV plant considering all sky conditions and the actual operational status of the plant. For this purpose, artificial intelligence techniques have been combined with solar resource forecasting techniques based on the processing of sky camera images. The sky camera, TSI-880, allows to forecast global horizontal irradiance (GHI) levels up to 3 h in advance. This forecasting is fed into a model developed with the Support Vector Machine (SVM) technique and Artificial Neural Network (ANN). Finally, the forecasting of the solar resource and the SVM model allow forecasting the power generated by the PV plant up to 3 h in advance.

Fig. 2 describes the steps followed in the methodology. Several

actions necessary to carry out the forecasting of the PV plant output can be seen in the graph. In general terms, the methodology can be summarised in two main stages: the development of the PV plant production model and the application of this model for production forecasting.

The model development stage is divided into the data analysis phase, the variable selection phase, the training or model development phase and the model validation phase. The application stage of the selected model for plant production forecasting is divided into the database generation, model application and forecasting phase. It should be noted that the predictive model is based on the forecasting of the solar resource at the PV plant with a model based on previously developed sky camera images. In the following subsections these actions will be presented in more detail.

3.1. Sky camera-based model for the GHI forecast

3.1.1. GHI estimation per pixel

The image provided by the all-sky camera contains information in the red, green and blue channels that make up the RGB colour space. The correlation between the digital levels and the GHI has been obtained following the methodology presented in detail in Refs. [42,48].

In order to find such a correlation, it is necessary to consider that the pixel levels behave differently depending on the proximity to the solar disc and the time of day. As demonstrated in Ref. [50], the area around the solar disc always appears more saturated than the rest of the image, and its extent varies with time and solar elevation. Therefore, to take all this into account in the correlation between digital levels and GHI, it is necessary to calculate the position of the sun and the distance to it of each pixel in each image [42]. Thus, the treatment of the pixels in the image will be done differently according to the distance to the sun, for which three different zones were defined: zone 1, the closest to the sun; zone 2, an intermediate zone; and zone 3, the furthest away from the sun.

Another step for the correct correlation between digital levels and GHI is the identification of the clear or cloudy sky condition for each pixel. For this, the methodology based on the processing of the different colour channels of sky camera images presented in Ref. [40] is used. Finally, a correlation between the digital pixel levels, the distance to the sun position and the sky condition with the GHI is obtained.

3.1.2. Determination of cloud motion vectors (CMV)

To make a forecasting, a sequence of three consecutive sky camera images, spanning about 3 min, is used. The correlation between these
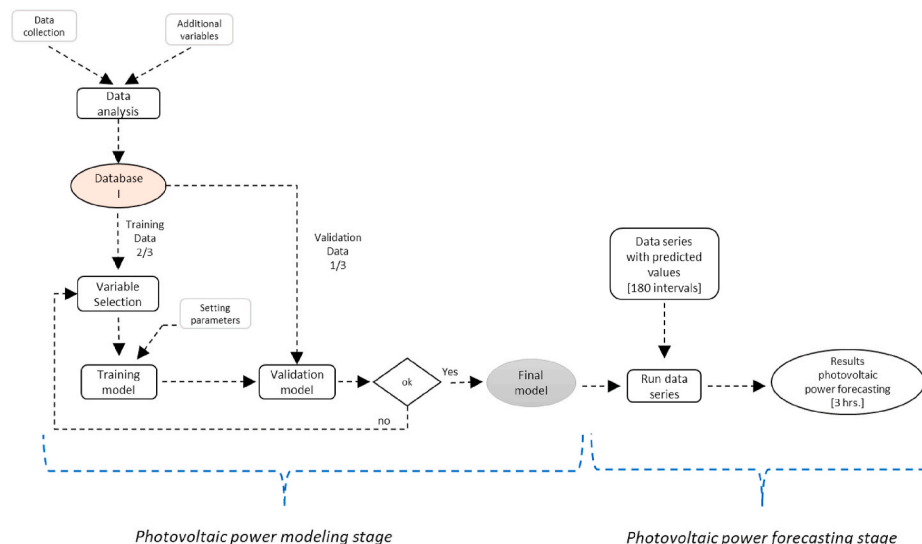

Fig. 2. Flowchart of the methodology for the PV production forecast by SVM technique.

three images makes it possible to establish the behavioural pattern of cloud movement at a given time. In order to study cloud movement, the following steps are taken [48,51,52].

- The picture taken with the sky camera is divided into different sectors, since the movement of the clouds will depend on the sector covered by the sky camera.
- The cloud motion vector is calculated for each sector by applying the maximum cross-correlation method.
- Different quality tests are applied to ensure the correct determination of the cloud motion.

Once the characteristic motion of each sector of the image is obtained, it is applied to the last image taken to estimate future pixel motions.

However, in this paper we do not intend to estimate cloud motion, but rather the variation of irradiance levels. Therefore, it is necessary to estimate the global component of the solar irradiance at each image pixel. The characteristic motion of each sector of the image will be applied to the estimated irradiance at each pixel, allowing the forecast of the global irradiance motion at each pixel of the sky image.

3.1.3. Motion of pixels and GHI forecasting

The CMV is applied to the last image received and re-applied to the result obtained. This process is repeated up to 180 times, obtaining the movement of the pixels from the minute in which the image was taken to the 180th minute in the future. Therefore, each application of the CMV is 1 min of forecasting. Specifically, the CMVs are applied to the estimated irradiance levels in the image, i.e., the motion is applied to each pixel and thus to the GHI values.

As the pixels move from 1 min to the next, the average value of the irradiance values after each movement, in each of the zones, has to be calculated. The final value of the GHI forecasting will be obtained from the average value of the three image zones [48,52].

3.2. Supervised learning models

3.2.1. Artificial neural network (ANN)

The type of neural network used is the multilayer fed perceptron (MLP) [53]. This type of neural network performs non-polynomial regression. In the MLP model, there is an input layer, a hidden layer and an output layer, each consisting of elementary units called neurons or nodes (see Fig. 3). The neurons in the input layer only receive the input signals and distribute them to the network. In subsequent layers, each neuron receives a signal, which is a weighted sum of the outputs of the previous layer's nodes. For node i in a subsequent layer, l, the total input x is given by

$$x_i^l = \sum_{j=0}^{N^{l-1}} w_{ij}^l y_j^{l-1} \tag{4}$$

being N the number of nodes, w_{ij} the weight for the connection between node j and node i and y_j the output of node j. The weight w_{i0} represents the bias of each node i with a constant unit activation ($y_0 \equiv 1$) [12,13, 54]. A nodal activation function is used to obtain the output of a node in a hidden or output layer, using the node's input as an argument. The sigmoid activation function from 0 to 1 has been used for the hidden neurons.

$$y = \frac{1}{1 + e^{-x}} \tag{5}$$

while the linear function ($y = x$) has been used for the output nodes. Although the number of input and output neurons is determined by the number of variables, the definition of the number of hidden neurons is not so simple. In current applications, the number of intermediate neurons is often decided in a heuristic way. In this work, the optimal number of intermediate neurons has been determined empirically as the minimal number of neurons for which estimation performance on a test set is satisfying [13], for more information see Refs. [15,23,55]. [13].

3.2.2. Support Vector Machines (SVM)

Support vector machines (SVM) is a supervised learning algorithm used in many classification and regression problems. Although originally developed as a binary classification method, its application has been extended to multiple classification and regression problems. It was developed by Vladimir Vapnik and his team at AT&T Labs [56].

Basically, Support Vector Machines are based on the Maximal Margin Classifier, which, in turn, is based on the hyperplane concept. The SVM algorithm searches for a hyperplane that separates different classes of data points with the widest possible margin between them.
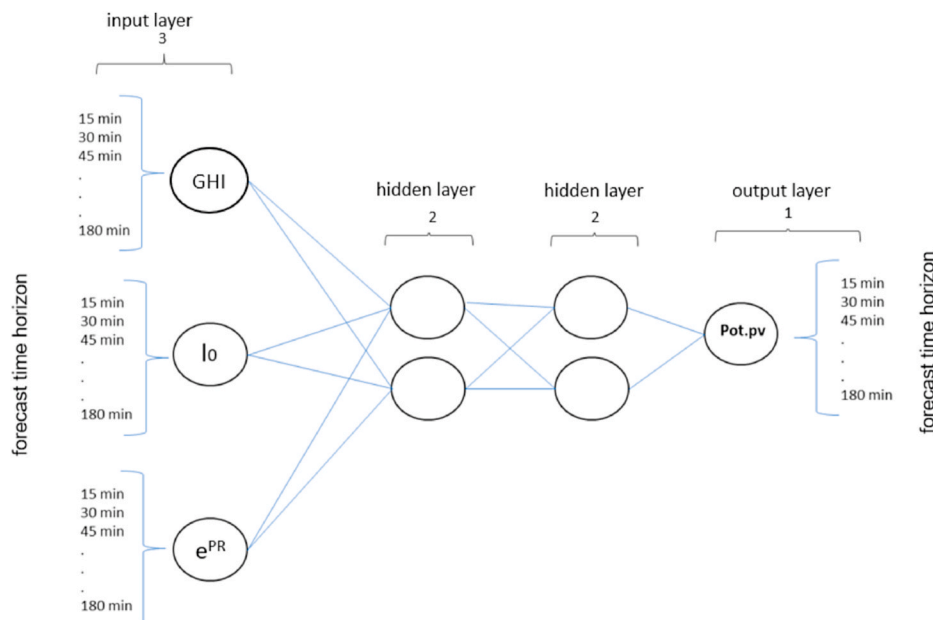

Fig. 3. Flowchart of the methodology for the PV production forecast by ANN technique.

The margin is defined as the maximum width of the region parallel to the hyperplane that has no interior data points. The algorithm can only find this hyperplane in problems that allow linear separation; in most practical problems, the algorithm maximises the flexible margin by allowing a small number of misclassifications. Thus, the SVM hyperplane equation may be defined as follows [57,58].

$$f(x) = w^T \bullet S(x) + b \tag{6}$$

where w^T is the transpose of the weight vector, b is the hyperplane bias and $S(x)$ is the kernel space for the input variables defined by x. As Zendehboudi et al. states in Ref. [58], to determine w, the minimization equation (7) can be used with the restrictions (8) and (9).

$$min\left\{\frac{1}{2}w^2 + c\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)\right\} \tag{7}$$

$$f(x_i) - \left\{w^T \bullet S(x_i) + b\right\} \leq \Psi + \xi_i; \ \ i = 1, 2, \ldots, N \tag{8}$$

$$\xi_i, \xi_i^* \geq 0; \ i = 1, 2, \ldots, N \tag{9}$$

where Ψ is equivalent to the accuracy placed on the training data samples, ξ_i and ξ_i^* are the positive slack variables, and c is an error penalization parameter which balances the fitting in the training stage and generalization in the implementation stage. Too large or too small a c could deteriorate the generalisability of the SVM at the implementation stage [59]. Finally, the equation can be solved by means of the Lagrange multipliers, expressing the solution in its dual form. Then the resulting SVM is called a least-square (LS) SVM and can be represented as follows.

$$f(x) = \sum_{i=1}^{N} \delta_i K(x, x_i) + b \tag{10}$$

where δ_i are the nonnegative Lagrange multiplier of equation (2), and $K(x, x_i)$ is a positive-definite kernel function. In general, the determination of an accurate SVM is highly dependent on the optimal choice of parameters for c, ξ and the kernel [57,58].

3.2.3. Development of ANN and SVM models

For the development of the models, 38.112 data have been used, corresponding to 2/3 of the total set.

3.2.3.1. Selection of the variables for the models.
For each modelling technique, the database was divided into two sets of 70% and 30% for training and validation, respectively. The input and output values of the database were normalized. To normalize the parameters, they were linearly scaled to lie in the range [0, 1], according to the following equation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{11}$$

where x is the parameter considered for normalization, x_{norm} is the normalized parameter, and x_{min} and x_{max} are the minimum and maximum values from the database for each parameter.

For the selection of the variables, the model was run several times, each time with a set of input variables, and the combination with the smallest number of variables that minimized the error was selected. This is a 'stepping forward' pruning method. It's a 'brute force' style of 'pruning', where the human must train the ANN many times to figure out which inputs can be dropped. The performance of models with 14 different combinations of input variables was analyzed. The combination of inputs selected was the one with the lowest number of inputs and the lowest error. Error quantification was performed using the statistical indicators root mean square error (nRMSE) and mean bias error (nMBE), according to the following equations.

$$nRMSE = \frac{100}{\overline{nEac}} \sqrt{\frac{1}{N} \sum (nEac_{est} - nEac_{mea})^2} \tag{12}$$

$$nMBE = \frac{100}{\overline{nEac}} \left[\frac{1}{N} \sum (nEac_{est} - nEac_{mea})\right] \tag{13}$$

Being $nEac_{est}$ and $nEac_{mea}$ PV plant estimated and measured energies, respectively, $\overline{nEac}$ the energy permeated from the Eac_{mea} and N the number of data points. Table 2 shows the results obtained for each of the combinations of variables executed with the two modelling techniques. First, the performance of the models was evaluated with all the available variables, upper part of the table. Subsequently, the errors made when using the possible combinations were seen, eliminating one of the variables, and so on until reaching two variables, as shown in the lower part of the table. To avoid generating multiple simulations when selecting the best combination of variables, fixed parameters were set for the ANN and SVM estimation techniques.

The choice of the selected parameters was made taking into account the number of variables and their errors. The chosen input variables are GHI, I_0, e^{PR}, which fulfils the condition of less variables and less errors.

3.2.3.2. ANN model.
Fig. 6 shows the flowchart used to develop the network models. The developed network consists of a multilayer perceptron with 3:2:2:1 architecture, i.e., three neurons in the input layer, two in the hidden layer and one in the output layer. The input variables to the network are: global irradiance incident on the solar field (GHI), extraterrestrial irradiance (I_0) and the global loss factor, PR. The output variable is the predicted PV power. The network training has been performed with the back-propagation learning method. The network architecture is shown in the figure below.

3.2.3.3. SVM model.
The input variables to the SVM are the same as those used for the development of the ANN. The values of the SVM parameters and their errors are shown in Table 3.

The best fitting parameters for the SVM are Hyperplane Fit Tolerance, C = 10, and fitting parameters $\xi = 0.01$, Radial basis function (RBF).

4. Results

In order to validate the forecasting models for PV production, 19.056 data were used. This database represents the 30% of the total set. Of these, 58% correspond to clear sky conditions, 35% to partly cloudy skies and 7% to overcast skies. This shows that the predominant conditions in the photovoltaic plant are clear skies and the least are overcast skies. Tables 4–6 show the results of the ANN and SVM PV production forecasting models in terms of nRMSE and nMBE. The errors of the solar resource prediction model, GHI, which is an input parameter of the

Table 2
Results of the input variables selection to find the best set to generate the models.

Variables	nRMSE (%)	nMBE (%)
GHI, T, cos θz, sin θz, I_0, PR, e^{PR}	3.31	0.00
T, cos θz, sin θz, I_0, PR, e^{PR}	23.54	0.57
GHI, cos θz, sin θz, I_0, PR, e^{PR}	3.32	0.01
GHI, T, sin θz, I_0, PR, e^{PR}	4.16	0.03
GHI, T, cos θz, I_0, PR, e^{PR}	3.47	0.04
GHI, T, cos θz, sin θz, PR, e^{PR}	3.54	0.00
GHI, T, cos θz, sin θz, I_0, e^{PR}	3.59	0.07
GHI, T, cos θz, sin θz, I_0, PR	3.45	0.00
GHI, T, cos θz, sin θz, I_0	14.55	−0.18
GHI, T, cos θz, e^{PR}	3.71	0.00
GHI, T, PR	6.25	−0.01
GHI, T, e^{PR}	3.91	−0.01
GHI, I_0, e^{PR}	3.70	0.04
GHI, T	31.49	−0.48

Table 3
Values of the SVM parameters and errors.

C	ξ	nRMSE %
10	1	41.5
100	1	41.50
1000	1	41.50
10	0.1	7.093
100	0.1	7.093
1000	0.1	7.093
10	0.01	0.68
100	0.01	0.78
1000	0.01	0.68

production models, are also included. The tables show the differences between the errors of each model in order to better assess the advantages of each methodology over the other. Figs. 4–6 show the nRMSE values

from the validation process for the ANN, SVM and GHI models as a function of the forecast time horizon for the different sky conditions. Finally, a comparison of the results with other models reported in the literature has been included.

4.1. Model validation results under clear sky conditions

Clear sky conditions are those conditions where the sky shows an absence of clouds. These are the ideal operating conditions for solar power plants.

Table 4 shows the results of the model validation and Fig. 4 shows the nRMSE versus the forecast time horizon for clear-sky conditions. In addition, both the figure and the table also show the errors made in the prediction of the solar resource with the sky camera system.

The results of the ANN and the SVM are practically the same for all

Table 4
Results of models' validation for clear sky conditions and different forecast time horizons. The columns ANN-SVM, ANN-GHI and SVM-GHI represent the subtractions of the error indicators of ANN minus SVM, ANN minus GHI and SVM minus GHI, respectively, in order to visualise the best predicting technique. In order to visualise the behaviour of the uncertainty of each model with respect to the forecasting horizon in a simple way, a colour scale has been defined. This colour scale associates a colour to each result according to the rest of the values obtained for the parameter considered.

| Forecast horizon [min]	ANN nRMSE [%]	ANN nMBE [%]	SVM nRMSE [%]	SVM nMBE [%]	GHI nRMSE [%]	GHI nMBE [%]	ANN-SVM nRMSE [%]	ANN-SVM nMBE [%]	ANN-GHI nRMSE [%]	ANN-GHI nMBE [%]	SVM-GHI nRMSE [%]	SVM-GHI nMBE [%]
15	8.09	2.50	8.13	2.84	9.95	4.16	-0.04	-0.34	-1.86	-1.66	-1.82	-1.32
30	7.91	2.44	7.89	2.69	9.38	3.77	0.02	-0.25	-1.47	-1.33	-1.49	-1.08
45	7.71	2.27	7.60	2.48	9.13	3.51	0.11	-0.21	-1.42	-1.24	-1.53	-1.03
60	7.74	2.10	7.62	2.25	9.31	3.47	0.12	-0.15	-1.57	-1.37	-1.69	-1.22
75	7.86	2.00	7.51	2.17	9.59	3.43	0.35	-0.17	-1.73	-1.43	-2.08	-1.26
90	7.88	1.68	7.33	1.85	9.85	3.27	0.55	-0.17	-1.97	-1.59	-2.52	-1.42
105	8.22	1.35	7.85	1.45	10.15	3.12	0.37	-0.10	-1.93	-1.77	-2.30	-1.67
120	9.00	1.35	8.52	1.40	10.55	3.03	0.48	-0.05	-1.55	-1.68	-2.03	-1.63
135	9.55	1.18	8.95	1.20	11.02	2.94	0.60	-0.02	-1.47	-1.76	-2.07	-1.74
150	10.09	1.05	9.47	1.03	11.64	2.94	0.62	-0.02	-1.55	-1.89	-2.17	-1.91
165	10.76	0.75	10.06	0.70	12.29	2.93	0.70	-0.05	-1.53	-2.18	-2.23	-2.23
180	11.63	0.58	10.76	0.49	12.93	2.94	0.87	0.09	-1.30	-2.36	-2.17	-2.45

Table 5
Results of models' validation for partially cloudy sky conditions and different forecast time horizons. The columns ANN-SVM, ANN-GHI and SVM-GHI represent the subtractions of the error indicators of ANN minus SVM, ANN minus GHI and SVM minus GHI, respectively, in order to visualise the best predicting technique. In order to visualise the behaviour of the uncertainty of each model with respect to the forecasting horizon in a simple way, a colour scale has been defined. This colour scale associates a colour to each result according to the rest of the values obtained for the parameter considered.

| Forecast horizon [min]	ANN nRMSE [%]	ANN nMBE [%]	SVM nRMSE [%]	SVM nMBE [%]	GHI nRMSE [%]	GHI nMBE [%]	ANN-SVM nRMSE [%]	ANN-SVM nMBE [%]	ANN-GHI nRMSE [%]	ANN-GHI nMBE [%]	SVM-GHI nRMSE [%]	SVM-GHI nMBE [%]
15	16.63	2.14	17.05	2.58	15.52	4.68	-0.42	-0.44	1.11	-2.54	1.53	-2.10
30	15.35	2.06	15.99	2.39	14.46	4.76	-0.64	-0.33	0.89	-2.70	1.53	-2.37
45	15.96	1.44	16.55	1.68	14.04	4.41	-0.59	-0.24	1.92	-2.97	2.51	-2.73
60	15.06	1.60	15.74	1.75	14.50	4.39	-0.68	-0.15	0.56	-2.79	1.24	-2.64
75	15.74	1.20	16.43	1.30	15.23	3.92	-0.69	-0.10	0.51	-2.72	1.20	-2.62
90	16.18	0.94	16.76	1.02	16.11	3.99	-0.58	-0.08	0.07	-3.05	0.65	-2.97
105	16.17	0.76	16.50	0.84	15.99	3.81	-0.33	-0.08	0.18	-3.05	0.51	-2.97
120	16.92	0.57	17.17	0.59	16.68	3.54	-0.25	-0.02	0.24	-2.97	0.49	-2.95
135	18.00	0.08	18.05	0.13	17.35	3.36	-0.05	-0.05	0.65	-3.28	0.70	-3.23
150	19.78	-0.86	19.86	-0.94	17.98	2.91	-0.08	-0.08	1.80	-2.05	1.88	-3.85
165	19.92	-0.79	20.11	-0.88	18.19	2.61	-0.19	-0.09	1.73	-1.82	1.92	-3.49
180	20.93	-0.82	20.44	-0.81	19.38	2.53	-0.01	-0.01	1.55	-1.71	1.06	-3.34

Table 6

Results of models' validation for all sky conditions and different forecast time horizons. The columns ANN-SVM, ANN-GHI and SVM-GHI represent the subtractions of the error indicators of ANN minus SVM, ANN minus GHI and SVM minus GHI, respectively, in order to visualise the best predicting technique. In order to visualise the behaviour of the uncertainty of each model with respect to the forecasting horizon in a simple way, a colour scale has been defined. This colour scale associates a colour to each result according to the rest of the values obtained for the parameter considered.

Forecast horizon [min]	ANN		SVM		GHI		ANN-SVM		ANN- G		SVM- G	
	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]	nRMSE [%]	nMBE [%]
15	19.38	1.48	19.46	2.11	17.92	3.85	-0.08	-0.63	1.46	-2.37	1.54	-1.74
30	17.89	1.82	18.02	2.36	16.59	4.09	-0.13	-0.54	1.30	-2.27	1.43	-1.73
45	17.42	2.20	17.65	2.61	15.89	4.20	-0.23	-0.41	1.53	-2.00	1.76	-1.59
60	17.10	2.12	17.29	2.49	16.31	4.24	-0.19	-0.37	0.79	-2.12	0.98	-1.75
75	17.33	1.97	17.42	2.28	16.82	4.18	-0.09	-0.31	0.51	-2.21	0.60	-1.90
90	17.54	1.85	17.71	2.12	16.71	4.25	-0.17	-0.27	0.83	-2.40	1.00	-2.13
105	17.95	2.05	17.98	2.30	17.52	4.24	-0.03	-0.25	0.43	-2.19	0.46	-1.94
120	18.44	1.59	18.48	1.76	17.89	4.07	-0.04	-0.17	0.55	-2.48	0.59	-2.31
135	19.06	1.32	18.98	1.49	17.73	3.75	0.08	-0.17	1.33	-2.43	1.25	-2.26
150	19.54	1.11	19.53	1.23	18.42	3.70	0.01	-0.12	1.12	-2.59	1.11	-2.47
165	20.28	0.80	20.26	0.93	18.91	3.61	0.02	-0.13	1.37	-2.81	1.35	-2.68
180	21.43	0.61	21.12	0.76	19.57	3.59	0.31	-0.15	1.86	-2.99	1.55	-2.83
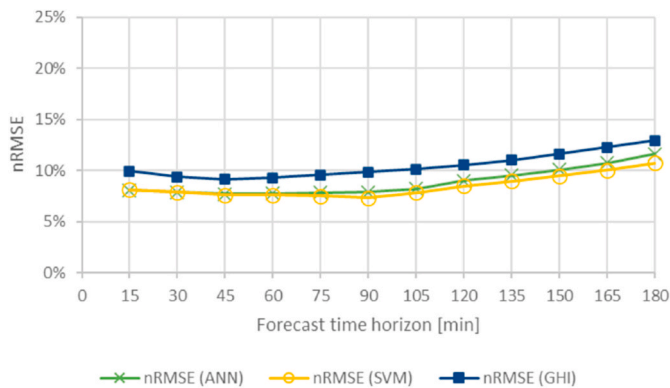
Fig. 4. nRMSE values from the validation process for the ANN (green) and SVM (yellow) models, and solar resource forecasting provided by the sky camera-based model (blue) as a function of the forecast time horizon for clear sky conditions.
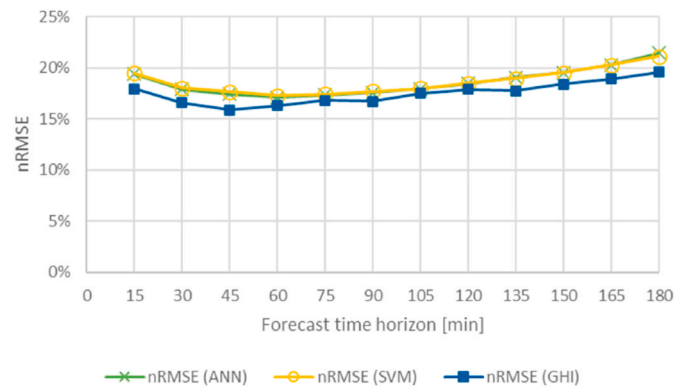
Fig. 6. nRMSE values from the validation process for the ANN (green) and SVM (yellow) models, and solar resource forecasting provided by the sky camera-based model (blue) as a function of the forecast time horizon for all sky conditions.
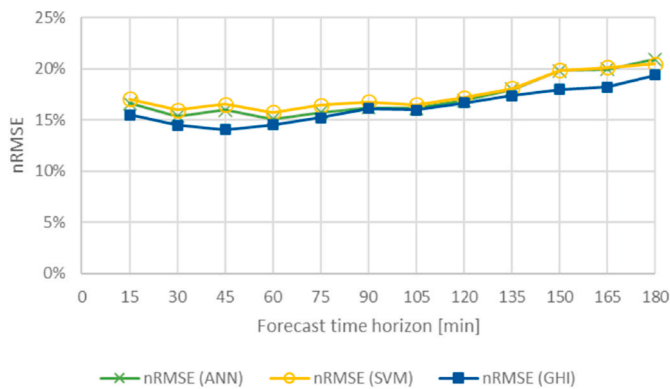
Fig. 5. nRMSE values from the validation process for the ANN (green) and SVM (yellow) models, and solar resource forecasting provided by the sky camera-based model (blue) as a function of the forecast time horizon for partly cloudy sky conditions.

forecast time horizons considered. As shown in Table 4, the SVM model shows a slight improvement over the ANN model in terms of nRMSE and nMBE for long-range forecast time horizons. The average value of the difference between ANN and SVM is 0.4% and 0.13% for nRMSE and nMBE, respectively. The nRMSE value of the ANN, between 15 and 120 min, is always less than 9%. Starting at min 120, the nRMSE increases from 9% to 12% for min 180.

Fig. 4 shows that the error of the models increases as the forecast time horizon increases, as does the uncertainty of the GHI. On the one hand, the fact that the behavior and magnitude of the production uncertainty is similar to that of the solar resource indicates that the solar resource is the main source of uncertainty in the production models. As observed in the results, the production models improve the nRMSE values with respect to the GHI prediction values by approximately 2%, see Table 4 and Fig. 4. On the other hand, the increase in the uncertainty of the models as the forecast horizon considered increases is a behavior that was reported in Ref. [49]. As reported in Ref. [49], the viewing range of sky cameras is between 20 and 30 km. In general terms, this fact limits the forecasting capabilities of GHI beyond the 60 min time horizon. For longer time horizons, predictions with methods based on satellite images improve with respect to systems based on sky cameras.

4.2. Model validation results under partly cloudy sky conditions

Partly cloudy conditions occur when clouds are present but do not cover the sky completely. In these sky conditions there are usually transient phenomena in which clear skies and clouds are interspersed over the photovoltaic plant, causing variations in the solar resource and, therefore, in productivity. In other words, the photovoltaic power varies from low values (presence of clouds) to high values (absence of clouds). These variations in the productivity of PV plants are a difficult to predict. This problem puts the quality of the grid power supply at risk and makes the operation of both the plant and the grid more complex. The present paper aims to provide a reliable tool to assist in the PV production prediction to facilitate operational tasks under all sky conditions.

Table 5 and Fig. 5 show the results obtained in the validation stage of the photovoltaic energy production forecast based on the time horizon of the forecast. It is observed that the uncertainty has increased with respect to the results obtained for clear sky conditions by approximately 6 points. This increase in error is due to the uncertainty introduced in the GHI prediction by the presence of clouds, which, as previously mentioned, is the main source of uncertainty in the models. For one thing, cloud cover can consist of different layers at different heights and only the bottom layer is visible to the sky camera, increasing the uncertainty in the GHI prediction. On the other hand, cloud thickness, cloud height, cloud density, cloud speed, and cloud formation or dispersion are factors that make GHI difficult to predict. Therefore, the behavior of the uncertainty in the GHI prediction continues to determine the behavior of the error in photovoltaic production. Again, it is observed that the limitations of the field of view of the sky camera increase the uncertainty from the forecast time horizon of 60 min onwards.

The ANN-SVM column of Table 5 shows that for any forecast time horizon, the SVMs have a slightly higher nRMSE and nMBE, with the average value of this difference for all forecast time horizons being of the order of 0.40% and 0.14%, respectively. In the first 120 min of the forecast time horizon, the nRMSE remains practically constant at around 16%, and presents a slight overestimation that varies from 2% to 0.6% from minute 15 to 120. From minute 135 to 180, the nRMSE increases from 18% to 21%, and the nMBE error is not significant.

The ANNs introduce a slight error in the output by using the predicted GHI as an input variable to the networks, as can be seen in the ANN-global irradiance column. The average nRMSE value for any forecast time horizon is of the order of 0.85. In contrast, the nMBE of the global irradiance is higher than the error committed by the ANN. The mean value for the entire forecast time horizon is of the order of 2.64%.

4.3. Model validation results under all sky conditions

In this section, the results of the validation of the models are analyzed under all sky conditions, whether they are: clear sky, partly cloudy or overcast. Table 6 and Fig. 6 show the results of the error made by the different forecast models.

For any forecast time horizon, the results provided by the ANNs are practically the same as those of the SVMs, the difference in their errors is negligible. The average value of nRMSE in the first 120 min of the forecast time horizon is of the order of 18% and from 135 to 180 min the error increases to 20%. For the entire forecast interval, the nRMSE of the ANNs and SVMs is less than 2% higher than the nRMSE of the global irradiance prediction.

This fact makes the model more sensitive to GHI uncertainty. As mentioned above, the GHI forecast is the main source of uncertainty in the prediction of PV production.

As previously mentioned, the small difference between the nRMSE of the production forecast and the solar resource forecast indicates that the main source of uncertainty in the models is the solar resource. Therefore, the behaviour of the GHI prediction error with the sky camera system

conditions the uncertainty of the PV production estimation. This is expected behaviour. In the PV production prediction model, the solar resource is the most variable and difficult to predict input parameter, as opposed to the PR and I_0. The low error in production forecast compared to that provided by the solar resource is thanks to considering the PR within the production models, which is an indicator of the operational status of the PV plant and its actual performance.

4.4. Comparison with other methodologies found in the literature

A literature review of PV production prediction models has been carried out in order to compare the results with those obtained in this paper. In general terms, the literature review has shown that there is no consensus on the use of parameters for the validation of the models. Table 7 shows a selection of models whose uncertainty indicator is the RMSE. The table shows the list of models where the supervised learning methodology used, their prediction horizon and their uncertainty in terms of nRMSE are specified. As can be seen, models developed with different learning techniques and different prediction horizons have been selected, covering horizons of less than 24 h.

In general, the results of the validation phases depended on the size of the selected databases, the operational state of the PV plant and the local meteorological sky conditions. The lowest error of the models presented in Table 7 corresponds to the Bouzerdoum et al. model [66]. This model for PV production prediction was developed based on the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Support Vector Machine (SVM) methods. The inputs used were irradiance on the array plane, modules temperature, ambient temperature at array side, string voltage, current, power and AC power. It does not have information provided by sky cameras, which limits nowcasting in case of partly cloudy or cloudy days. For the validation of the model, data from clear days were used. In these cases, the nRMSE value reported was 9.40%, while the RMSE values for clear days of the models proposed in this paper (see Table 4) were 7.74% and 7.62% for ANN and SVM, respectively, considering forecast horizons of 60 min.

The model developed by Hossain et al. [63] was based on the Extreme Learning Machine (ELM) technique and uses daily average solar irradiance, ambient temperature, module temperature, wind speed, PV output data as input data. As it does not use images from sky cameras, it makes a larger error in its short-term predictions (see Table 7).

Qing et al. [64] developed a model based on Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) Networks. As inputs they used solar irradiance, temperature, humidity, cloud cover, numerical weather prediction (NWP). They did not use a sky camera. The results of the validation showed an error in the prediction of the production of 18.64%.

Table 7

Table with forecasting models found in the literature summarising the methodology used, the model output, the forecast horizon and its uncertainty [60, 61].

Ref.	Methodology	Forecast horizon	nRMSE
[62]	Deep Convolutional Neural Network (DCNN)	15-min	14.09%
[63]	Extreme Learning Machine (ELM)	Day	[17.89, 35.39%]
		1h	[54.96, 90.1%]
[64]	Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) Networks	1h	18.64%
[65]	Historical Similar Mining (Hisimi) Model	1–24-h	10.14%
[66]	Seasonal Autoregressive Integrated Moving Average (SARIMA) And Support Vector Machine (SVM)	Hourly	9.40%
[67]	Back-Propagation Neural Network (BPNN)	4 h	19.02%
[68]	Long Short-Term Memory (LSTM) Networks	1 min	21%

The model developed by Monteiro et al. [65] uses of input variables historical values of variables corresponding to the PV plants. They developed the mesoscale numerical weather prediction (NWP) to build a database of meteorological parameters. In addition, a transition model was developed using previous energy production data. The simulation results show that the model uncertainty is 10.14% for a forecast 1–24 h ahead. They do not use information from sky cameras or satellites to predict the solar resource.

Venugopal et al. [62] established a model based on a convolutional neural network (CNN) to predict the output power of photovoltaic panels in the next 15 min. In addition, the input data were the photovoltaic production and images of the terrestrial sky in the last 15 min. This photovoltaic production prediction model achieved uncertainty levels of 14.09% for all sky conditions.

As can be seen in Table 7, the results obtained in the PV production prediction models are similar on average to those reported in this paper, 17% nRMSE for both ANN and SVM. However, in this case, the error is 1% higher than the GHI prediction error with the sky cam system, considering a prediction horizon of 60 min. This is achieved thanks to include the performance ratio as input parameter, which refers to the operating conditions of the plant, e.g., its soiling levels. Therefore, the main achievement of the models presented in the present paper is to predict the photovoltaic production in an error of 1% with respect to the errors yielded by the sky camera solar resource prediction methodology. The GHI sky camera prediction methodology used in this paper is an extensively validated methodology that provides better results than those based on satellite imagery for prediction horizons shorter than 60 min, see Refs. [40,41,48,50–52].This fact makes these models a fundamental tool to facilitate the tasks of PV plant and grid operators to predict and mitigate production ramps due to cloud transients, which are difficult to anticipate, in time and magnitude, for time horizons longer than several hours.

The databases used for the generation and validation of the models found in the literature were created from the measurements recorded around PV plants located in different parts of the globe. The different PV plant technologies employed, with their different operating and cleanliness states, the different local meteorological conditions, and the different sizes of databases, complicate the comparison of the uncertainties of the models reported by the authors (see Table 7). In conclusion, it is difficult to compare the uncertainty results with those models reported in the literature. They often use different parameters to indicate the uncertainty and are validated with databases that differ in the conditions under which they were generated. Therefore, it is necessary to make future efforts to reach a consensus both on the parameters to define the levels of uncertainty of the models, and to have reliable and complete databases with which to replicate the reported models and validate them.

5. Conclusions

In this paper, a methodology has been developed to forecast the production of photovoltaic plants for forecast time horizons from 15 to 180 min, within 15 min intervals. Digital levels of image sequences taken with sky cameras have been used to forecast global solar irradiance with a methodology previously developed and widely reported in the literature. Two models have been developed to forecast the production of the photovoltaic plant, one based on Artificial Neural Networks and the other on Support Vector Machines. The input variables of both models were the global solar irradiance predicted by the sky camera-based methodology, the extraterrestrial solar irradiance and the plant performance ratio. The output variable was the power produced by the photovoltaic plant. For the development of the models, the database of the photovoltaic plant located in the CIESOL building of the University of Almería, Spain, was used. This plant has an excellent location to produce photovoltaic electricity, with 56% of the total data set corresponding to clear days and 35% partly cloudy days. The database was

separated into two groups: 2/3 of the database was used for model development and 1/3 of the database was used for model validation.

The results of the validation process were similar for both models, ANN and SVM, for all sky conditions and for the entire forecast interval considered, from 15 to 180 min. In general, during the first 2 h of the forecast interval, the nRMSE and nMBE values were 18% and 2%, respectively, for both models of PV production. For clear sky conditions, the nRMSE and nMBE values obtained were approximately 8% and 2% during the first 2 h of production forecast, with the mean value of the error difference between ANN and SVM being of the order of 0.4% and 0.13% for nRMSE and nMBE. For partly cloudy sky conditions, the PV production forecasts show an nRMSE of around 16% and an nMBE of 1%, keeping the difference between both models at similar values.

In general, it was observed that the nRMSE value of both production models differs by less than 1% from the value obtained for the global solar irradiance forecast for a forecast interval of 60 min. This is achieved thanks to the inclusion of the performance ratio as an input parameter, which refers to the operating conditions of the plant. Therefore, the main achievement of the models presented in this paper is to predict the photovoltaic production with an error of 1% with respect to the errors produced by the sky camera solar resource prediction methodology, a widely validated methodology that provides better results than those based on satellite imagery for forecast horizons of less than 60 min.

The validation values of both models are in agreement with the validation results of other models reported in the literature, around 17% of nRMSE. For clear sky conditions, the models developed for this work improve the nRSME by almost 2% over those reported in the literature. However, care needs to be taken when comparing the results with other published works. The bibliographic review questioned the possibilities of comparison between models given that the conditions in which the databases were generated are different, in terms of photovoltaic technology, state of maintenance of the plants, local climatic conditions, inputs parameters, and database sizes, among others.

The results obtained make the models developed in this paper a fundamental tool to facilitate the tasks of photovoltaic plant and network operators to predict and mitigate production ramps due to cloud transients, which are difficult to anticipate, in time and magnitude. Users of self-consumption systems with net balance will be able to know 3 h in advance the amount of energy produced that will be injected into the grid or consumed from the grid. If the self-consumption is without surplus, it will allow them to know the amount of energy they can introduce or extract from the batteries so as not to have consumption problems due to cloud transit phenomena.

CRediT authorship contribution statement

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.

Acknowledgements

This paper is dedicated to the memory of our dear colleague and friend Prof. Francisco Javier Batlles Garrido (1959–2022).

References

[1] A. Marzo, A. Salmon, J. Polo, J. Ballestrín, G. Soto, G. Quiñones, J. Alonso-Montesinos, E. Carra, M. Ibarra, J. Cardemil, E. Fuentealba, R. Escobar, Solar extinction map in Chile for applications in solar power tower plants, comparison with other places from sunbelt and impact on LCOE, Renew. Energy 170 (2021) 197–211, https://doi.org/10.1016/j.renene.2021.01.126.

[2] H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama, IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), 2022. https://www.ipcc.ch/report/ar6/wg2/. (Accessed 22 June 2022).

[3] IRENA, RENEWABLE CAPACITY STATISTICS 2021 International Renewable Energy Agency (IRENA), Abu Dhabi, 2021. www.irena.org.

[4] IEA, Trends in photovoltaic applications task 1 strategic PV analysis and Outreach. International energy agency (IEA). Report IEA-PVPS T1-41:2021, 2021. www.iea-pvps.org.

[5] C.J. Sarasa-Maestro, R. Dufo-López, J.L. Bernal-Agustín, M. Sasso, C. Roselli, Analysis of Photovoltaic Self-Consumption Systems, (n.d.). https://doi.org/10.3390/en9090681.

[6] A. Ciocia, A. Amato, P. di Leo, S. Fichera, G. Malgaroli, F. Spertino, S. Tzanova, Self-consumption and self-sufficiency in photovoltaic systems: effect of grid limitation and storage installation, Energies 14 (2021), https://doi.org/10.3390/en14061591.

[7] M. Kumar, A. Kumar, Performance assessment and degradation analysis of solar photovoltaic technologies: a review, Renew. Sustain. Energy Rev. 78 (2017) 554–587, https://doi.org/10.1016/J.RSER.2017.04.083.

[8] M. Santhakumari, N. Sagar, A review of the environmental factors degrading the performance of silicon wafer-based photovoltaic modules: failure detection methods and essential mitigation techniques, Renew. Sustain. Energy Rev. 110 (2019) 83–100, https://doi.org/10.1016/J.RSER.2019.04.024.

[9] M.M. Fouad, L.A. Shihata, E.S.I. Morgan, An integrated review of factors influencing the perfomance of photovoltaic panels, Renew. Sustain. Energy Rev. 80 (2017) 1499–1511, https://doi.org/10.1016/J.RSER.2017.05.141.

[10] E. Fuentealba, P. Ferrada, F. Araya, A. Marzo, C. Parrado, C. Portillo, Photovoltaic performance and LCoE comparison at the coastal zone of the Atacama Desert, Chile, Energy Convers. Manag. 95 (2015) 181–186, https://doi.org/10.1016/j.enconman.2015.02.036.

[11] J.S. Stein, B. Farnung, PV Performance Modeling Methods and Practices, EEUU, Albuquerque, New Mexico, 2017.

[12] G. López, F.J. Batlles, J. Tovar-Pescador, Selection of input parameters to model direct solar irradiance by using artificial neural networks, Energy 30 (2005) 1675–1684, https://doi.org/10.1016/J.ENERGY.2004.04.035.

[13] G. López, M.A. Rubio, M. Martínez, F.J. Batlles, Estimation of hourly global photosynthetically active radiation using artificial neural network models, Agric. For. Meteorol. 107 (2001) 279–291, https://doi.org/10.1016/S0168-1923(01)00217-9.

[14] J.L. Bosch, G. López, F.J. Batlles, Daily solar irradiation estimation over a mountainous area using artificial neural networks, Renew. Energy 33 (2008) 1622–1628, https://doi.org/10.1016/J.RENENE.2007.09.012.

[15] A. Marzo, M. Trigo, J. Alonso-Montesinos, M. Martínez-Durbán, G. López, P. Ferrada, E. Fuentealba, M. Cortés, F.J. Batlles, Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation, Renew. Energy 113 (2017) 303–311, https://doi.org/10.1016/j.renene.2017.01.061.

[16] S. Rosiek, F.J. Batlles, Modelling a solar-assisted air-conditioning system installed in CIESOL building using an artificial neural network, Renew. Energy 35 (2010) 2894–2901, https://doi.org/10.1016/J.RENENE.2010.04.018.

[17] G. López, S.M. Sarmiento-rosales, C.A. Gueymard, A. Marzo, Effect of cloudiness on solar radiation forecasting, in: ISES Solar World Congress SWC/SHC 2019, International Solar Energy Society, 2019, pp. 2098–2108, https://doi.org/10.18086/swc.2019.43.05.

[18] J.L. Gómez, A.O. Martínez, F.T. Pastoriza, L.F. Garrido, E.G. Álvarez, J.A.O. García, Photovoltaic power prediction using artificial neural networks and numerical weather data, Sustainability 12 (2020) 10295, https://doi.org/10.3390/SU122410295, 12 (2020) 10295.

[19] L. Ramos, M. Colnago, W. Casaca, Data-driven analysis and machine learning for energy prediction in distributed photovoltaic generation plants: a case study in Queensland, Australia, Energy Rep. 8 (2022) 745–751, https://doi.org/10.1016/j.egyr.2021.11.123.

[20] B. Zazoum, Solar photovoltaic power prediction using different machine learning methods, Energy Rep. 8 (2022) 19–25, https://doi.org/10.1016/j.egyr.2021.11.183.

[21] M. Abuella, B. Chowdhury, Solar power forecasting using support vector regression, in: 2016 International Annual Conference of the American Society for Engineering Management, ASEM 2016, 2016.

[22] C.F. Yen, H.Y. Hsieh, K.W. Su, M.C. Yu, J.S. Leu, Solar power prediction via support vector machine and random forest, in: E3S Web of Conferences, vol. 69, 2018, https://doi.org/10.1051/e3sconf/20186901004.

[23] M. Trigo-Gonzalez, M. Cortés, J. Alonso-Montesinos, M. Martínez-Durbán, P. Ferrada, J. Rabanal, C. Portillo, G. López, F.J. Batlles, Development and comparison of PV production estimation models for mc-Si technologies in Chile and Spain, J. Clean. Prod. 281 (2021), 125360, https://doi.org/10.1016/j.jclepro.2020.125360.

[24] A. Zendehboudi, M.A. Baseer, R. Saidur, Application of support vector machine models for forecasting solar and wind energy resources: a review, J. Clean. Prod. 199 (2018) 272–285, https://doi.org/10.1016/j.jclepro.2018.07.164.

[25] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, F. Qin, Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization, J. Clean. Prod. 277 (2020), 123948, https://doi.org/10.1016/j.jclepro.2020.123948.

[26] J.M. Álvarez-Alvarado, J.G. Ríos-Moreno, S.A. Obregón-Biosca, G. Ronquillo-Lomelí, E. Ventura-Ramos, M. Trejo-Perea, Hybrid techniques to predict solar radiation using support vector machine and search optimization algorithms: a review, Appl. Sci. 11 (2021) 1–17, https://doi.org/10.3390/app11031044.

[27] B. Wang, J. Che, B. Wang, S. Feng, A solar power prediction using support vector machines based on multi-source data fusion, in: 2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings, 2019, pp. 4573–4577, https://doi.org/10.1109/POWERCON.2018.8601672.

[28] A.T. Eseye, J. Zhang, D. Zheng, Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information, Renew. Energy 118 (2018) 357–367, https://doi.org/10.1016/j.renene.2017.11.011.

[29] S. Preda, S.V. Oprea, A. Bâra, A. Belciu, PV forecasting using support vector machine learning in a big data analytics context, Symmetry (Basel) 10 (2018), https://doi.org/10.3390/sym10120748.

[30] J. Shi, W.J. Lee, Y. Liu, Y. Yang, P. Wang, Forecasting power output of photovoltaic systems based on weather classification and support vector machines, IEEE Trans. Ind. Appl. 48 (2012) 1064–1069, https://doi.org/10.1109/TIA.2012.2190816.

[31] A. Catalina, A. Torres-Barrán, C.M. Alaíz, J.R. Dorronsoro, Machine learning nowcasting of PV energy using satellite data, Neural Process. Lett. 52 (2020) 97–115, https://doi.org/10.1007/s11063-018-09969-1.

[32] H. Escrig, F.J. Batlles, J. Alonso, F.M. Baena, J.L. Bosch, I.B. Salbidegoitia, J.I. Burgaleta, Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast, Energy 55 (2013) 853–859, https://doi.org/10.1016/J.ENERGY.2013.01.054.

[33] J. Alonso-Montesinos, M. Martínez-Durbán, J. del Sagrado, I.M. del Águila, F.J. Batlles, The application of Bayesian network classifiers to cloud classification in satellite images, Renew. Energy 97 (2016) 155–161, https://doi.org/10.1016/J.RENENE.2016.05.066.

[34] F.J. Batlles, J. Alonso, G. López, Cloud cover forecasting from METEOSAT data, Energy Proc. 57 (2014) 1317–1326, https://doi.org/10.1016/J.EGYPRO.2014.10.122.

[35] J. Alonso-Montesinos, F.J. Batlles, J.L. Bosch, Beam, diffuse and global solar irradiance estimation with satellite imagery, Energy Convers. Manag. 105 (2015) 1205–1212, https://doi.org/10.1016/J.ENCONMAN.2015.08.037.

[36] J. Alonso-Montesinos, F.J. Batlles, Solar radiation forecasting in the short- and medium-term under all sky conditions, Energy 83 (2015) 387–393, https://doi.org/10.1016/J.ENERGY.2015.02.036.

[37] F.J. Olmo, A. Cazorla, L. Alados-Arboledas, M.A. López-Álvarez, J. Hernández-Andrés, J. Romero, Retrieval of the optical depth using an all-sky CCD camera, Appl. Opt. 47 (2008) H182–H189, https://doi.org/10.1364/AO.47.00H182. Vol. 47, Issue 34, Pp. H182-H189.

[38] G. Lopez, A. Martin, J. Alonso-Montesinos, F.J. Batlles, J. Barbero, Identification of cloud contamination in sun photometric data using ground-based sky imagery, J. Aeronaut. Aero. Eng. (2016), https://doi.org/10.4172/2168-9792.C1.014, 05.

[39] M. Martínez-Chico, F.J.J. Batlles, J.L.L. Bosch, Cloud classification in a mediterranean location using radiation data and sky images, Energy 36 (2011) 4055–4062, https://doi.org/10.1016/j.energy.2011.04.043.

[40] J. Alonso, F.J. Batlles, G. López, A. Ternero, Sky camera imagery processing based on a sky classification using radiometric data, Energy 68 (2014) 599–608, https://doi.org/10.1016/j.energy.2014.02.035.

[41] J. Alonso, A. Ternero, F.J. Batlles, G. López, J. Rodríguez, J.I. Burgaleta, Prediction of cloudiness in short time periods using techniques of remote sensing and image processing, in: Proceedings of the SolarPACES 2013 International Conference, vol. 49, 2014, pp. 2280–2289, https://doi.org/10.1016/j.egypro.2014.03.241.

[42] J. Alonso-Montesinos, F.J. Batlles, The use of a sky camera for solar radiation estimation based on digital image processing, Energy 90 (2015) 377–386, https://doi.org/10.1016/j.energy.2015.07.028.

[43] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. 11 (2007) 1633–1644, https://doi.org/10.5194/hess-11-1633-2007.

[44] Solar Radiation Measurement, Kipp, Zonen, n.d, https://www.kippzonen.com/. (Accessed 29 June 2022).

[45] C.N. Long, E.G. Dutton, BSRN Global Network Recommended QC Tests, V2.X, EPIC3Bremerhaven, PANGAEA, 2010. https://epic.awi.de/id/eprint/30083/. (Accessed 11 August 2019).

[46] V. Oyj, USER GUIDE vaisala HUMICAP ® humidity and temperature transmitter series HMT360, 2021. www.vaisala.com. (Accessed 29 June 2022).

[47] J. Kyrouac, A. Theisen, Biases of the MET Temperature and Relative Humidity Sensor (HMP45) Report, 2018.

[48] J. Alonso-Montesinos, R. Monterreal, J. Fernandez-Reche, J. Ballestrín, G. López, J. Polo, F.J. Barbero, A. Marzo, C. Portillo, F.J. Batlles, Nowcasting system based on sky camera images to predict the solar flux on the receiver of a concentrated solar plant, Rem. Sens. 14 (2022) 1602, https://doi.org/10.3390/RS14071602, 14 (2022) 1602.

[49] Duffie, Beckman, Solar Engineering of Thermal Processes, third ed., 2006 n.d.).

[50] J. Alonso, F.J. Batlles, C. Villarroel, R. Ayala, J.I. Burgaleta, Determination of the sun area in sky camera images using radiometric data, Energy Convers. Manag. 78 (2014) 24–31, https://doi.org/10.1016/j.enconman.2013.10.050.

[51] J. Alonso, F.J.J. Batlles, Short and medium-term cloudiness forecasting using remote sensing techniques and sky camera imagery, Energy 73 (2014) 890–897, https://doi.org/10.1016/j.energy.2014.06.101.

[52] J. Alonso-Montesinos, F.J. Batlles, C. Portillo, Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images, Energy Convers. Manag. 105 (2015) 1166–1177, https://doi.org/10.1016/j.enconman.2015.09.001.

[53] D.E. Rumelhart, J.L. McClelland, Learning internal representations by error propagation - MIT press books, parallel distributed processing: explorations in the microstructure of cognition: foundations, 1987, 318–362, https://ieeexplore.ieee.org/document/6302929. (Accessed 22 July 2022).

[54] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Inc., USA, 1995.

[55] M. Trigo-González, F.J. Batlles, J. Alonso-Montesinos, P. Ferrada, J. del Sagrado, M. Martínez-Durbán, M. Cortés, C. Portillo, A. Marzo, Hourly PV production

estimation by means of an exportable multiple linear regression model, Renew. Energy 135 (2019) 303–312, https://doi.org/10.1016/j.renene.2018.12.014.

[56] C. Cortes, V. Vapnik, Support-vector networks editor, Machine Leaming 20 (1995) 273–297. https://link.springer.com/content/pdf/10.1007/BF00994018.pdf. (Accessed 13 January 2022).

[57] M. Trigo-Gonzalez, M. Cortés, J. Alonso-Montesinos, M. Martínez-Durbán, P. Ferrada, J. Rabanal, C. Portillo, G. López, F.J. Batlles, Development and comparison of PV production estimation models for mc-Si technologies in Chile and Spain, J. Clean. Prod. 281 (2021), 125360, https://doi.org/10.1016/J.JCLEPRO.2020.125360.

[58] A. Zendehboudi, M.A. Baseer, R. Saidur, Application of support vector machine models for forecasting solar and wind energy resources: a review, J. Clean. Prod. 199 (2018) 272–285, https://doi.org/10.1016/J.JCLEPRO.2018.07.164.

[59] J. Zeng, W. Qiao, Short-term solar power prediction using a support vector machine, Renew. Energy 52 (2013) 118–127, https://doi.org/10.1016/J.RENENE.2012.10.009.

[60] S. Sobri, S. Koohi-Kamali, N.A. Rahim, Solar photovoltaic generation forecasting methods: a review, Energy Convers. Manag. 156 (2018) 459–497, https://doi.org/10.1016/J.ENCONMAN.2017.11.019.

[61] C. Liu, M. Li, Y. Yu, Z. Wu, H. Gong, F. Cheng, A review of multitemporal and multispatial scales photovoltaic forecasting methods, IEEE Access 10 (2022) 35073–35093, https://doi.org/10.1109/ACCESS.2022.3162206.

[62] V. Venugopal, Y. Sun, A.R. Brandt, Short-term solar PV forecasting using computer vision: the search for optimal CNN architectures for incorporating sky images and PV generation history, J. Renew. Sustain. Energy 11 (2019), 066102, https://doi.org/10.1063/1.5122796.

[63] M. Hossain, S. Mekhilef, M. Danesh, L. Olatomiwa, S. Shamshirband, Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems, 2017, https://doi.org/10.1016/j.jclepro.2017.08.081.

[64] X. Qing, Y. Niu, Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM, Energy 148 (2018) 461–468, https://doi.org/10.1016/j.energy.2018.01.177.

[65] C. Monteiro, T. Santos, A. Fernandez-Jimenez, I.J. Ramirez-Rosado, M. Sonia Terreros-Olarte, Short-term power forecasting model for photovoltaic plants, Based on Historical Similarity 6 (2013) 2624–2643, https://doi.org/10.3390/en6052624.

[66] M. Bouzerdoum, A. Mellit, A.M. Pavan, A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant, 2013, https://doi.org/10.1016/j.solener.2013.10.002.

[67] L. Zhang, Q. Bai, Y. Zhi, S. Zhen, K. Cang, B. Liang, Research on distributed photovoltaic ultra-short-term prediction technology based on spatial correlation, J. Shanxi Electr. Power. 5 (2017) 22–26.

[68] J. Zhang, R. Verschae, S. Nobuhara, J.F. Lalonde, Deep photovoltaic nowcasting, Sol. Energy 176 (2018) 267–276, https://doi.org/10.1016/J.SOLENER.2018.10.024.