IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 5, 2020, accepted October 22, 2020, date of publication November 5, 2020, date of current version November 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3036140

A Day-Ahead Irradiance Forecasting Strategy for the Integration of Photovoltaic Systems in Virtual Power Plants

GUILLERMO MORENO, PEDRO MARTIN, CARLOS SANTOS, FRANCISCO J. RODRÍGUEZ, (Member, IEEE), AND ENRIQUE SANTISO
Department of Electronics, University of Alcala, 28805 Alcala de Henares, Spain

Corresponding author: Carlos Santos (carlos.santos@uah.es)

This work was supported in part by the PROMINT-CM Project, funded by the Comunidad de Madrid and the European Social Fund under Grant S2018/EMT-4366, in part by the HELIOS Project, funded by the Ministerio de Ciencia, Innovación y Universidades under Grant RTC-2017-6231-3, and in part by the Ministerio de Ciencia, Innovación y Universidades through the INERCIA0 Project under Grant RTI2018-098865-B-C33.

ABSTRACT Encouraged by the considerable cost reduction, small-scale solar power deployment has become a reality during the last decade. However, grid integration of small-scale photovoltaic (PV) solar systems still remains unresolved. High penetration of Renewable Energy Sources (RESs) results in technical challenges for grid operators. To address this, Virtual Power Plants (VPPs) have been defined and developed to manage distributed energy resources with the aim of facilitating the integration of RESs. This paper introduces a hybrid irradiance forecasting approach aimed at facilitating the integration of PV systems into a VPP, especially when a historical irradiance dataset is exiguous or non-existent. This approach is based on Artificial Neural Networks (ANNs) and a novel similar hour-based selection algorithm, has been tested for a real PV installation, and has been validated also considering irradiance measurements from an aggregation of ground-based meteorological stations, which emulate the nodes of a VPP. Under a reduced historical dataset, the results show that the proposed similar hour-based method produces the best forecasts with regard to those obtained by the ANN-based approach. This is particularly true for one-month and two-month datasets minimizing the mean error by 16.32% and 9.07% respectively. Finally, to demonstrate the potential of the proposed approach, a comparative analysis has been carried out between the hybrid method and the most used benchmarks in the literature, namely, the persistence method and the method based on similar days. It has been demonstrated conclusively that the proposed model yields promising results regardless the length of the historical dataset.

INDEX TERMS Virtual power plants, hybrid irradiance forecasting, solar power integration, similarity matching.

NOMENCLATURE

ACRONYMS

AI	artificial intelligence
ANN	artificial neural network
BR	Bayesian regularization
DG	distributed generation
EMS	energy management system
ESS	energy storage system
ICT	information and communication technologies
k-NN	k-nearest neighbors
LM	Levenberg-Marquardt
MLP	multilayer perceptron
NWP	numerical weather prediction
PV	photovoltaic
RES	renewable energy source
RNN	recurrent artificial neural network
SCG	scaled conjugate gradient
SH	similar hour-based method
TSA	time series analysis
VPP	virtual power plant

The associate editor coordinating the review of this manuscript and approving it for publication was Fabio Mottola.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

NOTATION

BNI	beam normal irradiance $[W/m^2]$
CCF	cloud cover factor [-]
DHI	diffuse horizontal irradiance $[W/m^2]$
GHI	global horizontal irradiance $[W/m^2]$
MAE	mean absolute error $[W/m^2]$
MAPE	mean absolute percentage error [%]
NRMSE	normalized root-mean-square error [%]
PICP	Prediction Interval Coverage Probability [-]
α	elevation angle [rad]
θ_z	zenith angle [rad]
Ψ	azimuth angle [rad]
δ	declination angle [rad]
Φ	latitude angle [rad]
S	spatial pixel resolution [km]
H	cloud height [km]
(X, Y)	site location [pixel]
(R_X, R_Y)	relative cloud position [pixel]
(P_X, P_Y)	cloud location [pixel]
E	extra-terrestrial radiation $[W/m^2]$
T	temperature [°C]
D	difference vector [-]
u	uncertainty threshold [km]
ω	weight [-]
ED	Euclidean distance [-]
t	prediction hours [h]
z	past hours [h]
s	similar hours [h]
c	candidate hours [h]
N	normal distribution [-]
μ	mean value [-]
σ	standard deviation [-]
Y_t	measured data $[W/m^2]$
$\hat{Y}_t$	forecast value $[W/m^2]$

I. INTRODUCTION

The adoption of photovoltaic (PV) power generation is rising steeply worldwide [1]. There are several reasons behind its success: (a) the cost of photovoltaic power has plummeted since PV modules, storage systems and balance of system costs have been steadily dropping [2]. This has led to an increasing competitiveness in comparison to the conventional non-renewable resources; (b) PV peak power generation coincides with the time of higher load demand; (c) the increasing concern about climate change has definitely spread throughout the world and the electricity sector is playing a central role to fully decarbonize the power system. As a result, governments have implemented supportive policies to encourage investments in renewable sources of energy [3]; (d) the search for energy independence in most developed countries; and (e) most importantly the continued progress and improved accuracy of forecasting strategies of PV generation.

In general, power forecasting for renewable energy sources (RESs) has posed a considerable challenge during the last decade. This has particularly been the case for non-predictable resources such as solar energy where the power generation constantly fluctuates on account of meteorological factors such as cloud cover, temperature, wind speed or humidity level, which are stochastic in nature. This inherent uncertainty has hindered PV power integration at a high penetration level [4], [5]. This drawback can be overcome by including energy storage systems (ESSs) whereby this intermittent source of energy becomes more dispatchable [6]–[8]. However, a more technical and economic solution has been put forward: the aggregation of several PV systems into the so-called Virtual Power Plant (VPP) [9]–[11]. This approach allows prosumers [12], to maximize revenue opportunities by participating in the energy market mechanisms and by taking part in the operation of distribution and transmission networks in terms of the active control and services VPPs can provide, e.g. voltage regulation and frequency balancing, among others.

A VPP usually integrates four components [9]: (a) Distributed Generation (DG) units based on RESs and small scale fossil fuel conventional dispatchable generators; (b) ESSs; (c) responsive or flexible loads; and (d) information and communication technologies (ICTs) which play an essential role in the technological core of a VPP: the energy management system (EMS). The EMS coordinates the power flows among the different units in the VPP. Through a bidirectional communication strategy, which should be based on existing open standards such as the IEC 61850 [13], the VPP not only obtains information about the current state of the different nodes but also sends the commands related to specific targets, e.g. minimization of the generation costs, maximization of profit or reduction of greenhouse gases, to name a few. A crucial part, in the VPP general concept, involves obtaining accurate and rapid forecasts of the power generated by RESs with stochastic nature [14]. The purpose of forecasts is threefold: firstly, the power predictions allow VPP operators to meet regulatory requirements increasing the reliability and efficiency of the of the VPP; secondly, accurate predictions contribute to grid stability; and finally, more favourable trading conditions on the electricity markets can be achieved thereby maximizing revenue.

In the literature, there is a broad range of studies aimed at obtaining accurate forecasts. In this regard, [15]–[19] present comprehensive reviews of well-established techniques developed to forecast PV power generation. According to different factors, forecasting methods can be categorized into different groups. Regarding the forecasted parameter, two different approaches have been implemented: direct [20] and indirect [21]. Through historical datasets of weather conditions and PV power generation, the direct method predicts the generated power. Indirect forecasting, on the other hand, firstly predicts the solar irradiance and then, the output power is calculated by using a performance model of the PV plant. This approach is based on several methods including Numerical Weather Prediction (NWP) models, image-based systems, statistical-based alternatives and hybrid or ensemble

methods. As for the time horizon, four categories can be found [22], [23]: very short term forecast also called nowcasting (from 1 min to several minutes), short-term forecast (from 1 hour to several hours), medium term forecast (from 1 month to 1 year) and long-term forecast (up to several years). As far as the model approach is concerned, four types have been widely used: (a) statistical models based on time series analysis (TSA) which tries to identify patterns between historical datasets and the output parameters; (b) artificial intelligence (AI) models mainly based on artificial neural networks (ANNs); (c) physical strategies which use solar and PV models for solar power forecasting; (d) hybrid models which explore different algorithm combinations with the aim of improving forecast accuracy and reducing computational burden for online forecasting applications [22]. Another ongoing challenge in solar power forecasting consists in assessing the uncertainty of the results. To assist with this, deterministic forecasting, also called point forecasting, produces a single value for each timestamp within the time horizon without considering either the upper and lower bounds or the percentage of confidence for each value. Probabilistic forecasting, on the other hand, provides additional accurate information about the expected values in terms of the range of plausible values and the probability associated to each of them [17]. Finally, regarding the spatial horizon, forecasting techniques can be applied to a single plant or to an ensemble with the last option being of major interest because it usually provides greater accuracy.

As mentioned above, the indirect forecasting approach aims firstly at predicting the solar irradiance, mainly global horizontal irradiance (GHI), and then by using the physical model of the PV system, the output power is calculated. This is the strategy used in this paper for one reason: weather-related variables and irradiance datasets can be obtained from ground-based meteorological stations. Likewise, cloudiness and temperature forecasts are freely available from weather forecast web services such that of AEMET in Spain [24]. This clearly facilitates stable and accurate forecasts even in the initial stage of the PV system operation [21]. As opposed to indirect forecasts, direct approaches require an extensive historical dataset for the derivation of the power forecast model, which reduces the possibility of accurate predictions when a new VPP node is integrated. It would be interesting to forecast other irradiance-related parameters such as Diffuse Horizontal Irradiance (DHI) and Beam Normal Irradiance (BNI). However, there are no datasets for these variables since the GHI is usually the only parameter measured by meteorological stations. That is the reason why indirect forecasting methods are mainly developed for GHI predictions and for the other irradiance-related variables are virtually non-existent, especially for DHI [25] in which issues relating to sensor calibration and spatial representativeness are difficult to address [26].

Approaches to irradiance forecasting can range from the most basic such that of similar day-based method to the most demanding in terms of computational load such as Recurrent Artificial Neural Networks (RNNs) which require computationally demanding training algorithms. The similar day-based approach provides an appropriate choice for irradiance forecasting. Similar day-based forecasting involves mining a dataset with the aim of finding days or even hours which are similar to the forecast day/hour in terms of certain parameters such as cloudiness and temperature [27]. The success of this alternative relies on the low computational burden it imposes on the forecast algorithm. On the other hand, ANNs have been extensively used for daily solar irradiance forecasts [28]–[33]. The forecast performance of an ANN relies on the learning algorithm along with the data available for the training process, the transfer function, the architecture, the nonlinear mapping capacity and the choice of input variables. The main limitation of ANNs stems from the fact that they required an extensive dataset for training purposes for better generalization and accuracy. Conversely, similarity matching works better than ANNs for short datasets, especially when a time granularity of one hour is considered. This enhancement is demonstrated conclusively in this work. Therefore, in the context of a VPP and at the earlier stages of its operation when limited data is available, hybrid strategies, which combine different methods, can improve the overall forecasting accuracy. In general, hybrid techniques have been widely used in diverse industrial applications [34], [35], delivering good results.

In this paper, an irradiance short-term forecasting strategy is presented, with the aim of facilitating the integration of PV systems in VPPs, especially when the lack of a comprehensive dataset hinders the forecasting performance of the algorithms causing inaccurate results. This strategy is based on a hybrid approach which combines an ANN and a novel similar hour-based forecasting algorithm. The outputs of both forecasting methods are dynamically weighted, according to the type of the day and some accuracy metrics, to provide the final forecast. The forecasting approach relies on no-cost temperature and cloudiness forecast maps generated by the AEMET via NWP, the irradiance measurements from a real PV installation located in the Polytechnic School of Alcala University and different ground-based meteorological stations emulating the role of VPP nodes.

The main contributions of this paper are summarized as follows: (i) the proposed forecasting approach is implemented in the context of a VPP considering the challenges it poses and drawing on its strengths; (ii) the input data for the forecasting algorithms comes from weather forecasts regularly published, free of charge, by the AEMET; (iii) the similar hour-based approach, which produces accurate irradiance forecasts for a reduced dataset, this usually being the case when a new node is integrated in the VPP; and (iv) the ensemble of ANNs and the similar hour-based approach which, through a dynamically weighted function that depends on the type of day, produces encouraging results.

The paper is organized as follows. Section II introduces a general description of the irradiance forecasting hybrid approach. In section III the data description and

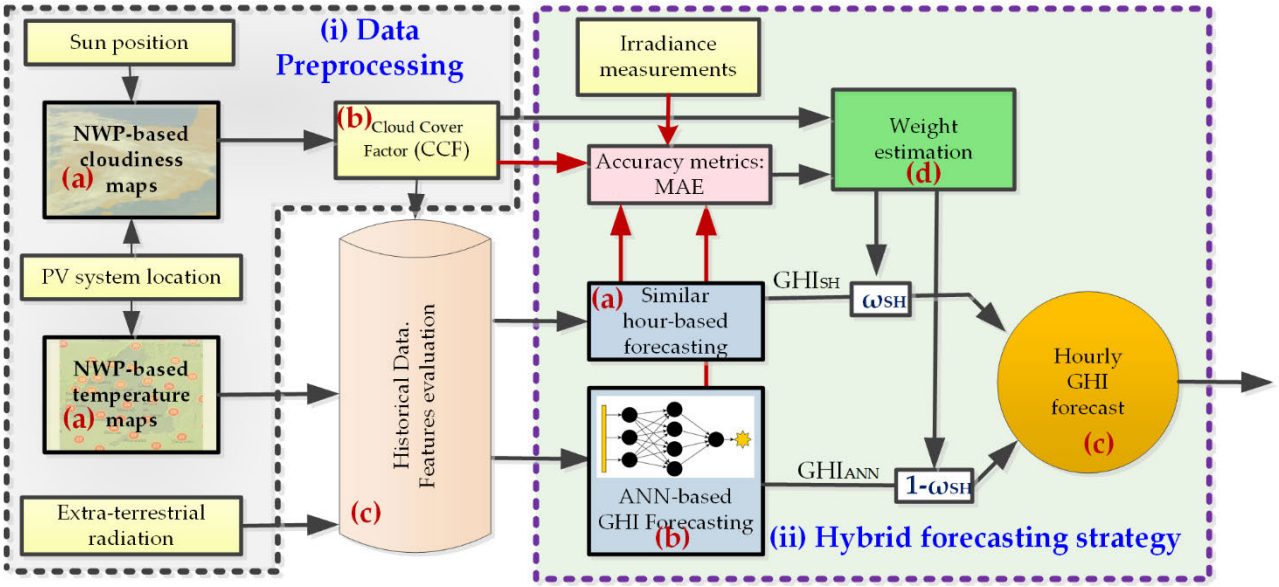


FIGURE 1. Block diagram of the forecasting approach, which consists of two parts: (i) a data pre-processing stage and (ii) the hybrid forecasting approach. The forecasting strategy is based on ANNs and a novel similar hour-based algorithm. The final forecast is obtained by dynamically weighting the outputs as a function of the type of the day.

pre-processing are presented. Section IV analyses in depth the algorithms involved in the hybrid approach. The experimental results are presented in Section V. Finally, some conclusions are drawn in the final section.

II. IRRADIANCE FORECASTING APPROACH. GENERAL DESCRIPTION

An important feature of irradiance forecasting models is that, in general, they rely on extensive historical dataset. However, when a VPP is to be operated in a cost-efficient manner at its initial stage or when a node is first integrated in an existing VPP, the lack of data reduces the accuracy of the day-ahead estimation of the irradiance and, as a result, the accuracy of the power forecasts. This leads to uncertainties, which result in financial penalties imposed by the grid operator. NWP-based GHI forecasts have proved to be a tool for indirect solar power prediction. Furthermore, when it comes to VPPs, an aggregation of small-scale PV installations makes it crucial to implement an EMS, which must include accurate forecasts such as, for instance, the NWP-based GHI forecasts for each location or site. However, this carries a cost, which depends on the number of sites taking part in the VPP. In order not to incur costs, which would cause a decline in profits, free access NWP-based cloudiness and temperature maps are used in this paper. This information along with known parameters such as the sun position, the location of the PV sites and the extraterrestrial radiation, constitute the inputs of the GHI forecasting hybrid approach proposed in this paper.

Fig. 1 shows the approach, which is based on ANNs and a novel similar hour-based algorithm. The outputs of both techniques are dynamically weighted according to the type of the day and some accuracy metrics. Thus, uncertainties in the 24-hour ahead final GHI forecasting, are reduced. The forecasting method consists of two parts: (i) a data pre-processing stage; and (ii) the hybrid forecasting approach.

i. The first part has 3 steps: (a) new data acquisition and transformation to provide the input data of the algorithm; (b) CCF calculation; and (c) data merging within the historical dataset collected. In step (a), to manage the ESS of some sites at night, weather forecast maps are downloaded at 22:00 hours from [24], gathering information on the day-ahead weather variables such as area of cloud cover and temperature. To turn the information from the maps into numerical data a transformation process is required. In step (b) the lack of an extensive dataset, especially at the earlier stages of the VPP operation, makes it essential to optimize the data available in order to provide the forecasting algorithms with the most relevant and correlated information. Hence, the data pre-processing stage becomes crucial. The 24-hour-ahead cloud cover maps are used to define a parameter, referred in this paper as Cloud Cover Factor (CCF), Section III-A, which is also based on the sun position. The CCF contains information about the shadows on the PV installation generated by a particular cloud area. In general, the shadowed area will be larger than the corresponding cloud area. Secondly, temperature maps are used to obtain the temperatures for the 24-hour target forecasting period. These temperatures can be validated by using both real measurements taken in the PV installation and those from the closest ground-based meteorological stations. Finally, the extra-terrestrial radiation gives information about the radiation at the top of the Earth's atmosphere.

In this paper it is assumed that the determining factor for the loss of radiation is the CCF, disregarding the influence of other factors such as the air molecules, the distance the solar radiation has travelled through the air mass, etc., which are assumed to modify the GHI in lesser proportion. This clearly introduces an estimation error the predictable effect of which is somewhat mitigated by giving more importance to those days closer to the target day. This is achieved by considering the temperature, since it is a parameter that depends on the season of the year. Finally, in step (c), once the information of the day-ahead weather variables has been processed, the historical dataset is updated with this information and the forecasting strategy can then be implemented.

ii. The second part focuses on the forecasting strategy and has 4 steps: (a) similar hour-based forecasting method; (b) artificial neural network forecasting approach; (c) hybrid forecasting strategy; and (d) weight estimation. In step (a) the novel method referred to in this paper as similar hour-based forecasting, Section IV-A, is implemented. It is based on the traditional method of similar days. The similar hour-based method performs reliably dealing with the information extracted from a reduced historical dataset. In step (b) the ANN, Section IV-B, forecasts from the same dataset. However, ANNs generalize better when an extensive historical dataset is available. In step (c) both methods are combined, thereby providing an accurate forecast irrespective of the dataset size. This makes the efficient management of a VPP possible from the very beginning or when a new VPP node is added.

The combined GHI forecasting output is the weighted sum of the individual GHI forecasting outputs of the two methods explained above (d). The weights depend on the type of the day, i.e. sunny, cloudy and overcast, and the Mean Absolute Error (MAE). Metrics based on mean and squared values have been selected to assess the performance accuracy because they are the most commonly used indexes in solar radiation techniques [36]. Error mean values are used for selection purposes, to minimize the forecasting error of each node comprising the VPP, irrespective of the length of the database considered, instead of penalizing atypical values. It is worth mentioning that results did not change excessively with the inclusion of atypical values. The type of the day is defined by means of the CCF, which shows to what extent a cloud area on the NWP-based cloudiness maps creates shadows on the PV installation.

The weights are calculated by using (1) where d_s, d_c, and d_o, are mutually exclusive flags which can take the values of either 0 or 1, representing with a value of 1, the type of day determined by the CCF, i.e. sunny, $d_s = 1$, cloudy, $d_c = 1$, or overcast, $d_o = 1$, and ω_d is the value of the weights which minimize the MAE in past hybrid predictions obtained for this type of day. These weights are updated daily by incorporating the latest and most up-to-date information from the dataset, thereby improving the accuracy of forecasting results.

$$\begin{aligned} \omega_{SH} &= d_s\omega_{d_s} + d_c\omega_{d_c} + d_o\omega_{d_o} \\ \omega_{ANN} &= 1 - \omega_{SH} \end{aligned} \tag{1}$$

III. DATA DESCRIPTION AND PRE-PROCESSING

As mentioned above, the inputs to the algorithm are based on weather forecasts, provided by the AEMET at different spatial and temporal scales, and the extra-terrestrial radiation, E, which is deterministic and can be evaluated by using known expressions. For instance, Duffie and Beckman's equation was considered to determine the extra-terrestrial radiation. Nevertheless, other expressions available in the literature [37] are equally accurate.

The weather forecasts are based on the NWP model HARMONIE-AROME. This model is commonly utilized for weather forecasts in Spain and other European countries [38]. NWP models include GHI and DNI forecasts, both being necessary to model irradiance on the inclined surfaces of the solar panels. However, the cost of purchasing this data is a deterrent for small-scale PV systems. Other weather forecasts include cloudiness, temperature, pressure and wind, all of them in the shape of weather maps. Cloudiness forecasts contain relevant information regarding irradiance. To turn this information into numerical values some data pre-processing must be applied. The term used in this paper for these numerical values is the above-mentioned cloud cover factor (CCF), in order not to confuse it with other parameters such as cloudiness index, which is modelled in a different way. Studies which rely on images from satellite or ground level cameras to find the shadows cast by the clouds already exist [39]–[42]. In this regard, the CCF provides the same information but without cost.

A. CLOUD COVER FACTOR EVALUATION FROM CLOUDINESS FORECAST MAPS.

The CCF is dimensionless and represents, in the context of the weather maps, the amount of cloud cover per pixel in each cloudiness forecast map, showing a negative correlation with the GHI, being 0 when the sunlight is not blocked by clouds and 1 when the sun is totally covered. At a particular time of day, e.g. 22:00, the 24 NWP-based cloudiness images, representing the cloudiness forecast for the next 24 hours, are downloaded from the AEMET web service. Fig. 2 represents a zoom area of the cloudiness forecast for the Community of Madrid (centre of the map) at 16:00 on March 1st, 2020. This image was downloaded on February 29th, 2020. The colormap on the right represents the percentage of cloudiness. The spatial resolution in the map is given by the pixel size in the image, being represented by a square with sides approximately equal to 2.5km.

Another important parameter for the CCF calculation is the cloud height. Unfortunately, this parameter cannot be obtained from Fig. 2 and some data pre-processing must be

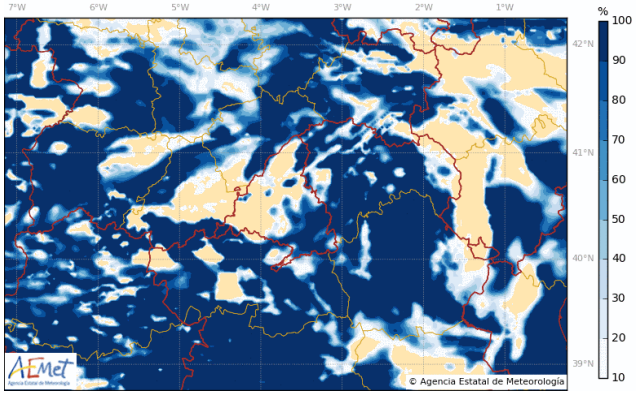


FIGURE 2. NWP-based cloudiness forecast map for the Community of Madrid (center of the map), from the spanish website, ©AEMET. Values range from 0% (absence of clouds) to 100% (heavy clouds).

done to identify those clouds that prevent the sun radiation from reaching the site, i.e. the clouds between the sun and the site. Therefore, both the cloud location and the sun's position are required. Fig. 3 shows the hourly position of the sun at the spring equinox for a site located in the Northern Hemisphere, and the most relevant variables used for the CCF evaluation. For the site-related pixels in the map, the aim is to quantify the CCF at a particular time of the day. Considering that the site location is known (X, Y), those pixels in the map covered by the clouds (P_X, P_Y) can be worked out, as a function of the cloud height (H), by using the following equations:

$$P_X = X + R_X \approx X + \frac{H}{S}\frac{sin\,(\psi)}{tan\,(\alpha)} \tag{2}$$

$$P_Y = Y + R_Y \approx Y + \frac{H}{S}\frac{cos\,(\psi)}{tan\,(\alpha)} \tag{3}$$

where R_X and R_Y represent the relative position of the pixel with respect to the site, S is the spatial resolution in the map $(2.5km)$, and α and ψ are the elevation and azimuth angles, respectively.

Since the cloud height cannot be extracted from the cloudiness forecast, (2) and (3) are evaluated with the greatest cloud height considered (15 km) [43]. By doing so, the cloudiness values of the pixels on the way from (X, Y) to (P_X, P_Y) are evaluated and the average of those values is stored.

However, the CCF can vary widely over time when clouds are present on the map. To reduce this variation, a smoothing procedure based on a parameter called uncertainty threshold (u) is applied. This variable is used to expand the selected area in every direction, selecting a larger area on the map to obtain a greater number of pixels that provides a smoother variation of CCF values and enables the identification of the type of the day. The value of u is selected through an iterative process in which the threshold is gradually increased, i.e. the selected area on the map is expanded. The process terminates when the CCF variation is smooth and its value is consistent with the GHI measured. The value of u that minimizes the forecasting error was set to 8 pixels, covering an area of 20 km.

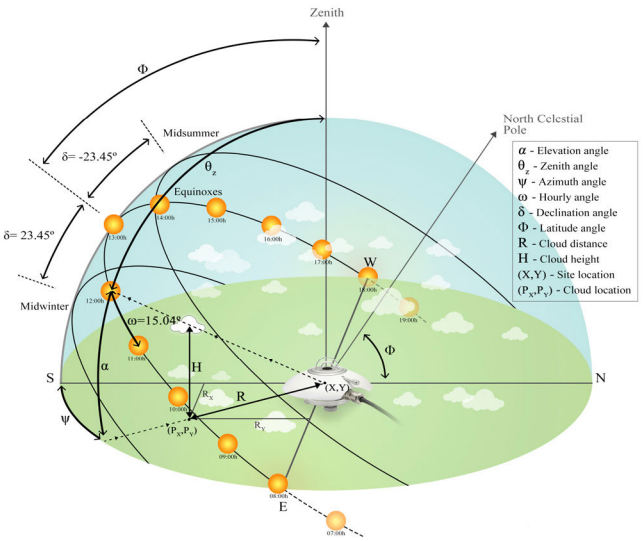


FIGURE 3. Relevant variables for the CCF assessment with respect to the sun position and the site location.

IV. HYBRID APPROACH FOR THE HOURLY GHI FORECASTING

In this section, the similar hour-based and ANN-based methods comprising the hybrid approach for GHI forecasting are explained.

A. SIMILAR HOUR-BASED APPROACH

The underlying behaviour behind meteorological events is difficult to model although it can be assumed that weather conditions repeat themselves in time. Therefore, searching for similarities is the key to predicting when certain weather conditions will repeat in the future.

A similar day-based approach intrinsically considers the weather conditions of a whole day to forecast the GHI in a moment of the day. Since the weather conditions do not follow a marked trend during the day, it seems perfectly reasonable to use the meteorological variables forecast at the target hour for similarity matching. Furthermore, if the candidates for the similarity study also depend on the extra-terrestrial radiation, which replaces the conventional time variables for the day and hour, the number of potential hours to be considered is significantly increased.

In the model proposed in this paper, for similarity matching, a time-window of three hours around the forecast hour is considered, because the distance travelled by the sunlight through the atmosphere depends on the position of the sun in the sky (elevation and azimuth) which, in turn, influences in the GHI loss variation Therefore, the same hour as that of the target hour along with the adjacent hours, i.e. the previous and the following one, are extracted from each day in the historical dataset since these hours are similar in terms of the sun's position in the sky. The Sun position for a particular hour progressively changes as the days go by, becoming closer to the Sun position for the adjacent hours. However,

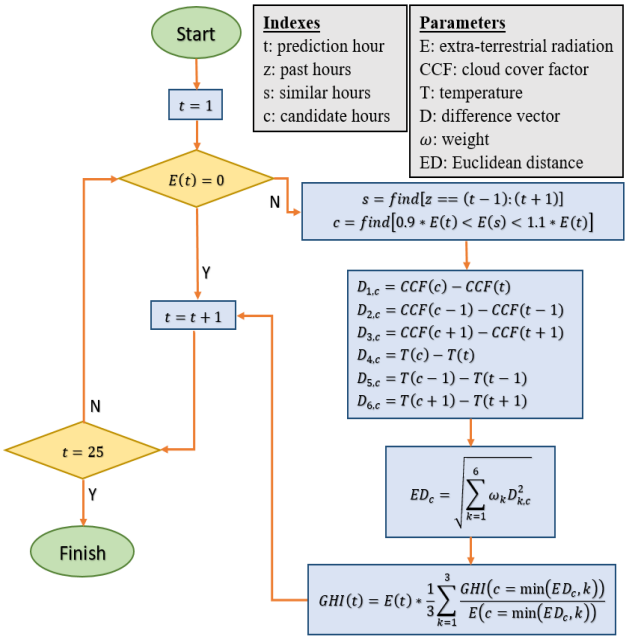


FIGURE 4. Flowchart of the proposed similar-hour based forecasting algorithm for a day-ahead prediction.

this strategy requires the removal of the hours with highly dissimilar values for the extra-terrestrial radiation. To that end, only the hours with values for the extra-terrestrial radiation, within a range of $\pm 10\%$ from that of the forecast hour are selected as potential candidates for similar hours. This strategy is referred to as similar hour-based approach in this work, and it makes a significant difference with respect to the day-based version. This is one of the main contributions of this paper since, to the best of the author's knowledge, this is the first time this approach has been used for irradiance forecasting. This algorithm, in contrast to the similar day-based methods, uses extra-terrestrial radiation to filter the most important time instants for the prediction, and delivers outstanding results in the early stage of the VPP node. Furthermore, the accuracy of this method is improved as new VPP nodes are aggregated, reducing the global error to a greater extent compared to other forecasting techniques.

The similar hour-based methodology is depicted in the flowchart of Fig. 4. Once the candidate hours (c) have been chosen from the historical dataset (z), the algorithm searches for similarities with the forecast hour (t) in terms of the CCF and temperature (T). Temperatures annually vary from minimum values in winter to maximum values in summer. Needless to say, the further back in time the potential candidates for similar hour are located, the less likely it is that the hours become similar hours.

The Euclidean distance (ED) is used as a measure of similarity. Firstly, the difference vectors (D) are obtained by evaluating the differences between the meteorological parameters, i.e. the CCF and the temperature (T), for the forecast hour (t) and those from the candidate hours (c). As justified above, the adjacent hours, i.e. $(t-1)$ and $(t+1)$, to the forecast hour (t) for the difference vector (D) calculation are also considered. As a result, a total of 6 variables are evaluated for the difference vectors (D), whose contribution to the similarity matching process is not, however, the same. Therefore, in a second step, a set of weights (ω), representing the relative importance the similar hour-based algorithm gives to each difference vector (D), is used. The selection of the weights (ω) is based on the principle of minimum error for the historical dataset of past forecasts, which are updated daily. Finally, the Euclidean weighted distance (ED) is calculated to find the most similar hours in the past to the forecast hour (t). Since NWP models have an inherent error that can adversely affect the forecast accuracy, the irradiance of the three candidates with the smallest Euclidean distance (ED) are averaged and taken as the final solution. Before the average can be calculated, an adjustment in the values of the irradiance of the three candidates is necessary. This adjustment is proportional to the difference between the extra-terrestrial radiation (E) for the chosen hour and that of the forecast hour (t). The algorithm iteratively repeats the similarity matching as long as the extra-terrestrial radiation (E) is greater than zero for the forecast hour (t).

B. ARTIFICIAL NEURAL NETWORK FORECASTING

The second forecast method making up the hybrid approach consists of an ANN, which produces accurate forecasts from extensive datasets. This clearly complements the similar hour-based method, which yields better results for a reduced dataset for which the ANN performance is only moderate. In this work, a multilayer perceptron (MLP) has been developed due to its simplicity and good overall performance. This type of ANN is the most widely used technique for day-ahead irradiance forecasting [44], excluding RNN because the time horizon is too large for a proper forecast using the observations as input [45] and the length of the dataset is too short to obtain patterns in a day-ahead forecasting [32].

The neural network has three layers: (i) an input layer has 7 neurons one for each predictor variable, namely, the CCF and the temperature at the time $(t-1)$, t and $(t+1)$, being t the forecast time, and the extra-terrestrial radiation; (ii) a hidden layer with 10 neurons which minimizes the forecast error by using a logarithmic sigmoidal activation function; and (iii) and output layer with one neuron with a linear transfer function which provides hourly GHI forecasting values.

The dataset comprises data from December 4th, 2019 to May 31st, 2020. To simulate the real scenario of the initial stage of a VPP, the ANN is trained daily with the historical dataset collected to the date under consideration. For operational purposes, the minimum amount of days in the dataset is established as 7 and the training and validation process is repeated every day. Consequently, the training dataset is variable and increases as the VPP operation time increases and more data is available.

The number of hidden layers is evaluated through cross-validation using the dataset available. Adding more layers

TABLE 1. MLP ANN characteristics of the hybrid method forecasting.

Parameter	MLP
Inputs	T, CCF, E
Output	GHI
Number of layers	3
Input neurons	7
Hidden neurons (one layer)	10
Output neurons	1
Hidden layer activation function	logistic sigmoid
Output layer activation function	linear
Objective function	MAE
Learning algorithm	Levenberg-Marquardt
Minimum gradient	1e-4
Maximum iterations	500

does not yield better results, but instead increases computational time. The Levenberg-Marquardt (LM) algorithm is used for the ANN learning process, because it ensures greater accuracy in comparison to other alternatives such as LM, Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG).

A sensitivity analysis of the ANN has also been carried out, to determine the relationship between the inputs and the output. As is to be expected, the extra-terrestrial radiation provides the greatest amount of information. The CCF is the second most important input, followed by the temperature (T). Temperature measured at time instants (t+1) and (t-1) are also relevant on account of the different patterns during mornings and evenings: depending on the temperature gradient, the ANN can recognize both periods of time.

To achieve better generalization, a popular technique in the context of short dataset is implemented, [32], [46]. It consists in training several independent ANNs for the same target variable. The average of the outputs of the set of independent ANNs is taken as the final prediction. Assuming that the error follows a normal distribution $N(\mu, \sigma) \sim N(0, 1)$, the average of the output values of up to 30 independent ANNs allows the prediction to be correctly validated. The characteristics of the ANN are summarized in Table 1.

C. COMBINING THE SIMILAR HOUR-BASED ALGORITHM WITH ANN-BASED FORECASTING

Once the forecasting methods have been introduced, the hybrid approach is explained in this section. The hybrid method consists in evaluating the final prediction as a weighted value of the individual forecasting outputs as shown in (4). The weights are determined as a function of the type of the day (sunny, cloudy and overcast), which depends on the values of the CCF at the forecast hour. By using the k-nearest neighbors (k-NN) algorithm, the days are classified taking into account the CCF and the corresponding set of weights are associated with the type of day. This simple classification algorithm allows the weight selection to be carried out automatically and independently for each site, selecting the set of weights that minimizes the MAE of previous forecasts.

$$GHI = \omega_{SH} GHI_{SH} + (1 - \omega_{SH}) GHI_{ANN} \tag{4}$$

The weights are updated daily for each type of day to optimize the final forecast. The historical data for the weight evaluation consists of up to a 2-month slide window with the most recent data. There are two reasons behind this value for the window width: (a) ANN performance improves as more data is available, which means that previous forecast should be disregarded; (b) it is expected that the strong seasonality in the weather directly influences the weights.

The novelty of the hybrid approach lies in the development of a model which has the ability to adapt itself to the amount of historical data. As previously stated, for a reduced dataset, the similar hour-based method outperforms ANN-based strategy, whereas the converse applies for larger datasets. Combining both methods, therefore, an accurate prediction can be obtained irrespective of the size of the dataset.

Initially, the similar hour-based forecasting output has a great influence on the prediction because this method considers extra-terrestrial radiation to filter the most important time instants for the prediction. When more data is available, the ANN-based forecasting gains more influence. In this way, a smooth transition of the weights is achieved.

V. RESULTS

This section presents the results obtained from the implementation of the similar hour-based and hybrid GHI forecasting strategies in two different scenarios: firstly, the approach is applied to an experimental setup (a real PV installation) that plays the role of a VPP node; and secondly, an aggregation of different PV installations in the shape of ground-based meteorological stations making up a VPP, is considered. The results from using other techniques, such as the persistence model, the similar day-based approach and neural networks, are also included for comparison purposes, proving the effectiveness of the proposed approaches.

A. EVALUATION OF THE HYBRID FORECASTING APPROACH FOR A REAL VPP NODE

To validate the proposed algorithm in the context of a single VPP node scenario, measurements taken from a recently installed photovoltaic facility located at the Polytechnic School of the University of Alcalá (Spain) are used. These measurements, mainly comprising GHI and temperature values, were recorded during the period between December 4th, 2019 to May 31st, 2020, and constitute the 6-month period of historical dataset for the algorithm validation. The first forecast is provided by using only a week of the historical dataset, which allows GHI predictions to be made from the earliest stages of the VPP node operation. The 24-hour ahead

GHI forecasting process is repeated on a daily basis, updating the data used in the process, with the collected data of that day. This process is carried out until all the data in the historical dataset is used. In this way, the performance of the prediction algorithm is assessed daily starting from the second week of the PV system operation until the 6-month period is covered.

To analyse the accuracy of the prediction algorithm, the overall error is calculated using the following performance indicators:

$$MAPE = \frac{\frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \hat{Y}_t\right|}{\frac{1}{T}\sum_{t=1}^{T} Y_t} 100[\%] \quad (5)$$

$$NRMSE = \frac{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \hat{Y}_t\right)^2}}{\frac{1}{T}\sum_{t=1}^{T} Y_t} 100[\%] \quad (6)$$

where Y_t is the measured data, $\hat{Y}_t$ is the forecast value and T is the length of the time series. MAPE shows the normalized average error between the measurements and the forecast, while NRMSE represents the normalized square error. The normalisation parameter used is the average of the measurements. Normalized indicators are used due to the fact that they allow a fair comparison of the results obtained as the validation is developed, since the dependence on the magnitude is removed. The use of both indicators is justified on the grounds that NRMSE is more sensitive to outliers than MAPE, which allows for a more comprehensive comparative study to be carried out.

In order to validate the benefits of the proposed forecasting strategy, described in section IV-C, a comparison is made with other widely used forecasting methods, as well as with the proposed similar hour-based approach. The following methods are analysed:

- Persistence model: it is the simplest method which assumes that the 24-hour ahead GHI forecast is equal to the GHI measurements taken the previous day [32].
- Similar day-based approach: this method calculates the difference vectors for the CCF and the temperature considering the weather forecasting of the day to be predicted and the previous 14 days within the historical dataset. The selection of the number of days is not arbitrary and aims at minimizing the error. The Euclidean distance of all the difference vectors is then calculated for each of the 14 days and the day that minimizes this distance is chosen as the following day's GHI prediction [47], [48].
- Proposed similar hour-based algorithm, which is described in detail in Section IV-A and separately implemented without hybridization.
- ANN-based forecasting approach, described in Section IV-B, and individually evaluated.

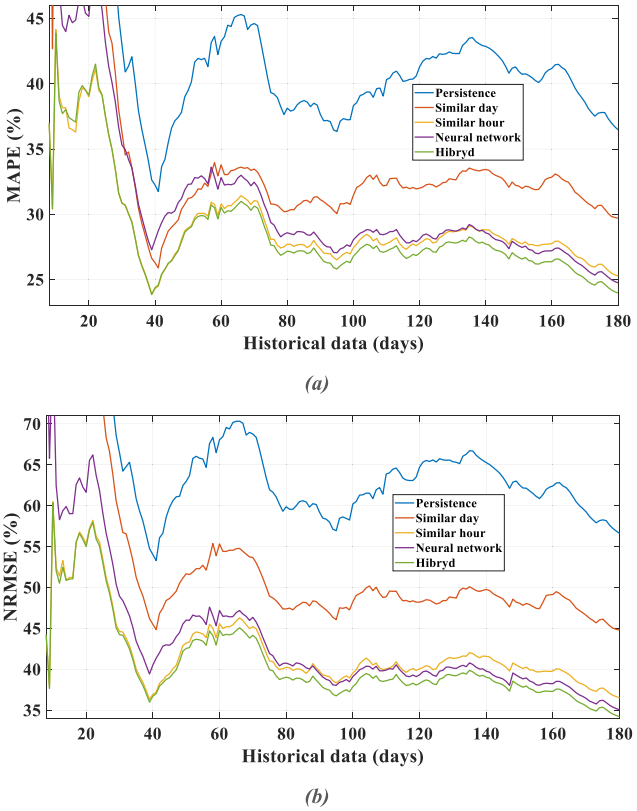


FIGURE 5. Forecasting errors of the methods under study (persistence, similar day, similar hour, neural network and hybrid model proposed) for the Alcala University site in terms of (a) MAPE and (b) NRMSE.

- Proposed Hybrid approach, described in detail in Section IV-C.

Fig. 5 depicts the forecast errors, using the MAPE and NRMSE indexes, depending on the days from the historical dataset employed. In order to improve the graph visualization, the highest errors made by certain forecasting methods when implemented for a reduced amount of historical data, are neglected. The maximum values in the neglected period for the persistence model, and similar day-based and ANN-based approaches are for the MAPE = [96.8 66.2 69.7] and for the NRMSE = [126.2 116.4 85.1].

From the figure it can be seen that the persistence method, on account of its simplicity, is the one that introduces the greatest error. The similar day-based forecast significantly improves the persistence prediction achieving a similar degree of accuracy to that of the ANN-based forecast for a reduced amount of historical data. However, as more data is available, the performance enhancement of the ANN is noticeable, especially with respect to that of similar day-based forecasting. The proposed similar hour-based method performs much better than the other methods in the case where little historical data is available. As the amount of historical data increases, it can be appreciated that its performance keeps improving relative to the similar day-based approach, and with similar accuracy to that of the ANN-based

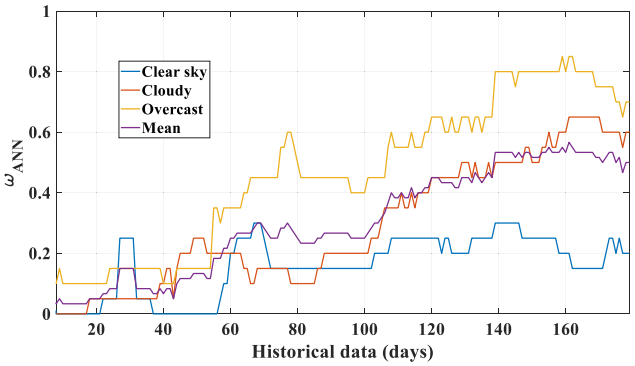


FIGURE 6. Weights used in the hybrid method for the ANN as a function of the number of days in the historical dataset, for the Alcala University site. It can be seen that the weights progressively give more importance to the ANN output, following a marked trend.

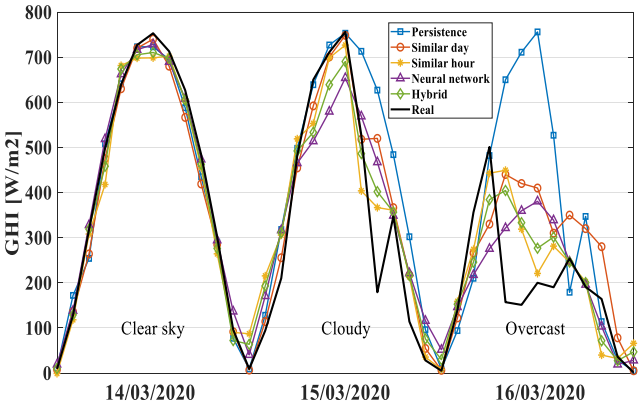


FIGURE 7. GHI forecasting values at the Alcala University site for the different methods, over three consecutive days, showing different weather conditions.

method. On the other hand, the ANN-based forecasts deliver better results than the similar hour-based method when a sufficiently extensive historical dataset is available, this being the reason that the ANN can generalize better. Finally, the hybrid method presented in this paper practically outperforms all the previous ones irrespective of the amount of data, because it manages to combine the advantages of neural networks and the similar hour matching.

The proposed hybrid method has been adjusted for different weights depending on the type of day as described in section 4.3. These weights evolve as the historical record increases as shown in Fig. 6.

Fig. 6 demonstrates how, as the historical dataset increases, the neuronal network carries more weight in the final prediction. This is justified by analysing Fig. 5(a), in which the accuracy of the ANN-based technique compared to that of the similar hour-based approach, gradually improves as the amount of historical data increases. This effect is more significant on cloudy and overcast days, since on clear sky days the performance of the similar hour-based technique is slightly better than that of the ANN-based approach. This

evolution is noticeable by analyzing the mean weight of the three day types in which the initial weight associated to the ANN output starts from approximately $\omega_{ANN} = 0.04$ reaching up to approximately $\omega_{ANN} = 0.51$ when the total historical dataset is completed.

Finally, Fig. 7 shows the GHI forecasting output of each method for three consecutive days between 14th and 16th March 2020. These days have been deliberately chosen because they represent the three types of day considered in the hybrid method. Moreover, in this case, approximately half of the historical dataset is employed, 102 days. It is observed that the error produced by all the forecasting methods increases as the cloudiness grows. This is because cloudy days are the most complicated to forecast since clouds have a strong impact on the GHI and it is difficult to predict their exact location for a 24-hour horizon. However, it can be seen from the results, that with the proposed hybrid strategy a considerable improvement in the forecast accuracy is achieved.

B. EVALUATION OF THE HYBRID FORECASTING APPROACH FOR AN EMULATED VPP

The potential of the proposed forecasting algorithms having already been demonstrated for a single photovoltaic installation, i.e. a single VPP node, in this section additional improvements are described for a set of photovoltaic facilities, grouped under the concept of VPP. As no additional photovoltaic installations are currently available, 6 ground-based meteorological stations located in the Community of Madrid in Spain (see Fig. 8) are used to emulate the new nodes. This is possible because meteorological stations publish, free of charge, all the required data utilized in the forecasting approach. The location of the stations is depicted in Fig. 8.

As in the previous section, the MAPE and NRMSE indexes are used to quantify the accuracy of the forecast outcomes for all the methods previously presented. The main difference is that in this case, once the GHI forecasts are produced for each station, all the VPP GHI variables are calculated by adding the corresponding GHI forecasts from each station, including the PV installation. In this way, the individual GHI measurements taken at each station, i.e. node of the VPP, are compared to the total GHI forecast.

Fig. 9 displays the evolution of the GHI forecasting error, for the ensemble of stations, quantified by the MAPE and NRMSE indexes. As depicted in Fig. 5, the graphs do not show the highest errors made by certain methods for the reason stated above. The maximum values in this omitted area for the persistence model, and for the similar day-based and ANN-based approaches are $MAPE = [93.263.5\ 62.8]$ and $NRMSE = [119.7\ 110.774.5]$. As in the case of a single installation, it can be seen that the persistence method exhibits the poorest performance followed by the similar day-based approach. The ANN-based method improves as the amount of historical data increases. In this case, however, it is not able to outperform the proposed method based on similar hours, irrespective the number of the days in the historical dataset. The novel methods proposed in this paper, i.e. similar hour-based and hybrid, achieve the highest accuracy regardless of

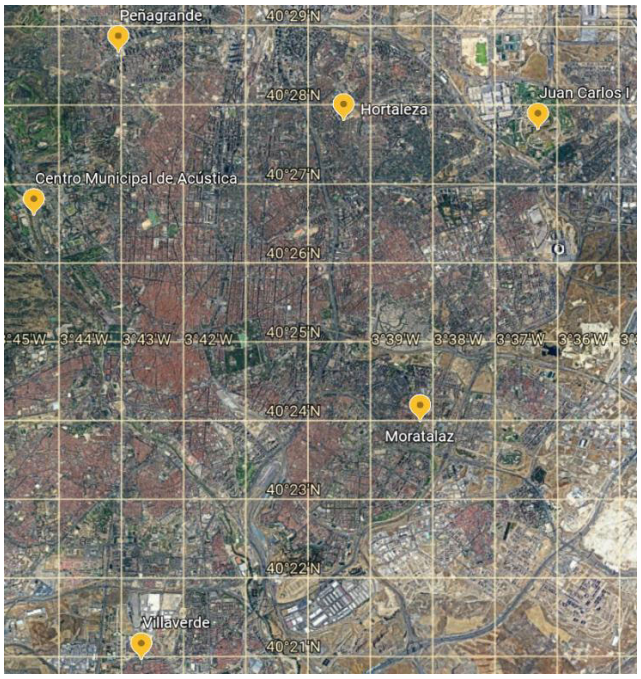


FIGURE 8. Location of the different ground-based meteorological stations in the Community of Madrid used in the study.

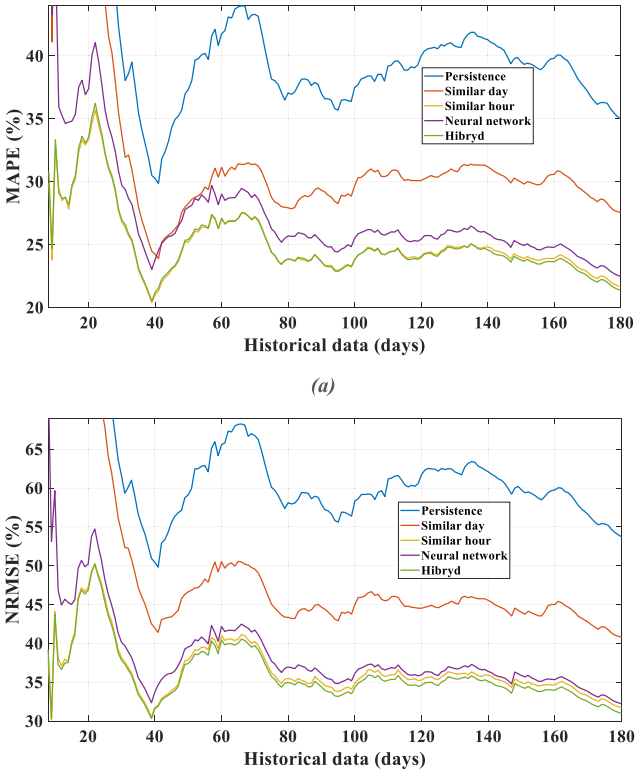


FIGURE 9. Forecasting error of the methods under study (persistence, similar day, similar hour, neural network and hybrid model proposed) for the VPP in terms of (a) MAPE and (b) NRMSE.

the number of days in the dataset, with the exception of the hybrid method, in which accuracy slightly increases as the historical dataset builds up.

To analyze the improvement of the proposed forecasting algorithms within the VPP scheme, the results obtained in each node are presented in Table 2. It can be appreciated that all the nodes in the VPP based on the meteorological stations, have similar performance to the one of the PV installation shown in Section V-A, in which the method based on similar hours introduces higher forecast error, in terms of MAPE and NRMSE, than the method based on ANNs when there is enough historical data available. However, when the similar hour-based GHI forecast for the entire VPP is calculated, the deviations from the GHI actual values, in the individual predictions for each node, tend to be compensated to a greater extent than when the ANN-based method is used.

This is because the method based on similar hours generalizes as a function of the node being considered and, consequently it can be assumed that the forecast errors follow different distributions. This makes certain errors partially compensate with each other when the GHI forecasts of the ensemble of VPP nodes are added. In contrast to the similar hour-based algorithm, in the ANN-based model the relationship between the inputs and the output identifies similar GHI patterns irrespective of the VPP node and, as a result, no error compensation takes place. The MAPE reduction of the similar hour-based strategy comparing the arithmetic mean of the 7 VPP nodes and the whole VPP is 3.31%, and in the case of NRMSE is 5.40%. As for the neural networks the reduction is 1.58% for the MAPE and 2.53% for the NRMSE. In the hybrid method, this reduction ranges between the two previous values as expected, 2.38% for the MAPE and 3.75% for the NRMSE.

Fig. 10 presents the average weight of all the nodes considered according to the historical dataset and the type of day. Although this average weight is not directly applied, because the set of weights for each VPP node are calculated individually, it provides an insight about how the weights for the different nodes evolve. It can be appreciated that this evolution is very similar to that of a single node shown in the Fig. 6. The mean weight for the three day types and the 7 VPP nodes evolves from an initial weight of $\omega_{ANN} = 0.10$ to approximately $\omega_{ANN} = 0.60$ when the total historical dataset is completed.

Finally, Fig. 11 shows the GHI forecasting for the entire VPP and for each method considering three days of the historical dataset: from 14th to 16th March 2020. These days are representative of the three types of day considered in the hybrid method. As in the case of one VPP node, it can be seen that the error produced in all the forecasting methods increases as the cloudiness grows. However, it can be observed to what extent, better forecasts are obtained with the proposed methods.

TABLE 2. Quantification of MAPE and NRMSE after 180 days of historical data.

Method	Sites	University	Ctro. Mpal. de Acustica	Hortaleza	Juan Carlos I	Moratalaz	Peñagrande	Villaverde	Mean	**VPP**
Similar hours	MAPE (%)	25.27	25.40	25.71	24.28	25.00	24.63	24.37	24.95	**21.64**
	NRMSE (%)	36.51	37.26	39.15	36.60	36.45	36.52	37.11	37.09	**31.69**
ANN	MAPE (%)	24.76	24.63	25.85	24.09	24.46	24.48	24.04	24.06	**22.48**
	NRMSE (%)	35.05	35.15	37.18	34.71	34.76	35.38	34.74	34.74	**32.21**
Hybrid	MAPE (%)	23.98	23.92	24.71	23.18	23.67	23.53	23.27	23.75	**21.37**
	NRMSE (%)	34.21	34.57	36.90	34.38	34.12	34.42	34.59	34.74	**30.99**

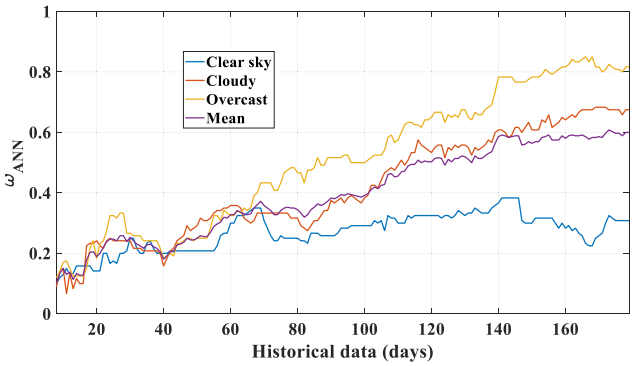


FIGURE 10. ANN mean weight for the VPP as a function of the day type and the historical data. It can be seen that weights give progressively more importance to the forecast provided by the ANN.

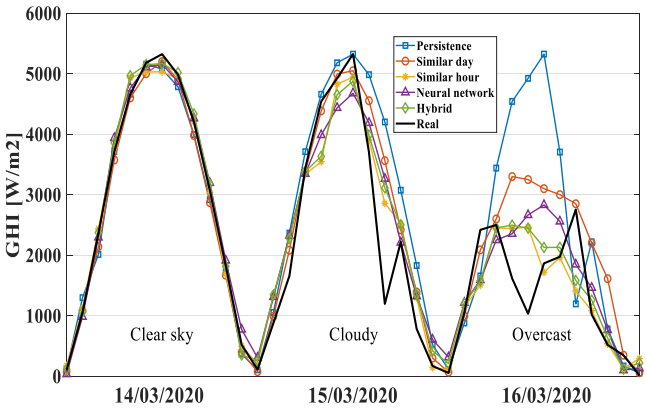


FIGURE 11. GHI accumulated forecasting in the VPP for the different methods and for three consecutive days, showing different weather conditions.

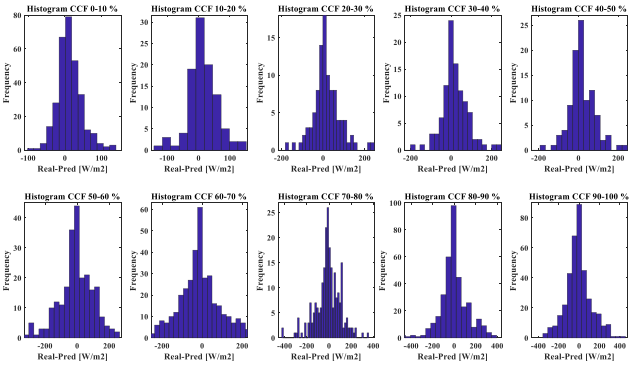


FIGURE 12. Distribution of the error for every subset considered. To create prediction intervals in the forecasting strategy, a Laplace distribution is assumed.

TABLE 3. Prediction intervals (p_s) for each subset and for the whole dataset, and their PICP.

CCF	MAE $[W/m^2]$	p_s $[W/m^2]$	*PICP*
0 - 10 %	34.02	±31.17	0.63
10 - 20%	44.08	±40.39	0.61
20 - 30 %	54.73	±50.15	0.60
30 - 40 %	58.52	±53.62	0.62
40 - 50 %	72.41	±66.35	0.58
50 - 60 %	80.87	±74.10	0.59
60 - 70 %	87.18	±79.88	0.55
70 - 80 %	89.81	±82.29	0.59
80 - 90 %	92.21	±84.49	0.60
90 - 100 %	91.67	±84.00	0.58
WHOLE DATASET	75.96	±69.60	0.62

C. UNCERTAINTY QUANTIFICATION

It is very important to specify a probabilistic range for the predictions, in order to assess the degree of uncertainty. To this purpose, in this section, statistical prediction intervals are considered based on the work carried out in [49].

Firstly, the dataset is split into 10 subsets as a function of the CCF. Looking at the error distribution of the hybrid-based forecast in Fig. 12, a Laplacian distribution can be reasonably assumed for each subset. Secondly, under this assumption, a prediction interval for each subset is defined, $I_{pred} \pm p_s$, in terms of the MAE, and the percentile p of probability $(1-s)$ is considered, knowing that $p_s = \pm MAE \cdot \ln(2s)$ for a Laplacian distribution.

In this particular case, the reliability of the prediction interval under a confidence value of 60%($s = 0.2$), is evaluated. The prediction interval, p_s, which is determined as a function of the MAE of each subset, is then calculated. With this interval, the Prediction Interval Coverage Probability (PICP) [50], can be worked out. The PICP indicates the percentage of values that are inside the interval, and it needs to be close

to the confidence level.

$$PICP = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t, where \epsilon_i = \begin{cases} 1 & if\ x_i \in [L_i, U_i] \\ 0 & if\ x_i \notin [L_i, U_i] \end{cases} \quad (7)$$

In Table 3, it can be observed that the PICP is close to the confidence level for every subset. It can also be appreciated that the prediction interval increases with the presence of clouds, showing that overcast days are the most difficult days to forecast.

VI. CONCLUSION

Solar irradiance forecasts are of paramount importance for the integration of PV systems in a VPP in an effective way. However accurate irradiance predictions generally rely on extensive datasets, which are not always available when a VPP begins operating or when a new VPP is first integrated. Furthermore, data access usually carries a cost, which is driven up as the number of VPP increases leading to a decline in profits. There is not a simple way to overcome these limitations with only one approach which performs efficiently irrespective of the dataset size. For this reason, this paper presents a hybrid approach comprising two methods based on similar hours and ANNs. The outputs of both forecasting methods are dynamically weighted, according to the type of the day (sunny, cloudy and overcast) and the MAE. The proposed forecasting approach uses temperature and cloudiness forecast maps freely generated by the AEMET via NWP along with irradiance measurements obtained from both a real PV installation located in the Polytechnic School of Alcala University and a group of different ground-based meteorological stations operating in the Community of Madrid (Spain). Both, the similar hour-based approach and the hybrid method have demonstrated better performance than widely employed forecasting techniques, namely persistence method, and similar day-based and ANN-based approaches, when limited historical data is available. For a 7-node VPP configuration and for a 6-month period of historical data, a MAPE of 21.64% and a NRMSE of 31.69% for the similar hour-based technique, and a MAPE of 21.37% and a NRMSE of 30.99% for the hybrid strategy are obtained.

Under a reduced historical dataset, the results show that the proposed similar hour-based method produces the best forecasts relative to those obtained by the ANN-based approach. For one-month and two-month datasets the mean error is reduced by 16.32% and 9.07% respectively. Finally, to demonstrate the potential of the proposed approach, a comparative analysis between the hybrid method and the most commonly used benchmarks in the literature, namely, the persistence method and the method based on similar days, has been carried out. It has been concluded that the proposed model yields promising results regardless the length of the historical dataset.

Future work will address the estimation of the power generated by the PV facilities within the structure of the VPP. To this end, the forecasting techniques presented in this paper will be used, weighting each station according to the rated power. Finally, as the historical dataset of the installation increases in length, the computational time of the algorithm will grow in importance, augmenting the interest in the implementation of advance optimization techniques for some steps in the algorithm such as the calculation of the weights.

REFERENCES

[1] (Oct. 2019). *Market Report Series: Renewables 2019. Analysis and forecasts for 2024*. International Energy Agency (IEA). Paris, France. Accessed: Sep. 30, 2020. [Online]. Available: https://webstore.iea.org/renewables-2019
[2] M. Perez, R. Perez, K. R. Rabago, and M. Putman, "Overbuilding & curtailment: The cost-effective enablers of firm PV generation," *Sol. Energy*, vol. 180, pp. 412–422, Jan. 2019.
[3] B. Lin and J. Zhu, "The role of renewable energy technological innovation on climate change: Empirical evidence from China," *Sci. Total Environ.*, vol. 659, pp. 1505–1512, Apr. 2019.
[4] R. Shah, N. Mithulananthan, R. C. Bansal, and V. K. Ramachandaramurthy, "A review of key power system stability challenges for large-scale PV integration," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 1423–1436, Jan. 2015.
[5] F. Falko, R. Brecha, and G. Luderer, "Analyzing major challenges of wind and solar variability in power systems," *Renew. Energy*, vol. 81, pp. 1–10, Sep. 2015.
[6] X. Li, D. Hui, and X. Lai, "Battery energy storage station (BESS)-based smoothing control of photovoltaic (PV) and wind power generation fluctuations," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 464–473, Apr. 2013.
[7] M. Katsanevakis, R. A. Stewart, and J. Lu, "Aggregated applications and benefits of energy storage systems with application-specific control methods: A review," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 719–741, Aug. 2017.
[8] S. Grillo, M. Marinelli, S. Massucco, and F. Silvestro, "Optimal management strategy of a battery-based storage system to improve renewable energy integration in distribution networks," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 950–958, Jun. 2012.
[9] S. Ghavidel, L. Li, J. Aghaei, T. Yu, and J. Zhu, "A review on the virtual power plant: Components and operation systems," presented at the IEEE Int. Conf. Power Syst. Technol. (POWERCON), Sep. 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7754037
[10] H. Saboori, M. Mohammadi, and R. Taghe, "Virtual power plant (VPP), definition, concept, components and types," presented at the Asia–Pacific Pow and Ener. Engr. Conf., Mar. 2011. [Online]. Available: https://ieeexplore.ieee.org/document/5749026
[11] D. Pudjianto, C. Ramsay, and G. Strbac, "Virtual power plant and system integration of distributed energy resources," *IET Renew. Power Gener.*, vol. 1, no. 1, pp. 10–16, Mar. 2007.
[12] G. Ritzer and N. Jurgenson, "Production, consumption, prosumption: The nature of capitalism in the age of the digital 'prosumer,'" *J. Consum. Culture*, vol. 10, no. 1, pp. 13–36, Mar. 2010.
[13] N. Etherden, V. Vyatkin, and M. H. J. Bollen, "Virtual power plant for grid services using IEC 61850," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 437–447, Feb. 2016.
[14] A. Tascikaraoglu, O. Erdinc, M. Uzunoglu, and A. Karakas, "An adaptive load dispatching and forecasting strategy for a virtual power plant including renewable energy conversion units," *Appl. Energy*, vol. 119, pp. 445–453, Apr. 2014.
[15] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Appl. Sci.*, vol. 10, no. 2, p. 487, Jan. 2020.
[16] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manage.*, vol. 156, pp. 459–497, Jan. 2018.
[17] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78–111, Oct. 2016.
[18] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renew. Sustain. Energy Rev.*, vol. 75, pp. 242–263, Aug. 2017.
[19] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy Combustion Sci.*, vol. 39, no. 6, pp. 535–576, Dec. 2013.

[20] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, Jan. 2018.

[21] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, and J. Sumita, "Forecasting electric power generation in a photovoltaic power system for an energy network," *Electr. Eng. Jpn.*, vol. 167, no. 4, pp. 16–23, Jun. 2009.

[22] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in PV output power forecast," *Sol. Energy*, vol. 136, pp. 125–144, Oct. 2016.

[23] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, Dec. 2015.

[24] *Spanish Agency of Meteorology (AEMET) Numerical Weather Predictions*. Accessed: Sep. 30, 2020. [Online]. Available: http://www.aemet.es/es/eltiempo/prediccion/modelosnumericos/harmonie_arome_ccaa?opc2=mad#

[25] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," *Renew. Energy*, vol. 132, pp. 871–884, Mar. 2019.

[26] C. Furlan, A. P. de Oliveira, J. Soares, G. Codato, and J. F. Escobedo, "The role of clouds in improving the regression model for hourly values of diffuse solar radiation," *Appl. Energy*, vol. 92, pp. 240–254, Apr. 2012.

[27] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2253–2262, Sep. 2015.

[28] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combustion Sci.*, vol. 34, no. 5, pp. 574–632, Oct. 2008.

[29] F. Wang, Z. Mi, S. Su, and H. Zhao, "Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters," *Energies*, vol. 5, no. 5, pp. 1355–1370, May 2012.

[30] R. Marquez, V. G. Gueorguiev, and C. F. M. Coimbra, "Forecasting of global horizontal irradiance using sky cover indices," *J. Sol. Energy Eng.*, vol. 135, no. 1, pp. 011017-1–011017-5, Feb. 2013.

[31] B. Amrouche and X. Le Pivert, "Artificial neural network based daily local forecasting for global solar radiation," *Appl. Energy*, vol. 130, pp. 333–341, Oct. 2014.

[32] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018.

[33] Y.-Y. Hong, J. J. F. Martinez, and A. C. Fajardo, "Day-ahead solar irradiation forecasting utilizing Gramian angular field and convolutional long short-term memory," *IEEE Access*, vol. 8, pp. 18741–18753, 2020.

[34] C. H. Fajardo, J. Mula, and R. Poler, "Adaptive and hybrid forecasting models—A review," *Engineering Digital Transformation*. Cham, Switzerland: Springer, Sep. 2018, pp. 315–322.

[35] Z. Hajirahimi and M. Khashei, "Hybrid structures in time series modeling and forecasting: A review," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 83–106, Nov. 2019.

[36] A. Teke, H. B. Yıldırım, and Ö. Çelik, "Evaluation and performance comparison of different models for the estimation of solar radiation," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1097–1107, Oct. 2015.

[37] O. Behar, A. Khellaf, and K. Mohammedi, "Comparison of solar radiation models and their validation under Algerian climate—The case of direct irradiance," *Energy Convers. Manage.*, vol. 98, pp. 236–251, Jul. 2015.

[38] L. Bengtsson, U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, and K. I. Ivarsson, "The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system," *Monthly Weather Rev.*, vol. 145, no. 5, pp. 1919–1935, Apr. 2017.

[39] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, Mar. 2012.

[40] C. Huang, N. Thomas, S. N. Goward, J. G. Masek, Z. Zhu, J. R. G. Townshend, and J. E. Vogelmann, "Automated masking of cloud and cloud shadow for forest change analysis using landsat images," *Int. J. Remote Sens.*, vol. 31, no. 20, pp. 5449–5464, Oct. 2010.

[41] S. Le Hégarat-Mascle and C. André, "Use of Markov random fields for automatic cloud/shadow detection on high resolution optical images," *ISPRS J. Photogramm. Remote Sens.*, vol. 64, no. 4, pp. 351–366, Jul. 2009.

[42] Z. Zhu and C. E. Woodcock, "Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change," *Remote Sens. Environ.*, vol. 152, pp. 217–234, Jul. 2014.

[43] Y. Huang, S. Siems, M. Manton, A. Protat, L. Majewski, and H. Nguyen, "Evaluating Himawari-8 cloud products using shipborne and CALIPSO observations: Cloud-top height and cloud-top temperature," *J. Atmos. Ocean. Technol.*, vol. 36, no. 12, pp. 2327–2347, Dec. 2019.

[44] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, Nov. 2013.

[45] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.

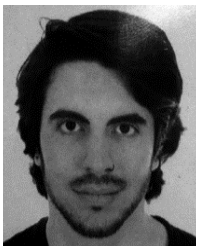
[46] S. Haykin, "Multilayer Perceptrons," in *Neural Networks and Learning Machines*, 3rd ed. Hamilton, ON, Canada: Prentice-Hall, 2009, p. 175.

[47] E. Akarslan and F. O. Hocaoglu, "A novel method based on similarity for hourly solar irradiance forecasting," *Renew. Energy*, vol. 112, pp. 337–346, Nov. 2017.

[48] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi, and A. K. Srivastava, "A novel approach to forecast electricity price for PJM using neural network and similar days method," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2058–2065, Nov. 2007.

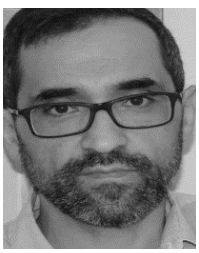
[49] J. Fonseca, T. Oozeki, H. Ohtake, T. Takashima, and O. Kazuhito, "On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation," *J. Electr. Eng. Technol.*, vol. 10, no. 1, pp. 30–40, May 2015.

[50] D. W. van der Meer, J. Widén, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," *Renew. Sustain. Energy Rev.*, vol. 81, no. 1, pp. 1484–1512, Jan. 2018.



GUILLERMO MORENO was born in Tarifa, Cadiz, Spain, in 1990. He received the University degree in industrial engineering technology from the University of Malaga, Malaga, Spain, in 2015, and the master's degree in industrial engineering from the University of Alcala, Madrid, Spain, in 2018.

Since 2017, he has been a Research Assistant with the Electronics Engineering Department, University of Alcala. His research interests include design of photovoltaic installations, photovoltaic power forecasting, and virtual power plants management and operation.



PEDRO MARTIN received the Ph.D. degree in telecommunications engineering from the University of Alcala, Spain, in 2000, and the degree in telecommunications from the Polytechnic University of Madrid, Spain, in 1994.

He currently works as an Associate Professor with the Department of Electronics, the University of Alcala. He is a member of the research group Electronics Engineering applied to the Renewable Energies. His research interets include design of embedded systems for power electronic systems and smart grids.

CARLOS SANTOS received the B.S. degree in telecommunications engineering and the M.Sc. degree in electrical engineering from the University of Alcala, Spain, in 2010 and 2011, respectively, and the Ph.D. degree in electronics in 2016.

He is currently with the Postdoctoral Research Contract, Department of Electronics, University of Alcala. His research interests include fusion algorithms, forecasting strategies, varying-time sampling control techniques, energy management systems, virtual power plants, and smart grids.

FRANCISCO J. RODRÍGUEZ (Member, IEEE) received the B.Sc. degree in technical telecommunication engineering from the University of Alcalá, Alcalá de Henares, Spain, in 1985, the M.Sc. degree in telecommunication from the Technical University of Madrid, Madrid, Spain, in 1990, and the Ph.D. degree in electronics engineering from the University of Alcalá, in 1997.

He worked in the private electronic industry for two years. Since 1986, he has been a Lecturer with the Department of Electronics, University of Alcalá, where he is currently a Professor. He is the author of more than 160 refereed publications in international journals, book chapters, and conference proceedings. Also, he has directed more than 55 investigation projects funded by public institutions and private industry. His research interests include smartgrids, microgrids, renewable energies, control electronics, and real-time processing.

ENRIQUE SANTISO received the Ph.D. degree in telecommunications engineering, in 2003.

He is currently a University Professor. Since January 1990, he has been teaching and researching with the Electronics Department, University of Alcalá. He has taught subjects related to sensing, data acquisition and processing, control, power, and digital systems. He has participated in eight published books, related to the teaching given. The teaching work was supported by the recognition of five teaching periods (five years). His research interests, which began in 1991, include mobile robotic systems (positioning and guidance), sensory systems, data acquisition, control, and power. He has directed two research projects in public calls, participated in one European project and 28 national projects. He has participated in 38 research and development projects with companies. He has 20 international articles and 41 participations in conferences. His research work has been recognized with the award of five research sections (six-year periods). He is also focusing his research on the management of electrical energy and the optimization of production systems.