Contents lists available at ScienceDirect

Solar Energy

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

Review

Review of photovoltaic power forecasting

J. Antonanzas [a,*], N. Osorio [b], R. Escobar [b,c], R. Urraca [a], F.J. Martinez-de-Pison [a], F. Antonanzas-Torres [a]

[a] EDMANS Group, Department of Mechanical Engineering, University of La Rioja, Logroño, Spain
[b] Center for Solar Energy Technologies, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile
[c] Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile

A B S T R A C T

Variability of solar resource poses difficulties in grid management as solar penetration rates rise continuously. Thus, the task of solar power forecasting becomes crucial to ensure grid stability and to enable an optimal unit commitment and economical dispatch. Several forecast horizons can be identified, spanning from a few seconds to days or weeks ahead, as well as spatial horizons, from single site to regional forecasts. New techniques and approaches arise worldwide each year to improve accuracy of models with the ultimate goal of reducing uncertainty in the predictions. This paper appears with the aim of compiling a large part of the knowledge about solar power forecasting, focusing on the latest advancements and future trends. Firstly, the motivation to achieve an accurate forecast is presented with the analysis of the economic implications it may have. It is followed by a summary of the main techniques used to issue the predictions. Then, the benefits of point/regional forecasts and deterministic/probabilistic forecasts are discussed. It has been observed that most recent papers highlight the importance of probabilistic predictions and they incorporate an economic assessment of the impact of the accuracy of the forecasts on the grid. Later on, a classification of authors according to forecast horizons and origin of inputs is presented, which represents the most up-to-date compilation of solar power forecasting studies. Finally, all the different metrics used by the researchers have been collected and some remarks for enabling a fair comparison among studies have been stated.

© 2016 Elsevier Ltd. All rights reserved.

Contents

---

* Corresponding author.
E-mail address: antonanzas.javier@gmail.com (J. Antonanzas).

Nomenclature

ANFIS	Adaptive Neuro-Fuzzy	MARE	Mean Absolute Relative Error
ANN	Artificial Neural Network	MBE	Mean Bias Error
AR	Auto-Regressive	MLP	Multi-Layer Perceptron
ARIMA	Auto-Regressive Integrated Moving Average	MLR	Multivariate Linear Regression
ARMA	Auto-Regressive Moving Average	MOS	Model Output Statistics
ARX	Auto-Regressive eXogenous	MRE	Mean Relative Error
CC	Cloud Cover	NARX	Non-linear AR-eXogenous
CDF	Cumulative Density Function	nRMSE	normalized Root Mean Square Error
CMV	Cloud Motion Vectors	NWP	Numerical Weather Predictions
CRPS	Continuous Ranked Probability Score	P	PV power
DTC	Distribution Transformer Controllers	PDF	Probability Density Function
ECMWF	European Center for Medium range Weather Forecasts	PFA	Probabilistic Finite Automata
EIM	Energy Imbalance Markets	PHANN	Physical Hybrid Artificial Neural Network
ELM	Extreme Learning Machines	PSO	Particle Swarm Optimization
GA	Genetic Algorithm	PTU	Program Time Unit
GCT	Gate Closure Time	PV	photovoltaic
GFS	Global Forecasting System	QR	Quantile Regression
GHI	Global Horizontal Irradiance	QRF	Quantile Regression Forests
GMDH	Group Method of Data Handling	RF	Random Forests
ICP	Interval Coverage Probability	RH	Relative Humidity
IEA	International Energy Agency	RMSE	Root Mean Square Error
I_{ex}	extraterrestrial irradiance	SARIMA	Seasonal Auto-Regressive Integrated Moving Average
I_{POA}	Irradiance on Plane Of Array	SVM	Support Vector Machine
k-NN	k-Nearest Neighbors	SVR	Support Vector Regression
KSI	Kolmogorov–Smirnov Integral	SOM	Self Organized Map
k_t	clearness index	T	air temperature
k_{cs}	clear sky index	VAR	Vector Auto-Regression
LM	Linear Model	VARX	Vector Auto-Regression eXogenous
LS	Least Square	W	wind speed
MAE	Mean Absolute Error	WD	Wavelet Decomposition
MAID	Mean Absolute Interval Deviation	WT	Wavelet Transform
MAPE	Mean Absolute Percent Error		

1. Introduction

Recently, the 2015 United Nations Climate Change Conference (COP21), now known as the Paris Agreement, has become a milestone in fighting global warming. The 196 countries that signed the document agreed to make efforts to limit the global warming to less than $2\,°C$ with respect to pre-industrial levels, which implies reducing the anthropogenic greenhouse emissions to zero during the second half of the 21st century. Reaching those goals involves an electrification of many current thermal systems, among many other actions. This Agreement stresses the necessity of generating energy via renewable sources and motivates the research on how to manage and integrate into the grid these variable generation systems.

Focusing on solar technology, photovoltaics have experienced enormous growth over the last years, amounting to a total installed capacity of around 177 GW worldwide by the end of 2014 (IEA, 2015) and growth is projected to continue at a similar rate in the future. Moreover, photovoltaic (PV) prices have seen a strong reduction, bottoming below \$1.5/Wp for fixed-tilt systems, boosting more installations (GTM). PV has already become a key agent in some electricity markets, reaching an annual 8% of solar share in Italy or close to 7% in Germany, and the number of countries where that percentage is greater than 1% is about 20 (IEA, 2015). In this context, the high penetration of PV in electric systems poses many economic benefits, but may also threaten the stability of the power grid without accurate forecasts.

PV production mainly depends on the amount of solar global irradiation incident on the panels, but that irradiation is not uniform over time. Solar resource variability and the uncertainty associated to forecasts are behind most of the problems that must be handled to maintain the stability of the power grid. A part of the fluctuations are deterministic and explained by the rotational and translational movements of the Earth with respect to the Sun, which are accurately described by physical equations. However, there also exists unexpected changes in the amount of solar irradiance arriving at the Earth's surface, mainly derived from the presence of clouds, which stochastically block the Sun's rays and grant PV power forecasting a certain level of uncertainty.

The ability of precisely forecasting the energy produced by PV systems is of great importance and has been identified as one of the key challenges for massive PV integration (EPIA, 2012; PV GRID, 2014). It is decisive for grid operators, since deviations between forecasted and produced energy must be supplied by the rest of technologies that form the energy portfolio. Some of the units that build the electric system act as operating reserve generators. Thus, a proper PV forecast would be able to lower the number of units in hot standby and, consequently, reduce the operation costs. Table 1 depicts the flexibility of conventional power plants and the ability to respond to such deviations.

An accurate forecast is not only beneficial for system operators (and, eventually, for all customers from the grid) since it reduces costs and uncertainties, but also for PV plant managers, as they avoid possible penalties that are incurred due to deviations between forecasted and produced energy.

The importance of the issue has boosted the development of many studies worldwide to obtain accurate forecasts. Two main approaches can be found in the forecasting of PV plant production: indirect and direct. Indirect forecasts firstly predict solar irradiation and then, using a PV performance model of the plant, obtain the power produced. On the other hand, direct forecasts directly calculate the power output of the plant. Also, many other studies only focus on the prediction of solar irradiation, since it is the most difficult element to model and have other applications apart from solar power forecasting. Both forecasts (power and irradiation) are approached via similar techniques. This review paper was based on those articles that have as the output the power produced by the plants, to establish a boundary in the scope and since that variable can be directly used by grid operators and plant managers. This work is limited to the study of scientific articles; the analysis of commercial forecasting tools is out of the scope of this review.

This paper presents a complete review of the state-of-the-art techniques to produce power forecasts for photovoltaics. There are some previous review articles with also a wide scope (forecasting techniques, sources of inputs, performance metrics, temporal and spatial coverage, ...), such as the work developed by Inman et al. (2013) and IEA (2013), but the rate at which new studies are developed requires that a new review showing current trends is conducted. Some of these new trends are the focus on the economic impact of forecasting, the importance of probabilistic forecasting and the necessity of agreement for a common suite of performance metrics. Other more recent reviews are only focused on a specific aspect of forecasting, such as ensemble forecasting (Ren et al., 2015) or different forecasting techniques (Wan et al., 2015).

The article is structured in such a way that it tackles some of the issues that arise when planning a forecast, such as the necessity to issue and improve solar power forecasts, the different techniques that can be used, spatial and temporal coverage, information that should be provided, measurement of accuracy and previous work developed by other researchers.

Thus, the paper is structured as follows: Section 2 explains some basic concepts that are used throughout the paper. Section 3 sets the foundations and main motivations of the study as it talks about the importance of forecasting, showing possible economic consequences of improved forecasts. Then, Section 4 shows the

Table 1
Flexibility of conventional power generation systems. Values for Nuclear, Hard coal and Combined cycle gas are for inflexible plants. *Source:* IEA (2014).

	Nuclear	Hard coal	Lignite	Combined cycle gas	Pumped storage
Start-up time 'warm' (h)	na	5–7	2–8	2–4	<0.1
Ramp rate (%/min)	0	0.6–4	0.6–6	0.8–6	15–25
Minimal possible load (%)	100	40–60	40–60	40–50	5–6

main approaches to forecasting power output: physical, statistical or hybrid. Section 5 discusses the benefits and characteristics of forecasting for either a single PV plant or for an ensemble of them. Section 6 talks about the different options to present the forecast: a single value or a probabilistic term. It also discusses about the implications it may have for grid operation. Section 7 discusses about the different time horizons that are necessary to be taken into account for a proper grid operation. In contrast to most of review papers about solar forecasting, we have classified the studies according to the forecast horizon instead of the techniques used. Here are collated and summarized all the articles found about solar power forecasting. Finally, a review of the metrics that are used to evaluate forecasts and the convenience of each of them is given in Section 8, along with some recommendations for a better comparability of studies. Moreover, at the end of certain sections and subsections a short summary is presented, which depicts the main findings and conclusions about each topic.

2. Basic considerations

In this section some basic concepts about solar irradiation and solar power generation are explained, which will ease the comprehension of the remaining parts of the text.

2.1. Clear sky models

As mentioned above, solar irradiance is mainly influenced by the presence of clouds, whose presence difficulties irradiance predictions. However, it is possible to approximate the irradiance under clear sky conditions, that is, in the absence of clouds. Such value can be used to calculate solar indices, normalize metrics and obtain the production of a solar plant under stationary conditions. Usually, clear sky models are fed with meteorological variables and solar geometry, using Radiative Transfer Models to establish the connections between the inputs. There exists a large number of clear sky models, which differ from each other mainly in the inputs needed by each model. Some of the most widely used clear sky models are the Solis model (Mueller et al., 2004), the European Solar Radiation Atlas (ESRA) model (Rigollier et al., 2000), the Ineichen model (Ineichen and Perez, 2002) and the Reference Evaluation on Solar Transmittance 2 (REST2) model (Gueymard, 2008). For an in-depth description of these and other clear sky models, readers are referred to the aforementioned references and to Inman et al. (2013) and Badescu et al. (2013). Some models only require one input (ESRA, Ineichen) whereas others require a large number of them (Solis). As detailed in Ineichen (2006), the choice of a clear sky model for a determined location is driven by the availability and quality of input data, which is the main limiting factor. The selection of a specific model is of secondary importance. Antonanzas-Torres et al. (2016) made an assessment of the impact of atmospheric components in clear sky models at different elevations, determining their accuracy according to the spatial resolution of the inputs.

2.2. Clear sky and clearness indices

There are two parameters, the clear sky index (k_{cs}) and the clearness index (k_t), which are widely used to classify weather conditions and to calculate smart persistence models. They are obtained in a similar way, but differ in the normalization variable. The clear sky index is the ratio of measured irradiance to the modeled clear sky irradiance I_{cs}.

$$k_{cs} = \frac{I}{I_{cs}} \tag{1}$$

The clearness index is normalized with respect to the extraterrestrial irradiance I_{ex}. Thus, it avoids the difficulty of modeling the interactions of irradiance with the atmosphere.

$$k_t = \frac{I}{I_{ex}} \tag{2}$$

where $I_{ex} = I_0 \cdot \epsilon_0 \cdot cos\theta$. I_0, the solar constant, takes a value of 1360 W/m^2, ϵ_0 is the eccentricity of the ellipse described by the Earth in its movement around the Sun and θ is the solar zenith angle.

2.3. Time definitions of forecasts

Regarding the temporal aspect of forecasts, it is important to introduce three concepts: forecast horizon f_h, forecast resolution f_r and forecast interval f_i. The forecast horizon is the amount of time between present time t and the effective time of predictions. The forecast resolution describes the frequency at which the forecasts are issued and the forecast interval denotes the time range of predictions.

2.4. Origin of inputs

Studies can also be classified according to the origin of inputs. To this extent, two main approaches can be found: models that use endogenous data, formed by current and/or lagged time-series of records of the production of a PV plant, and models that use exogenous data, which may come from local measurements (temperature, relative humidity, wind speed and direction, …), information from total sky imagers, satellite images, Numerical Weather Predictions (NWP) (temperature, relative humidity, irradiance, cloud cover, wind speed and direction, pressure, etc.), values from other meteorological databases and values from neighboring PV plants.

2.5. Persistence models

Persistence models are commonly used as benchmark for more developed models since they are the simplest. They assume that conditions (irradiance, power output, clear sky index, etc.) remain the same between t and $t + f_h$. Thus, many studies present their results using the skill score (ss), which shows improvement (or worsening) with respect to the said persistence/baseline models. However, comparisons with respect to skill score between different investigations must be performed carefully due to the abundance of reference models. They can only be carried out when persistence models have the same "persistence" variable. Hereafter there is a list of the most common persistence models.

The "naive persistence" assumes that the forecasted power for the time horizon will be the same as the last value measured. For instance, for a 1 h ahead horizon, the power at 14.00 will be the same as the power at 13.00. A variation of this model would be to consider the same value as that of the previous day (or the closest day with available measurements) at the same time (Fernandez-Jimenez et al., 2012; Monteiro et al., 2013b; Lorenz et al., 2011a). These models are recommended when time series are considered stationary.

$$P_p(t + f_h) = P(t) \tag{3}$$

Since solar irradiance time-series are not stationary, naive persistence is normally limited to intra-hour applications. To overcome this problem, other approach has been proposed, which can be applied not only at longer time horizons but also in intra-hour applications. It consists of decomposing solar power production in a stationary and in a stochastic component. The stationary

term is normally associated to the clear sky production and the stochastic term, to the cloud-induced changes (Coimbra and Pedro, 2013).

$$P(t) = P_{cs}(t) + P_{st}(t) \tag{4}$$

where $P_{cs}(t)$ is the expected power output under clear sky conditions and $P_{st}(t)$ represents the stochastic term.

These approaches derive in the so called "smart persistence", since a more stationary variable is selected, such as the clear sky index, the persistence of cloudiness or the ramp persistence.

Pedro and Coimbra (2012) detailed the following method.

$$P_p(t + f_h) = \begin{cases} P_{cs}(t + f_h) & \text{if } P_{cs}(t) = 0 \\ P_{cs}(t + f_h)\frac{P}{P_{cs}(t)} & \text{otherwise} \end{cases} \tag{5}$$

In low variability periods and short time horizons, P_{cs} is very accurate.

Coimbra and Pedro (2013) also showed another smart persistence model based on the above-mentioned decomposition of solar power production, keeping the stochastic part of the time series unvaried.

$$P_p(t + f_h) = P_{cs}(t + f_h) + P_{st}(t) \tag{6}$$

Zhang et al. (2015a) preferred a persistence of cloudiness approach for their forecasts.

$$P_p(t + f_h) = P(t) + SPI_t\{P_{cs}(t + f_h) - P_{cs}(t)\} \tag{7}$$

where SPI_t is the solar power index (the ratio of actual power to clear sky power).

For short time horizons, the ramp persistence was used (Lipperheide et al., 2015). It extends the variation in power output over the last second to persist over the forecast horizon.

$$P_p(t + f_h) = P(t) + f_h\{P(t) - P(t - 1s)\} \tag{8}$$

3. The economics of forecasting

The main purpose of improving the accuracy of solar power forecasts is to reduce the uncertainties related to this type of variable energy source, which would directly result in a safer and easier grid management. Moreover, curtailment applied to photovoltaics could be reduced (Bird et al., 2014). Plant managers also find motivation in issuing better predictions as they can better plan maintenance stops and generate more precise bids. As solar penetration increases in the energy portfolio, the impact of incorrect forecasts in the grid can be larger. Thus, in some electricity markets, solar producers can face penalties when deviations between forecasted and produced energy exceed a tolerance band. Deviation penalties promote issuing accurate forecasts to maximize the revenue. As pointed out in Mathiesen et al. (2013), under certain market situations (i.e. when the locational marginal prices of the real time market are five times higher than the day ahead market prices), an overprediction of just 10% would offset the economical benefit due to the aforementioned deviation costs. To deepen in the understanding of such penalties, readers are referred to Jacobs (2012), Botterud et al. (2012) and Mathiesen et al. (2013).

In spite of the important consequences forecast accuracy may pose, only a few studies have addressed the influence of the said accuracy on grid operation. Also, not all electricity grids react in the same manner to improved forecasts. Behind the scarcity of studies the following reasons can be found: complex power system modeling, difficulty of allocating costs or benefits, relatively low solar penetration into energy portfolio, poor understanding on how system operators can use the information provided and variety of trading systems (Brancucci Martínez-Anido et al., 2014; Zhang et al., 2015b). Nevertheless, penetration of solar energy into

the grid has begun to grow and is expected to become an important agent in the energy portfolio. Thus, it is important to understand the implications forecasts have. Utility companies, as well as distribution system operators, independent system operators, etc. can profit from solar forecasts. Another interesting aspect about solar integration is the variety of solar power producers, from utility-scale to distributed plants. As discussed in Section 5, aggregated distributed PV shows less variability than single plants. Nevertheless, most of distributed PV plants are "behind the meter", so that grid operators only perceive their impact as a reduction in demand. This can cause errors in load forecasting because grid managers ignore the exact amount of distributed PV (Mills et al., 2013). Thus, in the future, the task of load forecasting and solar power forecasting will fuse into net demand forecasting (Hong et al., in press).

The variability of solar energy and the uncertainty of the forecasts indirectly lead to the existence of operating reserves in electric systems in order to resolve differences between demand and production. Variability of the solar resource causes ramp events, which can not be followed by large plants or considered by trading blocks (with a time unit of 15 min). Moreover, the cloud induced changes may result in additional power needs since they can alter the supply–demand profile. However, this is not automatic: a well forecasted cloud induced power variation would not imply additional power reserve unless the time gradient was high. Also, variability of solar resource is behind some challenges that arise on grid operation with increasing levels of solar penetration. During sunrise, traditional generation plants must apply downward ramps to allow solar penetration, whereas during sunset, those ramps are positive to supply the energy that was previously coming from the Sun. Additionally, during midday, when solar production is at its maximum, conventional plants must be able to largely reduce their generation capacity. If any of these cases is not met, solar production would need to be curtailed to levels where the stability of the power system is not at risk (Brancucci Martinez-Anido et al., 2016).

What follows next is a brief summary of the different actions taken during operation and a classification of the reserves. Starting from the longest time horizon, the first actions addressed are unit commitment and scheduling, which are undertaken to cover the general load pattern of the day. With a lower time horizon, in the range of tens of minutes to hours, load following is done to offset deviations from the scheduled load pattern. It is normally addressed by economic dispatch and it may derive in starting and stopping quick-start generation systems. At shorter time horizons, we find regulation, which consists of balancing the variations in the supply/demand curve in the range of seconds to a few minutes. Allocated to each action, we can find different types of reserves: regulating, following and contingency reserves (primary, secondary and tertiary) and ramping reserves (secondary and tertiary) (Ela et al., 2011).

One of the first studies that focused on the economic implications of solar forecasting was the worked developed by Mills et al. (2013). There, they analyzed a single power system (the Arizona Public Service Company) and studied the required reserves to maintain stability and optimal commitment and economic dispatch introducing renewable energies. Day ahead and hour ahead markets were detailed, as well as a high PV and a low PV scenario. Due to the existence of base load nuclear plants, in the case with high PV production and low demand, the system posed severe operational challenges. The system modeled was stand-alone and no trading with neighboring utilities was considered. They neither considered probabilistic forecasts. They created a scenario where nuclear plants were flexible to see how the system could be improved. With no flexibility, the curtailment peaked at 17.8%, while introducing flexibility made it drop to 3.4%. A sensitivity analysis showed that integration costs ranged from $1 to $4.4

per MWh-PV for the cases considered. The major part of the costs were due to the maintenance of balancing reserves in the hour ahead market. To stress the need of flexibility in power systems, they ran a worst case scenario with constant nuclear power, low ramp rates and penalties on renewable curtailment, obtaining integration costs of $9.6 per MWh-PV. They concluded that for a practical PV integration, trading with other buyers is essential, as well as flexibility in thermal plants.

A different approach was introduced by Cormode et al. (2014), who also focused their study in the state of Arizona. They proposed the economic evaluation of forecasts, from the plant manager perspective, based on solar power curtailment, in order to avoid penalties derived from violations of Ramp Rate Rules (RRR). RRR were defined as a variation of 10%/min of the rated plant capacity and the cost of violations was stated as $0.1 per s and MW. They analyzed several forecasting scenarios: curtailment of expected clear sky production to certain levels, forecasts from NWP models, short-term predictions from a set of irradiance sensors, a kind of persistence method which limited upward ramps to the RRR (easy-up model) and a perfect forecast to set the limits. They evaluated scenarios based on the net revenues, which were obtained after discounting the penalties from the gross revenues. Thus, a trade-off appears between energy output and costs from violations of the RRR, where results are highly dependent on the said penalties. Results showed that a free running plant (no curtailment applied) would see the gross revenue reduced to 80% due to penalties. The best performing model was the easy-up model, whose net revenues corresponded to 90% of the gross incomes. The benefits of the model could be further enhaced if predictions from the set of irradiance sensors were used to trigger downward ramps.

De Georgi et al. (2015) also adopted a plant manager point of view and assessed the impact of forecasting accuracy on imbalance costs in the Italian electricity market. Bids with the predicted energy have to be made in the Day-Ahead Market and penalties apply when there are hourly deviations between actual and scheduled injected energy. If produced energy falls in the tolerance band [−10%, 10%], the producer perceives the incomes from the energy produced at the price established. Nevertheless, when mismatches exceed that tolerance, actions are taken. If they underestimate their production (they produce more than scheduled), the difference in energy between scheduled and produced is paid at a lower value, whereas if they overestimate their production (they produce less than scheduled) they have to repay the missing energy. The three possible situations are described below:

$$F(i) = \begin{cases} C_E \cdot E_T(i) & \text{if } |E_P - E_T| < 10\%E_P \\ C_E \cdot E_T(i) + (C_E + C_I) \cdot (E_T(i) - E_P(i)) & \text{if } E_P > E_T \\ C_E \cdot E_T(i) + (C_E - C_I) \cdot (E_T(i) - E_P(i)) & \text{if } E_P < E_T \end{cases}$$

(9)

where $F(i)$ is the money received, C_E, the energy price, C_I, the penalty, E_T, the actual energy, and E_P, the energy declared in the schedule.

They showed that the probability to inject the energy within the tolerance band was around 40% for the Least Square (LS) Support Vector Machine (SVM) and LS-SVM Wavelet Decomposition (WD) models (see Section 7). The economic incomes were calculated by normalizing the benefits with the maximum possible economic revenue (having all the production within the tolerance). They concluded that the LS-SVM WD model permitted the highest incomes among the models studied, with a 76% with respect to the maximum, dropping that amount to 53% for the other models. It was due to its higher accuracy and its tendency to underestimate the power output.

Brancucci Martínez-Anido et al. (2014) further deepened the investigations of economic implications considering the impact of

improving solar forecasts on a basis of 25% of solar integration. They modeled the Independent System Operator-New England (ISO-NE) system. Several scenarios were analyzed: no solar power, no solar power forecasting, current solar power forecasting, cases where improvements on forecasts of 25%, 50%, 75% were performed and a perfect forecast. Time horizons used were 4 h and day ahead. Results showed that even in the case with no solar forecasts, net generation costs were reduced by 12.3%. When they were taken into account, the reduction rose to 22.9%. This difference is explained by a more efficient commitment of conventional plants, which allowed a reduction in solar power curtailment from 34.5% to 11%. An improvement of 25% in the solar power forecasts would lead to a reduction in net generation costs of 1.56%, compared to the baseline forecasts. The improvement came from a more efficient commitment of power plants, not from a lowering on solar power curtailment. They conclude that the remaining scenarios (50%, 75% improved forecasts and a perfect forecast) did not provide any significant reduction in net generation costs in their system. This is in part due to the existence of large gas-fired plants in the ISO-NE, which can easily adapt their loads to overcome possible forecast errors. However, they highlighted that improved hour ahead and sub-hourly forecasts could still lead to cost reductions due to a better economic dispatch. Regarding electricity prices, base forecasts could lower the price of no solar by 7%, seeing only slight reductions afterwards for the improved forecasts scenarios. Solar energy introduces high variability in electricity prices: if no forecasts are taken, solar energy increases 40% price variability, while the increment climbs up to 116% if baseline forecasts are used (small variation with respect to improved forecasts). The relatively low variability when no forecasts were used was due to an overcommitment of generation units which limited the amount of solar energy into the grid.

Similarly, Brancucci Martinez-Anido et al. (2016) continued with the study of the ISO-NE, but this time they worked on day ahead forecasts. They focused on a twofold goal: determining the effects on bulk power systems of increasing solar power penetration and solar power forecasting improvements. In contrast to the earlier study, here they selected solar penetration levels of 4.5% (actual case) and some multiples (9%, 13.5% and 18%). They observed that with increasing solar power on the grid, some conventional power plants were displaced, especially gas combined cycles and gas and oil steam turbines (ST) (reductions up to 46% for 18% solar penetration). Moreover, they did not only reduce their production, but they also changed their operation, increasing the ratio of ramping to generation. However, gas and oil gas turbines (GT) and internal combustion (IC) generators increased their share due to their capacity of responding to fast ramp changes. Solar penetration also increased the starts and shutdowns of conventional plants. Gas combined cycles, which reduced their production 46% with 18% of solar share, saw their start and shutdown costs more than doubled. Moreover, no solar curtailment appeared for the 4.5% penetration, while it increased to 9.5% for the 18% penetration scenario. As for the impact of improved forecasts, the four scenarios from the previous study were considered. Increasing accuracy lead to reductions in solar curtailment, upward ramping of conventional power plants and in start and shutdown costs. Cost savings would climb up to $0.77 per MWh for the 13.5% solar share and 50% of forecast improvement, which translates to over $13.22M annually (subjected to fuel prices). Cost savings were derived from decreasing fuel and variable operation and maintenance costs.

Contrary to previous studies, which analyzed how improved forecasts could lower balancing reserves, the work by Zhang et al. (2015b) focused on target values to allow a certain reduction in reserves. They investigated 4 forecast horizons: day ahead, 4 and 1 h ahead and 15 min ahead. The baselines proposed were the persistence of cloudiness, for horizons between 0 to 4 h, and for longer
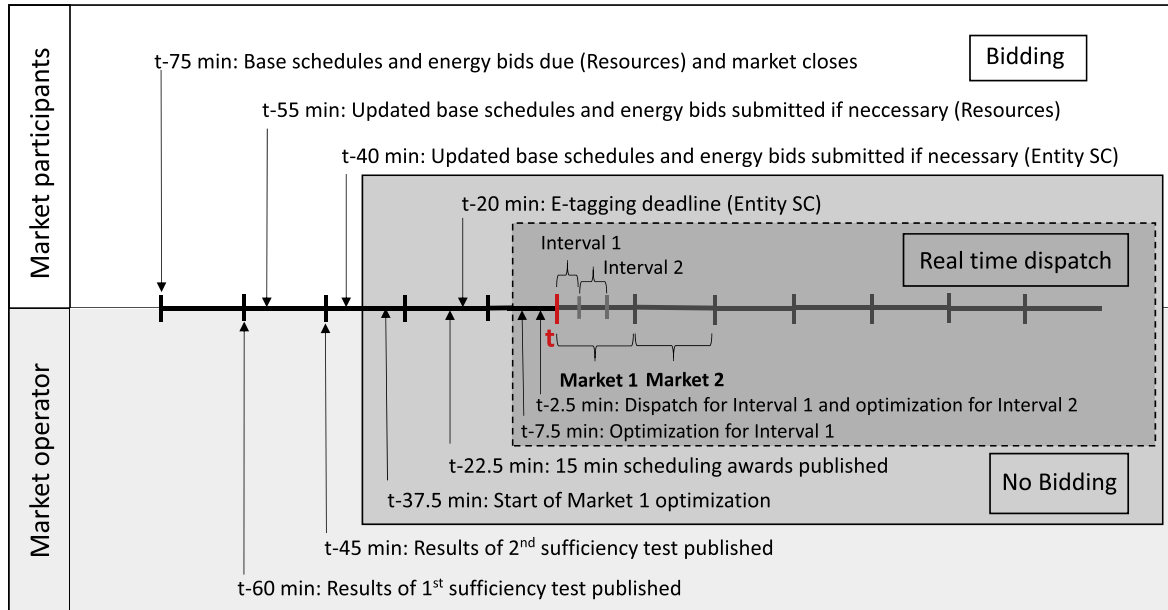
Fig. 1. Schedule of the western interconnection EIM. *t* represents the start of the operating hour (only in this figure). *Source:* Kaur et al. (2016).

lead times, the North American Mesoscale (NAM) forecast system. With these models they predicted irradiance and then, using the PV-Lib toolbox (Stein, 2012), they converted it into power. The methodology to obtain the target metrics is as follows: calculate the cost of reserves for the baseline forecasts; establish a reduction (in the case of study, 25%) in reserves to set the target reserve. With the focus on that target value, run several iterations with improved forecasts, both for ramping (%y) and non-ramping periods (%x), and calculate the difference between the new reserve thus obtained and the target reserve. Then, with the combination of forecasting improvements for both periods leading to the minimum difference, obtain the set of target metrics. For the 15 min ahead, 1 and 4 h ahead forecasts, the spinning reserves were assumed to be lowered, since they are capable of responding to changes in that period of time. They calculated them as the 95% confidence interval of solar power forecast errors at the respective lead times. However, in the day ahead forecast, both spinning and non-spinning reserves were used to obtain the targets. The spinning reserve was defined as the 70% confidence interval of the day ahead solar power forecast errors, whereas the non-spinning reserve was stated as the difference between the 95% and 70% confidence interval of the day ahead solar power forecast errors. They analyzed several cases, covering point and regional forecasts and different climate zones. For each of them they showed the baseline and target metrics to achieve the said reduction in reserves. To highlight the importance of the outcomes, they detailed that a 25% reduction in California Independent System Operator (CAISO) day-ahead spinning reserves would lead to annual savings of $5 million. Moreover, as said in Hodge et al. (2015), reserve savings only represent around 5–10% from what could be obtained from optimum unit commitment and economic dispatch, so possibilities for improvement are large.

Finally, Kaur et al. (2016) established the benefits of using improved solar power forecasts in comparison to baseline models in Energy Imbalance Markets (EIM). The objective of EIM is to reduce the use of ancillary services or additional reserves to balance the load/demand curve by permitting energy trading between interconnected balancing areas. They benefit from the different demand and generation profiles across the EIM area to reduce the required reserve capacity, improve reliability and reduce costs. An in-depth study of the Western Interconnection EIM was performed, analyzing all relevant time horizons for market operators and participants. A representation of the timeline for the real-time operation can be found in Fig. 1.

They studied lead times of 24 h, 75 and 5 min, with time intervals and resolutions of 1 h, 15 and 5 min. Models considered were the naive persistence, the smart persistence (SP) (corrected for clear sky solar irradiance), a Support Vector Regression (SVR) based model optimized with Genetic Algorithms (GA) and a reforecast model based on a linear regression to correct structured errors and bias in forecasts. They used the n-sigma method for determining the reserves, which takes into account the standard deviation in generation and demand. Reserves are influenced by the Gate Closure Time (GCT) (which defines the time when the option to submit or modify a bid expires) and the Program Time Unit (PTU) (the time window for which bids are submitted). They classified reserves depending on GCT. For GCT > 75 min, the scheduled relative reserve was:

$$R_r = \frac{\sigma(e)}{GHI} \tag{10}$$

where $\sigma(e)$ is the standard deviation of forecast errors for a given PTU and Global Horizontal Irradiance (GHI). They used an irradiation driven approach, although they claimed that in practice, power measurements should be used. The model fed with the reforecasted NWP permitted a reduction of 17.84% in the required reserves compared to the SP. However, for GCT < 75 min they calculated flexibility reserves, which are a function of change in standard deviation with respect to the magnitude of solar power production.

$$\sigma(e)^{f_r} \geqslant max_l(\sigma(e_l)) \tag{11}$$

where $\sigma(e)^{f_r}$ is the standard deviation in *n* forecast errors for a given PTU with a forecast resolution f_r. *n* is ratio between the f_i and f_r and $l \in \{1, 2, \ldots, n\}$. At this time frame, SVR-GA outperformed both *naive persistence* (reductions of 22.62–67.47%) and SP (average reductions of 21%) with respect to flexibility reserves. Biggest improvements were shown for forecast intervals of 5 min. Also, penalties derived from wrong forecasts (defined as errors larger than 7.5%) and imbalances were studied, determining that the proposed SVR-GA method reduces by 66.93% and 19.65% the probability of imbalance compared to *naive* and SP, respectively. They highlighted that flexibility in short time

operation can reduce the amount of reserves needed and the probability of imbalance in the network.

3.1. Summary

The improvement of solar power forecasts leads to the reduction of net generation costs in the electricity system and to the reduction of curtailment applied to photovoltaics. To foster these benefits it is important to allow trading with other markets and to establish flexibility in thermal plants. Huge potential for operation costs reduction still exists and the hourly and sub-hourly domains still have to be analyzed. It has been shown that the economic consequences of improved forecasts depend on the type of generation plants present in each specific market. Improved forecasts also permit plant managers to maximize their incomes.

4. Forecasting techniques

As presented in the Introduction, there are two main approaches for solar power forecasting. The first option consists in using analytical equations to model the PV system. Normally, most efforts are dedicated to obtain accurate irradiance forecasts, which is the main factor related to the power production. This approach is denoted as PV performance, physical, parametric or "white box" method. Contrarily, the second option consists in directly predicting the power output using statistical and machine learning methods. Additionally, forecasts can also be made based on a mix of both methods, which is denoted as hybrid model or "grey box". Most of the studies found used a direct approach.

4.1. PV performance models

Strictly speaking, the physical conversion of GHI or Irradiance on Plane Of Array (I_{POA}) into power output is not a forecasting technique on its own. The forecasting effort has been previously made in the prediction of irradiance and other necessary variables, such as temperature or wind. Predictions of irradiance follow similar patterns to those of power, inasmuch as comparable techniques are found depending on the time horizon. Temperature and wind predictions normally come from NWP models. In this section we will only give a general overview of PV performance models. We will not cover the different irradiance forecasting techniques found in literature. For that, readers are referred to some reviews on irradiance forecasting (Inman et al., 2013; Diagne et al., 2013; Gueymard and Ruiz-Arias, 2016; Urraca et al., 2016).

The main advantage of the PV performance method over the statistical is that, as no historical data is needed, it is possible to obtain the power output of a plant prior to construction. Knowing the technical specifications of the plant and NWP, the power output can be stated. Nevertheless, the major disadvantage of parametric models is the high dependence on NWP, which lack sufficient spatial and temporal resolution and have been reported as one of the main sources of error of this approach (Brabec et al., 2011; Dolara et al., 2015b; PVCROPS, 2015). For instance, whereas an error of 1.2% was reported for the plant modeling, it increased to 10% when irradiance predictions were incorporated to the model (PVCROPS, 2015). As a way to avoid these errors, Model Output Statistics (MOS) are applied to weather forecasts. The simplest MOS technique to improve temporal resolution consists in directly interpolating the values. There exist other more accurate methods, such as the incorporation of the clear sky index along with the solar zenith angle (Lorenz et al., 2008) or a method based on stepwise linear regression to select the variables that best represent the errors (Verzijlbergh et al., 2015). Nevertheless, the application of MOS requires some historical weather data, which may not be always available, diminishing part of the advantages of PV performance models. Additionally, each PV performance model is site-specific and because all technical specifications of equipment are rarely known, simplifications have to be made, impacting on the model accuracy.

Authors that used a white box approach in their way to forecast solar power output are: Lorenz et al. (2007, 2008, 2011b,a), Kudo et al. (2009), Pelland et al. (2011), Lonij et al. (2013), Lipperheide et al. (2015), Chu et al. (2015), Larson et al. (2016) and PVCROPS (2015).

To deepen in the comprehension of PV performance models, readers are referred to Dolara et al. (2015b) and Ayompe et al. (2010).

4.2. Statistical models

Statistical models do not need any internal information from the system to model it. It is a data-driven approach which is able to extract relations on past data to predict the future behavior of the plant. Thus, quality of historical data is essential for an accurate forecast. They have proven superior to PV performance models in the modeling of a PV plant (Graditi et al., 2016). Contrary to the parametric approach, for this approach a large historical dataset is typically required (meteorological and power measurements, i.e.), for which the plant must have been working already for some time. This method benefits from the ability of correcting systematic errors associated to the measurement of inputs.

The selection of a proper training dataset becomes crucial in the accuracy of the model developed. Several authors have studied the relation of the selection of the set of inputs in the final result. Almeida et al. (2015) compared training sets based on previous samples, on similar clearness index and on the similarity between the empirical distribution on the intradaily irradiance forecast for the day to be predicted and for each day of the dataset. They concluded that the best combination for the training set was the selection of 30 days (less days proved to be insufficient and more days did not yield better results) and based on the similarity of the empirical distribution mentioned before. Other authors found promising results first classifying the training set with respect to weather conditions (sunny, cloudy, overcast) and then training each day to be forecasted with the proper dataset (Mellit et al., 2014; Shi et al., 2011; Chen et al., 2011; Bouzerdoum et al., 2013).

Optimization of the set of inputs can be performed by the application of GA (Pedro and Coimbra, 2012), Particle Swarm Optimization (PSO), Genetical Swarm Optimization (GSO) (Ogliari et al., 2013), firefly optimization (Haque et al., 2013), stepwise regression (Ramsami and Oree, 2015), fuzzy logic (Simonov et al., 2012) or Principal Component Analysis (PCA) (Fonseca et al., 2014c). These techniques determine the combination of variables that yield best results and establish a trade-off between complexity and accuracy. The above-mentioned studies showed better results when optimization of inputs was applied.

What follows next is a broad classification of the main techniques used in the papers analyzed. For a more in depth description of the techniques, readers are referred to the corresponding authors or to Inman et al. (2013) and Diagne et al. (2013).

4.2.1. Regressive methods

These set of techniques estimate the relationship between a dependent variable (solar power output) and some independent variables, called predictors. Depending on how time series are treated (linear or non-linear and stationary or non-stationary), a further classification appears. Stationary time-series designate those

time series which fluctuate around a static mean while non-stationary series do not show such mean.

- **Linear stationary models**. Here we find Auto-Regressive (AR) models (Bacher et al., 2009), which model the output as a linear combination of lagged values of the predictors; simple Moving Average (MA) models (Li et al., 2014), used when data presents a constant variance over an equilibrium position around the mean, for which the average of the past data is used as the prediction; double MA models (Li et al., 2014), used when there is a trend; Auto-Regressive Moving Average (ARMA) models (Chu et al., 2015), which consider both lagged past values and errors; AR eXogenous (ARX) models (Bacher et al., 2009), which add exogenous data to an AR model, and Auto-Regressive Moving Average with eXogenous variables (ARMAX) models (Li et al., 2014), which introduce external variables in the time series analysis (i.e. from NWP). To deal with probabilistic analysis, there exist some adaptations, such as the Vector AR (VAR) or the Vector ARX (VARX) (Bessa et al., 2015).
- **Linear non-stationary models**. Auto-Regressive Integrated Moving Average (ARIMA) models (Pedro and Coimbra, 2012), which consider the union of AR and MA components, and Seasonal ARIMA (SARIMA) (Bouzerdoum et al., 2013), which introduce a seasonal component.
- **Non-linear stationary models**: Non-linear AR-eXogenous (NARMAX) model.

It has been proved that feeding the above-mentioned methods with exogenous data tend to improve results. Bacher et al. (2009) achieved better results with ARX than with AR, Li et al. (2014) concluded that their ARMAX model outperformed ARIMA, and Bessa et al. (2015) also got better results with their VARX model than with the VAR, although a probabilistic analysis yielded worse results in some quantiles.

4.2.2. Artificial Intelligence (AI) techniques

In this section, the main AI techniques are described:

- **Artificial Neural Networks (ANNs)**. The ANNs are the most used machine learning techniques in solar power forecasting (Fig. 2). They are inspired in the neuron operation, where a group of neurons are interconnected to form a neural network (NN). Connections have numeric weights, whose final value is given by the training phase, and all together predict the output. This range of techniques has proven useful in a wide variety of

situations and with a large number of inputs. There exist a large number of topologies. The classification by the number of hidden layers (simple perceptron or multi-layer) is the main one. From the different ANNs, the Multi-Layer Perceptron (MLP) are highlighted, which can be considered a universal proximate of functions and shows wide applicability, but many others are found: Time Delay NN, Elman recurrent NN, Radial Basis Function NN, Adaptive Neuro-Fuzzy (ANFIS), Adaptive Resonance Theory (ART), ART-2, etc.

- **k-Nearest Neighbors (k-NN)**. It is one of the simplest machine learning methods. It is based on an algorithm for pattern recognition, which compares the current state with training samples in a feature space. Euclidean distances are thus calculated and the first k nearest neighbors are selected for the predictions (Pedro and Coimbra, 2012).
- **Support Vector Machines (SVM)**. They are a supervised modeling method, firstly introduced by Vapnik and Lerner (1963) and further developed by Cortes and Vapnik (1995) to be used in classification problems. When they are applied to regression problems, they are known as Support Vector Regression machines (SVR). They stand out for their strong generalization capacity and for their ability to deal with non-linear problems. They work as a multiple linear regression using transformed predictors but keeping low complexity and a good fitting of data. There are three main parameters which dominate the performance of the technique and should be fine-tuned: ϵ, which is used in the accuracy term, the cost parameter C, which deals with the trade-off between accuracy and complexity, and γ, which regulates the kernel function, used to transform the predictors to a higher-dimensional feature space. SVM/SVR have shown great potential in several studies (Rana et al., 2015; Fonseca et al., 2011a).
- **Random Forests (RF)**. Firstly developed by Breiman (2001), they consist of an ensemble of decision/ regression trees, whose results show the mean prediction of individual trees. Normally, focusing on regression trees, they are characterized by overfitting to the training data set, which results on low bias but a high variance. It is common practice to grow several trees for each case study. By averaging results from several trees, it is possible to reduce the variance at the expense of slightly increasing the bias. Nevertheless, the overall performance of the model improves. Simple regression trees are very sensitive to noise of training dataset. Simply averaging the results of several trees to overcome this problem does not work, since being trained with the same dataset, they would be very correlated. This issue
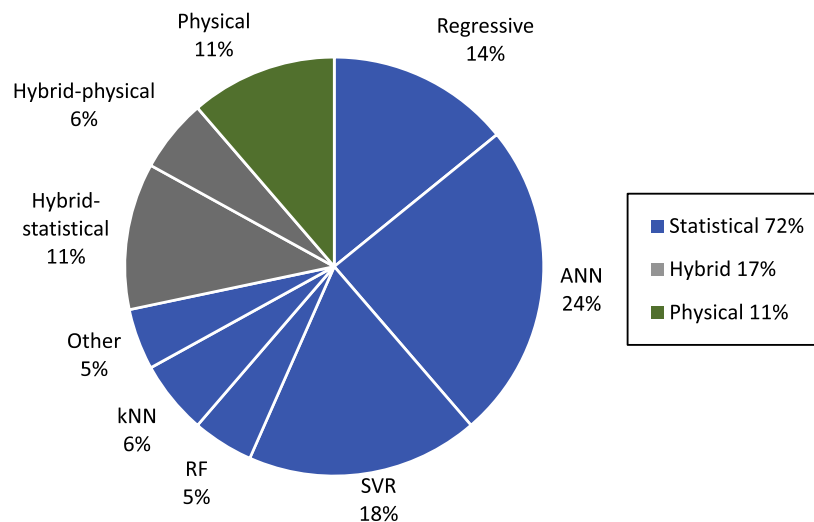

Fig. 2. Distribution of studies with respect to the technique used. Sample set: studies listed in Table 2.

is tackled with the bagging method, which consists of growing unpruned trees, each with a bootstrapped sample of the whole training set. The out-of-bag samples left out of the model are used to assess each model's performance. Nevertheless, trees can be still correlated, inasmuch as if some variables are very strong predictors for the output, they will be selected in most of the bagged trees. Random forests deal with this problem with feature bagging, which consists of selecting a random subset of features at each node and thus, they get to reduce correlation.

Also, there are many benchmarking studies to determine which techniques and under which conditions work best for the proposed situations (Pedro and Coimbra, 2012; Fernandez-Jimenez et al., 2012; Alessandrini et al., 2015; Zamo et al., 2014a).

4.3. Hybrid models

Single models can omit some information due to the way each technique transforms data. Thus, it is also common to combine techniques to foster their strengths so as to improve accuracy, which is denoted as hybrid, blended, combined or ensemble models. Models can be blended by several ways, such as bagging, boosting, voting or stacking.

Two approaches can be followed, either from combining two or more statistical techniques (hybrid-statistical) or from joining a statistical technique to a PV performance model (hybrid-physical). Several works can be found using the former approach. Bouzerdoum et al. (2013) combined SARIMA with SVM in their hour ahead predictions, whereas Ramsami and Oree (2015) used ARIMA with ANN, Vaz et al. (2016) applied an ensemble technique of ANN and Non-linear ARX (NARX) and Hossain et al. (2013) developed a methodology to select the best combination of regression techniques to construct an ensemble model, to name some of them. They all proved better performance than stand-alone techniques. The latter approach makes use of regressive or AI techniques with the objective of reducing the error associated to NWP or to forecast in time domains where some NWP models lack enough resolution, such as the hour ahead horizon. Normally, the variable predicted by the statistical technique is GHI, although some authors predicted k_t and then converted it into power (Bracale et al., 2013; Mora-Lopez et al., 2011). Some studies, however, constructed a black box model to directly predict power but included in its structure some physical expressions. That is the case of the Physical Hybrid Artificial Neural Network (PHANN) method, which combined an ANN with a clear sky model (Gandelli et al., 2014; Dolara et al., 2015a). Almonacid et al. (2014) also combined some physical expressions, such as the calculation of the clearness index and cell temperature, with NAR and ANN techniques.

4.4. Summary of forecasting techniques

Fig. 2 shows the distribution of studies analyzed regarding the technique used. As seen, the most common approach among the papers analyzed is the use of statistical techniques, especially ANN, accounting for the 24% of the studies.

5. Spatial horizon: single plant and regional forecasts

Forecasts can be made for a single PV system or for an ensemble of them. Normally, grid operators prefer regional forecasts since they are more useful to keep the balance between demand and supply in the electric system. To better understand the differences between point and regional forecasts, first we study the short term power output variability.

As detailed in Mills and Wiser (2010), variability of solar resource at different time scales poses several problems to the integration of solar energy. Forecasts of power output see the consequences of the unpredicted solar variability as deviations between forecasted and what is really produced. Balancing reserves have to be scheduled for such unforeseen changes. Variability is normally defined as the difference in the output of a plant from one averaging interval (with duration Δ_t) to another, commonly referred as "deltas" or step changes. Then, the standard deviation of the deltas over a long period for a certain Δ_t can be calculated. The overall average variability is normally represented by the 99.7^{th} percentile. Once calculated the variability for a single plant, the study can be extended to an ensemble of them, referred to the standard deviation of the variation of the aggregate output. It is seen that the variability of the ensemble is lower than the variability of scaling a single plant. They named this effect the "diversity filter" (12).

$$D^{\Delta_t} = \frac{\sigma_{\Delta P}^{\Delta_t}}{\sum_{i=1}^{N} \sigma_{\Delta P_i}^{\Delta_t}} \qquad (12)$$

If it is assumed that all N plants from the ensemble have similar size and variability, it is seen that the diversity filter depends on the correlation of the deltas. The step changes, in turn, depend on spatial and temporal scales. Thus, if plants are totally correlated, the diversity filter would be 1 and, consequently, the variability of the ensemble would equal the sum of that of individual systems. However, perfectly uncorrelated plants would lead to a global variability $\sqrt{N}$ times that of a single site. Correlation between sites is inversely proportional to distance between sites (D) and time scales. It is also influenced by the speed of clouds (CS) in the area, which will determine the relevant fluctuation time interval (Perez and Hoff, 2013). As an example, if an ensemble of equal PV plants were distributed in a 10x10 grid with 20 km spacing, the aggregated variability would be 6 times less than the variability of a single site, for $\Delta_t < 15$ min (Mills and Wiser, 2010). Perez et al. (2011) proposed an empirical Eq. (13) to determine correlation coefficients between sites from their study across USA considering the distance, time interval and cloud speed.

$$\rho_{pair} = e^{Ln(0.2)D/1.5\Delta_t CS} \qquad (13)$$

Later on, Arias-Castro et al. (2014) proposed a mathematical method to determine ramp rates correlations. Since it also took into account wind direction, an anisotropic model was obtained, which proved superior to previous isotropic models for up to 60 s.

As pointed out in Perez and Hoff (2013), real situations are conformed by dispersed generation with different systems and arbitrary distances. Site-pair correlation varies and depends on PV plant sizes, where smoothing effect can even take place intra-array. Thus, they proposed a general equation for the variability of non-homogenous PV plants (14), depending exclusively on site correlation and standard deviations of the plants in the fleet.

$$\sigma_{\Delta_t}^{fleet} = \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{\Delta_t}^i \sigma_{\Delta_t}^j \rho_{\Delta_t}^{i,j}} \qquad (14)$$

As seen, variability in power output is reduced when an ensemble of plants is considered. Since forecast errors increase with the variability of the signal to be forecast, one would expect that regional forecasts would also benefit from spatial averaging effects. Several authors have worked on this issue (Lorenz et al., 2007, 2008, 2009, 2011b,a, 2014; Zamo et al., 2014a; Zhang et al., 2015a; Fonseca et al., 2014a,b,c,d).

E. Lorenz and colleagues, through several papers (Lorenz et al., 2007, 2008, 2009, 2011b,a, 2014), made an in-depth study of the German PV plants, analyzing both single site and regional forecasts. They proved how spatial averaging and smoothing effects

lead to lower errors at the regional frame. They saw that the reduction of errors over an area was dependent on the correlation of errors in the plants of the ensemble. That correlation follows an exponential curve with the distance between plants. It rapidly decreases with increasing distance at the beginning and then it slows down for distances of 200 km or more (Fig. 3). The accuracy of the forecasts is inverse to the error correlation. Low values of correlation yield better area forecasts.

Whereas the average normalized Root Mean Squared Error (nRMSE) of single sites was 13%, that value dropped to 5% for the ensemble of German plants. They defined the error reduction factor f, stated as the ratio of the Root Mean Squared Error (RMSE) of the ensemble to the RMSE of single plants. Thus, they found a $f = 0.7$ for a subset of 11 PV plants, lowering to $f = 0.61$ when the region considered covered an area of 220×220 km or down to $f = 0.4$ for the area of Germany.

Zhang et al. (2015a) also analyzed the differences between single site and ensemble forecasting. nRMSEs for a 100 MW plant were 22% and 17% for day-ahead and 1 h ahead horizons, while those values dropped to 4% and 2% respectively for the aggregate of plants in the Western Interconnection (64,495 MW).

Regional forecasts can be approached by several ways, depending on the availability of data. As described in Fonseca et al. (2014d), there can be four different cases:

- **Knowledge of power generation from all PV systems**: If production from all the PV plants in the region is known, they can be modeled separately and the forecasts from each site can be summed (bottom-up strategy), benefiting from the smoothing effect (Fonseca et al., 2014a,c,d). This approach normally requires a big computational effort and the knowledge of data from all plants, which is difficult to obtain. Nevertheless, as shown in Fonseca et al. (2014d), it yielded the best results of the models proposed.
- **Knowledge of PV power only from some plants and not regionally**: In some cases, it is only possible to monitor the production of some PV plants and the regional production is not measured. Upscaling from a representative set of PV plants is then a suitable option. As described in IEA (2013), if done properly, there are almost no differences in accuracy compared to the modeling of all the plants. Upscaling is applied in two steps. In the first one, a subset must be selected so that its behavior is representative of the power output of the whole ensemble. One way of getting it is varying a random number of plants until their forecast errors or correlation parameters are similar to that of the ensemble. Another option could be, instead of making random subsets, to make subsets according to plant
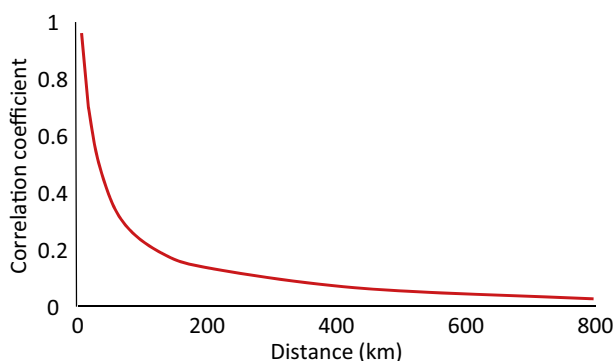

Fig. 3. Correlation coefficient of forecast errors of two systems in Germany over the distance between them. *Source:* Adapted from Lorenz et al. (2008).

characteristics so that they are representative of the complete set. The second step consists of the upscaling of the power output of the said subset of plants to the complete set.

Lorenz et al. (2011a) stressed the idea of a proper selection of the representative subset to obtain accurate results. They detailed an upscaling technique which took into account the geographic location (latitude ϕ and longitude λ) of the PV plants.

Fonseca et al. (2014d) proposed a method based on stratified sampling for selecting the subset of representative plants. This method takes into account the installed capacity and PV system location. Instead of making subsets with random number of plants, they applied (15) to select the sampling size n

$$n = \left\{ \frac{t(\alpha/2;\infty)s}{d} \right\}^2 \qquad (15)$$

where d is the error margin for the mean regional PV power provided by the sampling, t, the t-Student distribution and s, the standard deviation of the sampled values. Based on previous experiences, they considered $d = 0.05$ kW h/kW$_{cap}$ and $s = 0.15$ kW h/kW$_{cap}$. For the ensemble of 273 PV plants, they obtained a theoretical sampling size of 39 plants. The sampling size thus obtained is independent of the number of PV plants in the region. With the sampling size fixed, they selected the PV plants according to their location. Thus, if a prefecture contained 20% of the installed capacity, they selected 20% of the samples from there, with the aim of building a balanced set of samples. Further corrections were applied to account for the weight of the capacity of each prefecture with respect to the total. Then, the regional power P_{reg} was calculated with the upscaling Eq. (16).

$$P_{reg} = \frac{P_{set}}{Cs_{tot}} Ct_{reg} \qquad (16)$$

where P_{set} is the forecasted power of the sampling set, Cs_{tot}, the set capacity and Ct_{reg}, the total PV power capacity of the region. This method performed similarly to the bottom-up strategy, with the additional advantage of not needing to simulate all PV plants.

Another option is to establish a virtual reference PV plant and suppose that all other plants from the ensemble will perform similarly (Zamo et al., 2014a). There, they found that summing individual forecasts resulted in slightly better results than those obtained with the virtual plant (RMSE of 5.8% vs. 6%). On the other hand, both approaches improved single site forecasts (average RMSE 10–12%).

- **Knowledge of the regional production but not from individual systems**: In this case, the regional production is forecasted directly. Fonseca et al. (2014c) constructed their regional forecasts based on NWP from the Japan Meteorology Agency for the location of each PV plant and then, performed a PCA to avoid redundant information (i.e. NWP of PV plants located close to each other). It was observed that the proposed methodology, based on SVR, outperformed the bottom-up approach for areas with radius inferior to 100–150 km (depending on the region), when both forecasting techniques equaled. They found one exception: in the largest region considered, covering different climates, summing individual production yielded better results than their proposed methodology. This contrasts with the findings by Lorenz et al. (2011a) and Zhang et al. (2015a), who obtained best results for the largest area considered.
- **No PV power data is available**: if no PV power data is available, it is still possible to derive production from irradiance forecasts. However, it is necessary to provide the installed capacity of PV plants P_{cap}^{reg} and some information (tilt angle and azimuth).

$$P_{pred}^{reg} = \frac{P_{cap}^{reg}}{I_{ref}} k \sum_{i=1}^{n} a_i I_{POA,i} \qquad (17)$$

$$\sum_{i=1}^{n} a_i = 1 \qquad (18)$$

where I_{ref} is the irradiance at which the capacity of PV plants is rated, k is the balance of system coefficient and a, the weight given to the forecast of solar irradiation that reaches the PV modules with a given tilt angle and orientation.

5.1. Summary

Forecasts can be made for a single plant or for an ensemble of them, depending on the purpose for which they are issued. Regional forecasts benefit from the so-called smoothing effect, reaching error reduction factors down to 0.12. Such reduction of errors is dependent on the correlation of errors of the plants in the ensemble, which follows an exponential curve with the distance between plants. Regional forecasts can be approached by several ways, depending on the information available. In spite of the bottom-up strategy (summing individual forecasts) outperforming the others, satisfactory results can be obtained with other less computational demanding techniques.

6. Deterministic vs. probabilistic

Energy forecasts have been applied for a long time, not only predicting production (wind, solar) but also forecasting load. Each domain has its own peculiarities and differences in accuracy can be found between them. As seen in Fig. 4, solar power forecasts are the least mature of the energy forecasts analyzed by Hong et al. (in press), due to the relatively low solar penetration in electricity markets so far. However, wind forecasting shows a high level of maturity for its similarity to meteorological forecasting, which is more developed.

Traditionally, solar power forecasts have output a single value for each forecast horizon, what is denoted as a deterministic or point forecast. As seen in Fig. 5, the purpose of point forecasts is to determine at time t the power production at time $t + f_h$. This approach ignores information that is very valuable for utility managers, like the upper and lower bounds of possible forecasts or the percentage of confidence for each value.
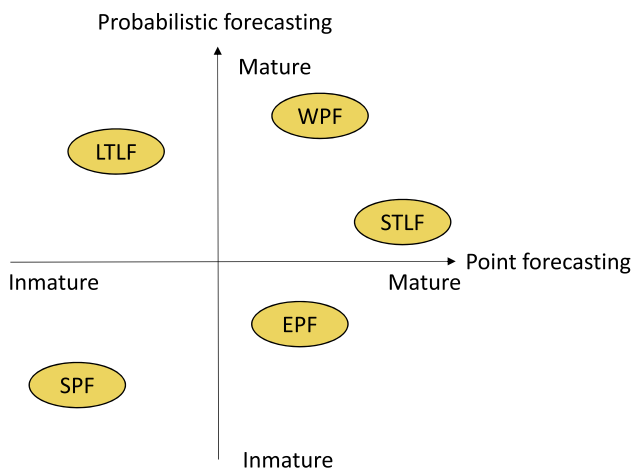

Fig. 4. Maturity of different energy forecasts. SPF stands for solar power forecasting, LTLF, for long term load forecasting, EPF, for electricity price forecasting, WPF, for wind power forecasting and STLF, for short term load forecasting. *Source:* Hong et al. (in press).

Thus, some authors have recently started to introduce probabilistic forecasts (normally used in wind and load forecasts), which add relevant information about the expected values. They are especially useful for activities with implicit uncertainty and where risk must be managed, like balancing generation and demand in the electricity market. Some of the benefits of probabilistic forecasts are the better allocation of power reserves to overcome solar power uncertainty and greater economical revenue in the day-ahead market compared to deterministic predictions (Alessandrini et al., 2015). Probabilistic forecasts provide a broader knowledge of the predictions, inasmuch as a range of plausible values is determined as well as the probability associated to each of them. Probability Density Functions (PDFs) are normally used at this point. A further classification can still be made, according to the adopted approach. Thus, we can find the prediction error approach, which provides the uncertainty of the error derived from the use of a deterministic method, and the direct approach, which directly outputs the statistical representation of the prediction (Bracale et al., 2013).

Probabilistic forecasts have been successfully used in Bacher et al. (2009), Almeida et al. (2015), Alessandrini et al. (2015), Bessa et al. (2015), Monteiro et al. (2013b), Zamo et al. (2014b), Zhang et al. (2015a), Bracale et al. (2013), Lorenz et al. (2009), Fonseca et al. (2015), Huang and Perry (2015), Nagy et al. (2016), Jafarzadeh et al. (2013), AlHakeem et al. (2015), Rana et al. (2015), Sperati et al. (2016), Mohammed et al. (2015) and Golestaneh et al. (2016).

Lorenz et al. (2009) presented their results in prediction intervals, indicating the range in which the measured power was expected to fall with a certain probability. To obtain the accuracy of the forecasts depending on cloud situations, they made weather specific prediction intervals based on an error analysis. The clear sky index k_{cs} and the cosine of the solar zenith angle ($cos\theta$) were used to model the Standard Deviation of Errors (SDE). Then, they obtained the upper and lower bounds of solar irradiance.

$$I_{upper/lower,limit} = I_{pred} \pm 2 \cdot sde(cos\theta, k_{cs}) \qquad (19)$$

where the $+$ symbol is used for the upper limit and the $-$ symbol, for the lower limit. Readers are reminded that the authors used a PV performance model to convert solar irradiance into PV power output. For single stations, 95% of the measured values fell within the prediction intervals, whereas for the ensemble of stations studied that value dropped to 91%, presumably due to the constant error reduction factor used for the ensemble of stations. Finally, they applied the error propagation presented in Lorenz et al. (2007), which also takes into account the uncertainty of the tilt conversion model and the PV system modeling. Bacher et al. (2009) studied the uncertainty of the forecasts in relation to the type of weather, concluding that the uncertainty of partly cloudy days was higher than for overcast or sunny days.

Monteiro et al. (2013b) created a model based on historical similarity and apart from giving the point forecast, they also showed the uncertainty associated with it. They grouped the power output into different power intervals, covering the whole range of possible power outputs from the studied plant. Then, they presented the results showing the probability of occurrence that the PV power production were included in each of the power intervals. If for a certain hour two consecutive intervals summed most of the probability, then it was stated that the uncertainty of that point forecast was low. Bracale et al. (2013) obtained the power output PDF using a Bayesian auto regressive time series model. The actual power measurements were located always between the 5 and 95th percentiles. They found it difficult to select a single parameter of the PDF (mean value, percentiles, ...) to represent the whole distribution.
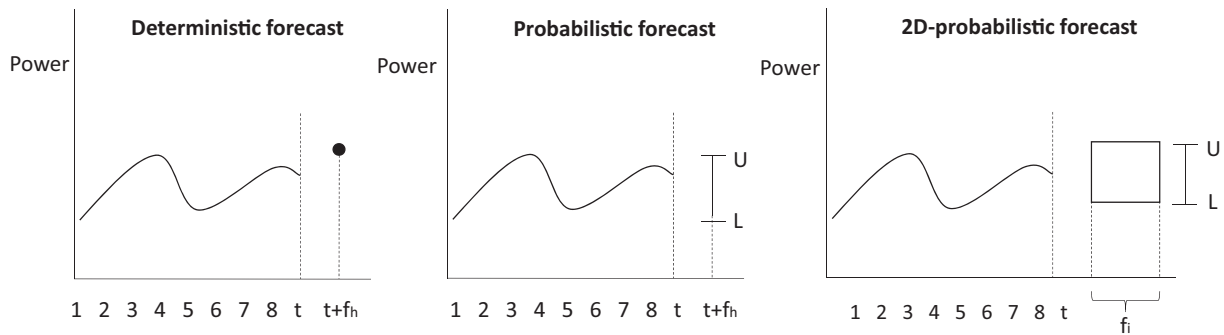
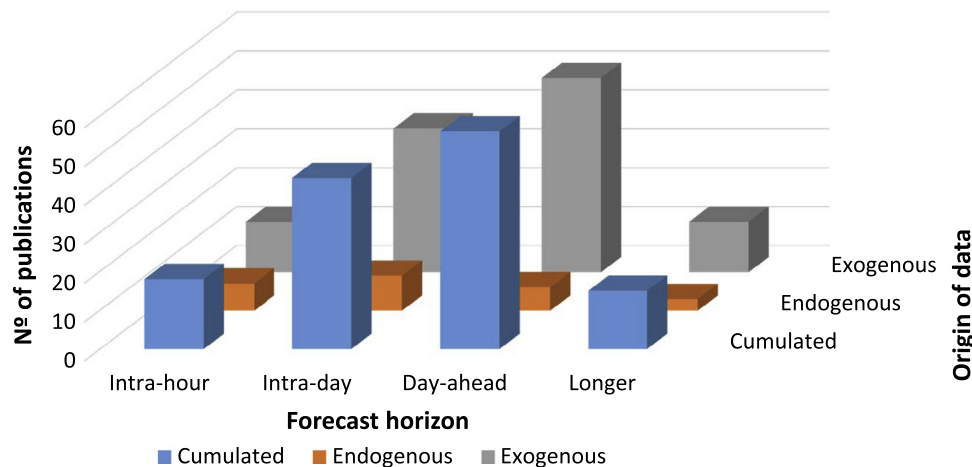**Fig. 5.** Different forecasting schemes. *Source:* Rana et al. (2015).

Fig. 6. Number of publications in recent years with respect to the time horizon and origin of inputs. Sample set: studies listed in Table 2.

Zamo et al. (2014b) applied several nonparametric statistical quantile regression methods to approach the problem. To try to improve forecasts, they performed a calibration technique based on the rank histogram. This graphical tool ranks each observation with the corresponding predicted members. The graphic is then obtained by computing several pairs of predicted and measured values. A flat rank histogram is a necessary but not sufficient condition for a forecast to be reliable. The calibration method works as follows: create the rank histogram based on the training subset, associate to each forecast quantile of the test set the associated proportion of observations below that quantile computed on the training sample. Said association gives some points of the Cumulative Density Function (CDF). Then, linear interpolation is performed between those points and extrapolation is performed based on assumptions, creating the corrected CDF. Results were explained according to the reliability and potential Continuous Ranked Probability Score (CRPS) terms (see Section 8) and for different set of quantiles. They showed that the climatological model used as benchmark was beaten by all studied techniques in terms of CRPS. With respect to the calibration technique, it was observed that in some cases it worsened the predictions. They found two explanations: the interpolation and extrapolation between available quantiles could introduce large errors if not enough quantiles were known. In addition to this, differences between training and test rank histograms could lead to an inadequate correction.

Alessandrini et al. (2015) also drew a rank histogram but with a different purpose from the above mentioned case. They used it to see if the observations were equally distributed in the predicted PDF and calculated the missing rate error to know if the predictions were under or over dispersive. To convert probabilistic forecasts to deterministic ones, they selected the median value of the PDF (solution that was also adopted by Almeida et al. (2015) and Bessa et al. (2015)), arguing that it provides lower Mean Absolute Error (MAE) than the mean. Other metrics used by these authors are compiled in Section 8. Sperati et al. (2016) also worked with PDFs and to further improve accuracy, they calibrated them via the variance deficit and ensemble model output statistics. They evaluated results in terms of statistical consistency (missing rate error), spread/skill consistency (with their respective diagrams), forecast value (Brier Skill Score (BSS) and Relative Operating Characteristic (ROC) Skill Score (ROCSS)), reliability and sharpness, and the CRPS.

Almeida et al. (2015) used quantile regression forests to obtain the quantile $Q_{.5}$ (median), from which they derived the point forecasts and all their performance statistics, and a 80% confidence interval (between quantiles $Q_{.1}$ and $Q_{.9}$). They analyzed both the amplitude and the accuracy of the confidence interval (see Section 8) and results were grouped in three classes according to the daily clearness index.

Bessa et al. (2015) used spatio-temporal correlations of distributed PV systems in a smart grid for their forecasts. Whereas in the deterministic analysis (derived from the median value) they obtained improvements ranging 6.5% and 3.3% for forecast horizons between 1 and 6 h, respectively, in the probabilistic analysis they obtained negative CRPS improvements for lead times between 4 and 6 h. To explain that difference, they plotted the quantile loss with respect to the quantiles for horizons of 1 and 5 h. They observed that for 1 h lead time, the quantile loss of the proposed model was lower than the baseline, which did not consider information from any other system. However, for 5 h lead time, the quantile loss of the proposed model was higher than the reference

Table 2
Recent publications on solar power forecasting.

Authors/year	Forecast horizon	Forecast resolution	Method	Variables
Lorenz et al. (2007)	Out to 3 days	1 h	NWP with post-processing for GHI. Then, a GHI→P PV performance model. Regional forecasting	Forecasted GHI and T
Yona et al. (2007)	1–3 h	1 h	ANN for GHI. Then, a GHI→P PV performance model	Past values of GHI
Lorenz et al. (2008)	Out to 3 days	1 h	NWP with post-processing for GHI (corrections as function of k_{cs} and solar zenith angle). Then, a GHI→P PV performance model. Regional forecasting	Forecasted GHI and T
Bacher et al. (2009)	1–36 h	1 h	AR with only lagged PV values, LM with NWP and ARX with NWP and lagged P	Past values of P and forecasts of GHI
Kudo et al. (2009)	18–31 h	1 h	Direct and indirect approach, based on day-before data	Past values of P and GHI and forecasts of GHI and T
Li et al. (2009)	15–45 min	15 min	Advance Grey-Markov chain	Actual and past values of P
Hassanzadeh et al. (2010)	1, 5, 10, 15, 30 and 60 min	–	GHI modeling via the sum of a deterministic component and a Gaussian noise signal and Kalman filtering. Then, GHI→P PV performance model	Past values of GHI
Mellit and Massi Pavan (2010)	Day-ahead	1 h	MLP for GHI. Then, a GHI→P PV performance model	Daily $\overline{GHI}$ and $\overline{T}$, day of month
Tao et al. (2010)	Day-ahead	1 h	NARX with NWP. Weather classification	Clear sky radiation for next day, forecasted highest and lowest T and day type index
Berdugo et al. (2011)	10–30 min and 1–3 h	10 min and 1 h	k-NN with analogs from neighboring plants	Analogs from 11 PV systems as far as 70 km from each other
Chen et al. (2011)	Day-ahead	1 h	ANN trained with NWP classified via SOM	Daily $\overline{P}, \overline{GHI}, \overline{T}, \overline{RH}, \overline{W}$, day of month
Chupong and Plangklang (2011)	Day ahead	1 h	Elman NN fed with clear sky model and NWP	NWP of T_{max}, T_{min} and cloudy index
Ding et al. (2011)	Day-ahead	1 h	ANN models trained with closest historical data with respect to forecasted weather conditions	Past values of P and high, low and average T from similar and forecasted days
Fonseca et al. (2011a)	1 h	–	SVM fitted with NWP	T, RH, low, mid and upper level cloudiness, I_{ex}
Fonseca et al. (2011b)	Up to 2 h and up to 25 h	–	SVM fitted with NWP. Comparison of accuracy of different forecasting horizons	T, RH, low, mid and upper level cloudiness, I_{ex}
Lorenz et al. (2011b)	Out to 3 days	1 h	NWP with post-processing for GHI. Then, a GHI→P PV performance model. Upscaling for regions	Forecasted GHI and T
Lorenz et al. (2011a)	Out to 2 days ahead	1 h	NWP and GHI→P PV performance model. Correction for snow-covered panels	Forecasts of GHI, T, snow depth
Mora-Lopez et al. (2011)	Day-ahead	–	PFA and multivariate regression to predict k_t. Then, a GHI→P PV performance model	P, I_{POA}, module temperature
Pelland et al. (2011)	0–48 h	1 h	NWP with MOS to remove bias and spatial averaging via Kalman filtering. Then, a GHI→P PV performance model	Forecasts of GHI and T, system characteristics (tilt, azimuth, performance, …)
Shi et al. (2011)	Day-ahead	1 h	Selection of SVM models based on previous data and NWP	Past values of P from similar weather situations
Al-Messabi et al. (2012)	10, 60 min	–	Dynamic ANN	Actual and past values of P
Chow et al. (2012)	Real time, 10–20 min	10 min	MLP	GHI, solar elevation and azimuth angle and dry-bulb T
Fernandez-Jimenez et al. (2012)	1–39 h	1 h	NWP in cascade from two models with different resolutions. Then, benchmark of different techniques: ARIMA, k-NN, ANN and ANFIS	Past values of P and forecasts of surface sensible and latent heat flux, surface downward shortwave and longwave radiation, top outgoing shortwave and longwave radiation and T
Mandal et al. (2012)	1 h	–	WT to decompose past values of P. Then, ANN and back transformation with WT	Actual and past values of P, GHI and T
Pedro and Coimbra (2012)	1–2 h	1 h	Benchmark of ARIMA, k-NN, ANN and GA-ANN. Study of variability periods	Past values of P
Simonov et al. (2012)	Day-ahead	1 h	Fuzzy logic pre-processing to filter NWP used as inputs for ANN	Julian day and forecasts of GHI, T, W, RH and pressure

Table 2 (*continued*)

Authors/year	Forecast horizon	Forecast resolution	Method	Variables
Yang and Xie (2012)	5, 15 min, 1, 24 h	–	ARX with spatio-temporal data to predict GHI. Then, a GHI→P PV performance model. Onsite historical data and from neighboring PV plants	Past values of GHI at different plants and CC percentage
Bouzerdoum et al. (2013)	1 h	–	Combination of SARIMA and SVM to take into account seasonality	Past values of P, I_{POA}, T, module temperature
Bracale et al. (2013)	1 h	–	Probabilistic forecasting based on Bayesian AR to predict future k_t from historical k_t PDF. Then, a physical $k_t \rightarrow P$ with Monte Carlo simulation	Past values of clearness index, T, RH, W, CC
Haque et al. (2013)	1–12 h	1 h	Combination of WT, fuzzy adaptive resonance theory mapping NN and firefly optimization algorithm	Actual and past values of P and GHI
Hossain et al. (2013)	6 h	–	Ensemble modeling using 3 regression algorithms: MLP, SVM and LMS	NWP of several variables
Jafarzadeh et al. (2013)	1–3 h	1 h	Interval type-2 Takagi–Sugeno-Kang fuzzy systems for probabilistic forecasts	NWP of GHI and T
Lonij et al. (2013)	15–90 min	15 min	Forecasts in cloudy days based on spatio-temporal correlations with distributed PV plants. Determination of cloud speed	P from neighboring plants, forecasts of W
Monteiro et al. (2013a)	Day ahead	1 h	Analytical model based on irradiance and PV attenuation indexes and MLP model	Past values of P, NWP of several meteorological variables
Monteiro et al. (2013b)	1–24 h	1 h	Historical similar mining mechanism based on transitions between actual and previous values of variables. Generation of probability matrix for future intervals	Past values of P, forecasted values of GHI and T, solar hour
Ogliari et al. (2013)	Day-ahead	1 h	ANN fed with NWP variables selected by GSO	Julian day, hour, predictions of T, W, RH, pressure and CC
Oudjana et al. (2013)	1–7 days	1 day	Benchmark of linear and multiple regression and ANN	Past values of GHI and T
Takahashi and Mori (2013)	30 min	–	Different ANN models, using Deterministic Annealing	Present and past values of P and T_{cell}
Urquhart et al. (2013)	30 s–15 min	30 s	Determination of cloud maps based on sky images	Sky images, values of P
Yona et al. (2013)	Day-ahead	1 h	Forecasts of GHI based on NWP, fuzzy theory and ANN. Then, a GHI→P PV performance model	Forecasts of T, RH and CC
Zeng and Qiao (2013)	1–3 h	1 h	Least Square SVM, comparison with AR and ANN	Past values of meteorological variables
Almonacid et al. (2014)	1 h	–	Time Delay NN to predict GHI and T. Then, ANN to predict P	Past values of GHI, T and P
Fonseca et al. (2014a)	Day ahead	1 h	Characterization of errors of regional forecasts with SVR	Past values of P and NWP of T, RH and CC in three levels, I_{ex}
Fonseca et al. (2014b)	Day ahead	1 h	Different methods for regional forecasting based on SVR, NWP and PCA	Past values of P and NWP of T, RH and CC in three levels, I_{ex}
Fonseca et al. (2014c)	Day ahead	1 h	Principal component analysis, SVR and NWP	Past values of P and NWP of T, RH and CC in three levels, I_{ex}
Fonseca et al. (2014d)	Day ahead	1 h	Comparison of four methodologies based on SVR and NWP for regional forecasting	Past values of P, NWP of T, RH and CC in three levels
Gandelli et al. (2014)	Day-ahead	1 h	PHANN model, combining ANN and a physical clear sky solar radiation model	Forecasts of T, W, RH, pressure, CC, inputs of clear sky model
Li et al. (2014)	Day-ahead	1 h	Benchmarking of time series models (ARMAX model with NWP variables, ARIMA, simple and double moving average, etc.) and ANN	Past values of P and forecasts of daily $\bar{T}$, dew T, highest and lowest T, precipitation, W, wind direction, RH, pressure and insolation duration
Long et al. (2014)	Intra day up to 3 days	1 day	Benchmark of several techniques: ANN, SVM, k-NN and MLR	P, GHI, W, RH, rain, insolation time, dew temperature
Lorenz et al. (2014)	15 min to 5 h	15 min and 1 h	Benchmark of physical and SVR approach comparing persistence, satellite CMV and NWP inputs	Past values of P, NWP, satellite CMV, information on PV system tilt and orientation

Reference	Horizon	Resolution	Description	Inputs
Masa-Bote et al. (2014)	Day ahead	1 h	Comparison of PV performance model fed with NWP and ARIMA fed with endogenous data	Past values of P and NWP of GHI and T
Mellit et al. (2014)	Day-ahead	1 h	Classification of future days with forecasted GHI. Then, ANN trained with similar weather situations to predict P	Actual and past values of P, GHI and cell temperature and forecasts of GHI and T
Russo et al. (2014)	15–60 min	15 min	Genetic Programming, no cloud data is used. No memory system	Actual value of P and time difference with respect to sunrise
Wu et al. (2014)	1 h	–	Hybrid model combining ARIMA, SVM, ANN, ANFIS and GA	Past values of P, NWP of GHI
Yang et al. (2014)	Day-ahead	3 h	Classification of historical data with SOM and learning vector quantization. Then, application of SVR model trained with similar data to the forecasted by the NWP to predict P	Forecasted maximum T, probability of precipitation, weather description, month, past similar GHI values
Zamo et al. (2014a)	28–45 h	1 h	Benchmark of statistical regression techniques: LM, binary regression tree, bagging boosting, RF, SVM and generalized additive models. Upscaling	Several NWP variables
Zamo et al. (2014b)	66–72 h	1 h	Probabilistic predictions using LM QR and QRForest. Assessment of the use of one or several NWP models	Several NWP models and variables
Alessandrini et al. (2015)	0–72 h	1 h	Analog Ensemble and QR, providing a likely set of P	Past values of P and forecasts of GHI, T, CC, solar azimuth and elevation angles
AlHakeem et al. (2015)	1–6 h	3 h	WD of actual PV measurements, ANN optimized via PSO and bootstrap method to quantify uncertainty	Actual values of P, predictions of GHI and T
Almeida et al. (2015)	Day ahead	1 h	QRF considering spatio-temporal variability indexes and different training sets	Past values of P, I_{POA}, T, W and several NWP variables
Bessa et al. (2015)	1–6 h	1 h	VAR and VARX models with information from surrounding smart meters in a smart grid to account for the presence of clouds. Point and probabilistic forecasts	Inputs of clear sky model, actual and past values of onsite P and from other systems
Chu et al. (2015)	5–15 min	5 min	Benchmark of techniques: cloud tracking, ARMA, k-NN and reforecast with ANN	Sky images, past values of P
De Felice et al. (2015)	1–10 days	–	SVR model using estimations and predictions of GHI and T	Estimations of GHI and T and NWP of GHI and T
De Georgi et al. (2015)	1, 3, 6, 12 and 24 h	–	Hybrid model based on LS SVM with WD	Past values of P and/or records of T, I_{POA} and module temperature
Dolara et al. (2015a)	24–72 h	1 h	PHANN model with mobile window forecast, updating training set	Forecasts of T, W, RH, pressure, CC, inputs of clear sky model
Fonseca et al. (2015)	Day ahead	1 h	Probabilistic forecasts based on the maximum likelihood estimation and similarity of input data using SVR and NWP	Past values of P and NWP of T, RH and CC in three levels, I_{ex}
Huang and Perry (2015)	1–24 h	1 h	GEFCom 2014. Gradient boosting for deterministic forecasts and k-NN regression for probabilistic	Several NWP variables from the ECMWF and past values of P onsite and from neighboring PV plants
Leva et al. (2015)	Day-ahead	1 h	ANN fed with historical data and NWP	Past values of P, GHI and NWP variables
Lin and Pai (2015)	1 month	1 month	Evolutionary seasonal decomposition LS-SVR for monthly prediction of the PV power output in Taiwan	Past values of P
Lipperheide et al. (2015)	20–180 s	20 s	Benchmark of cloud speed persistence model, AR and persistence methods	Past values of P and measurements of GHI to derive CMV
Lu et al. (2015)	Day ahead	1 h	Machine learning method combining NWP from 3 models to generate inputs for a GHI→P PV performance model	Past values of GHI and several NWP variables
Mohammed et al. (2015)	1–24 h	1 h	Probabilistic forecasting based on an ensemble of statistical techniques	Several NWP variables from the ECMWF and past values of P
Ramsami and Oree (2015)	Day-ahead	1 h	Hybrid ANN with input optimization via stepwise regression with NWP variables	Forecasts of GHI, T, RH, W, wind direction, pressure, rainfall and sunshine duration
Rana et al. (2015)	30 min, 1, 2, 3, 4, 6 h	–	2D Interval forecasts based on SVR	Past values of P, GHI, T, RH and W
Schmelas et al. (2015)	1–12 h	–	Three different models: physical, statistical (MLP) and hybrid (MLP-physical)	NWP of GHI, T, RH

Table 2 (*continued*)

Authors/year	Method	Forecast resolution	Forecast horizon	Variables
Zhang et al. (2015a)	Persistence of cloudiness for 1 h ahead	–	1 h and day ahead	Actual and past values of P and NWP
De Georgi et al. (2016)	Combination of LS SVR and Group Method of Data Handling to form a hybrid model	–	1, 3, 6, 12, 24 h	Past values of P and meteorological records
Do et al. (2016)	Determination of the minimum training period for three models: ANN, regressive and persistence	–	1 h	Actual values of P, measurements of CC and T
Golestaneh et al. (2016)	ELM to generate probabilistic forecasts	–	10, 60 min	Past values of P and meteorological records
Larson et al. (2016)	NWP of GHI and CC from two models. Then, MOS with spatial averaging to predictions of GHI and finally, a GHI→P model	1 h	Day ahead	NWP of GHI and CC
Li et al. (2016)	Comparison of ANN and SVR using hierarchical approach (summing individual predictions from inverters)	–	15 min, 1, 24 h	Past values of P, NWP of T, W and wind direction and solar geometry
Nagy et al. (2016)	GEFCom 2014. Voted ensemble of QRF and stacked RF	1 h	1–24 h	Several NWP variables from the ECMWF and past values of P
Rana et al. (2016)	Comparison of ensemble of NNs vs. SVR under different sets of inputs and using a correlation-based feature selection algorithm	5 min	5–60 min	Past values of P and/or meteorological data (GHI, T, RH, W)
Soubdhan et al. (2016)	Kalman filter with parameter tuning via AR and Expectation Maximization algorithm	–	1, 5, 10, 30, 60 min	Values of P and/or measurements of T and CC
Sperati et al. (2016)	Probabilistic forecasts based on a PDF modeled via NN and calibrated with the variance deficit and ensemble model output statistics	1 h	0–72 h	Past values of P, meteorological records and predictions from the ECMWF and solar geometry
Vaz et al. (2016)	ANN-NARX with NWP and measurements from neighboring PV plants	–	15 min–1 month	Forecasted GHI and T and P from neighboring plants

for the quantiles 5–20%, penalizing the CRPS. This highlights that an improvement in point forecasts does not necessarily mean an improvement in probabilistic metrics.

AlHakeem et al. (2015) applied bootstrap confidence intervals to quantify the uncertainty of previously calculated deterministic predictions, providing a range of possible future values.

Moreover, Rana et al. (2015) applied 2D forecast. 2D interval forecasts provide a range of expected values for a future time interval. There, they obtained an interval of likely values of solar power, defined by the upper and lower bounds of the distribution.

Golestaneh et al. (2016) obtained predictive densities, described with quantiles 0.05 to 0.95. They highlighted the importance of the reliability and sharpness as a way of evaluating probabilistic forecasts.

As seen, not many authors have applied yet probabilistic forecasts and several challenges are still to be solved. Some of these issues are whether an improved point forecasting model can lead to a better probabilistic forecast and whether the combination of different probabilistic forecasts leads to more accurate models (Hong et al., in press).

7. Temporal horizon

The main way in which forecasts can be classified is according to the time horizon. As will be discussed later on, predictions made for the diverse time horizons are important for different aspects of grid operation, such as maintenance of grid stability, scheduling of spinning reserves, load following or unit commitment.

What follows is a complete classification of studies regarding the time horizon. A general description detailing the main characteristics of each study and their most relevant results was performed. At the end of the section, Table 2 is presented, which summarizes all the studies described in this section, as well as some graphics detailing distribution of publications according to their approach, inputs and spatial scope (Fig. 7), the temporal and spatial horizon/resolution of some sources of inputs (Fig. 8), and forecasting errors (Fig. 9). Since the RMSE is the most widely used metric to present the results in the studies observed, in order to enable some comparisons between models, we have tried to describe results with respect to this metric. It has been proved that longer lead times tend to increase errors, especially under unstable weather conditions. Skill scores (33) were also provided when available. In Section 8 there is a complete assessment of other metrics used, which add very relevant information, as discussed there.

Fig. 6 depicts the distribution of recent publications with respect to the origin of inputs and forecast horizon. As seen, as the time horizon increases, the proportion of studies that incorporate exogenous variables, mainly derived from NWP, becomes larger. Most of studies focused on day-ahead, where most of the electricity is traded, and on intra-day horizons.

7.1. Intra-hour or nowcasting

Very short term forecasting, also denoted as intra-hour or nowcasting, covers forecast horizons from a few seconds to 1 h. It is important to assure grid quality and stability and to correctly schedule spinning reserves and demand response. It becomes a crucial issue when considering island grids or systems with low quality power supply where high solar penetration is present. It also provides benefits at the distribution system operation level, since it allows the reduction of the number of tap operations at the transformers. Nevertheless, robust systems with distributed PV plants do not suffer that much from very short term power generation variability, since aggregating production over large areas tend to balance fluctuations, as seen in Section 5. From the plant

manager point of view, it is important to know the most updated production predictions to be able to reformulate bids in sub-hourly markets (like in EIM) and to learn about ramping rates that affect their plants.

At this time, the main factor causing changes in solar irradiation is the presence of clouds. Cloud generation and movement obey physical rules, but turbulent processes make them appear to be stochastic and very difficult to model (Larson, 2013). Several efforts have been made to understand and predict cloud movement. Bosch and Kleissl (2013) established a relation between cloud motion vectors and irradiance and power measurements in a PV plant, whereas Chow et al. (2015) determined cloud speed and stability using a ground-based sky imaging system. The difficulty of detecting clouds in the Sun region, as well as determining the thickness of clouds was reported (Urquhart et al., 2015; Chu et al., 2015). Peng et al. (2015) faced the problem of cloud thickness determination using several total sky imagers and established spatio-temporal correlations between them, which proved to be useful. Some studies, as will be shown later in this subsection, used cloud-tracking techniques to obtain power forecasts.

There are two main approaches, according to the origin of the inputs, for addressing the intra-hour forecast horizon: models that use only endogenous data (past records from the PV plant) (Chu et al., 2015; Hassanzadeh et al., 2010; Li et al., 2009; Lipperheide et al., 2015; Russo et al., 2014; Al-Messabi et al., 2012; Soubdhan et al., 2016; Rana et al., 2016) and models that use exogenous data: outputs from total sky imagers (Urquhart et al., 2013; Chu et al., 2015), satellite images (Lorenz et al., 2014), NWP (Lonij et al., 2013; Lorenz et al., 2014; Li et al., 2016), meteorological measurements (Soubdhan et al., 2016; Chow et al., 2012; Rana et al., 2015; Takahashi and Mori, 2013; Rana et al., 2016; Golestaneh et al., 2016) or models that incorporate information from nearby PV plants (Berdugo et al., 2011; Lonij et al., 2013; Vaz et al., 2016; Yang and Xie, 2012). As seen, there are some studies that fall into both groups due to the fact that some are comparative studies which analyze different set of inputs.

Models using endogenous data have the advantage of simplicity in data collection, as no sky image processing nor communication with other PV plants are necessary. Regardless of the origin of input data, several studies showed that outperforming persistence models is difficult when working at short time frames. For instance, Urquhart et al. (2013) obtained worse performance for all time horizons considered. Lipperheide et al. (2015) could not beat persistence for time horizons less than 28s, whereas Lonij et al. (2013) underperformed for predicting less than 30 min ahead. Collaboration from neighboring PV plants was not effective

for horizons of less than 15 min in any of the cases. Normally, the skill of the forecasts increased with the time horizon and not many are the studies that worked with time frames inferior to 1 min (Chow et al., 2012; Urquhart et al., 2013; Lipperheide et al., 2015). In comparison to solar irradiance forecasting, where many studies make use of total sky imagers, these devices have not been yet widely applied to solar power forecasting.

7.1.1. Endogenous

From the authors that used endogenous data, two claim that PV plants are "no memory" systems (production in $t + 1$ is only related to operation in t) when talking about forecasting. Li et al. (2009), already starting with that assumption, applied an advance Grey-Markov chain, which is a probability analysis that can forecast the state trend of a data sequence based only in the current state and in a transfer probability matrix, to forecast 2 short periods during one test day. Their model output forecast power ranges and the probability for each range and period of time. They obtained an average relative deviation for time horizons of 15–45 min of −2.78% and −1.26% for each of the two time periods considered. Russo et al. (2014) considered 74 different inputs conformed by several variables and their lagged values (back to 165 min of the current time) and used Genetic Programming with the software tool *Brain Project* (Russo, 2012). Forecast horizons varied from 15 to 60 min in 15 min intervals. Results showed that the model based solely on current power production and on the time difference with respect to sunrise was almost as accurate as more complex ones, concluding that all important information for the forecast is captured by these two inputs. Errors ranged 21.8–35.0% for horizons of 15–60 min. It must be said that metrics used by these authors are different from the most common ones. As will be further discussed in Section 8, an error of i.e. 21.8% means that the error is 0.218 times the standard deviation of the output.

Nevertheless, other authors considered lagged values of plant operation or meteorological records in their best models. Chu et al. (2015) showed that when considering lagged values of variables back to 30 min, forecasts in the 5–15 min horizon could be improved. Also, Lipperheide et al. (2015) used lagged values of power measurements in an auto-regressive model, which proved to be the best performing model for horizons of 94–180 s (performance skills of 5.8–2.9% for 100–180 s horizons, respectively). That AR model was used as benchmark for their proposal of cloud speed persistence. Cloud motion vectors were derived from reference cells which measured the time delay between clouds arriving at different sensors in the plant (Bosch and Kleissl, 2013). With this model, they obtained skill forecasts of 13.9–7.6% for the horizons
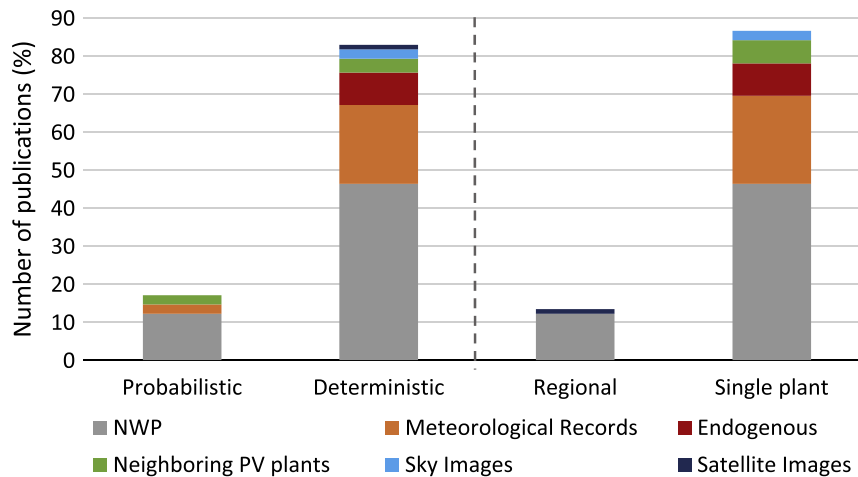

Fig. 7. Number of publications according to their approach and spatial scope. Sample set: studies listed in Table 2.
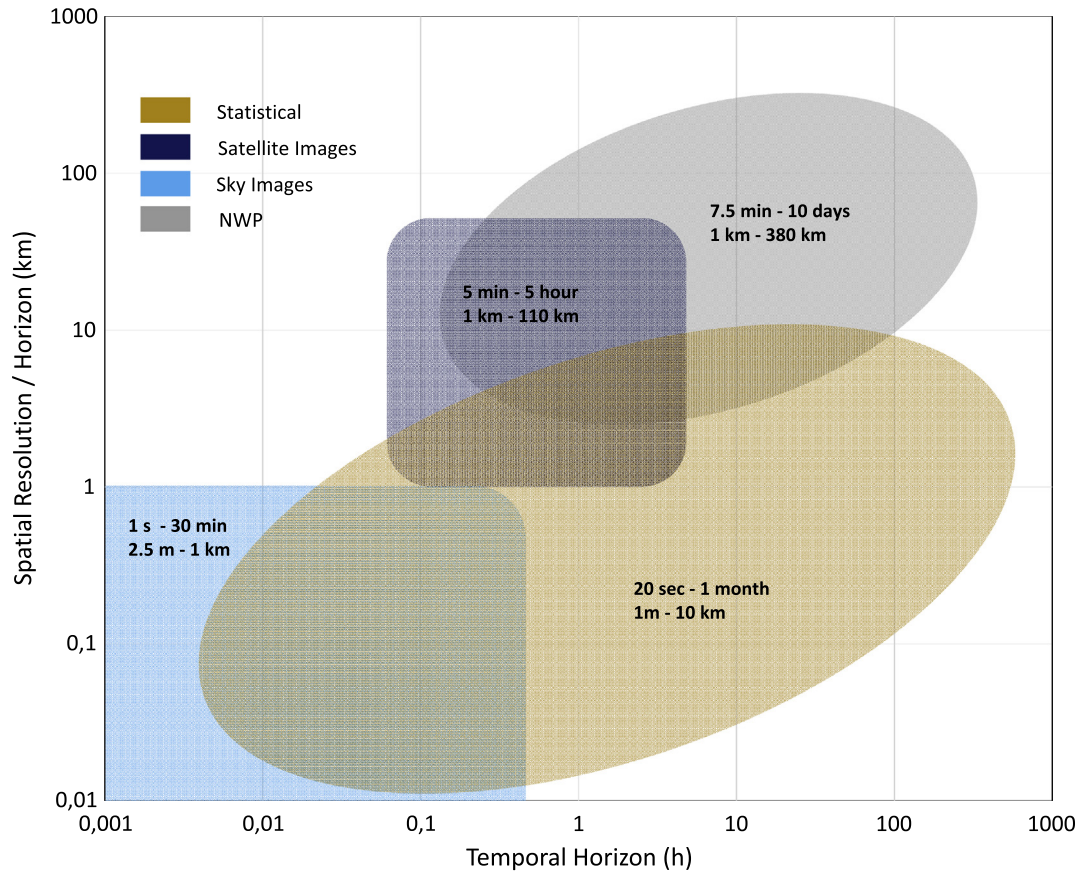
Fig. 8. Distribution of different techniques and sources of inputs with respect to their spatial resolution/horizon and the temporal horizon for which they are used in power forecasting. Time and spatial limits were fixed based on the sample set. Inspired by Inman et al. (2013), Diagne et al. (2013). Sample set: studies listed in Table 2.
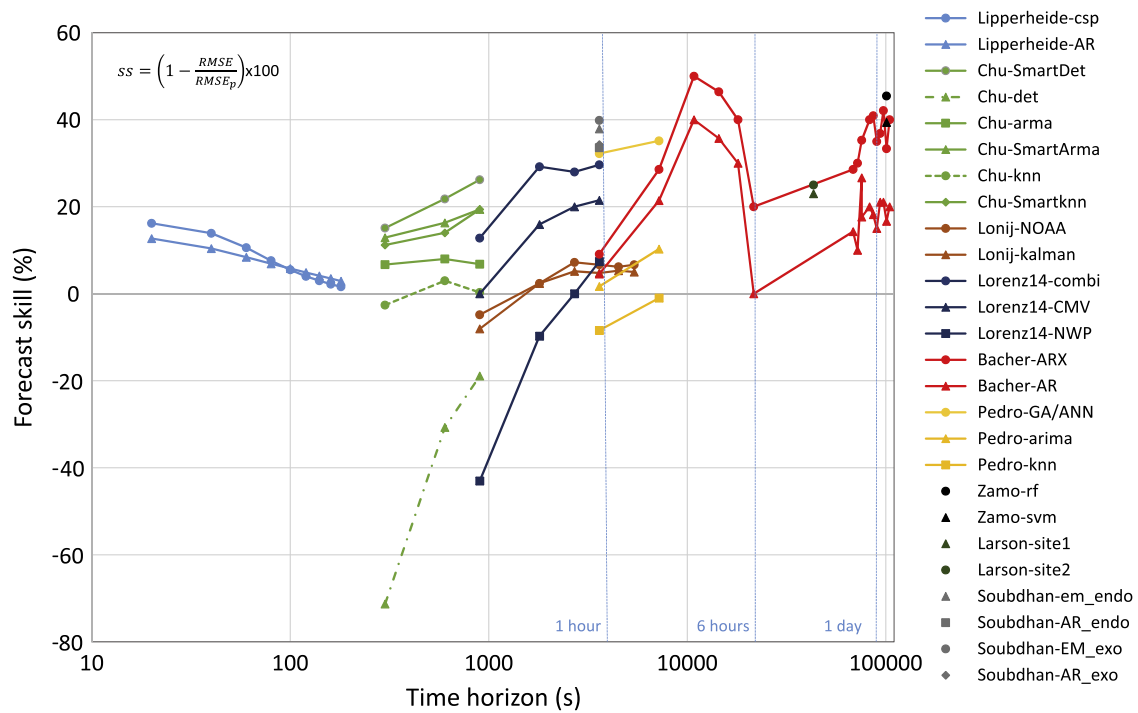

Fig. 9. Comparison of forecasting skill *ss* of different approaches.

40–80 s, respectively. Ramp persistence, however, outperformed for predictions up to 28 s ahead.

Machine learning techniques are widely use for forecasting. Al-Messabi et al. (2012) used dynamic neural networks for predicting 10 and 60 min ahead using as inputs only the actual power yield and lagged power values back to 80 min. They observed best results for the shortest horizon considered.

Soubdhan et al. (2016) performed predictions for time horizons of 1, 5, 10, 30 and 60 min. They considered two sets of inputs, one that consisted only of past PV measurements and another which also incorporated Cloud Cover (CC) and air temperature (T) values from a nearby meteorological station. They used a Kalman filter, where parameter tuning was addressed via two methods: an AR model and an Expectation–Maximization (EM) algorithm. Results showed that the EM algorithm outperformed the AR model, with nRMSEs of 8.29% and 8.87% with endogenous inputs, respectively, and 8.03% and 8.77% for the case of exogenous inputs and EM and AR, respectively. Skill scores of 36.81% and 38.80% were also obtained for the EM algorithm with endogenous and exogenous inputs, respectively. Nevertheless, the Mean Bias Error (MBE) was larger for the case with exogenous inputs. Similarly, Rana et al. (2016) also study the consequences of including meteorological records in the model, in comparison to models with only endogenous data. They compared two techniques: an ensemble of NNs and a SVR model. Variable selection was performed via a correlation-based feature selection algorithm. Univariate models (only endogenous data) performed similarly. The ensemble showed a slightly higher performance than SVR (Mean Relative Error MRE: 7.26 ± 1.37% for ensemble NN). The same trends were observed for multivariate models, which reached also similar levels of accuracy (univariate models slightly outperformed multivariate models). They outperformed baseline models for all the time horizon considered (5–60 min).

Other authors that forecasted at this time horizon using endogenous data preferred an indirect approach. Hassanzadeh et al. (2010) calculated solar irradiation via the sum of a deterministic component and a Gaussian noise signal. The filter was modeled with spectral analysis and ARMA. On partly cloudy days, solar irradiation was modeled with a Kalman filter. Once solar irradiation was obtained, they applied a simple PV performance model to obtain power forecasts in the horizons of 1, 5, 10, 15, 30 and 60 min, which coincide with the sampling rates. Results, averaged for a whole day, showed an increase in nRMSE from 23% to 37% for time horizons from 1 to 60 min, respectively, for the spectral function method. They concluded that sampling rate should be kept as short as possible to account for possible changes in cloud generation.

7.1.2. Sky images

Only two studies were found to use sky imaging products to directly forecast power output. Urquhart et al. (2013) started working with sky images derived from two Total Sky Imagers (TSI) located in a large power plant. They worked with time horizons ranging 30 s–15 min in 30 s intervals. Cloud maps, as well as a histogram of power normalized by the expected clear sky power output, were generated. Clear and cloudy modes were used then to assign values of normalized power to the shaded and unshaded cells of the cloud maps. Thus, the complete power output of the PV plant could be stated. However, this approach lead to underperformance when compared to persistence. Average RMSEs higher than 40% were obtained for each of the two test days.

Encouraged by the previous study, Chu et al. (2015) performed a comparison between techniques (cloud tracking, ARMA and k-NN), applying also a reforecast with ANN. That optimization took as inputs lagged generation values (back to 30 min in 5 min intervals) and predictions from the baseline models. Results showed

that the cloud tracking technique (described above) performed the worst, with forecast skills of −71.3% and −18.9% for 5 and 15 min, respectively, when no optimization was applied. They found two reasons for this poor performance: 2D treatment of clouds, simplifying their 3D nature, and misidentification of cloudy/clear pixels, especially near the Sun. However, when reforecasting was performed, that method outperformed ARMA (forecast skills of 12.9–19.4%), k-NN (forecast skills of −2.66% to 0.3%) and persistence, showing forecast skills of 15.1% and 26.2% for 5 and 15 min horizons, respectively. When no reforecast was made, best performing model was ARMA, with forecast skills up to 8%.

7.1.3. Meteorological records

Chow et al. (2012) built a MLP to obtain real time and 10 to 20 min ahead production using a small number of variables (dry-bulb temperature, solar elevation and azimuth angles and solar radiation). They obtained correlation coefficients (values showing 95% confidence intervals) of 0.84–0.94, 0.89–0.99 and 0.82–0.95 for real time, 10 and 20 min ahead, respectively. Takahashi and Mori (2013) also used ANN and studied 3 different models for 30 min ahead predictions: a Radial Basis Function Network and Generalized Radial Basis Function Networks with clustering via k-means and Deterministic Annealing. They used as inputs actual and lagged values of Power (P) and cell temperature (T_{cell}). Results showed that their Deterministic Annealing model outperformed the other models, having a standard deviation of errors 12% lower than the former model.

Rana et al. (2015) also worked with a machine learning approach, in this case SVR. They introduced the 2D interval forecasts for predicting intra-hour and intra-day. 2D interval forecast can be defined as an interval forecast for a range of future values, what is a continuous probabilistic forecast for a defined future period (more on this topic in Section 6). They also studied different sets of inputs depending on how frequently data was sampled. Results for the 30 min forecast horizon showed a MRE of 6.47% and an Interval Coverage Probability (ICP) of 65.58%, which represented big improvements with respect to baselines (more about these metrics in Section 8).

Golestaneh et al. (2016) worked with Extreme Learning Machines (ELM) to derive predictive densities. They presented their results taking into account the reliability and the sharpness, which they claimed to be the most representative attributes of a probabilistic forecast. They outperformed all baseline models presented with their approach.

7.1.4. NWP

Li et al. (2016) proposed a hierarchical approach to derive 15 min, 1 and 24 h ahead forecasts. They used and compared ANN and SVR based models to predict power output at the inverter level (11 inverters of 500 kW) and then, added individual forecasts to obtain the predictions at the plant level (6 MW). They used lagged values of power output, NWP of several variables as well as solar geometry values to train the models. The hierarchical approach performed better than the 'traditional' one for 15 min and 1 h ahead using either ANN or SVR, whereas for 24 h ahead, the hierarchical and traditional approach were similar.

7.1.5. Neighboring PV plants

Regarding the studies that used data from nearby PV plants, it can be highlighted that none of them could beat the persistence for a time horizon inferior to 15 min. This is because the spatio-temporal correlation with other PV plants is low at this time frame. The way authors approached collaboration between neighboring plants is diverse. Berdugo et al. (2011) preferred to share a small part of information (in the form of analogs) so as to keep most of the operation parameters private between the 11 plants

considered, which were up to 70 km apart. k-NN was used to uncover patterns in energy generation. There was a small tendency for the collaborative method to outperform the non-collaborative approach. It was claimed that if more neighbors were available, that tendency would be clearer. Yang and Xie (2012) used an ARX with spatio-temporal correlations from 5 nearby plants, as far as 200 km from the studied site, to predict solar irradiation and then, using a simple PV performance model, obtained the power output. They outperformed persistence for 1 h horizons, not in the 5–15 min horizons. On the other hand, Lonij et al. (2013) assumed a clear sky production of the PV plant, which was later modified by the presence of clouds (determined by the variations in the output of 80 PV rooftops over a 50×50 km area), whose velocity was obtained by several ways (NWP, Kalman filters or persistence). The model outperformed the persistence for forecast horizons between 30 and 90 min, with forecast skills of -4.8% and 2.4% for 15 and 30 min horizons, respectively.

Finally, Vaz et al. (2016), who applied NARX techniques, used local meteorological data and measurements from close PV plants (5 plants separated around 7 km) in the Netherlands. They studied several input combinations. The best performing model used information from all sites and meteorological data. For 15 min horizons (nRMSE of 0.09), accuracy of NARX was similar to persistence, but for horizons of 30 min (nRMSE of 0.13) on, the proposed model clearly outperformed the baseline.

7.1.6. Summary

Nowcasting can be approached via several ways, where models using only endogenous data or meteorological records are dominant. Methods that incorporate sky images are very promising, although further work has to be address to correctly identify clouds. NWP are not normally used at this time frame since they lack enough temporal resolution. Information shared from other PV plants has been taken into account in several studies, although for short lead times (less than 15 min) they do not provide satisfactory results due to the low spatio-temporal correlation between plants. It has been found that for short time horizons, beating persistence models is a difficult task.

7.2. Intra-day

Intra-day forecasts cover from 1 to 6 h and are important for load-following purposes. Also, they are essential for grid operators that control different load zones or who trade outside of the boundaries of their area (Pedro and Coimbra, 2012).

While for intra-hour forecasts just three of the studies (Lonij et al., 2013; Lorenz et al., 2014; Li et al., 2016) used NWP, for intra-day predictions many of them did so. NWP are a common way to add future information about the atmosphere conditions to the forecasting models. Nevertheless, at short forecast horizons, NWP are not used since they lack the necessary granularity for these issues. NWP are normally taken into account for forecast horizons longer than 4 h, although there are exceptions. At this time horizon, methods using satellite images also appear. As described in Kühnert et al. (2013), satellite images may come from Meteosat Second Generation (MSG) geostationary satellites, operated by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), the Geostationary Operational Environmental Satellite (GOES) for the American region or the Geostationary Meteorological Satellite (GMS) for the Japanese region. From the gathered images, the cloud-index images are obtained. Studying differences between time distributed images, it is possible to derive the Cloud Motion Vectors (CMV). Then, cloud movement is extrapolated to generate forecasts of GHI for the following hours. Finally, output maps are smoothed to reduce

errors. Normally, accuracy of thus created forecasts is adequate up to 5 h.

As in the previous subsection, a broad classification of the works can be made according to the origin of the inputs. Endogenous data was used by Bracale et al. (2013), Hassanzadeh et al. (2010), Pedro and Coimbra (2012), Russo et al. (2014), Yona et al. (2007), Zhang et al. (2015a), De Georgi et al. (2015), Al-Messabi et al. (2012) and Soubdhan et al. (2016), whereas other researchers introduced exogenous inputs, either coming from satellite images (Lorenz et al., 2014), NWP (Alessandrini et al., 2015; Bacher et al., 2009; Fernandez-Jimenez et al., 2012; Fonseca et al., 2011a; Fonseca et al., 2011b; Mellit et al., 2014; Lonij et al., 2013; Monteiro et al., 2013b; Pelland et al., 2011; Lorenz et al., 2007; Lorenz et al., 2008; Lorenz et al., 2009; Lorenz et al., 2011b; Lorenz et al., 2011a; Lorenz et al., 2014; Hong et al., in press; Yona et al., 2013; Li et al., 2014; Huang and Perry, 2015; Wu et al., 2014; Nagy et al., 2016; Jafarzadeh et al., 2013; Schmelas et al., 2015; AlHakeem et al., 2015; Li et al., 2016; Sperati et al., 2016; Mohammed et al., 2015), meteorological records (Mandal et al., 2012; Haque et al., 2013; Soubdhan et al., 2016; De Georgi et al., 2015; Almonacid et al., 2014; Rana et al., 2015; Bouzerdoum et al., 2013; Zeng and Qiao, 2013; Do et al., 2016; De Georgi et al., 2016) or from nearby PV plants (Berdugo et al., 2011; Lonij et al., 2013; Vaz et al., 2016; Yang and Xie, 2012; Bessa et al., 2015; Hong et al., in press; Huang and Perry, 2015).

7.2.1. Endogenous

Models with endogenous data are still widely used for this time horizon. Yona et al. (2007) followed an indirect approach, firstly forecasting GHI via NN and then fitting that prediction into a simple PV performance model. They obtained monthly Mean Absolute Percent Errors (MAPEs) ranging 11–17% for the three horizons considered (1–3 h) with a recurrent NN. An indirect approach was also performed by Bracale et al. (2013). In this work, one hour ahead probabilistic forecasts were obtained from the power output probability density function. To generate that, they first constructed the clearness index probability density function, which was derived from historical data. Then, they predicted the future mean value at time $t + 1$ using a Bayesian auto regressive time series model (relating solar irradiance to meteorological variables). Finally, a PV performance model (relating clearness index to power output) and a Monte Carlo simulation lead to the power predictions. Mean Absolute Relative Errors (MAREs) of 14.5% and 18.0% were obtained for winter and summer, respectively, and real measurements were always between percentiles 5–95% of the predicted power.

The work developed by Pedro and Coimbra (2012) represents a complete assessment of forecasting techniques. They compared ARIMA, k-NN, ANN and GA-ANN for the forecast horizons of 1–2 h using solely past records of power output. They divided the power output in two parts: a part that is explained by a clear sky model and a stochastic part. All the proposed models focused on the prediction of that stochastic part, which is the only one that presents a certain level of uncertainty. They separated results in three variability periods. For low variability, best performing model was k-NN. However, that situation reversed for medium and high variability periods, were GA-ANN clearly outperformed the other techniques. Forecast skills ranged 32.2–35.1% and nRMSE 13.07–18.71% for GA-ANN and the 1 and 2 h horizons, respectively, whereas forecast skills were negative for the k-NN.

7.2.2. Meteorological records

Hybrid techniques have also been applied in forecasting, such as the work by Mandal et al. (2012). They predicted 1 h ahead power output using a combination of wavelet transform and ANN. Inputs were past power measurements, irradiance and temperatures at

times t, t-12 and t-20 h. Past power values were transformed by the Wavelet Transform (WT) to obtain a decomposed approximation signal and detail coefficients. Then, all the variables were fed in the ANN. Finally, the forecast values of decomposed approximation signal and detail coefficients were restored by the WT. Results showed that the hybrid approach outperformed simple ANN, obtaining MAPEs of 4–13% for the Back Propagation NN. Haque et al. (2013) also used WT to decompose present and past values of P and then fed them, in combination with GHI measurements, into a fuzzy Adaptive Resonance Theory Mapping (ARTMAP) NN. Optimization was carried out by a firefly optimization algorithm. With this new methodology, they could outperform all the benchmark models, consisting of variations of WT and different ANN architectures, and obtained MAPEs of 12.11–13.13% depending on the season.

Bouzerdoum et al. (2013) also used a hybrid method, a combination of SARIMA and SVM. The SARIMA model analyzed the linear components of the power, whereas the SVM model was in charge of finding the non-linear patterns from the residuals of the SARIMA model. The hybrid model was found to outperform the other two techniques separately, reaching a nRMSE of 9.40%. Hybrid models were also used by De Georgi et al. (2015), who applied a combination of LS SVM with WD for predicting 1, 3, 6, 12 and 24 h ahead using two sets of inputs: the first one only contained past power measurements and the second added the module and ambient temperature and the I_{POA}. As benchmark, an ANN with WD and LS SVM with no WD models were used. The model proposed outperformed baselines, with nRMSE ranging 9.6–15.28% and 10.66–19.65% for horizons of 1 to 6 h and for the input sets containing all variables and only power measurements, respectively. Best results for input datasets that contained more variables were also obtained in De Georgi et al. (2014). Errors increased up to 22.76% for the longer lead times considered. They performed a complete assessment of the model performance calculating several probabilistic metrics.

Almonacid et al. (2014) created a ANN to predict 1 h ahead module power output, with GHI and cell temperature in $t + 1$ as inputs. To obtain those inputs, they first created two non-linear autoregressive models for forecasting GHI and air temperature from current and previous measurements, with which they obtained later the cell temperature. That data was fed to the ANN model, obtaining a nRMSE of 3.38%. Rana et al. (2015), whose methodology was previously presented in Section 7.1, obtained an average MRE of 9.65% and standard deviation of 1.93% for the 1–6 h horizons, whereas average ICP was 73.07% and standard deviation of 5.83%. These values represent an improvement of 9–44% in terms of MRE and 16–60% in terms of ICP with respect to baselines and ANN model used for benchmarking. Finally, Zhang et al. (2015a) performed hour and day ahead forecasts. They studied 4 different scenarios, considering geographical aggregation and different locations. They analyzed results via several metrics (see Section 8), obtaining a nRMSE of 17% to 2% for the single plant and for a relatively large ensemble of them (64,495 GW), respectively. This finding confirmed that spatial averaging reduces errors (Lorenz et al., 2007, 2008, 2009, 2011a,b, 2014; Mills and Wiser, 2010). Also, a striking MaxAE greater than 70% was found for the single plant and 25% for the relatively large ensemble of them. Such large errors can have a economic impact on grid operation.

Do et al. (2016) investigated on the minimum training period to achieve accurate one hour ahead forecasts. Their inputs were actual PV measurements and values of CC and T. They analyzed three models (ANN, regressive and persistence models) and two different climatic regions. Results showed that the tropical region required a shorter training set (3 months) than the template region (6 months), but once trained, errors were lower for the latter (nRMSE 10.69% in comparison to 11.97% for the regression

method). The regression model was found to outperform the ANN and persistence.

De Georgi et al. (2016) combined a LS SVM model with a novel NN, named Group Method of Data Handling (GMDH), to build an hybrid model, denominated Group least square support vector machine. Three set of inputs, differentiated by the way past predictions were incorporated to the input set, were also evaluated. Results showed that the hybrid model outperformed single models, although only slight differences could be found with respect to the GMDH model (in terms of nMAE). Between the single models, the NN outperformed the LS SVM model.

7.2.3. NWP

Many authors used forecasted variables in their prediction models. Here, we begin resuming most of the work developed by E. Lorenz and colleagues with respect to power forecasting. Even if forecast horizons are diverse, they are collected here for clarity purposes. Lorenz et al. (2007) took NWPs from the European Center for Medium range Weather Forecasts (ECMWF). Since they are in 3 h bases, they analyzed different spatial and temporal interpolation techniques to downscale irradiance forecasts to 1 h. Once horizontal irradiance was known, they converted it into I_{POA} and then, using a PV performance model, into PV power output. They studied single plants, an ensemble of 11 plants and another ensemble of 4500 plants distributed over Germany. Averaged results for up to 3 days ahead and for a single site showed a nRMSE of 13%, while it reduced to 9% for the ensemble of 11 plants and further down to 5% for the case of Germany. They concluded that the increase on forecast accuracy is highly dependent on the size of the region of study. In Lorenz et al. (2008) they further developed the model previously explained, whose goal was the hourly forecast of regional PV plants. In contrast to the previous study, they developed a statistical model to derive the characteristics of the PV plants, since in real situations many variables (exact location, orientation, tilt angle, etc.) are not known. They studied an ensemble of 460 PV plants. Irradiance forecasts were best modeled and corrected as a function of the clear sky index and the solar zenith angle. PV power output for an ensemble of plants was calculated applying a representative system model. They obtained nRMSE of 11%, 6% and 5% for a single plant, an ensemble over a 220×220 km area, and the whole of Germany, respectively. They also studied the possibility of upscaling the results from a subset to the whole ensemble of PV plants. They found that with 30% of the plants they could create a representative subset, obtaining similar accuracy in forecasts for all plants as if they were studied separately. Lorenz et al. (2011a) further developed the upscaling modeling. Upscaling is an important procedure since simulation of all PV systems is very computational demanding and system information is normally not known. Moreover, if performed correctly, upscaling results in almost no loss in accuracy. Here, they grouped representative subsets and compared them to actual distribution, showing in some grids stronger concentrations than actual values. To correct that, they also took into account the distribution of system orientations (azimuth and tilt angles) and the mix of module types. Additionally, a scaling factor was added to single systems, calculated as the ratio between overall installed power and installed power of the subset for that grid point. Results showed skill scores in the range of 30–40% and 40–50% for intraday and day-ahead forecasts. Overestimation was found in winter months, due to snow covering PV modules. This fact was corrected in Lorenz et al. (2011b). There, they introduced a method to correct systematic deviations due to snow. For that reason, they also used temperature and snow depth predictions from the ECMWF. An algorithm that first separated days with high probability of snow cover was created. Then, an empirical equation predicted power output when snow cover was present, considering a factor which

denoted the probability of snow covering PV modules. For single sites, the RMSE could be reduced to one third of the original RMSE (not considering snow cover) and for regional forecasts (size of $2° \times 2°$), the RMSE was lowered to half the prior errors for intra-day forecasts during winter months. When the whole year and region were considered, the new snow-approach also outperformed the original model, with nRMSEs of 3.9% and 4.6% for intra-day and day-ahead, in comparison to the original model, with nRMSEs of 4.9% and 5.7% for the same time horizons. Lorenz et al. (2014) continued with the study of the German PV system, in this case considering 921 PV plants, from which they knew tilt and orientation. They studied time horizons of 15 min to 5 h ahead using PV power measurements, satellite CMV and NWP. Additionally, local and regional forecasts were considered separately. For local and regional forecasts, CMV outperformed NWP for horizons up to close to 4 h. They also studied the combination of different source of inputs (power measurements, CMV and NWP) in a linear regression model (resulting in a considerable improvement of the combined model with respect to basic models, especially in regional forecasts) and with SVR, obtaining similar results.

Bacher et al. (2009) downloaded GHI forecasts from the HIRLAM mesoscale model of the Danish Meteorological Institute to be used as part of the inputs in a ARX model for the forecast of 1–36 h ahead power production of 21 small PV plants in Denmark in 1 h intervals. They compared three different models: AR with past endogenous data, a Linear Model (LM) with the NWP and an ARX with both past data and NWP. Solar power was normalized with respect to a clear sky model to gain stationarity. Results showed that for 1 h horizons, AR outperformed LM, indicating that the most important variable at that time frame was the solar power. For horizons from 2 to 6 h, the LM slightly outperformed the AR, concluding that both had a similar accuracy. Nevertheless, when both group of inputs were combined in the ARX, they resulted in the best performing model, with an averaged improvement with respect to RMSE of 35%.

Pelland et al. (2011) preferred a physical approach in their study. Using Canada's Global Environmental Multiscale model, they applied MOS to spatially average values and to remove bias using a Kalman filter. Post-processing lead to a skill score of 28% with respect to the base NWP in terms of GHI. Once the GHI was known, 12 different model combinations to convert GHI into I_{POA} were tested, as well as two different PV performance models. They concluded that the selection of the model to convert from GHI to I_{POA} had no big impact on final accuracy, since the biggest error came from the prediction of GHI. Also, both PV models performed similarly, although the linear model slightly outperformed the PVSAT, which was also simpler to train. They obtained RMSEs in the range of 6.4–9.2% for the whole time horizon (0–48 h). As pointed out, errors could be further decreased if the horizon of 0–6 h had only been considered for the optimization.

Fonseca et al. (2011a) studied the influence of considering numerically predicted cloudiness from the Japan Meteorology Agency (at three different levels in the atmosphere), in combination with other NWP variables (normalized temperature and relative humidity) and calculated values (extraterrestrial irradiation), in the prediction of 1 h ahead solar power output. For that purpose, they utilized SVR. Apart from outperforming persistence, they found out that not taking into account cloudiness predictions into the SVR model increased the RMSE by 32.6%. A final year-averaged RMSE of 9.48% was obtained for the best performing model. Moreover, Fonseca et al. (2011b) applied the technique described above (SVR with NWP variables) and observed that the yearly averaged RMSE increased 13% from up-to-2 h ahead forecasts to up-to-25 h ahead predictions. Longer lead times increased errors, especially for unstable weather conditions.

Fernandez-Jimenez et al. (2012) made use of two NWP models in a cascade structure. First, predictions from a low resolution model, the Global Forecasting System (GFS) (National Oceanic and Atmospheric Administration, 2003), were downloaded, whose outputs were used as inputs in a higher spatial and temporal resolution model, the MM5 (Dudhia et al., 2005). Then, those outputs fed a set of different forecasting techniques: ARIMA, k-NN, ANN and ANFIS. They predicted horizons of 1–39 h with 1 h resolution. Best performing model was a MLP, with an RMSE of 13.17% averaged for all the time horizons.

Yona et al. (2013) applied fuzzy theory to perform irradiance forecasts using NWP of Relative Humidity (RH) and CC. Then, using a PV performance model, converted prediction of GHI and T into PV power. Jafarzadeh et al. (2013) also used fuzzy systems, in this case, Takagi–Sugeno-Kang (TSK), for 1–3 h ahead forecasting. They used NWP of GHI and T, but instead of being numerical, they were linguistic variables, such as *low* or *mild* for T and *mostly cloudy* or *sunny* for GHI. Using interval type-2 triangular membership functions, they provided probabilistic forecasts which accounted for the uncertainty in predictions.

Monteiro et al. (2013b) used a mesoscale NWP to create a database of weather variables. Considering also past records of power production, they created a model based on transitions recorded in the past (for each variable in the dataset, two values were kept, the present and the previous one). For that purpose, they applied a historical similar mining mechanism, which assigns different weights to cases depending on their similarity with the current case. Later, they calculated power transitions and, once they were known, a probability matrix was created, containing the probabilities for each future forecast interval. Model parameters were optimized via GA. This model not only provided point forecasts, but also the uncertainty associated to them. Results showed that a skill forecast of 36.3% (nRMSE 10.14%) could be obtained and an improvement with respect to MLP (the second benchmark model) of 0.8%, averaged for the whole forecast horizon (1–24 h).

Alessandrini et al. (2015) also performed a probabilistic forecast taking data from the Regional Atmospheric Modeling System (RAMS) and historical data from the PV plant. They forecasted 0–72 h ahead using Analog Ensemble, a method which provides a set of likely PV predictions. The model selected the 20 past situations with lowest distance between forecasted and concurrent values and, based on that, the ensemble forecast was created. The Analog Ensemble model outperformed in statistical consistency both Quantile Regression (QR) and persistence, although similar results with QR in terms of MAE and CRPS were obtained. It slightly outperformed ANN from a deterministic point of view. Authors highlighted the little computational effort it took to make these forecasts. Sperati et al. (2016) continued with probabilistic forecasting in the same time horizon. They took forecasts of meteorological variables from the ensemble prediction system derived from the ECMWF and applied NN to remove bias from those variables and obtain PDFs of solar power. To further improve the predictions, they used two methods: the variance deficit, in order to adjust the ensemble spread, and the ensemble model output statistics, which seeks the minimization of the CRPS. They analyzed the results focusing on the reliability, resolution and uncertainty terms. Results showed that both methods obtained similar levels of statistical consistency.

Mellit et al. (2014) performed weather classification in accordance to forecasted GHI. That variable, along with forecasted cell temperature and current P, were fed into a ANN trained for the specific weather conditions to predict P in day-ahead basis. MAPEs of 1.8%, 4.0% and 3.6% were obtained for sunny, partly cloudy and overcast days.

Wu et al. (2014) constructed a hybrid model combining ARIMA, LS SVM, ANN and ANFIS techniques for predicting 1 h ahead. As

inputs they only needed historical power measurements and NWP of GHI. To calculate the weights of each single forecast they applied a GA. It was shown that the hybrid model outperformed each of the techniques that conformed it, obtaining nRMSEs ranging 3.43–6.57% depending on the test case.

Schmelas et al. (2015) compared three different forecasting techniques (physical, statistical *MLP* and hybrid, which incorporated physical information) for the energy prediction of a positive energy building. Predictions were made from 1 to 12 h ahead. Special focus was applied to the study of shadows, which end up being very problematic. NWP of *T*, RH and GHI were obtained from the Deutscher Wetterdienst (DWD). Best results were obtained for the hybrid model.

AlHakeem et al. (2015) forecasted 1, 3 and 6 h ahead. They applied Wavelet Decomposition to actual PV power values, which were added to predictions of GHI and *T* into a ANN, optimized via PSO. Results were reconstructed and the uncertainty of forecasts was obtained using the bootstrap method (see Section 6).

Zeng and Qiao (2013) predicted 1–3 h ahead using LS SVM. They used several meteorological variables to model atmospheric transmissivity and then, converted it to solar power as a function of the latitude and time of the day. The model proposed outperformed other benchmarking approaches (AR and Radial basis function NN).

7.2.4. Neighboring PV plants

Regarding the use of information from neighboring PV plants, two new studies in this time horizon are found (Bessa et al., 2015; Huang and Perry, 2015). The same studies as in the intra-hour domain also persist, whose performances were described above. The work developed by Bessa et al. (2015) focused on point and probabilistic forecasts in a smart grid, for time horizons of 1 to 6 h. They used a Vector Auto-Regression framework (VAR), which is an auto-regressive network with information both temporal and spatial. They worked at Low Voltage (LV) levels with VAR and at Medium Voltage (MV) levels both with VAR and VARX, which included information from the smart meters at the LV level to account for the influence of clouds. Also, with the aim of benchmarking, a classical AR was also calculated. At the secondary level (MV), VAR and VARX models outperformed AR in around 7.5–1.6% and 10–2.5% for lead times of 1 to 6 h, respectively. As pointed out, improvements decayed with the time horizon, showing that distributed information is more important for the first 3 h. This is because the area of study is small and cloud events cross the area in a short period of time. A similar conclusion was reached by Vaz et al. (2016). At the primary level, VAR outperformed AR, revealing the same trend: improvements dropped with the time horizon. An average improvement of around 12.5–2% over AR could be found for horizons of 1–6 h, respectively. High variability between cases (smart meters in LV and Distribution Transformer Controllers (DTC) in MV) was reported. For the probabilistic forecasts, VARX model outperformed AR in all lead times and VAR for horizons between 2 and 6 h at the DTC level using the CRPS. Nevertheless, at the primary level, improvements were negative for lead times 4–6 h. This is explained by a poor performance of VAR in some quantiles, which impacts the overall performance.

To conclude this part we present, as a summary, the main findings of the Global Energy Forecasting Competition 2014 GEFCom2014, detailed in Hong et al. (in press). Participants had to make probabilistic forecasts for the horizon of 1–24 h with hourly resolution. They were provided information from NWP and had to predict the output of three solar power plants in Australia. It must be highlighted that the five best models were nonparametric and they usually applied variations of QR and gradient boosting. Moreover, in a similar competition, the American Meteorological Society 2013–14 Solar Energy Prediction Contest, aiming at the prediction of daily solar energy in 98 sites, the top participants also used

Gradient Boosted Regression Tress in their models (Aggarwal and Saini, 2014), which reveals their good performance. Back in the GEFCom2014, whereas four of the five best models used information from the neighboring PV plants to issue the forecasts, the team that ranked second (Nagy et al., 2016) only used onsite information (in addition to the aforementioned NWP). This proves that a proper selection of onsite inputs and techniques can lead to satisfying results without falling in the complexity and higher dimensionality derived from including offsite information.

Huang and Perry (2015) took part in the GEFCom2014 and used both NWP variables from the ECMWF and production from neighboring plants. They chose the gradient boosting machine for the deterministic forecast, which they fed with variables from NWP, variables obtained from physical equations and data from the neighboring PV plants. However, they went for k-NN regression for the probabilistic analysis. They obtained an overall quantile score of 0.01211 using the pinball loss function (see Section 8). Nagy et al. (2016) employed regressor ensembling to issue the forecasts. As said before, they did not use information from neighboring PV plants. They tried four ensembling methods (voting, bagging, boosting and stacking predictors) and obtained best results with a voted ensemble of QR Forests (QRF) and a stacked RF. Their overall quantile score was 0.01241. Mohammed et al. (2015) used the data from the GEFCom2014 an applied an ensemble of statistical techniques (decision tree regressor, random forest regressor, k-NN regressor (uniform), k-NN regressor (distance), ridge regression, lasso regression and gradient boosting regressor) combining their results via a naive model, a normal distribution and a normal distribution with different initial settings. For individual models, best results were obtained with random forests and gradient boosting (average pinball loss of 0.0194 and 0.0193, respectively). Nevertheless, when ensemble models were created, they outperformed all single models. The combination via the normal distribution with different initial settings obtained the best results, with an average pinball loss of 0.0148.

7.2.5. Summary

There exists a wide range of options to issue intra day forecasts which have proven satisfactory results. Because of this, it is difficult to select a single best performing technique. Most selected models were ANN, SVR and regressive methods. Important lessons can be extracted also from the forecasting competitions described herein, since all researchers worked with the same database. There, gradient boosting machines obtained best results. On the other hand, NWP variables were extensively used at this time frame. The used of cloudiness predictions and variable selection algorithms showed to improve forecasts. Taking into account information from neighboring PV plants decreased forecasting errors when applied.

7.3. Six hours to day ahead

Forecasts from 6 h up to day ahead are normally used for planning and unit commitment. They cover horizons from 6 to 48 h, depending on when forecasts are issued. If for intra-hour forecasts there were only three studies introducing NWP variables, that amount increased to 57% in intra-day forecasts and further to 79% for the day ahead. These numbers give an insight on the benefits of using NWP when time horizons increase, since NWP are able to feed models with some future meteorological trends. Thus, when the same classification as in previous sections is shown, it is seen that only a few authors limited to endogenous data (Kudo et al., 2009; Rana et al., 2015; De Georgi et al., 2015; Haque et al., 2013; Masa-Bote et al., 2014; Long et al., 2014), whereas most of them took into account NWP (Alessandrini et al., 2015; Bacher et al., 2009; Mellit et al., 2014; Chen et al., 2011; Dolara

et al., 2015a; Fernandez-Jimenez et al., 2012; Lorenz et al., 2007; Lorenz et al., 2008; Lorenz et al., 2009; Lorenz et al., 2011b; Lorenz et al., 2011a; Fonseca et al., 2011b; Lu et al., 2015; Monteiro et al., 2013a; Pelland et al., 2011; Almeida et al., 2015; Ramsami and Oree, 2015; Shi et al., 2011; Tao et al., 2010; Zamo et al., 2014a; Zhang et al., 2015a; Hong et al., in press; Leva et al., 2015; Ogliari et al., 2013; Simonov et al., 2012; Yang et al., 2014; Gandelli et al., 2014; Ding et al., 2011; Yona et al., 2013; Larson et al., 2016; Huang and Perry, 2015; Fonseca et al., 2014a; Fonseca et al., 2014b; Fonseca et al., 2014c; Fonseca et al., 2014d; Fonseca et al., 2015; Chupong and Plangklang, 2011; Hossain et al., 2013; Masa-Bote et al., 2014; De Felice et al., 2015; Schmelas et al., 2015; AlHakeem et al., 2015; Li et al., 2016; Monteiro et al., 2013b; Mohammed et al., 2015; Sperati et al., 2016). Meteorological records were also used (Mellit and Massi Pavan, 2010; Mora-Lopez et al., 2011; Oudjana et al., 2013; De Georgi et al., 2016). Some others considered information from neighboring PV plants (Bessa et al., 2015; Vaz et al., 2016; Yang and Xie, 2012; Hong et al., in press; Huang and Perry, 2015).

7.3.1. Endogenous

Kudo et al. (2009) compared an indirect and direct approach to issue 18–31 h ahead forecasts. Results showed that the direct method performed better, achieving a mean error of 25.6%, in comparison to 28.1% of the indirect method.

Masa-Bote et al. (2014) compared an indirect and a direct approach. For the former, they used NWP of GHI and T and a PV performance model which took into account shading, optical and system losses. However, for the latter approach they used an ARIMA model but using only endogenous data. The forecasts were applied to a self-sufficient solar house, were it was also possible to storage energy and there was a ADSM system to schedule the deferrable loads along the day (programmed to maximized self-consumption). Results showed that the physical approach performed better than the ARIMA, obtaining MAPEs of 37.2% and 42.3%, respectively. Although other studies proved that the direct approach was more accurate than the indirect, it has to be stressed that the two models studied here were not fed with the same inputs (endogenous/NWP) and thus, results cannot be directly compared. An interesting conclusion of this work is that when the energy exchanged with the grid is forecasted (taking into account self-consumption with the ADSM technique), its mean daily error is 2%, compared to errors of 40% if all the energy produced were fed into the grid.

Long et al. (2014) carried out a benchmark of techniques for predicting daily solar power from the same day and up to three days ahead. They analyzed two scenarios: in the first one, records of dew T, RH, insolation time, W, precipitation, GHI and P with lagged values fed models created with ANN, SVM, k-NN and Multivariate Linear Regression (MLR). In the second approach, only records of power were used. For the scenario with all the variables, the MLR outperformed the others in most horizons, while in the case with only power values, ANN outperformed for intra day and day ahead predictions. None of the models proved superior to other for all time horizons.

7.3.2. Meteorological records

Mellit and Massi Pavan (2010) chose an indirect approach in their study of day-ahead forecasting. First, they predicted solar irradiance using present values of mean daily solar irradiance and air temperature with a MLP structure. A simple PV performance model considering fixed module efficiency and balance of system converted solar irradiance into power output. They took for the test case 4 sunny days. MAEs were below 5% for all days. Also, Mora-Lopez et al. (2011) employed an indirect approach to forecast day ahead production. They only used past values of

power, solar irradiance and temperature recorded onsite. They proposed a Probabilistic Finite Automata (PFA) and multivariate regression. The method was divided in 3 steps: in the first one, a linear regression was applied to predict the clearness index. In the second step, the PFA was created with the significant variables of each group. The values of irradiance were predicted using the PFA and the predicted values of clearness index. Finally, taking a physical equation, they predicted power output. They obtained an average mean prediction error of 16.25%.

The work developed by Oudjana et al. (2013) focused on forecasting 1 to 7 days ahead. They created 6 different models, according to the inputs considered and techniques used. Each set of inputs (temperature, irradiance and both) fed linear/multiple regression and ANN, obtaining best results for the ANN with the 2 inputs.

7.3.3. NWP

Apart from the work developed by E. Lorenz et al., described in Section 7.2, other studies are found. Many authors applied weather classification prior to training, so that models were fed with similar data to that forecasted. Chen et al. (2011) developed an online forecasting model for 24 h ahead using a Self Organized Map (SOM) to classify weather (sunny, cloudy and rainy) in accordance to meteorological variables (solar irradiance, total cloud amount and low cloud amount). Thus, the ANN-RBFN could be better trained. They tested several days and obtained MAPEs of 9.45%, 9.88% and 38.12% for sunny, cloudy and rainy days, respectively. A similar approach was followed by Shi et al. (2011), who also classified days according to their characteristics (sunny, cloudy, foggy and rainy). Then, using the historical data from the previous time step and the weather report of the next day, they built different models with SVM depending on the type of day to predict the power output. They obtained and average RMSE of 10.5% and a MRE of 8.64%. The sunny model performed the best and the cloudy the worst, with 4.85% and 12.42% in terms of MRE, respectively. Similarly, Ding et al. (2011) selected the closest historical record with respect to weather conditions to the forecasted day and then trained a ANN with that data for day-ahead predictions. They obtained a MAPE of 10.06% and 18.89% for sunny and rainy days, respectively. Weather classification of the database was also done by Yang et al. (2014), who proposed a methodology for day-ahead forecasting that included three stages. In the first one, after collecting historical data, a SOM and a learning vector quantization were applied to classify the data in different weather types. Then, they trained several SVR models and using the NWP for the forecast horizon, they selected, with the aid of a fuzzy inference model, the most suitable SVR model. They obtained an average nRMSE of 7% for their proposed model, outperforming both traditional SVR and ANN. Fonseca et al. (2015) generated prediction intervals based on input data similarity (computed via the Euclidean distance). From the selected "close" situations, they studied the forecasts errors and represented them following Laplacian and Gaussian distributions. Associating prediction intervals to reserves of power to deal with PV power uncertainty led to the proposed model outperforming the reference model. High forecast error coverages (ranging 97–98%) were obtained with substantially less reserve power than the reference.

Tao et al. (2010) created a NARX model that used NWP but no irradiance predictions, reaching a MAPE of 16.47% and outperforming a ANN model used for benchmarking for day ahead forecasts. Chupong and Plangklang (2011) neither used irradiance predictions from NWP (only T_{max}, T_{min} and cloudy index), but fed their Elman NN with the output from a clear sky model. They obtained a MAPE of 16.83% for day ahead predictions. Hossain et al. (2013) proposed an ensemble method combining three different regression techniques (selected among a set of ten regression tech-

niques), formed by MLO, SVM and least median square, for predicting power output 6 h ahead.

Monteiro et al. (2013a) compared two models for day ahead predictions. The first approach was analytical and modeled the differences between real and clear sky production via the irradiance and PV attenuation indexes. The second method was a MLP model. Both approaches showed similar errors, reaching RMSEs around 12%.

Zamo et al. (2014a,b) made a very complete study about the incorporation of NWP in the power forecasts. The latter will be described in subSection 7.4, as the forecast horizon is 2 days. In their first work, they selected a time horizon of 28–45 h to perform a benchmark of statistical regression methods. They used ARPEGE model, developed by Meteo France, and selected several variables. For benchmarking they selected 3 models: persistence (2 days before), linear model with the clear sky global irradiation as the only predictor and linear model with the downward solar irradiation forecast by ARPEGE as the only predictor. The more complex models were LM with all the predictors, binary regression tree, bagging boosting, RF, SVM and generalized additive models. Results showed that out of the baseline models, persistence was the best, and random forests outperformed in the other set of techniques. They also studied an upscaling technique, which consisted of creating a virtual reference power plant. Results showed this approach obtained comparable results to summing the power predictions of each single plant.

Regional forecasting was also a task performed by Fonseca and collegues, who focused on predicting in Japan, considering day ahead horizons (18–31 h, daylight values only). They started with the evaluation of regional forecast errors in comparison to single site predictions (Fonseca et al., 2014a). They applied a method consisting of summing individual forecasts and observed a smoothing effect of 30–45% depending on the region considered, with nRMSEs of 0.059–0.069 kW h/kW$_{cap}$. Then, Fonseca et al. (2014c) continued with the study of regional forecasts using data from 453 PV systems grouped in four regions. NWPs from the Japan Meteorological Agency of T, RH and CC were used, as well as values of I_{ex}, of the 60 previous days to the forecast day. To select the most relevant variables, a principal analysis component (PCA) was applied to a covariance matrix made of all the input variables, which had already been proven successful (Fonseca et al., 2014b). They selected those variables that explained 90% of the cumulative variance as inputs for a SVR model. Thus, they eliminated redundant information and simplified the predicting task. Results showed that the PCA model yielded better results than models without it. Skill scores of up to 80% and annual nRMSEs of 7% were reported. Moreover, this model also proved superior to the methodology consisting of summing the individual predictions of every PV system, except for a large region covering different climates. Fonseca et al. (2014d) further investigated the study evaluating four methods for different scenarios regarding availability of data. They considered the same variables as in the previous studies and focused on two regions, with different climates and size, and a third one formed by those two (273 PV systems in total). Two of the methods, summing individual forecasts (M1) and directly forecasting regional PV (PCA approach) (M3) had already been introduced in former articles and two new ones were proposed. One of them was based on stratified sampling (M2), suitable for situations when the individual production of some systems is known but regional generation is not measured. Thus, a small set of PV systems are chosen as representatives, from whom the regional production is derived. The last model is proposed for cases where no PV power data is available (M4). Then, PV power forecasts are derived from irradiance forecasts (for more information about these models, see Section 5). Results showed that for areas with similar weather conditions and little snow fall, all models performed similarly. Nevertheless, for the Chubu region (large and with variety of climates), models M1 and M2, which better account for the smoothing effect, outperformed the others. M2 advantages M1 in the sense that just a small subset of systems has to be modeled. Skill scores and nRMSEs ranging 79.2–84.3% and 0.255–0.345 kW h/kW$_{cap}$ were obtained for all the cases, respectively. Method M3 was the most stable one and M4 yielded poor results in winter months. Even if irradiance forecasts benefited from the smoothing effect, this was not directly translated into power. They found two reasons for the decoupling of power forecasts with respect to irradiance forecasts: the snow accumulated on PV panels (causing poor performance in Lorenz et al. (2011b) as well), which irradiance forecasts do not take into account, and the characteristics (installation, efficiencies, etc.) of the PV plants, which are not fully modeled as in M1 model. However, as authors highlighted, each model can be applied in situations were others cannot, so none of them should be neglected.

Dolara et al. (2015a) predicted 24–72 h ahead using a method consisting of PHANN. This model was firstly introduced by Gandelli et al. (2014) but in the new study they applied a mobile window forecast, which updated the training set with new data. They presented the model as a "grey-box", taking the best characteristics of both the physical method and the ANN. It combined an ANN with a PV performance model describing the clear sky solar radiation. The aim of using the clear sky model was to know the time period between sunrise and sunset and to compute the maximum solar radiation available. Then, they combined NWP with the said model. The PHANN outperformed the ANN in most days, getting a 50% reduction in the nMAE for many days.

Lu et al. (2015) presented day ahead predictions based on a machine learning blended method that combined NWP from three different models (NAM, Rapid refresh *RAP* and High-Resolution Rapid Refresh *HRRR*) as inputs for a PV performance model. Results showed that the blended model outperformed single models, permitting a reduction of 30% in MAE compared to single models. Such reduction was mainly explained by the cancelation of systematic bias errors in individual forecasts. Zhang et al. (2015a), who obtained day-ahead forecasts by 3TIER, based on NWP, got a nRMSE of 22% to 4% for a single plant and for a large ensemble, respectively. Hour ahead forecasts were more accurate than day-ahead. They also revealed that the accuracy difference between forecasts at different time horizons (hour to day ahead) increased with the area of aggregation in terms of MAE.

Day ahead predictions was a task also addressed by Ramsami and Oree (2015). They created a model with stepwise regression-feedforward NN hybrid model. The regression method was adjusted to determine the most important variables from the NWP to be used later on the ANN. This hybrid method outperformed the linear regression and ANN taken separately, showing RMSEs of 2.74%. Simonov et al. (2012) and Ogliari et al. (2013) worked with ANN for their day-ahead forecasts. The former used ANN with a fuzzy model to preprocess a subset of weather data to filter the inputs of the ANN, while the latter optimized the variable selection via a combination of GA and PSO, called Genetical Swarm Optimization *GSO*. For the latter, the yearly absolute error was 20.01% and the ANN-GSO model outperformed the simple ANN used as baseline. Li et al. (2014) compared different time series models, highlighting the superiority of ARMAX fed with NWP compared to ARIMA for day-ahead prediction. They obtained a MAPE 82.69% compared to 104.10% of the ARIMA. De Felice et al. (2015) predicted 1–10 days ahead without using local meteorological measurements. Instead, they obtained irradiance records from the Satellite Application Facility on Climate Monitoring (CMSAF) and temperature values from the E-OBS dataset, whereas predictions of those variables were got from the ECMWF. They applied

SVR and observed larger errors for longer forecast horizons and for regions with higher weather variability.

Leva et al. (2015) applied an ANN fed with historical irradiance and power measurements and NWP to forecast day-ahead production. They analyzed the sensitivity of the model to different input sets, obtaining an average nRMSE of around 16%. Almeida et al. (2015) also presented 24 h ahead predictions using QRF. Their methodology consisted of the collection of previous power measurements and the download of NWP variables from the forecasting service Meteogalicia, with which they trained the QRF. For the NWP variables they interpolated from surrounding cells to obtain point forecasts. They used several indexes to quantify spatial (terrain ruggedness, topographic and roughness indexes) and time variability (standard deviation of collection forecasts). They also studied different training sets: based on previous samples, on the similarity of the clearness index and on the similarity between the empirical distribution function of the intradaily irradiance forecast for the day to be predicted and for each day included in the database. They studied 17 different scenarios depending on the variables and training set selected. Results showed that selecting the training set based on the similarity of the empirical distribution yielded best results. Other conclusions drawn from the assessment are that 30 days seemed to be enough for the training set, Sun geometry variables were included in top scenarios as well as solar irradiance predictions. The accuracy of the model was not compromised if a small set of predictors were used and variability indexes slightly improved predictions. They obtained a skill score of 33–36%. They also presented the coefficient of variation of MBE and MAE to account for possible penalizations: if markets penalize daily energy error, the metric to be used is the cvMBE, but if they penalize the hourly energy error, then it is the cvMAE. They obtained cvMBE inferior to 1.3% and cvMAE less than 9.5%.

Larson et al. (2016) investigated day ahead production of two PV plants using NWP from two models. The NAM forecast system outputs GHI, but the Regional Deterministic Prediction System (RDPS) does not, so the forecasted CC allowed them to obtain GHI. Then, they applied MOS via spatial averaging to correct said GHI. Finally, a PV performance model allowed them to obtain power output from the GHI. On clear days, they could not beat persistence, while on overcast days their model proved superior. Averaged for all conditions they outperformed the persistence, with best nRMSE in the range of 9.3%. For single plant forecasts, over prediction errors larger than 20% of the plant rated capacity occurred during less than 9% of the total study time (over three years, hourly aggregation), although averaging predictions with a neighboring PV plant reduced such deviations due to the smoothing effect.

7.3.4. Neighboring PV plants

Yang and Xie (2012) obtained a nRMSE of 16.90%, averaged for the case studies proposed. They pointed out that for day ahead forecasts, the model which considered information only from the closest PV plant was more likely to obtain better results than models that took into account information from more PV plants. Contrarily, hour ahead forecasts were more accurate when data from the surrounding PV plants was considered. Vaz et al. (2016) showed almost no improvement with respect to persistence (RMSE of 19% and 20%, respectively) in the 24 h ahead forecast, contrasting with the results obtained for shorter time horizons.

7.3.5. Summary

As in other time horizons, the diversity of techniques used makes it difficult to extract clear conclusions about which are the best techniques or sets of predictors. However, it can be pointed out the supremacy of models that use NWP variables. The selection of the training database according to weather conditions instead of on previous days proved satisfactory results. Incorporating Sun geometry variables into the suite of predictors also showed to improve accuracy. A trade-off between complexity of the model and accuracy may be established to limit computational times.

7.4. Two days ahead or longer

Forecasts of 2 days ahead or longer are used for unit commitment, transmission management, trading, hedging, planning and asset optimization (Zamo et al., 2014a). They are also important for planning plant maintenance in a cost-effective way, that is, when the expected production is low. They cover horizons longer than 48 h. In comparison to other forecast horizons, here, not so many studies are found. Using the same classification as above, some studies predicted using endogenous data (Lin and Pai, 2015; Long et al., 2014; De Georgi et al., 2015). The rest used exogenous data, either from NWP (Alessandrini et al., 2015; Dolara et al., 2015a; Lorenz et al., 2007, 2008, 2009, 2011b,a; Zamo et al., 2014b; De Felice et al., 2015; Sperati et al., 2016), meteorological records (Oudjana et al., 2013; De Georgi et al., 2015; Long et al., 2014) or from neighboring PV plants (Vaz et al., 2016).

7.4.1. Endogenous

Lin and Pai (2015) predicted monthly power output for the ensemble of PV plants in Taiwan using only past values of *P*. To account for seasonal effects, they applied a seasonal decomposition (SD) to detrend and seasonally adjust the time series. Then, LS SVR were used, as the work by De Georgi et al. (2015), in which they applied GA to select the best set of inputs. Results showed that the proposed SD LS SVR model outperformed the benchmarking models (SARIMA, ARIMA, generalized regression NN and LS SVR) with a MAPE of 7.84%. Second best model was SARIMA, which proved the importance of seasonal decomposition.

7.4.2. NWP

With respect to the models that used NWP, the only new one (the rest were explained in previous sections) is the work developed by Zamo et al. (2014b). Similarly to the twin study they conducted (Zamo et al., 2014a), in this one they adopted a probabilistic focus to forecast daily production 2 days ahead (66–72 h ahead) and compared several statistical quantile regression methods. They worked with 35 different NWP models and concluded that no one performed better than the others. Conclusions were not clear either whether to use just one NWP or the ensemble. They studied the rank histogram calibration, which tried to improve the reliability of the forecasts through correcting the cumulative distribution factor with information from the rank histogram done with the training set. Results proved that calibration worsened the forecasts. They explained it by the interpolations made in the method, which in quantiles with low amount of data could be wrong. Additionally, dissimilarities between training set and test set can introduce more errors. Despite having large amount of data, it was not possible to state that corrected forecasts were better than uncorrected ones. They benchmarked against a simple monthly climatology model and all QR models performed better, reducing the CRPS by 25–50%.

7.4.3. Neighboring PV plants

Vaz et al. (2016), whose model was previously explained in Section 7.1, obtained better results with the NARX model compared to the persistence for 4, 7 and 15 days ahead (nRMSE of around 20%), whereas for larger horizons (20 days to 1 month), RMSEs went up to 24%, equaling persistence.

7.4.4. Summary

This time frame does not see as many publications as the previous since electricity markets are more active at shorter time horizons. NWP variables are commonly incorporated whereas the use of information from neighboring PV plants decreases due to the low spatio-temporal correlation of PV plants at this horizon.

7.5. Summary of papers

Table 2 depicts all recent studies on solar power forecasting, previously detailed in this section, showing the forecast horizon and resolution, the main characteristics of the model proposed and the variables used.

Fig. 7 depicts a classification of studies regarding their approach (deterministic/probabilistic) and spatial scope (single plant/regional) and the origin of inputs used. As seen, roughly 80% of the studies are for single plants and follow a deterministic analysis. NWP, as observed, constitute the main source of inputs.

Fig. 8 represents the distribution of different forecasting techniques and sources of inputs considering their spatial resolution/horizon and the temporal horizon in which they are applied. NWP, satellite images and sky images are plotted against their spatial resolution, whereas the statistical approach is represented against its spatial horizon (using only endogenous data). If inputs from NWP models, satellite or sky images are fed into statistical models, the spatial horizon of the statistical approach would expand, but that situation is not represented in Fig. 8 for clarity.

Finally, to conclude this section, Fig. 9 depicts a comparison of forecasting skills (with respect to naive persistence) of those studies that presented it or allowed its calculation. What first comes to sight is the small number of articles included in the comparison. The reason for this is the absence of agreement on which are the most adequate metrics to show results. A wide variety of indicators are calculated and finding a common metric that enables comparison becomes difficult. As a general rule, the forecasting skill increases with larger lead times, with some exceptions, such as the work by Lipperheide et al. (2015), who focused on the cloud movement over a small area penalizing forecasts at large time horizons. In the hour ahead horizon, several studies are found, achieving up to 40% of forecast skill.

8. Performance metrics

The performance and accuracy of a certain model can be assessed via several metrics. Metrics permit the comparison between different models and locations. Each one focus on a certain aspect of a point distribution. Thus, there is not a unique metric valid for all situations; instead, each one adds some information about the accuracy of the model. In the bibliography, several metrics can be found, although there are a group of them that are more commonly used. In the last years, some authors have focused on different metrics, which set the foundations for more complete assessments. Zhang et al. (2015a) made a through review of forecasting metrics.

8.1. Classical statistical metrics

Normalized Error (nE)

$$nE = \frac{P_{pred} - P_{meas}}{max(P_{pred})} \tag{20}$$

where P_{pred} is the predicted power output and P_{meas} is the measured power output.

Mean Absolute Error (MAE). It shows the average distance between the measured values and the model predictions. It is suitable for evaluating uniform forecast errors.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_{pred} - P_{meas}| \tag{21}$$

It can be normalized with respect to diverse factors, such as the predicted or measured power; then it is called Mean Absolute Relative Error (MARE), rMAE or MAE/Avg.

Mean Bias Error (MBE). It tells if the model over/underestimates.

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(P_{pred} - P_{meas}) \tag{22}$$

Standard Deviation of Errors (SDE).

$$SDE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{pred} - P_{meas} - MBE)^2} \tag{23}$$

Root Mean Square Error (RMSE). It penalizes large errors in a square order.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_{pred} - P_{meas})^2} \tag{24}$$

The above mentioned equations are related to each other by

$$RMSE^2 = MBE^2 + SDE^2 \tag{25}$$

Mean Absolute Percentage Error (MAPE). Like the MAE, it should be used to evaluate uniform forecast errors.

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{P_{pred} - P_{meas}}{P_0}\right| \tag{26}$$

where P_0 is the capacity of analyzed PV plants.

Median Absolute Percentage Error (MdAPE). It is less sensitive to outliers than the MAPE (De Felice et al., 2015)

$$MdAPE = median\left(\left|100\frac{P_{pred} - P_{meas}}{P_{meas}}\right|\right) \tag{27}$$

Coefficient of Determination (R^2) or Pearson's coefficient. It shows how correlated the forecasted and real values are.

$$R^2 = 1 - \frac{Var(P_{meas} - P_{pred})}{Var(P_{pred})} \tag{28}$$

Correlation Coefficient (ρ)

$$\rho = \frac{(Cov(P_{meas}, P_{pred}))^2}{Var(P_{meas})Var(P_{pred})} \tag{29}$$

Skill Score (ss) or forecasting skill was firstly defined as the ratio of the uncertainty U of solar forecasts to solar variability V. Marquez and Coimbra (2013) defined those two terms, which have been adapted here for PV power, (30) and (31).

$$V = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\Delta k(t))^2} \tag{30}$$

where Δk are the power step-changes depending on which I (I_{cs} or I_0) has been used in the calculation.

$$U = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left\{\frac{P_{pred,t} - P_{meas,t}}{P_{cs,t}}\right\}^2} \tag{31}$$

The skill score is then defined:

$$ss = 1 - \frac{U}{V} \tag{32}$$

As the uncertainty of a forecast diminishes, *ss* approaches to 1. A persistence forecast presents $ss = 0$. Thus, it is seen that the *ss* behaves as a measure of performance quality using the persistence model as a baseline. The closer to 1, the better the forecasting model. The authors demonstrated that the aforementioned ratio of U/V is robust through all time window subsets and represents a statistical invariant for different time horizons. When applied to regional forecasts, it can provide an estimate of the amount of stochastic solar variability that can be reduced. They also concluded that if a model is well trained for an specific climatology, the performance metric should not vary much if the model is applied to other locations with similar climates, reducing thus the influence of climatic variability over performance metrics. Moreover, it was noted that *ss* could be approximated with the RMSE, following (33):

$$ss \approx 1 - \frac{RMSE}{RMSE_p} \qquad (33)$$

8.2. Recently applied metrics

Classical metrics are commonly used by most authors. However, Zhang et al. (2015a) stressed the idea that these classical indicators may not fulfill the requirements of grid operators. For instance, large forecasting errors can have dramatic economic and stability consequences in real operation. Thus, metrics that highlight large errors are necessary. With this aim, they performed a complete assessment on more suitable metrics to evaluate power forecasts, taking into account different forecast horizons, geographic locations, etc. To this purpose, they classified metrics into 4 categories: statistical metrics for different time and geographic scales, uncertainty quantification and propagation metrics, ramp characterization and economic metrics. What follows summarizes these findings:

8.2.1. Statistical metrics

MAE and RMSE are the most common metrics used to evaluate forecasts. Nevertheless, they do not differentiate between two set of values with equal mean and variance but different skewness (which measures the asymmetry in the distribution) and kurtosis values. Thus, they overlook some characteristics of the distributions that could affect system operation. For example, an over-forecasting tendency (positive skewness) could derive in a suboptimal commitment of large thermal units. This would result in the use of more dispatchable (expensive) power plants to correct such deviations. Also, excess kurtosis can give insights about the relative frequency at which extreme events are produced so that system operators can be warned. These two metrics are not stand-alone, they have to be used in combination with other metrics (Zhang et al. 2015a).

$$skew = \frac{N}{(N-1)(N-2)} \sum_{i=1}^{N} \left(\frac{nE - \bar{nE}}{SD} \right)^3 \qquad (34)$$

$$kurt = \left\{ \frac{N(N-1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} \left(\frac{nE - \bar{nE}}{SD} \right)^4 \right\}$$
$$\cdot \frac{3(N-1)^2}{(N-2)(N-3)} \qquad (35)$$

Maximum absolute error (MaxAE) shows the largest forecast error. A big value on this metric could derive in a large economic impact on grid operation.

$$MaxAE = \max_{i=1,2,\dots,N} |P_{pred} - P_{meas}| \qquad (36)$$

Hossain et al. (2013) used the Mean Absolute Scaled Error (MASE), which they claimed to be scale free and have little sensitivity to outliers. The smaller MASEs indicate better forecasts.

$$MASE = \frac{MAE}{(1/N-1)\sum_{i=2}^{N} |P_{meas,i} - P_{meas,i-1}|} \qquad (37)$$

Kolmogorov–Smirnov Integral (KSI). It is a nonparametric test to determine if two data sets are significantly different. A small value of KSI means that the predicted and real values behave statistically similar, which translates into a good forecast. A KSI value of zero indicates that the CDFs of the two data sets are equal (Espinar et al., 2009).

$$KSI = \int_{x_{min}}^{x_{max}} D_n dx \qquad (38)$$

where D_n is the difference between two CDFs.

Similarly, the OVER parameter determines the statistical similarity on large forecast errors between the forecasted and actual power curve.

$$OVER = \int_{p_{min}}^{p_{max}} t dp \qquad (39)$$

where p_{max} and p_{min} and the maximum and minimun values of power generation, respectively, and t is defined by

$$t = \begin{cases} D_j - V_c & if \ D_j > V_c \\ 0 & if \ D_j \leqslant V_c \end{cases} \qquad (40)$$

where V_c is the critical value and D_j, the difference between two CDFs. Both KSI and OVER can be normalized by the term a_c ($a_c = V_c(p_{max} - p_{min})$) to obtain KSIPer and OVERPer, respectively.

As proposed in Beyer et al. (2009) in their work on solar modeling, KSI and OVER can be combined in one parameter, the KSD, to enable a continuous classification of the results (41):

$$KSD = w_1 KSI + w_2 OVER \qquad (41)$$

where w_1 and w_2 are weight parameters. Moreover, they also proposed a new metric (RIO), resulting from the combination of the above-proposed KSD and the RMSE, to provide information from the CDFs and the distance between pairs.

$$RIO = \frac{KSD + RMSE}{2} \qquad (42)$$

8.2.2. Uncertainty quantification

For the uncertainty quantification, Zhang et al. (2015a) proposed the Rényi entropy (43), and the standard deviation of the power forecast errors. As pointed out, classical statistical metrics, such as MAE and RMSE, are only unbiased if they are based on a Gaussian distribution.

Rényi entropy

$$H_\alpha(X) = \frac{1}{1-\alpha} log_2 \sum_{i=1}^{N} p_i^\alpha \qquad (43)$$

where α ($\alpha > 0$ and $\alpha \neq 1$) is the order of the Rényi entropy and p_i is the probability density of the *i*th discrete section of the distribution. Normally, large values of Rényi entropy mean a higher uncertainty in the forecasts.

8.2.3. Characterization of ramps

With respect to metrics to characterize ramps, the swinging door algorithm was proposed (Florita et al., 2013). It is a flexible and simple method, based on a threshold parameter (ϵ) which represents the width of the "door" or ramp. Small values of ϵ identify

many fluctuations, whereas a bigger value would only show the largest changes.

8.2.4. Economic metrics

The way grid operators face solar variability is through reserves. Operating reserves have an associated cost and the greater the penetration of solar energy, the bigger the energy reserves have to be to face possible variations. Thus, an accurate forecast can reduce the amount of operating reserves and, consequently, the operating costs of the system. The authors proposed the 95th percentile of forecast errors as a measure of the operating reserves needed.

The study carried out by Zhang et al. (2015a) concluded that all the metrics proposed were able to capture uniform improvements in solar forecasts. When ramp forecasts were improved or the threshold of the ramps was changed, the metrics of skewness, kurtosis and Rényi entropy were also sensitive to those changes. Based on the sensitivity analysis and on the nonparametric statistical results, they selected a smaller set of metrics to robustly assess the accuracy of forecasts. They recommended the use of the MBE, standard deviation, skewness, kurtosis, distribution of forecast errors, Rényi entropy, RMSE and OVERPer.

8.2.5. Other metrics

Apart from the metrics shown above, some authors have introduced other metrics for error analysis. This way, Russo et al. (2014) proposed this error:

$$E = 100\sqrt{\sum_{i=1}^{N_i} \frac{(P_{meas,i} - P_{pred,i})^2}{\sigma_i^2 N_i}} \tag{44}$$

They claimed that this metric is more correct than using MAE, RMSE or similar metrics because it proportionally correlates the learning error with the actual variability of the output. The interpretation of this error is as follows: for example, an error of 30% means that the modeling error is 0.3 times the standard deviation of the output.

Alessandrini et al. (2015) also proposed a different set of metrics from the most common ones. They used the missing rate error to analyze statistical consistency. It represents the fraction of observations lower/higher that the lowest/highest ranked prediction above or below the expected missing rate of $1/(n+1)$ (n is the number of ensemble members). A large positive missing rate error indicates an under dispersive ensemble. The same authors made use of binned-spread/skill diagrams to assess the uncertainty in probabilistic predictions. In these diagrams, the ensemble spread is compared to the RMSE of the ensemble average over small class intervals of spread. The ensemble is capable to predict its own error if there is good correlation in the diagram. Another metric used was the Continuous Ranked Probability Score (CRPS), which compares the cumulative distribution functions.

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\infty}\left(F_i^{pred}(x) - F_i^{meas}(x)\right)^2 dx \tag{45}$$

where $F_i^{pred}(x)$ is the CDF of the probabilistic forecast and $F_i^{meas}(x)$ is the CDF of the observation for the ith ensemble prediction/observation pair and N is the number of available pairs. A low value of CRPS indicates a good forecast. Zamo et al. (2014b) also used the CRPS and distinguished between the reliability and potential term of the CRPS. The reliability term measures the statistical consistency between the predicted and the observed distribution. The potential term is the CRPS of a perfectly reliable model. It gathers the intrinsic variability of the variable to predict and the resolution of the forecasting system. The resolution mentioned before refers to the ability to output different forecasts for different observations.

Continuing with Alessandrini et al. (2015), they also used the Brier Score (BS) for the probabilistic analysis, which is comparable to the RMSE of a deterministic forecast. It measures the difference between the predicted probability of a distribution and its occurrence.

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_n - o_n)^2 \tag{46}$$

where p is the forecasted probability of a categorical event, o_n, the categorical observation and N, the total number of (p_n, o_n) pairs. A low value of BS shows a good performance of the forecast model. The Brier Skill Score (BSS) was also defined to show improvements with respect to a reference model.

Sperati et al. (2016) further continued using these metrics and also applied the Relative Operating Characteristic (ROC) Skill Score (ROCSS). It is based on the ROC curve, which plots the false alarm rate. ROCSS values close to 1 mean accurate forecasts.

Rana et al. (2015), who worked with interval forecasting, used some other metrics to evaluate their models. They applied the Mean Absolute Interval Deviation (MAID), which measures the deviation of the predicted interval from the actual interval. Low MAID means a forecast with lower error.

$$MAID = \frac{1}{2N}\left\{\sum_{i=1}^{N}\left|U_{meas,k,i}^{P} - U_{pred,k,i}^{P}\right| + \left|L_{meas,k,i}^{P} - L_{pred,k,i}^{P}\right|\right\} \tag{47}$$

where $U_{meas,k,i}^{P}$ and $L_{meas,k,i}^{P}$ are the upper and lower bounds of the actual k-length interval and $U_{pred,k,i}^{P}$ and $L_{pred,k,i}^{P}$ the predicted upper and lower bounds of the interval for i-th example in the dataset. It can be normalized by the range of target values to obtain the percentage error. The Interval Coverage Probability (ICP) was also applied by these authors. This metric measures the probability that the k values of the time series $P_{t+1},\ldots,P_{t+k}$ for the next k-length interval will fall between the upper bound $U_{meas,k,i}^{P}$ and lower bound $L_{meas,k,i}^{P}$ of the predicted interval, averaged over all observations in the data set. The higher the value of ICP, the lower the forecast error.

$$ICP = \frac{1}{N \cdot k}\sum_{i=1}^{N}\sum_{j=i+1}^{i+k}c_j \cdot 100\% \tag{48}$$

where N is the number of samples, k, the length of the interval and c_j:

$$c_j = \begin{cases} 1 & if\ P_j \in [U_{meas,k,i}, L_{meas,k,i}^{P}] \\ 0 & otherwise \end{cases} \tag{49}$$

Almeida et al. (2015) used a modified version of the MAE and MBE, denoted as the coefficient of variation of the MAE and MBE, to assess models in markets which penalize the hourly energy error or the daily energy error, respectively.

$$cvMAE = \frac{MAE}{\bar{P}_{meas}} \tag{50}$$

$$cvMBE = \frac{MBE}{\bar{P}_{meas}} \tag{51}$$

where $\bar{P}_{meas}$ is the mean of the measured power.

Also, they assessed the performance of the confidence interval output by their model regarding its accuracy and its amplitude. The amplitude of the interval was calculated as its area normalized by the area of the observations. It gives an insight about the amount of energy forecasted in relation to the measured energy.

Greater values of this metric means more uncertainty of the quantile $Q.5$.

$$Q1Q9_{sum} = \frac{\sum_{i=1}^{N}(Q_{.9i} - Q_{.1i})}{\sum_{i=1}^{N}P_{meas,i}} \quad (52)$$

The performance statistics for the accuracy were calculated separately for the $Q_{.9}$ and for the $Q_{.1}$. They computed the instants when the observations were higher or lower than the respective quantiles.

Lastly, Hong et al. (in press) discussed several metrics to evaluate the different models that participated in the competition. After analyzing the suitability of MAE, KSI, CRPS and pinball loss function, they opted for the latter due to its adequacy for probabilistic forecasts and ease of implementation and communication. They calculated the pinball loss function (L) with all percentiles from the 1st to the 99th.

$$L(q_a,y) = \begin{cases} (1 - \frac{a}{100})(q_a - y) & if\ y < q_a \\ \frac{a}{100}(y - q_a) & if\ y \geqslant q_a \end{cases} \quad (53)$$

where q_a is a quantile forecast, $a/100$, the target quantile, y, the observation used for forecast evaluation and $a = 1, 2, \ldots, 99$. Then, the score can be averaged over the target quantiles, time periods and forecast horizons. As for the CRPS, the lower the score, the more precise the forecasts are.

8.3. Comparability of studies

Independently of the metrics used to assess the performance of the proposed models, there are some other factors that hamper comparisons among studies.

- **Climatic variability**: High climatic variability normally leads to higher forecast errors than areas with a more stable climate. It is recommended to test a same technique in different locations to know its robustness. Also, the creation of a general database covering different climatic situations and production of several PV plants would be desirable to allow researchers to test their models under the same circumstances and, thus, enable fair benchmarks between techniques.
- **Day/night values and normalization**: To make a fair comparison between studies it is important to state clearly which time frame has been taken into consideration and whether only daylight values, both day and night or only hours in which GHI is larger than threshold have been considered. Grid operators normally demand forecasts for all hours of the day. However, most of the studies compiled here only considered daylight hours. Also, another added difficulty for comparison is normalization of errors. There is no agree on which denominator should be used. It can be performed with respect to the plant peak power, the average power, weighted average or a range of measured values. Hoff et al. (2012) performed an assessment to address the issues raised above for irradiance predictions. Working on RMSE and MAE metrics, they calculated the impact of including night values or not, which in the case of RMSE normalized with the average, resulted in an increase of 41%, whereas in the rest of the cases the addition of night values decreased the error percentage, with reductions up to 50% for the MAE normalized with the capacity. The MAE normalized with the average proved to remain unchanged with the inclusion of night values, which along other subjective appreciations, it was proposed as the best practical metric from those studied. The RMSE normalized by the capacity also ranked high in their classification.
- **Sample aggregation**: The way samples are aggregated also affects results. Averaging samples over larger times leads to

smaller errors (Kaur et al., 2016; Russo et al., 2014). Russo et al. (2014) obtained smaller errors for values averaged to 1 h in comparison to aggregation in 15 min.
- **Spatial aggregation**: Not only time aggregation alters results, but also spatial aggregation. As discussed over the paper, spatial aggregation of plants reduces the ensemble error, so regional results cannot be compared to single site results.
- **Testing period**: Some authors tested over a long period of time covering all sky conditions. However, other authors tested their models on either only sunny days (Mellit and Massi Pavan, 2010) or only cloudy days (Lonij et al., 2013), which also increases difficulty to perform comparisons. Test samples should not have been previously used in the training stage. If the amount of samples is reduced, it is recommended to apply some kind of cross validation to avoid overfitting.
- **Specific plant attributes**: Distribution of errors along the day is different for fixed tilt modules than for dual-axis tracker modules. Tuohy (2015) showed that the MAE and variability of dual-axis modules is higher, especially during mid-morning and mid-afternoon hours.

9. Conclusions

Solar power forecasting becomes a crucial task as solar energy starts to play a key role in electricity markets. The complexity of issuing reliable forecasts is mainly caused by the uncertainty in the solar resource assessment. Moreover, energy markets work within different time frames and, thus, specific forecasts are needed for each time horizon. Several models appear to issue forecasts as accurate as possible.

From the collection of studies shown in this paper, the following trends and conclusions about solar power forecasting can be stated:

- The forecast horizon where most research has been done is the day-ahead. The reason for this behavior is that most of the energy is traded in day-ahead markets, when planning and unit commitment take place. As energy markets evolve, such as the case of EIM, intra-hour trading will become more important and thus, more research will focus on that time horizon and with a higher applicability in electricity markets.
- As the time horizon increases, the proportion of studies that use NWP in their models in comparison to those which use other variables also grows, reaching a 79% for the day-ahead lead time.
- The statistical approach was not only used in most occasions, but it also proved superior when compared to the parametric approach. Most recent papers used machine learning techniques, due to the ease of modeling without the need of knowing PV plant characteristics.
- Spatial averaging reduces variability of the solar resource and generates regional forecasts that are more reliable than single site ones. This is caused by the smoothing effect, which cancels errors with opposite sign in different PV plants. Regional forecasting results are also more useful for grid operators, which must handle the production of several distributed plants, whose impact is treated as an ensemble. However, point forecasts are needed by plant operators to assist them in market bidding.
- Traditionally, most of solar power forecasts were deterministic, that is, for each forecast horizon they provided a single value. Nevertheless, state-of-the art papers are introducing probabilistic forecasts, which enable a better risk assessment and decision making. Compared to load or wind power forecasting, the state of probabilistic solar power forecasting is still inmature and several challenges are yet to be solved.

- The economic impact of solar power forecasting has not been analyzed yet in depth. Recent papers have shown the potential for reducing balancing reserves with improved forecasts, yielding important economic savings and strengthening the grid. Economic assessments of improved forecasts are system specific because they greatly depend on the type of energy generators present in each electricity system. Also, reliable forecasts allow avoiding penalties to plant managers caused by deviations between scheduled and produced.
- Classical statistical metrics have been proved to omit important information in the evaluation of forecasting models. New metrics, not only statistical but also to quantify uncertainty, characterize ramps and evaluate economic impacts have been introduced recently, improving the capacity of model assessment. Guidelines for a proper model evaluation have been stated, not only referring to the metrics used but pointing out other crucial aspects for model comparisons as well.

Conflict of interest

The authors declare that they do not have any conflict of interests.

Acknowledgments

References

Aggarwal, S.K., Saini, L.M., 2014. Solar energy prediction using linear and non-linear regularization models: a study on AMS (American Meteorological Society) 2013–14 solar energy prediction contest. Energy 78, 247–256.

Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog ensemble for short-term probabilistic solar power forecast. Appl. Energy 157, 95–110.

AlHakeem, D., Mandal, P., Haque, A., Yona, A., Senjyu, T., Tseng, T., 2015. A new strategy to quantify uncertainties of wavelet-GRNN-PSO based solar PV power forecasts using bootstrap confidence intervals. In: IEEE Power and Energy Society General Meeting, 26–30 June 2015, Denver.

Almeida, M.P., Perpiñán, O., Narvarte, L., 2015. PV power forecast using a nonparametric PV model. Sol. Energy 115, 354–368.

Al-Messabi, N., Li, Y., El-Amin, I., Goh, C., 2012. Forecasting of photovoltaic power yield using dynamic neural networks. In: Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, 10–15 June 2012, pp. 1–5.

Almonacid, F., Perez-Higueras, P., Fernandez, E., Hontoria, L., 2014. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. Energy Convers. Manage. 85, 389–398.

Antonanzas-Torres, F., Antonanzas, J., Urraca, R., Alia-Martínez, M., Martínez-de-Pisón, F.J., 2016. Impact of atmospheric components on solar clear-sky models at different elevation: case study Canary Islands. Energy Convers. Manage. 109, 122–129.

Arias-Castro, E., Kleissl, J., Lave, M., 2014. A Poisson model for anisotropic solar ramp rate correlations. Sol. Energy 101, 192–202.

Ayompe, L.M., Duffy, A., McCormack, S.J., Conlon, M., 2010. Validated real-time energy models for small-scale grid-connected PV-systems. Energy 35 (10), 4086–4091.

Bacher, P., Madsen, H., Nielsen, H., 2009. Online short-term solar power forecasting. Sol. Energy 83, 1772–1783.

Badescu, V., Gueymard, C.A., Cheval, S., Oprea, C., Baciu, M., Dumitrescu, A., Iacobescu, F., Milos, I., Rada, C., 2013. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. Renew. Energy 55, 85–103.

Berdugo, V., Chaussin, C., Dubus, L., Hebrail, G., Leboucher, V., 2011. Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems. In: Next Generation Data Mining Summit (NGDM '11), Athens, Greece; 4 September 2011.

Bessa, R., Trindade, A., Silva, C., Miranda, V., 2015. Probabilistic solar power forecasting in smart grids using distributed information. Electr. Power Energy Syst. 72, 16–23.

Beyer, H.G., Polo, J., Suri, M., Torres, J.L., Lorenz, E., Müller, S., Hoyer-Klick, C., Ineichen, P., 2009. Management and Exploitation of Solar Resource Knowledge: Report on Benchmarking of Radiation Products.

Bird, L., Cochran, J., Wang, X., 2014. Wind and Solar Energy Curtailment: Experience and Practices in the United States. NREL.

Bosch, J., Kleissl, J., 2013. Cloud motion vectors from a network of ground sensors in a solar power plant. Sol. Energy 95, 13–20.

Botterud, A., Zhou, Z., Wang, J., Bessa, R., Keko, H., Sumaili, J., Miranda, V., 2012. Wind power trading under uncertainty in LMP markets. IEEE Trans. Power Syst. 27, 2.

Bouzerdoum, M., Mellit, A., Massi Pavan, A., 2013. A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Sol. Energy 98, 226–235.

Brabec, M., Pelikán, E., Krc, P., Eben, K., Malý, M., Juruš, P., 2011. A coupled model for energy production forecasting from photovoltaic farms. In: ES1002: Workshop March 22nd–23rd.

Bracale, A., Caramia, P., Carpinelli, G., Di Fazio, A., Ferruzzi, G., 2013. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and Control. Energies 6, 733–747.

Brancucci Martinez-Anido, C.B., Botor, B., Florita, A., Draxl, C., Lu, S., Hamann, H., Hodge, B.M., 2016. The value of day-ahead solar power forecasting improvement. Sol. Energy 129, 192–203.

Brancucci Martínez-Anido, C.B., Florita, A., Hodge, B.M., 2014. The Impact of Improved Solar Forecasts on Bulk Power System Operations in ISO-NE. NREL/CP-5D00-62817. National Renewable Energy Laboratory, Golden, CO.

Breiman, L., 2001. Random forests. Machine Learn. 45 (1), 5–32.

Chen, C., Duan, S., Cai, T., Liu, B., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol. Energy 85, 2856–2870.

Chow, C., Belongie, S., Kleissl, J., 2015. Cloud motion and stability estimation for intra-hour solar forecasting. Sol. Energy 115, 645–655.

Chow, S., Lee, E., Li, D., 2012. Short-term prediction of photovoltaic energy generation by intelligent approach. Energy Build. 55, 660–667.

Chu, Y., Urguhart, B., Gohari, S., Pedro, H., Kleissl, J., Coimbra, C., 2015. Short-term reforecasting of power output from a 48 MWe solar PV plant. Sol. Energy 112, 68–77.

Chupong, C., Plangklang, B., 2011. Forecasting power output of PV grid connected system in Thailand without using solar radiation measurement. Energy Proc. 9, 230–237.

Coimbra, C., Pedro, H., 2013. Stochastic-learning methods. In: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment, first ed. Academic Press, Waltham, pp. 383–406.

Cormode, D., Lorenzo, A., Holmgren, W., Chen, S., Cronin, A., 2014. The economic value of forecasts for optimal curtailment strategies to comply with ramp rate rules. In: IEEE 40th Photovoltaic Specialist Conference (PVSC). 8–13 June, Denver, Co.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learn. 20 (3), 273.

De Felice, M., Petitta, M., Ruti, P., 2015. Short-term predictability of photovoltaic production over Italy. Renew. Energy 80, 197–204.

De Georgi, M.G., Congedo, P.M., Malvoni, M., 2014. Photovoltaic power forecasting using statistical methods: impact of weather data. IET Sci. Measure. Technol. 8 (3), 90–97.

De Georgi, M.G., Congedo, P.M., Malvoni, M., Laforgia, D., 2015. Error analysis of hybrid photovoltaic power forecasting models: a case study of mediterranean climate. Energy Convers. Manage. 100, 117–130.

De Georgi, M.G., Malvoni, M., Congedo, P.M., 2016. Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine. Energy 107, 360–373.

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

Ding, M., Wang, L., Bi, R., 2011. An ANN-based approach for forecasting the power output of photovoltaic system. Proc. Environ. Sci. 11 (Part C), 1308–1315.

Do, M.T., Soubdhan, T., Robyns, B., 2016. A study on the minimum duration of training data to provide a high accuracy forecast for PV generation between two different climatic zones. Renew. Energy 85, 959–964.

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

Dolara, A., Leva, S., Manzolini, G., 2015b. Comparison of different physical models for PV power output prediction. Sol. Energy 119, 83–99.

Dudhia, J., Gill, D., Manning, K., Wang, W., Bruyere, C., 2005. PSU/NCAR Mesoscale Modeling System. Tutorial Class Notes and User's Guide: MM5 Modeling System Version 3. National Center for Atmospheric Research, Camp Springs, Maryland.

Ela, E., Milligan, M., Kirby, B., 2011. Operating Reserves and Variable Generation. NREL/TP-5500-51978. National Renewable Energy Laboratory, Golden, CO.

EPIA, September 2012. Connecting the Sun. Solar photovoltaic on the road to large-scale grid integration.

Espinar, B., Ramírez, L., Drews, A., Beyer, H.G., Zarzalejo, L., Polo, J., Martín, L., 2009. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Sol. Energy 83, 118–125.

Fernandez-Jimenez, L., Muñoz-Jimenez, A., Falces, A., Mendoza-Villena, M., Garcia-Garrido, E., Lara-Santillan, P., Zorzano-Alba, E., Zorzano-Santamaria, P., 2012. Short-term power forecasting system for photovoltaic plants. Renew. Energy 44, 311–317.

Florita, A., Hodge, B.-M., Orwig, K., 2013. Identifying wind and solar ramping events. In: IEEE 5th Green Technologies Conf., Denver, Colorado.

Fonseca, J.G., Oozeki, T., Takashima, T., Koshimizu, G., Uchida, Y., Ogimoto, K., 2011a. Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan. Progress in Photovoltaics: Research and Applications.

Fonseca, J.G., Oozeki, T., Takashima, T., Koshimizu, G., Uchida, Y., Ogimoto, K., 2011b. Photovoltaic power production forecasts with support vector regression: a study on the forecast horizon. In: Photovoltaic Specialists Conference (PVSC), 2011 37th IEEE.

Fonseca, J.G., Oozeki, T., Ohtake, H., Shimose, K., Takashima, T., Ogimoto, K., 2014a. Characterizing the regional photovoltaic power forecast error in Japan: a study of 5 regions. IEEJ Trans. Power Energy 134 (6), 537–544.

Fonseca, J.G., Oozeki, T., Ohtake, H., Shimose, K., Takashima, T., Ogimoto, K., 2014b. Forecasting regional photovoltaic power generation – a comparison of strategies to obtain one-day-ahead data. Energy Proc. 57, 1337–1345.

Fonseca, J.G., Oozeki, T., Ohtake, H., Shimose, K., Takashima, T., Ogimoto, K., 2014c. Regional forecasts and smoothing effect of photovoltaic power generation in Japan: an approach with principal component analysis. Renew. Energy 68, 403–413.

Fonseca, J.G., Oozeki, T., Ohtake, H., Takashima, T., Ogimoto, K., 2014d. Regional forecasts of photovoltaic power generation according to different data availability scenarios: a study of four methods. Progress in Photovoltaics: Research and Applications.

Fonseca, J.G., Oozeki, T., Ohtake, H., Takashima, T., Ogimoto, K., 2015. On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation. J. Electr. Eng. Technol. 10 (3), 1342–1348.

Gandelli, A., Grimaccia, F., Leva, S., Mussetta, M., Ogliari, E., 2014. Hybrid model analysis and validation for PV energy production forecasting. In: International Joint Conference on Neural Networks (IJCNN), July 6–11, Beijing, China.

Golestaneh, F., Pinson, P., Gooi, H.B., 2016. Very short-term nonparametric probabilistic forecasting of renewable energy generation–with application to solar energy. IEEE Trans. Power Syst. (99) http://dx.doi.org/10.1109/TPWRS.2015.2502423.

Graditi, G., Ferlito, S., Adinolfi, G., 2016. Comparison of photovoltaic plant power production prediction methods using a large measured dataset. Renew. Energy 90, 513–519.

GTM Reseach/SEIA U.S. Solar Market Insight, Report Q2, Executive summary, 2015.

Gueymard, C.A., 2008. REST2: high-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – validation with a benchmark dataset. Sol. Energy 82, 272–285.

Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol. Energy 128, 1–30.

Haque, A., Nehrir, M., Mandal, P., 2013. Solar PV power generation forecast using a hybrid intelligent approach. In: Power and Energy Society General Meeting (PES). IEEE, pp. 1–5.

Hassanzadeh, M., Etezadi-Amoli, M., Fadali, M., 2010. Practical approach for sub-hourly and hourly prediction of PV power output. North American Power Symposium (NAPS).

Hodge, B.-M., Florita, A., Sharp, J., Margulis, M., Mcreavy, D., 2015. The Value of Improved Short-Term Wind Power Forecasting. NREL/ TP-5D00-63175. National Renewable Energy Laboratory, Golden, CO.

Hoff, T., Perez, R., Kleissl, J., Renne, D., Stein, J., 2012. Reporting of irradiance modeling relative prediction errors. Prog. Photovolt: Res. Appl. http://dx.doi.org/10.1002/pip.2225.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R.J., in press. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. International Journal of Forecasting. Working paper available from <http://www.drhongtao.com/articles>.

Hossain, M.R., Oo, A.M.T., Ali, A.B., 2013. Hybrid prediction method for solar power using different computational intelligence algorithms. Smart Grid Renew. Energy 4, 76–87.

Huang, J., Perry, M., 2015. A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting. Int. J. Forecast. http://dx.doi.org/10.1016/j.ijforecast.2015.11.002.

IEA-Photovoltaic and solar forecasting: state-of-the-art. Report IEA-PVPS T14-01:2013.

IEA-The power of transformation, Paris: IEA, 2014.

IEA-Trends 2015 in photovoltaic applications, Executive summary. Report IEA-PVPS T1-27:2015.

Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. Sol. Energy 80 (4), 468–478.

Ineichen, P., Perez, R., 2002. A new airmass independent formulation for the Linke turbidity coefficient. Sol. Energy 73 (3), 151–157.

Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. Prog. Energy Combust. Sci. 39, 535–576.

Jacobs, D., 2012. Renewable Energy Policy Convergence in the EU: The Evolution of Feed-In Tariffs in Germany, Spain and France. Ashgate, Farnham, ISBN: 9781409439097.

Jafarzadeh, S., Fadali, M.S., Evrenosoglu, C.Y., 2013. Solar power prediction using interval type-2 TSK modeling. IEEE Trans. Sustain. Energy 4 (2).

Kaur, A., Nonnenmacher, L., Pedro, H.T.C., Coimbra, C.F.M., 2016. Benefits of solar forecasting for energy imbalance markets. Renew. Energy 86, 819–830.

Kudo, M., Takeuchi, A., Nozaki, Y., Endo, H., Sumita, J., 2009. Forecasting electric power generation in a photovoltaic power system for an energy network. Electr. Eng. Jpn. 167 (4).

Kühnert, J., Lorenz, E., Heinemann, D., 2013. Satellite-based irradiance and power forecasting for the German energy market. In: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment, first ed. Academic Press, Waltham, pp. 267–297.

Larson, V.E., 2013. Forecasting Solar Irradiance with Numerical Weather Prediction Models. Solar Energy Forecasting and Resource Assessment. Elsevier.

Larson, D., Nonnenmacher, L., Coimbra, C.F.M., 2016. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renew. Energy 91, 11–20.

Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., Ogliari, E., 2015. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Math. Comput. Simul. http://dxdoi.org/10.1016/j.matcom.2015.05.010.

Li, Y., He, L., Nie, R., 2009. Short-term forecast of power generation for grid-connected photovoltaic system based on advanced grey-Markov chain. In: International Conference on Energy and Environment Technology.

Li, Z., Rahman, S.M., Vega, R., Dong, B., 2016. A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. Energies 9, 55.

Li, Y., Su, Y., Shu, L., 2014. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. Renew. Energy 66, 78–89.

Lin, K.P., Pai, P.F., 2015. Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. J. Cleaner Product. http://dxdoi.org/10.1016/j.jclepro.2015.08.099.

Lipperheide, M., Bosch, J., Kleissl, J., 2015. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. Sol. Energy 112, 232–238.

Long, H., Zhang, Z., Su, Y., 2014. Analysis of daily solar power prediction with data-driven approaches. Appl. Energy 126, 29–37.

Lonij, V., Brooks, A., Cronin, A., Leuthold, M., Koch, K., 2013. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. Sol. Energy 97, 58–66.

Lorenz, E., Heinemann, D., Wickramarathne, H., Beyer, H.G., Bofinger, S., 2007. Forecast of ensemble power production by grid-connected PV systems. In: 20th European PV Conference, Milano, Italy, 03.09.–07.09.2007.

Lorenz, E., Heinemann, D., Kurz, C., 2011a. Local and regional photovoltaic power prediction for large scale grid integration: assessment of a new algorithm for snow detection. Prog. Photovolt.: Res. Appl. 20 (6), 760–769.

Lorenz, E., Hurka, J., Karampela, G., Heinemann, D., Beyer, H.G., Schneider, M., 2008. Qualified forecast of ensemble power production by spatially dispersed grid-connected PV systems. In: 23rd European Photovoltaic Solar Energy Conference, Valencia, Spain, 1–5 September.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009. Irradiance forecasting for the power prediction on grid-connected photovoltaic systems. IEEE J. Select. Top. Appl. Earth Observ. Remote Sensing 2, 1.

Lorenz, E., Kühnert, J., Wolff, B., Hammer, A., Kramer, O., Heinemann, D., 2014. PV power predictions on different spatial and temporal scales integrating PV measurements, satellite data and numerical weather predictions. In: 29th EUPVSEC, 22.–26. September 2014, Amsterdam, Netherlands.

Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., Kurz, C., 2011b. Regional PV power prediction for improved grid integration. Prog. Photovolt.: Res. Appl. 19, 757–771.

Lu, S., Hwang, Y., Khabibrakhmanov, I., Marianno, F., Show, X., Zhang, J., Hodge, B.M., Hamann, H., 2015. Machine learning based multi-physical-model blending for enhancing renewable energy forecast – improvement via situation dependent error correction. In: European Control Conference (ECC), Linz, 15–17 July 2015.

Mandal, P., Madhira, S., Ulhaque, A., Meng, J., Pineda, R., 2012. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Proc. Comput. Sci. 12, 332–337.

Marquez, R., Coimbra, C.F.M., 2013. Proposed metric for evaluation of solar forecasting models. ASME J. Sol. Energy Eng. 135, 0110161-9.

Masa-Bote, D., Castillo-Cagigal, M., Matallanas, E., Caamaño-Martín, E., Gutiérrez, A., Monasterio-Huelín, F., Jiménez-Leube, J., 2014. Improving photovoltaics grid integration through short time forecasting and self-consumption. Appl. Energy 125, 103–113.

Mathiesen, P., Kleissl, J., Collier, C., 2013. Case studies of solar forecasting with the weather research and forecasting model at GL-Garrad Hassan. In: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment, first ed. Academic Press, Waltham, pp. 357–381.

Mellit, A., Massi Pavan, A., 2010. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol. Energy 84, 807–821.

Mellit, A., Massi Pavan, A., Lughi, V., 2014. Short-term forecasting of power production in a large-scale photovoltaic plant. Sol. Energy 105, 401–413.

Mills, A., Botterud, A., Wu, J., Zhou, Z., Hodge, B.M., Heaney, M., 2013. Integrating Solar PV in Utility System Operations. ANL/DIS-13/18. Argonne National Laboratory.

Mills, A., Wiser, R., 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. LBNL-3884E. Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division, Berkeley, CA.

Mohammed, A.A., Yaqub, W., Aung, Z., 2015. Probabilistic forecasting of solar power: an ensemble learning approach. Intell. Decis. Technol. – Smart Innovat. Syst. Technol. 39, 449–458.

Monteiro, C., Fernandez-Jimenez, L.A., Ramirez-Rosado, I., Muñoz-Jimenez, A., Lara-Santillan, P., 2013a. Short-term forecasting models for photovoltaic plants: analytical versus soft-computing techniques. Math. Problems Eng.

Monteiro, C., Santos, T., Fernandez-Jimenez, L., Ramirez-Rosado, I., Terreros-Olarte, M., 2013b. Short-term power forecasting model for photovoltaic plants based on historical similarity. Energies 6, 2624–2643.

Mora-Lopez, L., Martinez-Marchena, I., Ppiliougine, M., Sidrach-deCardona, M., 2011. Machine learning approach for next day energy production forecasting in grid connected photovoltaic plants. In: World Renewable Energy Congress, Linköping, Sweden 8–13 May 2011.

Mueller, R., Dagestad, K., Ineichen, P., Schroedter-Homscheidt, M., Cros, S., Dumortier, D., 2004. Rethinking satellite-based solar irradiance modeling: the SOLIS clear-sky module. Remote Sensing Environ. 91 (2), 160–174.

Nagy, G., Barta, G., Kazi, S., Borbély, G., Simon, G., 2016. GEFCom 2014: probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. Int. J. Forecast. http://dxdoi.org/10.1016/j.ijforecast.2015.11.013.

National Oceanic and Atmospheric Administration, 2003. The GFS atmospheric model. Available on-line: <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on442.pdf>.

Ogliari, E., Grimaccia, F., Leva, S., Mussetta, M., 2013. Hybrid predictive models for accurate forecasting in PV systems. Energies 6, 1918–1929.

Oudjana, S., Hellal, A., Hadj-Mahammed, I., 2013. Power forecasting of photovoltaic generation. Int. J. Electr. Comput. Energet. Electron. Commun. Eng. 7, 6.

Pedro, H.T.C., Coimbra, C.F.M., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy 86, 2017–2028.

Pelland, S., Galanis, G., Kallos, G., 2011. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. Prog. Photovolt.: Res. Appl.

Peng, Z., Yu, D., Huang, D., Heiser, J., Yoo, S., Kalb, P., 2015. 3D cloud detection and tracking system for solar forecast using multiple sky imagers. Sol. Energy 118, 496–519.

Perez, R., Hoff, T., 2013. Solar resource variability. In: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment, first ed. Academic Press, Waltham, pp. 133–148.

Perez, R., Hoff, T., Kivalow, S., 2011. Spatial and temporal characteristics of solar radiation variability. In: Proc,. of International Solar Energy (ISES) World Congress, Kassel, Germany.

PVCROPS, October 2015. Estimation of the PV power that can be integrated in current EU networks. Deliverable D5.5.

PV GRID, Final Project Report, August 2014.

Ramsami, P., Oree, V., 2015. A hybrid method for forecasting the energy output of photovoltaic systems. Energy Convers. Manage. 95, 406–413.

Rana, M., Koprinska, I., Agelidis, V., 2015. 2D-interval forecasts for solar power production. Sol. Energy 122, 191–203.

Rana, M., Koprinska, I., Agelidis, V., 2016. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers. Manage. 121, 380–390.

Ren, Y., Suganthan, P., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting – a state-of-the-art review. Renew. Sustain. Energy Rev. 50, 82–91.

Rigollier, C., Bauer, O., Wald, L., 2000. On the clear sky model of the ESRA e European Solar Radiation Atlas e with respect to the Heliosat method. Sol. Energy 68 (1), 33–48.

Russo, M., 2012. Towards general purpose neuro-genetic programming socket based formal modeller. In: Lee, G., Howard, D., Kang, J.J., Slezak, D. (Eds.), ICHIT (1). In: Lecture Notes in Computer Science, vol. 7425. Springer, pp. 195–202.

Russo, M., Leotta, G., Pugliatti, P., Gigliucci, G., 2014. Genetic programming for photovoltaic plant output forecasting. Sol. Energy 105, 264–273.

Schmelas, M., Feldmann, T., Fernandes, J., Bollin, E., 2015. Photovoltaics energy prediction under complex conditions for a predictive energy management system. J. Sol. Energy Eng. 137.

Shi, J., Lee, W.J., Liu, Y., Yang, Y., Wang, P., 2011. Forecasting power output of photovoltaic system based on weather classification and support vector machine. In: Industry Applications Society Annual Meeting (IAS). IEEE.

Simonov, M., Mussetta, M., Grimaccia, F., Leva, S., Zich, R., 2012. Artificial intelligence forecast of PV plant production for integration in smart energy systems. Int. Rev. Electr. Eng. 7, 1.

Soubdhan, T., Ndong, J., Ould-Baba, H., Do, M.T., 2016. A robust forecasting framework based on the Kalman filtering approach with a twofold parameter tuning procedure: application to solar and photovoltaic prediction. Sol. Energy 131, 246–259.

Sperati, S., Alessandrini, S., Delle Monache, L., 2016. An application of the ECMWF ensemble prediction system for short-term solar power forecasting. Sol. Energy 133, 437–450.

Stein, J.S., 2012. The photovoltaic performance modeling collaborative (PVPMC). In: 38th IEEE Photovoltaic Specialists Conference (PVSC), Austin, Texas.

Takahashi, M., Mori, H., 2013. A hybrid intelligent system approach to forecasting of PV generation output. J. Int. Council Electr. Eng. 3 (4), 295–299.

Tao, C., Shanxu, D., Changsong, C., 2010. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. In: 2nd IEEE International Symposium on Power Electronics for Distributed Generation Systems.

Tuohy, A. et al., 2015. Solar forecasting: methods, challenges, and performance. Power Energy Mag., IEEE 13 (6), 50–59.

Urquhart, B., Ghonima, M., Nguyen, D., Kurtz, B., Chow, C.W., Kleissl, J., 2013. Sky-imaging systems for short-term forecasting. In: Kleissl, J. (Ed.), Solar Energy Forecasting and Resource Assessment, first ed. Academic Press, pp. 195–232.

Urquhart, B., Kurtz, B., Dahlin, E., Ghonima, M., Shields, J., Kleissl, J., 2015. Development of a sky imaging system for short-term solar power forecasting. Atmos. Measure. Tech. 8, 875–890.

Urraca, R., Antonanzas, J., Alia-Martínez, M., Martínez-de-Pisón, F.J., Antonanzas-Torres, F., 2016. Smart baseline models for solar irradiation forecasting. Energy Convers. Manage. 108, 539–548.

Vapnik, V., Lerner, A., 1963. Pattern recognition using generalized portrait method. Autom. Remote Control 24, 774–780.

Vaz, A., Elsinga, B., van Sark, W., Brito, M., 2016. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands. Renew. Energy 85, 631–641.

Verzijlbergh, R., Heijnen, P., de Roode, S., Los, A., Jonker, H., 2015. Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications. Sol. Energy 118, 634–645.

Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., Hu, Z., 2015. Photovoltaic and solar power forecasting for smart grid energy management. CSEE J. Power Energy Syst. 1 (4).

Wu, Y.K., Chen, C.R., Rahman, H.A., 2014. A novel hybrid model for short-term forecasting in PV power generation. Int. J. Photoenergy 2014, 9 569249.

Yang, H.T., Huang, C.M., Huang, Y.C., Pai, Y.S., 2014. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. IEEE Trans. Sustain. Energy 5, 3.

Yang, C., Xie, L., 2012. A novel ARX-based multi-scale spatiotemporal solar power forecast model. In: Proceedings of the North American Power Symposium (NAPS), USA, September.

Yona, T., Senjyu, T., Funabashi, T., 2007. Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system. In: Power Engineering Society General Meeting, 24–28 June 2007, Tampa, FL.

Yona, A., Senjyu, T., Funabashi, T., Kim, C.H., 2013. Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction. IEEE Trans. Sustain. Energy 4, 2.

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014a. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part I: Deterministic forecast of hourly production. Sol. Energy 105, 792–803.

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014b. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. Sol. Energy 105, 804–816.

Zeng, J., Qiao, W., 2013. Short-term solar power prediction using a support vector machine. Renew. Energy 52, 118–127.

Zhang, J., Florita, A., Hodge, B.M., Lu, S., Hamann, H., Banunarayanan, V., Brockway, A., 2015a. A suite of metrics for assessing the performance of solar power forecasting. Sol. Energy 111, 157–175.

Zhang, J., Hodge, B.M., Lu, S., Hamann, H., Lehman, B., Simmons, J., Campos, E., Banunarayanan, V., Black, J., Tedesco, J., 2015b. Baseline and target values for regional and point PV power forecasts: toward improved solar forecasting. Sol. Energy 122, 804–819.