Galván, I., Huertas, J., Rodríguez, F., Arbizu, C., Pozo, D., Aler, R. (2021). Evolutionary-based prediction interval estimation by blending solar radiation forecasting models using meteorological weather types. *Applied Soft Computing*, 109, 107531.

DOI: 10.1016/j.asoc.2021.107531

Applied Soft Computing 109 (2021) 107531

Contents lists available at ScienceDirect

Applied Soft Computing

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

Evolutionary-based prediction interval estimation by blending solar radiation forecasting models using meteorological weather types

Inés M. Galván [a], Javier Huertas-Tato [c], Francisco J. Rodríguez-Benítez [b], Clara Arbizu-Barrena [b], David Pozo-Vázquez [b], Ricardo Aler [a,*]

[a] EVANNAI Res. Group, Computer Science Department, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain
[b] MATRAS Res. Group, Andalusian Institute for Earth System Research (IISTA), Department of Physics, University of Jaén, 23071, Jaén, Spain
[c] AIDA Res. Group, ETSISI, Universidad Politécnica de Madrid, Calle Alan Turing s/n, 28031 Madrid, Spain

ARTICLE INFO

Article history:
Received 19 February 2020
Received in revised form 12 February 2021
Accepted 11 May 2021
Available online 26 May 2021

Keywords:
Prediction intervals
Solar forecasting
Blending approaches
Multi-objective optimization

ABSTRACT

Recent research has shown that the integration or blending of different forecasting models is able to improve the predictions of solar radiation. However, most works perform model blending to improve point forecasts, but the integration of forecasting models to improve probabilistic forecasting has not received much attention. In this work the estimation of prediction intervals for the integration of four Global Horizontal Irradiance (GHI) forecasting models (Smart Persistence, WRF-solar, CIADcast, and Satellite) is addressed. Several short-term forecasting horizons, up to one hour ahead, have been analyzed. Within this context, one of the aims of the article is to study whether knowledge about the synoptic weather conditions, which are related to the stability of weather, might help to reduce the uncertainty represented by prediction intervals. In order to deal with this issue, information about which weather type is present at the time of prediction, has been used by the blending model. Four weather types have been considered. A multi-objective variant of the Lower Upper Bound Estimation approach has been used in this work for prediction interval estimation and compared with two baseline methods: Quantile Regression (QR) and Gradient Boosting (GBR). An exhaustive experimental validation has been carried out, using data registered at Seville in the Southern Iberian Peninsula. Results show that, in general, using weather type information reduces uncertainty of prediction intervals, according to all performance metrics used. More specifically, and with respect to one of the metrics (the ratio between interval coverage and width), for high-coverage (0.90, 0.95) prediction intervals, using weather type enhances the ratio of the multi-objective approach by 2%–3%. Also, comparing the multi-objective approach versus the two baselines for high-coverage intervals, the improvement is 11%–17% over QR and 10%–44% over GBR. Improvements for low-coverage intervals (0.85) are smaller.

1. Introduction

Solar energy capacity has shown a large increase in the last years. However, its intermittency and dependence on the weather makes its integration in the electricity grid more challenging than operable sources. Those difficulties can be tackled to some extent by computing solar radiation forecasts as accurately as possible. Depending on the application, particular forecasting horizons may be preferable. For instance, solar radiation short-term forecasting addresses horizons up to six hours, which can be useful for the management of concentrating solar power plants [1], to participate in the intra-day electricity markets or forecasting ramp events [2]. Operating power systems and participating in electricity market with high penetration of solar energy demands a continuous improvement of solar power forecasting [3–5]. Many kinds of models have been proposed for forecasting solar radiation based on time series of previous radiation measurements. Some of them are statistical models [6], while others are based in machine learning techniques. In [7,8] extensive overviews of the use of these techniques can be found. More concretely, in [9], Support Vector Machines are used to include simple but very effective models, such as Smart Persistence, that extrapolates current solar radiation into the future, based on the solar daily cycle. Another kind of widely used models are Numerical Weather Prediction (NWP), which are based on mathematical models of the atmosphere. In [10] the focus is on the analysis and correction of GHI forecasts from three operational NWP models

* Corresponding author.
E-mail addresses: igalvan@inf.uc3m.es (I.M. Galván), javier.huertas.tato@upm.es (J. Huertas-Tato), fbenitez@ujaen.es (F.J. Rodríguez-Benítez), carbizu@ujaen.es (C. Arbizu-Barrena), dpozo@ujaen.es (D. Pozo-Vázquez), aler@inf.uc3m.es (R. Aler).

https://doi.org/10.1016/j.asoc.2021.107531

Nomenclature

AIW	Average Interval Width
$(AIW_x, PICP_x)$	Value of AIW and $PICP$ for a solution x from the Pareto Front
$B^{(1)}$	Bias for hidden neurons
$B^{(2)}$	Bias for the outputs neurons
C_t^h	The Global Horizontal Irradiance forecasting by CIAD-Cast Model at time t for horizon h
d_i	The desired output for sample i
D^+	Set of solutions that fulfill the target $PINC$
D^-	Set of solutions that do not fulfill the target $PINC$
f_{act}	The sigmoid activation function
GHI_{cs}	The clear sky Global Horizontal Irradiance
$GHI(t)$	The Global Horizontal Irradiance at time t
h	Forecasting horizon
I	Function that, for every instance M_i returns the lower and upper bounds of an interval with a given nominal coverage (PINC)
I^h	Function I for horizon h
I_{ps}	The MLP associated to the ps point in the Pareto front, whose output is a prediction interval I
M_i	Vector of input variables for instance i
n	Number of observations
P_t^h	The Global Horizontal Irradiance forecasting by Smart Persistence Model at time t for horizon h
p_i^{Low}	Lower bound of the PI for sample i
p_i^{Upp}	Upper bound of the PI for sample i
$PICP$	Prediction Interval Coverage Probability
$PINC$	Prediction Interval Nominal Coverage
ps	A point (AIW, PICP) of the Pareto front
q_1 and q_2	$\frac{1-PINC}{2}$ and $\frac{1+PINC}{2}$ quantiles, respectively
$Ratio$	Ratio between $PICP$ and AIW
S_t^h	The Global Horizontal Irradiance forecasting by Satellite Model at time t for horizon h
CWC	Coverage width-based criterion
$(T_t^{h(0)}, T_t^{h(1)}, T_t^{h(2)}, T_t^{h(3)})$	Binary variables that identify which of the four weather-types () is forecast at time t for horizon h. Only one of them can be set to 1 (while the remaining ones are set to 0)
$W^{(1)}$	Weights of the MLP from the input to the hidden layer
$W^{(2)}$	Weights of the MLP from the hidden to the output layer
W_t^h	The Global Horizontal Irradiance forecasting by WRF-Solar Model at time t for horizon h
$X^{(1)}$	Activation of the neurons for the hidden layer
$Y(p^{low}, p^{upp})$	Outputs of the MLP

within the continental United States with the aim of improving forecasting. In other works, the NWP models are used as inputs to machine learning models [11] or combined with other kind of models by means of machine learning to produce better forecasts [12]. Satellite based images can also be used to forecast radiation, by estimating cloud motion vectors from consecutive cloud index maps, and using them to forecast future cloudiness and solar radiation [6,13]. Different models have shown to be more appropriate for some prediction horizons. For instance, Smart Persistence is difficult to beat for short horizons, while NWP models tend to perform better for longer horizons (the break-even point may depend on the location) [14]. While single models offer reasonably accurate forecasts [15], previous research has shown that the integration (or blending) of several of those models can produce predictions better than any of the base models. It is the case of the work previously mentioned [12], where irradiance measurements, satellite and NWP are combined; or the studies presented in [16,17], where different forecasting models are combined to improve global and direct irradiance.

The works mentioned above deal with point forecasts, but in the case of solar energy, it is also important to estimate the uncertainty of the forecasts. Probabilistic solar forecasting has attracted more attention in the last years [18]. These types of forecast provide more information to users than deterministic forecasts, specially for users that can use this information to manage the risk and improve the system efficiency. There are many examples of how probabilistic renewable energies forecasts (including solar) provide more value than deterministic ones in the management of power systems [19,20]. Several works have also shown that probabilistic wind power forecasts are also useful for the participation in the electricity market. For instance, [21] concentrates on a particular test case (the Horns Rev wind farm) over a period of 1 year, in order to study the performance of an ensemble-based probabilistic forecasting methodology. In [22] it is studied the economic impact of several forecasting systems for wind energy electricity producers. Another interesting study shows how probabilistic forecasting can be used to improve the bidding strategy of ancillary services of renewable power plants [23]. In [24], uncertainties in load demand and renewable energy sources are taken into account, in order to perform optimal reactive power dispatch.

The aim of the present work is to estimate uncertainty for a blending approach that integrates different forecasting sources. A way of representing uncertainty are prediction intervals (PI), which offer a natural way to bound uncertainty by specifying a lower and an upper limit, that cover the quantity of interest with a given probability (set by user requirements). Several ways of estimating PI's can be found in the literature but a recent evolutionary-computation approach, called LUBE (Lower Upper Bound Estimation) has displayed competitive performance over standard approaches [25]. The LUBE approach uses a Multilayer Perceptron (MLP) with two outputs, the lower and upper bounds, which can be trained using single-objective evolutionary computation techniques [26] or multi-objective evolutionary methods [27].

In this work, in order to estimate PI's in the blending of models context, the MLP LUBE model is used, where the inputs are themselves models specialized on forecasting solar radiation, and the outputs are the lower and upper bounds of the PI. Four models will be used as inputs: Smart Persistence, WRF-Solar [28], CIAD-Cast [29], and Satellite [14], which are known to be complementary for different prediction horizons. The LUBE multi-objective approach described in [27] will be used for this purpose. The multi-objective approach has been selected in preference to the single-objective one because it allows to obtain a set of solutions with different coverage trade-offs in a single run.

However, there are some differences with the basic method used in [27]. In particular with the selection of the solution from the set of solutions, so that intervals with better coverage are obtained. This will be explained in Section 2.

The synoptic meteorological situation has a clear impact on the cloudiness conditions and, therefore, on the availability of the solar resources [14,30]. In middle-latitudes, changes in synoptic weather conditions are related to the occurrence, position, and strength of low and high pressure centers. In a recent work, [31] presented a regime-switching short term wind power forecasting method based on the analysis of meteorological conditions. The author firstly identified three atmospheric modes, based on the reanalysis data, relevant for wind power forecasting. Then, the modes were used to improve wind forecasts. The method was tested in the United Kingdom, showing improved skills. In [32,33] a methodology is proposed to improve day-ahead solar forecast by developing meteorological situation-dependent corrections for different forecasting models.

Within this context, a second goal of this article is to explore whether adding information about synoptic weather conditions to the blending model (in addition to the aforementioned four input models) is able to improve the quality of the PI's. So far, to our knowledge, the analysis of synoptic conditions has not been used to improve short-term solar radiation forecasting, either deterministic or probabilistic. To this end, the four weather types (WT) identified for the study area by [30] will be considered in this work. The four WT's account for: (1) overcast conditions (low pressure systems over the study area); (2) clear sky conditions (high pressures); (3) broken clouds conditions (transient perturbations) and (4) local clouds (moderate high pressures systems). While knowing the WT is relevant for point solar forecasting, the aim of this work is to study, specifically, its influence on PI's, because knowing in advance the expected sky condition may lead to more appropriate intervals (i.e. broken clouds sky conditions may increase uncertainty, which in turn requires larger intervals, while clear sky situations should result in narrower intervals). In order to use WT for estimating PI's, the multi-objective LUBE approach is also used, but adding WT as an additional input to the MLP blending model.

The present study has been carried out to estimate PI's for Global Horizontal Irradiance (GHI) at four forecasting horizons (15, 30, 45 and 60 min) at a location in the Iberian peninsula (Seville), using measurements from March 2015 to February 2017. For comparison purposes two methods have been used. The first is Quantile Regression (QR) [34], a common linear method widely used in the context of probabilistic forecasting [35]. The second is Gradient Boosting Regression (GBR) [36], a non-linear method which is able to estimate quantiles, out of which PI's can be computed. In this work, QR and GBR will be used in the same conditions than MOPSO (Multi-objective Particle Swarm Optimization), that is, using the four models (Smart Persistence, WRF-Solar, CIAD-Cast, and Satellite) as inputs and also the WT.

The rest of the paper is organized as follows. The multi-objective LUBE approach is summarized in Section 2. Section 3 includes a brief description of the four forecasting models and a description of the WT's used in this work. Section 4 explains how the approach has been used to estimate PI's from the blending of the four aforementioned models, and how weather types are included in the MLP blending model. In Section 5, the data used, the methodology followed, the evaluation metrics, and the results of the experiments are presented. Finally, Section 6 draws the main conclusions from the experiments carried out.

2. General concepts about prediction interval estimation using MOPSO

The approach that will be used in this work to estimate PI's is the LUBE multi-objective approach described in [27], with some modifications related to the selection of the optimal solution, that will be explained later in this section.

The approach consists on a MLP with two outputs, the upper and lower bounds of the intervals. The MLP is trained using a Multi-objective Particle Swarm Optimization algorithm (MOPSO) [37]. MOPSO is a multi-objective version of the Particle Swarm Optimization (PSO) evolutionary algorithm [38]. Other multi-objective evolutionary algorithms could have been used to train the MLP, such as the NSGAII algorithm [39], but initial experiments showed MOPSO to perform better. The algorithm PSO has been designed for finding the optimum value for a fitness function that has a number of parameters that take values either in a discrete or in a real-valued parameter space. However, PSO is mostly used for real-value parameter optimization (and so is MOPSO).

The parameters (or weights) of the MLP are codified in the particles and the search is guided by the optimization of two objectives: interval width and coverage. The specifics about PI's and swarm optimization of MLP's are detailed as follows.

Suppose there is a set of n observations composed by several pairs (M_i, d_i) where M_i is a vector of input variables and d_i is the ith desired output. A function I has to be found where for each input M_i, two outputs are returned: p_i^{Low} and p_i^{Upp}, where p_i^{Low} and p_i^{Upp} represent the lower and upper bounds of the PI. d_i should be contained within both boundaries with probability equal to the desired coverage, called PI's Nominal Coverage (or $PINC$). Therefore, the function $I(M_i)$ for any set of observations M_i is described by Eq. (1).

$$I(M_i) = \left[p_i^{Low}, p_i^{Upp}\right] \text{ such that } prob(p_i^{Low} < d_i < p_i^{Upp}) = PINC \quad (1)$$

A way of approximating Eq. (1) is to maximize the Prediction Interval Coverage Probability ($PICP$) defined in Eq. (2). However, it is trivial to obtain 100% $PICP$ values using very wide PI's. Therefore, a search for a useful I function can be formulated as a multi-objective optimization problem where $PICP$ is maximized (Eq. (2)) and the Average Interval Width (AIW), defined in Eq. (3) is minimized.

$$PICP = \frac{1}{n}\sum_{1}^{n} \chi I(M_i) \quad (2)$$

$$\chi I(M_i) = \begin{cases} 1 & \text{if } p_i^{Low} < d_i < p_i^{Upp} \\ 0 & \text{otherwise} \end{cases}$$

$$AIW = \frac{1}{n}\sum_{1}^{n} |p_i^{Upp} - p_i^{Low}| \quad (3)$$

To approximate function I, a MLP is used. The MLP has three fully connected layers: input, hidden and output. Therefore it has two weight vectors, represented by two matrices $W^{(1)}$ (weights from the input to the hidden layer) and $W^{(2)}$ (weights from the hidden to the output layer). Thus, the outputs of the MLP, Y, can be expressed by Eq. (4)

$$X^{(1)} = f_{act}(M \cdot W^{(1)} + B^{(1)})$$
$$Y = f_{act}(X^{(1)} \cdot W^{(2)} + B^{(2)}) \quad (4)$$

where $X^{(1)}$ is the activation of the neurons of the hidden layer, $f_{act}(V)$ is the sigmoid activation function, $B^{(1)}$ are the bias added for hidden neurons and $B^{(2)}$ are the bias for the outputs neurons. In this case, the vector output Y represents the lower and upper bands of PI, $Y = (p^{low}, p^{upp})$.

Usually, MLP is typically trained using the backpropagation algorithm, although evolutionary computation techniques can also be used [40]. In the case of this article, backpropagation is unsuitable for two reasons: it does not support multiple objectives when learning and it is a supervised training algorithm, which would require the target values. The actual irradiance is available, but not the upper and lower bounds (p_i^{low} and p_i^{upp}) of the intervals. Then, the MLP has to be optimized through an alternative method and a Multi-Objective Particle Swarm Optimization (MOPSO) technique is proposed to solve both issues.

Particle Swarm Optimization (PSO) [41] is an iterative stochastic optimization algorithm for finding the optimum value of a function (usually called the fitness function) with several real-valued parameters. In PSO, search is conducted by a population (swarm) of candidate solutions (particles), which are the parameters to be optimized. The main advantage of PSO for this work is that it is a derivative-free method which is able to optimize any function, even if it is not non-differentiable. Since the estimation of PI's is formulated in this work as a multi-objective problem (more than one objective to be optimized), the multi-objective version of PSO has been used (MOPSO) [37]. The output of MOPSO is not a single solution, but a set of them, called the Pareto front or non-dominated set of solutions. A solution is called non-dominated if none of the objectives can be improved without degrading some of the other objectives. Thus, the Pareto front contains the solutions that are the best ones at least for one of the objectives [42]. The MOPSO algorithm tries to generate a non-dominated set of solutions well spread along the Pareto Front.

The particles used in this approach are the combination of $W^{(1)}$, $W^{(2)}$, $B^{(1)}$ and $B^{(2)}$, that is, the weights and biases of the MLP. The fitness function is defined using two objectives: *PICP* and *AIW* (Eqs. (2) and (3)). As it has been previously mentioned, *PICP* is to be maximized and *AIW* minimized. Therefore, the fitness function can be formulated as Eq. (5), where both goals in the function have to be minimized.

$$Fitness = (AIW, 1 - PICP) \tag{5}$$

In this case, the Pareto front is composed of a set of MLP's (each one represented by its weights and biases) (see Fig. 1). Each point or solution x in the Pareto front, with $(AIW_x, PICP_x)$, represents a MLP that, for that particular $PICP_x$ value, its AIW_x is as narrow as could be found by the optimization algorithm. From this Pareto front, a single solution can be selected according to the desired coverage (*PINC*). For instance, Fig. 1 shows how a point has been selected for $PINC = 0.9$ (horizontal black line at $1 - PINC = 0.1$ selects a point with $AIW = 0.39$). However, in general there might not be a solution in the front with exactly the required value of *PINC*. In [27] that point was selected as the solution closest to the target *PINC*, either above or below, using the training Pareto front (i.e. the one with the minimum $|PICP - PINC|$ value). In this work, this criterion has been modified because it was observed that in some cases, the solution with the *PICP* value closest to the target would not reach that target value (i.e. $PINC < PICP$). To avoid this problem, if there are solutions for which $PICP \geq PINC$, then the one with minimum $|PICP - PINC|$ will be selected. Otherwise, the one with minimum $|PICP - PINC|$ will be selected. In other words, if there are solutions for which $PINC < PICP$, these are preferred, even if there are other solutions with smaller $|PINC - PICP|$ but with $PINC > PICP$. Algorithm 1 displays the process to select the solution (*ps*) from the Pareto front in more detail. The inputs to the algorithm are: the set of MLP's in the Pareto front, the evaluation of each of these MLP's on the validation set according to *AIW* and *PICP* (validation front, *VF*), and the target *PINC*. The algorithm first computes D^+ and D^-, which is the set of solutions that fulfill/not fulfill the target *PINC*, respectively. If D^+ is not empty, then the solution *ps* whose *PICP* is closest to the target is selected from D^+. Otherwise, the solution *ps* with minimum $|PICP_k - PINC|$ is selected from D^-. The method returns the actual MLP I_{ps} corresponding to the selected solution *ps*, together with its evaluation on the validation set $(AIW_{ps}, PINC_{ps})$. This process guarantees that, if there are solutions that fulfill the required *PINC*, the best one among them will be selected.

Input: $MLP = \{I_1, I_2, \ldots, I_p\}$: set of MLP's in the Pareto front
Input: $VF = \{(1, AIW_1, PICP_1), (2, AIW_2, PICP_2), \ldots, (p, AIW_p, PICP_p)\}$: where $(k, AIW_k, PICP_k)$ is the kth MLP in the Pareto front evaluated on the validation set
Input: *PINC*
Result: $(I_{ps}, AIW_{ps}, PICP_{ps})$

$D^+ \leftarrow \{(k, AIW_k, PICP_k) \in VF \mid (PICP_k - PINC) \geq 0\}$
$D^- \leftarrow \{(k, AIW_k, PICP_k) \in VF \mid (PICP_k - PINC) < 0\}$
if $D^+ \neq \emptyset$ **then**
  $(ps, AIW_{ps}, PICP_{ps}) \leftarrow \underset{(k, AIW_k, PICP_k) \in D^+}{\operatorname{argmin}} (|PICP_k - PINC|)$
end
else
  $(ps, AIW_{ps}, PICP_{ps}) \leftarrow \underset{(k, AIW_k, PICP_k) \in D^-}{\operatorname{argmin}} (|PICP_k - PINC|)$
end
return $(I_{ps}, AIW_{ps}, PICP_{ps})$

Algorithm 1: Process to select a solution from the Pareto front given a target *PINC*

3. Forecasting models and meteorological weather types

In this section, the four models which are the inputs to MOPSO, QR and GBR are briefly described. Also, the four weather types (WT) that will be used to improve the quality of the PI's will be presented.

3.1. Description of models

- **Smart Persistence**: This model provides forecasts by assuming that the clear sky index keeps constant along the forecasting period. The forecasts are given by Eq. (6):

$$GHI(t + h) = \frac{GHI(t)}{GHI_{cs}(t)} \cdot GHI_{cs}(t + h) \tag{6}$$

  where $GHI(t)$ is the measured irradiance when the forecasts is issued and GHI_{cs} is the corresponding clear sky irradiance. The European Solar Radiation Atlas (ESRA) clear sky model [43] was used.
- **Satellite-based model**: Cloud Index (CI) maps, obtained based on the Heliosat −2 method [44], are used to provide satellite-based short-term GHI forecasts. Firstly, consecutive CI maps are used to estimate the Cloud Motion Vectors (CMVs), then the underlying atmospheric flow is estimated to finally provide the cloudiness and solar radiation forecasts. The methodology here used is fully described in [14], and includes the use of the streamlines of the cloud speed vectors.
- **WRF-Solar**: is a particular physical configuration of the WRF numerical weather prediction model version 3.6 [28]. The model is configured with one domain of 5 km spatial resolution (30 s of model time-step) and 50 vertical levels [14]. Initial and boundary conditions are taken from the National Centers for Environmental Prediction (NCEP) GFS [0.5

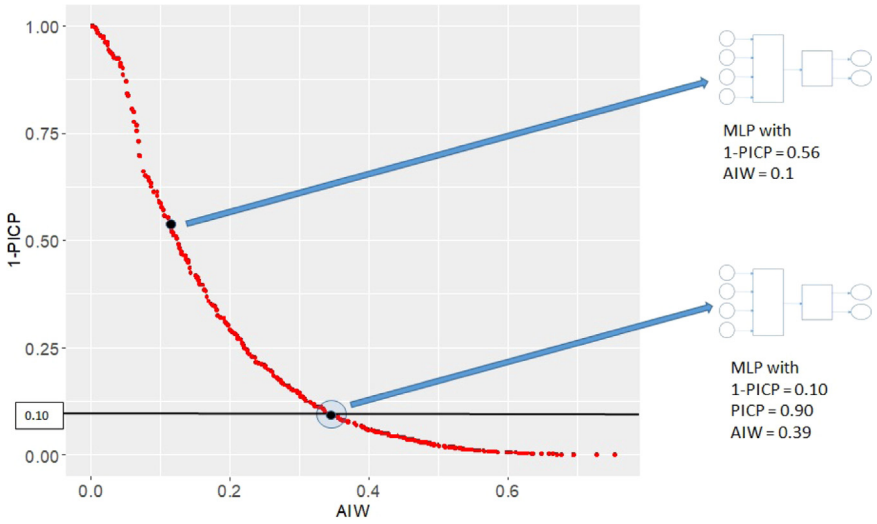

Fig. 1. Pareto front: Each point represents a MLP that estimates PI's with $1 - PICP$ and AIW values.

Deg] dataset (National Centers for Environmental Prediction, NCEP, 2006). For each day, three 18-h simulations are run starting at 00, 06 and 12 UTC, discarding the first 6 simulated hours as spin-up. Outputs are saved every 15 min, at the moment when the satellite passed over the study area.

- **CIADCast**: The CIADCast model [14,29] is a hybrid model which aims to combine the accuracy of the cloud representation in satellite images with the dynamical capabilities of the WRF model. In the CIADCast model, CI retrievals from satellite are advected and diffused to obtain cloudiness forecasts and, therefore, radiation forecasts. To this end, CI maps are ingested in the model for every satellite gathering time (at a particular vertical layer of each column of the model) as mass mixing ratios. Then, the WRF-Solar model is used to advect and diffuse the CI values as dynamical tracers mainly horizontally, but also vertically. Finally, CI forecasts are de-normalized to provide GHI forecasts. The CIADCast forecasts are run simultaneously with the WRF-Solar model, as explained in the previous paragraph.

3.2. Meteorological weather types

In this work, the four WT's described in [30] are considered in the probabilistic forecasting procedure. The methodology used to identify these WT's is based on the use one-minute GHI measurements. Particularly, based on this measurement, several statistics (accounting for the radiation variability) are computed over a period of three hours. The statistics are then used as inputs of a clustering procedure to identify the WT's. Therefore, a specific WT can be assigned to every three hours period. Remarkably, this methodology was specifically designed to account for the solar irradiance temporal variability in the study area. The WT corresponding to the period when the forecast is issued can be easily identified based the three previous hours measurements. As summary, the four WTs are:

- **Overcast**: this first WT accounts for the presence of a synoptic perturbation over the study area. In terms of cloudiness, this means overcast conditions and scarce variability of the GHI along a period of several hours/days.
- **Transient weather conditions**: this second WT is usually observed some hours/days before or after a synoptic perturbation passes over the study area. Broken sky conditions are then observed, with a high variability of the GHI. These sky conditions are the most challenging for solar radiation forecasting.
- **Local weather features**: this third WT accounts for the presence of moderate high pressure anomalies over the study area, which allows the development of local weather features such as, for instance, convection. As a consequence, certain amount of cloudiness can be observed over the study area.
- **Clear sky**: this last WT accounts for the presence of a high pressure system over the study area. As a consequence, clear sky is mostly observed, with little if any GHI variability in period of several hours/days.

Each forecast is assigned to one of these four categories, according to the WT observed at the time when the prediction was issued.

4. Prediction intervals for the blending of forecast models using weather types

In this section, the multi-objective LUBE approach described in Section 2 is applied to estimate PI's for the forecasting of GHI at different prediction horizons h, using as inputs the GHI predictions provided by four different models: Smart Persistence (P), Satellite (S), WRF-Solar (W) and CIADcast (C). Two proposals are made to estimate PI's: the base proposal that uses only the four models (as predictors), while the weather-type proposal that includes also information about the kind of weather expected for the next hour (see Fig. 2).

The base proposal consists in computing functions I^h (being h the forecasting horizon) that attempts to fulfill Eq. (7). As it is shown in Fig. 2 (left), the inputs to the MLP at instant t are the GHI predictions of the four models for horizon $t+h$ (P_t^h, S_t^h, W_t^h and, C_t^h). The outputs of the MLP are the lower and upper bounds of the intervals at time $t+h$ (p_{t+h}^{Low} and p_{t+h}^{Upp}).

$$I^h(P_t^h, S_t^h, W_t^h, C_t^h) = \left[p_{t+h}^{Low}, p_{t+h}^{Upp}\right] \quad \text{such that } prob(p_{t+h}^{Low} < ghi_{t+h} < p_{t+h}^{Upp}) = PINC \tag{7}$$

The weather-type proposal is formally similar to the base one, except that weather-type information is added to the inputs of MLP to approximate the functions I^h. Information about weather conditions, i.e. sky conditions, can greatly influence the prediction intervals in short-term horizons. For instance, clear sky conditions may results in narrower intervals, while broken cloud conditions require wider ones. Incorporating weather type information gives

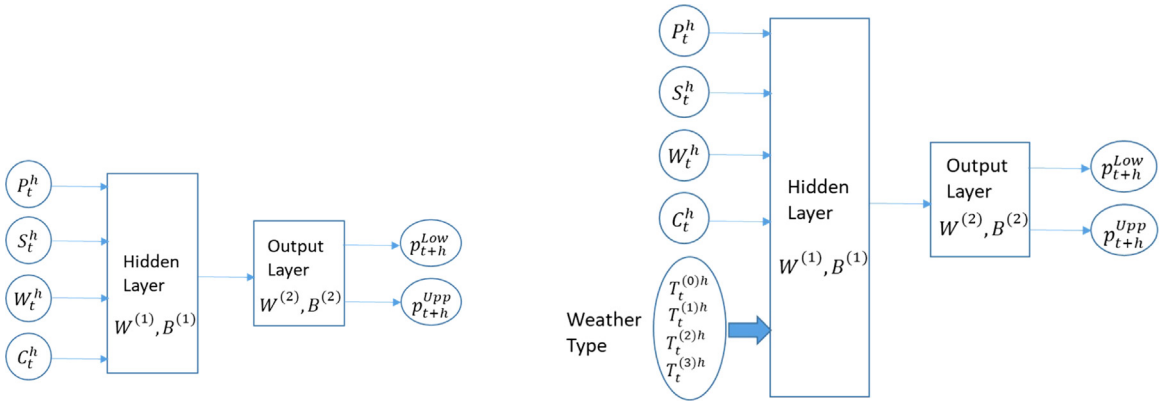

Fig. 2. Structure of MLP's for the LUBE approach: base proposal (left), weather type proposal (right).

context to function I^h and may result in better prediction intervals. In this work, the four weather-types described in Section 3 are used.

The functions I^h for the weather type proposal are represented by Eq. (8). Additionally to the GHI predictions of the four predictors for horizon $t+h$, the weather type information is included as inputs (see right part in Fig. 2). Weather-type is a discrete variable, and a common approach to use discrete attributes in neural networks is to use one-hot-encoding, where a discrete variable of q values is split into q binary variables. Thus, the four variables $(T_t^{h(0)}, T_t^{h(1)}, T_t^{h(2)}, T_t^{h(3)})$ of Eq. (8) are binary variables that identify which of the four weather-types is forecast at time t to be present for horizon h. Only one of them can be set to 1 (while the remaining ones are set to 0).

$$I^h(P_t^h, S_t^h, W_t^h, C_t^h, \mathbf{T_t^{h(0)}}, \mathbf{T_t^{h(1)}}, \mathbf{T_t^{h(2)}}, \mathbf{T_t^{h(3)}}) = \left[p_{t+h}^{Low}, p_{t+h}^{Upp}\right] \quad \text{such that } prob(p_{t+h}^{Low} < ghi_{t+h} < p_{t+h}^{Upp}) = PINC \tag{8}$$

As mentioned in Section 2, in this work we use a multi-objective algorithm to train MLP's and the output of the algorithm is a Pareto front, where each point is a function I^h approximated by a MLP (see Fig. 1). Given a user-defined desired *PINC*, a particular I^h appropriate for that *PINC* is extracted from the front using the process described in Algorithm 1. It is important to remark that in this work and for both approaches, base and WT, the multi-objective optimizer is run for every horizon h, thus obtaining a different Pareto front for every h.

5. Experimental validation

In this section, the performance of the proposals for the estimation of PI's is shown and analyzed. In what follows, the base approach is denoted as *MOPSO − Base* and the weather type proposal as *MOPSO−WT*. In order to compare with baseline methods, QR and GBR have been also used to estimate PI's for each forecasting horizon h. In this work, QR and GBR has been done using the four predictors as inputs (*QR−Base*, *GB−Base*) and also using the information about weather type (*QR−WT*, *GB−WT*) in the same way than for the MOPSO approach. A brief explanation of using QR and GBR to estimate PI's is included in Section 5.2. The following subsections describe the dataset used to fit and evaluate the methods (Section 5.1), the methodology employed to run the experiments and the metrics used to compare the quality of the PI's (Section 5.2), and the empirical results and discussion (Section 5.3).

Table 1
Number of instances for every horizon.

Horizon	Instances
15	7858
30	7773
45	7654
60	7515

5.1. Dataset

The dataset used in this work contains GHI measurements at Seville station and their forecasts made by the four different forecast models: Smart Persistence, Satellite WRF-Solar and CIADcast. Forecasts are obtained at four time horizons: 15, 30, 45, and 60 min. GHI has been measured with a Kipp & Zonen CMP6 pyranometer with a 1 min sample rate. The observation covers the period from March 2015 to February 2017. A more detailed description of the dataset is provided in [30].

Every instance in the dataset follows this pattern: $(M_t, d_t) = ((P_t^h, S_t^h, W_t^h, C_t^h), GHI_{t+h})$ where the P_t^h, S_t^h, W_t^h, C_t^h are the predictions of the four predictors issued at time t and making predictions for time $t+h$. GHI_{t+h} is the actual value measured at time $t+h$. t change with a frequency of 15 min. To select the relevant hours of the day, and ensure the quality of the data, records with a zenithal angle larger than 75 degrees have been removed. In order to remove the influence of the daily cycle, the *GHI* has been normalized to the clear sky index. Table 1 shows the number of data instances for every horizon.

5.2. Methodology and evaluation measures

Cross-Validation

In this work, the cross-validation (CV) methodology has been applied to evaluate the different methods. Standard CV partitions the dataset randomly, but in our case it is not the most appropriate approach, because there is temporal dependency between samples, and this is known to result in overly optimistic evaluations. Therefore, the CV employed in this work consists on splitting the data into 4 groups, one for each week of every month. Therefore, fold 1 contains the first week of each month (January, February, …). Fold 2, the second week, and so on. This guarantees that, at least, training and testing partitions will never contain instances belonging to the same week, which allows a more realistic analysis of the performance of the methods. A validation set is also used to select the optimal parameters of the approaches (number of hidden neurons for MLP and number

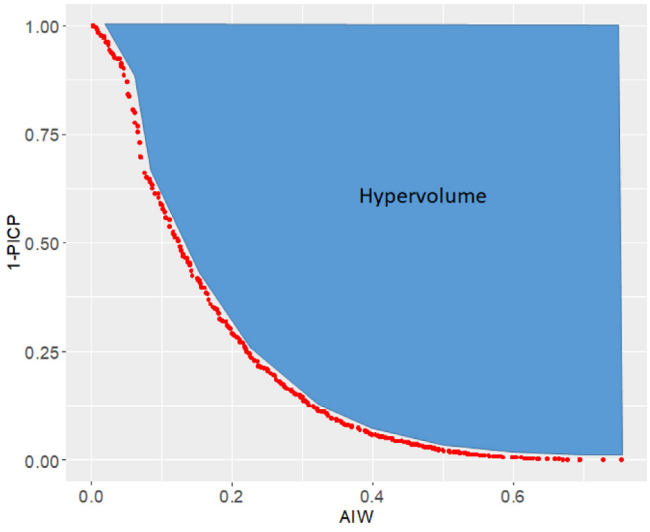

Fig. 3. Hypervolume of a Pareto front.

of iterations for PSO) and to select a solution from the Pareto front for a target *PINC*. For this purpose, each training set is again divided into training and validation sets. Given that each training set contains three weeks, the validation set uses the third week, and the remaining two weeks are used for fitting the models. In summary, training contains two weeks of every month, validation contains one week, and the remaining week is used for testing.

Given the stochastic nature of PSO algorithm and GBR method, 30 runs of PSO and GBR have been carried out using different random seeds for each fold of the CV. For QR, only one run has been made due the non-stochastic nature of QR.

Grid Search for MOPSO parameter selection

For MOPSO approaches ($MOPSO-Base$, $MOPSO-WT$) we have performed experiments to select the optimal parameters. The most important parameters are the number of hidden neurons for the MLP and the number of iterations for PSO algorithm. Experiments have been done using 3, 5, 10, 15, 20 and, 30 hidden neurons and the number of iterations for PSO algorithm goes from 200 to 30 000, saving solutions (Pareto fronts) each 200 iterations.

The best configuration is selected using the hyper-volume indicator [45]. This metric is commonly used in the context of multi-objective optimization algorithms for comparing two sets of solutions (i.e. two Pareto fronts). The hyper-volume evaluates the quality of a set of solutions using a single value and it measures the volume (area in this case) of the dominated portion of the objective space. In Fig. 3, the hyper-volume of the Pareto front is shown. It is observed that as the front gets closer to the origin (smaller width and higher value of *PICP*), higher hyper-volume values are obtained. In our problem, as the hyper-volume increases, the solutions in the front will be better. In this work, this indicator is evaluated on the validation set (that is, the Pareto front of MLP's obtained by MOPSO using the training set, are evaluated on a different set, the validation set). This makes the selection of the best configuration independent of the training set, and prevents overfitting to some extent.

The procedure for optimal parameter selection is a grid search, where all possible combinations of parameters are evaluated. The grid search is carried out for each of the 30 runs, with the aim of finding the best configuration of parameters for each case. Each combination of parameters (hidden neuron and iteration) results in a Pareto front, which is evaluated using the validation set of each fold. Algorithm 2 displays in detail how grid search is applied. Basically, it uses a nested loop to explore all possible combinations of iterations and number of hidden neurons. For each combination, MOPSO is run and the Pareto front of MLP's is obtained. Then, each MLP in the front is evaluated on the validation set in order to obtain its *AIW* and *PICP* (validation front). Finally, the hyper-volume of the validation front resulting from this particular combination of iterations and neurons is computed. The method returns the combination of maximum hyper-volume.

Input: TS: Training Set
Input: VS: Validation Set
Result: $(Best\_iter, Best\_hidden)$

$Hidden = \{3, 5, 10, 15, 20, 30\}$
$Iter = \{200, 400, 600, ..., 29800, 30000\}$
for $h \in Hidden$ **do**
  for $i \in Iter$ **do**
    # Obtain $MLP = \{I_1, I_2, \ldots, I_p\}$: set of MLP's in the Pareto front
    $MLP \leftarrow MOPSO(i, h, TS)$
    # Evaluate MLP on the validation set (VS)
    $VF \leftarrow \{(1, AIW(I_1, VS), PICP(I_1, VS)), \ldots, (p, AIW(I_p, VS), PICP(I_p, VS))\}$
    # Compute hyper-volume of the Validation Front (VF)
    $hv[i, h] \leftarrow HyperVolume(VF)$
  end
end
$(Best\_iter, Best\_hidden) \leftarrow \underset{(i,h)}{\text{argmax}}(hv[i, h])$
return $(Best\_iter, Best\_hidden)$

Algorithm 2: Grid Search process to select the best number of iterations and hidden neurons using the validation set.

Evaluation metrics

In this work three different measures have been used to evaluate the quality of PI's: *CWC*, *Ratio*, and *AIW*. The first one (*CWC*) was proposed in [46] and it evaluates if the *PICP* value of the solution is larger or equal than the target *PINC*, together with the width of the interval. Eq. (9) describes this measure for a target *PINC* value.

$$CWC = AIW \cdot \left(1 + (PICP < PINC) \cdot (PICP) \cdot e^{-b(PICP-PINC)}\right) \tag{9}$$

where b is set to 50 (as it was proposed by the authors) and $PICP < PINC$ is one if the condition is satisfied and zero otherwise. *CWC* is expected to be small for good solutions. As it is observed, *CWC* exponentially penalizes solutions with *PICP*s below the target *PINC*, otherwise the *CWC* is equal to the width. If the *PICP* value of the interval does not satisfy the target *PINC*, *CWC* is basically dominated by the exponential difference ($e^{-b*(PICP-PINC)}$), and the *AIW* has almost no weight on the final *CWC* value. This is not completely fair for solutions that just miss the target *PINC* (i.e. solutions with *PICP* slightly below *PINC*), specially if the *AIW* is small. Thus, in this work, the *Ratio* measure is also used to provide another view on the results.

Ratio measures the ratio between the *PICP* and the average width (*AIW*), and it is calculated following Eq. (10). Basically, it measures the tradeoff between *PICP* and width. Higher values of *Ratio* are obtained by the best solutions because they achieve high *PICP* values by means of narrow intervals. High *PICP* values achieved using large intervals will result in low *Ratio* values. Contrary to *CWC*, even if a solution obtains a *PICP* slightly smaller than the target *PINC*, the *Ratio* measure will still show it as a good solution, if the *AIW* is small enough.

$$Ratio = \frac{PICP}{AIW} \tag{10}$$

Finally, the average interval width (*AIW*) given by Eq. (3) will also be provided to compare the different approaches.

Baseline Methods: Quantile Regression and Gradient Boosting

For comparison purposes, QR [34] and GBR [36] have been used as baseline. QR is a linear method while GBR is non-linear. Both methods obtain models that estimate some given quantiles by minimizing the quantile loss (see Eq. (11)).

$$L(\hat{y}_i, y_i) = \max(q(\hat{y}_i - y_i), (q-1)(\hat{y}_i - y_i)) \tag{11}$$

where q is the desired quantile, y_i is the ground truth, and $\hat{y}_i$ is the prediction of the model.

Gradient Boosting produces regression models which are ensembles of trees. The ensemble is trained sequentially, by adding one base model (a tree) at a time. Gradient Boosting can produce ensembles that minimize any loss (or cost) function by training each base model, so that it points in the negative gradient direction (of the loss function).

For both methods, QR and GBR, the procedure to estimate PI basically consists on generating two models (linear for QR and non-linear for GBR), one for the upper bound of the interval and another one for the lower bound. Formally, this procedure is as follows. Let q_1 and q_2 be the $\frac{1-PINC}{2}$ and $\frac{1+PINC}{2}$ quantiles, respectively. Quantile q_1 leaves a $\frac{1-PINC}{2}$ probability tail to the left of the distribution and quantile q_2 leaves $\frac{1-PINC}{2}$ probability tail to the right of the distribution. Therefore, the interval $[q_1, q_2]$ has a coverage of *PINC*. QR and GBR are used to fit two models that, given some particular input X_i, returns q_1 and q_2 with which the interval $[q_1, q_2]$ can be constructed.

Similarly to MOPSO, QR and GBR have been tested using as inputs the GHI predictions of the four predictors for horizon $t+h$, (P_t^h, S_t^h, W_t^h and, C_t^h), and using also the weather type information as inputs to the linear models: (P_t^h, S_t^h, W_t^h, C_t^h, $T_t^{h(0)}$, $T_t^{h(1)}$, $T_t^{h(2)}$ and, $T_t^{h(3)}$).

QR and GBR both need to be trained for each target *PINC* (unlike MOPSO, which is run just once, and then a solution for the desired *PINC* is drawn from the front).

In the case of GBR, an exhaustive experimentation has been done in order to determine the optimal parameters. For GBR the three main parameters are the number of trees in the ensemble (*ntrees*), the maximum depth of each tree (*depth*) and the shrinkage or learning rate (*shrinkage*). Experiments using *ntrees* = 100, 200, 500, 1000, 1500, ..., 5000, *depth* = 2, 4, 6, ..., 14 and *shrinkage* = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5 have been carried out following a methodology similar to MOPSO. That is, all possible combinations of parameters are evaluated for each run and each fold. The study of best parameters has also been carried out for each prediction horizon, as for the MOPSO approach, but also for each *PINC* target value. For GBR, the best combination of parameters has been selected using the following criterion: the solutions whose coverage or *PICP* value for the validation set is greater than or equal to the desired coverage *PINC* are selected. From all these solutions, the one with the smallest *AIW* is chosen.

5.3. Experimental results

In this section we present and compare the results obtained with the different approaches presented in Section 4 to estimate PI's for the blending of the four predictors (Smart Persistence, WRF-solar, CIADcast, and Satellite) to forecast GHI at four time horizons (15, 30, 45 and 60 min). Three different values for the target *PINC* have been used: *PINC* = 0.85, *PINC* = 0.90 and *PINC* = 0.95.

For MOPSO approaches (*MOPSO − Base*, *MOPSO − WT*) we have performed experiments to select the optimal parameters for each forecasting horizon using the grid search explained before (see Algorithm 2). Table 2 displays the best number of hidden neurons and iterations selected on average (the average of the 30 runs). It can be seen that *MOPSO − WT* usually requires more hidden neurons than *MOPSO − Base*. This is reasonable, given the extended number of inputs of the former. No particular trend can be observed on the number of hidden neurons depending on the horizon. Finally, the number of iterations is usually closer to the maximum value (30 000 iterations). Some experiments were run by hand using extra iterations, but no significant differences were observed.

Table 2
Average number of hidden neurons and iterations selected for each MOPSO method and horizon.

Horizon	*MOPSO − Base*		*MOPSO − WT*	
	Hidden	Iterations	Hidden	Iterations
15	6.8	29 753	10.9	29 727
30	8.3	29 620	8.9	29 833
45	7.4	29 393	9.9	29 480
60	6.2	29 600	9.8	29 307

Table 3
Average number of trees, depth and shrinkage selected for GBR approaches, for each horizon and each *PINC* value.

Horizon	*GB − Base*			*GB − WT*		
	Trees	Depth	Shrinkage	Trees	Depth	Shrinkage
PINC = 0.85						
15	3857	5.87	0.005	4167	6.13	0.0017
30	3517	4.67	0.003	3070	5.6	0.0118
45	2817	3.67	0.0103	2723	5.07	0.0067
60	3333	7.93	0.001	3373	3.4	0.00457
PINC = 0.90						
15	4000	4	0.001	3633	5.47	0.001
30	3583	2.13	0.001	3517	4	0.001
45	2833	4.93	0.0016	3517	3.93	0.001
60	1573	3.33	0.0043	2683	3	0.0019
PINC = 0.95						
15	2020	3.53	0.0032	1570	11.2	0.0015
30	2140	2	0.0041	4217	2	0.0007
45	150	11	0.0005	1973	2.47	0.0041
60	220	13.6	0.0008	4113	2	0.0011

Extensive parameter tuning has also been carried out for the GBR approaches (*GB − Base* and *GB − WT*). Best parameters are displayed in Table 3. No remarkable trend can be observed in this case. Contrariwise to MOPSO, different models have to be trained for different *PINC* values, and therefore, parameter tuning has also to be performed for each *PINC*.

Table 4 shows the averages of *Ratio*, *CWC* and *AIW* obtained by *MOPSO − Base*, *MOPSO − WT*, *QR − Base*, *QR − WT*, *GB − Base* and *GB − WT* blending approaches (for the testing dataset) for different values of *PINC* and for each forecasting horizon. MOPSO and GBR methods are reported in terms of the average of the 4 folds of CV and of the 30 runs, while for QR the average is over the 4 folds of CV (because QR is not stochastic). The best results have been boldfaced. In order to improve the understanding of the results, they have also been represented graphically in Figs. 4, 5, and 6. Statistical significance tests have also been carried out by means of the Wilcoxon signed rank test [47]. The test has been applied to the three measures, for all forecasting horizons and for the three target *PINC* values. Results of the statistical significance are shown in Table 5. Statistical comparisons have been done for *MOPSO − WT* vs. *MOPSO − Base*. The significance test was also performed between and *MOPSO − WT* vs. *QR − WT* and *MOPSO − WT* vs. *GB − WT* because the approaches using weather types as inputs generally perform better than without.

Table 4
Test averages of *Ratio*, *CWC* and *AIW* for each *PINC* desired value and for each forecasting horizon. MOPSO and GBR methods are reported in terms of the average of the 4 folds of CV and of the 30 runs. For QR the average is over the 4 folds of CV.

Method	*Ratio*				*CWC*				*AIW*			
	15	30	45	60	15	30	45	60	15	30	45	60
PINC = 0.85												
MOPSO − Base	3.829	3.546	3.380	**3.309**	0.317	0.340	0.352	0.407	0.223	0.240	0.252	**0.257**
MOPSO − WT	3.832	3.555	**3.420**	3.297	0.325	0.296	**0.348**	**0.333**	0.222	0.241	**0.249**	0.259
QR − Base	3.220	3.104	2.993	2.893	0.508	0.559	0.593	0.599	0.263	0.272	0.282	0.292
QR − WT	3.121	3.002	2.896	2.788	0.619	0.702	0.724	0.844	0.270	0.279	0.289	0.299
GB − Base	3.891	**3.596**	3.307	3.182	**0.233**	0.439	0.445	0.536	0.219	**0.236**	0.257	0.266
GB − WT	**3.911**	3.544	3.355	3.078	0.287	**0.255**	0.555	0.385	**0.218**	0.241	0.251	0.276
PINC = 0.90												
MOPSO − Base	3.121	2.995	2.878	2.815	0.424	0.455	0.508	0.477	0.289	0.301	0.313	0.320
MOPSO − WT	**3.247**	**3.094**	**2.974**	**2.859**	0.404	0.391	**0.483**	**0.436**	**0.278**	**0.292**	**0.303**	**0.316**
QR − Base	2.601	2.571	2.508	2.441	0.723	0.765	0.750	0.794	0.344	0.348	0.357	0.367
QR − WT	2.633	2.648	2.589	2.509	0.884	0.818	0.818	0.813	0.337	0.336	0.344	0.356
GB − Base	3.211	2.665	2.597	2.152	0.469	0.379	0.586	0.829	0.280	0.338	0.346	0.418
GB − WT	3.145	2.832	2.704	2.450	**0.383**	**0.319**	0.548	0.742	0.286	0.319	0.333	0.366
PINC = 0.95												
MOPSO − Base	2.479	2.387	2.329	2.275	**0.535**	0.641	0.654	0.592	0.384	0.398	0.408	0.418
MOPSO − WT	**2.560**	**2.416**	**2.366**	**2.340**	0.575	**0.495**	**0.573**	**0.589**	**0.372**	**0.394**	**0.402**	**0.407**
QR − Base	2.008	1.986	1.957	1.925	1.162	1.131	1.137	1.094	0.469	0.474	0.482	0.491
QR − WT	2.201	2.181	2.146	2.165	1.023	1.040	0.981	1.063	0.428	0.432	0.440	0.435
GB − Base	1.693	1.653	1.400	1.418	1.081	1.163	1.414	1.382	0.561	0.574	0.677	0.668
GB − WT	1.705	1.701	1.647	1.674	1.004	1.088	0.771	1.145	0.556	0.558	0.577	0.567

Table 5
Statistical significance tests for *MOPSO − WT* vs. *MOPSO − Base*, *MOPSO − WT* vs. *QR − WT* and *MOPSO − WT* vs. *GB − WT* at different *PINC* values and horizons. Measures considered: *Ratio*, *CWC*, and *AIW*.

Methods	*Ratio*				*CWC*				*AIW*			
	15	30	45	60	15	30	45	60	15	30	45	60
PINC = 0.85												
MOPSO − WT vs. *MOPSO − Base*	=	=	=	=	=	=	=	+	=	=	=	=
MOPSO − WT vs. *QR − WT*	+	+	+	+	+	+	+	+	+	+	+	+
MOPSO − WT vs. *GB − WT*	=	=	+	+	−	−	+	+	−	=	=	+
PINC = 0.90												
MOPSO − WT vs. *MOPSO − Base*	+	+	+	+	+	+	=	=	+	+	+	+
MOPSO − WT vs. *QR − WT*	+	+	+	+	+	+	+	+	+	+	+	+
MOPSO − WT vs. *GB − WT*	+	+	+	+	=	+	+	+	+	+	+	+
PINC = 0.95												
MOPSO − WT vs. *MOPSO − Base*	+	+	+	+	=	+	+	+	+	=	+	+
MOPSO − WT vs. *QR − WT*	+	+	+	+	+	+	+	+	+	+	+	+
MOPSO − WT vs. *GB − WT*	+	+	+	+	+	+	+	+	+	+	+	+

Symbol + means that the first method is significantly better than the second one; = means that no significant difference was found, and symbol − that the second one is significantly better than the first one. Next, those results will be analyzed in detail.

Table 4 and Figs. 4, 5, and 6 show that the quality of PI's (with respect to *Ratio*, *CWC*, and *AIW*) obtained with MOPSO (*MOPSO − Base* and *MOPSO − WT*) is better than those obtained with QR (*QR − Base* and *QR − WT*). That can be observed for all measures, for all forecasting horizons and for all *PINC* values. It is noticeable that QR has a poor performance for *CWC* compared to the MOPSO blending approaches, which implies that QR has difficulties to achieve the target *PINC*. The statistical comparison between *MOPSO − WT* vs. *QR − WT* (Table 5) confirms that *MOPSO − WT* is always significantly better than *QR − WT*. Comparing MOPSO and GBR, we can observe in Table 4 and Figs. 4, 5, and 6 that MOPSO approaches (*MOPSO − Base* and *MOPSO − WT*) perform better than GBR approaches (*GB − Base* and *GB − WT*) for all horizons and all measures when *PINC* value is 0.95. That is confirmed observing the statistical comparison between *MOPSO − WT* vs. *GB − WT* shown in Table 5. GBR method has difficulties finding good solutions for high demanding coverage values, even being worse than QR approaches, at least for *Ratio* and *AIW* measures, as it can be clearly seen in Figs. 4 and 5. For *PINC* values of 0.90, MOPSO approaches also perform better than GBR approaches, except for the horizon 15 and 30 and the *CWC* measure, where *GB − WT* provides the best result in terms of average, although as can be seen in Table 5, that differences are not statistically significant. Comparing GBR and QR approaches, for *PINC* = 0.90 GBR is better than QR, except for horizon 60, where GBR and QR perform quite similar. The superiority of GBR approaches is mainly observed for less demanding coverage (*PINC* = 0.85) and short forecast horizons (15 and 30 min). This superiority is observed for all measures and in terms of the average. However, these differences are only statistically significant for *CWC* and horizons 15 and 30 and for *AIW* and horizon 15. In the rest of cases, there are no significant differences. For horizons 45 and 60 min, MOPSO approaches (*MOPSO − Base* and *MOPSO − WT*) perform always better than GBR approaches for all measures, which is verified with statistical tests shown in Table 5. For this *PINC*, GBR performs better than QR in all horizons and all measures.

In order to analyze the influence of weather type on PI's, *MOPSO − WT* vs. *MOPSO − Base*, *QR − WT* vs. *QR − Base* and *GB − Base* vs. *GB − WT* will be compared. Regarding *Ratio*, it is observed in Table 4 and Fig. 4 that *MOPSO − WT* obtains the best values of *Ratio* for all forecasting horizons and target *PINC*s (except at

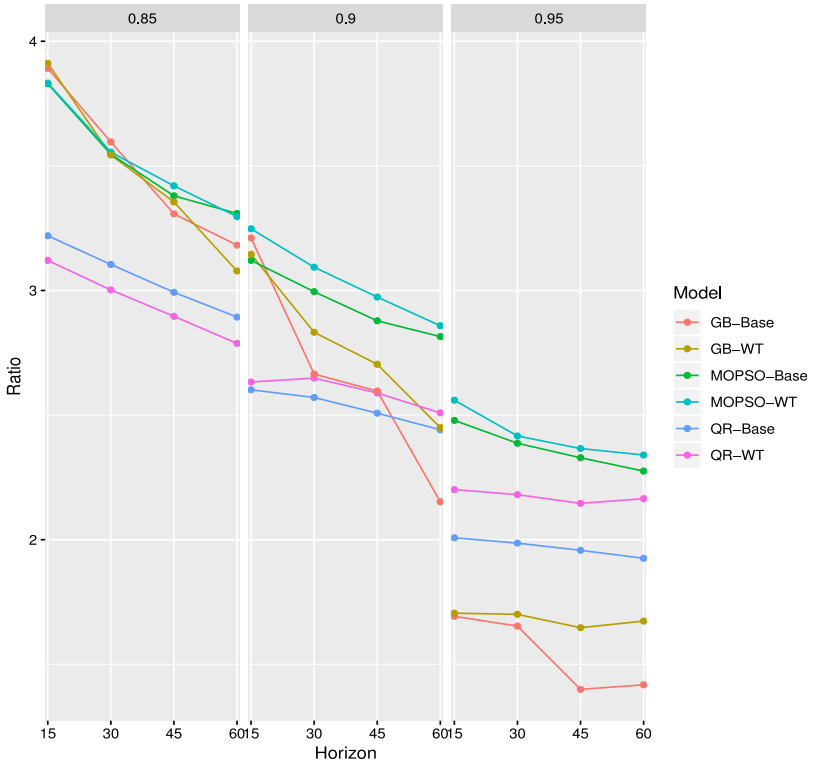

Fig. 4. Average *Ratio* of $MOPSO-Base$, $MOPSO-WT$, $QR-Base$, $QR-WT$, $GB-Base$ and $GB-WT$ along the forecast horizon (x-axis) for each *PINC* value (0.85, 0.90, 0.95).

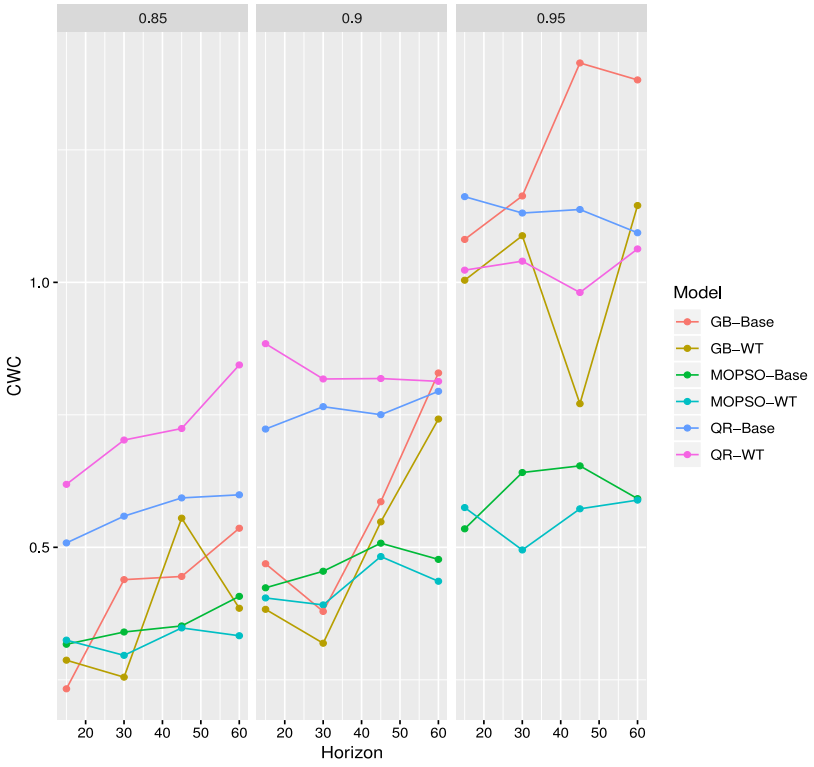

Fig. 5. Average *CWC* of $MOPSO-Base$, $MOPSO-WT$, $QR-Base$, $QR-WT$, $GB-Base$ and $GB-WT$ along the forecast horizon (x-axis) for each *PINC* value (0.85, 0.90, 0.95).

horizon 60 and *PINC* 0.85, where $MOPSO-Base$ is slightly better than $MOPSO-WT$, 3.309 versus 3.297). The statistical significance tests (Table 5) confirm that $MOPSO-WT$ is significantly better than $MOPSO-Base$ for *PINC*s 0.95 and 0.90 in all horizons, but not for *PINC* = 0.85 for which $MOPSO-WT$ and $MOPSO-Base$ are not significantly different. With respect to QR, Table 4 and Fig. 4 show that the use of weather conditions also helps to obtain better values of *Ratio*, but only for *PINC* values of 0.90 and 0.95. In this

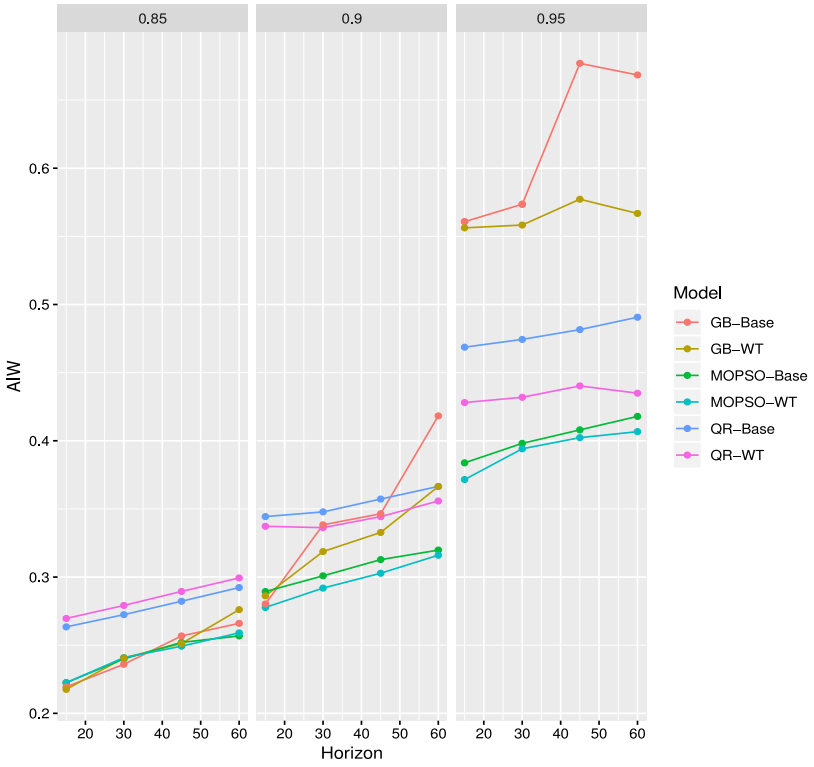

Fig. 6. Average *AIW* of $MOPSO-Base$, $MOPSO-WT$, $QR-Base$, $QR-WT$, $GB-Base$ and $GB-WT$ along the forecast horizon (x-axis) for each *PINC* value (0.85, 0.90, 0.95).

case, $QR-WT$ is worse than $QR-Base$ for $PINC = 0.85$. Regarding GBR, in most of cases the use of weather type information as inputs ($GB-WT$) also helps to obtain better results, except in certain cases, such as $PINC = 0.85$ and horizons 30 and 60, and $PINC = 0.90$ and horizon 15.

Similar results can be observed with respect to average *CWC* values in Table 4 and Fig. 6, where the use of weather types helps the $MOPSO-WT$ approach to obtain better results than $MOPSO-Base$, for all *PINC* values and forecasting horizons, except for horizon 15 and *PINC* values 0.85 and 0.95. In any case, and similarly to *Ratio*, for $PINC = 0.85$ weather type information does not make a sufficient contribution to make differences statistically significant for *CWC*. But they are statistically significant for *PINC* 0.90 and 0.95 in most of the cases (again, this is similar to the previous results concerning the *Ratio*). Regarding the QR blending approach, weather type is only useful for *PINC* 0.95 (see Table 4 and Fig. 6). For the rest of *PINC* values, $QR-WT$ obtains worse *CWC* values than $QR-Base$. In the case of GBR, $GB-WT$ performs always better than $GB-Base$, except for $PINC = 0.85$ and horizons 15 and 45 min.

With respect to *AIW*, Table 4 and Fig. 6 show that intervals are generally narrower when weather types are used. This is always true for the MOPSO, QR and GBR blending approaches, when *PINC* is 0.90 and 0.95, but not for 0.85. For the 0.85 *PINC* value, the averages of *AIW* for both $MOPSO-Base$ and $MOPSO-WT$ approaches are very similar for all forecasting horizons. This is confirmed by the statistical significance tests (Table 5). In the case of QR, weather type is only helpful for *PINC* values 0.90 and 0.95, but not for 0.85 (where *AIW* tends to be worse when weather type is used). And for GBR, the information about weather types allows to obtain narrower intervals, except for $PINC = 0.85$, where $GB-WT$ is slightly worse for horizons 30 and 60.

In summary, for *PINC* value 0.95, the best method is $MOPSO-WT$ for all metrics and all horizons. $MOPSO-WT$ enhances by around 2% the *Ratio*, as compared to the $MOPSO-base$ method. This improvement is 7% for *CWC* and 2% for *AIW*. The reported improvements are the average for all horizons. Similar improvements can be observed for $PINC = 0.90$: 3% (*Ratio*), 8% (*CWC*), and 3% (*AIW*). Comparing $MOPSO-WT$ and $QR-WT$ for $PINC = 0.95$, improvements are typically larger: 11% for *Ratio*, 46% for *CWC*, and 9% for *AIW*. For $PINC = 0.90$ the improvements are also around those values: 17% (*Ratio*) , 49% (*CWC*), 14% (*AIW*). Finally, the comparison of $MOPSO-WT$ vs. $GBR-WT$ for $PINC = 0.95$ shows even larger improvements for $MOPSO-WT$: 44% (*Ratio*), 43% (*CWC*), and 30% (*AIW*). For $PINC = 0.90$, improvements of $MOPSO-WT$ are smaller, but still quite significant: 10% (*Ratio*), 6% (*CWC*), and 9% (*AIW*).

For less demanding coverage ($PINC = 0.85$), weather-type does not offer improvement as large as before. Only for some horizons $MOPSO-WT$ is better than $MOPSO-Base$. But per-horizon averages display smaller percentage improvements: 0.28% for *Ratio*, 7.45% for *CWC*, and 0.03% for *AIW*. It is also for this 0.85 *PINC* value that $MOPSO-WT$ and $GBR-WT$ perform much more closely than for larger *PINC* values: only 2%, 5%, and 1% improvements (*Ratio*, *CWC*, *AIW*, respectively) are observed for $MOPSO-WT$ vs. $GBR-WT$. In fact, only for long horizons (45 and 60 min) does $MOPSO-WT$ outperforms $GBR-WT$. For instance, improvements of 7% (ratio) can be observed for 60-minute horizons. Finally, the improvement of $MOPSO-WT$ vs. $QR-WT$ is very similar to the one observed for $PINC = 0.90$ (19% in *Ratio*, 55% in *CWC*, 15% in *AIW*).

6. Summary and conclusions

In the case of solar energy forecasting, it is known that some forecasting models provide good predictions for short-horizons (for instance Smart Persistence), while others work better for longer horizons (for instance WRF-Solar). Previous research has shown that the blending or integration of individual models is

able to improve forecasts by taking advantage of the complementary synergies of each model. In this work, the estimation of uncertainty by means of PI's, in the context of blending models, has been addressed. That is, individual models are blended so that the output of the blending is not a point forecast, but a PI. With this purpose, a MLP blending model has been used, optimized with a LUBE multi-objective approach, because previous works have shown LUBE to have a very good performance for estimating PI's in different domains. The use of a multi-objective approach is also very pertinent in this context because PI estimation is inherently multi-objective, and optimization algorithms are able to obtain a Pareto front of solutions in a single run. This front represents the optimal trade-offs between interval coverage and width, out of which the user can select the solution(s) most appropriate for the desired PINC(s).

In this work two approaches have been studied. First, the MLP uses as inputs the predictions of four forecasting models (Smart Persistence, Satellite, WRF-Solar and CIADcast) and the outputs are the lower and upper bounds of the PI. Second, information about four meteorological weather types is added to the MLP's inputs in order to study whether the knowledge of weather type can reduce the uncertainty of PI's. Both approaches have been empirically validated to estimate PI's for GHI at different forecast horizons (15, 30, 45, and 60 min). The study was conducted at a location in the south of the Iberian Peninsula (Seville) in the period March 2015 to February 2017.

The performance of the approaches has been measured using the *CWC*, *Ratio* and *AIW* metrics, that evaluate the quality of the intervals from different points of view. *CWC* measures mostly whether the coverage of the interval satisfies the target coverage; *AIW* is the average interval width; and the *Ratio* measures the trade-off between coverage and width. QR and GBR have been used as baseline methods. An exhaustive methodology has been applied, that includes using 4-fold cross-validation, running 30 times for each configuration, and statistical significance tests for three different target coverage values ($PINC = 0.85, 0.90, 0.95$).

Regarding the importance of supplying weather-type information, results on the three metrics have shown that WT helps MOPSO to obtain better PI's mainly for *PINC* values 0.95 and 0.9. For those *PINC* values, the use of WT enhances MOPSO between 2% and 3% the *Ratio*, 7%–8% the *CWC*, and 2%–3% the *AIW*. When the target *PINC* is 0.85, (a less demanding coverage), WT influence on the quality of PI's is much smaller.

With regard to baselines, it can be concluded that MOPSO outperforms the QR method for all forecasting horizons, all *PINC* values, and the three metrics. Depending on the *PINC* value, the magnitude of improvement is between 11% and 19% for *Ratio*, 46%–55% for *CWC*, and 9%–15% for *AIW*. MOPSO also outperforms GBR for all metrics and all forecasting horizons *PINC* values 0.95 and 0.90. 10%–44% for *Ratio*, 46%–49% for *CWC*, 9%–30% for *AIW*. For $PINC = 0.85$, the superiority of MOPSO is only observed for large horizons (45 and 60 min) and improvements are smaller (2%, 5%, and 1% for *Ratio*, *CWC*, and *AIW*, respectively).

Therefore, the blending of models by means of MLP's, optimized using MOPSO, and specially including weather-type information, is an advantageous alternative for estimating PI's. This advantage is larger for high-coverage PI's ($PINC = 0.95, 0.90$) than for less demanding probability coverage ($PINC = 0.85$). These conclusions have been validated using statistical significance tests.

Despite the good performance of MOPSO, it is necessary to remark that evolutionary optimization methods, such as MOPSO, are known to have high computational costs. In this sense, it would be interesting as future research in the context of model blending, to adapt efficient MLP optimization algorithms for PI estimation, so that comparable results can be obtained in less time.

CRediT authorship contribution statement

Inés M. Galván: Conceptualization, Methodology, Software, Writing - original draft, Investigation, Validation, Writing - review & editing. **Javier Huertas-Tato:** Conceptualization, Methodology, Software, Writing - original draft, Investigation, Validation, Writing - review & editing. **Francisco J. Rodríguez-Benítez:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Investigation, Writing - review & editing. **Clara Arbizu-Barrena:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Investigation, Writing - review & editing. **David Pozo-Vázquez:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Investigation, Writing - review & editing. **Ricardo Aler:** Conceptualization, Methodology, Software, Writing - original draft, Investigation, Validation, Writing - review & editing.

Declaration of competing interest

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

Acknowledgments

The authors are supported by projects funded by Agencia Estatal de Investigación, Spain (PID2019-107455RB-C21 and PID2019-107455RB-C22/AEI/10.13039/501100011033). Also supported by Spanish Ministry of Economy and Competitiveness, project ENE2014-56126-C2-1-R and ENE2014-56126-C2-2-R (http://prosol.uc3m.es). The University of Jaén team is also supported by FEDER, Spain funds and by the Junta de Andalucía, Spain (Research group TEP-220). The authors are in debt with the National Centers for Environmental Prediction (NCEP), EUMETSAT, Faculdade de Ciencias da Universidade de Lisboa, Grupo de Energía Solar of the Universidad Politécnica de Madrid and Abengoa Solar for providing the data used in this work.

References

[1] J. Dersch, M. Schroedter-Homscheidt, K. Gairaa, N. Hanrieder, T. Landelius, M. Lindskog, S. Müller, L. Ramirez Santigosa, T. Sirch, S. Wilbert, Impact of DNI nowcasting on annual revenues of CSP plants for a time of delivery based feed in tariff, Meteorol. Z. (2019).

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

[3] D.S. Renné, Emerging meteorological requirements to support high penetrations of variable renewable energy sources: solar energy, in: Weather Matters for Energy, Springer, 2014, pp. 257–273.

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

[5] S.E. Haupt, Short-range forecasting for energy, in: Weather & Climate Services for the Energy Industry, Palgrave Macmillan, Cham, 2018, pp. 97–107.

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

[7] A. Mellit, Artificial intelligence technique for modelling and forecasting of solar radiation data: a review, Int. J. Artif. Intell. Soft Comput. 1 (1) (2008) 52–76.

[8] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, Renew. Energy 105 (2017) 569–582.

[9] S. Belaid, A. Mellit, H. Boualit, M. Zaiani, Hourly global solar forecasting models based on a supervised machine learning algorithm and time series principle, Int. J. Ambient Energy (2020) 1–25.

[10] P. Mathiesen, J. Kleissl, Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States, Sol. Energy 85 (5) (2011) 967–977.

[11] A. Torres-Barrán, Á. Alonso, J.R. Dorronsoro, Regression tree ensembles for wind energy and solar radiation prediction, Neurocomputing 326 (2019) 151–160.
[12] B. Wolff, J. Kühnert, E. Lorenz, O. Kramer, D. Heinemann, Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data, Sol. Energy 135 (2016) 197–208.
[13] E. Lorenz, J. Kühnert, D. Heinemann, Short term forecasting of solar irradiance by combining satellite data and numerical weather predictions, in: Proceedings of the 27th European PV Solar Energy Conference (EU PVSEC), Frankfurt, Germany, Vol. 2428, 2012, p. 44014405.
[14] F.J. Rodríguez-Benítez, C. Arbizu-Barrena, J. Huertas-Tato, R. Aler-Mur, I. Galván-León, D. Pozo-Vázquez, A short-term solar radiation forecasting system for the iberian peninsula. Part 1: Models description and performance assessment, Sol. Energy 195 (2020) 396–412.
[15] A. Koca, H.F. Oztop, Y. Varol, G.O. Koca, Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey, Expert Syst. Appl. 38 (7) (2011) 8756–8762.
[16] J. Huertas-Tato, R. Aler, F. Rodríguez-Benítez, C. Arbizu-Barrena, D. Pozo-Vázquez, I.M. Galván, Predicting global irradiance combining forecasting models through machine learning, in: International Conference on Hybrid Artificial Intelligence Systems, Springer, 2018, pp. 622–633.
[17] J. Huertas-Tato, R. Aler, I. Galván, F.J. Rodríguez-Benítez, C. Arbizu-Barrena, D. Pozo-Vázquez, A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: Model blending approaches based on machine learning, Sol. Energy 195 (2020) 685–696.
[18] D.W. Van der Meer, J. Widén, J. Munkhammar, Review on probabilistic forecasting of photovoltaic power production and electricity consumption, Renew. Sustain. Energy Rev. 81 (2018) 1484–1512.
[19] R. Doherty, M. O'malley, A new approach to quantify reserve demand in systems with significant installed wind capacity, IEEE Trans. Power Syst. 20 (2) (2005) 587–595.
[20] E. Ela, A. Tuohy, R. Entriken, E. Lannoye, R. Philbrick, Using probabilistic renewable forecasts to determine reserve requirements, in: Solar Integration Workshop, 2017.
[21] P. Pinson, H. Madsen, Ensemble-based probabilistic forecasting at horns rev, Wind Energy: Int. J. Progress Appl. Wind Power Convers. Technol. 12 (2) (2009) 137–155.
[22] S. Alessandrini, F. Davò, S. Sperati, M. Benini, L. Delle Monache, Comparison of the economic impact of different wind power forecast systems for producers, Adv. Sci. Res. 11 (1) (2014) 49–53.
[23] S. Camal, A. Michiorri, G. Kariniotakis, Probabilistic forecasting and bidding strategy of ancillary services for aggregated renewable power plants, in: 6th International Conference Energy & Meteorology, 2019.
[24] P.P. Biswas, P.N. Suganthan, R. Mallipeddi, G.A. Amaratunga, Optimal reactive power dispatch with uncertainties in load demand and renewable energy sources adopting scenario-based approach, Appl. Soft Comput. 75 (2019) 616–632.
[25] A. Khosravi, S. Nahavandi, D. Creighton, A.F. Atiya, Lower upper bound estimation method for construction of neural network-based prediction intervals, IEEE Trans. Neural Netw. 22 (3) (2011) 337–346.
[26] C. Wan, Z. Xu, P. Pinson, Direct interval forecasting of wind power, IEEE Trans. Power Syst. 28 (4) (2013) 4877–4878.
[27] I.M. Galván, J.M. Valls, A. Cervantes, R. Aler, Multi-objective evolutionary optimization of prediction intervals for solar energy forecasting with neural networks, Inform. Sci. 418 (2017) 363–382.
[28] P.A. Jimenez, J.P. Hacker, J. Dudhia, S.E. Haupt, J.A. Ruiz-Arias, C.A. Gueymard, G. Thompson, T. Eidhammer, A. Deng, WRF-Solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction, Bull. Am. Meteorol. Soc. 97 (7) (2016) 1249–1264.
[29] C. Arbizu-Barrena, J.A. Ruiz-Arias, F.J. Rodríguez-Benítez, D. Pozo-Vázquez, J. Tovar-Pescador, Short-term solar radiation forecasting by advecting and diffusing MSG cloud index, Sol. Energy 155 (2017) 1092–1103.
[30] F.J. Rodríguez-Benítez, C. Arbizu-Barrena, F.J. Santos-Alamillos, J. Tovar-Pescador, D. Pozo-Vázquez, Analysis of the intra-day solar resource variability in the Iberian Peninsula, Sol. Energy 171 (2018) 374–387.
[31] J. Browell, D.R. Drew, K. Philippopoulos, Improved very short-term spatio-temporal wind forecasting using atmospheric regimes, Wind Energy 21 (11) (2018) 968–979.
[32] S. Lu, Y. Hwang, I. Khabibrakhmanov, F.J. Marianno, X. Shao, J. Zhang, B.-M. Hodge, H.F. Hamann, Machine learning based multi-physical-model blending for enhancing renewable energy forecast-improvement via situation dependent error correction, in: 2015 European Control Conference (ECC), IEEE, 2015, pp. 283–290.
[33] H.F. Hamann, A Multi-scale, Multi-Model, Machine-Learning Solar Forecasting Technology, Tech. rep., IBM, Yorktown Heights, NY (United States). Thomas J. Watson Research Center, 2017.
[34] R. Koenker, K.F. Hallock, Quantile regression, J. Econ. Perspect. 15 (4) (2001) 143–156.
[35] P. Lauret, M. David, H. Pedro, Probabilistic solar forecasting using quantile regression models, energies 10 (10) (2017) 1591.
[36] J.H. Friedman, Stochastic gradient boosting, Comput. Statist. Data Anal. 38 (4) (2002) 367–378.
[37] C.A. Coello, G. Toscano, M. Salazar, Handling multiple objectives with particle swarm optimization, IEEE Trans. Evol. Comput. 8 (3) (2004) 256–279.
[38] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-International Conference on Neural Networks, Vol. 4, IEEE, 1995, pp. 1942–1948.
[39] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, in: International Conference on Parallel Problem Solving from Nature, Springer, 2000, pp. 849–858.
[40] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, M. Sánchez-Girón, Daily global solar radiation prediction based on a hybrid coral reefs optimization–extreme learning machine approach, Sol. Energy 105 (2014) 91–98.
[41] J. Kennedy, R. Eberhart, Y. Shi, Swarm Intelligence, Morgan Kaufmann Publishers, San Francisco, 2001.
[42] K. Deb, Multi-Objective Optimization using Evolutionary Algorithms, Vol. 16, John Wiley & Sons, 2001.
[43] C. Rigollier, O. Bauer, L. Wald, On the clear sky model of the ESRA - European Solar Radiation Atlas - with respect to the Heliosat method, Sol. Energy 68 (1) (2000) 33–48.
[44] C. Rigollier, M. Lefèvre, L. Wald, The method Heliosat-2 for deriving shortwave solar radiation from satellite images, Sol. Energy 77 (2) (2004) 159–169.
[45] E. Zitzler, L. Thiele, Multiobjective optimization using evolutionary algorithms: a comparative case study, in: International Conference on Parallel Problem Solving from Nature, Springer, 1998, pp. 292–301.
[46] A. Khosravi, S. Nahavandi, D. Creighton, A.F. Atiya, Comprehensive review of neural network-based prediction intervals and new advances, IEEE Trans. Neural Netw. 22 (9) (2011) 1341–1356.
[47] F. Wilcoxon, Individual comparisons by ranking methods, in: Breakthroughs in Statistics, Springer, 1992, pp. 196–202.