EPJ Plus

EPJ.org
your physics journal

Eur. Phys. J. Plus (2019) **134**: 594 DOI 10.1140/epjp/i2019-12966-5

Multimodel ensemble approach for hourly global solar irradiation forecasting

Nahed Zemouri, Hassen Bouzgou and Christian A. Gueymard

Società Italiana di Fisica

Springer

Eur. Phys. J. Plus (2019) **134**: 594
DOI 10.1140/epjp/i2019-12966-5

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

Multimodel ensemble approach for hourly global solar irradiation forecasting

Nahed Zemouri[1], Hassen Bouzgou[2,a], and Christian A. Gueymard[3]

[1] Department of Electronics, Faculty of Technology, University of M'sila, M'sila, Algeria
[2] Department of Industrial Engineering, Faculty of Technology, University of Batna 2 (Mostefa Ben Boulaïd), Batna, Algeria
[3] Solar Consulting Services, Colebrook, NH, USA

Received: 10 May 2019 / Revised: 27 August 2019
Published online: 2 December 2019

Abstract. This contribution proposes a novel solar time series forecasting approach based on multimodel statistical ensembles to predict global horizontal irradiance (GHI) in short-term horizons (up to 1 hour ahead). The goal of the proposed methodology is to exploit the diversity of a set of dissimilar predictors with the purpose of increasing the accuracy of the forecasting process. The performance of a specific multimodel ensemble forecast showing an improved forecast skill is demonstrated and compared to a variety of individual single models. The proposed system can be applied in two distinct ways. The first one incorporates the forecasts acquired from the different forecasting models constituting the ensemble via a linear combination (combination-based). The other one consists of a novel methodology that delivers as output the forecast provided by the specific model (involved in the ensemble) that delivers the maximum precision in the zone of the variable space connected with the considered GHI time series (selection-based approach). This forecasting model is issued from an appropriate division of the variable space. The efficiency of the proposed methodology has been evaluated using high-quality measurements carried out at 1 min intervals at four radiometric sites representing widely different radiative climates (Arid, Temperate, Tropical, and High Albedo). The obtained results emphasize that, at all sites, the proposed multi-model ensemble is able to increase the accuracy of the forecasting process using the different combination approaches, with a significant performance improvement when using the classification strategy.

1 Introduction

Forecasting solar irradiance has become a key step to correctly predict the performance of solar-energy conversion system and to guarantee the stability and safety of the electrical grid. Electric utilities need to control the generation systems based on foreknowledge of the expected load and generation capacity over the next minutes, hours and days, since some power generators can be started rapidly whereas others need plenty of time before being ready. Moreover, the price of electricity production is traded on a constantly and rapidly evolving market. This explains why accurate forecasts of electricity production and consumption are needed to lower electricity costs, which can be expected to ultimately benefit customers.

The normal procedure of predicting electrical power produced by a solar conversion system is based on the available solar irradiance, as well as several aspects that concern the system performance and related environmental issues (*e.g.*, soiling) at the time of interest. Using generators based on photovoltaic (PV) conversion technologies, the electricity produced is, in principle at least, linearly related to the incident solar irradiance. Hence, using an appropriate PV energy simulation tool, the prediction of power production is directly related to that of the incident irradiance. The latter is highly variable over time for meteorological reasons, however, which makes its exact prediction difficult. From a more general standpoint, forecasting solar irradiance with good accuracy can be expected to increase the effectiveness and value of the power output of any type of solar energy system.

In the literature, many methods have been proposed to forecast solar radiation, and —most importantly for the present applications— global horizontal irradiance (GHI). Based on detailed reviews [1–6], they can be classified into

[a] e-mail: bouzgou@gmail.com (corresponding author)

four main categories: i) Statistical and Artificial Intelligence (AI) models; ii) Remote-sensing models using ground-based or satellite-based cloud imagery; iii) Numerical weather prediction models; iv) Hybrid models, using two or more of the above techniques. The present investigation focuses on the latter type of technique, using an ensemble of various models to improve the overall forecasting skill of each of them. For that purpose, solar radiation data can be assimilated as a time series created by a stochastic process. Its prediction would thus be possible through appropriate mathematical modeling of the original stochastic process. Accordingly, this forecasting process can be mathematically understood as the conditional expectation of the historical data records and the related model. Among the many statistical models that have been proposed in the literature, the focus of this study is on those using traditional AI techniques and/or hybrid ones.

Mellit *et al.* [7] presented an adaptive wavelet-network to predict GHI on a daily basis. Hocaoğlu *et al.* [8] proposed a two-dimensional (2D) model of the hourly GHI using image-processing methods. Various filter-tap configurations were assessed, using feed-forward neural networks (NN), to optimize nine different linear filters. Cao and Lin [9] presented a diagonal recurrent wavelet neural network (DRWNN) to forecast GHI. Mellit and Pavan [10] investigated a multilayer perceptron (MLP) model to predict 24-hour-ahead solar irradiance using daily-mean GHI and air temperature. In parallel, Martín *et al.* [11] proposed a comparison study of several models (NN, autoregressive, and fuzzy logic) to predict half-daily values of GHI.

With respect to hybrid approaches, Benmouiza and Cheknane [12] proposed a combination of k-means clustering and ANN. The k-means algorithm extracts features from the data, which are then used to model the time series. Next, a nonlinear autoregressive (NAR) NN is applied to make forecasts. Salcedo-Sanz *et al.* [13] presented a time-based composite covariance optimized by a Gaussian process regression to blend multiple meteorological variables and forecast GHI in Spain. This innovative approach combines purely statistical methods to more conventional meteorological information. Benmouiza and Cheknane [14] investigated a hybrid method combining ARMA and TDNN models. The ARMA part is used to capture the linear component, whereas TDNN is used for the nonlinear one. Akarslan and Hocaoğlu [15] proposed two seasonal adaptive approaches to forecast GHI. The first one uses linear predictions for the winter and summer seasons, and an empirical method for the spring and fall seasons. In the second one, the clearness index is used to choose between the linear and empirical models.

Monjoly *et al.* [16] investigated various techniques of multi-scale decomposition of clear-sky index data, such as Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD) or Wavelet Decomposition (WD), combined with traditional forecasting models such as autoregressive processes or ANN. Bouzgou and Gueymard [17] proposed a new variable selection approach based on the minimum redundancy maximum relevance principle, with Extreme Learning Machine (ELM) to forecast GHI for three typical horizons (15 min, 1 hour, and 1 day ahead). That approach outperformed the other scenarios in terms of forecasting performance. In subsequent work [18], the authors proposed a wrapper mutual information methodology for short-term GHI forecasting. The proposed methodology outperformed the other tested techniques in terms of both accuracy and computational time. More recently, Cornejo-Bueno *et al.* [19], proposed several Machine Learning regression techniques to forecast GHI from geostationary satellite data. Different types of neural networks, support vector regression and Gaussian processes were selected as regression techniques to be evaluated. Their results showed the capability of machine learning regressors to achieve better forecasts using satellite measurements.

In recent years, ensemble learning techniques have also been used to improve solar radiation forecasting [20–22]. In [23], a multi-stage ANN is proposed. A first-stage neural network predicts the average atmospheric pressure of the next day from atmospheric pressure data of the previous day. A second-stage neural network predicts the insolation level of the next day from the average atmospheric pressure and weather data of the previous day. A third-stage neural network predicts the insolation of the next day from the insolation level and weather data of the previous day. The results obtained with the multi-stage and the single-stage neural networks are compared with the measured ones. In [24], the authors proposed an analog ensemble (AnEn) approach to generate probabilistic solar power forecasts. The AnEn was based on an historical set of deterministic numerical weather prediction (NWP) model forecasts and observations of the solar power. Similarly, the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble Prediction System (EPS) was applied to produce short-term probabilistic solar power forecasts (SPF) [25]. The EPS is based on the application of a meteorological model multiple times, starting from slightly perturbed initial conditions. The distribution of these different runs permits to estimate the forecasting uncertainty.

Hall and Hall [26] proposed a competitive two-model ensemble forecasting method: the first model is based on non-linear regressions describing the potential relationships between various meteorological variables, whereas the second one is based on a pattern recognition technique. The outcomes of these two models are combined and weighted to forecast solar radiation, either 1-hour or 3-hour ahead. Chaouachi *et al.* [27] proposed a bagging ANN model to forecast the short-term solar irradiance. The bagging ANN model is composed of three kinds of ANN: MLP, radial basis function (RBF), and recurrent neural network (RNN), in addition to historical data after bootstrapping. The performance of the bagging ANN models was found better than those of single models. Zemouri and Bouzgou [28] proposed a set of time series forecasting techniques based on the combination of Support Vector Regression methods to predict GHI in Algeria.

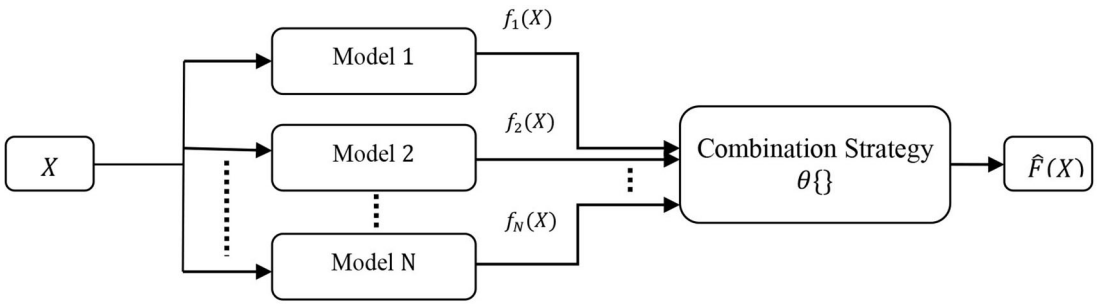

Fig. 1. Block diagram of the proposed forecasting system.

The literature just reviewed undoubtedly shows that many stochastic and AI methods are possible and can be combined efficiently. Although individual methods are used extensively, their "intrinsic force" is not inevitably the greatest possible option. The new methodology presented in what follows tries to improve forecasts by dynamically selecting the most suitable predictor for each time series realization using a classification approach.

The principle of combining different forecasts, or "ensemble forecasting", is inspired by the fact that the best possible models of real-world data generation processes cannot be precisely identified. Hence, the selection of a single existing model is a "hit and miss" proposition, particularly if the statistical properties of the time series or the performance of models vary over time [29]. It can be considered that a trustworthy method would be one that decreases model risk and seeks to improve the forecast accuracy by taking advantage of the different strengths of various potentially effective models, while compensating for their limitations.

The output of the novel approach proposed in this study is the forecast produced by the learning model distinguished by the maximum performance in the region of the *variable space*. The latter is also called "observation space" because each historical value of the GHI series (dimension/coordinate) corresponds to an observation of the historical recorded GHI. In this contribution, various ensemble types are evaluated using several statistical and/or AI core models.

The remainder of this paper is organized as follows: The next section gives a detailed description of the proposed multimodel ensemble system with a small recall of the forecasting techniques included in the ensemble. The solar radiation database used to assess the different approaches is presented in sect. 3. Section 4 presents the metrics used to evaluate the different methodologies. Finally, the experimental assessment of the proposed technique and the comparison with other approaches are described in sect. 5.

2 Methodology

Consider a set of N forecasting models $f_i(X)$ $(i = 1, 2, \ldots, N)$ trained independently on the available training data. The problem, schematized in fig. 1, is to define a combination method $\theta\{\}$, such that the resulting forecasts obtained after the combination of the diverse single models for a particular unseen data point (*i.e.*, not included in the training set) can be described as

$$\hat{F}(X) = \theta\{f_1(X), f_2(X), \ldots, f_N(X)\}. \tag{1}$$

2.1 Ensemble modeling

Ensemble methods are learning algorithms that construct a set of models and then classify/predict the new data points by taking a weighted combination of their predictions [29,30]. An Ensemble of forecasting techniques necessitates the designation of a set of different forecasting models. It can be constructed through the same forecasting model with different parameter settings or according to different learning methods. From another perspective, Ensemble methods are meta-algorithms that combine numerous machine-learning methods into one model with the goal of either reducing the variance (bagging) or the bias (boosting), or improving predictions (stacking). The existing methods can be divided into two groups:

- *Sequential ensemble methods* in which base l earners are created consecutively (*e.g.*, AdaBoost). The goal here is to exploit the dependence between different base learners. The general performance can be improved by pondering the bad examples (wrong outputs) with heavy weights, as conventional in boosting methods.
- *Parallel ensemble methods* in which the base learners are created in parallel (*e.g.*, Random Forest). The goal here is to exploit the potential independence between all base learners since the error can be strongly reduced by averaging.

Usually, a single base-learning algorithm can be used to produce homogeneous base models from models of the same type, leading to homogeneous ensembles. Alternatively, heterogeneous models, *i.e.*, models of diverse types, lead to heterogeneous ensembles. With the objective of being more precise than any of its individual single members, the single models in the Ensemble have to be as precise as possible and as diverse as possible [30]. Ensemble diversity, which is the actual difference between single models, is a central issue in ensemble theory. It is obvious that, to improve the combination process, the single models must be dissimilar. Otherwise, there would be no performance enhancement if duplicate single models were combined.

This study contemplates only the second approach, in which *heterogeneous modeling* is used. This consideration is supported by the interest in examining the efficiency of the approach for the forecasting of GHI over the short term (from 1 min to 1 hour ahead). At such high refresh rates, the exogenous variables that could be used to predict GHI with some degree of reliability (*e.g.*, cloud fraction, precipitable water, or aerosol optical depth) are typically not available from local observations or from gridded reanalysis modeled data (whose spatial resolution is too coarse anyway).

As mentioned above, several combinational approaches exist, and are designed according to two different strategies in what follows. In the first one, referred to as "combination-based", the final forecast is obtained after combining the forecasts achieved by the different single models. The other strategy, referred to as "selection-based", consists of a novel method (detailed below) that derives the final forecast by selecting the output (prediction) of the "best individual model" using a methodology that defines a proper partition of the variable space. In the present study, three families of models are used to construct the optimal ensemble, namely: i) Autoregressive-based models; ii) ANN-based models and; iii) Kernel-based models.

2.2 Individual methods

2.2.1 Autoregressive-based models

ARIMA models [31] constitute the origin of the majority of statistical forecasting techniques. An ARIMA(p, d, q) —usually called a Box-Jenkins model— supposes that every future value of a time series can be expressed as a linear function of some past observations, in addition to some white-noise terms. It utilizes a common differencing procedure of order d to make the time series stationary, and decomposes it into different models as follows:

– Autoregressive process, AR:

A stochastic process Y_t with real values is considered an $AR(p)$ process of order p if

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t. \tag{2}$$

The $AR(p)$ value at time t is then estimated using p past values and a random term ε_t.

– Moving Average process, MA:

A stochastic process Y_t is said to be an MA process of order q, denoted $MA(q)$, if there exists q coefficients $b_1, b_2, \ldots, b_q$, and a white noise ε_t, such that

$$y_t = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}. \tag{3}$$

The value of $MA(q)$ process at time t is therefore regressed on its own past errors.

– Autoregressive Moving Average, ARMA:

$AR(p)$ processes are special cases of ARMA. A stochastic process Y_t is considered an ARMA process of order p, q if

$$y_t = \varepsilon_t + (a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p}) + (b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \ldots + b_q \varepsilon_{t-q}). \tag{4}$$

This can also be formulated as $\phi(z) Y_t = \Theta(z) \varepsilon_t$, where $\phi(z) = 1 + a_1 z + a_2 z^2 + \ldots + a_p z^p$ and $\Theta(z) = 1 + b_1 z + b_2 z^2 + \ldots + b_q z^q$ are the polynomial terms of the AR and MA components of the ARMA(p, q) process Y_t, and z is the back-shift operator.

– Autoregressive Integrated Moving Average, ARIMA:

ARIMA is an expansion of the ARMA class of models where a differencing operation is added. A process Y_t is considered an ARIMA(p, d, q) if $(1 - z)^d Y_t$ is a causal ARMA(p, q). The ARIMA model can be formalized as

$$\phi(z)(1 - z)^d Y_{t_t} = \Theta(z) \varepsilon_t. \tag{5}$$

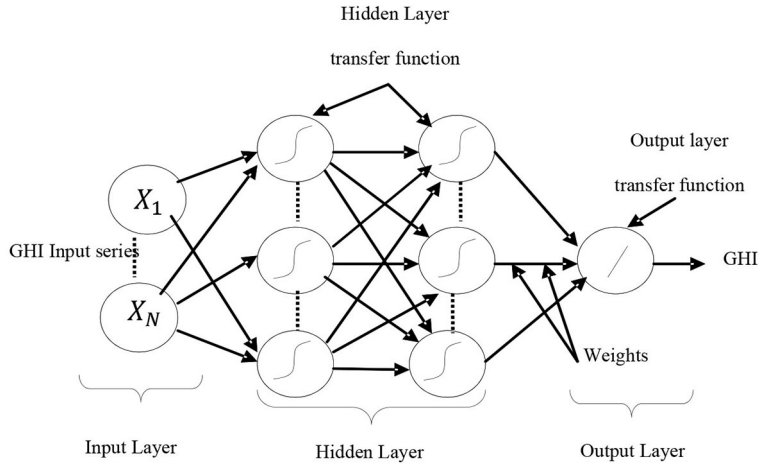

Fig. 2. Architecture of a multi-layer perceptron neural network.

– Seasonal Autoregressive Integrated Moving Average, SARIMA:

The SARIMA model can be described as follows:

$$\phi_p(z)\Phi_P\left(z^s\right)(1-z)^d(1-z^s)^D Y_t = \theta_q(z)\Theta_Q(z^s)\varepsilon_t, \tag{6}$$

where $\phi_p(z)$, $\Phi_P(z^s)$, $\theta_q(z)$ and $\Theta_Q(z^s)$ are the polynomial terms of orders p, P, q and Q, respectively, and d and D are the orders of non-seasonal and seasonal differencing, respectively.

2.2.2 Artificial neural network-based models

ANNs have been created to surmount the limitations experienced with more traditional methods when trying to solve complex problems. The ANN learns from certain examples by building an input-output function aimed at carrying out predictions. NNs contain an interconnection of several neurons. In the literature, several architectures of ANNs with different learning algorithms have been proposed [32]. A learning algorithm is essentially a procedure whose function is to regulate the weights of a given network. For a given learning dataset, this can be done by minimizing an objective function between the desired and estimated target function. The present study focuses on four types of ANNs, namely: MLP, RBF, ELM, and Elman Neural Network (ENN).

– Multi-layer perceptron neural networks, MLP:
An MLP is composed of several highly interconnected units (neurons) structured in layers and functioning in parallel with feed-forward information flow, as illustrated in fig. 2. The signals flow successively throughout the different layers from the input to the output layer. The layers between the input and the output layers are called hidden layers. In each layer, each unit computes a scalar product between the weights vector and the output vector issued from the preceding layer. Next, a transfer function is applied to the previous outcome to create an input for the following layer [32]. A general transfer function for the hidden layers is the sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}}\,. \tag{7}$$

– Radial basis functions neural networks, RBF:

An RBF neural network is composed of three layers (an input, a hidden layer, and an output layer). As illustrated in fig. 3, input neurons (*i.e.*, input variables) just transmit input variables X_j to the following layer. Each neuron in the hidden layer is characterized by a kernel function φ_j (typically a Gaussian function) described by a width σ_j and a centre c_j, such that

$$\varphi_j\left(\|y-c_j\|\right) = \exp\left(-\frac{1}{2}\left(\frac{\|y-c_j\|}{\sigma_j}\right)^2\right). \tag{8}$$

The output layer contains only one neuron providing the predicted value. The output function can be written as:

$$f(y) = \sum_{j=1}^{m} \lambda_j \varphi_j\left(\|y-c_j\|\right) \tag{9}$$

where λ_j and m are the weight and the number of RBF functions, respectively [33,34].

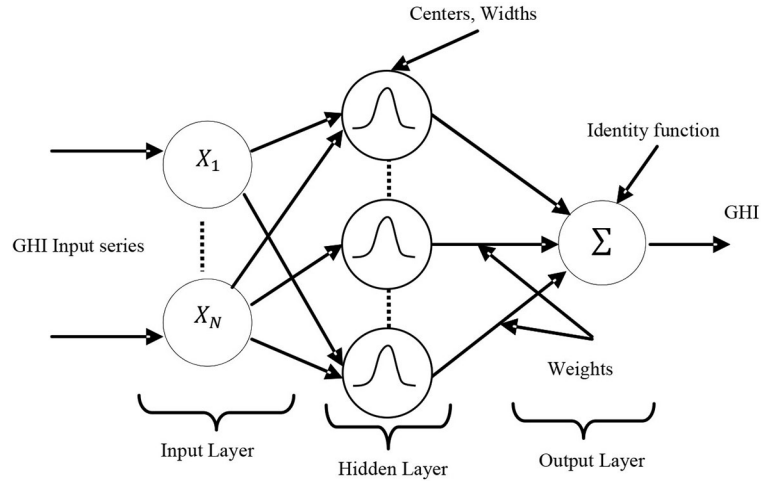

Fig. 3. Architecture of a Radial Basis Function Neural Network (RBFNN).

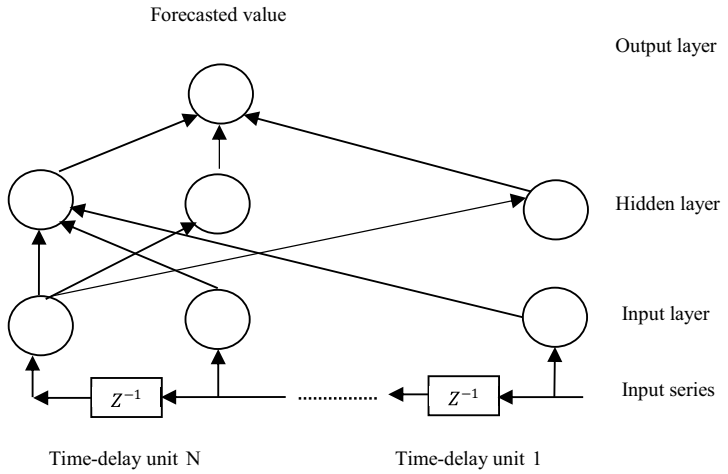

Fig. 4. Elman neural network architecture.

– Elman Neural Networks, ENN:

An ENN is a typical kind of dynamic neural network, based essentially on the structure of classical feed-forward artificial neural networks. It stores an internal state to elaborate the function of its dynamic characteristics and provide the system with the capability to adapt for any time-varying characteristic. An ENN can thus directly reflect the dynamic characteristics of a system, which is an important characteristic of an advanced forecasting tool. The ENN structure is illustrated in fig. 4.

Generally, ENNs are divided into four layers: input, hidden, undertake, and output layer. The input, hidden, and output layers are similar to those in conventional ANNs. The input layer nodes act simply as signal transmission units. The output layer node acts as a summation function, whereas the undertake layer (or "state layer") is utilized to remember the preceding output values of the hidden layer units that have to be returned back to the input of the network.

The ENN's storage capacity makes it more sensitive to historical data, allowing the internal feedback to enhance the network aptitude of dealing with dynamic variations in the data. In addition, an ENN can approximate any nonlinear function with any degree of accuracy and can avoid the impact of external noise. The mathematical equations in the nonlinear state space of an ENN can be written as

$$y(x) = g\left(\omega^3 x(x)\right), \tag{10}$$

$$x(k) = f\left(\omega^1 x_c(k) + \omega^2(u(k-1))\right), \tag{11}$$

$$x_c(k) = x(k-1), \tag{12}$$

where y is the output vector with dimension equal to M; x is the hidden layer node vector with M dimensions; u is the input vector with r dimensions; x is the feedback state vector with N dimensions; ω^3 are the weights linking the hidden layer with the output layer; g is the transfer function of the output node; f is the transfer function of the hidden layer nodes.
Generally, ENNs also use the back-propagation algorithm to determine the network weights.

– Extreme Learning Machines, ELM:

The ELM model has been proposed by Huang *et al.* [35]. This kind of Single-Layer Feedforward Neural Network (SLFN) can be utilized for both regression and classification problems. Suppose an ELM with Q hidden neurons and activation function $g(x)$ that is designed to learn N different samples (X_i, t_i), where $X_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$. The input weights and hidden biases are chosen randomly. Hence, the nonlinear system can be converted to a linear system as follows:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{13}$$

where $\mathbf{H} = \{h_{ij}\}$ $(i = 1, \ldots, N$ and $j = 1, \ldots, G)$ is the output matrix of the single hidden layer, $h_{ij} = g(W_j * Y_i + b_j)$ denotes the output of the j-th hidden neuron with respect to X_i; $W_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]^T$ is the weight vector between the j-th hidden neuron and the input neurons, and b_j is the bias of the j-th hidden neuron; $W_j * Y_i$ is the inner product of W_j and X_i; $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^T$ is the matrix of output weights, and $\beta_j = [\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}]^T$ $(j = 1, \ldots, K)$ denotes the weight vector connecting the j-th hidden neuron to the output neurons. Finally, $\mathbf{T} = [t_1, t_2, \ldots, t_N]^T$ is the matrix of targets (*i.e.*, desired output). Therefore, the output weights are determined by finding the least-squares solution of the linear system described by eq. (13), such that

$$\widehat{\boldsymbol{\beta}} = \mathbf{H}^{+}\mathbf{T} \tag{14}$$

where $\mathbf{H}^{+}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{H}$. Usually, ELMs tend to result in good generalization performance with an increased learning speed [28,36].

2.2.3 Kernel-based models

– Support Vector Machine, SVM:

SVMs are supervised non-parametric techniques, and hence do not require any assumption about data distribution. The goal of an SVM regression approach is to define a smooth function $f(x)$ that has (as a maximum) ε deviation from the learning samples y_i, which are extracted from a solar irradiance time series in the present case [37,38]. This can be achieved by mapping the original series from the input d-dimensional space into a transformed variable space, *i.e.*, $\Gamma(x) \in R^{d'}(d' > d)$, and fixing a linear estimate as follows:

$$f(x) = \omega^* * \Gamma(x) + b^*. \tag{15}$$

The linear function in the higher dimensional variable space is obtained by optimizing a cost function, whose objective is to minimize two criteria (Euclidean norm and error minimization), and is expressed as

$$\Psi(\omega, \xi) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right), \tag{16}$$

where C is a regularization parameter. That function is subject to the constraints below:

$$\begin{cases} y_i - (\omega \cdot \Gamma(x_i) + b) \leq \varepsilon + \xi_i, \\ (\omega \cdot \Gamma(x_i) + b) - y_i \leq \varepsilon + \xi_i^*, \end{cases} \quad \xi_i,\ \xi_i^* \geq 0 \text{ and } i \in [1, N], \tag{17}$$

where ξ_i and ξ_i^* are slack variables linked to samples that do not fit inside the ε-deviation tube. The formulation of the error function is designed to deal with an ε-insensitive loss function $|\gamma|_\varepsilon$ defined as

$$\begin{cases} 0, & \text{if } |\tau| \leq \varepsilon, \\ |\tau| - \varepsilon, & \text{otherwise}, \end{cases} \tag{18}$$

where τ is the deviation with regard to the desired output. This is equivalent to saying that the differences between the measured and the predicted values are accepted inside the ε-tube (error less than ε), whereas a linear penalty is imposed to the points outside the ε-insensitive tube. An example of linear SVM is shown in fig. 5.

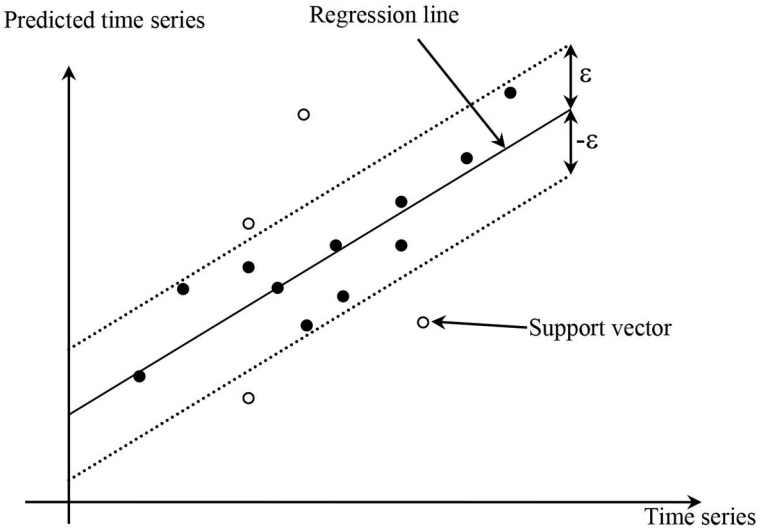

Fig. 5. Linear SVM regression with ε-tube.

The result of the optimization problem is a function represented in the input (original) variable space as

$$f(x) = \sum_{i \in S} (\alpha_i - \alpha_i^*) K(x_i, x) + b^*, \tag{19}$$

where $K(\cdot, \cdot)$ is a kernel function. S is the subset of indices $(i = 1, 2, \ldots, N)$ corresponding to the nonzero Lagrange multiplier α_i or α_i^*. According to their importance in determining a solution, the Lagrange multipliers evaluate each training point. Those training points with nonzero weights are called support vectors.

2.3 Combination strategies

A combination strategy tries to improve the global forecasting performance by combining various approaches. It attempts to increase the forecasting accuracy through a practical solution, but at the expense of additional computational complexity. Therefore, one should have good understanding of the properties of all ensemble models and should prudently use the accessible information with the intention of developing effective combination methodologies [39]. After generating a set of base learners, and instead of trying to determine the best single model, ensemble methods select a combination to get the best generalization capability, where the combination method has an important role. Dietterich [40] attributed the benefit obtained with the combination process to the following three fundamental classes of factors:

- *Statistical*: usually the hypothesis space is too large to discover small training datasets, whereas other different hypotheses can give the same precision on the training data. If the learning model selects one space among these hypotheses, there is a risk that a wrongly chosen hypothesis does not correctly predict the future data. Accordingly, by combining hypotheses, the risk of selecting an incorrect hypothesis can be reduced.
- *Computational*: several learning models make different types of local search that can be stuck in local optima. Although there is sufficient training data, it remains very hard to discover the best hypothesis. The combination of local searches launched from several distinct starting points can deliver a better estimate of the true unknown hypothesis. In this case, the combination of hypotheses reduces the risk of choosing a wrong local minimum.
- *Representational*: In several machine-learning problems, the right unknown hypothesis cannot be characterized. In that case, using a combination of hypotheses allows an expansion of the representable space of functions. Hence, the learning model can form a more precise estimate of the true unknown hypothesis.

The three classes of factors just described typically explain why the classical learning techniques tend to fail [30]. The rest of this section provides a short summary of the techniques used in this study for linear combination of multiple forecasts.

2.3.1 Simple average, SA

The easiest combination method is the simple average, where all individual forecasts have the same weight ($w_j = 1/n$). Several studies have revealed that, curiously, the simple average often improves the overall forecasting accuracy, even when some individual forecasts appear as outliers or extreme values [41,42].

2.3.2 Error-Based method, EB

Here, the time series to be forecasted is divided into two complementary subsets, namely the training and the validation sets. The individual models are trained on the training set and their resulting forecast errors (using the validation set) are recorded. The weight of the combination for each individual forecast is then calculated as inversely proportional to the forecast error of the corresponding model, *i.e.*,

$$W_j = \frac{e_j^{-1}}{\sum_{j=1}^{n} e_j^{-1}}\,. \tag{20}$$

2.3.3 Least-Squares Regression, LSR

In this method, the weights of a linear combination are assigned by minimizing the Sum of Squared Error (SSE) between the forecasted and measured data. The classic formulation of the linear combination can be expressed as follows:

$$\hat{Y} = UW, \tag{21}$$

where

$$\begin{aligned} U &= \left[\hat{y}_1^{(1)}\hat{y}_1^{(2)}\ldots\hat{y}_1^{(n)}; \hat{y}_2^{(1)}\hat{y}_2^{(2)}\ldots\hat{y}_2^{(n)}; \ldots; \hat{y}_M^{(1)}\hat{y}_M^{(2)}\ldots\hat{y}_M^{(n)}\right]^T, \\ W &= [w_1, w_2, \ldots, w_n]^T, \end{aligned} \tag{22}$$

where $\hat{Y}$ and Y are the predicted and measured series, respectively, and W is populated with the weights of the individual models. The forecast SSE can then be expressed as

$$\begin{aligned} \mathrm{SSE} &= \sum_{t=1}^{M}(y_t - \hat{y}_t)^2 \\ &= (Y - UW)^T(Y - UW) \\ &= Y^TY - 2W^TU^TY + W^TU^TUW. \end{aligned} \tag{23}$$

By minimizing SSE with respect to W, the weight vector is derived as $W = (U^TU)^{-1}U^TY$.

2.3.4 Differential Weighting method, DW

Similarly to the LSR method, the weights of a linear combination of forecasts are obtained by minimizing the variance of the combined forecast error [29]. In fact, this approach necessitates knowledge of the covariance matrix of forecast errors, which is unknown in practice. The weights are calculated with

$$W_i = \frac{\sum_{s=t-v}^{t-1}(e_s^{(i)})^2}{\sum_{j=1}^{n}\sum_{s=t-v}^{t-1}(e_s^{(j)})^2}\,, \tag{24}$$

where n is the number of individual models, t is the time period of forecast, w_i is the weight assigned to the i-th method based on the data's preceding period $(t-1)$, v is a constant, and e_t^i is the percentage forecast error at time t, which is defined as

$$e_t^{(i)} = \frac{y_t - \hat{y}_t^{(i)}}{y_t} \quad \forall\ i = 1, 2, \ldots, n. \tag{25}$$

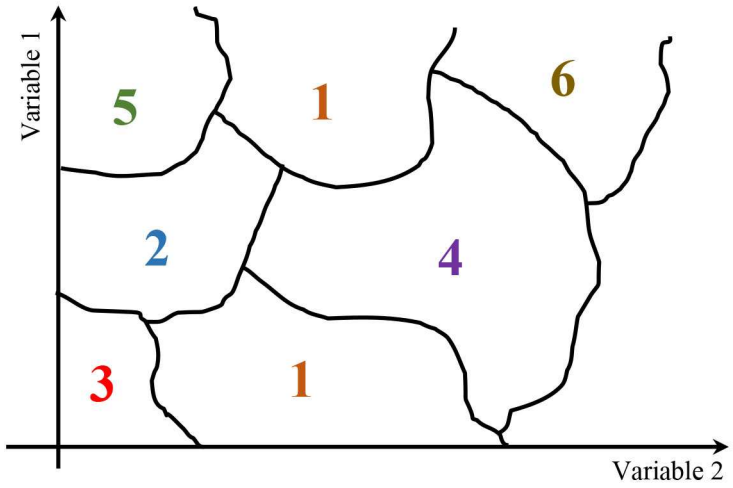

Fig. 6. Schematic partition of a two-dimensional variable space for an example composed of six models. Each region designates the individual forecasting algorithm that delivers the optimal forecasting accuracy value in the corresponding zone of the time series' variable space.

2.3.5 Outperformance method, OP

Proposed by Lemke and Gabrys [43], OP adopts a Bayesian framework of subjective probabilities to assign proper weights to the set of individual forecasts. The weight of an individual forecasting model is determined from the number of past times it outperformed the other models. To simplify, the combination of two individual models F1 and F2 is done as follows:

$$\hat{y}_k = w\hat{y}_k^{(1)} + (1-w)\hat{y}_k^{(2)}, \quad \forall\ k = 1, 2, \ldots, N. \tag{26}$$

In the present case, it is assumed that the performance of the two models is evaluated using some absolute error and is recorded M times. With this scenario, it is reasonable to accept that the model with a smaller absolute error outperforms the other models. If the number of times F_1 outperformed F_2 is a fraction k of M, it can be assumed that k follows a beta distribution $B(k|a_1, a_2)$ $(0 \leq k \leq 1; a_1, a_2 > 0)$ [43]. For every forecast realization i, the Bernoulli variable δ_i is expressed as

$$\left.\begin{aligned} \delta_i &= 1, \quad \text{if } F_1 \text{ outperforms } F_2, \\ &= 0, \quad \text{otherwise,} \end{aligned}\right\}. \tag{27}$$

The overhead framework can be extended to n models by assuming that the outperformance fractions k_i $(i = 1, 2, \ldots, n)$ of the individual models follow the n-parameters Dirichlet distribution, which is equivalent to the beta distribution.

2.4 Classification-based strategy

This strategy is motivated by the idea that diverse high-performance predictors can be connected with different zones of the input variable space of the time series. The best individual model is the one that minimizes the error in different portions of the variable space. The approach here consists in making a partition of the variable space, where each point in the variable space (single time series) is linked with the predictor of the ensemble that achieves the minimum forecasting error. As illustrated in fig. 6, the training step includes the identification (among the set of existing individual forecast models) of the best individual predictor $\hat{P}(X)$ for each series S of the variable space. The optimal division of the variable space into a set of regions is acquired using the analysis of the training data S_i $(i = 1, 2, \ldots, M)$, each being attributed to a single predictor. Hence, the goal of the selection element is to search for the best possible way to obtain the most appropriate partition of the variable space, so that

$$\hat{P}(X) = \underset{j=1,2,\ldots,N}{\operatorname{argmin}} \left\{ \left| f_j\left(S_i\right) - y_i \right| \right\}. \tag{28}$$

Nevertheless, this would not modify the achieved partition because the result of the "argmin" operator is not influenced by monotonic functions in the forecasting phase. Each unknown series $S \in \Re^d$ is introduced as input to the forecasting model. Its output is then the prediction $\hat{P}(X) \in \{1, 2, \ldots, N\}$ of the most precise predictor for the considered series. The forecasted value $\hat{F}(X)$ provided by various predictors can be expressed as

$$F(X) = f_{\hat{P}(x)}(X). \tag{29}$$

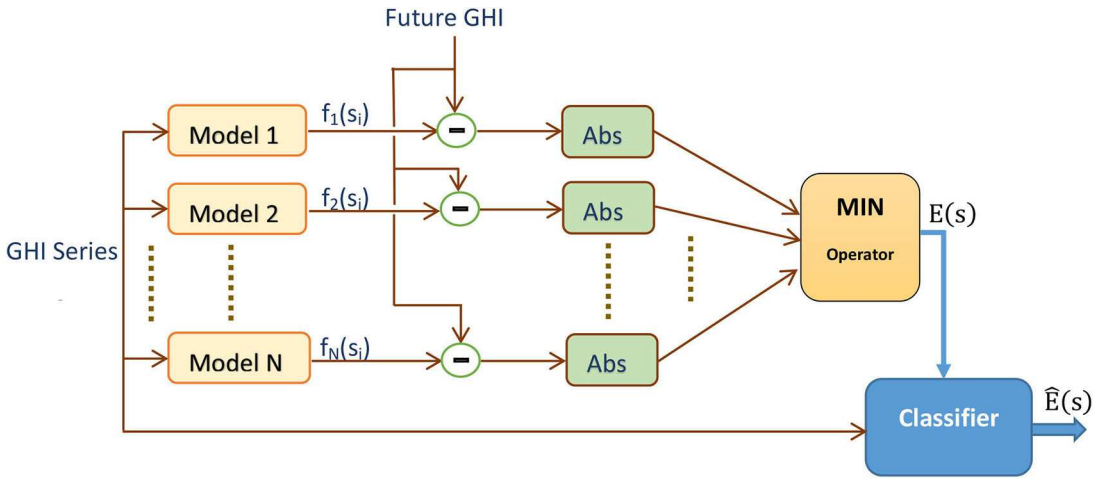

Fig. 7. Block diagram of the training process using the classification-based approach.

Table 1. Information on the four test stations used in this study. The mean and maximum measured GHI values are given in W/m^2. Key to acronyms: BSRN —Baseline Surface Radiation Network; NREL —National Renewable Energy Laboratory.

Site	Source	Elevation [m]	Longitude [°]	Latitude [°]	Max_GHI [W/m^2]	Mean_GHI [W/m^2]	Period	Climate
Alice Springs	BSRN	547	133.88	−23.79	1260.7	543.2	2009-2010	Arid
Brasilia	BSRN	1023	−47.71	−15.60	1335.1	475.6	2006–2007	Tropical
Golden	NREL	1829	−105.18	39.74	1209.3	459.4	2007–2008	Temperate
Ny-Ålesund	BSRN	11	11.93	78.92	899.0	184.3	2010–2011	High Albedo

From an operational standpoint, the classification approach can be used according to the utilization of the information enclosed in the accessible GHI records. This involves two parts: the training and the test. As is illustrated in fig. 7, the training phase consists of the determination of the most likely model for each point in the variable space. The best partition of the variable space (constituted of a set of zones) is achieved through the analysis of the training series, each being attributed to a single forecast model. The goal of the selection approach is to build the optimal partition in the best probable manner. Here, the optimality can be formulated in terms of forecasting error. In the present case, the Minimum Absolute Error (MAE) is selected for that task.

In the test phase, each series (for which the future GHI value is unknown) is provided as input to the classifier, which outputs the forecasted value of the best accurate model for this series [34,44].

3 Sources of data

A high-quality database of 1 min irradiance measurements from four radiometric stations that observe irradiance components (including GHI) with thermopile radiometers has been discussed [45]. The latter reference provides detailed information on these stations and the quality control process of their data, so only a brief description is given here in table 1. Three of the stations are part of the Baseline Surface Radiation Network (BSRN) [46], and one is the main station of the National Renewable Energy Laboratory (NREL) network (`https://midcdmz.nrel.gov/`). The four stations are respectively illustrative of distinct types of radiative climate: Arid, Temperate, Tropical, and High Albedo, which are conducive to different cloud regimes and different seasonal patterns, as shown in fig. 8. For this first example of application of the method, the original 1 min data have been compiled into hourly data to forecast the next hourly period, which is a common short-term horizon in practice. Further developments will investigate other applications, such as the 15 min horizon, which is also very important in electricity dispatching and price trading.

4 Performance metrics

Different metrics have been used in the literature to evaluate the accuracy of solar radiation models and forecasts [47, 48]. In the present study, the following statistical indicators are used.

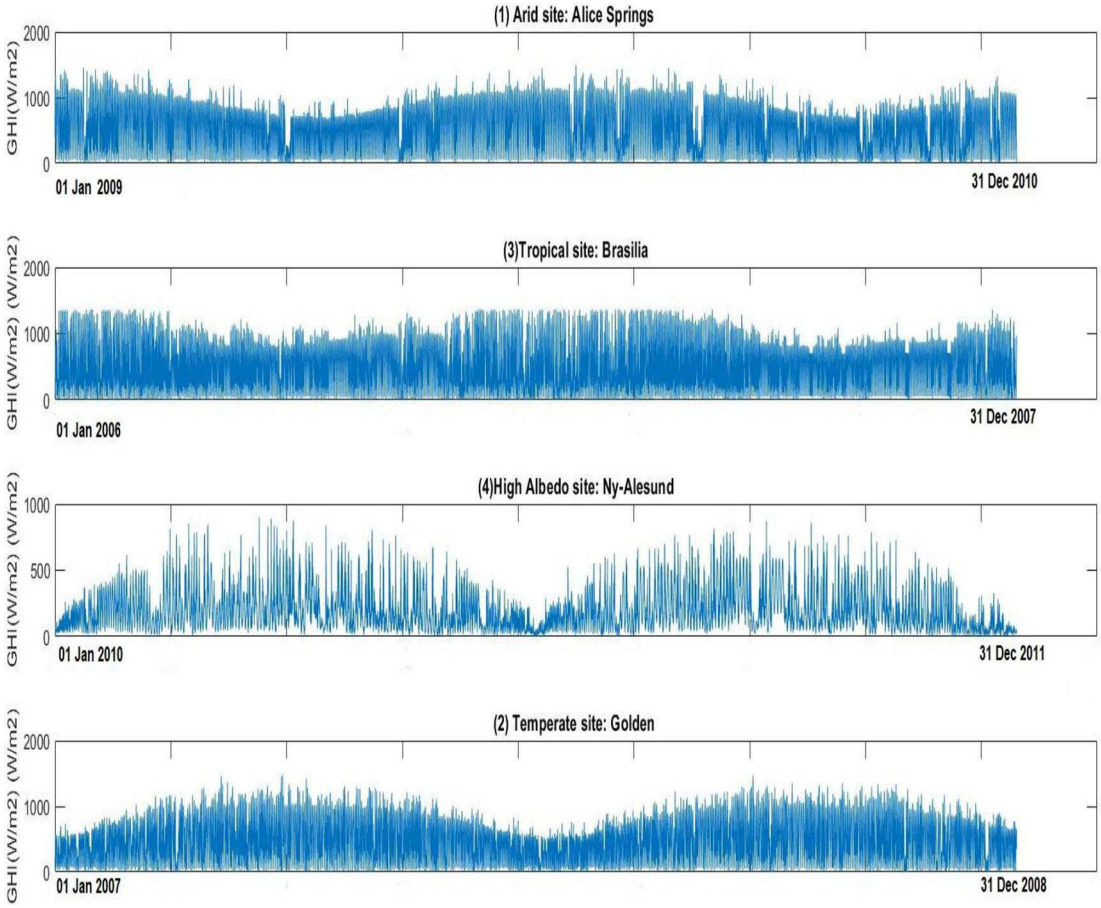

Fig. 8. Time series of GHI daytime data over two successive years, as recorded at four stations (Alice Springs, Brasilia, Ny-Ålesund, and Golden, from top to bottom).

– Coefficient of determination:

$$R^2 = 1 - \frac{var(y_t - \hat{y}_t)}{var(y_t)}\,. \tag{30}$$

– Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{M}\sum_{t=1}^{M}\frac{|y_t - \hat{y}_t|}{|y_t|} \times 100. \tag{31}$$

– Normalized Mean Squared Error (NMSE):

$$NMSE = \frac{\frac{1}{M}\sum_{t=1}^{M}(y_t - \hat{y}_t)^2}{var(y)}\,, \tag{32}$$

where $var(y)$ is the variance of measured values.

– Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{M}\sum_{t=1}^{M}(y_t - \hat{y}_t)^2}, \tag{33}$$

where y_t and $\widehat{y}$ are the measured and forecasted values, respectively, and M is the number of forecasted series.

– Forecasting skill (FS):

FS can be expressed as the ratio of the model's performance to that of the persistence model, in terms of RMSE:

$$FS = 1 - \frac{RMSE_m}{RMSE_p}\,. \tag{34}$$

Table 2. Statistical results obtained at all test stations by different autoregressive models and their combination methods. In each group of columns, the best results appear in boldface.

Site/Method	Individual models				Combination methods					
	ARMA	ARIMA	SARIMA	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs										
FS	0.09	**0.11**	−0.07	0.04	0.09	0.22	0.11	0.03	0.11	**0.38**
R^2	0.79	**0.80**	0.71	0.77	0.79	0.82	0.72	0.76	0.80	**0.90**
MABE	128.3	**115.8**	134.8	139.7	120.0	99.4	111.3	121.7	110.8	**71.9**
MAPE	25.0	**22.6**	26.3	27.4	23.4	19.7	21.8	23.7	21.6	**14.0**
RMSE	159.1	**155.4**	187.8	168.0	158.8	134.2	155.1	168.9	155.2	**108.1**
Brasillia										
FS	0.05	0.06	0.05	**0.14**	0.19	0.24	0.19	0.20	0.19	**0.43**
R^2	0.72	0.73	0.72	**0.77**	0.80	0.83	0.79	0.80	0.80	**0.90**
MABE	128.9	**113.6**	123.0	121.0	97.6	95.3	97.6	102.8	97.4	**62.2**
MAPE	25.5	**22.4**	24.3	24.0	19.3	18.8	19.3	20.3	19.2	**12.3**
RMSE	162.6	160.5	163.5	**147.2**	138.5	129.3	137.9	136.6	138.5	**97.1**
Ny-Ålesund										
FS	0.06	0.04	−0.06	**0.08**	−0.03	0.08	−0.01	−0.17	0.006	**0.27**
R^2	0.60	0.57	0.47	**0.61**	0.52	0.63	0.47	0.37	0.55	**0.76**
MABE	67.6	**60.5**	71.5	66.3	71.9	65.2	70.7	82.2	69.2	**44.7**
MAPE	32.7	**28.7**	34.0	32.0	34.8	31.5	34.2	39.8	33.4	**21.6**
RMSE	87.5	**83.1**	92.6	85.5	96.1	85.2	94.2	109.0	92.4	**67.7**
Golden										
FS	0.10	0.13	−0.02	**0.20**	0.11	0.19	0.12	−0.01	0.11	**0.40**
R^2	0.76	0.77	0.69	**0.81**	0.76	0.81	0.76	0.69	0.76	**0.89**
MABE	102.1	**93.9**	106.6	94.6	92.2	89.7	92.2	103.8	92.1	**57.5**
MAPE	23.9	**22.0**	24.9	22.1	21.6	21.01	21.6	24.3	21.6	**13.4**
RMSE	132.7	128.4	150.8	**117.7**	130.8	119.1	129.6	149.2	130.5	**87.6**

5 Results and discussion

For the experiments reported here, the GHI time series include two years of data at each of the four test stations. The GHI series, defined in a 50-dimensional variable space (*i.e.*, the last 50 hour points at any moment, excluding night periods), are subdivided into three sets: two learning sets (training and validation, approximately 25% of the whole dataset for each of them), and a test set (approximately half of the whole dataset). The training set is used to learn the different single models of the ensemble, whereas the validation set is required to train the different supervised combination/classification schemes proposed in this study (*i.e.*, SA, LSR, EB, DWS, OP and Class).

5.1 Results obtained with autoregressive predictors

AR predictors are the most general class of models for the forecasting of time series. Here, five models are used (AR, MA, ARMA, ARIMA and SARIMA). It is found that the best model is ARIMA at all sites, apparently because it takes advantage of two basic models (AR and MA) with the AR mathematical procedure known as "differencing operation" applied to make the series stationary. With respect to combination methods, the classification approach gives the best results at all sites (see table 2 and fig. 9).

5.2 Results obtained with neural predictors

Neural computing has emerged as a real-world technology, with effective applications in various fields. The majority of these applications are concerned with problems in pattern recognition, mainly in classification, regression and

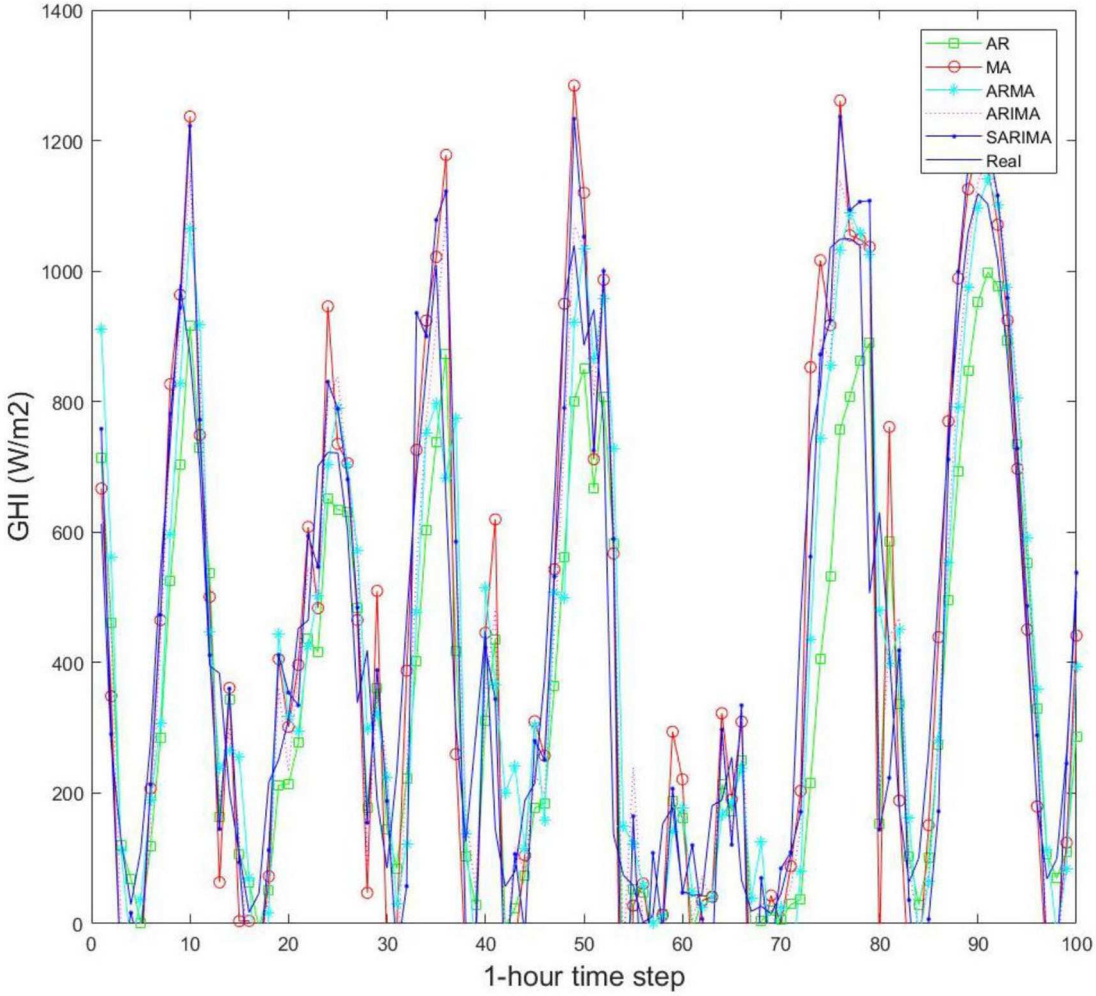

Fig. 9. Short window of measured and predicted time series by different autoregressive models at the Alice Springs test site.

clustering. They exploit feed-forward network architectures, such as the multi-layer perceptron or the radial basis function network [49]. In the present study, several architectures are investigated, namely, MLP, RBF, ELM, and ENN. Here, ENN appears to perform best in terms of all statistical indicators (see table 3 and fig. 10), which can be explained by the fact that an ENN reveals the temporal dynamic behavior of a time sequence. Contrasting with feedforward neural networks, ENNs can use their internal state (memory) to process various sequences of GHI measurements. This makes them more appropriate for this kind of task.

5.3 Results obtained with SVM predictors

The selection of this type of predictor results from the interest in a comprehensive and broad evaluation of the efficiency of the SVM technique when used in GHI forecasting problems. More precisely, three diverse types of SVM are considered: a linear SVM (SVM-Linear; equal to an SVM without kernel mapping), an SVM with polynomial kernel (SVM-Polynomial), and an SVM with Gaussian kernel (SVM-RBF). Adding the latter type of SVM makes it possible to assess the effect of kernel mapping on the SVM forecasting method. Moreover, valuable indications are then gained about which forecasting models should be selected to implement the diverse combination strategies and assess the efficiency of the ensemble. As reported in table 4, SVM-RBF performs best at all test sites, as revealed by all metrics. This can be related to the flexibility of the RBF kernel to model nonlinear phenomena (see fig. 11). With respect to combination methods, the classification-based strategy outperforms the other ones by a large margin.

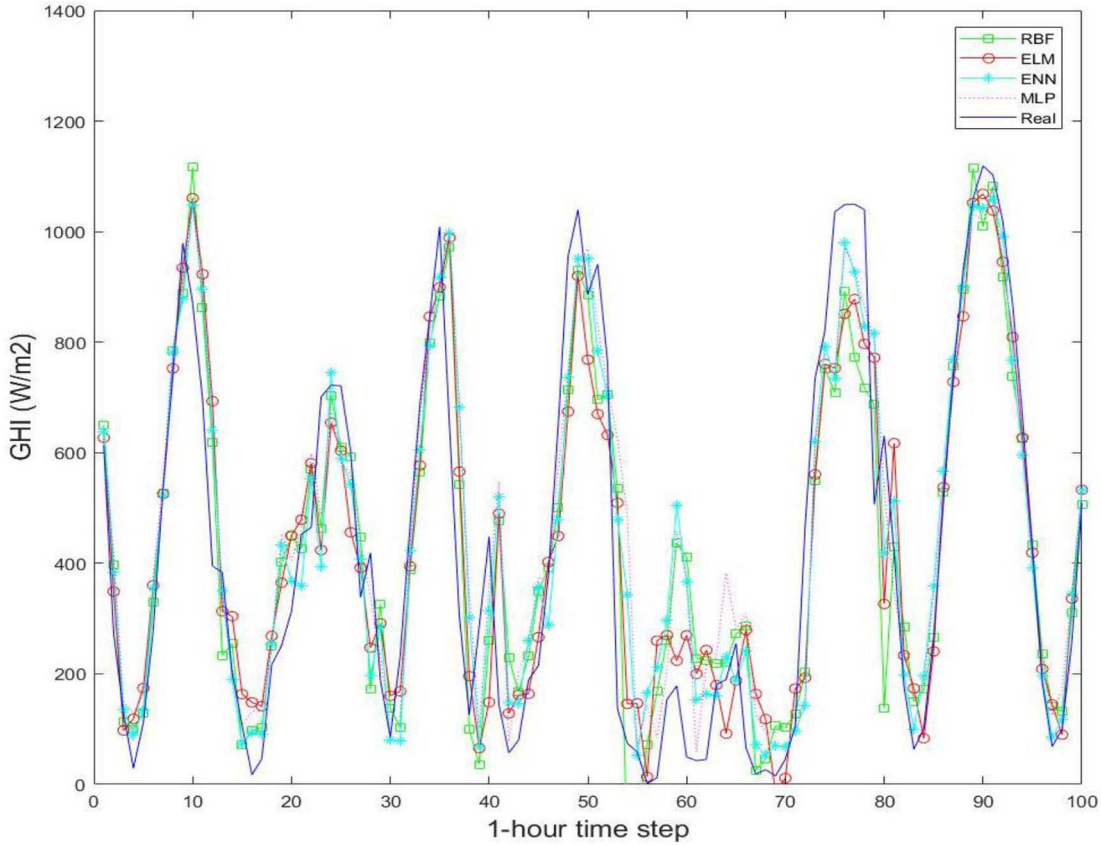

Fig. 10. Short window of measured and predicted time series by different ANN models at the Alice Springs test site.

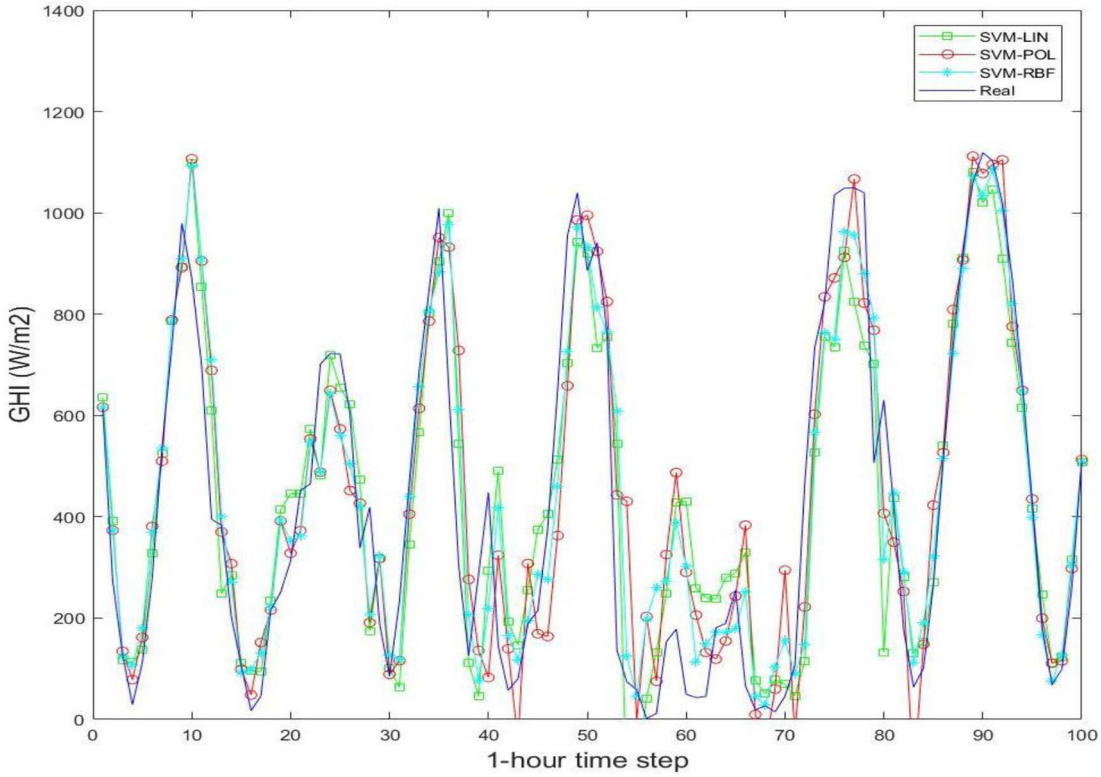

Fig. 11. Short window of measured and predicted time series by different SVM models for the Alice Springs test site.

Table 3. Forecasting results for the four test stations obtained with four ANN models and their combination methods. In each group of columns, the best results appear in boldface.

Site/Method	Individual models					Combination methods					
	MLP	RBF	ELM	ENN	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs											
FS	0.32	0.28	0.30	**0.34**	0.13	0.34	0.35	0.34	0.34	0.35	**0.49**
R^2	0.89	0.87	0.87	**0.89**	0.81	0.89	0.89	0.89	0.89	0.89	**0.93**
MABE	80.6	86.9	85.0	**75.4**	129.7	76.7	74.8	76.5	76.8	76.1	**50.7**
MAPE	15.7	16.9	16.5	**14.7**	25.3	14.9	14.6	14.9	14.9	14.8	**9.8**
RMSE	118.5	125.6	122.5	**114.6**	151.9	114.3	113.0	114.2	114.2	113.9	**88.1**
Brasilia											
FS	0.36	0.32	0.34	**0.38**	0.12	0.38	0.42	0.39	0.35	0.39	**0.54**
R^2	0.88	0.87	0.87	**0.88**	0.76	0.89	0.89	0.89	0.87	0.89	**0.93**
MABE	80.4	89.8	86.9	**80.2**	128.8	79.9	65.5	79.6	85.4	79.6	**50.6**
MAPE	15.9	17.7	17.2	**15.9**	25.5	15.8	12.9	15.7	16.9	15.7	**10.0**
RMSE	108.4	115.4	112.7	**106.1**	150.6	104.9	99.1	104.8	110.9	104.8	**78.8**
Ny-Ålesund											
FS	0.14	0.12	0.13	**0.15**	0.12	0.16	0.20	0.16	0.12	0.31	**0.37**
R^2	0.69	0.66	0.67	**0.69**	0.62	0.70	0.71	0.69	0.66	0.79	**0.83**
MABE	61.4	60.8	59.6	**57.7**	59.2	57.2	55.2	57.2	60.7	41.6	**48.4**
MAPE	29.7	29.4	28.8	**27.9**	29.6	27.7	26.7	27.7	29.3	20.1	**23.4**
RMSE	79.5	81.5	80.0	**78.2**	80.2	77.4	73.7	77.3	81.5	63.6	**58.1**
Golden											
FS	0.34	0.29	0.29	**0.34**	0.20	0.35	0.38	0.35	0.35	0.35	**0.51**
R^2	**0.877**	0.85	0.85	0.876	0.81	0.87	0.88	0.87	0.88	0.88	**0.93**
MABE	73.3	84.2	81.1	**71.5**	101.5	74.3	63.6	73.8	72.7	72.4	**47.1**
MAPE	17.1	19.7	19.0	**16.7**	23.7	17.4	14.9	17.3	17.0	16.9	**11.0**
RMSE	97.3	104.3	104.9	**96.7**	117.9	95.9	90.8	95.8	95.9	94.8	**71.5**

5.4 Results obtained with a hybrid ensemble of predictors

Here, for each family of predictors, the best performing one is selected. The best single model of the ensemble appears to be SVM-RBF. This could be expected to some extent, due to the better generalization capability of SVMs at learning problems. With respect to the combination strategies, as it was for previous combinations, the classification-based strategy outperforms the other ones by a large margin (see table 5 and fig. 12).

5.5 Results obtained with feature selection

Another way to investigate the performance of the proposed forecasting methodology is to use an additive processing step prior to introducing the input series to the different predictors. Here, this means that the feature selection method described in [36] is used, and that the best set of the selected historical variables is obtained using the hold-out principle of model selection [22]. For each case, the best set is obtained according to the performance of the forecasting technique in the validation set.

From table 6, it can be seen that there is a slight performance increase in comparison with the results reported in table 5 (without feature selection). An important exception is that of the autoregressive model (ARIMA), which degrades the forecasting performance. This can be explained by the fact that ARIMA models are based on polynomial modelling and need more historical variables to describe the time series behavior of the solar radiation signal. Nevertheless, when using the different combination techniques, the system maintains the same level of performance despite the poor results of the ARIMA technique. This confirms the usefulness of the combination strategy when there is enough diversity in the input techniques.

Another way to compare the performance of the different models consists in using the Taylor diagram [51], which is a graphical illustration of the statistical connection between two random variables. It characterizes three population

Table 4. Forecasting results obtained with three SVM kernel models and their combination methods at all four test stations. In each group of columns, the best results appear in boldface.

Site/Method	Individual models				Combination methods					
	SVM-RBF	SVM-Poly	SVM-Lin.	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs										
FS	**0.35**	0.27	0.27	0.14	0.33	0.35	0.34	0.35	0.34	**0.48**
R^2	**0.89**	0.86	0.86	0.81	0.89	0.89	0.89	0.89	0.89	**0.93**
MABE	**75.8**	86.2	89.5	127.3	78.6	76.05	78.3	75.8	77.9	**73.2**
MAPE	**14.7**	16.8	17.4	24.8	15.3	14.8	15.2	14.7	15.1	**10.7**
RMSE	**113.6**	126.6	126.9	149.2	115.8	113.5	115.67	113.6	115.3	**91.0**
Brasilia										
FS	**0.39**	0.38	0.37	0.16	0.39	0.42	0.39	0.39	0.39	**0.48**
R^2	**0.89**	0.88	0.88	0.78	0.89	0.89	0.89	0.89	0.89	**0.92**
MABE	**80.5**	81.4	84.1	123.5	81.1	66.7	81.1	80.5	81.3	**63.4**
MAPE	**15.9**	16.1	16.6	24.5	16.0	13.2	16.0	15.9	16.1	**12.5**
RMSE	**105.0**	106.3	108.3	144.3	104.5	99.0	104.5	104.8	104.5	**88.6**
Ny-Ålesund										
FS	**0.15**	0.13	0.07	0.11	0.14	0.18	0.14	0.15	0.15	**0.26**
R^2	**0.68**	0.66	0.63	0.40	0.68	0.70	0.68	0.68	0.68	**0.76**
MABE	**57.7**	59.6	65.2	78.2	58.4	56.1	58.3	57.5	58.2	**45.1**
MAPE	**27.9**	28.8	31.5	33.1	28.2	27.1	28.2	27.8	28.1	**21.8**
RMSE	**78.2**	80.5	85.8	99.5	79.3	75.7	79.	78.2	79.0	**68.1**
Golden										
FS	**0.34**	0.33	0.31	0.28	0.34	0.35	0.34	0.34	0.34	**0.45**
R^2	**0.87**	0.86	0.86	0.84	0.87	0.87	0.87	0.87	0.87	**0.91**
MABE	**72.3**	75.4	78.4	92.6	73.2	70.1	73.1	72.3	73.2	**56.6**
MAPE	**16.9**	17.6	18.3	21.6	17.1	16.4	17.1	16.9	17.1	**13.2**
RMSE	**97.2**	98.7	101.1	106.3	96.3	95.4	96.3	97.2	96.3	**80.0**

statistics linking two variables at the same time, namely: correlation, standard deviation, and centered root mean square difference. These graphs are specifically suitable to assess several aspects of complex models or in judging the relative skill of many different models.

In the two-dimensional space of a Taylor diagram, each point can describe the combination of the three statistics stated above because they are interrelated by

$$E'^2 = \sigma_t^2 + \sigma_m^2 - 2\sigma_t\sigma_m R, \tag{35}$$

where E' is the centered mean-square difference between the measured and forecasted series, R is the correlation coefficient between the measured and forecasted series, and σ_m and σ_f are the variances of the test and model fields, respectively. The centered RMS difference between the forecasted and measured GHI time series is relative to the distance between the prediction and the point on the x-axis identified as "Measured". Similarly, the standard deviation of the forecasted series is proportional to the radial distance between the prediction and the origin.

Using the present observations and forecast models, Taylor diagrams for the different sites are shown in figs. 13–16. These diagrams indicate that all correlation values between the measured and forecasted time series are in the range [0.95–0.99] for Alice Springs, Ny-Ålesund and Brasilia, and in the range [0.90–0.95] for Golden. These diagrams also clearly indicate that the classification strategy outperforms all other combination models. It is worth noting that the Taylor diagram does not provide information about overall biases, but simply describes the centered model error.

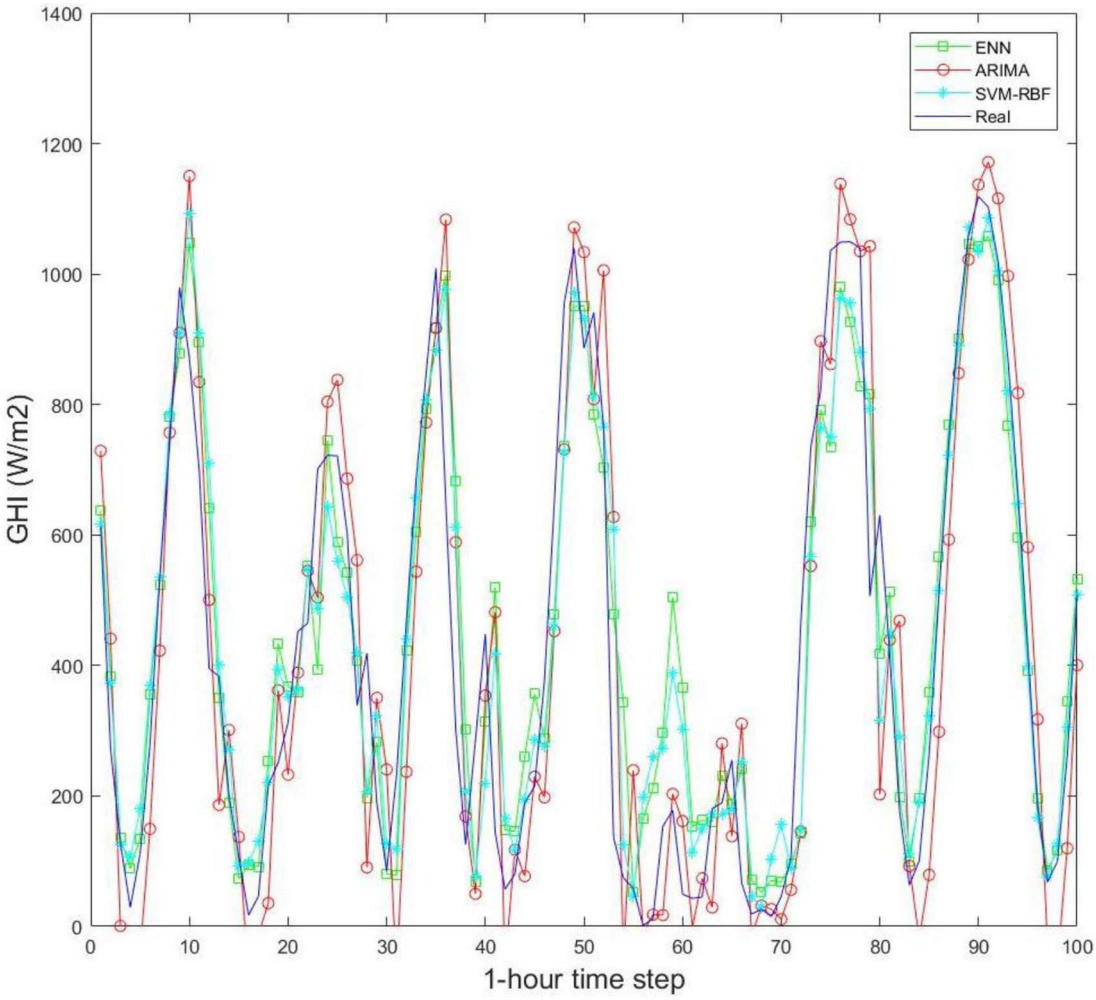

Fig. 12. Comparison between the best performing models for each prediction family at the Alice Springs test site.

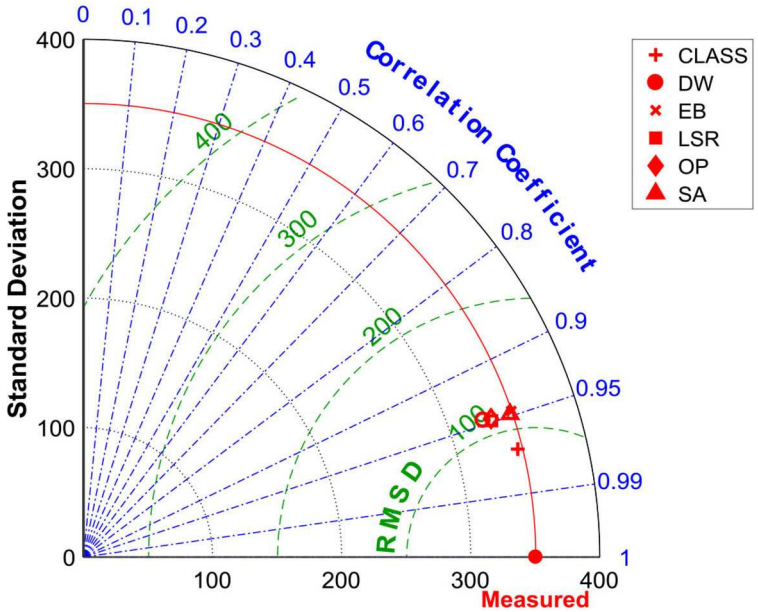

Fig. 13. Taylor diagram of measured and predicted GHI values by different combination models for Alice Springs.

Table 5. Forecasting results obtained with the best single model of each prediction family (SVM-RBF, ENN and ARIMA) and their combination methods at all four test stations. In each group of columns, the best results appear in boldface. Individual models include a typical clear-sky determination (CS) [50] and Persistence (Persist.).

Site/Method	Individual models					Combination methods					
	ARIMA	ENN	SVM-RBF	CS	Persist.	SA	LSR	EB	DWS	OP	Class
Alice Springs											
FS	0.11	0.34	**0.35**	0.15	0.17	0.35	0.36	0.36	0.35	0.36	**0.51**
R^2	0.80	0.89	**0.89**	0.58	0.82	0.89	0.90	0.89	0.89	0.89	**0.94**
MABE	115.8	**75.4**	75.8	140.1	121.0	75.9	71.4	73.9	75.7	73.4	**49.3**
MAPE	22.6	**14.7**	14.8	27.3	23.6	14.8	13.9	14.4	14.7	14.3	**9.6**
RMSE	155.4	114.6	**113.6**	260.0	145.4	113.6	110.8	112.1	113.5	111.4	**84.5**
Brasillia											
FS	0.06	0.38	**0.39**	0.26	0.11	0.40	0.42	0.41	0.40	0.39	**0.59**
R^2	0.73	0.88	**0.89**	0.71	0.75	0.89	0.89	0.88	0.89	0.88	**0.94**
MABE	113.6	**80.2**	80.5	91.5	131.4	73.2	65.6	73.1	77.0	73.4	**46.0**
MAPE	22.4	**15.8**	15.9	18.1	26.0	14.5	12.9	14.4	15.2	14.5	**9.1**
RMSE	160.5	106.1	**105.0**	186.3	152.5	102.8	99.2	100.4	101.8	104.1	**70.3**
Ny-Ålesund											
FS	0.04	0.15	**0.15**	−0.01	0.10	0.20	0.20	0.14	0.17	0.13	**0.38**
R^2	0.57	**0.69**	0.68	−1.29	0.70	0.50	0.74	0.41	0.70	0.66	**0.83**
MABE	60.5	57.8	**57.7**	155.6	58.8	78.3	26.4	61.0	56.6	62.8	**39.3**
MAPE	28.7	27.9	**27.9**	117.4	28.4	37.8	54.7	29.5	27.3	30.4	**19.0**
RMSE	83.1	78.2	**78.2**	187.9	75.2	98.1	71.1	79.6	76.8	80.8	**56.8**
Golden											
FS	0.13	**0.34**	0.34	0.20	0.17	0.36	0.37	0.36	0.35	0.36	**0.53**
R^2	0.77	**0.87**	0.87	0.60	0.79	0.88	0.88	0.87	0.87	0.88	**0.93**
MABE	93.9	71.5	**72.3**	106.0	102.5	67.0	65.0	66.8	68.8	66.9	**43.5**
MAPE	22.0	**16.7**	16.9	24.8	23.9	15.7	15.2	15.6	16.1	15.6	**10.1**
RMSE	128.4	**96.7**	97.2	197.0	121.5	94.2	92.6	93.3	95.5	93.8	**68.2**

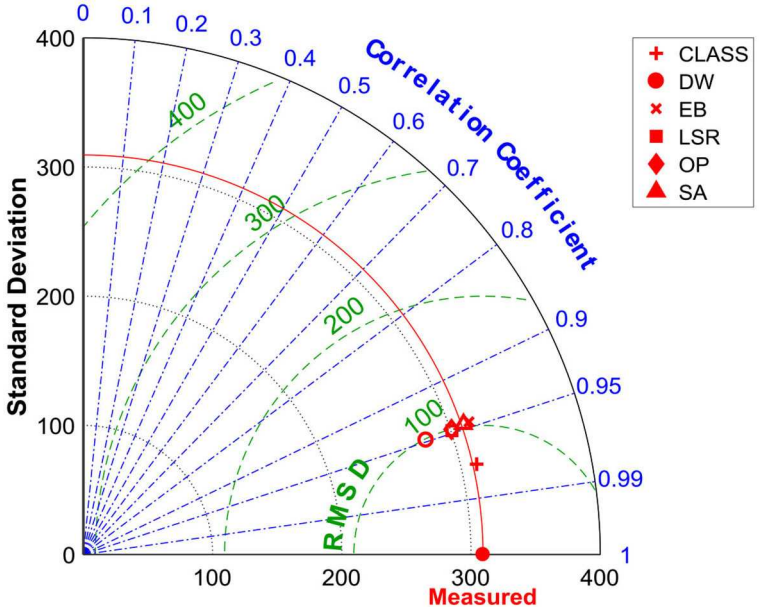

Fig. 14. Taylor diagram of measured and predicted GHI values by different combination models for Brasilia.

Table 6. Forecasting results obtained using two-blocs system (feature selection and regression technique). In each group of columns, the best results appear in boldface. SV: selected historical variables.

Site/Method	Individual models			Combination methods					
	ARIMA	ENN	SVM_RBF	SA	LSR	EB	DWS	OP	Class
Alice Springs	**SV = 15**	**SV = 12**	**SV = 19**						
FS	−1.62	0.35	0.35	−0.09	0.37	0.28	0.35	0.31	**0.48**
R^2	−0.72	0.89	0.89	0.70	0.90	0.88	0.89	0.88	**0.93**
MABE	386.5	77.5	75.5	159.2	14.4	91.0	75.6	89.4	**57.1**
MAPE	75.3	15.1	14.7	31.0	74.1	17.7	14.7	17.4	**11.1**
RMSE	460.4	113.0	113.2	191.6	109.3	124.9	112.9	120.1	**91.1**
Brasillia	**SV = 12**	**SV = 11**	**SV = 16**						
FS	−1.31	0.38	0.38	−0.001	0.43	0.31	−0.02	0.30	**0.50**
R^2	−0.65	0.895	0.89	0.69	0.90	0.85	0.68	0.85	**0.92**
MABE	334.1	81.4	81.3	144.8	65.5	94.5	146.8	97.3	**62.1**
MAPE	66.1	16.1	16.0	28.6	12.9	18.7	29.0	19.2	**12.2**
RMSE	397.4	105.5	105.3	172.3	96.7	118.3	175.7	119.6	**85.7**
Ny-Ålesund	**SV = 17**	**SV = 14**	**SV = 11**						
FS	−1.83	−0.20	−0.15	−0.17	0.12	−0.04	−0.15	−0.04	**0.23**
R^2	−0.66	0.78	0.78	0.75	0.84	0.75	0.80	0.82	**0.89**
MABE	113.5	43.1	39.9	47.3	29.6	40.0	41.7	40.3	**26.9**
MAPE	85.7	32.5	30.1	35.6	22.4	30.2	31.4	30.4	**20.3**
RMSE	145.1	61.5	58.9	60.3	44.6	53.6	59.1	53.5	**38.9**
Golden	**SV = 17**	**SV = 17**	**SV = 13**						
FS	−0.60	0.36	0.33	0.17	0.39	0.30	−0.39	0.31	**0.53**
R^2	0.23	0.88	0.87	0.80	0.89	0.85	0.41	0.86	**0.93**
MABE	193.2	70.8	78.2	99.7	14.8	81.9	169.7	81.6	**49.6**
MAPE	45.2	16.6	18.3	23.3	63.3	19.1	39.7	19.1	**11.6**
RMSE	236.6	94.0	98.9	121.3	89.6	102.4	206.7	101.5	**68.4**

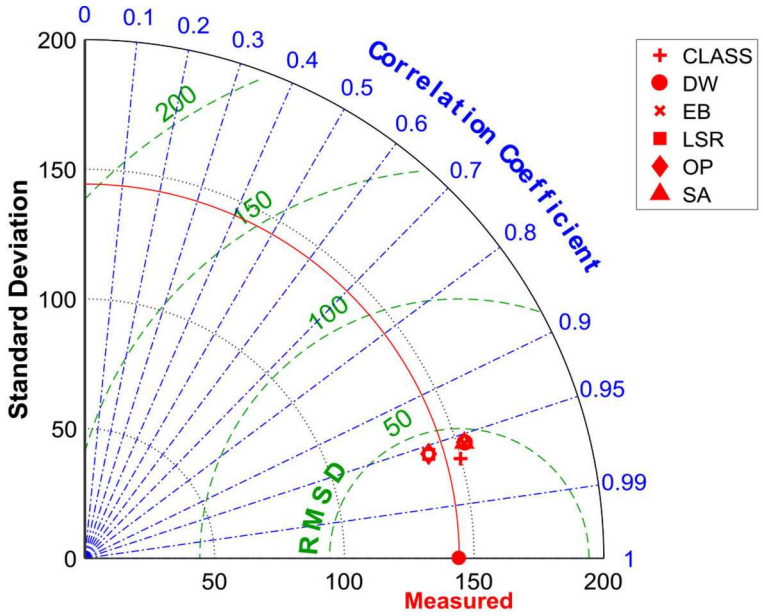

Fig. 15. Taylor diagram of measured and predicted GHI values by different combination models for Ny-Ålesund.

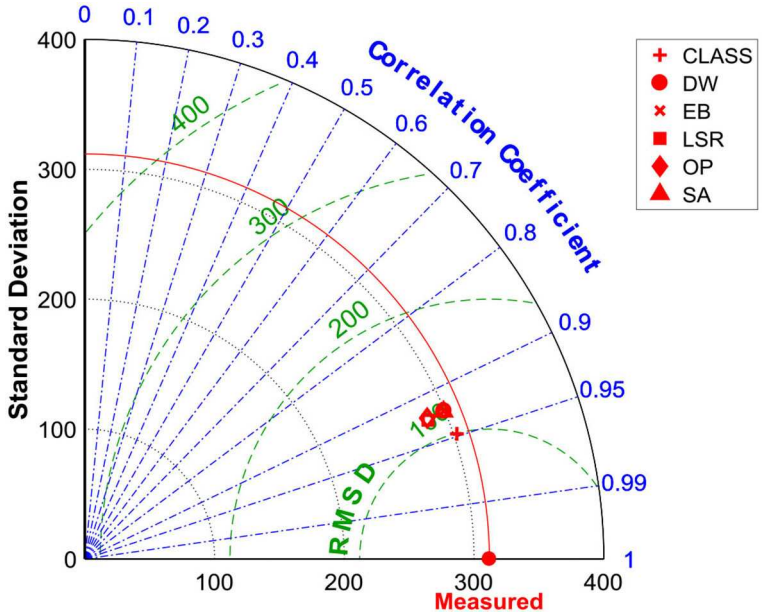

Fig. 16. Taylor diagram of measured and predicted GHI values by different combination models for Golden.

Table 7. Results of the hypothesis testing comparing the classification strategy to other combination strategies for Alice Springs.

Class *vs.* other fusion strategies	W-L-T	Fisher test p-value	K-S test p-value	t-Test p-value
LSR	27-22-1	–	0.840	0.900
OP	26-22-2	–	0.150	0.910
EB	44-6-0	–	0.004	0.020
DWS	36-14-0	–	0.030	0.650
SA	47-0-3	–	0.190	0.710

5.6 Hypothesis testing

The statistical hypothesis testing carried out in this study is principally based on the paired t-test and Fisher sign test. They are useful to assess whether a particular combination approach of the best individual models (*i.e.*, (ARIMA) of autoregressive model, (MLP) of artificial neural networks, and (SVM-RBF) of support vector machines) is better than the best fusion strategy. The latter is known to be the classification strategy [18,52]. To perform the statistical tests, the complete dataset is subdivided into M subsets so that each model is learned in subdivision i and tested in subdivision $i+1$. To obtain a group of 51 different subsets, K is fixed to 50. The Kolmogorov-Smirnov (K-S) test can then be executed to confirm the data normality. If indeed the data distribution is normal, a t-test is executed to compare the performance of the diverse combined strategies and uncover the best ones. If not, a non-parametric Fisher sign test is performed to achieve the same comparison.

Table 6 provides a comparison between the different fusion strategies and the proposed methods based on classification for Alice Springs. The level of significance, α, is fixed at 0.05, which corresponds to a 95% confidence. The results of the K-S test and those of either the t-test or the Fisher test are also presented. Furthermore, for each case, Win-Lose-Tie (W-L-T) values are provided. Here, the probabilistic p-value is the probability that, when the null hypothesis is true, the statistical performance of the proposed classification approach would be the same as, or better than, other scenarios. In other words, a lower p-value points to an improved performance of the proposed model. From table 7, it is evident that the Classification approach statistically outperforms all other combination strategies, as evidenced by the W-L-T values.

6 Conclusion

The topic of time series analysis and forecasting is a dynamic research area, having fundamental importance in numerous practical fields. Improving accuracy of time series forecasts is a challenging task that has been gaining

continuous research attention from the past few decades. In this contribution, a new methodology has been proposed to forecast the global horizontal irradiance (GHI) of the next hour, based on a multimodel ensemble. The methodology targets the exploitation of the diversity between an ensemble of varied forecast models, with the objective of improving the accuracy of the forecasting process. Several strategies have been described and implemented. The different strategies vary from each other with regard to: i) the combination procedure; ii) the technique used to exploit the accessible prior knowledge of the hourly GHI time series.

The diverse ensembles of multimodels utilized in the present experiments are based on different forecasting families, mainly: 1) Autoregressive models; 2) Neural networks models; 3) Support vector machine models. The methodologies proposed here can improve the forecasting performance of GHI by combining different techniques, just with the condition of diversity, *i.e.* using different forecasting families.

With the purpose of achieving a consistent and comprehensive evaluation of the effectiveness of the proposed methodology, different experiments (with different architectures) were considered. The experiments simulated several architectures, where the members of the ensemble were either based on the same forecasting family or the best model of each prediction family included in the ensemble.

The experimental assessment performed to assess the efficiency of the multimodel system clearly showed that it can deliver forecasting accuracy and skill better than the best single model incorporated in the ensemble, with a higher performance when using the classification strategy. This result is particularly encouraging because the improved skill of the proposed ensemble method is maintained at all four test sites investigated here, which are representative of diverse climates and overall cloud regimes. Further developments of the method will evaluate its skills for shorter-term forecasts, *e.g.*, for the 15 min horizon.

The principal investigators and personnel of four BSRN and NREL radiometric stations are thanked for their relentless efforts in maintaining these sites and providing the high-quality observations that were necessary for this study.

References

1. M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Renew. Sustain. Energy Rev. **27**, 65 (2013).
2. R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, Prog. Energy Combust. Sci. **39**, 535 (2013).
3. C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Renew. Energy **105**, 569 (2017).
4. T. Khatib, A. Mohamed, K. Sopian, Renew. Sustain. Energy Rev. **16**, 2864 (2012).
5. A.K. Yadav, S.S. Chandel, Renew. Sustain. Energy Rev. **33**, 772 (2014).
6. D. Yang, J. Kleissl, C.A. Gueymard, H.T.C. Pedro, C.F.M. Coimbra, Sol. Energy **168**, 60 (2018).
7. A. Mellit, M. Benghanem, S.A. Kalogirou, Appl. Energy **83**, 705 (2006).
8. F.O. Hocaoğlu, Ö.N. Gerek, M. Kurban, Sol. Energy **82**, 714 (2008).
9. J. Cao, X. Lin, Eng. Appl. Artif. Intell. **21**, 1255 (2008).
10. A. Mellit, A.M. Pavan, Sol. Energy **84**, 807 (2010).
11. L. Martín, L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony, Sol. Energy **84**, 1772 (2010).
12. K. Benmouiza, A. Cheknane, Energy Convers. Manag. **75**, 561 (2013).
13. S. Salcedo-Sanz, C. Casanova-Mateo, J. Munoz-Mari, G. Camps-Valls, IEEE Geosci. Remote Sens. Lett. **11**, 1936 (2014).
14. K. Benmouiza, A. Cheknane, Theor. Appl. Climatol. **124**, 945 (2016).
15. E. Akarslan, F.O. Hocaoglu, Renew. Energy **87**, 628 (2016).
16. S. Monjoly, M. André, R. Calif, T. Soubdhan, Energy **119**, 288 (2017).
17. H. Bouzgou, C.A. Gueymard, Sol. Energy **158**, 595 (2017).
18. H. Bouzgou, C.A. Gueymard, Renew. Energy **133**, 1055 (2019).
19. L. Cornejo-Bueno, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz, Sol. Energy **183**, 768 (2019).
20. Y. Liu, X. Yao, Neural Netw. **12**, 1399 (1999).
21. Y. Ren, P.N. Suganthan, N. Srikanth, Renew. Sustain. Energy Rev. **50**, 82 (2015).
22. H. Bouzgou, *Advanced Methods for the Processing and Analysis of Multidimensional Signals: Application to Wind Speed* (University of Batna, 2012).
23. Y. Kemmoku, S. Orita, S. Nakagawa, T. Sakakibara, Sol. Energy **66**, 193 (1999).
24. S. Alessandrini, L. Delle Monache, S. Sperati, G. Cervone, Appl. Energy **157**, 95 (2015).
25. S. Sperati, S. Alessandrini, L. Delle Monache, Sol. Energy **133**, 437 (2016).
26. J. Hall, J. Hall, in *Proceedings of the 40th ASES National Solar Conference 2011* (ASES, 2011).
27. A. Chaouachi, R. Kamel, K. Nagasaka, Jaciii **3**, 41 (2010).
28. N. Zemouri, H. Bouzgou, in *International Conference on Artificial Intelligence in Renewable Energy Systems, IC-AIRES2017* (2017) pp. 1–7.

29. R.T. Clemen, Int. J. Forecast. **5**, 559 (1989).
30. Z.H. Zhou, *Ensemble Methods: Foundations and Algorithms* (2012).
31. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, in *Time Series Analysis: Forecasting and Control* (Prentice Hall, New Jersey, 1994).
32. S. Haykin, *Neural Networks and Learning Machines* (2008).
33. J. Park, I.W. Sandberg, Neural Comput. **3**, 246 (1991).
34. H. Bouzgou, N. Benoudjit, Appl. Energy **88**, 2463 (2011).
35. G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Neurocomputing **70**, 489 (2006).
36. H. Bouzgou, C.A. Gueymard, Sol. Energy **158**, 595 (2017).
37. V.N. Vapnik, *Statistical Learning Theory* (1998).
38. K.R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, IEEE Trans. Neural Networks **12**, 181 (2001).
39. L.I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms: Second Edition* (2014).
40. T.G. Dietterich, Neural Comput. **10**, 1895 (1998).
41. R.L. Winkler, Int. J. Forecast. **5**, 605 (1989).
42. J.S. Armstrong, in *A Handbook of Human Resource Management Practice* (2001) p. 849.
43. C. Lemke, B. Gabrys, Neurocomputing **73**, 2006 (2010).
44. L. Bruzzone, F. Melgani, IEEE Trans. Geosci. Remote Sens. **43**, 159 (2005).
45. C.A. Gueymard, J.A. Ruiz-Arias, Sol. Energy **128**, 1 (2016).
46. A. Ohmura, E.G. Dutton, B. Forgan, C. Fröhlich, H. Gilgen, H. Hegner, A. Heimo, G. König-Langlo, B. McArthur, G. Müller, R. Philipona, R. Pinker, C.H. Whitlock, K. Dehne, M. Wild, Bull. Am. Meteorol. Soc. **79**, 2115 (1998).
47. C.F.M. Coimbra, J. Kleissl, R. Marquez, in *Solar Energy Forecasting and Resource Assessment* (Academic Press, 2013) pp. 171–194.
48. C.A. Gueymard, Renew. Sustain. Energy Rev. **39**, 1024 (2014).
49. C.M. Bishop, in *Neural Networks Pattern Recognition* (1996).
50. R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, F. Vignola, Sol. Energy **73**, 307 (2002).
51. K.E. Taylor, J. Geophys. Res. Atmos. **106**, 7183 (2001).
52. T.W. Anderson, Biometrics **41**, 815 (1985).