



**HAL**  
open science

# Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting

Luis Mazorra Aguiar, B. Pereira, Philippe Lauret, F. Díaz, Mathieu David

## ► To cite this version:

Luis Mazorra Aguiar, B. Pereira, Philippe Lauret, F. Díaz, Mathieu David. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, 2016, 97, 10.1016/j.renene.2016.06.018 . hal-01334564

**HAL Id: hal-01334564**

**<https://hal.science/hal-01334564>**

Submitted on 18 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting

L. Mazorra Aguiar <sup>a,\*</sup>, B. Pereira <sup>a</sup>, P. Lauret <sup>b</sup>, F. Díaz <sup>a</sup>, M. David <sup>b</sup>

<sup>a</sup> University Institute for Intelligent Systems and Numerical Applications in Engineering, University of Las Palmas de Gran Canaria, Edificio Central del Parque Tecnológico, Campus de Tafira, 35017, Las Palmas de Gran Canaria, Spain

<sup>b</sup> Laboratoire de Physique et Ingénierie Mathématique pour l'Energie et l'environnement (PIMENT), University of La Réunion, Campus du Moufia 15, Avenue René Cassin, 97715 Saint Denis Messag 9, France

## A B S T R A C T

Isolated power systems need to generate all the electricity demand with their own renewable resources. Among the latter, solar energy may account for a large share. However, solar energy is a fluctuating source and the island power grid could present an unstable behavior with a high solar penetration. Global Horizontal Solar Irradiance (GHI) forecasting is an important issue to increase solar energy production into electric power system. This study is focused in hourly GHI forecasting from 1 to 6 h ahead. Several statistical models have been successfully tested in GHI forecasting, such as autoregressive (AR), autoregressive moving average (ARMA) and Artificial Neural Networks (ANN). In this paper, ANN models are designed to produce intra day solar forecasts using ground and exogenous data. Ground data were obtained from two measurement stations in Gran Canaria Island. In order to improve the results obtained with ground data, satellite GHI data (from Helioclim 3) as well as solar radiation and Total Cloud Cover forecasts provided by the European Centre for Medium Range Weather Forecasts (ECMWF) are used as additional inputs of the ANN model. It is shown that combining exogenous data (satellite and ECMWF forecasts) with ground data further improves the accuracy of the intra day forecasts.

## 1. Introduction

According to the trend shown for renewable energy generation, the proportion of these kinds of energies in the power system will increase in the next years. Renewable energies show a fluctuating generation profile because of their dependence on meteorological conditions. Since grid operators need to keep under control these variations, in order to accommodate the input/output balance of the system, forecasting methods are necessary to improve the coupling of renewable sources. This becomes an important issue when facing the fact that photovoltaic and solar thermo electrical plants are widely spread in most of the present power systems.

The output power of a PV plant is mainly correlated with the global solar irradiance received on the plan of the modules. The calculation of this received irradiance requires the knowledge of

the direct and diffuse components of the solar irradiance. But in many cases as in our study, no measurement of these two components exists. Thus, these two components are commonly derived from the global horizontal irradiance (GHI) with decomposition models [1–3]. When no assumption is done on the inclination of the PV modules, the key factor becomes the GHI. So, in this work, we will only focus on the forecasting of the GHI. Indeed, a reliable Global Horizontal Irradiance (GHI) forecasting model is considered as an important tool to avoid unstable behavior in the electrical grid, maintaining the balance between demand and supply [4,5].

Isolated power systems, such as the one in Gran Canaria Island, need to generate all the electricity demand solely with their own resources. The island power grid could present an unstable behavior if there is a large scale solar energy introduction. This is aggravated when a great variability in solar radiation conditions are shown in the Canary Islands, caused by the diversity of the climatic areas that emerge as a by product of the differences in its orography [6,7].

GHI forecasting has been developed in the last years using a

\* Corresponding author.

E-mail address: [luis.mazorra@ulpgc.es](mailto:luis.mazorra@ulpgc.es) (L.M. Aguiar).

wide range of methods. The most common forecasting models used in this field are:

- Statistical models, models based on GHI time series prediction: most common statistical models used in GHI forecasting are linear models, such as autoregressive (AR) and autoregressive moving average (ARMA), and machine learning techniques, such as Artificial Neural Networks (ANNs) [8].
- Forecasting using Ground based sky images: models based on sky images obtained with 180° vision angle cameras. Sky images lead to know cloud cover conditions for a few minutes ahead.
- Satellite images models. Geostationary satellites get atmosphere images all around the Earth with temporal resolution less than an hour. The high development occurred in the recent years in satellite data retrieving makes this technique a very useful tool to improve GHI forecasting [9,10].
- Numerical Weather Prediction models (NWP) based on physical models for estimating atmosphere conditions, including clouds formation and dissolution. Physical models are described with differential equations solved with numerical methods. NWPs models offer time horizons forecasting from few hours to 15 days ahead [4,11].

Depending on the forecasting time horizon and granularity, forecasting results are used for different purposes and are based on different input parameters [12,13]. Ground based sky images models show high precision information about cloud cover conditions and movement for intra hourly forecasting horizons [14]. On the other hand, forecasting models based on cloud motion vectors derived from geostationary satellite images offer accurate results for hourly prediction range up to 6 h ahead [9,10]. NWPs forecasting precision varies depending on the selected time scale and the geographic zone of the study. Heinemann [4] showed that acquiring the radiation values for clear skies without deviation is possible. NWPs models associated with a post processing method using hourly ground measurements showed accurate results for predictions from 6 h onwards [15].

The statistical method proposed in this paper is based on Artificial Neural Network (ANN) for GHI hourly forecasting from 1 to 6 h ahead. ANNs are considered a very attractive method for GHI forecasting because of their capacity to establish relationships between an input and output datasets [8]. ANNs have been used for solar radiation prediction using past ground measurements. Many papers have described ANN accurate results for GHI predictions with time horizons from few hours ahead, such as explained in Ref. [16–19], to 24 h ahead daily irradiation forecasting [20,21]. Statistical models based on ANN offer the possibility to combine historical GHI ground data with other meteorological data. Rehman [20] used air temperature and ground relative humidity for daily irradiation forecasting. Kemmoku [22] designed an ANN to forecast daily irradiation using atmospheric pressure and several meteorological parameters. Sfetsos & Coonick [23] used GHI ground data, air temperature, wind velocity and pressure for solar irradiance hourly forecasting. Moreover, several papers studied the influence of sunshine duration, air temperature, relative humidity, latitude and longitude to obtain GHI hourly and daily forecasting [18,24,25].

ANNs may even lead to more accurate GHI forecasts by combining ground measurement time series analysis with exogenous data, such as ground irradiance derived from satellite images and different NWPs data. In the last years, several studies used different techniques such as autoregressive models (AR) [26–28], artificial neural networks [29,30] and genetic algorithms for choosing the relevant information from satellite derived data [28].

In this paper satellite derived and ground data are combined with solar radiation and total cloud cover predicted by the

European Centre for Medium Range Weather Forecasts (ECMWF). The satellite derived data is selected as in Ref. [30], using correlation analysis results. In this work, we used the Bayesian framework [31] which improves ANNs learning process. In addition, the Bayesian approach provides techniques to optimize the ANN architecture and to select the most relevant inputs [32,33].

The organization of this paper is as follows. In section 2, we present the ground data, satellite derived data and ECMWF data used in the present paper. The clear sky model is described in Section 3 and we analyzed Ground data distribution in Section 4. In Section 5, we detail the forecasting methods theory, in Section 6 we explain the ANN architecture optimization and Section 7 is devoted to the selection of the most relevant satellite inputs. In Section 8, we show the error metrics used in this paper and in Section 9 we detail the results of the different models and discuss the influence of the different exogenous data to improve GHI forecasting. Finally, Section 10 gives main remarks and conclusions.

## 2. Available data

The datasets used in the present study to build the forecasting model include ground, satellite and numerical predicted data. All the data will be converted into an hourly basis in order to compare and contrast all the information in a coherent manner. The temporal availability for all type of data was restricted to the year 2005 because this is the only year in which ground and satellite data overlaps: ground data have been processed up to the year 2005, meanwhile the satellite data recollection started on the same year.

The application of the ANN models requires the division in training and testing datasets. Therefore, due to the restriction of only one year, we took the decision to select three weeks of a month for training and one week for testing purposes since this study only possess one year available. In this manner we can assure that training and testing datasets share similarities regarding meteorological behaviors, while keeping a ratio of 75% training and 25% testing [28].

### 2.1. Ground data set

Ground data are obtained from two measurement stations located in Gran Canaria Island (Spain) and managed by the Canary Islands Technology Institute (Instituto Tecnológico de Canarias). These stations are part of a network of 23 stations in the Canary Islands, validated by the elaboration of a solar map of the zone [6,34]. The selected stations are representative for the climatic division present on this island [6,35]. This climatic contrast between the northern region, with higher cloudiness, and the southern one, with clearer days and less variability, is well known and produced by the strong influence of the trade winds of the area.

The equipment used to acquire global solar horizontal irradiance (GHI) is a secondary standard pyranometer, the CMP 11 of Kipp & Zonen, with 3% accuracy for daily sum of GHI. Data are recorded for every 5 s, 1 min average and later assembled into an hourly basis. In order to maintain enough quality of the data, all the measurement data were treated with the SERI QC control software [36,37]. This treatment consists on filtering values of radiation that are negative or higher than the top of the atmosphere.

Furthermore, in global solar radiation forecasting, it is usual to remove night data and anomalous readings from the equipment at dawn and sunset, or other shadowing effects. Therefore, a filter based on the solar zenith angle (SZA) was introduced to establish a frontier of 80° for all data. This threshold is validated in the bibliography [38], and also concurs with the careful observations of the shadow effect for each station.

## 2.2. Satellite derived data set

Satellite information used in this work was retrieved from the Helioclim 3 database version 5 (HC3v5). All this information has been processed by the Heliosat 2 method, where the images taken from the Meteosat geostationary satellite network are converted into solar radiation in the area of study [39,40].

The selected area contains the entire island of Gran Canaria as well as a significant portion of sea at the north east, motivated by the knowledge and influence of the trade winds in the Canary Islands (see Fig. 1). This area is defined, in decimal degrees, by the coordinates latitude [+28.7500 to +27.2500], and longitude [ -16.0000 to -14.5000], resulting in a grid of  $61 \times 55$  pixels of information, where each pixel possesses a spatial resolution of  $3 \times 3 \text{ km}^2$ . The information provided consists of global horizontal irradiation on top of the atmosphere (ToA). All of these parameters are handled in  $\text{Wh/m}^2$  with a temporal resolution of 15 min, assembling then into an hourly manner in order to be in accordance with the hourly ground data.

An initial assessment of the satellite information shows that the relative Root Mean Square Error between these data and the ground data is coherent. These estimations have an average 12.2% rRMSE at CO\_Pozo Izquierdo and 27.8% rRMSE at C1 Las Palmas. Further assessment on the satellite data values indicates a consistency with the climatic behavior of the island, following a tendency very similar to the ground data values [30].

## 2.3. ECMWF data set

In this work, we used the predicted data provided by the ECMWF (European Centre for Medium Range Weather Forecast), granted by the Laboratoire de Physique et Ingénierie Mathématique pour l'Energie et l'environnement (PIMENT) from the Université de La Réunion.

This information was selected for the island of Gran Canaria for an area located from  $27.5^\circ$  to  $28.5^\circ$  of latitude north, and  $15^\circ$  to  $16^\circ$  longitude west Fig. 2. For each pixel of the selected grid there is a spatial resolution of  $16 \times 16 \text{ km}^2$  and 21 vertical levels of resolution. On the other hand, all the data concerning the ECMWF for the year 2005 comes within 3 h intervals, therefore an interpolation of the

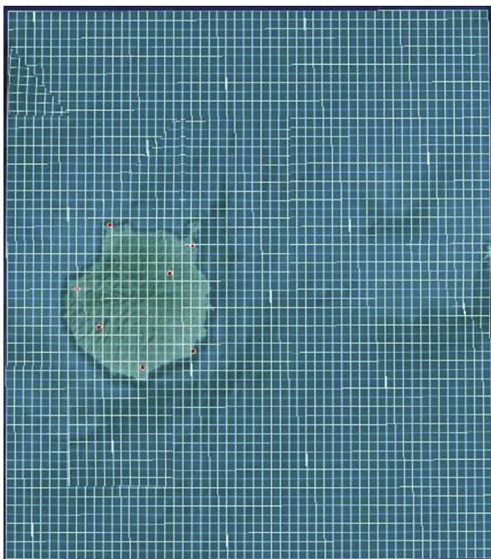


Fig. 1. Geographic distribution of satellite-derived data obtained from HC3v5 for Gran Canaria Island. Courtesy of MINES ParisTech/ARMINES.

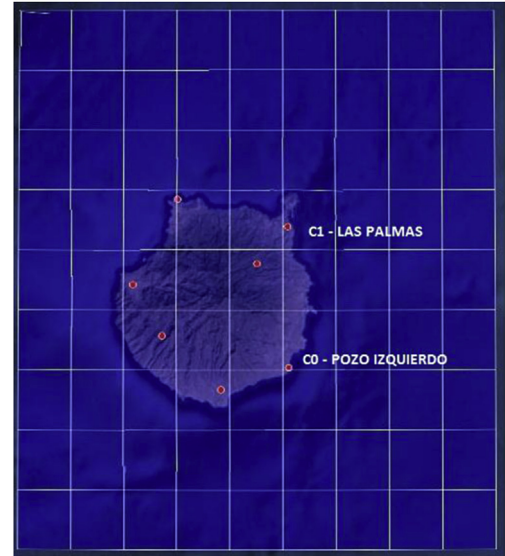


Fig. 2. ECMWF grid for the domain.

values into an hourly basis was necessary to compare data sets. This was achieved using a method that takes into account the conservation of the solar energy [41].

This ECMWF extraction provides a lot of information for a great number of meteorological variables, but in this study we only extracted the following variables described by *Latitude, Longitude* and *Time*:

- Total Cloud Cover (TCC), with values between 0 and 1 using a cloud index,
- Surface Solar Radiation Downwards (SSRD), for accumulative values of  $\text{J/m}^2$  within two instants.

## 3. Data transformation: use of the clear sky index

In order to work with statistical models, all datasets have to be transformed into stationary time series, which mean that auto correlation structures are constant over time [42]. The approach used in this work was the Clear Sky index ( $k^*$ ), extensively applied in the bibliography [13,15], in order to remove the seasonal and daily trend from the hourly time series. This index is obtained by applying the following formula:

$$k^* = \frac{I_g}{I_{cs}} \quad (1)$$

where  $I_g$  is the hourly measured GHI and  $I_{cs}$  is the hourly irradiance computed by a clear sky model for a determined location and time.

Many Clear Sky models have been proposed depending on the different climatic parameters used as inputs [43,44]. This survey is based on the Bird model [45], which is well known to provide accurate results with only a few meteorological inputs [46]: Aerosol Optical Depths (AOD500 nm and AOD380 nm), water vapour and Ozone atmospheric content. AODs and water vapour column were obtained from AERONET network [47]. We obtained data from 2008 to 2014 in order to calculate a climatological annual mean of the atmospheric parameters for all the years. On the other hand, the Ozone was retrieved from the World Ozone Monitoring Mapping provided by the Canadian Government [48] and also aggregated in an annual mean.

An evaluation of the Bird Clear Sky model was made for two measurement stations in Gran Canaria Island: CO Pozo Izquierdo

[27.8175 °N, 15.4244 °E, 47 m] and C1 Las Palmas [28.1108 °N, 15.4269 °E, 17 m]. The assessment of the accuracy of the Bird clear sky model was calculated with the Ineichen method [49], allowing the detection of clear sky hours from the ground dataset in order to compare them with the values given by the clear sky model. This accuracy is quantified by the relative Root Mean Square Error (% rRMSE). The results of the Bird method for the selected locations are listed in Table 1.

The model accuracy for the number of clear sky hours found for each site is around 4% rRMSE. These errors are in the same order as those obtained through the McClear model, which uses AOD, Ozone and water vapour from MACC project [50]. This means that the variability of these atmospheric input parameters can be considered as negligible for the case of Gran Canaria and climatological annual mean values are enough.

This result does not indicate any difference between southern and northern stations, while the number of clear sky hours is clearly higher in Pozo Izquierdo (southern), which is consistent with the behavior in the climate of the island. On the other hand, the variability of each site is given by the method proposed by Hoff and Perez [51], with the standard deviation of the change in the clear sky index, where southern station also takes a lower value compared with northern one, maintaining the coherence with the known climate patterns in the area of study.

### 3.1. Ground data analysis

In order to characterize ground data of each location we decided to classify days in all the available data by mean of intraday clear sky index average, Eq (2), and intraday variability, Eq (3) and (4), as explained in Ref. [26,52]. With these two parameters we could clearly classify the days in nine different frames, see Fig. 3. We divided intraday means into A, B and C days, where A is the heavy clouded type and C data type is the clearer one. In the other hand, the classification by roman numbers I, II and III follow the variability, where a I day stands for a stable variability and a type III describes a great variability through the day. By this distribution an example of a stable and clear day will fit into a type CIII.

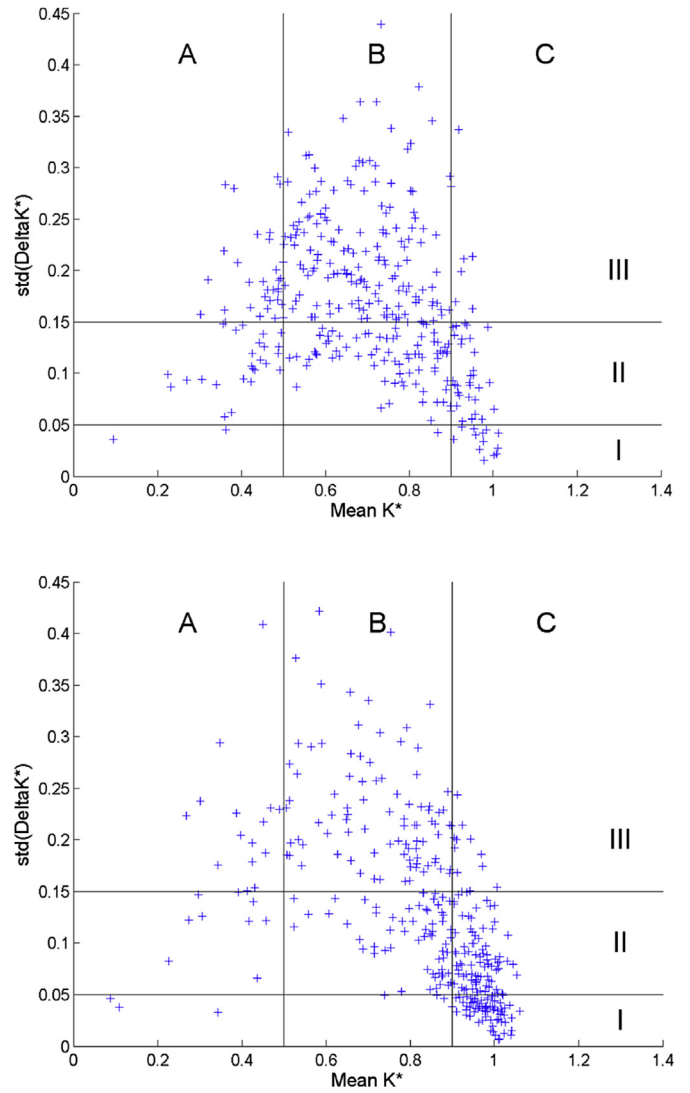
$$\bar{k}^* = \frac{\sum_{t=1}^{H_d} k^*(t)}{H_d} \quad (2)$$

$$\Delta k^*(t) = k^*(t+1) - k^*(t) \quad (3)$$

$$std(\Delta k^*) = \sqrt{\frac{\sum_{t=1}^{H_d} (\Delta k^*(t) - \bar{\Delta k^*})^2}{H_d}} \quad (4)$$

**Table 1**  
Measurement stations characteristics, Bird model inputs and performance.

	C0	Pozo Izquierdo	C1	Las Palmas
Data provider	ITC		ITC	
GHI time step	1 min		1 min	
Period of record	2.003	2.005	2.002	Jan/Jan 2.003 Jul/Dec 2.004 2.005
Average pressure (Pa)	101202		101202	
Ozone Columns (cm)	Average	0.3	0.3	
H2O Columns (cm)	Average	1.815	1.815	
AOD 500 nm	Average	0.159	0.159	
AOD 380 nm	Average	0.184	0.184	
Ba (Asymmetric factor)		0.84	0.84	
Clear sky model accuracy	hourly data	Model	Bird	
		Number of hour of clear sky (Ineichen)	3224	1660
		rRMSE	3.97%	3.98%
Site variability [51]			0.1411	0.1752



**Fig. 3.** Distribution of days in the plane  $\bar{k}^* - std(\Delta k^*)$  of ground dataset for station C1 Las Palmas (up) and station C0-Pozo Izquierdo (down). Intraday mean is divided in A ( $0 < \bar{k}^* < 0.5$ ), B ( $0.5 < \bar{k}^* < 0.9$ ) and C ( $0.9 < \bar{k}^*$ ) days. Daytype variability is divided in I ( $0 < std(\Delta k^*) < 0.05$ ) great variability, II ( $0.05 < std(\Delta k^*) < 0.15$ ) and III ( $0.15 < std(\Delta k^*)$ ) a steady one.

All the data represented in Fig. 3 and on Table 2 show great coherence with the empirical observations of the weather of Gran Canaria Island. There are clearer days in the southern part (C0),

**Table 2**

Distribution of day type categories in Gran Canaria.

	C0	Pozo Izquierdo	C1	Las Palmas
Total A ( $0 < \bar{k}^* < 0.5$ )	27	7.78%	58	16.86%
A I ( $0 < std(\Delta k^*) < 0.05$ )	3		2	
A II ( $0.05 < std(\Delta k^*) < 0.15$ )	9		26	
A III ( $0.15 < std(\Delta k^*)$ )	15		30	
Total B ( $0.5 < \bar{k}^* < 0.9$ )	154	44.38%	234	68.02%
B I ( $0 < std(\Delta k^*) < 0.05$ )	3		1	
B II ( $0.05 < std(\Delta k^*) < 0.15$ )	62		78	
B III ( $0.15 < std(\Delta k^*)$ )	89		155	
Total C ( $0.9 < s$ )	166	47.84%	52	15.12%
C I ( $0 < std(\Delta k^*) < 0.05$ )	71		13	
C II ( $0.05 < std(\Delta k^*) < 0.15$ )	83		32	
C III ( $0.15 < std(\Delta k^*)$ )	12		7	

with 166 days in type C, than the northern one (C1), with 234 falling in the type B class.

#### 4. Forecasting models description

The solar radiation parameter used by the statistical models proposed in this work is the clear sky index. Indeed, the general function used to relate input and output datasets is as follows:

$$\widehat{k}^*(t+h) = F\left(k_G^*(t-i), \dots, k_G^*(t-i), k_{Ex1}^*(t), \dots, k_{Ex1}^*(t-j), \dots, k_{Exn}^*(t), \dots, k_{Exn}^*(t-j)\right) \quad (5)$$

Where  $\widehat{k}^*(t+h)$  is the clear sky index predicted for time horizon  $h$ , from  $h=1$  to  $h=6$  h with hourly granularity,  $k_G^*(t-i)$  is clear sky index from the ground dataset at the location of interest for 'i' past values and  $k_{Exn}^*(t-j)$  corresponds to the different exogenous clear sky indices (inferred from satellite or ECMWF data) for past values. One the most important decision to take into account by the modeler is the number of inputs for the model. The forecasting model may obtain less accurate results if the ANN architecture is unnecessarily complex because of irrelevant inputs.

The dataset for the training process consists in pairs of input output vectors with 'p' number of patterns,  $P = \{X_i, y_i\}_i^p$ . The general function  $F$  is established during the training process. In this work,  $X_i$  matrix contains ground radiation data, satellite derived radiation data and ECMWF predicted parameters, while  $y_i$  vector contains ground solar irradiance for the forecasting time horizon. Clear sky index is the parameter used during the training and testing process, however, results discussed in section 6 will be in terms of global solar irradiances, obtained from Eq. (1) as  $\widehat{I}_g = \widehat{k}^* \cdot I_{cs}$ .

##### 4.1. Naive models

In the literature, it is common practice to compare results obtained by the proposed models against some naive or reference models in order to discuss the improvement. In this study, we have worked with one naive model and a climatological one. The naive model is the so called Smart Persistence (Smart Pers) model [53]. The latter only works with ground historical radiation data from the location of interest, using clear sky index series. It consists in forecasting clear sky index for each time horizon 'h' as the mean of the 'h' previous clear sky values [54]. Smart persistence model is defined by Eq. (6).

$$\widehat{k}^*(t+h) = \text{mean}\left(k^*(t), \dots, k^*(t-h)\right) \quad (6)$$

##### 4.2. Climatological model

The climatological model [14], which is based on the historical mean (see Eq. (7)), provides a constant forecasting whatever the horizon is:

$$\widehat{k}^*(t+h) = \text{mean}\left(k^*(t)\right)_{\text{historical mean}} \quad (7)$$

##### 4.3. Forecasting ANN based model

Artificial Neural Networks (ANNs) are statistical methods capable to establish a relation between an input and output data sets. ANNs are composed by several individual units, called neurons, and connected by weights. These neurons received information from an external parameter or another neuron. During the training process the associated weights are modified to obtain the optimal weight distribution. Usually, each neuron applies an activation function to calculate the correspondent output. In this paper the function applied is the hyperbolic tangent function:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

One of the most common ANN model used in engineering applications is based on the Multilayer Perceptron (MLP). The MLP consists in one layer of inputs, one hidden layer made of several non linear neurons and one output layer with no feedback or lateral connections between them. The output layer is made of one neuron with an identity activation function (see Eq. (9)).

$$\widehat{k}^*(t+h) = \alpha_0 + \sum_{j=1}^h \alpha_j \cdot f\left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} \cdot k^*(t-i)\right] \quad (9)$$

The optimal weights are computed optimizing a cost function. The most common function is the average of square difference between the network output  $\widehat{k}^*(t+h)$  and the desired output, target set 'k\*' (see Eq. (10)). Firstly, a random weight set is established and then, during the training process, are optimized by minimizing a cost function by backpropagation algorithm. The backpropagation training algorithm used to minimize  $E(w)$  is the scale conjugate gradient.

$$E(w) = \frac{1}{2} \sum_{i=1}^N \left(\widehat{k}^*(t+h) - k_i^*\right)^2 \quad (10)$$

To improve the accuracy of the model, one of the most important issues is the network architecture. Too complex ANNs may lead to a good approximation of the training dataset, but poor results when a new test dataset is used (this problem is called overfitting). It is generally recommended to control the model complexity in order to achieve good generalization results [8,32]. Several model complexity control techniques exist and in this paper a Bayesian regularization framework is proposed [17,31,55]. The Bayesian approach considers a probability density function over the weight space and introduces new hyperparameters  $\alpha$  and  $\beta$  to control model complexity. The optimal values correspond to the maximum of the probability density function optimizing the following cost function, Eq. (11).

$$S(w) = \frac{\beta}{2} \sum_{i=1}^N \left(\widehat{k}^*(t+h) - k_i^*\right)^2 + \frac{\alpha}{2} \sum_{i=1}^m w_i^2 \quad (11)$$

As described in Lauret, 2006 [17], to find the optimal hyper parameters  $\alpha$  and  $\beta$  and the optimal weight vector  $w_{MP}$  we followed

an iterative procedure. At the beginning, weights are randomly initialized using a Gaussian distribution and the hyperparameters use small values. The optimal weights sets are computed using the standard scale conjugant gradient algorithm, minimizing  $S(w)$ . Once we obtained the first optimized weights set, hyperparameters are obtained using the Bayesian Framework numerical implementation. These steps are repeated until the regularized error is equal to half the number of data points,  $S(w) = N/2$ .

## 5. ANN architecture optimization

In this study, the NETLAB toolbox developed in Matlab was used for ANNs computation [56]. We focused on improving ANNs forecasting results combining ground data and exogenous data as inputs.

ANN model complexity decision is one of the most important problems to study in ANNs forecasting. In this paper, we focus on number of inputs and hidden units. Techniques such as Master optimization [57] or genetic algorithm [29] can be used to optimize the ANN models. However, in this work, based on our experience [32], we chose to use the Bayesian technique.

The Bayesian framework also offers different possibilities to study the model architecture. Automatic Relevance Determination (ARD) technique assigns a different regularization coefficient to each input and allows determining the most relevant inputs during the training process [32,33]. ARD information of six inputs was studied in order to choose the most relevant for GHI forecasting. The best result for all time horizons and both stations was obtained using 6 inputs of ground measurements inputs. Fig. 4 shows an example with a high relevance for all six inputs in CO station.

Once the number of ground lags inputs is fixed, Bayesian framework is used for choosing the number of hidden units by computing the probability of each model (evidence of the model) [30]. The highest evidence of the model is used to select the best option for each station and time horizon [8,31,58]. Most of results show low number units as we can see in Fig. 5 for station CO and time horizon 3 h. However the difference between %rRMSE and %rMAE for the testing dataset in each experiment were not so pronounced.

To avoid overfitting we used %rRMSE and %rMAE obtained on test datasets to decide the final result. Bayesian framework has

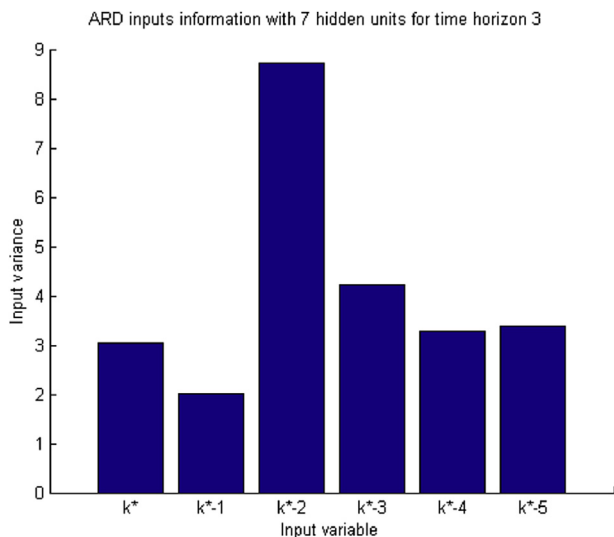


Fig. 4. ARD results using 6 h ground measurement lag data as inputs in ANNs at station CO Pozo Izquierdo for time horizon  $h = 3$  h.

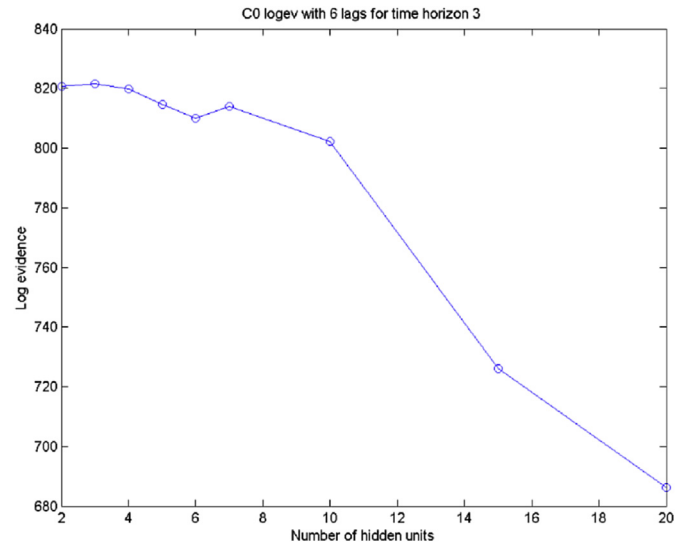


Fig. 5. Logev information using 6 h ground measurement lag data and satellite data as inputs in ANNs at station CO Pozo Izquierdo for time horizon  $h = 3$  h.

been widely used and tested for forecasting [32] obtaining very good results to avoid overfitting. In this case, Fig. 6 shows the final result obtained for training and testing datasets forecasting using classic ANN training and Fig. 7 shows the results obtained with Bayesian Framework, where overfitting problem is not present.

## 6. Selection of the satellite data inputs

ANNs have been trained using different combination of input data. Every ANNs forecast hourly GHI with six past ground clear sky data as inputs. The exogenous data added to the ground clear sky data are solar radiation data derived from Helioclim 3 and solar radiation and total cloud cover (TCC) predictions from ECMWF. Finally, in addition to the Smart persistence and climatology models, the ANN models computed and compared in this paper are the following:

- ANN model with only past ground data as inputs (denoted herein NN).
- ANN model with past ground data and satellite data as inputs (denoted herein NN + SAT).
- ANN model with ground data and ECMWF radiation data as inputs (denoted herein NN + ECMWF).
- ANN model with ground data, satellite data and ECMWF radiation data as inputs (denoted herein NN + ECMWF + SAT).

One of the most important decisions is to select the best satellite information to improve GHI hourly forecasting. We considered a maximum number of 30 satellite derived radiation data based on the calculation explained by Mazorra et al. [30]. In order to study weather conditions in Gran Canaria Island and establish the best satellite pixels between whole gridded data at the different stations, we computed the Pearson correlation between satellite and ground data [27,28]. The huge amount of satellite derived data makes the computation difficult, so we applied a median filter for each  $3 \times 3$  satellite pixels. Consequently, a Super pixel was created computing GHI median value of every  $3 \times 3$  group of pixels, Fig. 8.

In each station the correlation between ground data at  $t = 0$  and each satellite derived pixel with time lags from the same temporal period to a maximum of 3 h earlier, Eq. (12), was computed. In that way we establish a relation between gridded

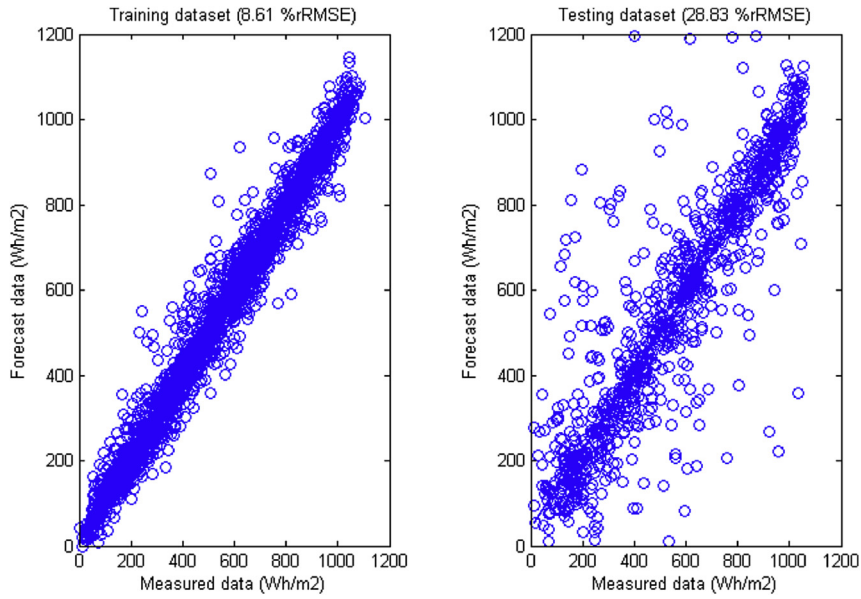


Fig. 6. Measured data vs Forecasted data for the training (left) and testing (right) datasets using classic ANN for station C0.

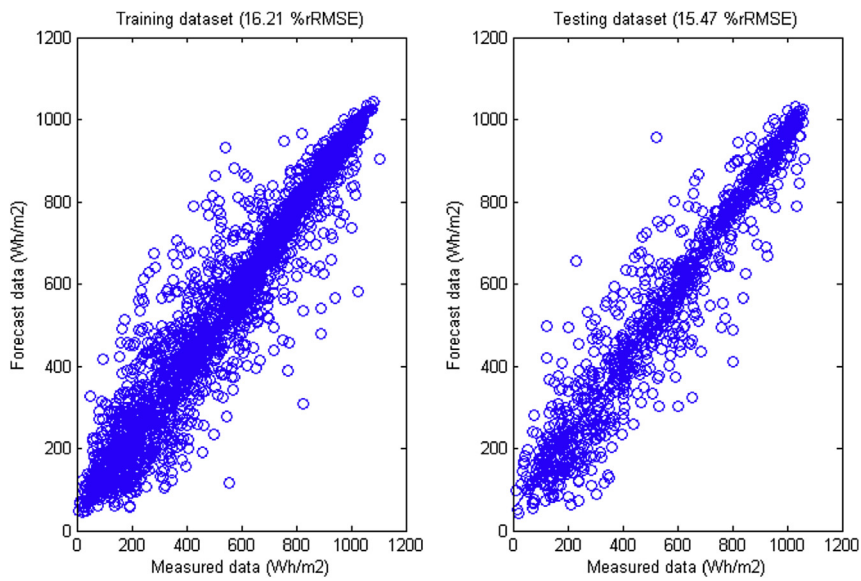


Fig. 7. Measured data vs Forecasted data for the training (left) and testing (right) datasets using Bayesian Framework for station C0.

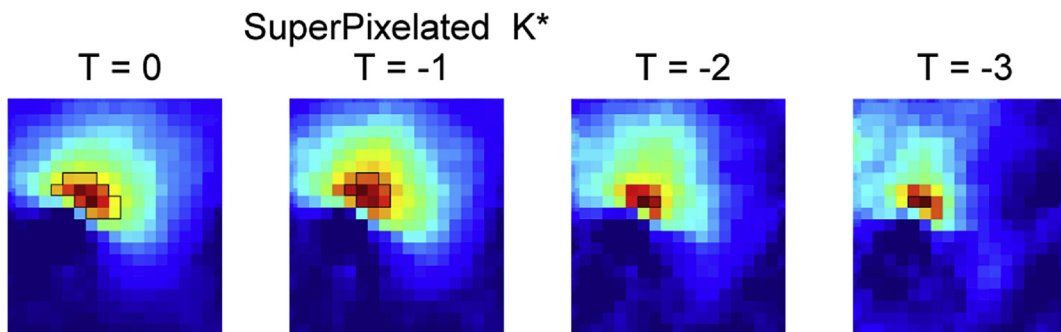


Fig. 8. Superpixel ( $3 \times 3$ ) selection at station C1-Las Palmas using Test-1 for time-lagged correlation images,  $t = 0, 1, 2$  &  $3$  h. Black area shows selected superpixels. Each pixel represents the intercorrelation with  $k^*$  between ground measurement and each satellite pixel.



pixels for different time lags and the ground data for the present moment. Therefore, we have information about incoming events from the surroundings.

$$C_{k^*}(i,j)_h = \text{corr}(k_{\text{ground}}^*(t), k_{\text{satellite}}^*(t-h)) \quad \text{for } h = 0, 1, 2 \& 3 \quad (12)$$

Thus, ANNs inputs represent the surrounding pixels with highest relate with ground station data for different time lags. As explained in Ref. [30], the distribution of pixels in the different time lagged images is according to the number of correlation values over 0.5 in each station [28]. To decide the best selection of pixel, we computed six different tests during the training process. As explained in Ref. [30], the distribution of pixels in the different time lagged images is according to the number of correlation values over 0.5 in each station [28]. For each station, four tests were calculated using the annual correlation results and two more test used a different distribution every quarterly group of images. The best was selected using the lowest %rRMSE and %rMAE for the testing dataset.

## 7. Performance error metrics

In order to evaluate the performance of each method, we used two standard error metrics widely used in the solar forecasting community: the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) [59,60]. Dividing both absolute error by the average of the hourly GHI data we compute their relative metrics (%rRMSE and %rMAE). In this paper, to study the quality of the forecasting methods, we provide the relative errors. To compare the relative improvement of the different models with respect to the persistence simple model, we calculate the forecast skill parameter [61].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (GHI_{\text{forecast},i} - GHI_{\text{measured},i})^2} \quad (13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |GHI_{\text{forecast},i} - GHI_{\text{measured},i}| \quad (14)$$

$$SKILL(\%) = 1 - \frac{RMSE_{\text{method}}}{RMSE_{\text{persistence}}} \times 100 \quad (15)$$

## 8. Forecasting results

Once we decided the optimal satellite derived pixels, the different ANN performance results were compared. Tables 3–8 present the forecast performance for both station in terms of %rRMSE, RMSE and %SKILL from time horizon 1 to 6 using annual

**Table 3**  
RMSE for testing dataset at station C0 Pozo Izquierdo for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	92.47	124.64	140.10	144.44	141.50	138.69
CLI	145.13	145.17	145.25	145.33	145.42	145.50
NN	88.20	113.39	125.13	127.12	131.68	130.03
NN + ECMWF	86.40	112.93	117.70	116.94	120.01	120.88
NN + SAT	<b>83.58</b>	<b>104.58</b>	113.51	116.15	<b>117.31</b>	120.60
NN + ECMWF + SAT	84.00	106.17	<b>110.51</b>	<b>114.93</b>	118.89	<b>120.43</b>

**Table 4**  
%rRMSE for testing dataset at station C0 Pozo Izquierdo for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	17.03	22.95	25.80	26.60	26.05	25.54
CLI	26.72	26.73	26.74	26.76	26.78	26.79
NN	16.24	20.88	23.04	23.41	24.25	23.94
NN + ECMWF	15.91	20.79	21.67	21.53	22.10	22.26
NN + SAT	<b>15.39</b>	<b>19.26</b>	20.90	21.39	<b>21.60</b>	22.21
NN + ECMWF + SAT	15.47	19.55	<b>20.35</b>	<b>21.16</b>	21.89	<b>22.17</b>

**Table 5**  
%SKILL for testing dataset at station C0 Pozo Izquierdo for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	0.00	2.65	6.53	14.06	19.68	21.76
CLI	-56.94	-13.38	3.09	13.53	17.46	17.91
NN	4.68	11.51	16.54	24.32	25.15	26.51
NN + ECMWF	6.62	11.86	21.49	30.38	31.78	31.68
NN + SAT	<b>9.67</b>	<b>18.38</b>	24.29	30.85	<b>33.31</b>	31.84
NN + ECMWF + SAT	9.16	17.08	<b>26.27</b>	<b>31.62</b>	32.51	<b>32.06</b>

**Table 6**  
RMSE for testing dataset at station C1 Las Palmas for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	118.95	169.11	190.69	195.34	190.21	182.18
CLI	163.64	163.70	163.78	163.86	163.95	164.03
NN	110.63	143.89	157.06	162.11	162.09	162.88
NN + ECMWF	110.32	139.15	148.96	149.10	148.03	148.30
NN + SAT	105.34	136.59	147.31	151.94	156.71	157.03
<b>NN + ECMWF + SAT</b>	<b>104.75</b>	<b>134.37</b>	<b>142.82</b>	<b>145.41</b>	<b>147.31</b>	<b>147.88</b>

**Table 7**  
%rRMSE for testing dataset at station C1 Las Palmas for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	27.42	38.98	43.96	45.03	43.85	42.00
CLI	37.72	37.74	37.76	37.77	37.79	37.81
NN	25.50	33.17	36.21	37.37	37.37	37.55
NN + ECMWF	25.43	32.08	34.34	34.37	34.12	34.19
NN + SAT	24.28	31.49	33.96	35.03	36.13	36.20
<b>NN + ECMWF + SAT</b>	<b>24.15</b>	<b>30.98</b>	<b>32.92</b>	<b>33.52</b>	<b>33.96</b>	<b>34.09</b>

**Table 8**  
%SKILL for testing dataset at station C1 Las Palmas for different time horizons (in bold the best result).

Model	1 h	2 h	3 h	4 h	5 h	6 h
SMART-PERS	0.00	-1.24	2.29	8.46	15.35	20.16
CLI	-37.57	1.99	16.08	23.21	27.04	28.11
NN	7.16	14.02	19.63	24.04	27.80	28.51
NN + ECMWF	7.42	16.85	23.78	30.14	34.06	34.91
NN + SAT	11.44	18.23	24.52	28.80	30.26	31.18
<b>NN + ECMWF + SAT</b>	<b>11.94</b>	<b>19.55</b>	<b>26.81</b>	<b>31.86</b>	<b>34.44</b>	<b>35.19</b>

testing data set. Each row corresponds to the forecasting methods used in this paper. In all cases, forecasting models show worse %rRMSE error as time horizon increases. The Smart Persistence in creases from 15%rRMSE in 'h 1' to 26%rRMSE in 'h 6' at C0 station, while at C1 station error results oscillate between 27% and 42%. On the other hand, all models based on ANNs architecture lead to error results around 22% at C0 and 34% at C1 for time horizon 'h 6'. Different error results, almost 12% difference for 'h 6', have been

observed between northern and southern stations. This result is consistent with cloud formation processes explained and observed in the north. Furthermore, it is also observed a dichotomy between the improvements of non linear methods depending on the station and sky conditions. At station C0, ANN models show an improvement between 1.5% and 4% with Smart Persistence model for 'h = 1' and 'h = 6'. While at C1, ANN models lead to a gain from 3% to 6%.

Models with exogenous data improve on ANN models based only on ground data (NN) for all cases and at both stations. Figs. 9 and 10 show the influence of satellite radiation data and ECMWF radiation and TCC data in terms of %rRMSE for both stations.

At station C0, NN + SAT obtain the best results for the first 3 h ahead compared to NN + ECMWF (improvement over 1%). This results are consistent with the fact that satellite data comes from time lags from 'h = 0' to 'h = 3' hours. In the last 3 h, NN + ECMWF is similar to the satellite data (NN + SAT). The combination of both exogenous data, NN + ECMWF + SAT, leads to the best error results in terms of %rRMSE in general.

At station C1, as well as C0, satellite data achieve better results for the first 3 h, while ECMWF improve almost 2% satellite data for 'h = 6'. The combination of both exogenous data (Satellite and ECMWF) leads to the best error results for all time horizons.

Tables 5 and 8 present the %SKILL for all time horizons in both stations. Even if in southern stations, the results with ANN + ECMWF + SAT are much better than in the north, the SKILL parameter shows very similar results for both areas. Furthermore, the SKILL forecast increases with time horizon, which means that the more far ahead in time, the better results we get with ANN + ECMWF + SAT method compared with persistence model. Moreover, Figs. 11 and 12 show the results for both stations in terms of %SKILL. It is easily observed that in station C0 the satellite data offer the best results for the first 3 h, while both models obtained very similar results for the last 3 h. In station C1, NN + SAT model obtain the best results from 'h = 1' to 'h = 3' while NN + ECMWF is the best model from 'h = 4' to 'h = 6'. For both stations, overall, the

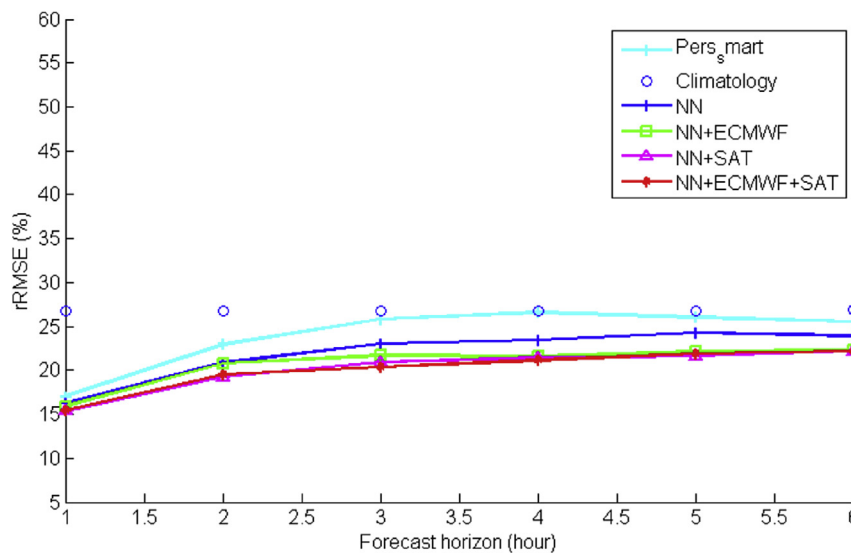


Fig. 9. %rRMSE for testing dataset at station C0 - Pozo Izquierdo for different time horizons using ANN models with satellite and ECMWF radiation data.

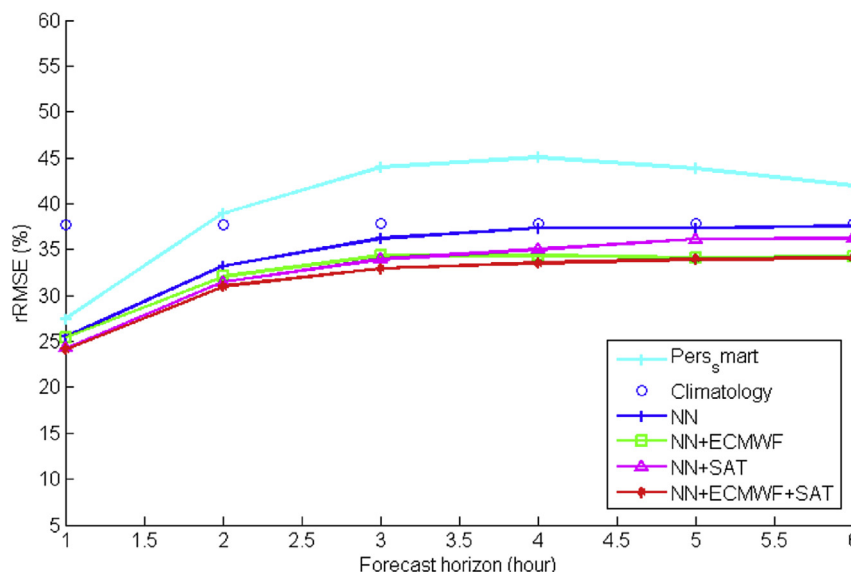


Fig. 10. %rRMSE for testing dataset at station C1-Las Palmas for different time horizons using ANN models with satellite and ECMWF radiation data.

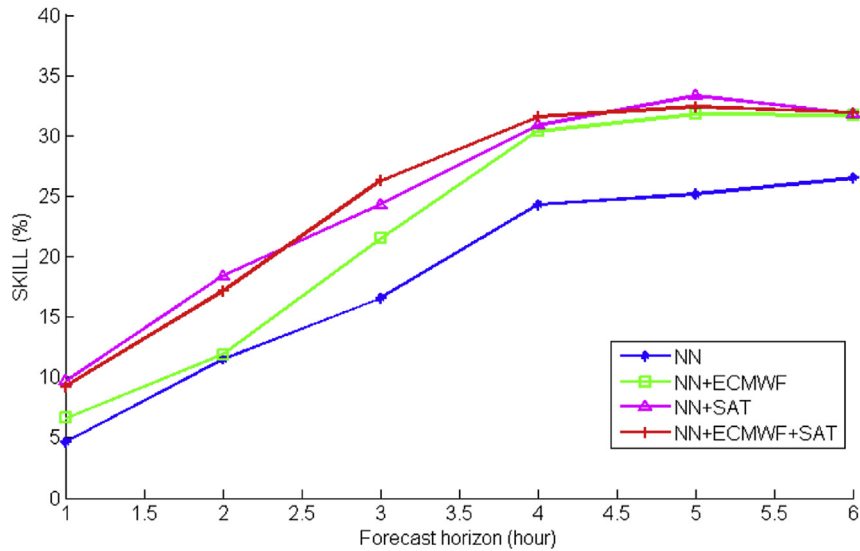


Fig. 11. %SKILL for testing dataset at station C0 Pozo Izquierdo for different time horizons using ANN models with satellite and ECMWF radiation data.

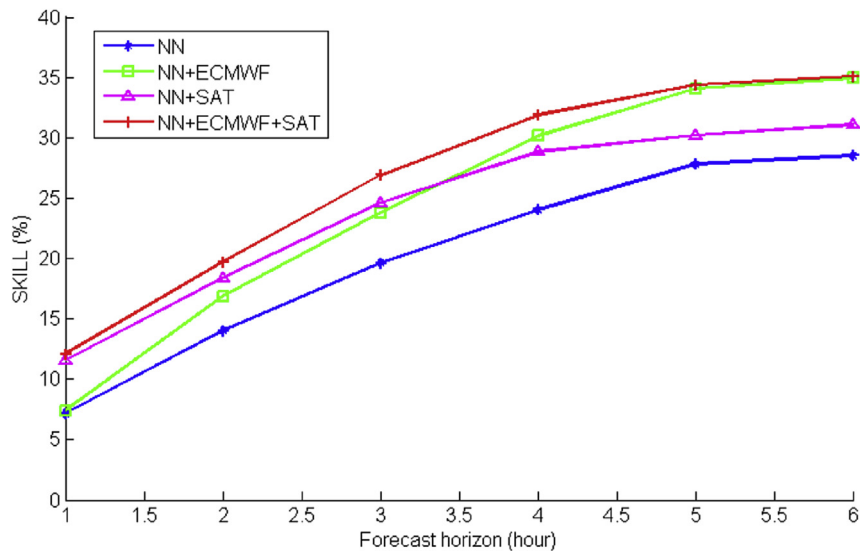


Fig. 12. %SKILL for testing dataset at station C1-Las Palmas for different time horizons using ANN models with satellite and ECMWF radiation data.

best model is the combination of both inputs, ECMWF and satellite data, NN + ECMWF + SAT.

As a consequence, the model recommended for the whole Gran Canaria Island is NN + ECMWF + SAT. This model improves NN model (only with ground data) from 1% for 'h 1' to almost 2% for 'h 6' at station C0 and around 1.5% for 'h 1' to almost 3.5% for 'h 6' at station C1.

## 9. Conclusions

This work proposes to combine ground measurements with exogenous inputs provided by satellite and NWP data in order to improve intra day solar forecasting.

In order to explain the influence of each exogenous data on forecasts, satellite data and data from ECMWF were first studied separately and in combination. At Station C0, for the first 3 h forecast, inputting satellite data (NN + SAT) obtained better results compared to inputting data from ECMWF (NN + ECMWF). However, in the last 3 h, similar results were obtained by including

ECMWF data. In this case, the best results were obtained by combining both models (NN + ECMWF + SAT) as compared with NN. In Station C1, this same effect is more clearly observable. In the first 3 h, the best model is the one that uses satellite data (NN + SAT) and in the last hours the best model is the ECMWF data model (NN + ECMWF). Similarly, the best results were obtained by combining these models (NN + ECMWF + SAT). The conclusion is that, in general combining exogenous satellite data and ECMWF data provides the best forecast results for both stations.

We would recommend working with the NN + ECMWF + SAT model on the entire island. With this model, at the C0 Pozo Izquierdo station, %rRMSE increases with the forecast horizon from 15.47% for 'h 1' to 22.17% for 'h 6', and for C1 Las Palmas, it fluctuated between 24.15% and 34.09%.

More work has to be done in order to cope with all meteorological situations. For instance, with more training data or by implementing a specific procedure (random sampling of the training cases), one can create a model ensemble that may be able to cope with the different meteorological conditions.

Lastly, future lines of research should carry out a more effective selection of satellite pixels. Similarly, wind and humidity would give us very important information about solar radiation behavior. We propose carrying out a detailed study of wind and humidity in the various layers of the atmosphere in order to determine which layer offers the best results and relates these information with satellite pixels.

## Acknowledgments

This work has been supported by the “Convocatoria 2014 – Proyectos I + D + I – Programa Estatal de Investigación, Desarrollo e Innovación orientada a los retos de la sociedad” of the Spanish Government, grant contracts CTM2014 55014 C3 1 R, as part of the project “Integración de nuevas metodologías en simulación de campos de vientos, radiación solar y calidad del aire” and by the “Cátedra Endesa Red de la Universidad de Las Palmas de Gran Canaria en su Convocatoria del curso 2014 2015”. The authors are very grateful to Canary Islands Technology Institute (Instituto Tecnológico de Canarias, I.T.C.) for providing ground measurements databases and to Dr. Philippe Blanc, Responsable des activités de recherche sur l'évaluation des ressources énergétiques renouvelables of MINES ParisTech/ARMINES, for his help with satellite databases.

## References

- [1] H. Kambezidis, B. Psiloglou, B. Synodinou, Comparison between measurements and models for daily solar irradiation on tilted surfaces in Athens, Greece, *Renew. Energy* 10 (4) (1997) 505–518.
- [2] G. Notton, C. Critofari, P. Poggi, Performance evaluation of various hourly slope irradiation models using Mediterranean experimental data of Ajaccio, *Energy Convers. Manag.* 47 (2) (2006) 147–173.
- [3] M. Diez-Mediavilla, A. de Miguel, J. Bilbao, Measurement and comparison of diffuse solar irradiance models on inclined surfaces in Valladolid (Spain), *Energy Convers. Manag.* 46 (13–14) (2005) 2075–2092.
- [4] D. Heinemann, E. Lorenz, M. Girodo, Forecasting of Solar Radiation. Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale, Nova Science Publishers, New York, 2006.
- [5] M. Wittmann, H. Breitzkreuz, M. Schroedter-Hoscheidt, M. Eck, Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants, *Sel. Top. Appl. Earth Observ. Remote Sens. IEEE J.* 1 (1) (2004) 18–27.
- [6] F. Díaz, G. Montero, J. Escobar, E. Rodríguez, R. Montenegro, An adaptive solar radiation numerical model, *J. Comput. Appl. Math.* 236 (18) (2012) 4611–4622.
- [7] M. Diagne, M. David, J. Boland, N. Schmutz, P. Lauret, Post-processing of solar irradiance forecasts from WRF model at Reunion Island, *Sol. Energy* 105 (2014) 99–108.
- [8] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, 1995.
- [9] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renné, T. Hoff, Validation of short and medium term operational solar radiation forecasts in the US, *Sol. Energy*, 84 (2010) 2161–2172.
- [10] A. Hammer, D. Heinemann, E. Lorenz, B. Lücke, Short-term forecasting of solar radiation: a statistical approach using satellite data, *Sol. Energy* 67 (1) (1999) 139–150.
- [11] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S. Müller, W. Traunmüller, Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe, *Sol. Energy* 94 (2013) 305–326.
- [12] V. Kostylev, A. Pavlovski, Solar power forecasting performance toward industry standards, in: *Proceedings of the 1st International Workshop on the Integration of Solar Power into Power Systems*, October 24, Aarhus, Denmark, 2011.
- [13] M. Sengupta, A. Habte, S. Kurtz, A. Dobos, S. Wilbert, E. Lorenz, T. Stoffel, D. Renné, C. Gueymard, D. Myers, S. Wilcox, P. Blanc, P. R., Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications, Technical Report NREL/TP-5D00 63112 Contract No. DE-AC36-08GO28308, 2015.
- [14] E. Lorenz, D. Heinemann, Prediction of solar irradiance and photovoltaic power, in: *Comprehensive Renewable Energy*, Elsevier, Oxford, 2012, pp. 239–292.
- [15] M. Diagne, M. David, P. Lauret, J. Boland, Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.* 27 (2013) 65–76.
- [16] L. Hontoria, J. Aguilera, P. Zufiria, Generation of hourly irradiation synthetic series using the neural network multilayer perceptron, *Sol. Energy* 72 (5) (2002) 446–2002.
- [17] P. Lauret, M. David, E. Fock, A. Bastide, C. Riviere, Bayesian and Sensitivity analysis approaches to modeling the direct solar irradiance, *Sol. Energy Eng.* 128 (2006) 394–405.
- [18] A. Mellit, A. Pavan, A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy, *Sol. Energy* 84 (2010) 807–821.
- [19] R. Inman, H. Pedro, C. Coimbra, Solar forecasting methods for renewable energy integration, *Prog. Energy Combust. Sci.* 39 (6) (2013) 535–576.
- [20] S. Rehman, M. Mohandes, Artificial neural network estimation of global solar radiation using air temperature and relative humidity, *Energy Policy*, 36 (2) (2008) 571–576.
- [21] J. Bosch, G. Lopez, F. Batlles, Daily solar irradiation estimation over a mountainous area using artificial neural networks, *Renew. Energy* 33 (7) (2008) 1622–1628.
- [22] Y. Kimmoku, S. Orita, S. Nakagawa, T. Sakakibara, Daily insolation forecasting using a multi-stage neural network, *Sol. Energy*, 66 (3) (1999) 193–199.
- [23] A. Sfetsos, A. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, *Sol. Energy*, 68 (2) (2000) 169–178.
- [24] A. Ghanbarzadeh, A. Noghrehabadi, E. Assreh, M. Behrang, Solar radiation forecasting based on meteorological data using artificial neural networks, in: *7th IEEE International Conference on Industrial Informatics, INDIN, 2009*.
- [25] M. Mohandes, S. Rehman, T. Halawani, Estimation of global solar radiation using artificial neural networks, *Renew. Energy*, 14 (1) (1998) 179–184.
- [26] R. Dambreville, Prevision du rayonnement solaire global par teledection pur la gestion de la production d'énergie photovoltaïque, Grenoble, 2014. These de l'Université de Grenoble.
- [27] R. Dambreville, P. Blanc, J. Chnussot, D. Boldo, Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model, *Renew. Energy*, 72 (2014) 291–300.
- [28] A. Zagouras, H. Pedro, C. Coimbra, On the role of lagged exogenous variables and spatio-temporal correlations in improving the accuracy of solar forecasting methods, *Renew. Energy*, 78 (2015) 203–2018.
- [29] R. Marquez, H. Pedro, C. Coimbra, Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs, *Sol. Energy*, 92 (2013) 176–188.
- [30] L. Mazorra Aguiar, B. Pereira, M. David, F. Diaz, P. Lauret, Use of satellite data to improve solar radiation forecasting with Bayesian artificial neural networks, *Sol. Energy*, 122 (2015) 1309–1324.
- [31] D. Mackay, *Information Theory, Inference, and Learning Algorithms*, Cambridge university Press, Cambridge, 2003.
- [32] P. Lauret, E. Fock, R. Randrianarivony, J. Manicom-Ramsamy, Bayesian neural network approach to short time load forecasting, *Energy Convers. Manag.* 49 (5) (2008) 1156–1166.
- [33] J. Bosch Saldana, Modelización del recurso solar utilizando redes neuronales artificiales y su aplicación a la generación de mapas topográficos de radiación, Almería, 2010. Tesis Doctoral. Universidad de Almería.
- [34] ITC Canarias, *Mapasolarcanarias*, 2011 [En línea]. Available: <http://mapasolar.itccanarias.org/mapasolarcanarias/>.
- [35] L. Mazorra, F. Díaz, G. Montero, R. Montenegro, Typical meteorological year (TMY) evaluation for power generation in gran Canaria island, Spain, in: *Proceedings of the 25th European Photovoltaic Solar Energy Conference and Valencia*, 2010.
- [36] E. Maxwell, S. Wilcox, M. Rymes, Users Manual for SERI QC Software, Assessing the Quality of Solar Radiation Data, NREL-TP-463 5608, National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO, 1993.
- [37] S. Younes, T. Muneer, Clear-sky classification procedures and models using a world-wide data-base, *Appl. Energy* 84 (2007) 623–645.
- [38] P. Blanc, Atlas du Potentiel Solaire Photovoltaïque et Thermodynamique en region PACA, Centre Energetique et Procédes, MINES ParisTech/ARMINES, Conseil General Alpes Maritimes, 2011.
- [39] C. Rigollier, M. Lefèvre, L. Wald, The method Heliosat-2 for deriving shortwave solar radiation from satellite images, *Sol. Energy* 77 (2004) 159–169.
- [40] P. Blanc, B. Gschwind, M. Lefèvre, L. Wald, The HelioClim project: surface solar irradiance data for climate applications, *Remote Sens.* 3 (2) (2011) 343–361.
- [41] B. Espinar, L. Wald, P. Blanc, C. Hoyer-Klick, M. Schroedter-Hoscheidt, R. Wanderer, Report on the Harmonization and Qualification of Meteorological Data: Project ENDORSE, Energy Downstream Service, Providing Energy Components for GMES Grant Agreement No.262892, Paris, France: Armines, 2011.
- [42] C. Chatfield, *Time Series Analysis, an Introduction*, Chapman&Hall, 2004.
- [43] M. Reno, C. Hansen, J. Stein, Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis, SANDIA Report SAND2012 2389, 2012.
- [44] S. Younes, R. Claywell, T. Muneer, Quality control of solar radiation data: present status and proposed new approaches, *Energy* 30 (2005) 1533–1549.
- [45] R. Bird, R. Hulstrom, Simplified the Clear Sky Moel for Direct and Diffuse Insolation on Horizontal Surfaces, Technical Report No. SERI/TR-642-761, Solar Energy Research Institute, 1981.
- [46] V. Badescu, C. Gueymard, S. Cheva, C. Oprea, M. Baci, A. Dumitrescu, F. Iacobescu, I. Milos, C. Rada, Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania, *Renew. Energy*, 55 (2013) 85–103.
- [47] B.N. Holben, T.F. Eck, I. Slutsker, D. Tanré, J.P. Buis, A. Setzer, E. Vermote, J.A. Reagan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, A. Smirnov, AERONET a federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.* 66 (1998) 1–16.

- [48] Canada's Environment, World Ozone Monitoring Mapping, 2015. Available: <http://es-ee.tor.ec.gc.ca/e/ozone/ozoneworld.htm>.
- [49] P. Ineichen, Comparison of eight clear sky broadband models against 16 independent data banks, *Sol. Energy* 80 (2006) 468–478.
- [50] M. Lefevre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald., McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, *Atmos. Meas. Tech.* 6 (2013) 2403–2418.
- [51] T. Hoff, R. Perez, Quantifying PV power output variability, *Sol. Energy*, 84 (2010) 1782–1793.
- [52] P. Lauret, R. Perez, L. Mazorra-Aguiar, E. Tapachés, H. Diagne, M. David, Characterization of the intraday variability regime of solar irradiation of climatically distinct locations, *Sol. Energy*, 125 (2016) 99–110.
- [53] R. Perez, A. Kankiewicz, J. Schlemmer, K. Hemker, S. Kivalov, A new operational solar resource forecast model service for PV fleet simulation, in: *Photovoltaic Specialist Conference (PVSC) IEEE 40th*, 2014.
- [54] T. Hoff, R. Perez, Modeling PV fleet output variability, *Sol. Energy* 86 (2012) 2177–2189.
- [55] P. Lauret, E. Fock, T. Mara, A node pruning algorithm based on the Fourier amplitude sensitivity test method, *IEEE Trans. Neural* 17 (2) (2006) 273–293.
- [56] I. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Springer, London, 2002.
- [57] C. Cornaro, M. Pierro, F. Bucci, Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast, *Sol. Energy* 111 (2015) 297–312.
- [58] W. Penny, S. Roberts, Bayesian Neural Networks for classification: how useful is the evidence framework? *Neural Netw.* 12 (1999) 877–892.
- [59] P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi, A benchmarking of machine learning techniques for solar radiation forecasting in an insular context, *Sol. Energy* 112 (2015) 446–457.
- [60] C. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.* 30 (2005) 79–82.
- [61] C. Coimbra, J. Kleissl, R. Marquez, Overview of solar-forecasting methods and a metric for accuracy evaluation, in: *Solar Energy Forecasting and Resource Assessment*, Elsevier, 2013, pp. 171–194.