Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 84 (2010) 1772–1781

SOLAR ENERGY

www.elsevier.com/locate/solener

Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning

Luis Martín [b,*], Luis F. Zarzalejo [a], Jesús Polo [a], Ana Navarro [a], Ruth Marchante [b], Marco Cony [b]

[a] *Energy Department, CIEMAT, Madrid 28040, Spain*
[b] *Investigaciones y Recursos Solares Avanzados (IrSOLaV), Tres Cantos 2 8045, Spain*

Received 16 February 2010; received in revised form 3 June 2010; accepted 3 July 2010
Available online 31 July 2010

Communicated by: Associated Editor Christian Gueymard

Abstract

Due to strong increase of solar power generation, the predictions of incoming solar energy are acquiring more importance. Photovoltaic and solar thermal are the main sources of electricity generation from solar energy. In the case of solar thermal energy plants with storage energy system, its management and operation need reliable predictions of solar irradiance with the same temporal resolution as the temporal capacity of the back-up system. These plants can work like a conventional power plant and compete in the energy stock market avoiding intermittence in electricity production.

This work presents a comparisons of statistical models based on time series applied to predict half daily values of global solar irradiance with a temporal horizon of 3 days. Half daily values consist of accumulated hourly global solar irradiance from solar raise to solar noon and from noon until dawn for each day. The dataset of ground solar radiation used belongs to stations of Spanish National Weather Service (AEMet). The models tested are autoregressive, neural networks and fuzzy logic models. Due to the fact that half daily solar irradiance time series is non-stationary, it has been necessary to transform it to two new stationary variables (clearness index and lost component) which are used as input of the predictive models. Improvement in terms of RMSD of the models essayed is compared against the model based on persistence. The validation process shows that all models essayed improve persistence. The best approach to forecast half daily values of solar irradiance is neural network models with lost component as input, except Lerida station where models based on clearness index have less uncertainty because this magnitude has a linear behaviour and it is easier to simulate by models.

Keywords: Energy meteorology; Solar radiation; Solar radiation forecasting; Solar thermal energy; Clearness index; Lost component

1. Introduction

Although considerable effort has been done to make use of solar energy efficiently from industrial revolution expecting fossil fuels would run out in the future, only minimal resources have been directed towards forecasting incoming solar energy at ground level (Veziroglu and dot, 2008). However, the necessity to have forecasting models which could optimize the integration of solar energy into electric grid is increasing as they gain recognition as an energetic source.

Currently, the potential market of solar energy is huge. Its development is being supported by agreements in Kyoto protocol and by progressive series of regulations regarding green energy (feed-in tariff) established in several countries like Spain and Germany (Patlitzianas et al., 2007). In the case of Spain, current legislation (Royal Decree 661/2007 25th of May) allows promoters minimize investment risks and contribute to the development of solar energy.

* Corresponding author. Tel.: +34 677740918; fax: +34 913466037.
E-mail address: luis.martin@irsolav.com (L. Martín).

0038-092X/$ - see front matter
doi:10.1016/j.solener.2010.07.002

Spanish legislation allows solar energy operators to choose between two possibilities to sell electricity production in the energy free market:

- *The tariff model* operator receives a fixed price for the electricity they produce independent on the time of the day. Renewable energy promoters have a guaranteed to feed-in all the electricity produced, to ensure economic attractiveness of the projects.
- *The premium model* where promoters receive the sum of electricity price negotiated in energy free market plus an additional prime to compensate the competitiveness of renewable energy production. In this scheme renewable energy producers act as any other conventional plant on the liberalized market.

The participation in the energy stock market is regulated by two basic rules. It is necessary to predict the amount of energy which will be produced, up to 72 h before, and deviations of energy production are strongly penalized.

1.1. Solar thermal energy with storage energy system

Solar thermal power plants (STPP) produce electricity focusing beam solar energy component into receivers to produce heat that moves a thermodynamic cycle. STPP with a storage energy system can work like a conventional power plant and compete in the energy stock market avoiding electricity production intermittence. The optimum size of thermal storage system to minimize the cost of energy production and the price of components is between 6 h and 9 h (Herrmann, 2009). The benefits of thermal energy storage are the following:

- *Profit maximization:* in periods of energy stock market low prices, it is possible to storage heat, and when prices are higher, the plant can run at full capacity even without incoming solar energy.
- *Reducing intermittence:* the production of electricity can be constant without fluctuations. This way, STPPs can work in the electric grid in the same way as a conventional power plant. Besides, components like the turbine suffer less due to the reduction of connections and disconnections.
- *Increasing plant utilization:* the overall annual production can be increased. In periods of high incoming solar energy, the excess of thermal energy can be stored to be used in periods when heat from the solar field is not enough to move the turbine to its nominal power.
- *Peak shaving:* the electricity national consumption peak during midday can be reduced without having to rely on other sources of energy.

A previous study demonstrates that participation of STPP promoters in the *premium model* can be more profitable than the tariff model (Wittmann et al., 2008). This way, to maximize the revenue at the energy stock market when prices are higher, STPP needs to use an optimized operation strategy. The thermal energy storage system is used to shift the STPP electricity production towards peak price periods of electricity in the market. Therefore, scheduling of delivery of electricity production will be defined by daily prices course and the strategy of operation of STPP needs a model to predict stock market electricity prices.

Therefore, the combination of a thermal energy storage system, an optimized strategy operation and a model to obtain reliable predictions of solar irradiance with the same time-step as the temporal capacity of thermal energy storage system is a way to improve profitability, management and operation of STPP.

1.2. Solar irradiance predictions

There are two main groups of models that can be used to obtain predictions of solar irradiance:

- *Physical models:* are based on mathematical equations to describe the physics and dynamic of the atmosphere. These equations don't have a unique solution due to its non-linearity. Therefore, numerical methods obtain approximate solutions and that's way these models are also known as numerical weather predictions models (NWP). Previous works with NWP models have been done only for hourly and daily temporal resolution (Renne, 2009). The errors from these models vary significantly and depend on the climate and dynamic of solar radiation at the location under study. In the case of Spain, there is a previous study which compares results from different NWP models (Lorenz et al., 2009). Mean errors for hourly prediction vary between 20.8% and 31.7% for first day, 21.3% and 36.8% for the second day and 22.4% and 40.9% for the third day of forecast in terms of relative root mean square deviation (rRMSD).
- *Statistical models:* are based on making relations between past observations and future values to predict solar irradiance. Depending on the information they use, it is possible to difference two groups:
  - *Classical models:* these models are defined as classical, reflecting its prominence in the period where information from NWP where unavailable. Currently, they are used for the short-term and medium-long term prediction, where information from NWP is unavailable (Wilks, 1995). These models are less complex than NWP because they need less information and computation time to make predictions. Using a radiometric database to train the models and real-time ground measurements of the location under study it is possible to make solar irradiance predictions. Previous works in this field have been done in daily-time-steps in the USA and Morocco with errors which vary from 16% and 28% in terms of rRMSD (Kemmoku and Nakagawa, 1999; Safi et al., 2002). However, none of these studies have been done in the Iberian Peninsula.

◦ *Statistical downscaling:* these models are another important application of statistical models which are used to improve output from NWP models of some variables or locations not represented with enough precision by the models (Wilks, 1995). The studies with these models are mainly for daily and hourly time-steps. Previous studies shows that for daily solar irradiance, forecasting errors vary between 18% and 25% in terms of rRMSD in the case of the EEUU (Jensenius and Cotton, 1981; Baker and Casper, 1981) and for hourly solar irradiance forecasting errors vary between 30% and 40% in terms of rRMSD (Perez et al., 2007) for Germany and EEUU. The results shown in previous NWP section for the southern of Spain are not considered in this section, because although they apply a post-processing to the results of the NWP models, the methodology they use doesn't make any relation between observed measurements and the variable to predict like in statistical downscaling methods.

Another way to use statistical models is in combination with other instruments used to estimate solar irradiance like meteorological satellites (Hammer et al., 1999) and sky cameras (Casa Nova et al., 2005). Both instruments generate images which can be used to estimate solar irradiance based on the strong impact of cloudiness on ground irradiance. Therefore, the description of temporal development of the cloudy situations is essential to forecast the irradiance. The methods to predict solar irradiance from satellite and sky camera images are based in atmospheric motion vector fields (AMV), neural networks and time series analysis. In the case of satellite images, prediction errors are 10% rRMSD for a forecasting horizon of 30 min; meanwhile, if the prediction horizon is increased to 6 h the error is 25% in term of rRMSD (Hammer et al., 2000).

1.3. Solar irradiance predictions with applications in solar thermal power plants

In this work we present a study of predictability of half daily global solar irradiance time series based on statistical models. The temporal horizon of the predictions is 3 days or six half days. The reasons to choose half daily-time step to make the predictions are the following:

- Daily solar irradiance isn't a suitable forecasting time step to be used in STPP because the temporal capacity of thermal energy storage is between six and nine hours.
- Half daily values consist of two values of accumulated global solar irradiance for each day which gives a high detail of the dynamic characteristic of solar irradiance divided by solar noon.
- There is no previous forecasting study on half daily-time-steps.

Furthermore, the main reason to choose statistical models to generate the predictions is that they are simpler than NWP models. Besides, in the case of daily solar irradiance forecasting time-steps both models have similar accuracy. The statistical models tested in this study are autoregressive models, neuronal network models and fuzzy logic which have been used in other renewable energy applications (Esen et al., 2008a,b; 2009a,b).

Forecasting half daily values of solar irradiance is just a first step. This predictions are used in combination with a thermal energy storage system, an optimized strategy operation of STPP and predictions of energy prices to schedule electricity power production in an hourly basis which is the temporal resolution demanded by legislation. Furthermore, forecasting of half daily values of solar irradiance is of great importance for the management, operation and maintenance of STPP.

2. Solar radiation dataset

The dataset used in this study is based on hourly global solar irradiance ground measurements from stations of the Spanish National Radiometric Network from *Agencia Española de Meteorología* (AEMet). The quality of radiometric data available and temporal acquisition of measurements is two crucial factors related to adjustment of proposed models. In this study an exhaustive quality analysis has been done on both important factors which influence high quality measurements. Manufacturer of pyranometers used is the Dutch firm *Kipp&Zonen* and belong to CM11 series (see Fig. 1).

Table 1 summarizes the main geographic characteristics of the stations as well as the available time period of the ground measurements. Fig. 2 shows the location of the stations in the Iberian Peninsula.

Half daily-time series are constructed from hourly values of solar irradiance and consist in accumulated solar irradiance values in the hours from solar raise until solar noon and from solar noon until solar dawn for each day. Then, for each day there are two values which give a high detail of the dynamic characteristic of solar irradiance divided by solar noon.

3. Methodology

In this work, forecasting methods based on time series analysis have been essayed to predict half daily values of solar irradiance for the next 3 days. Due to non-stationary behaviour of half daily global solar irradiance time series, it has been transformed to two new variables: clearness index (K_T) and lost component (LC). Clearness index is defined as the ratio between ground measured global solar irradiance and extraterrestrial solar irradiance. Lost component is defined as the difference between extraterrestrial solar irradiance and ground measured global solar irradiance.

In the next subsections the different statistical models used to generate the predictions are described briefly.

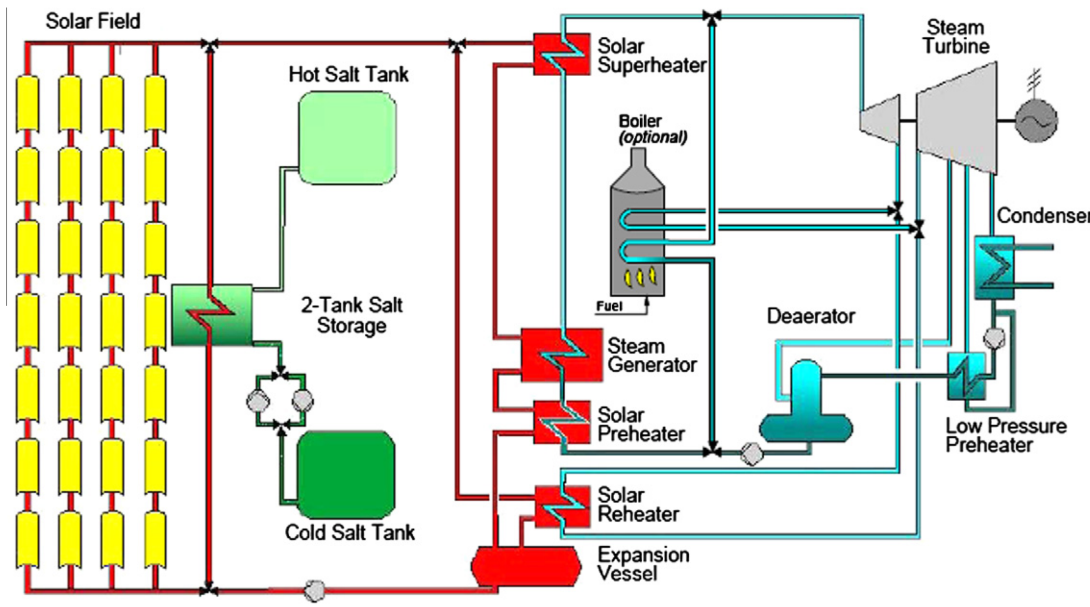

Fig. 1. Schematic configuration of Andasol-1 power plant.

Table 1
Geographic characteristics and time period of the ground data.

Station	Latitude (°)	Longitude (°)	Altitude (m)	Period
Murcia	38.00	−1.17	69	1994–2003
Albacete	39.00	−1.87	674	1996–2003
Madrid	40.45	−3.72	680	1994–2003
Lerida	41.63	0.60	202	1997–2002

3.1. Autoregressive models (AR)

Autoregressive models generate lineal predictions from the input and are also known as Infinite Impulse Response Filter (IIRF). The notation AR(p) refers to the autoregressive model of order p. The AR(p) is expressed by the following equation:

Fig. 2. Geographic location of the radiometric stations.

$$\hat{x}_t = c + \sum_{i=1}^{p} a_i x + \varepsilon_t \qquad (1)$$

where a_i are the parameters of the model, c is a constant which represent the mean value of the time series and ε_t is a white noise signal.

3.2. Neural networks (NN)

Neural networks models are an abstract simulation of neural biological systems which tries to synthetic its abilities (Haykin, 1998).

The most basic processing element is the neuron. A neural network model has a large number of these elements organised in three different kinds of layers: the input, the output and one or more hidden layers (see Fig. 3). The interconnection of each neuron is done through weights. The signal or input vector flows from the input towards the output layer. In each neuron there are two processing elements. First, a function which consists of a weighted sum is applied to the inputs of the neuron. In a second step, the output from the first step is used as input of a nonlinear function (sigmoid (S), tangent sigmoid (ST)) to know if the neuron will be activated or not.

The estimation of the weight parameters of the neural network is done with a gradient descent backpropagation algorithm based on a error correction method (Palit and Popovic, 2005). The training dataset is made of pairs of input and output vectors. The learning algorithm is applied to the neural network in a sequential way; each input vector is used to estimate an output. The error between the output given by the neural network and the theoretical output is used to correct the parameters of the neural network retropropaging the errors from the output layer to the input layer.

The different neural network configurations used to forecast solar irradiance is presented in Table 2.

Table 2
Neural network models tested.

NN Model	Number of layers	Neurons in each layer	Activation function
Model NN1	1	1	Sigmoid (S)
Model NN2	2	3–1	Tangent Sigmoid (ST)-S
Model NN3	3	5–3–1	ST-ST-S
Model NN4	4	8–5–3–1	ST-ST-ST-S
Model NN5	5	15–8–5–3–1	ST-ST-ST-ST-S

3.3. Adaptative-network-based fuzzy inference system (ANFIS)

ANFIS are a class of neural networks which are functionally equivalent to fuzzy logic inference systems. These models have various features that can be used to implement systems that mimic the behaviour of human thinking (Palit and Popovic, 2005).

The neuro-fuzzy model ANFIS (adaptative-network-based fuzzy inference system) proposed by (Jang, 1993), incorporates a five-layer network to implement if-then fuzzy logic rules of Takagi–Sugeno type. Training algorithm of ANFIS models use a combination of least squares estimation and backpropagation for parameter estimation.

4. Results and discussions

Next, the analysis of the results for each station is shown. In a first step, the best or optimum models for each station and each forecasting horizon are analysed. From these results, we have done a selection of a final model for each station. This selection is based on choosing the model with less uncertainty and less number of parameters.

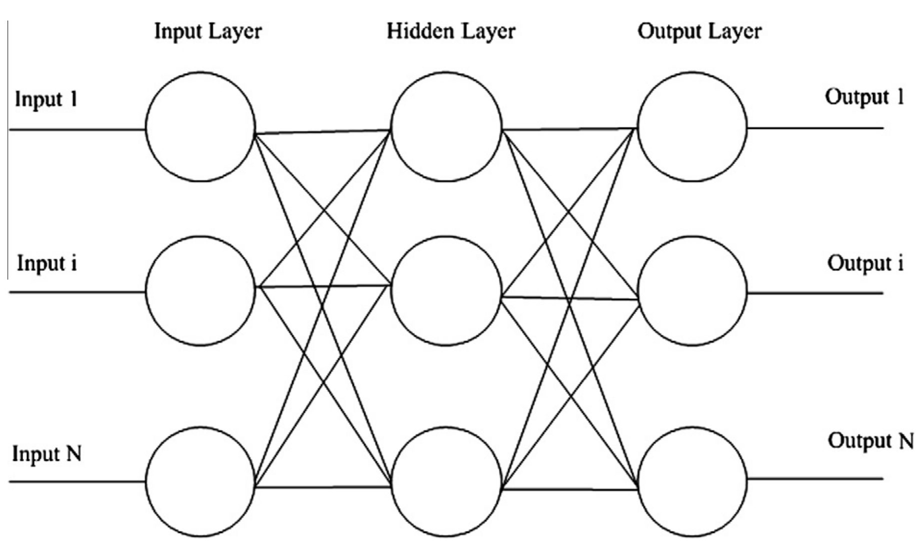

Fig. 3. The back propagation neural network.

The index to measure the errors of the models is the relative root mean squared deviation (rRMSD) and it is calculated with the following expression:

$$\mathrm{rRMSD} = \frac{\sqrt{\sum_{i=1}^{n}(x-\hat{x})^2/n}}{\bar{x}} \tag{2}$$

where N is the population size, x is the observed value, $\bar{x}$ is the mean of observed values and $\hat{x}$ is the predicted value.

The models essayed are compared with a reference model, which is supposed to be the most basic one, with the intention of measuring the improvement we can achieve. The basic model elected is the persistence model (PER) because it is the most extended model to contrast new proposed models. Persistence is based on the assumption that the value for the next temporal step is the same as the present value:

$$\hat{x}_{t+k} = x_t \tag{3}$$

where $\hat{x}_{t+k}$ stands for the prediction for the next k steps, and x_t is the observation at the temporal instant t.

The improvement over persistence in terms of rRMSD is obtained from the following expression:

$$\mathrm{Improvement} = \left(1 - \frac{\mathrm{rRMSD}_m}{\mathrm{rRMSD}_p}\right) \times 100 \tag{4}$$

where rRMSD_m and rRMSD_p are the relative root mean square deviation for the model essayed and for the model based on persistence respectively.

The dataset of each station is divided sequentially into two groups: the training and validation datasets. For each station, the first sequential 75% of the dataset is used to estimate its parameters. The assessment of the models proposed is performed over the last sequential 25% of the dataset.

Table 3
Optimum models for each radiometric station.

Station	Models	Prediction horizon (Half daily)					
		1	2	3	4	5	6
Murcia	AR	K_T AR(10)	K_T AR(9)	K_T AR(9)	K_T AR(9)	K_T AR(1)	K_T AR(1)
	NN	K_T NN3(8)	LC NN2(10)	LC NN4(9)	LC NN4(9)	LC NN4(9)	LC NN2(10)
	ANFIS	LC ANFIS(4)	K_T ANFIS(5)	LC ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)
Albacete	AR	LC AR(8)	LC AR(1)	LC AR(1)	LC AR(1)	LC AR(1)	LC AR(1)
	NN	LC NN3(2)	LC NN1(6)	LC NN1(6)	LC NN1(6)	LC NN1(6)	LC NN2(8)
	ANFIS	LC ANFIS(2)	LC ANFIS(1)	LC ANFIS(2)	LC ANFIS(1)	LC ANFIS(2)	LC ANFIS(1)
Madrid	AR	LC AR(10)	LC AR(9)	LC AR(9)	LC AR(9)	LC AR(9)	K_T AR(9)
	NN	LC NN3(3)	LC NN4(9)	LC NN4(9)	LC NN4(9)	LC NN4(9)	LC NN4(9)
	ANFIS	K_T ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)	LC ANFIS(4)
Lerida	AR	K_T AR(4)	K_T AR(1)	K_T AR(1)	K_T AR(1)	K_T AR(1)	K_T AR(1)
	NN	K_T NN1(9)	K_T NN2(9)	K_T NN1(9)	K_T NN1(9)	K_T NN1(9)	K_T NN2(9)
	ANFIS	K_T ANFIS(3)	K_T ANFIS(3)	K_T ANFIS(3)	K_T ANFIS(3)	K_T ANFIS(2)	K_T ANFIS(2)

Table 4
RMSD of the optimum models for each radiometric station.

Station	Models	Prediction horizon (Half daily)					
		1	2	3	4	5	6
Murcia	AR	20.65	23.67	24.83	25.44	25.57	25.71
	NN	***20.58***	***23.38***	***24.05***	***24.44***	***24.62***	***24.68***
	ANFIS	20.86	23.77	24.83	25.27	25.53	25.69
	Persistence	25.22	29.36	31.94	32.64	33.47	33.46
Albacete	AR	26.54	27.91	31.44	30.27	31.56	31.06
	NN	***26.37***	27.78	***30.24***	29.81	***31.26***	***30.39***
	ANFIS	27.36	***27.51***	30.55	***29.77***	31.37	30.42
	Persistence	27.90	32.06	35.91	36.30	38.39	38.19
Madrid	AR	***23.11***	26.82	28.32	29.24	29.75	30.10
	NN	23.21	***26.07***	***27.08***	***27.68***	***28.07***	***28.31***
	ANFIS	23.40	26.77	27.62	28.27	28.81	28.96
	Persistence	26.17	31.05	33.43	34.75	35.79	36.20
Lerida	AR	***23.45***	27.27	28.72	29.14	29.43	29.69
	NN	23.65	***26.98***	***28.54***	***28.88***	***29.13***	***29.31***
	ANFIS	23.52	27.22	28.81	29.21	29.37	29.66
	Persistence	27.45	33.78	36.50	37.45	38.24	38.99

4.1. Models results

For each station we have tested ten different autoregressive models AR(p) with order $p = 1\ldots10$. The different configurations of neural network models NN(p) have been tested with different input vector size from $p = 1\ldots10$. Training algorithm used is Leverage-Marquardt algorithm (Haykin, 1998). In the case of ANFIS(p), the model has been tested with different input vector size $p = 1\ldots6$.

Table 3 shows for each type of model and each station which is the optimum configuration to minimize the rRMSD for each forecasting horizon. Table 4 shows rRMSD of each optimum model and Table 5 shows the improvement in terms of rRMSD of the optimum models against model based on persistence. The selection of best models for each temporal horizon is marked in red color and underlined. The improvement of the optimum models for radiometric station and each forecasting horizon is shown in Figs. 4–7.

From the last figures and tables we obtain the following information:

- Autoregressive models from time series shows higher uncertainty than nonlinear models.
- Neural Networks and ANFIS models obtain better results from lost component time series. Except in Lerida station where clearness index time series is easier to simulate by models.
- Clearness index time series obtains better results in models of lower order compared to lost component.
- First temporal lag of the time series has the biggest information to make predictions. Indeed, when the size of input vector of the models is increased, the improvement of the models is limited to near 2% in terms of rRMSD.

Table 5
Improvement of the models against persistence.

Station	Models	Prediction horizon (Half daily)					
		1	2	3	4	5	6
Murcia	Improvement AR	18.12	19.39	22.25	22.06	23.60	23.16
	Improvement NN	***18.39***	***20.38***	***24.35***	***24.79***	***26.02***	***25.58***
	Improvement ANFIS	17.28	19.05	22.25	22.58	23.72	23.22
Albacete	Improvement AR	4.89	12.95	12.44	16.62	17.80	18.67
	Improvement NN	***5.50***	13.35	***15.78***	17.89	***18.58***	***20.43***
	Improvement ANFIS	1.95	***14.19***	14.92	***17.99***	18.29	20.34
Madrid	Improvement AR	***11.70***	13.63	15.30	15.87	16.88	16.85
	Improvement NN	11.30	***16.05***	***19.01***	***20.36***	***21.57***	***21.80***
	Improvement ANFIS	10.58	13.79	17.40	18.66	19.50	19.99
Lerida	Improvement AR	***14.58***	19.26	21.30	22.20	23.04	23.85
	Improvement NN	13.85	***20.12***	***21.80***	***22.88***	***23.83***	***24.82***
	Improvement ANFIS	14.32	19.41	21.05	21.99	23.20	23.92

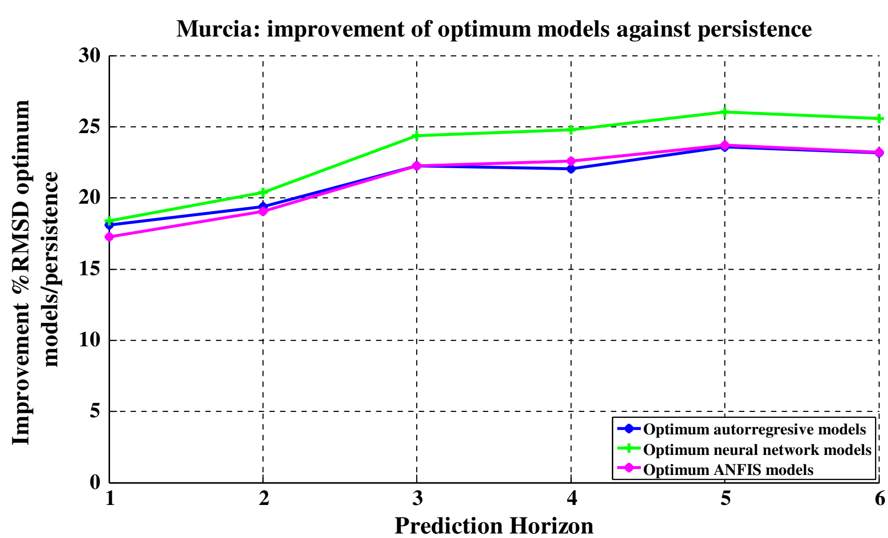

Fig. 4. Murcia: improvement of the optimum models essayed against persistence.

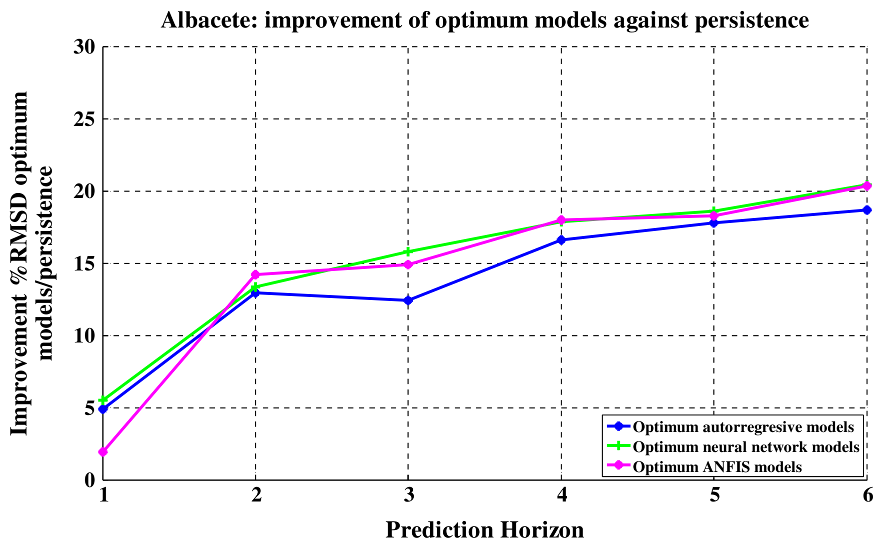

Fig. 5. Albacete: improvement of the optimum models essayed against persistence.

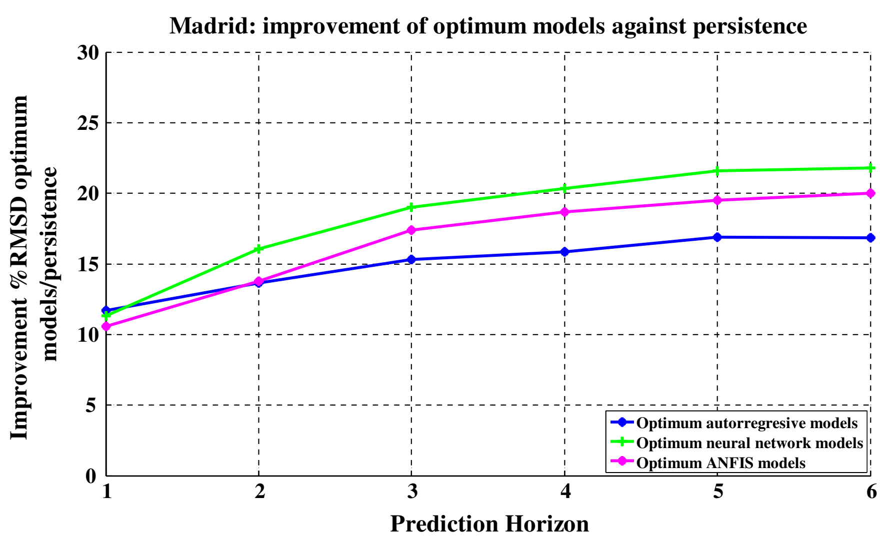

Fig. 6. Madrid: improvement of the optimum models essayed against persistence.

- The evaluation process shows a strong dependence of the forecast accuracy or precision of the models on the climatic conditions and temporal sequence of the data used to train the models.
- The relatives RMSD ranges from 20% to 32% for all forecasting horizons.
- The southern station tested is Murcia and shows similar mean forecasting errors from ECMWF used to predict hourly global solar irradiance. Errors vary between 20% and 25% in terms of RMSD.
- Albacete station shows errors around 6% in terms of RMSD higher than Murcia station which is quite close. For all horizons the errors for this stations varies between 26% and 32% in terms of RMSD. The causes of these high errors could be:
  - ○ Temporal discontinuity in the measurements. In fact, this is the main reason why Albacete station shows bigger predictive uncertainty.
  - ○ A different dynamic in the behaviour of solar irradiance in Albacete station.
  - ○ Problems with pyranometric instruments which could disturb the optimum behaviour of predictive models essayed.
  - ○ Small amount of measurements available for the training and validation of the models. However, in Lerida station the radiometric dataset available is also small and errors are in the same terms as other stations. Therefore, it is not possible to assert that the length of data available for the estimation of parameters models could have any influence in final errors.

- The other two stations show similar errors which vary between 23% and 28% in terms of rRMSD in the case of Madrid and between 23.5% and 29% in the case of Lerida.

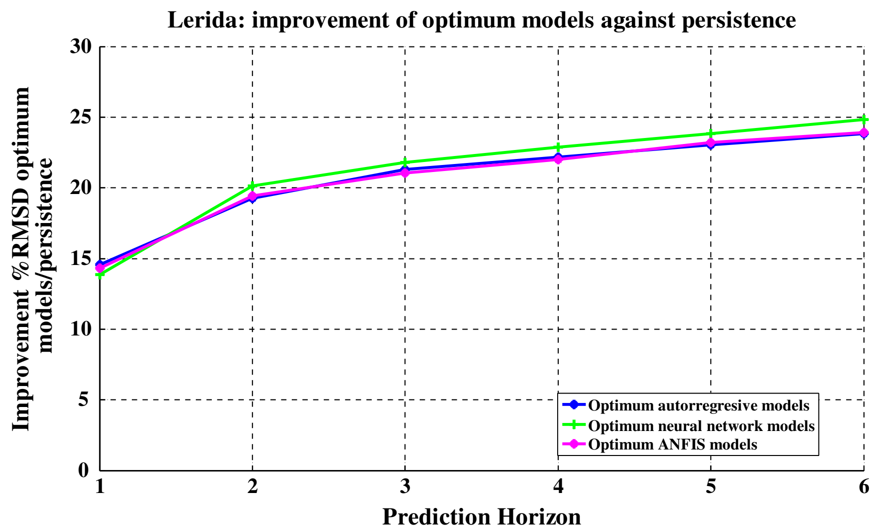

Fig. 7. Lerida: improvement of the optimum models essayed against persistence.

Table 6
Murcia: rRMSD of final model and persistence and improvement.

Models		Forecasting horizon (Half daily)					
		1	2	3	4	5	6
Murcia	Model LC NN2(10)	21.34	***<u>23.38</u>***	***<u>24.16</u>***	***<u>24.55</u>***	***<u>24.76</u>***	***<u>24.90</u>***
	Model K_T NN3(8)	***<u>20.58</u>***	23.62	24.66	25.30	25.54	25.70
	Persistence	25.22	29.36	31.94	32.64	33.47	33.46
Albacete	Model LC NN1(6)	26.55	27.78	***<u>30.24</u>***	29.81	***<u>31.26</u>***	***<u>30.39</u>***
	Model LC NN3(2)	***<u>26.37</u>***	28.35	30.61	30.26	31.46	30.63
	LC ANFIS(1)	28.17	***<u>27.51</u>***	31.80	***<u>29.77</u>***	32.37	30.42
	Persistence	27.90	32.06	35.91	36.30	38.39	38.19
Madrid	Model LC AR(9)	***<u>23.21</u>***	26.81	28.29	29.21	29.73	30.12
	Model LC NN4(9)	23.59	***<u>26.07</u>***	***<u>27.08</u>***	***<u>27.68</u>***	***<u>28.07</u>***	***<u>28.31</u>***
	Persistence	26.17	31.05	33.43	34.75	35.79	36.20
Lerida	Model K_T NN1(9)	***<u>23.65</u>***	27.09	***<u>28.54</u>***	***<u>28.88</u>***	***<u>29.13</u>***	29.33
	Model K_T NN2(9)	23.94	***<u>26.98</u>***	28.58	28.91	29.15	***<u>29.31</u>***
	Persistence	27.45	33.78	36.50	37.45	38.24	38.99

4.2. Final model selection

The selection of the final model, in case there are various models with similar errors, is based on the selection of the model with less number of parameters. The use of this criterion isn't just because the estimation of the simple models is easier, besides it is possible to avoid problems with redundant parameters (Pankratz, 1983). The redundancy happens when models with higher order are used and instead a model with lower order would be enough. Although both models can be used in the forecasting, the estimation of the parameters in the models with higher order result in instabilities, because the parameters of the models changes significantly for different training datasets from the same process (Farnum and Stanton, 1989). Other criterion applied to choose the final model is based on using the minimum number of models and this way simplify forecasting process.

Table 6 shows for each station in red[1] color and underlined the errors of the best models. The final model selected is marked with green color in the background. Errors of persistence model are also shown in the same table.

Final models are bases in neural networks models and improve errors of the model based on persistence in terms of rRMSD. For all radiometric stations errors vary between 20.58% for the first temporal horizon of prediction and 30.39% for the last horizon. The improvement of the final model against persistence varies between 4.84% for the first horizon and 25.58% for the last temporal forecasting horizon in terms of RMSD for all stations. Therefore, the improvement of final models against persistence is higher with increasing forecasting horizon.

[1] For interpretation of color in Table 6, the reader is referred to the web version of this article.

5. Conclusion

This work presents a comparison of statistical models applied on clearness index and lost component to forecast half daily values of solar irradiance for the next 3 days. These predictions have its application in STPP which uses a combination of thermal energy storage system and an optimized strategy operation. This way STPP can improve profitability and operation of STPP.

An storage energy system with capacity between six and nine hours is the optimum size to minimize the cost of components and cost of energy produced (Herrmann, 2009). Besides, STPP with storage capacity can work like a conventional energy power plant due to its capacity to deliver to the grid the electricity produced without intermittence and any stop due to decrease of incoming solar irradiance available.

This way, predictions of incoming solar energy in similar time-steps as the capacity of the storage system allows knowing previously and with the highest accuracy the thermal energy that will be available to be dispatched to the electric grid.

Neural Networks and ANFIS models obtain better results from lost component time series. Except in Lerida station where clearness index time series is easier to simulate by models.

Clearness index time series obtains better results in models of lower order compared to lost component. This is an indication that clearness index time series have less complexity or this variable is more lineal than lost component.

The evaluation process shows a strong dependence of the forecast accuracy or precision of the models on the climatic conditions and temporal sequence of the data used to train the models.

The main advantage of the methodology exposed here is its simplicity to obtain predictions of solar irradiance. In case that wouldn't exist a radiometric database in the location where predictions needs to be generated, it is possible to train the models with radiometric data estimated from satellite images. However it is important to note that the errors exposed here are an upper limit for solar irradiance forecasting for half daily values because no dynamic information from models has been used. So the improvement of these results can be achieved with more complex models which uses statistical techniques in combination with the output from NWP models.

Acknowledgements

The authors wish to thank AEMet for their collaboration with the ground-based measurements, and in particular to José Montero and Santiago Enríquez for their interest and help. Also to IEA/SHC Task 36 which is promoting research on this topic and where the authors are collaborating.

References

Baker, D. G., Casper, M. A., 1981. Subjective forecasting of received solar radiation. Proc. First Workshop on Terrestrial Solar Resource Forecasting and on the Use of Satellites for Terrestrial Solar Resource Assessment 8–11.

Casa Nova, J., Boaventura, C., de Moura Oliveira, P. B., 2005. Solar Irradiation Forecast Model Using Time Series Analysis and Sky Images. In: Proceedings of 5th Conference of the European Federation for Information Technology in Agriculture, Food and Environment (EFITA/WCCA 2005), Vila Real (Portugal). Available from: <http://www.efita.net/apps/accesbase/bindocload.asp?d=5941&t=0&identobj=5HFF23pk&uid=57305290&sid=57&idk=1>.

Esen, H., Inalli, M., Sengur, A., Esen, M., 2008a. Performance prediction of a ground-coupled heat pump system using artificial neural networks. Expert Systems with Applications 35, 1940–1948.

Esen, H., Inalli, M., Sengur, A., Esen, M., 2008b. Predicting performance of a ground-source heat pump system using fuzzy weighted pre-processing-based ANFIS. Building and Environment 43, 2178–2187.

Esen, H., Ozgen, F., Esen, M., Sengur, A., 2009a. Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. Expert System with Applications 36, 11240–11248.

Esen, H., Ozgen, F., Esen, M., Sengur, A., 2009b. Modelling of a new solar air heater through least-squares support vector machines. Expert Systems with Applications 36, 10673–10682.

Farnum, N., Stanton, L., 1989. Quantitative Forecasting Methods, PWS-Kent Publishing Company, Boston

Hammer, A., Heinemann, D., Hoyer, C., Lorenz, E. Satellite based short-term forecasting of solar irradiance–comparison of methods and error analysis, 2000.

Hammer, A., Heinemann, D., Lorenz, E., Luckehe, B., 1999. Short-term forecasting of solar radiation: a statistical approach using satellite data. Solar Energy 67, 139–150.

Haykin, S., 1998. Neural Networks: A Comprehensive Foundation Upper Saddle River. NJ, USA, pp. 1–842.

Herrmann, U. The PCTrough Performance Model, 2009. Workshop on Thermal Storage for Trough Power Systems.

Jang, J.S., 1993. ANFIS: Adaptative-Network-Based Fuzzy Inference System. IEEE Transactions on Semiconductors 83, 378–406.

Jensenius, J. S., Cotton, G. F., 1981. The development and testing of automated solar energy forecasts based on the model output statistics (MOS) technique. In: Proceedings of American Solar Energy Society, pp. 22–29.

Kemmoku, Orita., Nakagawa, Sakakibara., 1999. Daily insolation forecasting using a multi-stage neural network. Solar Energy 66, 193–199.

Lorenz, E., Remund, J., Müller, S. C., Traunmüller, W., et al., 2009. Benchmarking of different approaches to forecast solar irradiance. In: Proceedings of 24 European Photovoltaic and Solar Energy Conference and Exhibition, Hamburg (Germany).

Palit, A.K., Popovic, D., 2005. Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications. Springer-Verlag.

Pankratz, A., 1983. Forecasting with Univariate Box-Jenkins Methods: Concepts and Cases. Wiley Publishing Company, New York, USA.

Patlitzianas, K.D., Ntotas, K., Doukas, H., Psarras, J., 2007. Assessing the renewable energy producers' environment in EU accession member states. Energy Conversion and Management 48, 890–897.

Perez, R., Moore, K., Wilcox, S., Renne, D., et al., 2007. Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database. Solar Energy 81, 809–812.

Renne, D. Semi-Annual Status Report: November 2009. Task 36: Solar Resource Knowledge Management, 2009. IEA. IEA SHC Task 36 Solar Resource Knowledge Management.

Safi, S., Zeroual, A., Hassani, M., 2002. Prediction of global daily solar radiation using higher order statistics. Renewable Energy 27, 647–666.

Veziroglu, T.N., dot, a., 2008. 21st Century's energy: hydrogen energy system. Energy Conversion and Management 49, 1820–1831.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press.

Wittmann, M., Breitkreuz, H., Schroedter-Homscheidt, M., Eck, M., 2008. Case studies on the use of solar irradiance forecast for optimized operation strategies of solar thermal power plants. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE 1, 18–27.