

Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models

Khalil Benmouiza¹ · Ali Cheknane²

Received: 7 January 2015 / Accepted: 12 April 2015 / Published online: 25 April 2015
© Springer-Verlag Wien 2015

Abstract This paper aims to introduce an approach for multi-hour forecasting (915 h ahead) of hourly global horizontal solar radiation time series and forecasting of a small-scale solar radiation database (30- and 1-s scales) for a period of 1 day (47,000 s ahead) using commonly and available measured meteorological solar radiation. Three methods are considered in this study. First, autoregressive–moving-average (ARMA) model is used to predict future values of the global solar radiation time series. However, because of the non-stationarity of solar radiation time series, a phase of detrending is needed to stationarize the irradiation data; a 6-degree polynomial model is found to be the most stationary one. Secondly, due to the nonlinearity presented in solar radiation time series, a nonlinear autoregressive (NAR) neural network model is used for prediction purposes. Taking into account the advantages of both models, the goodness of ARMA for linear problems and NAR for nonlinear problems, a hybrid method combining ARMA and NAR is introduced to produce better results. The validation process for the site of Ghardaia in Algeria shows that the hybrid model gives a normalized root mean square error (NRMSE) equals to 0.2034 compared to a

NRMSE equal to 0.2634 for NAR model and 0.3241 for ARMA model.

1 Introduction

Solar energy is one of the most important renewable energies to generate electricity and meet our everyday needs. PV systems are used to convert this energy to a DC electrical power. However, sometimes it is not possible to estimate the PV system outputs in long-term because they depend strongly on the input parameters such as the amount of solar radiation and temperature. Thus, the solar radiation data should be measured continuously and accurately over the long-term. Unfortunately, in most areas of the world, solar radiation measurements are not easily available due to financial, technical, or institutional limitations. Therefore, many studies have carried out to develop methods to estimate the amount of the solar radiation (Zhang et al. 1998; Zhang 2003; Kaplanis 2006; Kaplanis and Kaplani 2007; Boland 2008; Wu and Chan 2011; Pandey and Soupir 2012; Badescu et al. 2013). In addition, forecasting of solar radiation is important for the integration of photovoltaic plants into an electrical grid. Proper solar irradiance forecasting helps the grid operators to optimize their electricity production and /or to reduce additional costs by preparing an appropriate strategy (Diagne et al. 2009). Forecasts of solar radiation can be either in short or long term. Forecasts for the near future can be done using relatively simple procedures with a good accuracy. In the other side, forecasts for the far future need more complicated models. This is known as a difficult problem, due to the non-linearity and complexity of modeling of the solar radiation series (Zhang 2003; André Luis et al. 2008; Wu and Chan 2011; Mellit et al. 2013; Khatib et al. 2012; Peled and Appelbaum 2013). Hence, many studies have been conducted

✉ Ali Cheknane
cheknanali@yahoo.com; a.cheknane@mail.lagh-univ.dz

Khalil Benmouiza
benkhalil3@gmail.com

¹ Unité de Recherche des Matériaux et Energies Renouvelables, Department of Physics, Faculty of Science, Abou Bekr Belkaid University, BP 119, Tlemcen 13000, Algeria

² Laboratoire des Semiconducteurs et Matériaux Fonctionnels, Université Amar Telidji de Laghouat, Bd des Martyrs, BP 37G, Laghouat 03000, Algérie

on this subject such as stochastic models (Boland 2008; Wu and Chan 2011) and neural network methods (Markham and Rakes 1998; Zhang et al. 1998; Mihalakakou et al 2000; Mellit et al. 2009; Wu and Chan 2011). These models treated the solar radiation sequence as a time series; they used mathematical models in the modeling phase to forecast future values.

The autoregressive–moving-average (ARMA) model is commonly used in time series prediction, the popularity of the ARMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology (Box and Jenkins 1970). However, ARMA model requires a stationary time series, while most real-time series are not stationary (Box and Jenkins 1970; Kwiatkowski et al. 1992; Wu and Chan 2011). We found using the augmented Dickey–Fuller (ADF) test (Dickey and Fuller 1981) that the solar radiation time series is not stationary. Hence, we need a detrending phase to make the time series stationary (Wu and Chan 2011). Therefore, Jain model (Baig et al. 1991; Kaplanis 2006), Baig et al. (1991), Kaplanis (2006), Kaplanis and Kaplani (2007), and high-degree polynomial models are tested in this paper to remove the trends of the solar radiation series. A test of stationarity of the residual series using the ADF test was applied to get the best model to use it in the simulation. The choice of the suitable order of ARMA model is reached using autocorrelation and partial correlation of the residual series as well as the Akaike Information Criterion (AIC) (Akaike 1974).

On the other hand, time series prediction using neural network approach is non-parametric, in the sense that it is not necessary to know any information about the process that generates the signal (Denton 1995; Markham and Rakes 1998; Zhang 2003). Among them, nonlinear autoregressive (NAR) neural networks which used only the past values of the time series to forecast future values. A good choice of the number of delays, neurons, and training algorithm can resolve the problem of the non-linearity of the time series.

However, both ARMA and NAR models present limitations in the forecasting phase. ARMA model shows good results for linear problems, but it could represent huge errors in the nonlinear problems; also, the outliers made the prediction by NAR networks difficult (Zhang 2003; Diagne et al. 2009; Wu and Chan 2011). Hence, hybrid models are proposed taking the advantages of the two models to provide better prediction results. Pelikan et al. (1992), and Ginzburg and Horn (1994) proposed a model combining several feed forward neural networks, improving the time series forecasting accuracy. Wedding and Cios (1996) described a combining method using radial basis function networks and the Box–Jenkins models. Luxhoj et al. (1996) presented a hybrid econometric and an ANN approach for sales forecasting. Zhang (2003) proposed a method using a hybrid combination between ARMA and ANN models to predict time series, André Luis et al. (2008) used Zhang (2003) model and

adjusted the model on the midpoint and an interval range series in the training set. Wu and Chan (2011) proposed a technique employing a combination of ARMA and time delay neural network (TDNN) for one-step ahead prediction based on Zhang (2003) model. In addition, many authors have already studied successfully the coupling between ANN and different traditional computing technologies such as fuzzy logic, wavelet-based analysis (Peled and Appelbaum 2013) and genetic algorithm methods (Mellit et al. 2009; Diagne et al. 2009; Boata and Gravila 2012; Chen et al. 2013). However, most of these models present limitations especially in long-term forecasting. Hence, in this paper, we propose a hybrid model of ARMA and NAR network for multi-step ahead prediction of solar radiation time series for better performance in long-term forecasting.

The follow-up of this paper is organized as follows. In Section 2, we present the proposed methodology as well as backgrounds of the ARMA, NAR, and the hybrid models. A comparison between the detrending models to get the most stationary series is also seen. In Section 3, we have presented the data used in the simulation and comparison results. In Section 4, we simulate the forecasting results of the hybrid model and compared them with other models. The last section is devoted to the conclusion and discussion of future works.

2 Background

This section introduces the adopted methodology in this paper as shown in Fig. 1. It consists of forecasting hourly solar radiation using hybrid ARMA and NAR neural network model. Also, a review of ARMA, NAR, and the proposed hybrid model is discussed.

2.1 The ARMA model

ARMA model of order (p, q) can be viewed as linear filters for digital signal processing. It is of the form,

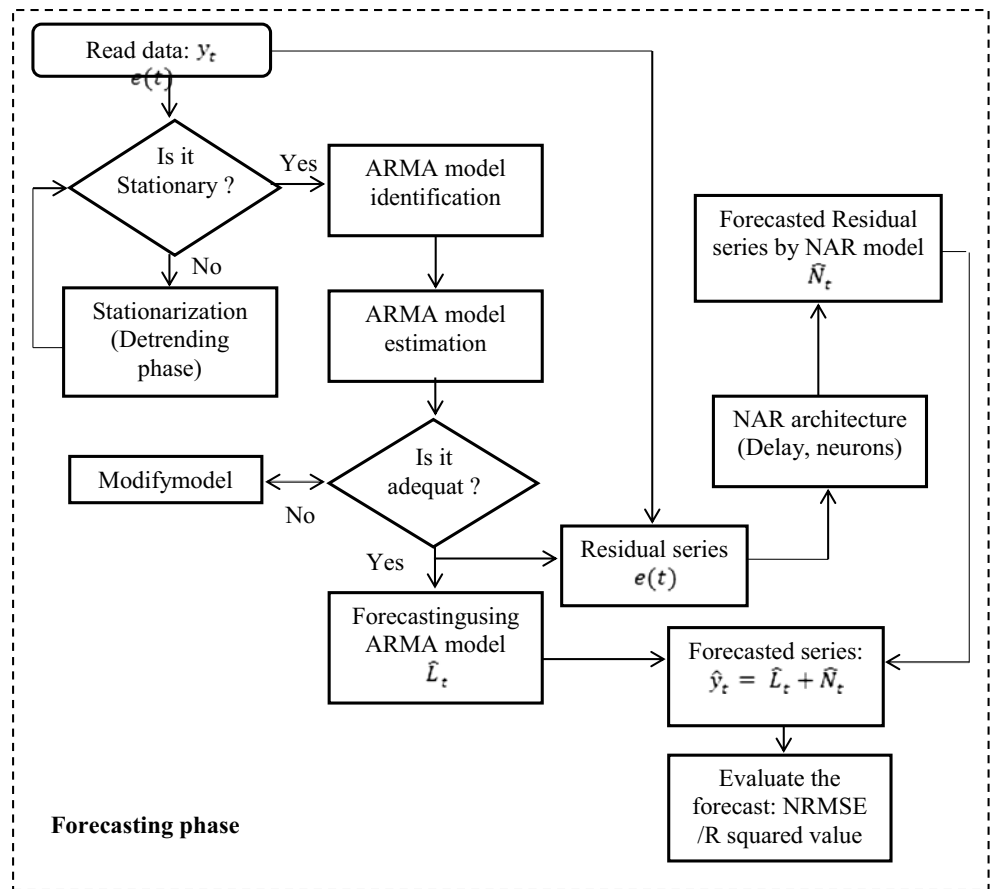
$$x_t = \sum_{i=1}^p \phi_i x_{t-i} + e_t + \sum_{j=1}^q \theta_j e_{t-j} \quad (1)$$

where, $\phi_i (i=1 \dots p)$ and $\theta_j (j=1 \dots q)$ are constants representing the autoregressive AR, and the moving average MA parameters of order p, q , respectively. x_t is the actual value and e_t represents the Gaussian white noise with mean zero in time t . To find the parameters of Eq. (1), the Box and Jenkins (1970) method is applied as expressed in what follows.

2.1.1 Stationarization

Time series modeling and forecasting requires explicitly a stationary time series (Makridakis et al. 1998; Voyant et al.

Fig. 1 The flowchart of the proposed methodology



2013). The condition of stationarity (weakly stationarity) implies a stable series. Which means that the mean $\mu(t)$ and the covariance $cov(x_t, x_{t+h})$ stay constant over time, as expressed by the following equations:

$$E[x_t] = \mu(t) = \mu. \tag{2}$$

$$cov[x_t, x_{t+h}] = E[(x_t - \mu)(x_{t+h} - \mu)] \tag{3}$$

Moreover, a strict stationary series needs a time invariant joint distribution of any observation of the processes. In addition, modeling and analysis of time series of classical models such as ARMA model without testing the stationarity can present real practical problems (Ineichen 2008).

Hence, several methods are demonstrated in the literature to check the stationarity (non-stationarity). The most widely used one is the test of a unit root in the time series (Dickey and Fuller 1981; Kwiatkowski et al. 1992). A unit root test is a test for a specific type of non-stationarity for autoregressive time series. The series is covariance stationary if and only if all the roots of the characteristic polynomials are outside the unit circle in the complex plane. In other words, if it exists a unit root, then the time series is not stationary. Otherwise, it is stationary.

The most widely used method to test unit root is the ADF test (Dickey and Fuller 1981), expressed by the following equation,

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{j=1}^p (\delta_j \Delta x_{t-j}) + e_t \tag{4}$$

where, α is a constant called a drift, β is the coefficient on a time trend, p is the lag order autoregressive process, γ is the coefficient presenting process root, δ_t represent the lag operator and e_t represents an independent identically distributed residual term with mean zero and variance $\sigma^2=0$.

The focus of testing is whether the coefficient γ equals to zero, what means that the original x_1, x_2, \dots, x_n process has a unit root. Hence, the null hypothesis of $\gamma=0$ (random walk process) is tested against the alternative hypothesis $\gamma < 0$ to obtain a stationary series.

The ADF statistic, used in the test, is a negative number. The stronger reject of the null hypothesis needs more negative test. In our simulation and using this stationarity test, we found that the solar radiation series is not stationary. Hence, a stationarization step is needed. A phase of detrending is introduced to obtain a stationary series. In this phase, we simulated different models to fit the solar radiation time series. For each model, the residual series between simulated series and the

original series had been tested using the ADF test. The most stationary series will be used in ARMA modeling. In this paper, the Jain model (Baig et al. 1991; Kaplanis 2006), Baig et al. (1991), Kaplanis (2006), Kaplanis and Kaplani (2007) and high-degree polynomial models are applied to remove trends of the solar radiation series as follow.

The Jain model The Jain model (Baig et al. 1991; Kaplanis 2006) proposed a Gaussian function to fit the recorded data and established the following relation for global irradiation.

Where, r_t is the ratio of hourly to daily global solar radiation, t is the true solar time in hours, m is the time of pick solar radiation hour of the day, and σ is the standard deviation of the Gaussian curve.

$$r_t = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t-m)^2}{2\sigma^2}\right] \tag{5}$$

The Baig model The Baig et al. (1991) model modified Jain’s model to fit the recorded data during the starting and ending periods of a given day. In this model, r_t was estimated by:

$$r_t = \frac{1}{2\sigma\sqrt{2\pi}} \left\{ \exp\left[-\frac{(t-m)^2}{2\sigma^2}\right] + \cos\left[180\frac{(t-m)^2}{S_0-1}\right] \right\} \tag{6}$$

where, S_0 is the length of the day (from sunrise to sunset), n_j is the number of the day at the site with latitude φ . δ is the sun declination.

$$S_0 = \frac{2}{15} \cos^{-1}[-\tan(\varphi)\tan(\delta)] \tag{7}$$

Several methods are found in the literature to estimate the standard deviation σ using recorded data (Kaplanis 2006). Bevington (1969) mentioned that the determination of σ does not need any recorded data and it depends only on the day length, as expressed in Eq. (8):

$$\sigma = 0.246S_0 \tag{8}$$

The r_t values are obtained to offer:

$$I_t = r_t \cdot H_n \tag{9}$$

Where, I_t is hourly solar radiation and H_n is the daily global solar radiation data.

Kaplanis model Kaplanis (2006) proposed another model to estimate hourly global solar radiation that is:

$$r_t = \alpha + \beta \cos\left(\frac{\cos(2\pi(t-m))}{24}\right) \tag{10}$$

where, α and β are parameters which have to be determined for any site and for any day (Kaplanis 2006). However, this

model presented some drawback in the estimation of solar radiation at noontime. Hence, Kaplanis and Kaplani (2007) proposed an improved model for more accuracy as presented in the following equation:

$$r_t = a + b \frac{e^{-\mu(n_j)\chi(t)} \cos(2\pi(t-m)/24)}{e^{-\mu(n_j)\chi(t=m)}} \tag{11}$$

Where, a and b are determined in the same way as Eq. (10), $\mu(n_j)$ is the solar beam attenuation coefficient and $\chi(t)$ is the distance of the solar beam travels within the atmosphere at time t .

High-order polynomial model This model is represented as follows:

$$I_t = a_0h^0 + a_1h^1 + a_2h^2 + \dots + a_nh^n \tag{12}$$

Least squares regression analysis was used to fit Eq. (12) to the data for each hour of the day to obtain the values of the regression constants $a_0, a_1 \dots a_n$ for each month of the year and h is the time (Al-Sadah et al. 1990).

The trends obtained from these models are simulated against the measured data to find the suitable model to be used in the prediction phase. For that, the monthly average hourly global solar radiation time series is then applied. The data are collected from the National Meteorological Office (ONM) of Algeria for the site of Oran (35.6911° N, 0.6417° W). Figure 2 shows the monthly average hourly global horizontal solar radiation of January 2010 in watt per square meter against the estimated models. We ignored data between 6:00 and 20:00 o’clock because there is no solar radiation measured during this period.

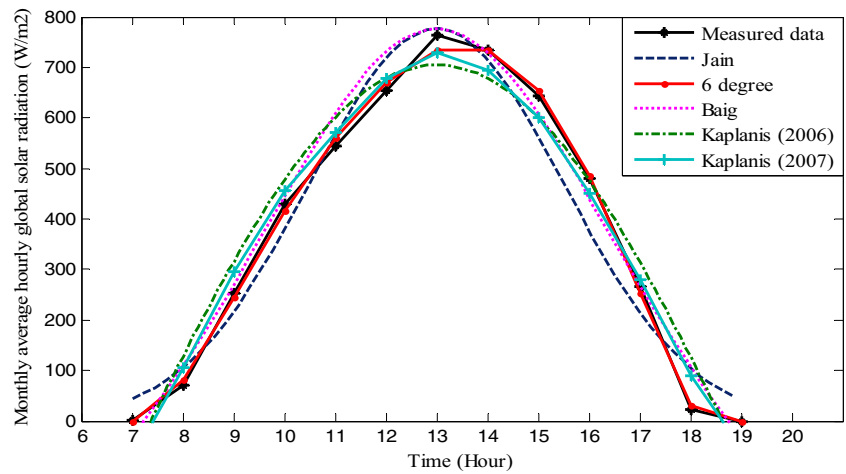
To choose the proper model, we have to check the stationarity of the series. Thus, the ADF test is applied to the residual series between measured and simulated data from different models. If the test result is below the critical values that means we should reject the null hypothesis and the time series is stationary.

Otherwise, it is not stationary. The statistical power is the probability tests to reject a false null hypothesis (Dickey and Fuller 1981). The test results are presented in Table 1. The performances of the five simulated models to predict monthly average hourly global solar radiation from mean daily global solar radiation are evaluated using the root mean square error (RMSE) and normalized root mean square error (NRMSE),

$$RMSE = \left[\left\langle (I_{i,predicted} - I_{i,measured})^2 \right\rangle \right]^{\frac{1}{2}} \tag{13}$$

RMSE and NRMSE provide information in the short-term performance of correlations by allowing a term-by-term comparison of the actual deviation between the predicted and

Fig. 2 Comparison between the measured monthly average hourly global horizontal solar radiation data of January 2010 for the site of Oran, Algeria, and the Jain (Baig et al. 1991; Kaplanis 2006), Baig et al. (1991), Kaplanis (2006), Kaplanis and Kaplani (2007), and 6-degree polynomial models



measured values. The model that has the lowest NRMSE is considered the best model.

$$NRMSE = \left(\frac{\left[\left\langle (I_{i,predicted} - I_{i,measured})^2 \right\rangle \right]^{\frac{1}{2}}}{\langle I_{i,measured} \rangle} \right) \tag{14}$$

The results of the statistical comparison of the simulated models are presented in Table 2,

From Fig. 2 and Table 2, it is clearly shown that Jain’s model fits the monthly average hourly global solar radiation series, but it presents a big NRMSE error versus other models that equal to 0.1490 especially at the beginning and at the end of the series. Hence, since the Baig’s model is based on Jain’s model, it was used to overcome this error. However, it still represents some gaps with NRMSE equal to 0.1146.

For the Kaplanis (2006) model, it used a different method than Jain and Baig models, but still had a big NRMSE equal to 0.1013. Using the improved approach by Kaplanis and Kaplani (2007), the NRMSE is reduced to 0.0735. The 6-degree polynomial model seems the best choice to fit the solar radiation time series, which represents the lowest NRMSE error equal to 0.0358.

Table 1 The ADF test for the detrending models

Detrending models	Statistical power	Significant level	Test results	Critical value
Jain’s model	0.0428	0.05	-2.0389	-1.9580
Baig’s model	0.0126	0.05	-2.6580	-1.9580
Kaplanis (2006) model	0.0117	0.05	-2.6918	-1.9580
Kaplanis and Kaplani (2007) model	0.0097	0.05	-3.4307	-1.9580
6-Degree polynomial model	0.001	0.05	-4.3232	-1.9580

In addition, from the results of Table 1, we can see that the test results are below the critical values. Therefore, the residual series of all these models is considered stationary. The statistical power of 6-degree polynomial model is the highest one, which implies that the residual series between this model and measured data has the lowest probability to incorporate a unit root. Hence, it is considered the most stationary residual series.

Since higher degree polynomial model provides the best performance in both detrending and fitting phases, we used this model for ARMA model in the detrending phase to predict future values.

2.1.2 Model identification

Model identification consists of specifying the appropriate structure, AR, MA, or ARMA and orders of the model (Box and Jenkins 1970). Identification is sometimes done by looking at the plots of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). After determining the ACF and PACF functions, we can choose the (p,q) order of the ARMA model, as expressed in Table 3,

Table 2 The RMSE and NRMSE between actual data and the other different models

	Error (RMSE)	Error (NRMSE)
Jain’s model	55.2255	0.1490
Baig’s model	42.8663	0.1146
Kaplanis (2006) model	37.1987	0.1013
Kaplanis and Kaplani (2007) model	22.5810	0.0735
6-degree polynomial model	13.3939	0.0358

Table 3 Different scenarios of choosing ARMA (p, q) parameters

	AR (p)	MA (q)	ARMA (p, q)
ACF	Tails off	Cuts off	Tails off
PACF	Cuts off	Tails off	Tails off

Akaike’s Information Criterion (AIC) (Akaike 1974) defined by Eq. (15), is another factor to decide ARMA (p, q) order. AIC provides a measure of the model quality by simulating the situation where the model is tested on a different data set. According to Akaike’s theory, the most accurate model has the smallest AIC.

$$AIC = \log V + \frac{2d}{N} \tag{15}$$

Where V is the loss function, d is the number of estimated parameters and N is the number of values in the estimation data set.

2.1.3 Parameter estimation

Once the orders of ARMA model obtained, estimation of the model parameters is straightforward. The parameters are estimated using maximum likelihood method (Box and Jenkins 1970). The last step of the ARMA model building is the diagnostic checking of the model adequacy. The plotting of residuals examines the goodness of the obtained model.

2.2 The nonlinear autoregressive (NAR) model

Recurrent neural networks have been widely used for modeling of nonlinear dynamical systems (Haykin 1998; Ljung 1998). Among various types of the recurrent neural networks, time delay neural networks (TDNN) (Haykin 1998; Wu and Chan 2011), layer

recurrent networks (Haykin 1998) and NAR (Markham and Rakes 1998; Chow and Leung 1996). TDNN is a straightforward dynamic network that consists of a feed-forward network with a tapped delay line at the input layer which the dynamics appear only in the input layer of a static multilayer feed-forward network. However, the NAR is a dynamic recurrent network, with feedback connections including several layers of the network. The next value of the dependent output signal is regressed on previous values of the output signal. The main advantage of using the NAR network comparing with the TDNN is that the input to the feed-forward network is more accurate which, provide more precise result for multi-step ahead prediction.

The NAR model is based on the linear AR model, which is commonly used in time-series forecasting. The defining equation for the NAR network is:

$$\hat{y}(t) = f(y(t-1) + y(t-2) + \dots + y(t-d)) \tag{16}$$

f is a nonlinear function, where the future values depend only on regressed d earlier values of the output signal as expressed in Fig. 3.

When using NAR network, the network performs only a one-step ahead prediction after it has been trained. Therefore, we need to use the closed loop network to perform a multi-step-ahead prediction and turn the network into a parallel configuration. The output of the closed loop NAR network is expressed as follows:

$$\hat{y}(t + p) = f(y(t-1) + y(t-2) + \dots + y(t-d)) \tag{17}$$

where p represents the forecasted steps in the future.

A crucial part of a neural network working is the training step. Because of the very similarity structure between NAR network and the Multilayer Perceptron (MLP), the back propagation method with some modification is then applied; training typically starts with random weights on its synapses. It is exposed to a training set of input data. The output of the

Fig. 3 Structure of NAR network

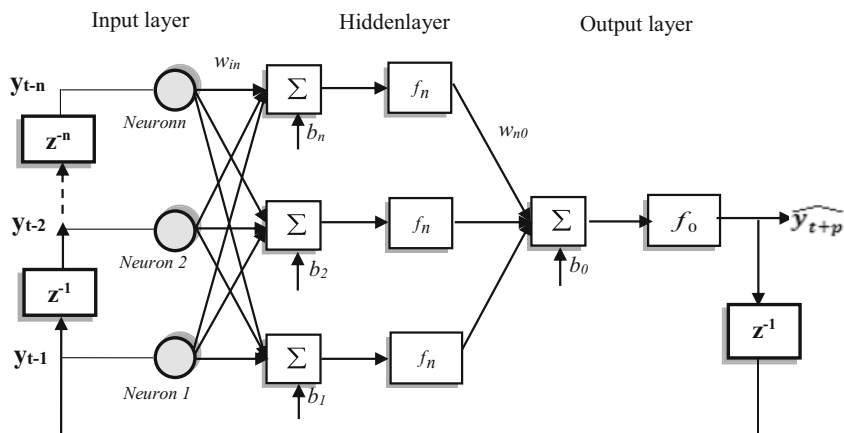


Fig. 4 a Comparison between measured hourly global horizontal solar radiation data (from 1 November 2010 to 31 December 2010) and the forecasted using hybrid model. **b** Comparison between measured hourly global horizontal solar radiation (from 1 November 2010 to the 30 November 2010) and forecasted by proposed model. **c** Comparison between measured hourly global horizontal solar radiation (from 1 December 2010 to the 31 December 2010) and forecasted by proposed model. **d** Comparison between measured hourly global horizontal solar radiation (from 1 November 2010 to the 14 November 2010) and forecasted by proposed model

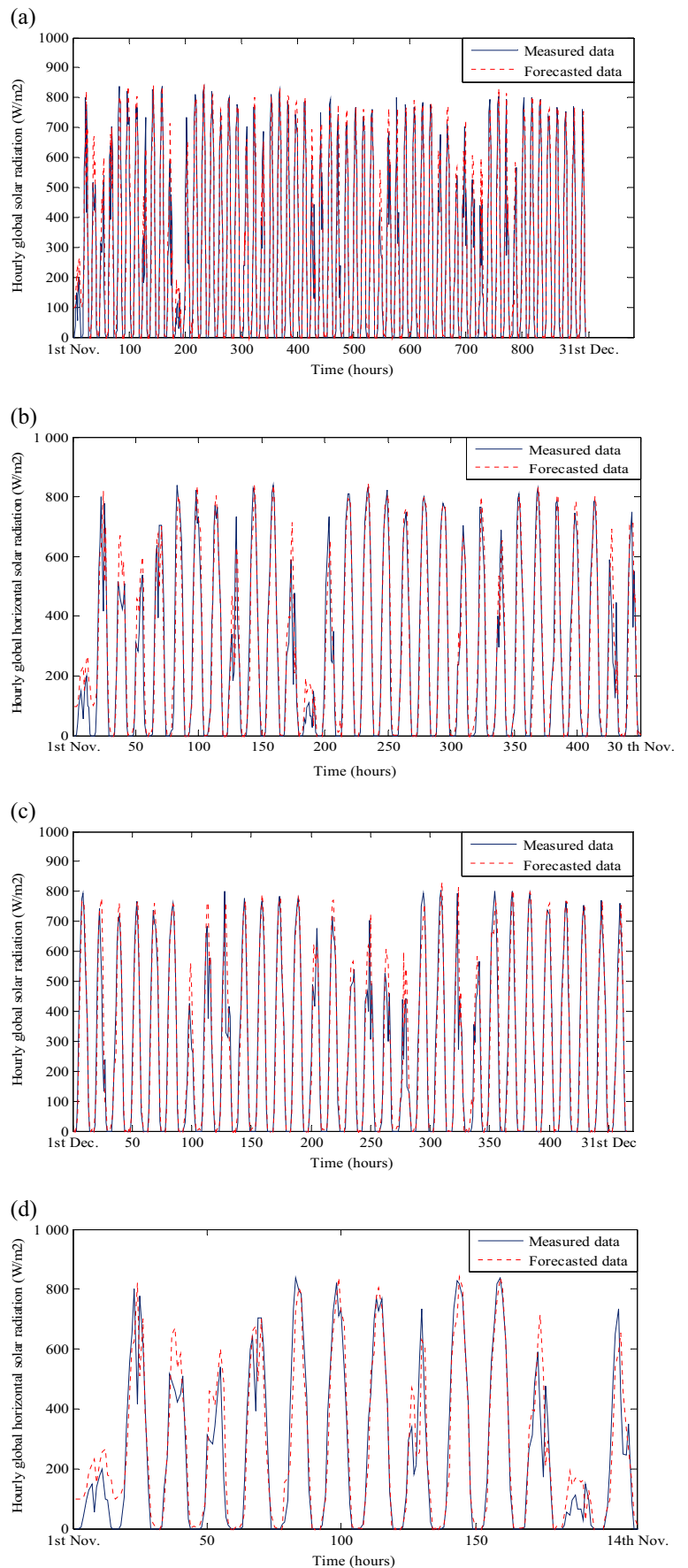
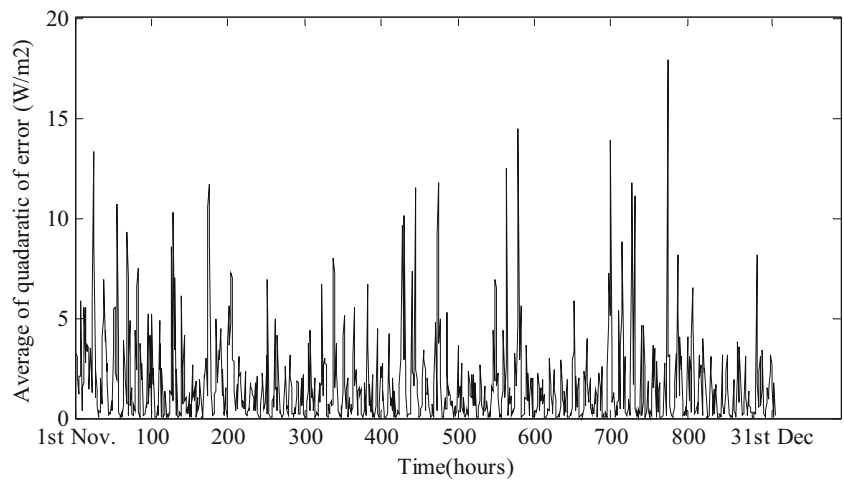


Fig. 5 The average of quadratic error between measured global horizontal solar radiation (from 1 November to 31 December 2010) and the forecasted using hybrid model



network is compared to the example (supervised training) and a learning procedure alters the network interconnections (weights).

Several training algorithms available in the literature, algorithms such as the Levenberg-Marquardt (Levenberg 1944; MacQueen 1967), and Bayesian Regularization (MacKay 1992), proved to be too computationally intensive to train larger networks. After a heuristic search, the scaled conjugate gradient algorithm presented in Moller (1993) is selected to train larger networks. Once the network is trained using the preselected inputs and outputs, all the synaptic weights are frozen and the network is ready to be tested on the new input information.

2.3 The hybrid model

ARMA model represented linear models and has achieved great popularity since the publication of Box

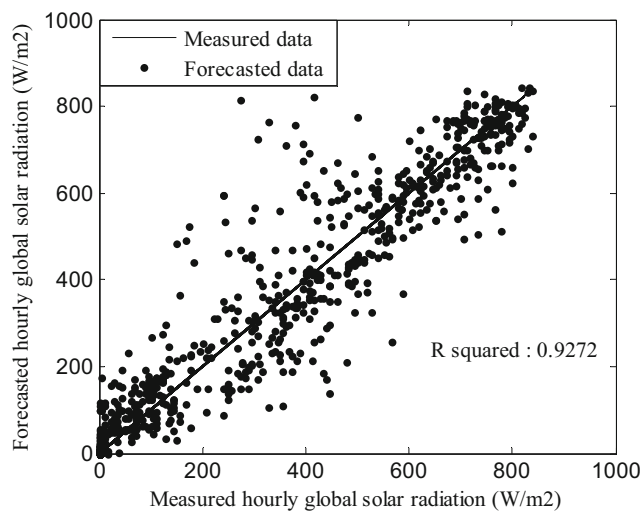


Fig. 6 The measured hourly global horizontal from (1 November 2010 to 31 December 2010) versus forecasted time series using hybrid model

and Jenkins (1970). However, this method may not be adequate for nonlinear problems, contrary of the NAR networks that can solve the complexity of nonlinear systems. However, not one of them can use for both linear and nonlinear problems (Zhang 2003; André Luis et al. 2008; Wu and Chan 2011). Hence, a hybrid models is applied taking the advantages of both ARMA and NAR models. We can simply detect the nonlinearity in a time series by using the surrogate data test for nonlinearity (Kugiumtzis 2000). The proposed hybrid model in this work is based on Zhang (2003) model. It is assumed that time series is composed of a linear autocorrelation structure and a non-linear part:

$$y_t = L_t + N_t \tag{18}$$

where, L_t denotes the linear part and N_t denotes the nonlinear part. The proposed method by Zhang (2003) consists of two stages. Firstly, ARMA model is used to predict future values at time t noted. The residual series between the time series and linear ARMA model series contains only nonlinear relationship. As expressed by the following equation:

$$v_t = y_t - \hat{L}_t \tag{19}$$

where, v_t denote the residual at time t from the linear model.

Table 4 The RMSE and NRMSE error for the site of Ghardaia, 2012; London, 2005; and Almeria, 2012

Location and year	Training/testing data set	ARMA (p,q) order	NRMSE		
			ARMA model	NAR model	Hybrid model
Ghardaia, 2012	(6222/1655)	(4,6)	0.3241	0.2634	0.2034
London, 2005	(3816/1068)	(3,5)	0.3338	0.2862	0.2154
Almeria, 2010	(7196/1564)	(6,5)	0.3001	0.2313	0.1910

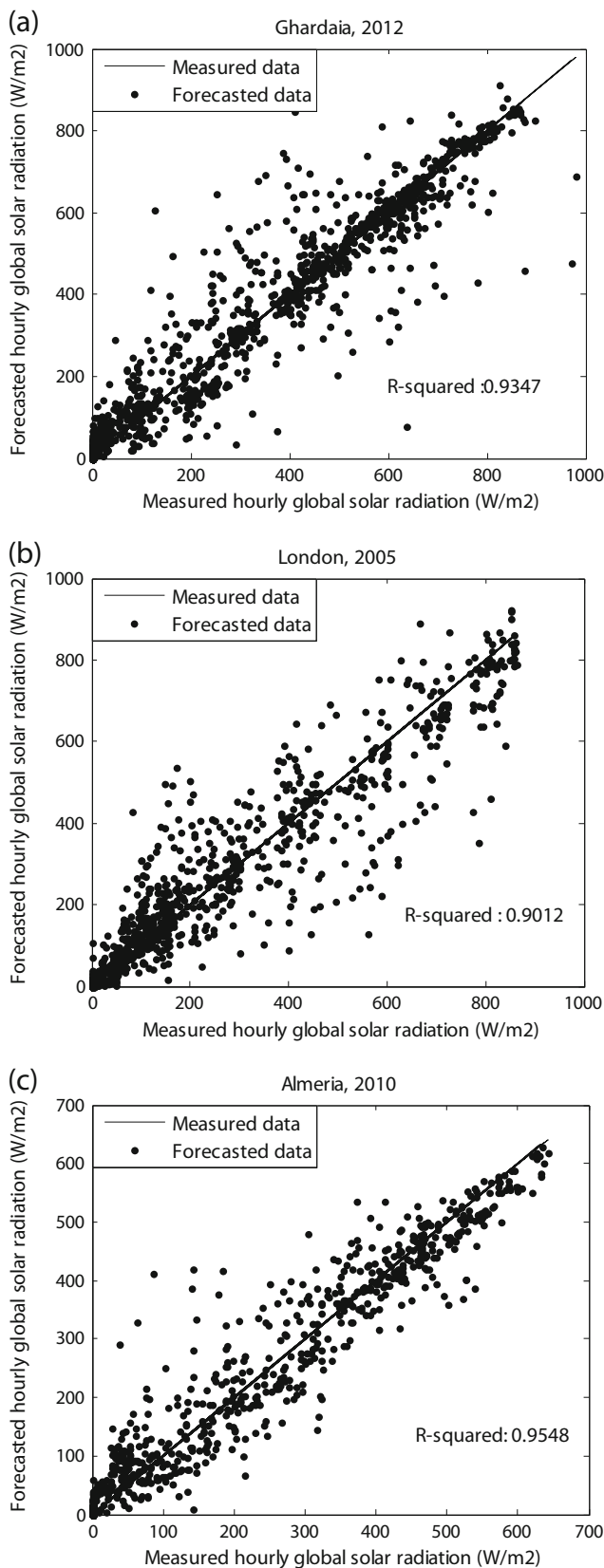


Fig. 7 a The measured test hourly global horizontal from versus forecasted time series using hybrid model for the site of Ghardaia, 2012. b The measured test hourly global horizontal from versus forecasted time series using hybrid model for the site of London, 2005 c The measured test hourly global horizontal from versus forecasted time series using hybrid model for the site of Almeria, 2010

With n input nodes, the NAR model for the residuals will be:

$$v_t = f(v_{t-1}, v_{t-2}, \dots, v_{t-n}) + e_t \tag{20}$$

where, f is a nonlinear function determined by the neural network and e_t is the random error. The forecasted series from Eq. (20) is denoted \hat{y}_t . Then the combined forecast will be expressed by the next equation:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{21}$$

In our simulation, we noted that the residual series v_t is often a random process that makes difficulties in the prediction of future values. The use of a 1D interpolation of v_t can solve this problem. Interpolation is a method of constructing new data point within a range of known data points. The obtained series of interpolation is then used to be forecasted by the NAR network.

3 Data selection

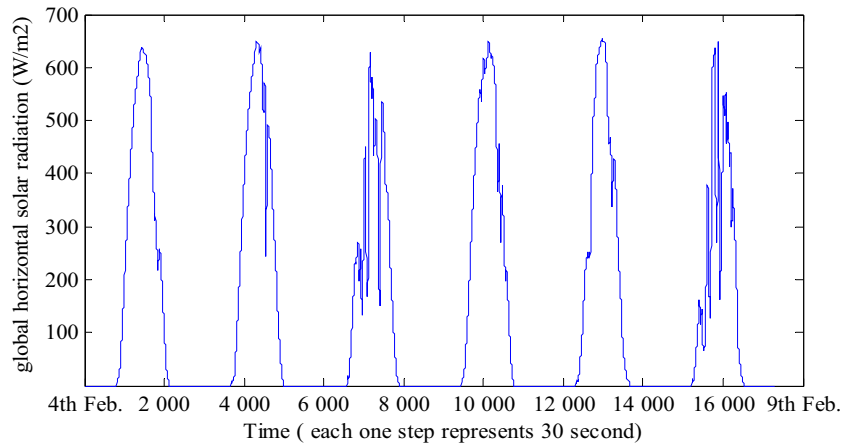
Our goal of the simulation is to select the best model for multi-hour ahead forecasting of the future values of hourly global solar radiation data. To evaluate the quality of the proposed model, the root mean square error (RMSE) and normalized root means square error (NRMSE) are chosen as the forecasting accuracy measures. Lewis (1982) considered that if the NRMSE values are between 0.2 and 0.5, the forecasted model is considered good model. Wu and Chan (2011) and Kostylev and Pavlovski (2011) found that the best performing model on an hourly time scale had an NRMSE of 0.17 for mostly clear days and 0.32 for

Table 5 Comparison between the NRMSE of the forecasting models taken from Wu and Chan (2011) and Huang et al. (2013) and the proposed ARMA + NAR model

Models	NRMSE
ARMA + TDNN (Wu and Chan 2011)	0.3
vs. ARMA + NAR	0.1835
CARDS (Huang et al. 2013)	0.165
vs. ARMA + NAR	0.1339

Secondly, by modeling the residuals using NAR method, nonlinear relationships can be discovered.

Fig. 8 The measured 30-s global horizontal solar radiation (from 4 February 2005 to 9 February 2005) for the site of Oran, Algeria



mostly cloudy days. In addition, the R -squared value gave by Eq. (22) is used as metric to judge the goodness of the forecast.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (I_{i,measured} - I_{i,predicted})^2}{\sum_{i=1}^n (I_{i,measured} - \bar{I}_{i,measured})^2} \right) \quad (22)$$

Moreover, an important task of the proposed method is chosen the proper training and testing data set to avoid the over fitting problem. Hence, the k -fold cross validation method (Kohavi 1995) has been used to check the performances. In this method, the data set is divided into k subsets, each time, one of the k subsets is used as the test set and the other $k-1$ subsets are put together to form a training set. Then, the average error across all k trials is computed until we reached the best training and testing data set (Klipp et al. 2008).

In the simulation phase, we tested several hourly global horizontal solar radiation time series in this work for different climatic locations in the world. From the National Meteorological Office of Algeria, we choose the site of Oran, Algeria (35.6911° N, 0.6417° W) for the year of 2010 and the site of Ghardaia, Algeria (32.4908° N, 3.6728° E) for the year of 2012. From the Soda service website (<http://www.soda-is.com/eng/index.html>), the site of London, England (51.5171° N, 0.1062° W) for the year of 2005 and from GeoModelSolar S.R.O. (data calculated from Meteosat MSG and MFG satellite data (2012 EUMETSAT) and data (2012 ECMWF and NOAA) by SolarGIS method) the site of Almeria, Spain (36.8300° N, 2.4300° W) for the year of 2010.

In addition, to evaluate the performance of the proposed methodology to forecast hourly solar radiation against the methods presented in literature, a comparison part between ARMA and NAR approach and other methods is needed. For that, two models that based in hybrid methodology are

Fig. 9 The measured test 30-s global horizontal solar radiation (9 February 2005) versus forecasted series using hybrid model for the site of Oran, Algeria

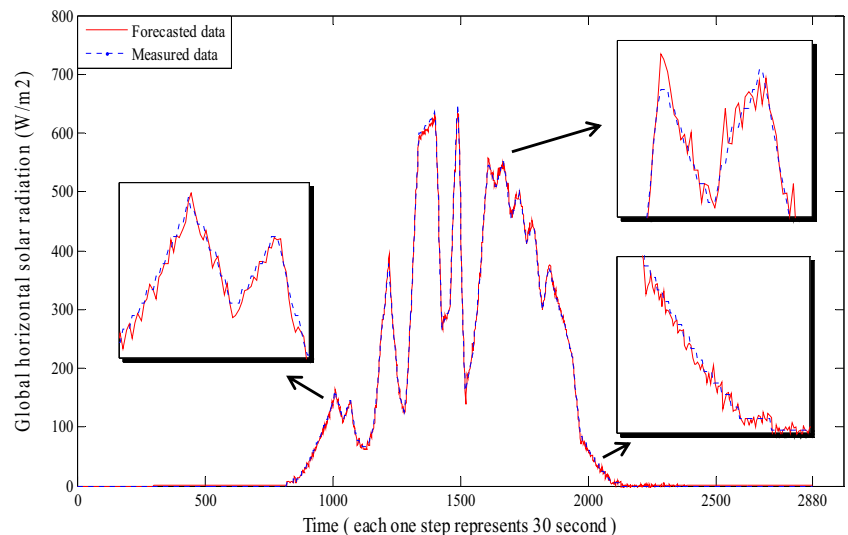
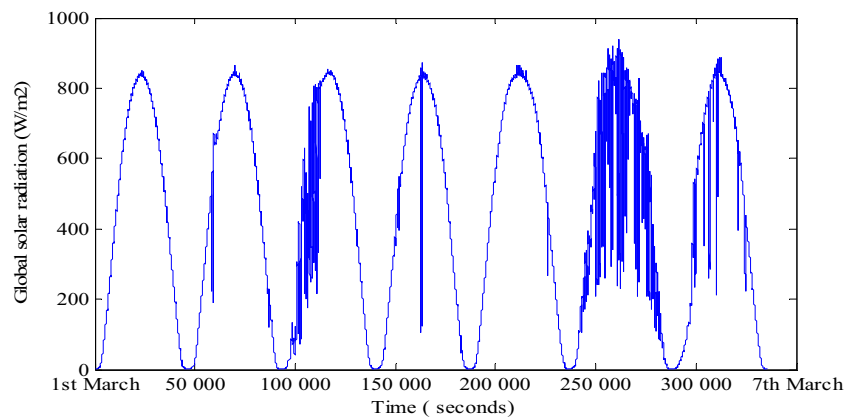


Fig. 10 The measured secondly global horizontal solar radiation (from 1 March 2013 to 7 March 2013) for the site of Sohar, Oman



selected. First, the hybrid model (ARMA and TDNN) proposed by Wu and Chan (2011). In this method, Al-Sadah et al. (1990) model is used to fit the monthly mean solar radiation series. Moreover, the hybrid model of ARMA with TDNN is selected for the forecasting purpose. Secondly, we have chosen the model developed by Huang et al. (2013), a coupled autoregressive and dynamical system (CARDS) model is used to forecast the solar radiation. In addition, the Fourier series is used to fit the solar radiation time series.

For the comparison between the method of this paper and other models, we used the same sample data used in Wu and Chan (2011) (Singapore, 2010; testing day: 2 February) and Huang et al. (2013) (Mildura, 2001; testing day: 25 January).

4 Results and discussion

The first time series used in the simulation is for the site of Oran, Algeria (35.6911° N, 0.6417° W) for the year of 2010. We ignored data between 5:00 and 21:00 o'clock because there is no solar radiation measured during this period. Using the k -fold cross validation method the data are divided into

two sets, training set (from 1 January 2010 to 31 October 2010) that represent 4,530 h, and test data set (from 1 November 2010 to the 31 December 2010) that represent 915 h (prediction horizon). The training data set is used exclusively for model development then the test sample is used to evaluate the established model.

The hybrid ARMA-NAR method is applied to do the forecasting. First, ARMA model is used to predict hourly global solar radiation time series, then the residual between ARMA and measured series is forecasted using NAR model. The obtained forecast is added to the one of ARMA models.

In the detrending phase, we used a 6-degree polynomial model to get a stationary residual series. From the autocorrelation, partial correlation, and the AIC test of the residual series, we established that the ARMA (5, 7) is the suitable model to use it in the simulation.

In addition, different algorithms of training and sets of delays and neurons were tested experimentally in the simulation of the nonlinear autoregressive neural network model.

We found that the use of 31 delays and 10 neurons in the hidden layer with the Levenberg-Marquardt

Fig. 11 The measured secondly global horizontal solar radiation (7 March 2013) versus forecasted series using hybrid model for the site of Sohar, Oman

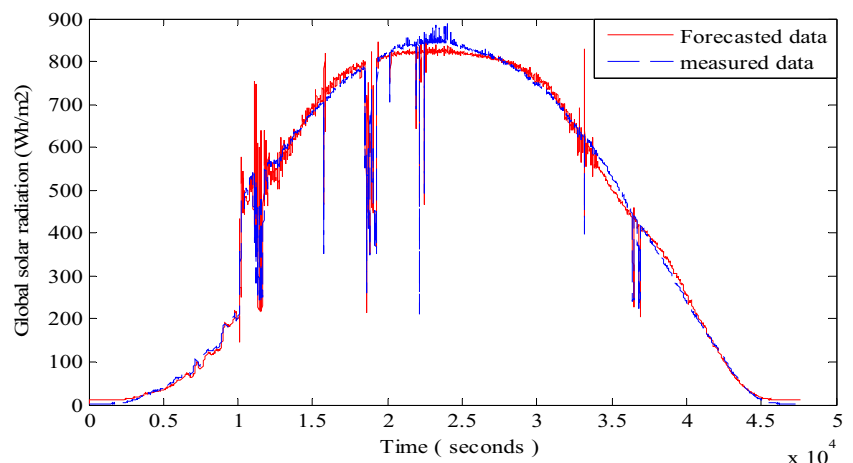
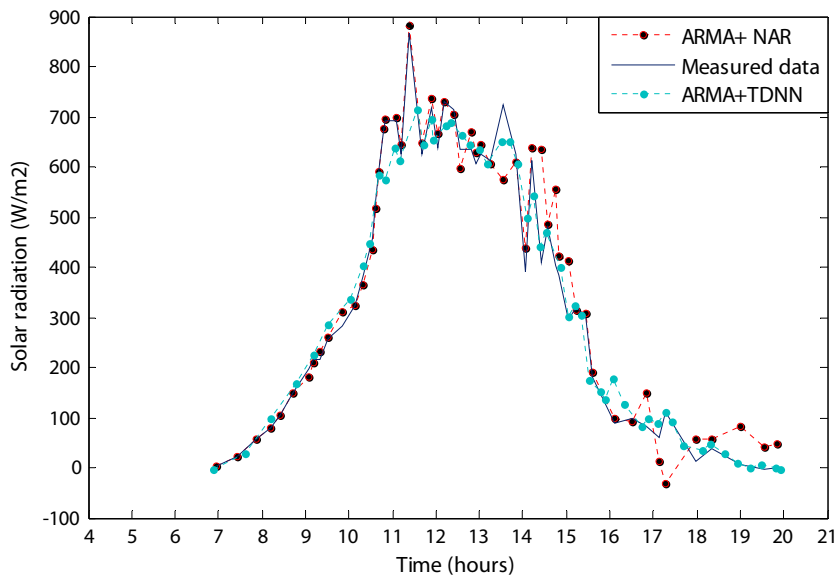


Fig. 12 Comparison between the 2nd February solar radiation data of Singapore 2010 taken from Wu and Chan (2011), the forecasting model using ARMA + TDNN (Wu and Chan 2011) and the proposed ARMA + NAR model



training method gives the fastest convergence with the smallest forecasting error.

The simulation results of the hybrid model to forecast hourly global solar radiation for the year of 2010 are presented in Fig. 4a; the blue line represents the measured hourly global horizontal solar radiation and the red dot one is the forecasted series by hybrid model. In addition, Fig. 4b–c represents the comparison results for the months of November 2010 and December 2010, respectively, and Fig. 4d for the first 2 weeks of November 2010. The blue line represents measured data, and the red dot line is the forecasted data.

The comparisons and performance of the forecasting hourly global horizontal time series using a hybrid model have been evaluated by calculating the RMSE errors between the actual data and forecasted one for the period of 1 November 2010 to 31 December 2010 (915-h-step ahead).

Moreover, the quadratic error expressed in Eq.(23) between measured and simulated hourly global solar radiation using the proposed method is demonstrated in Fig. 5.

In addition, Fig. 6 represents the measured time series versus the forecasted one.

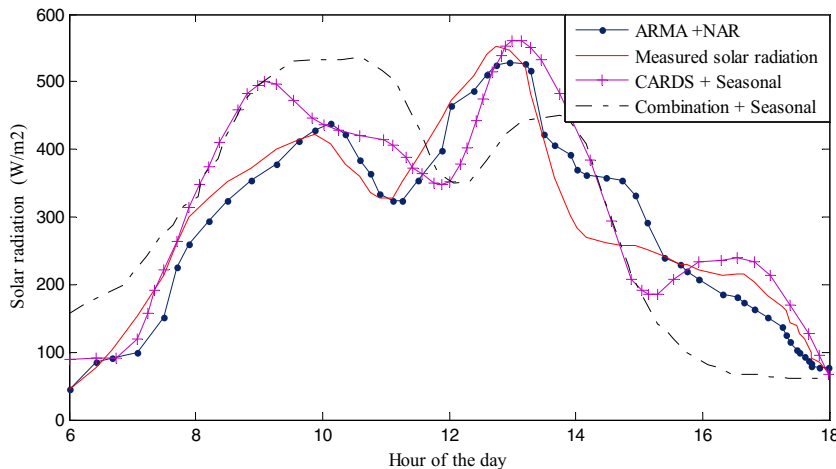
$$err = \left(\frac{(I_{i,predicted} - I_{i,measured})^2}{n} \right) \tag{23}$$

Where *err* is the quadratic error and *n* is the number of samples.

From Figs. 4a–d, 5, and 6 it was clearly shown that the hybrid model forecasted in good manner the measured solar radiation time series. From Fig. 4a, the total RMSE is equal to 71.82 W/m² and the NRMSE is 0.2103. With an *R*-squared value equal to 0.9272. Nevertheless, we can ensure that the comparison between forecasted and measured solar radiation time series presents some lag due to the presence of clouds.

In a same manner, we applied the hybrid method for the sites of Ghardaia (2012), London (2005), and Almeria (2010). The results of the *k*-fold cross validation as well as the RMSE and NRMSE errors between the measured and forecasted

Fig. 13 Comparison between the 25 January solar radiation data of Mildura taken from Huang et al. (2013, p.146), the forecasting model using CARDS+ Seasonal (Huang et al. 2013), Combination + Seasonal (Huang et al. 2013) and the proposed ARMA + NAR model



series are represented in Table 4. Moreover, the simulation results of the proposed hybrid model versus measured hourly solar radiation for the sites of Ghardaia, London, and Almeria are shown in Fig. 7a–c, respectively.

From the results of Fig. 7a–c and Table 5, the hybrid model is considered the suitable method to forecast such similar problems. The NRMSE error had the lowest values comparing with single ARMA and NAR models. In addition, the R -squared value was found to be high for all tested locations.

The above-mentioned models are simulated based on hourly scales. However, the uncertainty of solar radiation time series increases in small scales (less than 1 min time step). Hence, it is an important task to test the proposed hybrid model in small scales. For that, two small step solar radiation data are used. First, a sequence of 30-s solar radiation data for the site of Oran, Algeria (from 4 February to 9 February) was used as shown in Fig. 8. The data are divided into training dataset (from 4 February to 8 February) and testing dataset (9 February) (Fig. 9).

And second, a sequence of 1-s solar radiation data for a desert zone in Sohar, Oman (From 1 March to 7 March 2013) is used as shown in Fig. 10. We ignored the data between 19 o'clock and 6 o'clock because there is no solar radiation data measurement in this period. In addition, data are divided into training dataset (from 1 March to 6 March) and testing dataset (9 February).

The simulation results of the forecasted data compared with measured data are shown in Fig. 9. (Oran, Algeria) and Fig. 11 (Sohar, Oman). From Fig. 9, it is clearly shown that the hybrid model is good with an NRMSE equal to 0.1935. In addition, from Fig. 11 the hybrid model forecast in good manner with an NRMSE equal to 0.1767. However, forecasted data represent some fluctuations compared with measured data that are because it simulated in small scales, which reduce the forecasting accuracy.

4.1 Comparison with other models

For the comparison between the method of this paper and other models, we used the same sample data used in Wu and Chan (2011) (Singapore, 2010; testing day: 2 February) and Huang et al. (2013) (Mildura, 2001; testing day: 25 January).

Figures 12 and 13 show the simulation between the forecasting results using the ARMA and NAR method and other models. According to these figures and results of Table 5 we can see that the hybrid model provides better results with an NRMSE equal to 0.1835 against an average NRMSE of 0.3 for ARMA and TDNN model, and NRMSE of 0.1339 compared with the best NRMSE of the CARDS model that equals to 0.165. Finally, these results show the robustness and the accuracy of the proposed method in this paper.

5 Conclusion

In this paper, we introduced a hybrid model for multi-step ahead forecasting of hourly global horizontal solar radiation time. Firstly, ARMA model is applied to a stationary residual series that obtained from a detrending phase, the ADF test is used to choose the most stationary residual series. We concluded that the high polynomial degree fitting gives better results. Secondly, the NAR model is applied for the forecasting purpose that gives satisfactory results than the ARMA model. However, it takes much calculation time than the first model. The last approach is based on a hybrid method that combined both ARMA and NAR models. According to the fact that solar radiation series has linear and nonlinear components, the ARMA model was good to forecast the linear behavior of the solar radiation time series. Also, NAR network proved to be a suitable method to capture the non-linearity of the series. But, no one of them was suitable to extract full characteristics of global solar radiation series. Hence, as a conclusion of this works, the hybrid model is a good method to forecast such similar problems.

However, those models present a limitation in the forecasting in extremely bad weather condition, thus future works will be focused to test other hybrid models that can improve the reliability for the very cloudy sky.

Acknowledgments The authors would like to thank the University of Laghouat for the financial aspect of the present work. Also, the authors would like to thank Dr. Tamer Khatib and Dr. Hussian Kazem for providing experimental data used for further validation of the proposed model. These data were obtained from an experiment done in the frame of the research project no. ORG SU EI 11 010. Funded by the research council of the Sultanate of Oman.

References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 6:716–723
- Al-Sadah FH, Ragab FM, Arshad MK (1990) Hourly solar radiation over Bahrain. *Energy* 15:395–402
- André Luis S, de Maia Francisco AT, de Carvalho Teresa BL (2008) Forecasting models for interval-valued time series. *Neurocomputing* 71:3344–3352
- Badescu V, Gueymard C, Cheval S, Oprea C, Baciu M, Dumitrescu A, Iacobescu F, Milos I, Rada C (2013) Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renew Energy* 55:85–103
- Baig A, Achter P, Mufti A (1991) A novel approach to estimate the clear day global radiation. *Renew Energy* 1:119–123
- Bevington PR (1969) Data reduction and error analysis for the physical sciences. McGraw Hill Book Co, New York, p 336
- Boata RST, Gravila P (2012) Functional fuzzy approach for forecasting daily global solar irradiation. *Atmos Res* 112:79–88
- Boland JW (2008) Time series and statistical modelling of solar radiation. In: Badescu V (ed) Recent advances in solar radiation modelling. Springer-Verlag, Berlin, pp 283–312

- Box GEP, Jenkins G (1970) Time series analysis, forecasting and control. Holden -Day, San Francisco
- Chen SX, Gooi HB, Wang MQ (2013) Solar radiation forecast based on fuzzy logic and neural network. *Renew Energy* 60:195–201
- Chow TWS, Leung CT (1996) Non-linear autoregressive integrated neural network model for short term load forecasting. *IEE Proc Gener Transm Distrib* 143:500–506
- Denton JW (1995) How good are neural networks for causal forecasting? *J Bus Forecast* 14:17–20
- Diagne M, David M, Lauret P, Boland J, Schmutz N (2009) Review of solar irradiance forecasting methods and a proposition for small-scale in solar grids. *Renew Sust Energy Rev* 13:406–419
- Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49:1057–1072
- Ginzburg I, Horn D (1994) Combined neural networks for time series analysis. *Adv Neural Inf Process Syst* 6:224–231
- Haykin S (1998) Neural networks: a comprehensive foundation, 2nd edn. Prentice Hall, Upper Saddle River
- Huang J, Korolkiewicz M, Agrawal M, Boland J (2013) Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (CARDS) model. *Sol Energy* 87:136–149
- Ineichen P (2008) A broadband simplified version of the Solis clear sky model. *Sol Energy* 82:758–762
- Kaplanis S (2006) New methodologies to estimate the hourly global solar radiation; comparisons with existing models. *Renew Energy* 31:781–790
- Kaplanis S, Kaplani E (2007) A model to predict expected mean and stochastic hourly global solar radiation $l(h;n_j)$ values. *Renew Energy* 32:1414–1425
- Khatib T, Mohamed A, Sopian K (2012) A review of solar energy modeling techniques. *J Renew Sustain Energy Rev* 16:2864–2869
- Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2008) Systems biology in practice: concepts, implementation and application. John Wiley & Sons, West Sussex, p 327
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2 (12): 1137–1143. Morgan Kaufmann, San Mateo, CA
- Kostylev V, Pavlovski A (2011) Solar power forecasting performance—towards industry standards. Proceedings of the 1st International Workshop on the Integration of Solar Power into Power Systems, Aarhus, Denmark
- Kugiumtzis D (2000) Surrogate data test for nonlinearity including nonmonotonic transforms. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 62(1):R25–R28
- Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *J Econom* 54:159–178
- Levenberg K (1944) A method for the solution of certain problems in least squares. *Q Appl Math* 5:164–168
- Lewis CD (1982) International and business forecasting methods. Butterworths, London
- Ljung L (1998) System identification: theory for the user, 2nd edn. Prentice Hall PTR, Upper Saddle River
- Luxhoj JT, Riis JO, Stensballe B (1996) A hybrid econometric-neural network modeling approach for sales forecasting. *Int J Prod Econ* 43:175–192
- MacKay DJC (1992) Bayesian interpolation. *Neural Comput* 4:415–447
- MacQueen JB (1967) Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press, Berkeley, pp 281–297
- Makridakis S, Wheelwright SC, Hyndman RJ (1998) Forecasting: methods and applications, 3rd edn. John Wiley & Sons, New York
- Markham IS, Rakes TR (1998) The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. *Comput Oper Res* 25:251–263
- Mellit SA, Kalogirou L, Hontoria SS (2009) Artificial intelligence techniques for sizing photovoltaic systems: a review article. *Renew Sust Energy Rev* 13:406–419
- Mellit A, Massi Pavan A, Benganem M (2013) Least squares support vector machine for short-term prediction of meteorological time series. *Theor Appl Climatol* 111:297–307
- Mihalakakou G, Santamouris M, Asimakopoulos DN (2000) The total solar radiation time series simulation in Athens, using neural networks. *Theor Appl Climatol* 66:185–197
- Moller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw* 4:525–533
- Pandey PK, Soupir ML (2012) A new method to estimate average hourly global solar radiation on the horizontal surface. *Atmos Res* 115:83–90
- Peled A, Appelbaum J (2013) Evaluation of solar radiation properties by statistical tools and wavelet analysis. *Renew Energy* 59:33–38
- Pelikan E, de Groot C, Wurtz D (1992) Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks. *Neural Netw World* 2:701–712
- Voyant C, Muselli M, Paoli C, Nivet M (2013) Hybrid methodology for hourly global radiation forecasting in Mediterranean area. *Renew Energy* 53:1–11
- Wedding DK, Cios KJ (1996) Time series forecasting by combining RBF networks, certainty factors, and the Box–Jenkins model. *Neurocomputing* 10:149–168
- Wu J, Chan KC (2011) Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. *Sol Energy* 85:808–817
- Zhang G (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175
- Zhang G, Patuwo EB, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. *Int J Forecast* 14:35–62