Regular Paper

Time series analysis and short-term forecasting of solar irradiation, a new hybrid approach

N. Bigdeli, M. Salehi Borujeni, K. Afshar

EE Department, Imam Khomeini International University, Qazvin, Iran

ARTICLE INFO

Keywords:
- Clear sky irradiation
- Flower Pollination Algorithm (FPA)
- Recurrence plot
- ANFIS
- Multivariate regression model
- Solar irradiation correlation analysis

ABSTRACT

In this paper, nonlinear time series analysis and short-term prediction of solar irradiation were considered, simultaneously. The proposed methodology is to employ time series analysis methods as well as swarm and evolutionary algorithms in conjunction with well-known regression, fuzzy and neural network model structures to develop a simple but efficient and applicable model for solar irradiation forecasting. The employed experimental data was the hourly solar irradiation of Qazvin city in Iran for five years. At first, the solar irradiation data was normalized using the daily clear sky irradiation data which is an annually periodic time series. Then, the properties of normalized solar irradiation were characterized via time series analysis methods such as recurrence plots, autocorrelation and mutual information analysis. Based on these analyses, each year was divided into two seasons, the sunny and cloudy seasons which are noticeably different in dynamics. Next, a hybrid but simple model was developed to predict the solar irradiation in different seasons. For the sunny season, an optimized multivariate regression model was proposed; and for the cloudy season a bi-level model consisting of an optimized regression model and ANFIS was developed. The model parameters were tuned optimally by various evolutionary algorithms being GA, PSO, ABC, COA, and flower pollination algorithm (FPA). A Fourier-type model was also developed for modeling of the clear sky data. The results showed the out-performance of FPA method in tuning of the model parameters and convergence time. Besides, the performance of the proposed bi-level model was evaluated in comparison with some other model structures such as artificial neural networks, ANFIS networks, LSE-regression models, LS-support vector machines model, etc. The results showed that the proposed method performs considerably better than the other methods in forecasting the solar irradiation time series in both sunny and cloudy seasons.

1. Introduction

Solar cell is an energy source that is able to convert solar energy directly into electrical energy. Till 2030, 30% of the total energy will be supplied by renewable energy sources. That is, from this value, the share of solar cells will be about 10% of total energy [1] and up to 30% of the energy resources of the European countries [2].

Solar cells are usually integrated in power grids in conjunction with wind power plants, thermal power plants, hydroelectric power plants, and/or batteries. In such grids, fluctuations in solar irradiation and wind speed impose serious reliability and management problems. However, if we introduce more renewable energy, the power grid system might be destabilized due to the fluctuations of weather conditions. To keep the power grid system stable even when we introduce more renewable energy to the power grid system, we need to predict the outputs of renewable energy and compensate the fluctuations by thermal power plants, hydroelectric power plants, and/or batteries. Identifying the uncertainty of future renewable energy outputs is a key to realize such compensations [3–5]. On the other hand, due to its application in forecasting of the generated power in solar power plants, i.e. photovoltaic (PV) arrays, solar irradiation forecast is a dynamic challenging problem, yet. Two main challenges to high penetration rates of PV systems are variability and uncertainty, i.e. the fact that PV output exhibits variability at all timescales and the fact that this variability itself is difficult to predict. The maximum generated power of the PV arrays can be calculated as the multiplication of the solar irradiation at earth’s surface, the collector area (A), and the efficiency of the PV (η) [5]. A and η are the characteristics of PV which are different in various models. Besides, the effective area of PV (A) is proportional to the number of employed PV cells. Therefore, in the literature, solar irradiation forecasting has been considered as the main variable input for the PV array output power forecasting [2,6–35].

Various mathematical methods have been used to predict the solar irradiation such as numerical weather prediction (NWP) models. A NWP
model works on fields of atmospheric variables. These are, very extensive models of the atmosphere have been used to produce meteorological forecasts. To predict future states of the fields, the model is initialized using observed data and a set of mathematical equations for the physics and dynamics of the atmosphere are solved [16]. Another method is model output statistics (MOS). MOS is a statistical method applied to the output of NWP models in order to predict other variables or to improve the predictions [17]. Some recent studies used satellite images or sky images from the ground to predict the solar irradiation. The simplest methods for estimating solar irradiation from the satellite information rely on straightforward relationships between a normalized parameter of the solar irradiance (such as clearness or clear sky index) and the cloud index [7,10,11]. Many different models are also used to predict the solar radiation time series like the classic auto-regression, auto-regression and moving average and Kalman filtering [18,19]. But generally using intelligent models result in increasing the accuracy of prediction [9,12,14,20–22]. In [22], the methods for solar irradiation prediction using artificial neural network techniques were reviewed closely. Other methods, such as least-square support vector machines (LS-SVM) [23], fuzzy [24] and wavelet based models [25,26] have been also employed as the solar irradiation prediction tools. On the other hand, it has been shown in the literature that employment of hybrid methods improves the performance of the solar irradiation models significantly. For example in [27], the hybrids of improve K-means clustering and multilayer perceptron neural network was proposed for this purpose. In some other recent researches, employment of optimization methods for model improvement was considered [15,28]. For good reviews about the solar irradiation forecasting methods, one may refer to [13,29] as well.

From another point of view, the solar radiation sequence can be treated as a time series. This means that we can use time series analysis methods to investigate the underlying dynamics and employ mathematical methods to predict the next values. In the literature, there are some attempts to characterize the frequency content of the solar irradiation time series [30–35]. In [31,32], it has been shown that the solar irradiation captures yearly and intraday cycles that can be used for detrending the hourly solar irradiation or to model the daily profile of it [29]. In [33], some data mining algorithms have been employed to develop four data-driven models for solar irradiation prediction. In [34], it has been tried to find the most relevant input parameters for solar radiation prediction. In [35], time series models have been employed for large scale global solar radiation forecasting from geostationary satellites data.

1.1. The methodology and the paper outline

In this paper, the short-term prediction of solar irradiation has been considered. The proposed methodology is to employ time series analysis methods as well as swarm and evolutionary algorithms in conjunction with well-known regression, fuzzy and neural network model structures to develop a simple but efficient and applicable model for solar irradiation forecasting. Therefore, the proposed analysis and modeling procedure has been organized in two sections. In Section 2, the time series analysis results for the solar irradiation time series have been presented. At first, the solar irradiation has been normalized by removing the daily clear sky irradiation data which is an annually periodic time series. Then, the properties of the normalized solar irradiation have been investigated via time series analysis methods such as recurrence plots, as well as autocorrelation and mutual information analysis. Based on the presented results, the solar irradiation performs a deterministic seasonal behavior with long-term correlation in each season and so, each year has been divided into two seasons, the sunny and the cloudy seasons which are noticeably different in dynamics. The results of analyses imply the model structure and the model inputs as described in Section 3, where, a hybrid but simple model has been synthesized to predict the solar irradiation in different seasons. The proposed model includes 3 components: the first is related to the modeling of clear sky solar radiation trend which is a simple Fourier-type model described in Section 3.1. Next, for the sunny season, an optimized multivariate regression model is proposed in section 3.2; and for the cloudy season a bi-level model consisting of an optimized regression model and adaptive neuro-fuzzy inference system (ANFIS) has been developed in section 3.3. In order for optimizing the model parameters in both levels, the model has been tuned by various evolutionary algorithms being genetic algorithm (GA) [36], particle swarm optimization (PSO) [37,38], artificial bee colony (ABC) [39], cuckoo optimization algorithm (COA) [40], and flower pollination algorithm (FPA) [41,42]. The results show the out-performance of FPA method in tuning of the model parameters and convergence time. Finally, the performance of the model has been evaluated in Section 4 in comparison with some other model structures such as artificial neural networks (ANN), ANFIS networks, LSE-regression models, least square-support vector machines (LS-SVM) model, etc. The results show that the proposed method performs better than the other methods in forecasting the solar irradiation time series in both sunny and cloudy seasons. Section 5 provides the conclusions of the paper.

Based on the above explanations and compared to the existing works, the main contributions of this paper in can be expressed as:

i. A new hybrid modeling method based on the combination of time series analysis with evolutionary and swarm optimization methods in conjunction with well-known regression, fuzzy and neural network model structures for solar irradiation forecasting is provided.

ii. A new time series analysis method based on recurrent plots is applied to investigate graphically the hidden nonlinear dependencies, patterns and dynamic variations (seasonalities) of the solar irradiation time series; which up to the knowledge of the authors, has not employed for this purpose, yet.

iii. From the results of the performed analyses, two distinct seasons (sunny and cloudy seasons) was observed with different dynamics and the time period of them was determined precisely based on the recurrent plots.

iv. Based on the time series analysis results and taking the seasonality of the solar irradiation into account, two new model structures is proposed for sunny and cloudy seasons, which are simple, sufficiently accurate and applicable for routine practical purposes with low computational complexity. The clear sky solar irradiation was modeled via a simple Fourier-type model.

v. Just the past values of the solar irradiation time series has been encountered as the model’s inputs; while the model inputs have been determined systematically via calculating the embedding delay and embedding dimension as well as the correlation analysis of the time series.

vi. The proposed model structures have been well tuned via various evolutionary algorithms and a comparative study was provided to investigate the performance of the past and recent evolutionary algorithms in tuning such a forecasting model.

2. The experimental data and its time series analysis

2.1. The experimental data

In this paper, the hourly solar irradiation of Qazvin city in Iran is considered as our experimental data [43]. Seeking for the predictability of the data and in order for constructing a proper forecasting model, the data will be closely studied, in this section. The available data is the hourly mean data of solar radiation from Jan. 1, 2008 to Dec. 31, 2012 (43,824 h) as shown in Fig. 1; where Fig. 1(a) shows the solar irradiation for the whole period and in Fig. 1(b), the data from June
2008 to June 2010 (17,520 h) has been depicted to make the variations of the data clearer. Besides, in Fig. 1(c) some parts of Fig. 1(b) have been zoomed out to make a clearer insight about the daily variations of the solar irradiations. According to these figures, the solar irradiation is a non-stationary data in which a seasonal trend can be distinguished, truly. Besides, although the daily pattern of solar irradiation is approximately the same, but the daily peak of irradiation is various even in consecutive days. As characterized by the researches in the literature, the solar irradiation captures yearly and intraday cycles that can be used for detrending the hourly solar irradiation or to model the daily profile of it [29]. Fig. 2 shows the frequency content of the solar irradiation data of Fig. 1(a), performed via power spectral density (PSD) analysis of the time series. From this figure, it is shown that two peaks are observed in the PSD of the solar irradiation time series as:

\[ T_1 = \frac{1}{0.04167} \approx 24 \text{ hours} = 1 \text{ day} \]

\[ T_2 = \frac{1}{0.0001142} \approx 8760 \text{ hours} = 365 \text{ days} \]

Which are corresponding to intraday and yearly cycles, respectively.

Implied by these observations and to make the data more treatable, in this paper, some kind of special normalization has been applied to the data. In the literature, various methods have been proposed for detrending the hourly solar irradiance time series. Amongst, the so-called clear sky index is used commonly for this purpose. As expressed in the related literature [29,31,32], the clear sky radiation i.e. the radiation that reaches the top of the layers of the atmosphere is well defined and can be calculated via various methods such as those listed in [44]. Clear sky irradiation (or equivalently the clear sky index) is a deterministic variable which keeps its basic yearly shape in various places and can be calculated via empirical, statistical as well as algebraic relationships in terms of the state of the atmosphere in a given solar zenith angle at any given time [45]. Therefore, regarding the date of days, the clear sky irradiance remains the same for the same day and month, while the details (the detrended time series) are time-varying. Therefore, considering the same day and month, an estimate of the clear sky data can be used for detrending the solar irradiance data regardless of the year. In Fig. 3, the hourly clear sky irradiation (Fig. 3(a)), the daily extracted maximum clear sky irradiation (Fig. 3(b)), and the normalized data (Fig. 3(c)) have been illustrated for the data of Fig. 1(b). From these figures and as emphasized in [44,46], it can be observed that maximum daily clear sky irradiation performs an annually periodic time series which can be modeled analytically. Therefore, the measurements of the previous years may be employed for this purpose, yet.

Based on the provided discussions, in this paper, for each day, the maximum clear sky irradiation of Qazvin has been extracted from the hourly clear sky data taken from [42] and then the 24 h data of each day has been divided by this maximum daily clear sky value. Next, in this section, at first, the properties of the normalized solar irradiation such as its seasonality and predictability are investigated via recurrence plots. Then, the correlation analysis is performed for selection of the proper forecasting model inputs. In this context, the CRP toolbox of MATLAB [47] is employed as our tool.

2.2. Basics of recurrence plot analysis

RP was first presented by Eckman as a tool for analysis of time series, especially to identify hidden dependencies in the complicated data. With RP, one can graphically detect hidden patterns and structural changes in data or see similarities in patterns across the time series under study [48–50]. The first step in analyzing a time series with RP is to reconstruct its embedded phase space via method of delays. In this method, vectors in a new space, the embedding space,
are formed from time delayed values of the scalar measurements \[51\]. In outline, a data series \(S\) can be considered as a set of \(n\) scalar measurements being:

\[
S = \{s_1, s_2, \ldots, s_n\} \tag{2a}
\]

The vectors in the embedding space are defined as:

\[
x_i = \{s_i, s_{i+d}, s_{i+2d}, \ldots, s_{i+(d-1)d} \} \tag{2b}
\]

Where \(r\) and \(d\) are known as embedding delay and embedding dimension, respectively \[50\]. RP illustrates vector’s behavior by the following matrix:

\[
R_i = \Theta(x - \|i; s_i\|) \tag{3}
\]

Where \(x_i\) stands for the point in the reconstructed phase space at time \(i\) and \(\varepsilon\) is a threshold. \(\Theta\) is Heaviside function that assigns a black dot to one, and white dot to zero. Two-dimensional graphical representation of \(R_{ij}\) is called RP \[48,52,53\]. In this context, three types of main system namely periodic, chaotic and stochastic are considered that their pattern images are different. Periodic systems are indicated as\[55\]. RP illustrates vector’s behavior by the following matrix:

\[
H(A, B) = -\sum_{i=1}^{n} \sum_{j=1}^{m} (p(a_i, b_j) \times \log(p(a_i, b_j))) \tag{6}
\]

If \(A\) is a time series, \(B\) can be considered as the lagged vector of \(A\).

2.4. The embedding dimension

Another parameter that must be determined for the reconstruction phase space is embedding dimension. A good method for determining the minimal sufficient embedding dimension is the false nearest neighbor method which was proposed by Kennel. The idea behind this method is that with a right embedding dimension the neighborhoods in a reconstructed phase space are mapped onto neighborhoods again. But, for wrong embedding delay the topological structure is no longer preserved and the neighbor points are projected into neighborhoods of other points. These points are called false neighbors. Based on this idea, of the algorithm of false nearest neighbors is as following: for each point in the \(m\)-dimensional reconstructed phase space look for its nearest neighbor and calculate the distance between these two neighbors. Iterate both points and compute the distance of the iterated neighbors. If the ratio of these two distances are less than a threshold, this point is marked as having a false nearest neighbor \[55\].

The concept of the false nearest neighbors is explained in Fig. 4, as well. From this figure, it can be seen that in \(R^3\) space, \(b\) is the nearest neighbor of \(a\), and by enhancing the dimension (in \(R^4\) space) the two point distance increases so that these points are not neighbors anymore. Thus, \(b\) is \(a\)'s neighbor not for the dynamics of the system, but because the geometric structure of the attractor has been projected down onto smaller space. In this way \(b\) is a false near neighbor of \(a\). By this method, dimension \(d\) is the minimum dimension for a successful reconstruct of state space only if the neighbor points in \(R^d\) space are neighbors in \(R^{d+1}\) space too. The criterion that the embedding dimen-
2.4. RP analysis

Now, according to the obtained embedding delay and embedding dimension, the RP of the normalized solar irradiation data is plotted in Fig. 7. From these figure, two major issues is concluded. The first one is that the dynamics of solar irradiation data has a strong seasonality. In a more detail expression, in the dynamic behavior of the solar irradiation, two seasons is truly distinguished; the sunny season is related to data from 1 to 3600 of Fig. 1(b) (data 13,142–16,742 of Fig. 1(a)), as well as 8761–12,361 of Fig. 1(b) (data 21,903–25,503 of Fig. 1(a)), corresponding to June to September of years 2009 and 2010. The reconstructed phase spaces of these two intervals are very well similar. The second part of each year is related to the cloudy season which shows a completely different behavior with respect to the sunny season but into a noticeable extent shows similar dynamics in two years.

The second observation is related to the dynamic properties and predictability of the data. From Fig. 7(b), it is illustrated that the diagonal lines are long and continuous at sunny season; it proves that solar radiation is somewhat periodic and so predictable in this season. At the cloudy season, on the other hand, the diagonal lines are noticeably shorter and so the solar radiation data seems chaotic and less predictable at this season. That is, the solar irradiation is short-term predictable in the cloudy season. Figs. 8 and 9 show the time series of sunny and cloudy seasons; which emphasize the results of the RP analysis, as well.

2.5. Correlation analysis

One of the most common tools for determination of effective inputs for a forecasting model is the correlation analysis between the underlying time series. The linear cross correlation between two time series \(X=\{x_i\}_{i=1}^N\) and \(Y=\{y_i\}_{i=1}^N\) can be defined as:

\[
C_{X,Y}(k) = E(x_{n+k}^* y_n) - E(X)E(Y), \quad k = 1, \ldots, N
\]

Where, in Eq. (7), \(E(\cdot)\) and * stand for mathematical expectation and complex conjugate, respectively. Indeed, the cross correlation is employed to examine the linear correlation of the two time series. The more slow decays the correlation plot, the more linearly predictable is the \(Y\) in terms of \(X\).

In this section, the correlation analysis is performed separately for each season, respectively. Fig. 10(a) and (b) show the autocorrelation of the solar irradiation versus the time lag for sunny and cloudy seasons, respectively. From this figures one may conclude that linear correlation in the normalized solar irradiation data with its lagged values is noticeable and conservative for the same hour of day and up to four hours beforehand up to several days before the desired time. This linear correlation is, however, lower in the cloudy season.

Nonlinear correlation of the normalized data is also reflected via the mutual information (MI) analysis as is illustrated in Fig. 11. From this figure, for the sunny season the MI analysis is in accordance with the observation in the autocorrelation analysis results. This means that even a well-tuned regression model may be sufficient for prediction of solar irradiation in the sunny season. In the cloudy season, on the other hand, nonlinear dependency among the lagged values of solar irradiation exists but is not as strong as in the sunny season. Therefore, a more complicated nonlinear model with the ability of learning the hidden nonlinear relationship is required for solar irradiation prediction in this season.

3. The solar irradiation time series prediction

In this section, prediction of the solar irradiation will be done in three phases. At first, a prediction model is constructed for modeling and prediction of the maximum daily clear sky solar irradiation time
series. Next, two prediction models are developed for prediction of the normalized solar irradiation in sunny and cloudy seasons.

As it is well known, the maximum daily clear sky solar irradiation is an annually periodic quasi-sine shape time series which can be modeled in terms of several sine functions. In order for modeling of the normalized solar irradiation time series in the sunny and cloudy seasons, in this paper, a variety of model structures such as artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS) networks, Regression model, least square- support vector machines (LS-SVM) model and some others have been constructed, trained and evaluated. In case of sunny season, it will be seen in continue that well-tuned regression model (an optimized regression model) can precisely predict the normalized solar irradiation time series’ in the sunny season.

Prediction of the normalized time series in the cloudy season is, however, a more complicated task. In order for constructing a prediction model for this time series, various structures have been tested as well in this research and finally, a new bi-level hybrid method consisting of an optimized regression model and ANFIS is proposed. The details of the developed models will be described in continue.

3.1. Modeling of clear sky solar radiation trend

As stated earlier, the first step in forecasting the solar irradiation is to predict the maximum daily clear sky irradiation. As stated earlier, this time series is annually periodic. Due to approximately sine wave shape of this time series (Fig. 3), this time series has been modeled via the Fourier analysis of the annual time series.

The approximation of the maximum clear sky can be then modeled as:

\[ f(t) = 874.4 + 233.7 \cos(w \cdot t) + 67.86 \sin(w \cdot t) - 37.53 \cos(2w \cdot t) - 36.79 \sin(2w \cdot t) \]  

(8a)

In Eq. (8), \( t \) stands for the desired day of year and \( w \) is the main frequency of the time series i.e.:

\[ w = \frac{2\pi}{T} = \frac{2\pi}{365} \approx 0.0172 \]  

(8b)

Fig. 12 shows the modeled time series in comparison with the real one. In order to evaluate the model, the mean absolute percentage
error (MAPE) has been used as our index. Let $A_i$ and $P_i$ be the $i$th actual and model output, where $N$ is the number of points in the set, the MAPE index is defined as [57]:

$$ MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_i - A_i}{A_i} \right| \times 100\% $$

(9)

With this index, the model of Eq. (8) approximates the annual clear sky time series with the small error of 0.4936%.

3.2. Prediction of solar radiation in the sunny season

The second step in solar irradiation prediction is to predict the normalized solar irradiation time series in the sunny season. As stated earlier, a variety of linear and nonlinear model structures such as MLP artificial neural networks, ANFIS networks, least square-regression model, LS-SVM model and some others has been employed for modeling of this time series (refer to Appendices B to D for some

Fig. 9. Solar irradiation data in cloudy season: (a) in the period of two years (b) some parts of (a) zoomed out.

Fig. 10. Autocorrelation of the solar irradiation data in (a) sunny season; (b) cloudy season.

Fig. 11. MI of the solar irradiation data in (a) sunny season; (b) cloudy season.
Amongst, due to strong linear dependency in this time series, a well-tuned linear regression model performed more suitable for modeling and prediction of this time series. In order for constructing this model, a general form multivariate linear regression function has been considered as follows:

$$f(x) = \alpha_1 x_{1} + \alpha_2 x_{2} + \alpha_3 x_{3} + \ldots + \alpha_k x_{k}$$  \hspace{1cm} (10)

Where, $x_i$ stands for $i$th delayed input regressor's and $\alpha_i$ are the model parameters to be identified. Based on the auto-correlation as well as the mutual information plots of Figs. 10(a) and 11(a), the underlying model structure for the normalized solar irradiation in the sunny season has been considered as follows:

$$I_{\text{predicted}}(t) = \alpha_1 I_{\text{real}}(t-24) + \alpha_2 I_{\text{real}}(t-48) + \alpha_3 I_{\text{real}}(t-72)$$  \hspace{1cm} (11)

Where, $I_{\text{predicted}}$ and $I_{\text{real}}$ are the predicted and the real values of the normalized solar irradiation in the sunny season has been considered as follows:

$$I_{\text{predicted}}(t) = \alpha_1 I_{\text{real}}(t-24) + \alpha_2 I_{\text{real}}(t-48) + \alpha_3 I_{\text{real}}(t-72)$$  \hspace{1cm} (11)

Where, $I_{\text{predicted}}$ and $I_{\text{real}}$ are the predicted and the real values of the normalized solar irradiation in the sunny season; $\alpha_i$, $i = 1, 2, 3$ are the model parameters to be identified. In order to well tune these parameters, FPA has been successfully used in this paper. That is, the fitness function of Eq. (12) has been defined for this purpose, as follows:

$$\text{fitness}_i = \frac{1}{N} \sum_{t=1}^{N} \left| I_{\text{predicted}}(t) - I_{\text{real}}(t) \right| / \left( \text{mean}\left(I_{\text{real}}\right) \right)$$  \hspace{1cm} (12)

Where, the function of $\text{mean}(I_{\text{real}})$ is representative of the average value of the $I_{\text{real}}$ time series over the training data set. In order for minimizing the fitness function, Genetic Algorithm (GA) [36] with initial population of 200, Particle Swarm Optimization (PSO) [37,38] with 200 particles, Artificial Bee Colony (ABC) [39] with colony size of 200, Cuckoo Optimization Algorithm (COA) [40] with 25 nests and flower pollination algorithm (FPA) [41–43] with initial population of 20 have been employed as our optimization tool. Table 1 shows the results for tuning the parameters via various optimization algorithms running on a Core i7 Laptop, with 6 GB RAM and 32-bit windows 7 Professional operating system. As shown in Table 1, the FPA and COA methods have the least forecasting error in the sunny season. However, having a lower convergence time, the FPA method (see Appendix A for more details on this method) is chosen as the best method.

By minimizing this fitness function via FPA evolution for 2000 training data, the best tuning parameters are calculated and the prediction model would be:

$$I_{\text{predicted}}(t)=0.827I_{\text{real}}(t-24)+0.128I_{\text{real}}(t-48)+0.044I_{\text{real}}(t-72)$$  \hspace{1cm} (13)

Fig. 13 shows the evolution of the FPA fitness function versus the number of the generation iterations. As it is seen it has been converged in 87 iterations which is, very well.

### 3.3. Prediction of solar radiation in the cloudy season

Originated from the time series analysis results, it is observed that predicting solar irradiation data in the cloudy season is a more complicated task. Once again, in order for constructing a prediction

---

**Table 1**

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE%</th>
<th>Convergence iteration</th>
<th>Convergence time (s)</th>
<th>Values of $\alpha_1$, $\alpha_2$ and $\alpha_3$ in Eq. (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>3.82</td>
<td>59</td>
<td>5.910</td>
<td>$\alpha_1=0.680$, $\alpha_2=0.189$, $\alpha_3=0.132$</td>
</tr>
<tr>
<td>PSO</td>
<td>3.78</td>
<td>47</td>
<td>3.791</td>
<td>$\alpha_1=0.750$, $\alpha_2=0.091$, $\alpha_3=0.158$</td>
</tr>
<tr>
<td>ABC</td>
<td>3.65</td>
<td>33</td>
<td>3.111</td>
<td>$\alpha_1=0.825$, $\alpha_2=0.125$, $\alpha_3=0.048$</td>
</tr>
<tr>
<td>COA</td>
<td>3.64</td>
<td>130</td>
<td>1.596</td>
<td>$\alpha_1=0.827$, $\alpha_2=0.128$, $\alpha_3=0.044$</td>
</tr>
<tr>
<td>FPA</td>
<td>3.63</td>
<td>87</td>
<td>0.776</td>
<td>$\alpha_1=0.827$, $\alpha_2=0.128$, $\alpha_3=0.044$</td>
</tr>
</tbody>
</table>
model for this time series, various structures such as MLP artificial neural networks, ANFIS networks [59], Regression model, LS-SVM model and some others have been tested in this paper and finally, a new bi-level hybrid method consisting of regression model tuned by optimization algorithms and ANFIS is proposed for prediction of this time series.

The infrastructure of the propose model has been shown in Fig. 14. Based on this figure, the modeling procedure is done sequentially in two levels. That is, initially, the normalized solar irradiation is predicted for the 24 h of the desired day regression model which are then fed to an ANFIS model to make a more precise prediction. In order to implement the first level regression predictor, the proper regressor's have been selected considering the autocorrelation and mutual information plots of Figs. 10(b) and 11(b), respectively. Therefore, the regression model would be:

\[
\hat{I}_{\text{predicted}}(t) = \beta_1 \times I_{\text{real}}(t-24) + \beta_2 \times I_{\text{real}}(t-48) + \beta_3 \times I_{\text{real}}(t-72)
\]  

(14)

Where, \( \hat{I}_{\text{predicted}} \) and \( I_{\text{real}} \) are the predicted and the real values of the normalized solar irradiation in the desired time of the cloudy season; \( \beta_i \), \( i = 1,2,3 \) are the model parameters to be identified. Here, the \( \beta_i \) coefficients are identified recursively using the optimization algorithms being GA with initial population of 200, PSO with 200 particles, ABC with colony size of 200, COA with 25 nests and FPA with initial population of 20. In this method the cost function of Eq. (14) is minimized. That is,

\[
\text{fitness}_i = \frac{1}{n} \sum_{t=1}^{n} (\hat{I}_{\text{predicted}}(t) - I_{\text{real}}(t))^2
\]

(15)

Considering the correlation results of the data, each time, the data of the last two week i.e. 336 previous data are used to determine the \( \beta_i \) coefficients.

In the second level, the final forecast of the next 24 h irradiation is performed by an ANFIS model. The outputs of the regression model are considered as the inputs of this network as shown in the block diagram of Fig. 14 (where, \( I_{\text{predicted}}(t) \) is the predicted value of the normalized solar irradiation) and the network has been trained and evaluated with about 80% of the available data via the fuzzy tool box of MATLAB [58]. The number of input membership functions of the network has been considered as 3, 4, and 3 for three inputs and type of the membership functions has been considered as triangle (trimf), Gaussian combination (gauss2mf), and triangle (trimf) membership functions, respectively. The underlying membership functions have been shown in Fig. 15.

Applying various optimization algorithms, the results of regression tuning are shown in Table 2, for the cloudy season. For the cloudy season, as indicated in Table 2, the COA, ABC and FPA methods has the same MAPE rate. Similar to the sunny season, the FPA method is chosen as the best method, for its lower convergence time, too.

4. Results

Combining the three developed models, the 24 h prediction results of solar irradiation for two test weeks in the sunny and cloudy seasons have been shown in Figs. 16 and 17, respectively. Besides, Table 3 shows the clear sky ratio i.e. the ratio of the solar irradiance at the ground level to the clear sky irradiance for the two test weeks in the cloudy season. From this table it is observed that for week 1, the clear sky ratio ranges from about 92–93% at dotted cloudy (almost sunny) days to 65.8–72.6% at semi-cloudy days and 44.6% at cloudy day. For week 2, almost the same range is observed while the dotted and semi-cloudy days are more and the clear sky ratios of the days follow a different sequence. That is, in week 1, the first two days and the last two days are dotted cloudy days while the intermediate three days are semi-cloudy and cloudy days. In week 2, the first and the last days are cloudy while the five intermediate days are dotted or semi-cloudy days with various clear sky ratios. Therefore, the two test weeks take into account the sunny, dotted, semi-cloudy and cloudy days into account, simulta-

![Fig. 14. The schematic diagram of the proposed bi-level prediction model for solar irradiation in the cloudy season.](image)

![Fig. 15. The membership functions of the three inputs for ANFIS after training.](image)

<table>
<thead>
<tr>
<th>Method</th>
<th>MAPE%</th>
<th>Convergence iteration</th>
<th>Convergence Time (s)</th>
<th>Values of ( \beta_1, \beta_2 ) and ( \beta_3 ) in Eq. (13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>13.35</td>
<td>53</td>
<td>2.002</td>
<td>( \beta_1=0.641, \beta_2=0.352, \beta_3=0.004 )</td>
</tr>
<tr>
<td>PSO</td>
<td>13.34</td>
<td>36</td>
<td>2.880</td>
<td>( \beta_1=0.653, \beta_2=0.326, \beta_3=0.018 )</td>
</tr>
<tr>
<td>ABC</td>
<td>13.16</td>
<td>59</td>
<td>2.707</td>
<td>( \beta_1=0.666, \beta_2=0.323, \beta_3=0.008 )</td>
</tr>
<tr>
<td>COA</td>
<td>13.16</td>
<td>103</td>
<td>1.250</td>
<td>( \beta_1=0.666, \beta_2=0.323, \beta_3=0.008 )</td>
</tr>
<tr>
<td>FPA</td>
<td>13.16</td>
<td>75</td>
<td>0.652</td>
<td>( \beta_1=0.666, \beta_2=0.323, \beta_3=0.008 )</td>
</tr>
</tbody>
</table>
neously with different sequences. As shown in Figs. 16–17, the hybrid model has been able to predict the solar irradiation very well in both seasons. In order to evaluate the proposed method numerically and show its effectiveness several evaluation metrics has been used. These metrics are the modified mean absolute percentage error (\(\text{Mod\_MAPE}\)) and the mean absolute error (\(\text{MAE}\)) defined as follows [58]:

\[
\text{Mod\_MAPE}\% = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{I_{\text{real}}(i)-I_{\text{predicted}}(i)}{A_v} \right| \times 100
\]  

(16)

\[
A_v = \frac{1}{N} \sum_{i=1}^{N} I_{\text{real}}(i)
\]  

(17)

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |I_{\text{real}}(i)-I_{\text{predicted}}(i)|
\]  

(18)

where, \(I_{\text{real}}\) is real irradiation data and \(I_{\text{predicted}}\) is the predicted one; and \(N\) is number of test points of each season which is here equal to 168. Tables 4 and 5 show the prediction error indices of the solar irradiation in the sunny and cloudy season for the proposed method in comparison with several other methods (as described in Appendices B to D). The structure and specifications of these models has been described in these tables. From these tables, it is shown that the proposed method performs considerably better than the other methods in forecasting the solar irradiation time series in both sunny and cloudy seasons.

Table 3
The clear sky ratio for the cloudy test weeks.

<table>
<thead>
<tr>
<th>Day</th>
<th>Clear sky ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Week 1</td>
</tr>
<tr>
<td>1</td>
<td>0.933</td>
</tr>
<tr>
<td>2</td>
<td>0.931</td>
</tr>
<tr>
<td>3</td>
<td>0.726</td>
</tr>
<tr>
<td>4</td>
<td>0.658</td>
</tr>
<tr>
<td>5</td>
<td>0.446</td>
</tr>
<tr>
<td>6</td>
<td>0.928</td>
</tr>
<tr>
<td>7</td>
<td>0.924</td>
</tr>
</tbody>
</table>

Table 4
The error indices of solar irradiation prediction of the proposed method in comparison with a few methods for the sunny season.

<table>
<thead>
<tr>
<th>Method</th>
<th>(\text{Mod_MAPE}%)</th>
<th>(\text{MAE})</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>6.77</td>
<td>22.19</td>
<td>Hidden Layer:2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of nodes: [6,8]</td>
</tr>
<tr>
<td>ANFIS</td>
<td>4.57</td>
<td>14.98</td>
<td>Number of membership functions: [4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Type of membership functions: {trimf, trimf, trimf}</td>
</tr>
<tr>
<td>LS-SVM</td>
<td>5.27</td>
<td>17.26</td>
<td>Type of kernel function: RBF</td>
</tr>
<tr>
<td>Regression Tuned</td>
<td>4.14</td>
<td>13.56</td>
<td></td>
</tr>
<tr>
<td>with LSE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression Tuned</td>
<td>3.63</td>
<td>11.89</td>
<td>Initial population of 20</td>
</tr>
<tr>
<td>with FPA (proposed)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 16. Prediction of solar irradiation in the sunny season (dash) in comparison with the real value (line).

Fig. 17. Prediction of solar irradiation in the cloudy season (dash) in comparison with the real value (line) (a) for test week 1, (b) for test week 2.
Appendix B

ANFIS

Jang first introduced the ANFIS method by embedding the Fuzzy Inference System (FIS) into the framework of adaptive networks [60]. ANFIS network is used to model highly nonlinear systems and functions and mathematical models with high accuracy [60]. ANFIS network structure is demonstrated in Fig. B.1. There are five layers which perform as follows:

Layer one: Fuzzification process.
Layer two: Runs the fuzzy and multiplies the incoming signals and sends the product out.
Layer three: Normalizes the MF’s.
Layer four: Conclusion part of the fuzzy rules.
Layer five: Calculates the output by adding up the outputs of the fourth layer which is the defuzzification process.

Relations between different layers are as follows [61]:

\[ w_i = \mu_{\alpha_i}(x_1) \times \mu_{\beta_i}(x_2) i=1, 2 \]  

(B.1)

As \( \mu(x) \) is the membership function.

5. Conclusions

In this paper, nonlinear time series analysis and short-term prediction of solar irradiation have been considered simultaneously. Seeking for the predictability of the data and in order for constructing a proper forecasting model, at first, the properties of the solar irradiation has been characterized via time series analysis methods. For this purpose, initially, the solar irradiation time series has been normalized using the daily clear sky irradiation data which is an annually periodic time series. Then, the properties of normalized solar irradiation has been investigated via recurrence plots, as well as autocorrelation and mutual information analysis. The employed data are five years irradiation data of Qazvin city in Iran. Based on these analyses, each year has been divided into two seasons, the sunny and cloudy seasons which are noticeably different in dynamics. In continue, a hybrid but simple model has been synthesized to predict the solar irradiation in different seasons. Finally, the performance of the proposed has been compared with several well-known modeling schemes. The results show that the proposed method performs considerably better than the other methods in forecasting the solar irradiation time series in both sunny and cloudy seasons.

Therefore, the main contribution of the paper is to express a successful application of a hybrid modeling method based on the combination of time series analysis with evolutionary and swarm optimization methods for solar irradiation forecasting. Employment of such a modeling approach makes the proposed model structurally simple with small number of model inputs and parameters which are solely determined based the measured values of the solar irradiation time series. Therefore, no additional metrological measurement is needed for prediction. The other feature of the proposed method is its systematic approach which makes it possible to repeat the same procedure for any regional solar irradiation time series. The simple structure, good accuracy and systematic approach makes the proposed method quite applicable for the wide range of solar irradiation forecasting as it is required in forecasting of the generated power in solar power plants, i.e. photovoltaic (PV) arrays.

### Appendix A. Flower pollination algorithm

Flower pollination is the transport of pollen from a male flower to a female bloom. Pollination may take place in the form of biotic or abiotic. About 90% of flowering plants belong to biotic pollination, that is, pollen is transferred by a pollinator such as insects and animals. About 10% of pollination takes abiotic form which does not require any pollinators [41,42]. Based on the concept of flower pollination, Flower pollination algorithm (FPA) is developed by Yang in 2012. Inspired by the flow pollination process of flowering plants are the following rules [43]:

Rule 1: Biotic and cross-pollination is considered as global pollination process with pollen-carrying pollinators performing Lévy flights.
Rule 2: Abiotic and self-pollination are considered as local pollination.
Rule 3: Flower constancy can be considered as the reproduction probability is proportional to the similarity of two flowers involved.
Rule 4: Local pollination and global pollination is controlled by a switch probability \( p \in [0,1] \).

### Table 5
The error indices of solar irradiation prediction of the proposed method in comparison with a few methods for cloudy season.

<table>
<thead>
<tr>
<th>Method</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Average</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mod MAPE%</td>
<td>MAE</td>
<td>Mod MAPE%</td>
<td>MAE</td>
</tr>
<tr>
<td>ANN</td>
<td>17.49</td>
<td>20.64</td>
<td>18.81</td>
<td>22.10</td>
</tr>
<tr>
<td>ANFIS</td>
<td>15.21</td>
<td>18.1</td>
<td>17.92</td>
<td>20.9</td>
</tr>
<tr>
<td>LS-SVM</td>
<td>17.44</td>
<td>20.63</td>
<td>18.6</td>
<td>21.33</td>
</tr>
<tr>
<td>Regression Tuned with LSE</td>
<td>10.22</td>
<td>11.88</td>
<td>11.4</td>
<td>12.51</td>
</tr>
<tr>
<td>Regression Tuned with FPA</td>
<td>13.62</td>
<td>16.24</td>
<td>15.24</td>
<td>17.91</td>
</tr>
<tr>
<td>Hybrid bi-level (proposed)</td>
<td>7.82</td>
<td>9.18</td>
<td>9.91</td>
<td>11.27</td>
</tr>
</tbody>
</table>
\[ \bar{y}_i = \frac{w_i}{w_1 + w_2} \]  
(B.2)

\[ f_i = (px + qy + r_j) i=1, 2 \]  
(B.3)

where, \( r, p \) and \( q \) are consequent parameters.

\[ f = \pi d_1 + \pi d_2 \]  
(B.4)

**Appendix C**

**Artificial Neural Network (ANN)**

An ANN is composed of a number of interconnected neurons which are arranged in a few layers, called input, hidden and output layers (see Fig. C.1). The output of each node is a weighted sum of its inputs added to a constant term called bias. Forecasting with neural networks involves two steps: training and testing. In training, a proper ANN is constructed via some different stages. At first, proper inputs should be selected. This stage is an important stage which is usually done through input-output linear/non-linear correlation analysis and probably the experience of the ANN developer about the underlying system. The selected inputs and the output of the network are then normalized to feed them to the network. The next stage is to choose the numbers of layers and nodes of each layer as well as the transfer function of each neuron. This stage is normally a trial and error stage which is repeated until the best performance of the network is achieved. Final stage in training includes learning of the chosen network. A learning process in the neural network constructs an input–output mapping function between the selected inputs and output. The learning process is done via adjusting the weights and biases at each iteration based on the minimization of an error measure. Thus, learning stage entails an optimization process. There are different learning algorithms in the literature [62]. One of most efficient learning algorithms is the Levenberg–Marquardt algorithm. This algorithm is actually a modified Gauss–Newton method that converges 10–100 times faster than the well-known back propagation algorithm. The details of the Levenberg–Marquardt algorithm can be found in Ref. [62].

![Fig. B.1. The structure of an ANFIS network.](image)

![Fig. C.1. An artificial neural network structure.](image)
Appendix D

Least Square- Support Vector Machine (LS-SVM)

SVM is a powerful method of solving nonlinear classification issues and estimation and modeling functions that are expressed based on statistical learning theory concepts and structure of minimizing risks [63]. LS-SVM method that was introduced for the first time by Suykens [64] gets an overall response by solving a set of linear equations that makes it faster than SVM method. An $N$-plot series of training data such as $(x_i, y_i)$, for $i=1, 2, ..., N$ is assumed; where $x_i$ and $y_i$ are the $i$th member of input and output vector, respectively. Decision making function can be defined as Eq. (D.1) [65]:

$$y(x) = w^T \Phi(x) + b$$  \hspace{1cm} (D.1)

Where, $\Phi(x)$ is a nonlinear function that mapped input space into a high dimensional space, $W$ is the weight vector and $b$ is bias. For the function estimation issues, the fitness function $J$ should become minimum:

$$J = \frac{1}{2} \|w\|^2 + \frac{1}{2}c \sum_{i=1}^{N} \varepsilon_i^2  \hspace{1cm} (D.2)$$

With regard to:

$$y_i = w^T \Phi(x_i) + b + \varepsilon_i \hspace{1cm} (D.3)$$

Where, $C$ is regulation constant and $\varepsilon$ is the training error.

The final model obtained with the LS-SVM for function estimation is defined as Eq. (D.3):

$$y(x) = \sum_{i=1}^{N} \alpha_i \sum_{j=1}^{N} K(x_i, x_j) + b \hspace{1cm} (D.3)$$

That $\alpha_i$ is Lagrange coefficients and is obtained from minimizing following function:

$$L(w, b, \varepsilon, \alpha) = \frac{1}{2} \|w\|^2 + \frac{1}{2}c \sum_{i=1}^{N} \varepsilon_i^2 - \sum_{i=1}^{N} \alpha_i [w^T \Phi(x_i) + b + \varepsilon_i - y_i] \hspace{1cm} (D.4)$$

$K(x_i, x_j)$ is Kernal function and RBF-Kernal function which is used in this paper is defined as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \hspace{1cm} (D.5)$$

To predict by LS-SVM, LS-SVMLAB Toolbox of MATLAB software has been used [66].

References


