



A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco

Brahim Belmahdi^{1,a} , Mohamed Louzazni², Abdelmajid El Bouardi¹

¹ ETEE, Faculty of Science, Abdelmalek Essaadi University, B.P 1717, Tetuan, Morocco

² Science Engineer Laboratory for Energy, National School of Applied Sciences, Chouaib Doukkali University, El Jadida, Morocco

Received: 6 June 2020 / Accepted: 5 November 2020

© Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract The aim of this paper is to propose and introduce an innovative hybrid model to forecasted the daily global solar radiation (DGSR) in three different cities in Morocco. The DGSR depends on several meteorological parameters and it is difficult to determine its behaviour. The used data have been collected from three different cities with different meteorological data. In this context, we compared 3-year forecasting performance of two methods widely used in forecasting research with a hybrid method. The first model uses the artificial neural network (ANN) and the second one, known as autoregressive integrated moving average (ARIMA), is based on both statistic regression and time series analysis. The third model is based on the combination between ANN and ARIMA methodologies. The three methods have been trained within the same experimental data, thereby allowing a much needed homogeneous correlation that is completely missing in the existing research. The forecasted DGSR generated by the hybrid ARIMA–ANN approach shows a high correlation with experimental results and a relatively small error rate. The forecast of DGSR obtained by hybrid ARIMA–ANN method is compared with the single ANN and ARIMA methods and experimental data. In order to be more accurate, an empirical error is measured and contrasted to check the significance of the expected outcomes and the accuracy of the model. In order to assure results reliability, statistical errors are computed and compared to verify the validity of the forecasted results and the performance of the model. The compared results through statistical error are very accurate compared with a single model in term of R^2 , MBE, RMSE, NRMSE, MAPE, TS and Sd. Generally, the obtained value using the hybrid model is the most adequate which can adjust the experimental data with satisfactory precision. In the end, the ARIMA–ANN model can be used in forecasting photovoltaic power or temperature.

1 Introduction

Recently, renewable energy (RE) is recognized to become one of the most successful renewable energy resources with the capacity to change the country's power profile. The RE has been predominantly produced from several sources such as thermal, solar sunlight, wind, hydraulic and so on. Actually, the various country produce the electrical energy from RE for significantly reduced environmental damage, in fact the greenhouse effect has already

^a e-mail: belmahdi.brahim@gmail.com (corresponding author)

Table 1 Time scale category of forecasting

Forecast nature	Horizon	Actions
Very short-term or ultra-short-term forecasting	Seconds to 30 min.	Power system and smart grid planning [12]
Short term	Between 30 and 360 min, 1 day, or 7 days [13]	Electricity market, power system and control of renewable energy [12]
Medium-term	6–24 h, 1 day, 1 week and 1 month [12]	Maintenance scheduling of conventional or solar energy integrated power systems
Long term	24 h, 1 month to a year [13]	Power generation, transmission, distribution and solar energy rationing [14]

encouraged the discovery of new alternative energy systems [1]. Since the strong growth of the above resource is expected to enhance government, it is also important to identify the opportunities, difficulties and prospective of RE development. In Morocco, the government has implemented various RE projects to generate 42% of electricity by 2020, focusing primarily on solar and wind energy technologies [2]. The main difficulty in the development of these technologies is to employ solar energy as an electricity source in the photovoltaic generator (PVG), thermal solar energy (TSE) and solar concentration technologies (CPV). This difficulty has encouraged scientists and academics to identify the effective mechanisms for forecasting the value of solar radiation, since the generation production of solar photovoltaic system depends directly on global solar radiation [3]. We noticed that the precision of solar radiation models is closely linked to the performance of modelling the power production of installed solar systems and influences the management and scheduling of future sustainable energy installations [4]. Through the use of an appropriate model for forecasting solar radiation, it is feasible to control the power provided by the photovoltaic system. In fact, the assessment, evaluation and forecasting of global solar radiation are necessary due to the great importance for the performance of PVG, TSE and CPV in electrical energy production and its integration in the local electrical grid. In terms of improving and confirming that the power generated from RE source is well injected into the electrical grid without perturbation. Throughout this regard, the forecasting of global solar radiation will have an important impact on the development and management of future energy systems. Forecasting serves a significant task in controlling the performance of the electricity grid [5]. Different methods to forecast global solar radiation have been developed [6]. They depend on both available data and their specific forecasting horizon. Different forecasting categories are summarized in Table 1 [7]. Several scientists recommend three categories of the forecast horizon: short-term, medium-term and long-term [8]. Some others proposed a fourth category [9] depending on the criteria of the decision-making phase for intelligent or microgrids [10, 11], appropriately referred to as the very short-term or ultra-short-term prediction horizon. Nevertheless, no commonly accepted standard has been set.

Countless techniques for forecasting DGSR have been documented in the different scientific references. Predicting methodologies are commonly organized into three major categories, namely traditional models [15], machine learning and statistical regression methods [16]. The use of a combined methodology is suggested to enhance the performance of each single model. The traditional models, also called statistical or mathematical models, can be

classified into dynamic [17] and empirical [18] models. Dynamic models were used to forecast global solar radiation in long-term durations. Among empirical models, those based on the use of accessible weather data as inputs were usually preferred, thanks to their low computational cost and easy data requirement [15]. The basic principle on which they ground is the association between global solar irradiation and weather and/or climatic parameters, such as sunshine duration and air temperature. Hargreaves et al., [19] presented the first empirical model using temperature by assuming the temperature difference ΔT assigned to global solar radiation. Meanwhile, Bristow et al. [20] proposed the exponential relationship between global solar radiation and ΔT (Bristow-Campbell model), which might describe 70-90% of global solar variability in America. Later, Hassan et al. [21] compared 3 newly designed models and 17 existing DGSR models in Egypt, and recorded that the temperature-based model is the most reliable in term of greater forecasting precision with respect to traditional sunshine-based models. The models are tested and evaluated on a 20-years span observed dataset. Results show that the new model is particularly relevant in weather forecasting techniques. On the other hand, Fan et al. [15] examined and reviewed 14 emerging temperature models and introduced 6 new temperature models in China. The newly developed polynomial model assures reliable DGSR forecasting and it can be implemented in environments in which only air temperatures are available. Also, this model is used to characterize the mathematical equation relationship between global solar radiation and the associated environmental parameters. Commonly, the above methodology does not require historical data, instead depends heavily on detailed station placement descriptions and climate conditions. The input measurements are used to identify of DGSR dependent on meteorological conditions. These models can be both simple, if based on solar sunshine duration or more complex, if additional parameters such as temperature, wind speed, dust and relative humidity are included. Therefore, the traditional approaches will not be suitable to forecast global solar radiation throughout any specified temporal and geographical scale, particularly in short term.

Although empiric models are suitable to forecast global solar radiation in different conditions, their findings have been less accurate compared to machine learning models [22]. Recently, quite a variety of machine learning models have been constructed to forecast global solar radiation. Similarly to all artificial intelligence (AI) strategies, machine learning (ML) does not require any priory knowledge of the system, and it can deal with problems that cannot be depicted by concrete algorithms [22], and nowadays is among the most effective methods for time series data forecasting. However, the most adopted machine learning (ML) algorithm in the DGSR forecasting is the artificial neural networks (ANN). [22]. Notton et al. [23] implemented the artificial neural networks (ANN) to forecast both the global horizontal radiation (GHR) and direct normal radiation (DNR) over 1 h ($h + 1$) to 6 h ($h + 6$). From a deep review about the ANN method application in solar irradiation field [6], the suitability of the methodology clearly erases. The conclusions of that kind of assessment underline the ability of the hybrid ANN method.

The third category is based on statistic regression or probabilistic techniques focused on follow-up measurements or determining data, generally used for short-term and very short-term forecasting [24].

These methods rely more heavily on historical data and they are used to evaluate the intrinsic rule of forecasting by analysing past information. In addition, these methods are based on historical records associated with meteorological information, regardless of the fact that the past data will appear in the future [25]. The principal drawback of statistical methods is the fact that they ground on the hypothesis of linear stationary structures that are not suitable for nonlinear solar radiation. Consequently, the forecast performance of the statistical model depends on the time and reliability of the historical data. These methods are also known as

black box. Some of the most widely adopted time series analysis models are autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) [26], autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), autoregressive moving average with exogenous inputs (ARMAX), autoregressive integrated moving average with exogenous data (ARIMAX) [27]. Bacher et al. [28] presented the AR with exogenous inputs model to forecast the hourly value of solar photovoltaic power. They noticed that the AR model is operating well when the prediction period is up to two hours. Benmouiza et al., [26] used a combination between ARMA and nonlinear autoregressive (NAR) to forecast multihours ahead global horizontal solar radiation. The importance of the combination is to enhance the efficiency of the single methods. In fact, ARMA requires a stationary time series, whereas most real time series are not. On the contrary, ARIMA approach does not take into account the mechanism behaviour and incorporates non-stationary elements from time series information [29]. The ARMAX concept does not depend on the forecast of solar radiation, but it is considered as the traditional model of ARIMA [29]. ARIMA models will certainly explain the complexities of the data in a provided application. The effectiveness of the ARIMA model is attributed to its computational characteristics as well as the well-known Box-Jenkins methods in the model construction process [30]. In fact, a collection of analytical smoothing methods can be related to ARIMA models [31]. While ARIMA models are very robust because they can reflect many different patterns of time series, their key weakness is the pre-assumed linear structure of the systems. Forecasting performance may be increased if several independent models are adapted to the same data rather than using a single model. Several hybrid methods have also been suggested in the research, incorporating the benefits of two or more different models.

Currently, numerous researchers have applied a combination of the different single method to enhance the performance of the forecasted results and to overcome various problems such as the nonlinearity and complexity of the weather data. The hybrid technique becomes actually the most used strategy in forecasting due to its ability to forecast complex and nonlinear input parameters. The use of ARIMA–ANN was introduced by [32] to present more accurate forecasting with respect to individual models. The same method was used by Babazadeh et al. [33] to forecast the gasoline price consumption. This technique was applied to several fields such as water quality prediction [34], electricity price [35], stock index returns [36] and global solar radiation [37].

In this background, this paper aims to forecast the global solar radiation for three different regions in Morocco using the hybrid ARIMA–ANN model. The DGSR experimental data are collected from three different cities, namely Er-rachidia, Ifrane and Tanger. The three cities are characterized by different climate zoning from Mediterranean to hot desert climate passing from cold region of the second city. The motivation for using the ARIMA–ANN model is due to its strong capacity to model nonlinear and complex structures through time series analysis. The results obtained subsequently showed a strong matching with the low error experimental data. Also, the results are compared with single ARIMA model, ANN model and experimental data, respectively. In order to evaluate the results, specific error parameters are also determined and examined such as mean absolute percentage error (MAPE), mean bias error (MBE) and percentage MBE, root mean square error (RMSE) and percentage RMSE, standard deviation (SD) and percentage SD, normal root mean square error (NRMSE) and T-statistic (TS). The overall performance is evaluated by determination coefficient (R^2). Another analysis is based on the linear regression coefficients.

This paper is structured as it follows: Sect. 2 describes the material and methodology through the experimental data and the ANN, ARIMA and ARIMA–ANN models. Sections 3

and 4 present the evaluation of forecast performance and the empirical results and discussions, respectively. Concluding remarks are reported in Sect. 5.

2 Materials and methods

As part of its venture regarding electricity consumption, Morocco gives priority to increasing clean energies and sustainable development. Morocco has a very high solar potential, more than $2600 \text{ kWh/m}^2/\text{year}$ and connected with Spanish network via 400kv and 700 MW power lines. The Moroccan government has installed various renewable energy projects to get the target of 42% of electricity from sustainable energy by 2020 and 52% by 2030 [2]. The most important project aims to generate 2 gigawatts in five major projects installed in Ouarzazate, Ain Bni Methar, Boujdour, Foug Al Oued and Sebkhatah, using photovoltaic and concentrated solar power. A three-case study is chosen from the installed project named Er-rachidia, Ifrane and Tanger. The selected locations are evaluated and investigated, according to a variety of research analyses of the Moroccan environment, a government agency specializing in sustainable energy and efficiency has created an environment zone [38].

2.1 Experimental data

The three selected cities Er-rachidia, Ifrane and Tanger are characterized by different climate conditions. The first Er-rachidia (latitude = 31.930° , longitude = -4.424° , altitude = 1080 m) is characterized by a hot desert climate, dry and mostly clear year. Generally, temperatures vary from 3°C to 38°C . The summers in Ifrane (latitude = 33.500° , longitude = -5.167° , altitude = 1665 m) are short, warm and arid, the sky is mostly clear around the year. The temperature frequently varies from -3°C to 28°C . Finally, in the third city (Tanger) (latitude = 35.580° , longitude = -5.900° , altitude = 21 m), the summers are warm, humid and arid, and the sky is mostly clear, and the temperature typically varies from 9°C to 29°C .

The collected data have been measured every 10 min by Pyrometers instrument from Kipp and Zonen type CM11 (Fig. 1) [39] at three locations. The Pyranometer is characterized by excellent linearity, fast response time and low tilt error.

The historical measured data covered the period from January 2013 to December 2015. An example of the DGSR on a horizontal surface is illustrated in Fig. 2.

Figure 2 shows regular annual fluctuation in DGSR. From this figure, it emerges that the solar radiation time series is a non-stationary, being strongly affected by an annual phenomenon. Annual phenomenon can be differentiated. In fact, while the average of DGSR is generally the same, the regular peak of radiation is variable except on consecutive days.



Fig. 1 Pyrometers Kipp and Zonen CM11 type [39]

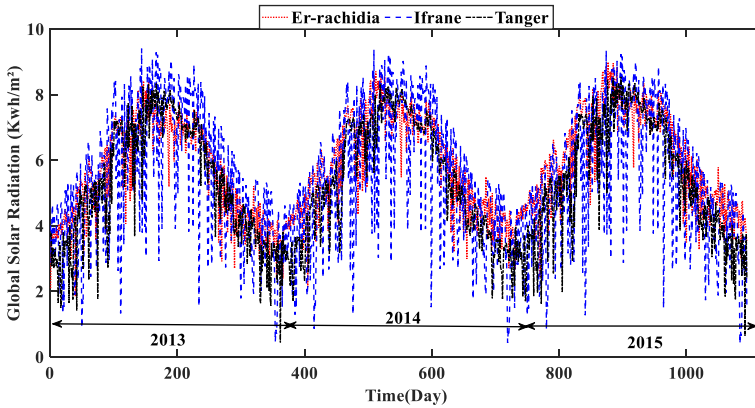


Fig. 2 Example of DGSR of the three selected locations (Er-rachidia, Ifrane and Tanger)

2.2 ARIMA model

The ARIMA model is widely used in several fields (econometrics and engineering) [40]. The ARIMA model calculates the significance of the generated time series as a linear composition from its historical measurements [36]. The common form of the ARIMA (p, d, q) includes a mixture of three types of models: p is the order of the autoregressive (AR) model; d is the degree of differencing to keep data stationary (I); and q is the order of the moving average (MA) model. The general of the ARIMA model is presented in the Eq. 1 [30].

$$\begin{cases} Y_t = (1 - B)^d (1 - B^s)^D X_t - \mu \\ \phi(B) = \varphi(B^s) Y_t = \theta(B) \Theta(B^s) Z_t, Z_t \in N(0, \sigma^s) \end{cases} \quad (1)$$

with $\Phi(z) = 1 - \sum_{i=1}^p \Phi_i z^i$, $\varphi(z) = 1 - \sum_{i=1}^p \varphi_i z^i$, $\Phi(z) = 1 - \sum_{i=1}^p \Phi_i z^i$ and $\Theta(z) = 1 + \sum_{i=1}^q \Theta_i z^i$ where $\emptyset, \theta, \Phi, \Theta$ present the polynomial coefficients, D and s represent the order of differentiation of the seasonal part period, the part of seasonal autoregressive and seasonal moving average part of the model.

The time series forecasting by ARIMA models could be carried out by four steps: classification, approximate, treatment and forecasting [41, 42]. In this paper, the four fundamental steps of ARIMA are selected carefully and have the following order: firstly, we begin with the identification by choosing the best fit model referred to the autocorrelation function (ACF) and partial autocorrelation function (PACF) to evaluate the practicable persistence possible arrangement in the DGSR data. The behaviour of the ACF and PACF analyses allows identifying the ARIMA model that explains the corresponding stationary time series. The second step is the model's approximate input (x) parameters using one of the determination methodologies. The third step is diagnostic; it involves the residual value of the chosen model with findings and measurements of the approved model. The last one is predictive; it produces forecasts and calculates a random error. In addition, the ARIMA is investigated through the Ljung-Box Q test [43], where the insignificant assumption specifies that there is no residual autocorrelation for k lag at the time of the test referred to Q is defined as:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\rho_k^2}{n - k} \quad (2)$$

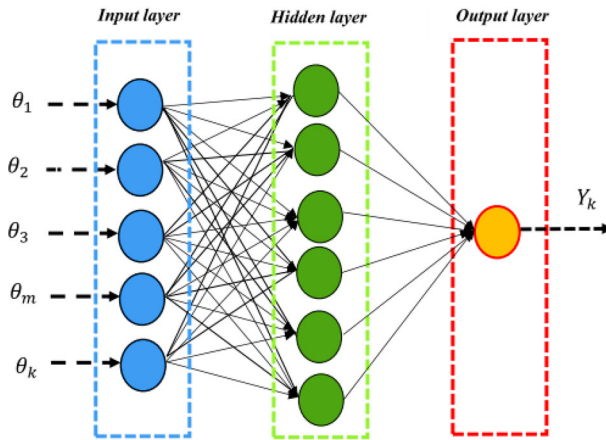


Fig. 3 A comprehensive form of ANN model

where n is the sample size, ρ_k is the sample autocorrelation at lag k , and h is the number of lags being tested.

2.3 Artificial neural network

The ANN is well recognized as an effective AI computing device that has already been largely used in all disciplines such as telecommunication, materials, medicine and neurology fields [44, 45]. The ANN procedure technique is essentially based on the input layer and data acquisition ability named hidden layers for the output layer. ANN has been widely used in single [46] and combined forecasting with statistical regression [47] to forecast photovoltaic system power. Results are better than other techniques [22]. In this context, to overcome several problems, a feed-forward network (FFN) based on a back-propagation learning (BPL) technique was selected to forecast the DGSR at case study cities as illustrated in Fig. 3.

The response variable Y_k is represented as

$$Y_k = \phi_0 \sum_{j=1}^n (w_{kj} + \theta_k) \tag{3}$$

where Φ_0 is the activation function of the hidden layer, Y_k is the output of the hidden layer k th, θ_k represents the bias value of the hidden layer, and w_{kj} is the synaptic weight value from input to x_j to the hidden layer k .

The forecast performance evaluation of the implemented models is categorized into two sample procedures; the first is a learning dataset that is used specifically for model creation, containing all inputs and forecasted outputs, the second is to validate model through testing dataset.

2.4 Proposed hybrid ARIMA–ANN model

ANN can be applied to nonlinear systems, ARIMA to linear ones. The mixing of the two models can overcome several problems of nonlinearity and stationary or non-stationary data [48]. The ARIMA–ANN model is the combination of linear presented by ARIMA and non-linear presented by ANN. In addition, the ARIMA model is adapted to the time series data

forecasting, and the error sequence is assumed to appear as nonlinear function and modelled by applying ANN. Further, the forecasting results are provided by the ARIMA and ANN models, which are combined to get the last step of the estimation. In term of efficiency and performance, this hybrid model is better than the individual ARIMA and ANN models, as illustrated by [32]. According to this paradigm, time series data are presumed to be a collection of linear and nonlinear subsystems, presented in the context by:

$$Y_t = L_t + N_t \quad (4)$$

while L_t represents a linear element and N_t represents a nonlinear element.

The residuals are necessary for identifying the adequacy of linear models and are modelled by ANNs and given as:

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \quad (5)$$

where f is the nonlinear function determined by ANN and ε_t is the random error.

The forecast equation can be written in the following equation.

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (6)$$

In conclusion, the mentioned procedure for the hybrid method constitutes of two steps. The first step is by using the ARIMA model to evaluate the linear aspect of the problem. In the second step, an ANN model is built to model the residuals of the ARIMA model. Since the ARIMA model could not control the nonlinear nature of the dataset, the residuals of the linear model must provide details on nonlinearity. The outputs from the ANN could be used as forecasts of the error conditions for the ARIMA model. The combined model incorporates the special characteristics and reliability of the ARIMA model including the ANN model to assess various developments. This may then be useful to forecast linear and nonlinear metrics independently through different models and then integrate the forecasts to increase overall modelling and forecasting accuracy. The steps of the proposed hybrid ARIMA–ANN model are presented in Fig. 4.

The statistical metrics are used to evaluated the performance of the used ARIMA–ANN model, which are applied in a variety of disciplines to determine the quality of the forecast models [46, 47, 49, 50] (“[Methodology Appendix](#)”). Generally, the assessment of the forecasted models has been based on the analysis of the statistical metrics used to check the accuracy, performance and efficiency of the models. It should be emphasized that the lower value of MAPE, MBE and PMBE, RMSE, PRMSE, Sd PSd, NRMSE prove the accuracy of the forecasted values. The lower values of Ts mean a suitable model’s performance. The Sd represents the ratio between measured and computed values: Sd = 0 means the absence of a linear relationship, while Sd = 1 shows the ideal linear relationship between measured and computed values. Finally, the best correlation coefficient R^2 must be close to 1 as possible.

3 Result and discussion

In this section, the results of the application of the ARIMA–ANN model forecasting method over a 3-years interval are applied for the three selected cities, analysed and discussed. To apply the ARIMA–ANN model requires that the time series is stationary. As it is well known, the global solar radiation presents annual and daily variations and/or oscillations. These periodicities make the time series non-stationary. Many authors [51–53] use these models to make the global radiation time series stationary. In this paper, we use variant of

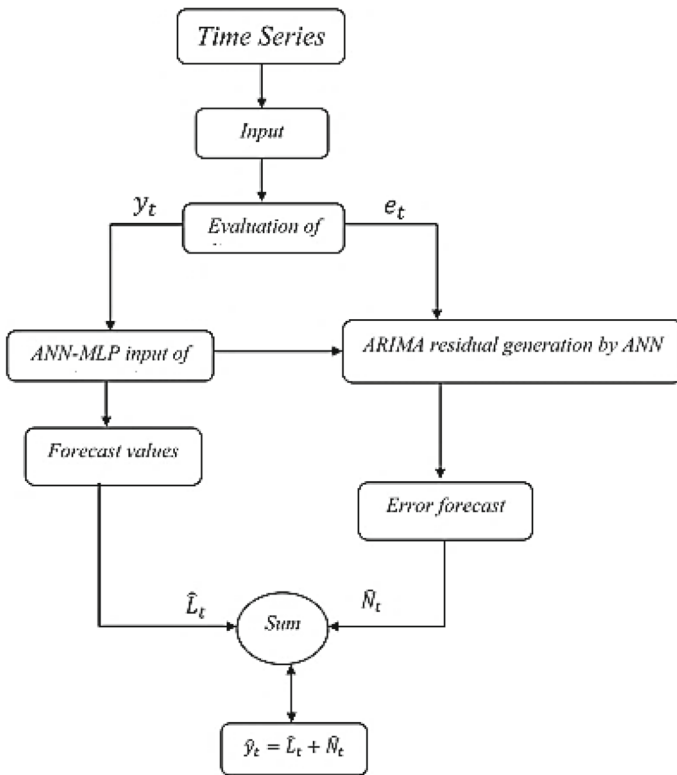


Fig. 4 Flowchart of hybrid ARIMA–ANN

this index, considering only the radiation outside the atmosphere (TOA), this way we get the clearness index (K_t). It is defined as the ratio of the global solar radiation at the earth surface to the equivalent extraterrestrial solar radiation on the earth ground surface (TOA) as described in the foregoing equation [54]:

$$K_t = \frac{DGSR}{TOA} \tag{7}$$

$$TOA = \int^{\text{day}} I_0 E_0 \sin(h) dt \tag{8}$$

where I_0 is the solar constant, h is the solar elevation and E_0 is the Earth–Sun distance correction.

This technique does not completely make the global solar radiation stationary. To tackle this problem, we have completed our method by using variation coefficients Cv_x :

$$Cv_x = \frac{\sigma}{\mu} \tag{9}$$

where σ is standard deviation and μ the average.

Figure 5 presents the results of the application of our global stationary methodology and its impact on the time series. Before any treatment (step 1), the variation coefficient of the time series is high ($Cv_x \sim 0.57$), while in steps 2 and 3, this coefficient is divided by two

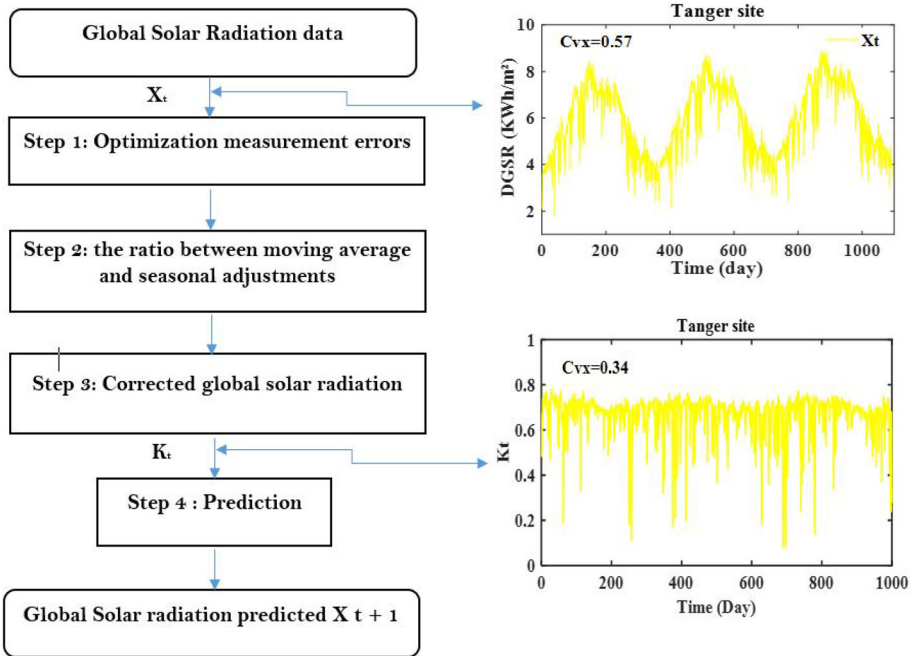


Fig. 5 Example of the daily behaviour of clearness index of Tanger site

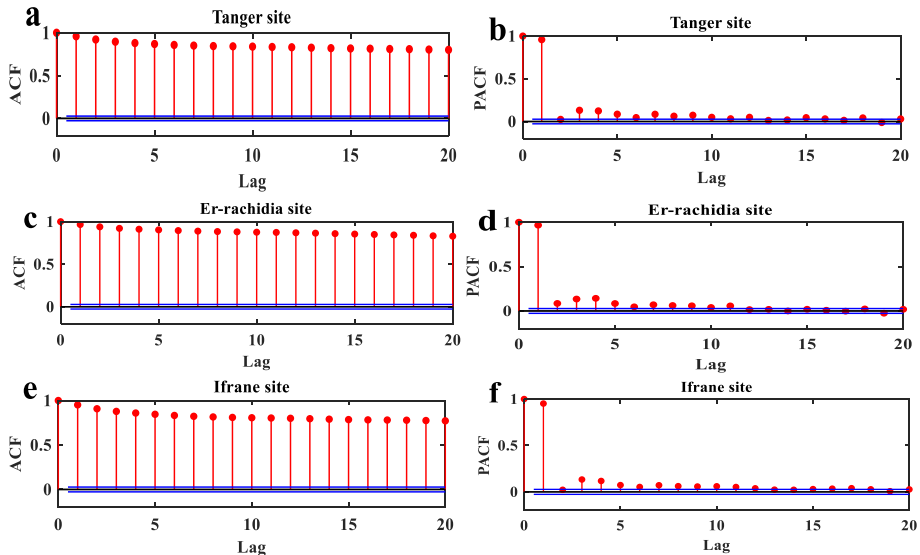


Fig. 6 ACF and PACF of clearness index time series, a–b Tanger, c–d Er-rachidia, and e–f Ifrane

($Cv_x \sim 0.34$). This coefficient and the shape of the curves tend to show that there is a better stationarity at the end of steps 2 and 3.

The corresponding ACF and PACF are shown in Fig. 6 for Tanger (panels a–b), Er-rachidia (panels c–d) and Ifrane (panels e–f), respectively.

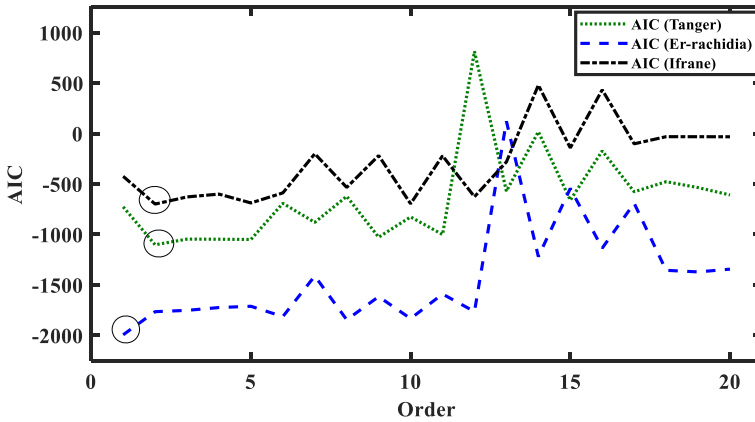


Fig. 7 Presentation of the AIC criterion of Tanger, Er-rachidia and ifrane

Table 2 ARIMA (2,1,1), ARIMA (1,1,1), ARIMA (2,1,1) models parameters

Sites	Parameter	Estimate	Standard error	T-Statistic	P value
Tanger	AR(1)	0.01569	0.033372	0.47017	0.63824
	AR(2)	0.91329	0.032123	28.431	0.0000
	MA(1)	0.97087	0.024834	39.093	0.0001
Er-rachidia	AR(1)	0.5936	0.045332	13.0946	0.0000
	MA(1)	-0.27977	0.050392	-5.5518	0.0002
Ifrane	AR(1)	0.98815	0.16427	6.0153	0.0017
	AR(2)	-0.17639	0.080795	-2.1832	0.0290
	MA(1)	-0.60461	0.15888	-3.8054	0.0001

It's highlighted that after a few lags, the ACF accumulates within 95% of the limit, suggesting a relatively stationary time series. For all the analysed cities, PACS has a major spike at LAG = 1, suggesting that an AR (2) or any of the higher order autocorrelation may be sufficient.

The akaike intelligence criterion (AIC) is described as the most commonly used. The criteria of goodness-of-fit based on the information criterion are presented in Eq. (10):

$$AIC = -2 \ln(L) + 2(p + q + K + 1) \tag{10}$$

where p is autoregression parameters, q is moving average parameters, L is likelihood, k is number of model parameters. The computed results of the AIC are showed in Fig. 7.

AIC results showed that the values for Tanger reach a lowest error when the order is equal to two. Thus, the correct configuration model for Tanger site is ARIMA (2,1,1); analogously ARIMA (2,1,1) is the best model for Ifrane site. AIC values for Er-rachidia site reach minimum when the order is equal to one. The appropriate ARIMA model FOR Er-rachidia is thus ARIMA (1,1,1). AIC results are shown in Fig. 7. The obtained data are reported in Table 2.

When the model fit is adequate and its parameters are forecasted, the diagnostic assessment for the residuals is applied to check if they fit well the data series. Throughout this evaluation examination, we investigate if the residual model collected from the ACF and PACF graphs is

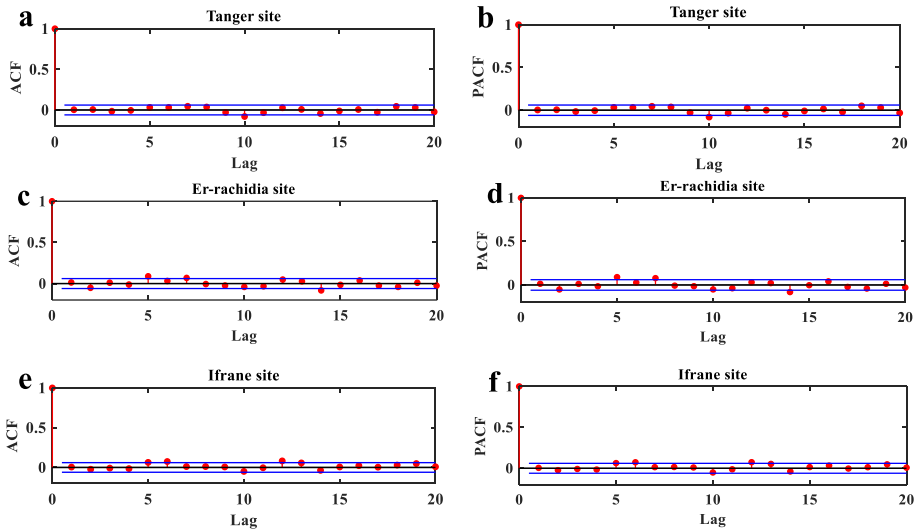


Fig. 8 ACF and PACF of the ARIMA (2,1,1), ARIMA (1,1,1) and ARIMA (2,1,1) residuals for **a–b** Tanger, **c–d** Er-rachidia, and **e–f** Ifrane

Table 3 Ljung-Box test of ARIMA (2, 1, 1), ARIMA (1, 1, 1) and ARIMA (2, 1, 1) models

Sites	Lag k	DF	Q -statistic	Chi square	P value
Tanger	8	6	6.1040	12.5916	0.4116
	17	15	19.7904	24.9958	0.1801
	22	20	28.4277	31.4104	0.1001
	38	36	46.4479	50.9985	0.1139
	45	43	54.2305	59.3035	0.1171
Er-rachidia	8	7	14.0671	18.2653	0.1080
	17	16	26.2962	35.5573	0.0334
	22	21	32.6706	39.4977	0.0867
	38	37	52.1923	53.9528	0.3551
	45	44	60.4809	61.4815	0.4178
Ifrane	8	6	7.3767	12.5916	0.2874
	17	15	21.8050	24.9958	0.1130
	22	20	26.5083	31.4104	0.1497
	38	36	46.9870	50.9985	0.1040
	45	43	51.3172	59.3035	0.1799

IID (independent and identically distributed). Figure 8 shows the ACF and PACF behaviour of the established residuals ARIMA (2, 1, 1), ARIMA (1, 1, 1) and ARIMA (2, 1, 1) models. As we can conclude that most of the significant increases are within the 95% CI.

However, the residuals model was evaluated by the Ljung-box analysis, and the obtained results are listed in Table 3. All chi-squares χ^2 are larger than Q statistics and all p values for lag numbers are more significant than 0.05. From the previous analysis, it follows that residuals are uncorrelated and represent white noise.

The correlation between experimental and forecasted data using ARIMA models is presented in Fig. 9. The circles show the experimental data points, and the line indicates the

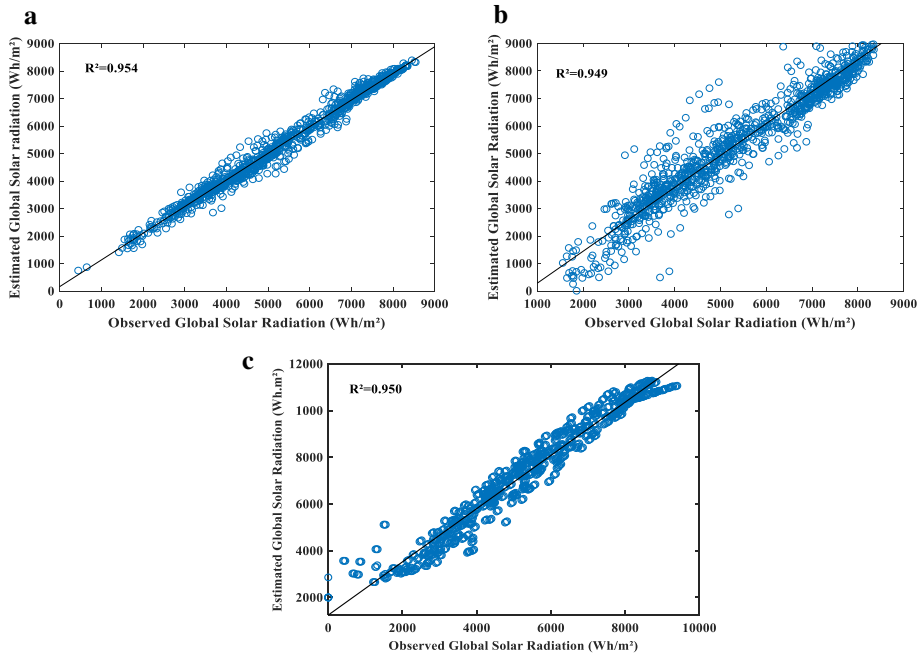


Fig. 9 Regression plot between experimental and forecasted DGSR obtained by ARIMA models for Tanger (a), Er-rachidia (b) and Ifrane (c)

relatively better match of the training data derived from the forecasted GSR. As shown from Fig. 9, the coefficient of determination is close to 1 for Tanger, Er-rachidia and Ifrane are 0.954, 0.949 and 0.950, respectively.

Previous results show that the chosen hidden layers scheme in the ANN can handle all data if the right number of neurones is selected [55–57]. In this study, a three-layer neural network is chosen for forecasting the clearness index. Based on Eq. (23), the current parameters of the ANN model are given as follows:

- The input neurons correspond to the number of lagged observations;
- The number of the output layer is one. After several trials, we found that the optimum neural network for Er-rachidia is one input, one hidden layer with one neuron and one output ($1 \times 1 \times 1$), for Tanger and Ifrane, the optimum network is two inputs, one hidden layer with two neurons and one output layer ($2 \times 2 \times 1$).

Figure 10 shows regression correlation analysis of the forecasted values using ANNs model. In term of accuracy, the ANNs have improved the performance in comparison with ARIMA model. The presented correlation coefficients reach to 0.969, 0.958 and 0.959 in Tanger, Er-rachidia and Ifrane, respectively.

The residuals of the ARIMA model which has nonlinear part are used as input to the multi-layer perceptron of ARIMA–ANN model, and the Levenberg–Marquardt (LM) is the trained algorithm. While the outputs have been normalized inside [1]. The proposed hybrid model has the particularity of using both the strength of ARIMA and ANN models to determine different patterns. As we can see from Fig. 11, ARIMA–ANN model performs better with respect to the single ARIMA and ANN models. The present coefficient of determination for Tanger, Er-rachidia and Ifrane is 0.986, 0.988, and 0.984, respectively. According to linear

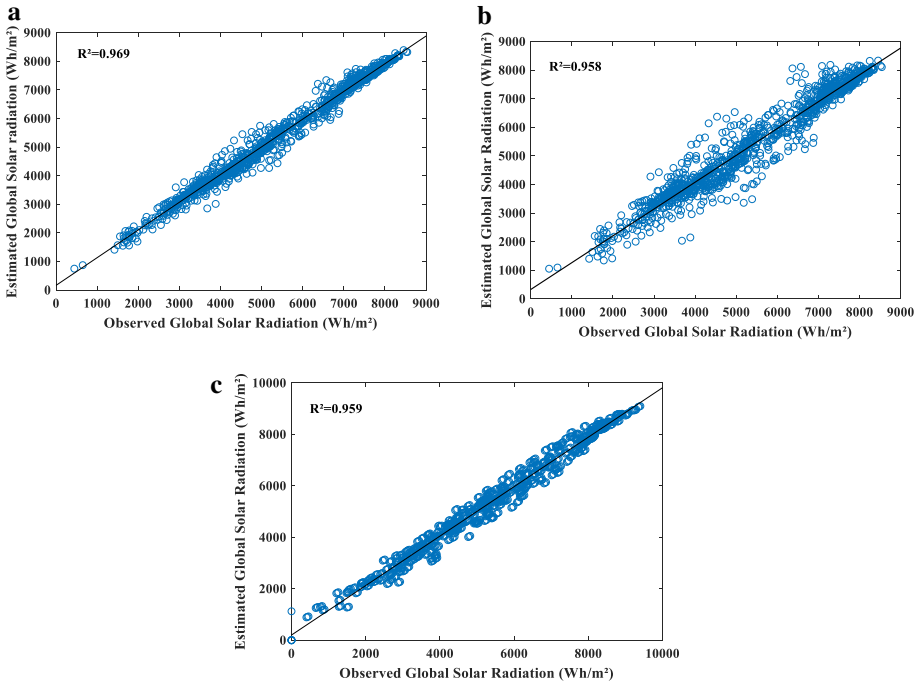


Fig. 10 Regression plot between experimental and forecasted DGSR obtained by ANN models for Tanger (a), Er-rachidia (b) and Ifrane (c)

regression analysis results, the forecasted values obtained by merging ARIMA and ANN models match better the measured data with respect to single ARIMA and ANN models. The correctness of the suitable forecasting model of DGSR is selected from the compared results obtained by ARIMA, ANN and hybrid ARIMA–ANN models, respectively. The conclusion is based on several compared terms such as statistical error measures. The forecasted values obtained by ARIMA and ANN with the measured data were better than the single model ARIMA and ANN models.

Table 4 shows several statistical indices which were computed to check previous results: MBE, RMSE, NRMSE, MAPE, TS, Sd, PSd and linear regression coefficients (R^2 , a, b) (see “[Methodological Appendix](#)” for more details). This evaluation also provided the benefit of determining which model values are statistically important or not at a given degree of level.

In ARIMA modelling, the MBE, PMBE, RMSE, PRMSE, NRMSE, MAPE, TS, Sd, PSd and linear regression coefficients for Tanger site are -3.430 Wh/m^2 (-0.064%), 713.365 Wh/m^2 (13.275%), 0.133 , 44.102 , 0.499 , 713.610 Wh/m^2 (13.280%) and linear regression are 1.169 , -328.159 , these indicators were calculated as -1.110 W h/m^2 (0.019%), 662.724 Wh/m^2 (11.363%), 0.114 , 44.483 , 0.554 , 663.026 Wh/m^2 (11.368%) and 1.194 , -497.753 , respectively for Er-rachidia. For Ifrane site, the MBE (PMBE) is -2.560 (-0.047%), the RMSE (PRMSE) and NRMSE are 1475.166 Wh/m^2 (27.215%) and 0.272 , the MAPE, TS, Sd (PSd) and linear regression are 105.960 , 1.073 , 1475.883 Wh/m^2 (27.228%) and 1.517 , -408.896 , respectively. Applying the ANN model, the obtained values of R^2 are very close to 1 representing optimum and best correlation between the forecasted and measured values. The PMBE, PRMSE and PSd range, respectively, from -0.301 to -0.065 , from 8.368 to 25.030

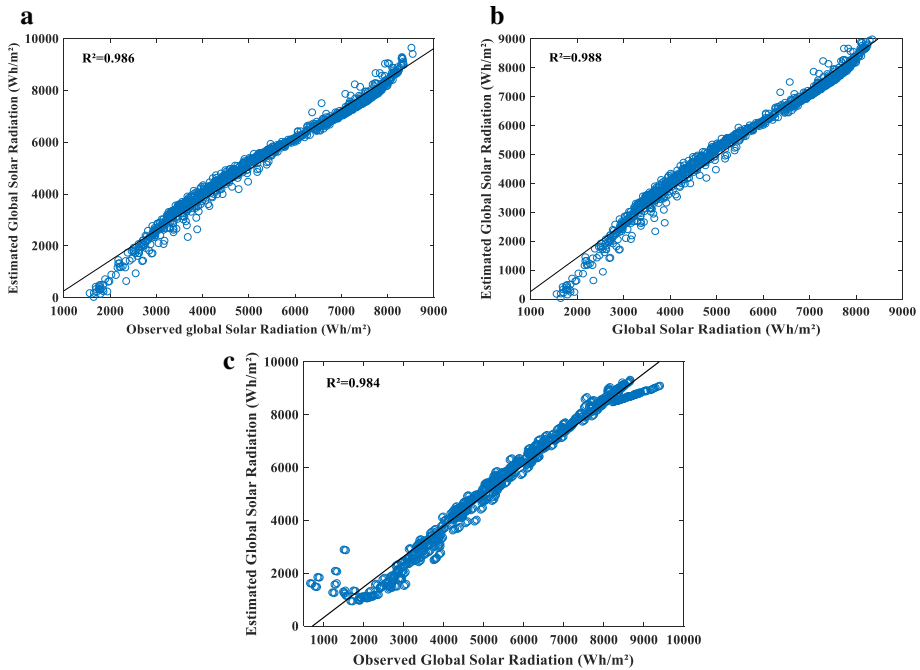


Fig. 11 Regression plot between experimental and forecasted DGSR obtained by hybrid ARIMA–ANN models for Tanger (a), Er-rachidia (b) and Ifrane (c)

and from 8.372 to 25.040 for optimum and suitable configuration of Tanger, Er-rachidia and Ifrane. The MBE, RMSE, NRMSE, MAPE, TS and Sd range, respectively, from -16.338 Wh/m² to -3.531 Wh/m², from 449.670 Wh/m² to 1356 Wh/m², from 0.084 to 0.250, from 29.793 to 97.227, from 0.246 Wh/m² to 574 Wh/m² and from 449.542 Wh/m² to 1357.228 Wh/m² for three sites. The value of MBE and NRMSE is very close to 0 indicating the accuracy between estimated and measured DGSR value. The constant ‘b’ is very close to 0 representing the perfect linear fit and linear relationship between the forecasted and the measured values. Other results are described in the same table. In the cases of hybrid model, the statistical indicator values for optimum configuration of Tanger are: -10.765 Wh/m² for MBE, 446.352 Wh/m² for RMSE, 0.083 for NRMSE, 25.544 for MAPE, 0.252 Wh/m² for TS and 446.862 Wh/m² for Sd these indicators were calculated as -0.084 for PMBE, 7.391 for PRMSE and 7.394 for PSD for Er-rachidia. For Ifrane site, MBE, RMSE, NRMSE, MAPE and linear regression are -18.899 Wh/m², 582.882 Wh/m², 0.108, 42.936 and 0.921, -3211.117 , respectively. The correlation between the used models in terms of precision and accuracy demonstrates that the hybrid ARIMA–ANN model has lower values than the single ARIMA and ANN approaches, respectively.

The current study results were compared to several literature recent works which use different models (Table 5). The highest R^2 value, corresponding to the current study, suggests that hybrid ARIMA–ANN performs well than other existing models.

Figures 12, 13 and 14 show the forecasted daily global solar radiation compared with the single ANN and ARIMA. For the three selected location, the hybrid ARIMA–ANN model has given higher accuracy and perform better than single ARIMA and ANN models, and is more effective with the experimental data values.

Table 4 Statistical analysis results for the optimum and suitable ARIMA, ANN and hybrid ARIMA-ANN models

Models	MBE	RMSE	NRMSE	MAPE	TS	Sd	R ²	A	B	
Tanger	ARIMA(2,1,1)	- 10.765	713.365	0.133	44.102	0.499	713.610	0.954	1.161	- 874.346
	ANN	- 3.531	449.670	0.084	29.793	0.262	449.542	0.969	0.938	328.159
	Hybrid	- 3.430	446.352	0.083	25.544	0.252	446.862	0.986	1.169	- 905.756
Er-rachidia	ARIMA(1,1,1)	- 1.110	662.724	0.114	44.483	0.554	663.026	0.949	1.194	- 1130.411
	ANN	- 4.901	510.450	0.088	30.730	0.376	510.669	0.958	0.915	497.753
	Hybrid	- 3.791	431.078	0.075	30.413	0.246	431.247	0.988	1.279	- 1628.165
Ifrane	ARIMA(2,1,1)	- 2.560	1475.166	0.272	105.960	0.574	1475.838	0.950	1.517	- 2802.222
	ANN	- 18.899	1356.706	0.250	97.227	1.073	1357.228	0.959	0.921	408.896
	Hybrid	- 16.338	582.882	0.108	42.936	0.398	582.842	0.984	1.595	- 3211.117

Table 5 Comparative study between optimum hybrid model and many existing models in the literature

Reference numbers	Method	Study area	Coefficient of determination (R^2)
[58]	ANN-PSO	Saudia Arabia	0.772
[59]	ARMA-ANN	France	0.782
[60]	SVR-POLY	Iran	0.786
[60]	SVM-RFB	Iran	0.791
[61]	SVM-WT	Iran	0.908
[62]	ANN-ANFIS	Iran	0.911
[63]	NN-ARX	Turkey	0.913
[64]	ARMA-ANN	Algeria	0.914
[65]	ANN-MTM	Algeria	0.924
[66]	GP-SA	Iran	0.925
[67]	WRF-KALMAN FILTER	China	0.957
[68]	ANN-WAVELET	Abu Dhabi	0.963
[69]	MLP-model	Algeria	0.977
The current study	ARIMA-ANN	Morocco	0.988

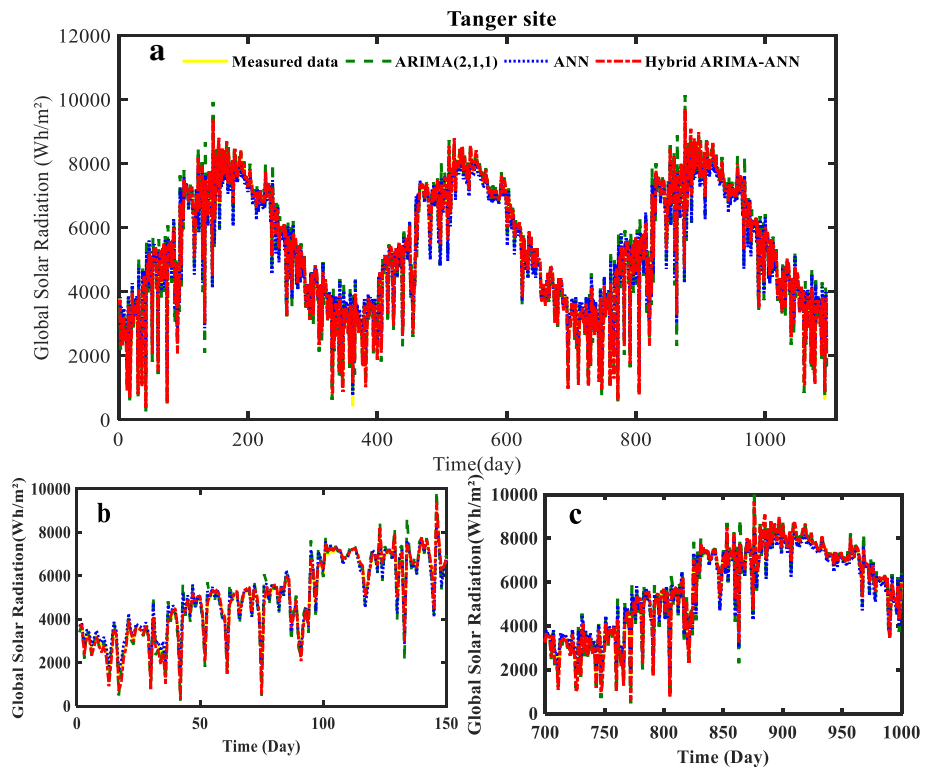


Fig. 12 Compared experimental and forecasted DGSR values for ARIMA (2,1,1), ANN and hybrid ARIMA-ANN for Tanger site (a). b, c Are enlargement of a

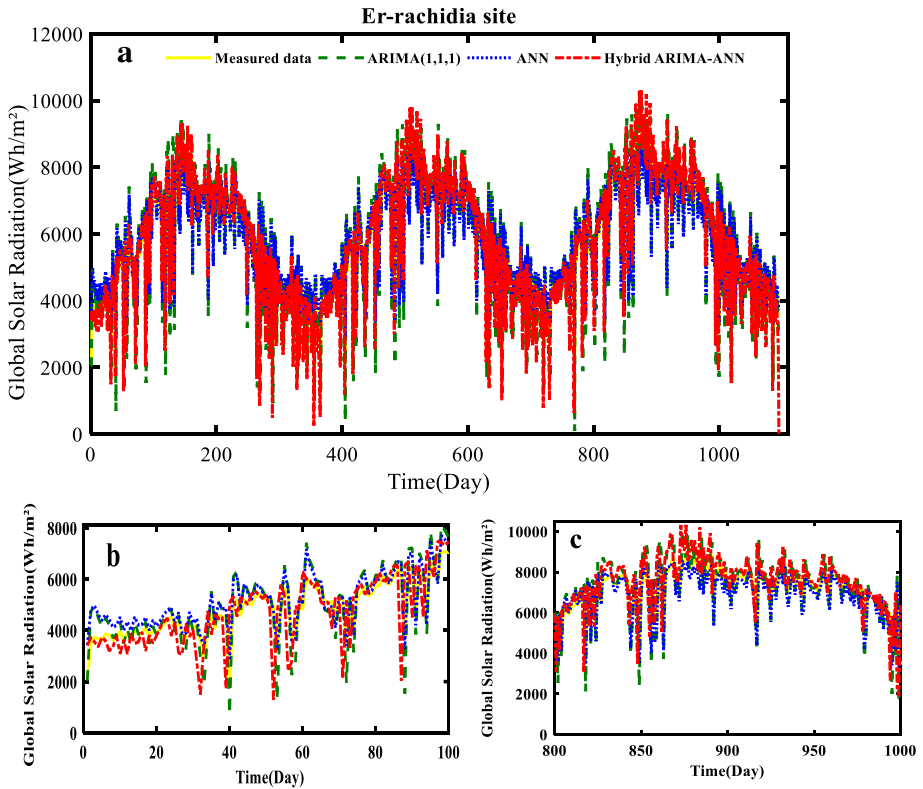


Fig. 13 Compared experimental and forecasted DGSR values for ARIMA (1,1,1), ANN and hybrid ARIMA-ANN for Er-rachidia site (a). b, c Are enlargement of a

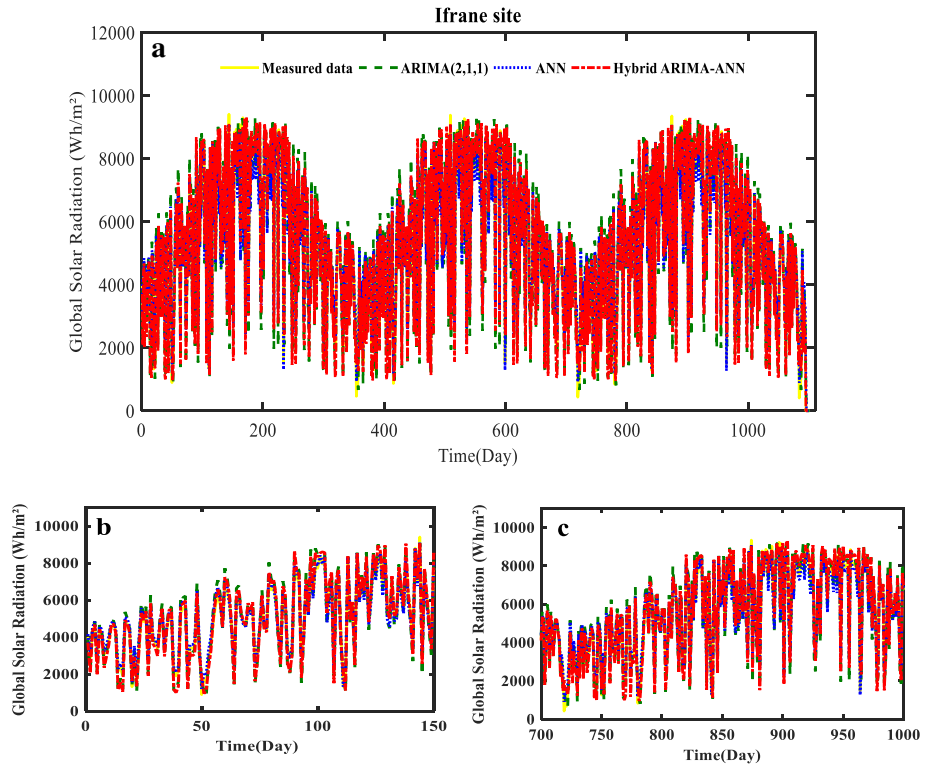


Fig. 14 Compared experimental and forecasted DGSR values for ARIMA (2,1,1), ANN and hybrid ARIMA–ANN for Ifrane site (**b**), **c** Are enlargement of **a**

4 Conclusion

In this paper, we have proposed an innovative hybrid model to forecast the daily global solar radiation in three different regions located in Morocco. The experimental used data are taken from three different stations for full years 2013, 2014 and 2015. The established approach of the proposed study has provided the weight to capture different patterns of ARIMA and AI models. Before applying the modelling approaches, the DGSR has been transferred to K_t to make data non-stationary. According to the non-stationary data, the optimum ARIMA and ANN models were processed. In time series data, the significant ACF, PACF and AIC criteria allowed to select the ARIMA (2. 1. 1), ARIMA (1.1.1) as adequate models of three sites.

The accuracy and performance of the proposed ARIMA–ANN model have been evaluated and checked, using various statistical measurement errors. Results obtained by hybrid ARIMA–ANN show a suitable matching between the observed and forecasted values, suggesting the ARIMA–ANN suitability to reproduce experimental data with satisfactory precision.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: All data included in this manuscript are available upon request by contacting the corresponding author.]

Methodological Appendix

Where G_{forecast} and G_{exp} are the forecasted and measured values of DGSR, respectively. $\overline{G}_{\text{forecast}}$ and $\overline{G}_{\text{exp}}$ present the average of forecasted and measured values of DGSR and N is number of observation data (Table 6).

Table 6 Statistical criteria equations

Equation	Description
$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left \frac{G_{\text{forecast},i} - G_{\text{exp},i}}{G_{\text{exp},i}} \right $	Evaluate the efficiency and performance of the forecasting model
$\text{MBE} = \frac{1}{N} \sum_{i=1}^N (G_{\text{forecast},i} - G_{\text{exp},i})$	Measures the median bias in the forecasting model
$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (G_{\text{forecast},i} - G_{\text{exp},i})^2}$	Presents how the errors are measured and how far apart the correlation graph coefficients are
$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (G_{\text{forecast},i} - G_{\text{exp},i})^2}}{\frac{1}{N} \sum_{i=1}^N G_{\text{exp},i}}$	Allows to evaluate datasets or models at varying sizes.
$Sd = \sqrt{\frac{N(\text{RMSE}^2 - \text{MBE}^2)}{N-1}}$	Is a function of the sum of variance or attenuation of the value range
$Ts = \sqrt{\frac{(N-1)\text{MBE}^2}{\text{RMSE}^2 - \text{MBE}^2}}$	Used to assess if there is a qualitative difference between the results of two classes that could be linked to some characteristics
$R^2 = 1 - \frac{\sum_{i=1}^N (G_{\text{forecast},i} - G_{\text{exp},i})^2}{\sum_{i=1}^N (G_{\text{forecast},i} - \overline{G}_{\text{exp}})^2}$	Used to measure the correct unidirectional correlation between both the experimental and the forecasted values
$Y = aG + b$	Linear regression coefficients
$a = \frac{G_{\text{forecast}}G_{\text{exp}} - \overline{G}_{\text{exp}}G_{\text{forecast}}}{G_{\text{exp}}^2 - \overline{G}_{\text{exp}}^2}$	
$b = \overline{G}_{\text{forecast}} - a\overline{G}_{\text{exp}}$	

References

1. S.A.R. Khan, K. Zaman, Y. Zhang, The relationship between energy-resource depletion, climate change, health resources and the environmental Kuznets curve: evidence from the panel of selected developed countries. *Renew. Sustain. Energy Rev.* **62**, 468–477 (2016)
2. “RENEWABLE ENERGY” Ministère de l’Industrie, du Commerce et de l’Économie Verte et Numérique”
3. Y. Ge, Y. Nan, L. Bai, A hybrid prediction model for solar radiation based on long short-term memory, empirical mode decomposition, and solar profiles for energy harvesting wireless sensor networks. *Energies* **12**, 24 (2019)
4. V. Badescu et al., Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renew. Energy* **55**, 85–103 (2013)
5. A.A. Babatunde, S. Abbasoglu, Predictive analysis of photovoltaic plants specific yield with the implementation of multiple linear regression tool. *Environ. Prog. Sustain. Energy* **38**(4), 13098 (2019)

6. A.R. Pazikadin, D. Rifai, K. Ali, M.Z. Malik, A.N. Abdalla, M.A. Faraj, Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): a review of five years research trend. *Sci. Total Environ.* **715**, 136848 (2020)
7. M. Guermoui, F. Melgani, K. Gairaa, M.L. Mekhalfi, A comprehensive review of hybrid models for solar radiation forecasting. *J. Clean. Prod.* **258**, 120357 (2020)
8. R.A. de Marcos, A. Bello, J. Reneses, Electricity price forecasting in the short term hybridising fundamental and econometric modelling. *Electr. Power Syst. Res.* **167**, 240–251 (2019)
9. N. Amral, C. S. Ozveren, and D. King, Short term load forecasting using Multiple Linear Regression,” in *2007 42nd International Universities Power Engineering Conference*, 2007, pp. 1192–1198
10. M.Q. Raza, M. Nadarajah, C. Ekanayake, On recent advances in PV output power forecast. *Sol. Energy* **136**, 125–144 (2016)
11. B. Belmahdi, A. El Bouardi, Simulation and optimization of microgrid distributed generation: a case study of University Abdelmalek Essaâdi in Morocco. *Procedia Manuf.* **46**, 746–753 (2020)
12. Y. Ren, P.N. Suganthan, N. Srikanth, Ensemble methods for wind and solar power forecasting—a state-of-the-art review. *Renew. Sustain. Energy Rev.* **50**, 82–91 (2015)
13. U.K. Das et al., Forecasting of photovoltaic power generation and model optimization: a review. *Renew. Sustain. Energy Rev.* **81**, 912–928 (2018)
14. M.K. Behera, I. Majumder, N. Nayak, Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Eng. Sci. Technol. Int. J.* **21**(3), 428–438 (2018)
15. J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, H. Bai, Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. *Renew. Sustain. Energy Rev.* **105**, 168–186 (2019)
16. L. Wu, G. Huang, J. Fan, F. Zhang, X. Wang, W. Zeng, Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Convers. Manag.* **183**, 280–295 (2019)
17. J. Almorox, C. Hontoria, Global solar radiation estimation using sunshine duration in Spain. *Energy Convers. Manage.* **45**(9), 1529–1535 (2004)
18. H. Khorasanizadeh, K. Mohammadi, M. Jalilvand, A statistical comparative study to demonstrate the merit of day of the year-based models for estimation of horizontal global solar radiation. *Energy Convers. Manage.* **87**, 37–47 (2014)
19. G.H. Hargreaves, Z.A. Samani, Estimating potential evapotranspiration. *J. Irrigat. Drainage Div.* **108**(3), 225–230 (1982)
20. K.L. Bristow, G.S. Campbell, On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agric. For. Meteorol.* **31**(2), 159–166 (1984)
21. G.E. Hassan, M.E. Youssef, Z.E. Mohamed, M.A. Ali, A.A. Hanafy, New temperature-based models for predicting global solar radiation. *Appl. Energy* **179**, 437–450 (2016)
22. C. Voyant et al., Machine learning methods for solar radiation forecasting: a review. *Renew. Energy* **105**, 569–582 (2017)
23. G. Notton, C. Voyant, A. Foulloy, L.J. Duchaud, L.M. Nivet, Some applications of ANN to solar radiation estimation and forecasting for energy applications. *Appl. Sci.* **9**, 1 (2019)
24. M. Paulescu, E. Paulescu, P. Gravila, V. Badescu, “Weather Modeling and Forecasting of PV Systems Operation,” *Green Energy and Technology*, (2013)
25. C. Voyant, G. Notton, Solar irradiation nowcasting by stochastic persistence: a new parsimonious, simple and efficient forecasting tool. *Renew. Sustain. Energy Rev.* **98**, 343–352 (2018)
26. K. Benmouiza, A. Cheknane, Small-scale solar radiation forecasting using ARMA and nonlinear autoregressive neural network models. *Theoret. Appl. Climatol.* **124**(3–4), 945–958 (2016)
27. V. Kushwaha, N.M. Pindoriya, A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew. Energy* **140**, 124–139 (2019)
28. P. Bacher, H. Madsen, H.A. Nielsen, Online short-term solar power forecasting. *Sol. Energy* **83**(10), 1772–1783 (2009)
29. Y. Li, Y. Su, L. Shu, An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* **66**, 78–89 (2014)
30. M. Geurts, G. E. P. Box, G. M. Jenkins, Time series analysis: forecasting and control. *J. Marketing Res.* (1977)
31. M. Khashei, M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* **11**(2), 2664–2675 (2011)
32. P.G. Zhang, Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**, 159–175 (2003)
33. R. Babazadeh, A Hybrid ARIMA–ANN approach for optimum estimation and forecasting of gasoline consumption. *RAIRO-Oper. Res.* **51**(3), 719–728 (2017)

34. D. Ömer Faruk, A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intell.* **23**(4), 586–594 (2010)
35. R. Adhikari, R.K. Agrawal, A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Comput. Appl.* **24**(6), 1441–1449 (2014)
36. M. Kumar, M. Thenmozhi, Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models. *Int. J. Bank. Account. Finance* **5**(3), 284–308 (2014)
37. A. Rabehi, M. Guermoui, and D. Lalmi, “Hybrid models for global solar radiation prediction: a case study,” *International Journal of Ambient Energy*, Taylor and Francis Ltd., pp. 1–10, 02-Jan-2018
38. “Invest in Morocco - Solar Energy.” [Online]. Available: <http://www.invest.gov.ma/?Id=24&lang=en&RefCat=2&Ref=145>. Accessed 24 Dec, 2019
39. CMP11 secondary standard pyranometer - Kipp & Zonen
40. B. Belmahdi, M. Louzazni, A. El Bouardi, One month-ahead forecasting of mean daily global solar radiation using time series models. *Optik* **219**, 165207 (2020)
41. R. Jamil, Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. *Renew. Energy* **154**, 1–10 (2020)
42. G. Box, G. Jenkins, G. Reinsel, G. Ljung, *Fifth Edition Time Series Analysis Forecasting and Control* (Wiley, Hoboken, 2016)
43. R.S. Tsay, *Analysis of Financial Time Series* (Wiley, Hoboken, 2010)
44. N. Daldal, M. Nour, K. Polat, A novel demodulation structure for quadrature modulation signals using the segmentary neural network modelling. *Appl. Acoust.* **164**, 107251 (2020)
45. R.R. Naik, N.S. Gandhi, M. Thakur, V. Nanda, Analysis of crystallization phenomenon in Indian honey using molecular dynamics simulations and artificial neural network. *Food Chem.* **300**, 125182 (2019)
46. S. Al-Dahidi, O. Ayadi, J. Adeeb, M. Louzazni, Assessment of artificial neural networks learning algorithms and training datasets for solar photovoltaic power production prediction. *Front. Energy Res.* **7**, 130 (2019)
47. M. Louzazni, H. Mosalam, A. Khouya, A non-linear auto-regressive exogenous method to forecast the photovoltaic power output. *Sustain. Energy Technol. Assessments* **38**, 100670 (2020)
48. Ü.Ç. Büyüksahin, Ş. Ertekin, Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **361**, 151–163 (2019)
49. C. F. M. Coimbra, J. Kleissl, R. Marquez, Overview of Solar-Forecasting Methods and a Metric for Accuracy Evaluation. in *Solar Energy Forecasting and Resource Assessment*, Elsevier Inc., 2013, pp. 171–194
50. M. Louzazni, A. Khouya, K. Amechnoue, M. Mussetta, A. Crăciunescu, Comparison and evaluation of statistical criteria in the solar cell and photovoltaic module parameters’ extraction. *Int. J. Ambient Energy* **41**(13), 1482–1494 (2018)
51. A. Mellit, A.M. Pavan, A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Sol. Energy* **84**(5), 807–821 (2010)
52. J. Mubiru, E.J.K.B. Banda, Estimation of monthly average daily global solar irradiation using artificial neural networks. *Sol. Energy* **82**(2), 181–187 (2008)
53. J. Mubiru, Predicting total solar irradiation values using artificial neural networks. *Renew. Energy* **33**(10), 2329–2332 (2008)
54. J.A.D.W.A. Beckman, *Solar Engineering of Thermal Processes*, 4th edn. (Wiley, Hoboken, 2013)
55. B. Amrouche, X. Le Pivert, Artificial neural network based daily local forecasting for global solar radiation. *Appl. Energy* **130**, 333–341 (2014)
56. C.G. Ozoegwu, Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. *J. Clean. Prod.* **216**, 1–13 (2019)
57. C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, Hybrid methodology for hourly global radiation forecasting in Mediterranean area. *Renew. Energy* **53**, 1–11 (2013)
58. M.A. Mohandes, Modeling global solar radiation using particle swarm optimization (PSO). *Sol. Energy* **86**(11), 3137–3145 (2012)
59. C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy* **39**(1), 341–355 (2012)
60. Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband, B. Khoshnevisan, Potential of radial basis function based support vector regression for global solar radiation prediction. *Renew. Sustain. Energy Rev.* **39**, 1005–1011 (2014)
61. K. Mohammadi, S. Shamshirband, C.W. Tong, M. Arif, D. Petković, S. Ch, A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers. Manag.* **92**, 162–171 (2015)
62. S. Amirkhani, S. Nasirivatan, A.B. Kasaeian, A. Hajinezhad, ANN and ANFIS models to predict the performance of solar chimney power plants. *Renew. Energy* **83**, 597–607 (2015)

63. A. Gani et al., Day of the year-based prediction of horizontal global solar radiation by a neural network auto-regressive model. *Theoret. Appl. Climatol.* **125**(3–4), 679–689 (2016)
64. K. Gairaa, A. Khellaf, Y. Messlem, F. Chellali, Estimation of the daily global solar radiation based on Box-Jenkins and ANN models: a combined approach. *Renew. Sustain. Energy Rev.* **57**, 238–249 (2016)
65. A. Mellit, M. Benghanem, A.H. Arab, A. Guessoum, A simplified model for generating sequences of global solar radiation data for isolated sites: using artificial neural network and a library of Markov transition matrices approach. *Sol. Energy* **79**(5), 469–482 (2005)
66. E.S. Mostafavi, S.S. Ramiyani, R. Sarvar, H.I. Moud, S.M. Mousavi, A hybrid computational approach to estimate solar global radiation: an empirical evidence from Iran. *Energy* **49**(1), 204–210 (2013)
67. Y. Che, L. Chen, J. Zheng, L. Yuan, F. Xiao, A novel hybrid model of WRF and clearness index-based Kalman filter for day-ahead solar radiation forecasting. *Appl. Sci.* **9**(19), 3967 (2019)
68. S. Hussain, A. AlAlili, A hybrid solar radiation modeling approach using wavelet multiresolution analysis and artificial neural networks. *Appl. Energy* **208**, 540–550 (2017)
69. A. Rabehi, M. Guermoui, D. Lalmi, Hybrid models for global solar radiation prediction: a case study. *Int. J. Ambient Energy* **41**(1), 31–40 (2020)