ISSN 0003-701X, Applied Solar Energy, 2022, Vol. 58, No. 1, pp. 137–151. © Allerton Press, Inc., 2022.
Russian Text © The Author(s), 2022, published in Geliotekhnika, 2022, No. 1, pp. 175–191.



# Forecasting of Solar Irradiances using Time Series and Machine Learning Models: A Case Study from India

**Sarita Sheoran[a], Raj Shree Singh[a], Sumanta Pasari[a, *, **], and Rakhee Kulshrestha[a]**

[a]*Department of Mathematics, Birla Institute of Technology & Science Pilani, Pilani Campus, Rajasthan, Jhunjhunu, 333031 India*

**e-mail: 86.sumanta@gmail.com*

***e-mail: sumanta.pasari@pilani.bits-pilani.ac.in*

Received August 13, 2021; revised February 28, 2022; accepted April 28, 2022

**Abstract**—With the focus on renewable energy resources due to environmental reasons, reliable forecasting of renewable energy has great societal importance. This study focuses on the analysis and forecasting of GHI data at two different locations in India, namely Pokhran and Bitta. Since the GHI time series plots exhibit seasonality and randomness, the time series SARIMA model along with two machine learning models, namely MLP and LSTM, are implemented for daily, weekly and monthly forecasting. The efficacy of these competitive models is assessed using MAPE and RMSE values. We also perform residual analysis as a post processing step of the implemented models. For monthly forecasting, the SARIMA model has the best performance, as it precisely captures monthly seasonality in comparison to the machine learning models. However, for short term daily forecasting, machine learning models provide much better estimates with MLP as the most desirable one. Since the SARIMA model fails to fully capture the high amount of fluctuation (mostly, seasonal fluctuation) in the daily and weekly observations, we additionally implement an ARIMA model with sliding windows to improve modelling efficacy. The present study therefore provides a clear guideline on the selection of forecasting models based on the desired time horizon.

**Keywords:** renewable energy, global horizontal irradiance, forecasting, time series, machine learning

**DOI:** 10.3103/S0003701X22010170

## INTRODUCTION

Energy sources that can naturally replenish on human timescale are classified as renewable sources of energy. Renewable sources like sunlight, wind, rain, tides and geothermal heat are virtually inexhaustible in duration but limited in amount that is available per unit of time. India is the second most populated country in the world and holds the seventh position for the largest country. It needs an enormous amount of energy to cater the economic development plans of its people. According to a report by the Ministry of Power, Government of India [1], in the year 2021, India has produced 60.9% of its total energy needs via fossil fuels, 12.1% by hydro power sources, 1.8% by nuclear power sources and the rest 25.2% by renewable energy sources. In a recent article by Lu et al. [2], the authors have claimed that the combined wind and solar energy can meet about 80% of India's electricity demands by 2040. The article also states that achieving that level of reliance on renewable energy would reduce $CO_2$ emissions by 85% and the overall power generation costs by $50 billion. Recently, the Ministry of New and Renewable Energy (MNRE), India has floated several open calls for proposals to garner ideas about renewable energy forecasting, prominent site identification, and operational management. The present study on solar energy forecasting at different time scales contributes to this endeavour.

## MOTIVATION

When traditional sources like fossil fuels are used for power generation, the output is predominantly controlled by the plant's generation capacity. However, when renewable sources like solar and wind energy are used for power generation, the generation also depends on weather conditions apart from the machines' capacity. The fundamental basis of managing existing and newly constructed power systems is power generation forecasting, failing to do so may lead to inappropriate operational practices and inadequate energy transactions [3]. Higher penetration of solar energy significantly increases the uncertainties of solar power systems, leading to complications in system operations and planning [4]. Thus, accurately forecasting the amount of solar irradiation provides a valu-

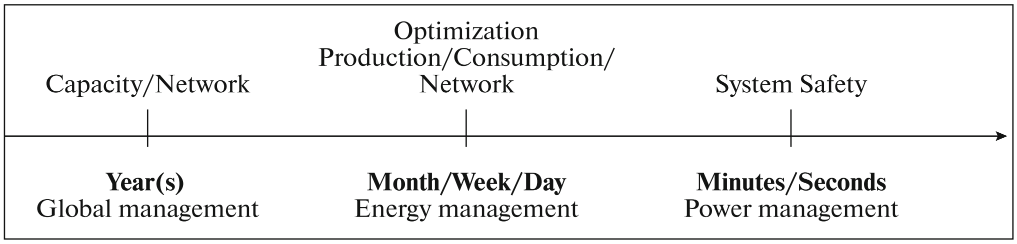


**Fig. 1.** Prediction scale for energy management in the electrical network [3].

able tool to ease the energy integration and it enables independent system operators to function more efficiently and run the power systems reliably. As the power generation from solar energy has heavy dependence on weather, seasonal changes, geographical location and time, the forecasting methods may not provide uniformly efficient results across all regions [5, 6].

The time frame of a forecasting model can vary from several minutes to several years. Based on the time horizon for power system operations, the solar energy forecasting models can be classified into three broad classes: short term, medium term and long term. Short term forecasts include forecasting from several hours to a couple of days; medium term forecasts deal with forecasting from several weeks to months, whereas long term forecasting provides forecasts from several months to years [7]. Therefore, temporal time horizons play crucial role for energy operations (fig. 1) in integration of intermittent solar and wind energy into current and future energy structure. Similarly, for energy suppliers, it helps to manage energy system ensuring the precise balance among electricity production, consumption and distributions [3].

## LITERATURE SURVEY

There have been several studies on the forecasting of solar irradiance using time series methods and machine learning techniques due to their intrinsic flexibility and robustness in data analysis. Here we mention a few time series models for solar energy forecasting, followed by a few machines learning models.

In 2010, Martin et al. [6] implemented different statistical models to generate solar energy predictions at four different locations in Spain. They compared the models' performance for different time series inputs namely, clearness index, lost component and GHI. The best results were obtained using the neural network models to forecast half daily values of solar irradiance with lost component as input. Similarly, Reikard [8–10] used ANN and window sliding based ARIMA models, and compared their performances. It was observed that when forecasting over short horizons, time series models (ARIMA) work best when estimated over moving windows of observations. This enables them to pick up the effect of seasonality on recent weather conditions. Yang et al. [11] proposed three forecasting methods to predict next hour solar irradiance values and associated cloud cover effects. The first method takes in GHI values and directly uses it to forecast at one-hour intervals through additive seasonal decomposition, followed by an ARIMA model. The second method forecasts DHI and DNI separately using additive seasonal decomposition, followed by an ARIMA model. The results of the two forecasting methods are then combined to predict GHI values using an atmospheric model. The third method considers cloud cover effects and uses an ARIMA model to predict cloud transients. Their results showed that the third method outperformed the other two. The method showed that the use of cloud cover techniques improves performance of the forecasting models. In 2014, Sahin et al. [12] used Australian monthly electricity consumption dataset from January 1956 to August 1995 and implemented ARIMA and SARIMA models. They also estimated uncertainty in the forecast through triangular fuzzy numbers. The interval of forecasting was generated from a lower and upper bound estimator. They found that the proposed approach when combined with SARIMA (2,1,2) (15,1,2)12 provides most satisfactory results. In 2019, Alsharif et al. [13] implemented ARIMA, SARIMA and Monte-Carlo prediction models for hourly solar radiation data of Seoul, South Korea over 37 years (1981–2017). The results demonstrate that the ARIMA (1,1,2) model can be used to represent daily solar irradiation, while the seasonal ARIMA (4,1,1) of 12 lags can be used to represent monthly solar irradiation. In 2019, Sharadga et al. [14] implemented ARMA, ARIMA, SARIMA and six neural network models for three different types of days, namely sunny, cloudy and rainy days. The results indicate that time series models of ARMA (3,4), ARIMA (2,1,3) and SARIMA (2,1,3) (2,0,1)14 provide the best fit representation.

Apart from time series models, many machines learning models such as SVR [15], SVM [15], ANN

[16], k-NN [16], MLP [5], and LSTM [5, 17] have provided satisfactory results for the solar energy forecasting at different time horizons, especially in short term forecasting. In this regard, the efficient utility of MLP and LSTM models has been demonstrated in various studies. A study on solar potential of Himachal Pradesh (India) was performed by Yadav and Chandal [18]. They used ANN based global solar radiation model to assess the solar potential. Gensler et al. [5] compared performances of MLP, LSTM, DBN and auto-LSTM machine learning models on a German solar power dataset. The best performing model is the auto-LSTM, closely followed by the DBN model.

Motivated by the above studies, the present study focuses on GHI modelling at two selected locations in India, namely Pokhran (Rajasthan) and Bitta (Gujarat). Machine learning and time series forecasting models are employed to provide a comprehensive idea of solar energy forecasting at different time horizons at these locations. Four models, namely SARIMA, ARIMA, MLP and LSTM are implemented for monthly, weekly and daily forecasting. The performance of these models is assessed using RMSE and MAPE values.

## STUDY AREA AND DATASET

The dataset for this study is obtained from National Solar Radiation Database (NSRDB; https://nsrdb.nrel.gov/) maintained by the US Department of Energy. The two locations of this study correspond to Dhirubhai Ambani Solar Park near Pokhran (Rajasthan) and Adani Power Limited near Bitta (Gujarat). For each location, the dataset is available for the period of January 1, 2000 to December 31, 2014. The features in the dataset include DHI and clearsky DHI, DNI and clearsky DNI, GHI and clearsky GHI, along with other environmental variables, such as temperature, pressure, relative humidity and solar zenith angle. The GHI dataset used in current study is measured in W/m$^2$. The primary focus of this study is to implement above mentioned four models for monthly, weekly and daily forecasting of GHI values based on the past data.

## METHODOLOGY

The Description of the Studied four Models is Provided Below

SARIMA $(p, d, q)(P, D, Q)s$ model: The expression for the SARIMA model [19] is given as:

$$\phi(B)(1-B)^d\Phi(B^s)(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t,$$

where, $\{X_t\}$ is a time series of forecast variable and $Z_t$ denotes the random noise; B is the backshift operator. The parameters for these types of models are as follows:

- $p$ and seasonal $P$ indicate number of autoregressive terms;
- $d$ and seasonal $D$ indicate order of differencing that must be performed to stationarize the time series;
- $q$ and seasonal $Q$ represent number of terms in moving average;
- $s$ indicates seasonal length.

The orders of seasonality for the SARIMA models are 12 for monthly forecasting, 52 for weekly forecasting and $73 \times 5$ for daily forecasting. Note that we have prepared five different sets each with 73 days of data so as to reduce the high amount of fluctuation (mostly, seasonal fluctuation) in the daily observations. After optimization, the predictions of these five individual models are combined to reduce the computational resources.

ARIMA-window sliding: For the daily forecasting, the dataset was split into training and testing data and the best ARIMA model (with minimum AIC) for the training dataset was fitted on to the sliding window model [8–10] to forecast the results for testing dataset. The dataset was observed to show a high positive autocorrelation peak at an interval of one year which declined during the following years. Therefore, a sliding window of 365 days was considered to make predictions for the next day and thus the results were accumulated for 3 years of test dataset. Similarly, a sliding window of 52 weeks was considered to make predictions for the weekly dataset.

MLP model: The MLP [5] is a particular feed forward neural network comprising three layers, namely the input layer, hidden layer and the output layer. The input layer is where we insert our data. Each node in the input layer can take one data value. After inserting the data, each node performs a non-linear transformation on the input values and sends it to the hidden layer. Each hidden layer node performs another nonlinear transformation to the hidden layer values and sends it to the output layer nodes. The output layer values are then compared with the expected values. The errors are calculated and consequently used to further tune the neural network weights to improve the overall accuracy through a back propagation algorithm. The MLPs are considered powerful tools for time series forecasting as they can simultaneously approximate arbitrary non-linear functions, handle noise, accept multivariate inputs and perform multistep forecasting. There are several different hyperparameters which require tuning in an MLP model. These include the number of hidden layer nodes, the activation functions, training time and optimization algorithms. The key parameters in the MLP and

**Table 1.** Model description with associated configuration

| Model | Activation function | Number of layers | Optimizing algorithm | Loss function | Epochs | Type of hidden layer |
|---|---|---|---|---|---|---|
| MLP | Rectified Linear Activation Function (ReLU) | 3 | Adaptive Moment Estimation (Adam) | Mean Squared Error Function (MSE) | 2000 | Dense |
| LSTM | Rectified Linear Activation Function (ReLU) | 3 | Adaptive Moment Estimation (Adam) | Mean Squared Error Function (MSE) | 2000 | LSTM |

LSTM models are summarized in Table 1, though we have tried many more combinations of the model parameters. The best fit model is chosen based on the residual analysis.

LSTM model: The LSTM [17] is a special kind of RNN which is used widely in sequence models. It has application in all kinds of sequence models such as handwriting recognition, speech recognition and text classification [3]. The LSTMs have an internal memory in the form of gates. This allows the LSTM to accumulate important information which may be required in future. It has all the features of RNN along with an internal memory to help better prediction. The number of neurons in each layer is governed by the dimensionality of the input matrix for daily, weekly and monthly datasets.

The accuracy of the forecasting models is evaluated and compared on the basis of the RMSE and MAPE error metrics. Lower the error values, better is the forecast. The formulae for RMSE and MAPE are given as:

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{t=N} \frac{(X_t - \widehat{X}_t)^2}{N}},$$

$$\mathrm{MAPE} = \sum_{t=1}^{t=N} \left| \frac{(X_t - \widehat{X}_t)}{X_t} \right| \frac{1}{N},$$

where, $X_t$ is the forecast variable; $\widehat{X}_t$ is the corresponding forecast value and N denotes the length of time series of the forecast variable.

The proposed methodology is carried out in a five-fold manner, as follows:

(i) Test for stationarity: We perform the ADF (Augmented Dickey Fuller) test to check for stationarity in the time series. The hypotheses are as follows:

H0: The series has a unit root.

H1: The series has no unit root.

If the null hypothesis (H0) is rejected, it implies that the time series does not have a unit root, indicating stationary data.

For the present data, we obtain the test statistic value 4.047, which is less than the critical values −3.471, −2.879, −2.576 at 1, 5 and 10% levels of significance, respectively. Therefore, we reject the null hypothesis and conclude that the dataset is stationary. Rejection of the null hypothesis enables us to determine the value of differencing parameter 'd' in ARIMA to be zero. Otherwise, we would have to perform some methods such as differencing, detrending and transformations to make the data stationary.

(ii) Computation of descriptive statistics: The dataset is initially split into two parts, namely testing and training data. For the present study, we consider training dataset from January 1, 2000 to December 31, 2011 and the testing dataset from January 1, 2012 to December 31, 2014 for the machine learning models. The dataset from January 1, 2006 to December 31, 2011 is considered as the training dataset for time series models. The training dataset is then utilized to compute various descriptive measures, such as sample mean, sample median, quartiles, sample standard deviation and coefficient of variation.

(iii) Pre-processing of data: Before the training dataset can be used for modelling, it needs to be split and sampled accordingly. The training dataset is hourly sampled. Later, it is sampled based on the mean value of the time horizon. For the monthly and weekly forecasts in SARIMA, the dataset is sampled monthly and weekly, respectively. To implement the daily SARIMA model, the training dataset is divided into five parts using the modulo function with each part having seasonality of 73, as otherwise the daily

**Table 2.** Descriptive statistics of the two datasets

| Solar Site | Latitude | Longitude | Mean | Standard deviation | Coefficient of variance | 50th percentile value |
|---|---|---|---|---|---|---|
| Bitta | 23.26° N | 69.02° E | 269.41 | 61.75 | 0.23 | 271.65 |
| Pokhran | 26.92° N | 71.91° E | 557.27 | 110.04 | 0.20 | 575.90 |

**Table 3.** Forecasting at different scales for Rajasthan

| Monthly forecasting | | | |
|---|---|---|---|
| Location | Model | RMSE | MAPE |
| Rajasthan | SARIMA | 15.437 | 0.02080 |
| | MLP | 24.873 | 0.03065 |
| | LSTM | 62.962 | 0.05608 |
| Weekly forecasting | | | |
| Rajasthan | SARIMA | 45.934 | 0.06336 |
| | ARIMA | 28.127 | 0.03110 |
| | MLP | 29.865 | 0.03092 |
| | LSTM | 27.775 | 0.02962 |
| Daily forecasting | | | |
| Rajasthan | SARIMA | 81.677 | 0.09372 |
| | ARIMA | 34.804 | 0.04158 |
| | MLP | 0.135 | 0.01462 |
| | LSTM | 0.098 | 0.01591 |

data exhibits high order of seasonality. For the MLP and LSTM models, the training dataset is converted to an input matrix and an output vector. For monthly model, the input matrix contains GHI values for all days in that month. Thus, the input dimension is 28, 30, or 31. The input dimension is 7 for weekly model, whereas it is 9 for daily model (the dataset contains GHI values recorded for 9 hours in a day).

(iv) Forecasting models: The training dataset of each location is used in time series and machine learning models. First, the models are evaluated on the testing dataset and depending on the performance, the models are optimised for better results. The ‘pmdarima’ library of Python is used to construct SARIMA and ARIMA models, whereas the ‘keras library’ is used to build the LSTM and MLP models. The LSTM [17] and MLP [5] models are trained using the input matrix and output vectors, and their performance is compared with the time series models using RMSE and MAPE values. The configuration for each of these models has been specified above in Table 1.

(v) Residual Analysis: As a post processing step, we perform the residual analysis to check whether there is any systematic bias in the implemented models. As the residuals of a forecast model should exhibit normal distribution with zero mean and a constant variance, we analyse residual plots, histogram plots and P-P plots (Section 14.8 and Section 14.9 [20]) of the standardised residuals corresponding to the best fit models for the selected three time horizons.

## RESULTS

Various measures of descriptive statistics of the daily dataset from January 1, 2000 to December 31, 2011 for the studied locations are summarized in Table 2. It is observed that Pokhran has the largest mean GHI value followed by Bitta, whereas Bitta has the largest coefficient of variation followed by Pokhran.

### *Monthly Forecasting*

The monthly forecasts from MLP, LSTM and SARIMA models of the two study sites are pictorially shown in Figs. 2 and 3. The MAPE and RMSE values of these models are listed in Tables 3 and 4. For

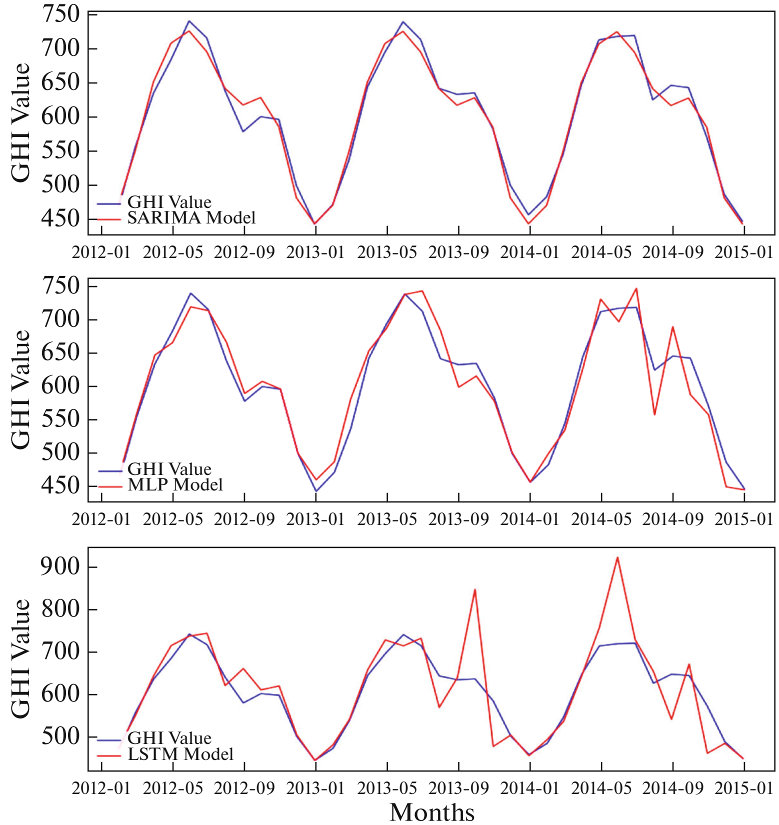


**Fig. 2.** Performance of the forecasting models for monthly dataset from Rajasthan.

**Table 4.** Forecasting at different scales for Gujarat

| Location | Model | RMSE | MAPE |
|---|---|---|---|
| **Monthly forecasting** | | | |
| Gujarat | SARIMA | 22.765 | 0.03036 |
| | MLP | 27.079 | 0.03203 |
| | LSTM | 34.789 | 0.04476 |
| **Weekly forecasting** | | | |
| Gujarat | SARIMA | 48.321 | 0.06259 |
| | ARIMA | 30.848 | 0.04164 |
| | MLP | 28.759 | 0.03443 |
| | LSTM | 27.616 | 0.03298 |
| **Daily forecasting** | | | |
| Gujarat | SARIMA | 84.718 | 0.10693 |
| | ARIMA | 33.671 | 0.04575 |
| | MLP | 0.151 | 0.01748 |
| | LSTM | 0.085 | 0.01436 |

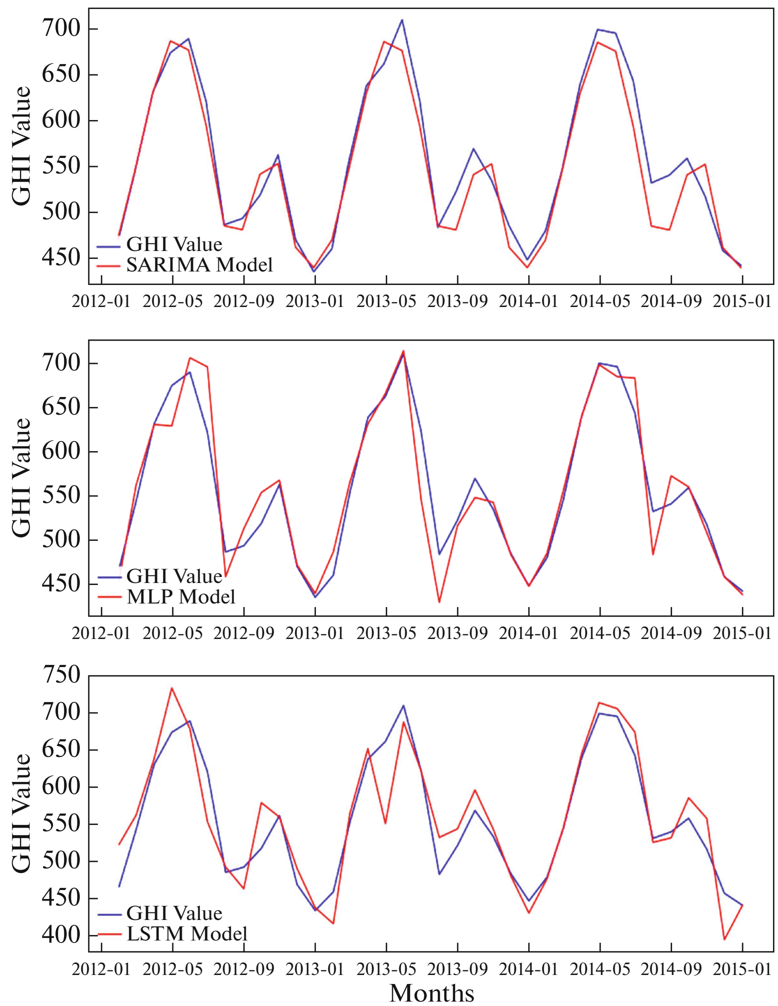


**Fig. 3.** Performance of the forecasting models for monthly dataset from Gujarat.

monthly forecasting, the SARIMA model outperforms the other two models due to its ability to precisely capture the monthly seasonality in the dataset.

*Weekly Forecasting*

The weekly forecasts from MLP, LSTM and SARIMA models of the two study sites are shown in Figs. 4 and 5. We observe that the SARIMA model fails to capture the high amount of fluctuation (mostly, seasonal fluctuation) in the weekly dataset. We additionally implement the ARIMA model along with sliding windows of length 52 (corresponding to number of weeks in a year), which improves the efficacy of time series models. The MAPE and RMSE values of these models are mentioned in Tables 3 and 4. For weekly forecasting, the LSTM model reflects the least error values followed by the ARIMA and MLP model. However, the ARIMA model based on sliding windows is preferable for the weekly forecasting since it requires much lesser training than the LSTM to provide similar accuracy. For the weekly dataset, the ARIMA model and neural networks both are capable of learning different inherent temporal dependencies that are difficult to model by SARIMA model.

*Daily Forecasting*

The daily forecasts from MLP, LSTM, SARIMA and sliding window based ARIMA models of the two study sites are shown in Figs. 6 and 7. The MAPE and RMSE values of these models are tabulated in Tables 3 and 4. It is observed that the machine learning models exceptionally outperform time series models, similar to the observation made in [21]. There could be two possible explanations for this: (1) the large amount of daily dataset provides sufficient training information for the neural networks to represent the underlying process; (2) At daily resolutions, as there is often seasonality of high order, nonlinear variability and intermittency due to cloud coverage and precipitation in

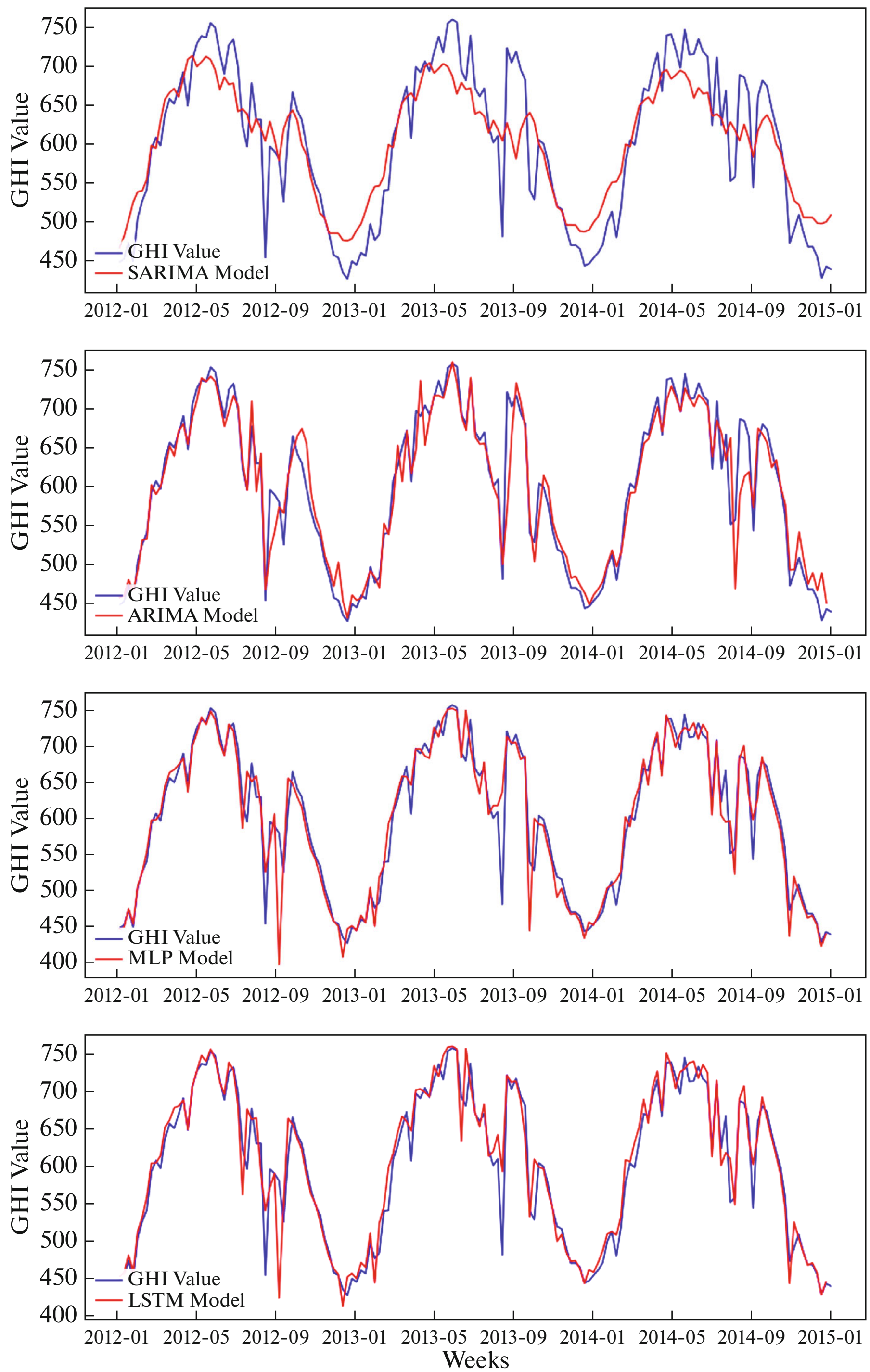


**Fig. 4.** Performance of the forecasting models for weekly dataset from Rajasthan.

the environment, the time series models fail to capture all of these features collectively.

### *Analysis of Residuals*

As mentioned earlier, here we perform residual analysis mainly to check whether there is any systematic bias in the implemented models. A forecast bias is a tendency for a model to consistently produce higher or lower forecast values than their actual values. Therefore, analysis of bias is an important post-processing step. The main characteristic of the unbiased models is the normality of the residuals. Therefore, if the residuals are not approximately of Gaussian shape,

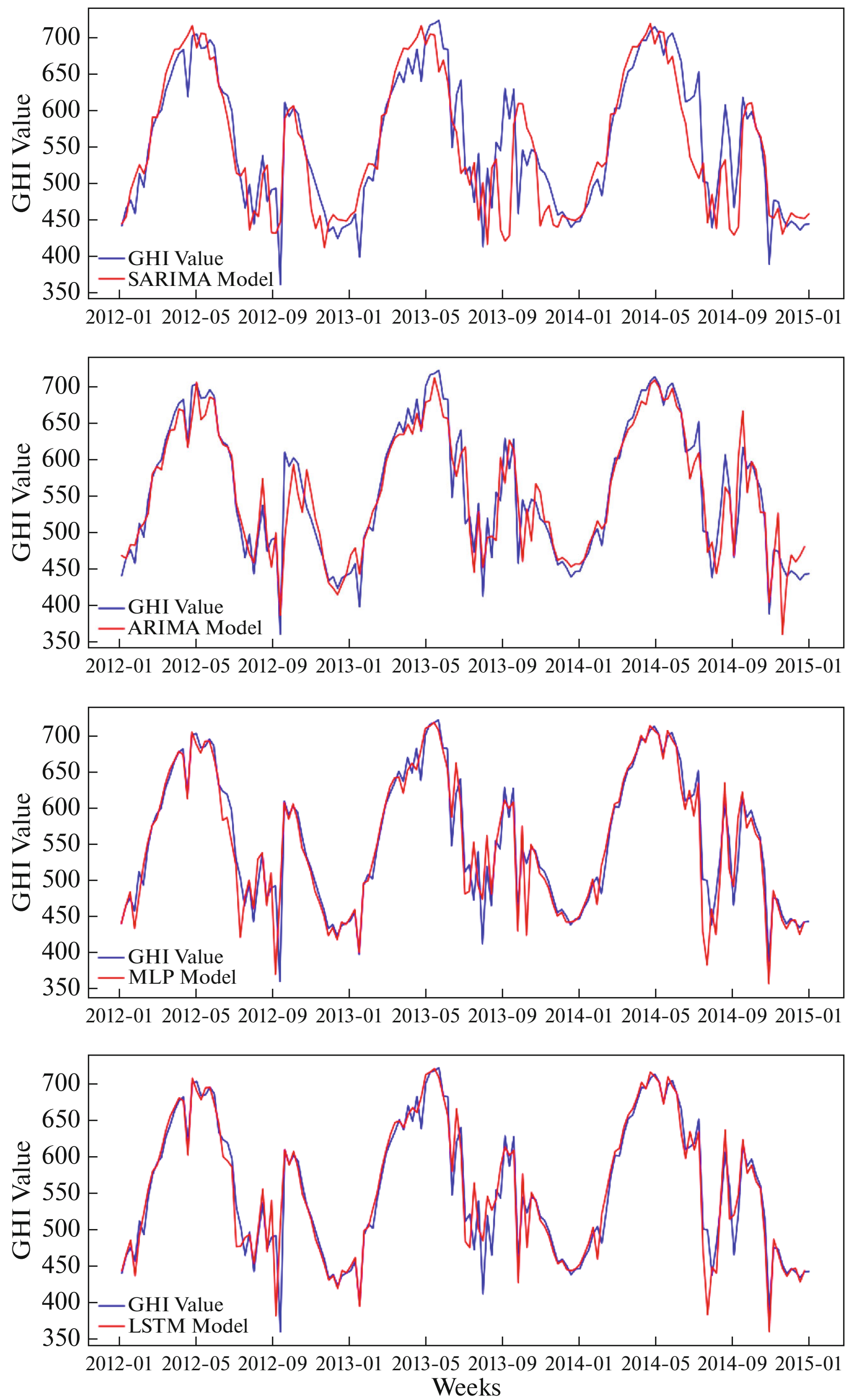


**Fig. 5.** Performance of the forecasting models for weekly dataset from Gujarat.

their randomness is lost, violating the fundamental assumption of a forecast model [20]. Here we perform residual analysis through standardised residual plots and P-P plots corresponding to the best fit models at three chosen time horizons (Figs. 8, 9 and 10). From the figures, it appears that the histogram plots are bell shaped, and the residual plots are symmetric about zero, exhibiting a normal distribution in each of the cases. However, when we carry out the Jarque–Bera normality test, the null hypothesis (that is, residuals follow normal distribution) is accepted for the monthly data with the p-value equals to 0.723; for weekly and daily data, the null hypothesis is rejected. The deviation of residuals from normal distribution in

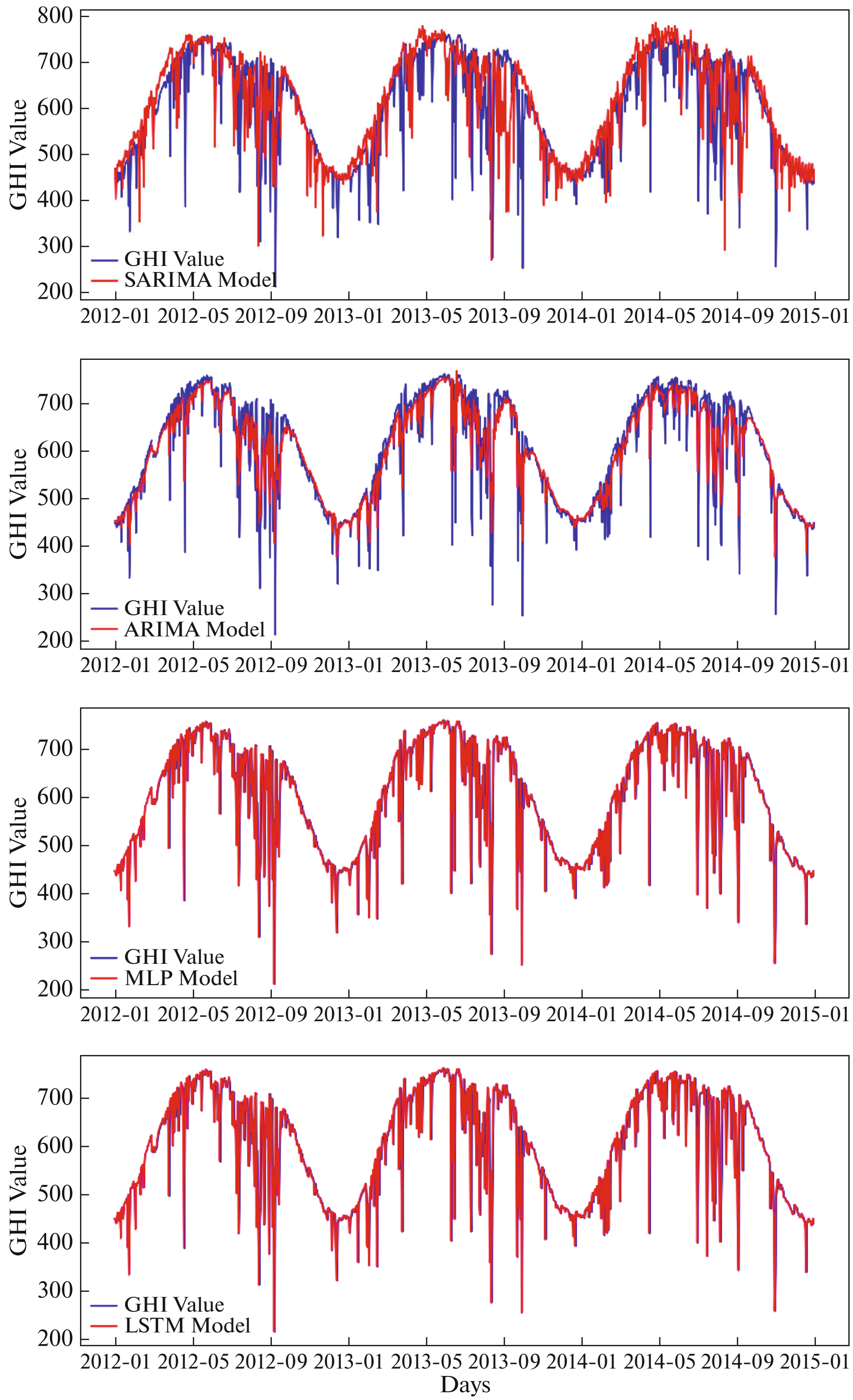


**Fig. 6.** Performance of the forecasting models for daily dataset from Rajasthan.

daily and weekly dataset is possibly due to much more nonlinear variability in weather patterns, causing presence of more outliers at shorter time horizons. Another reason could be the high order of seasonal fluctuations that involve differencing at the 52-week horizon and 365-day horizon, which typically requires a much longer time series to settle down (personal discussion with Gordon Reikard, USA Cellular).

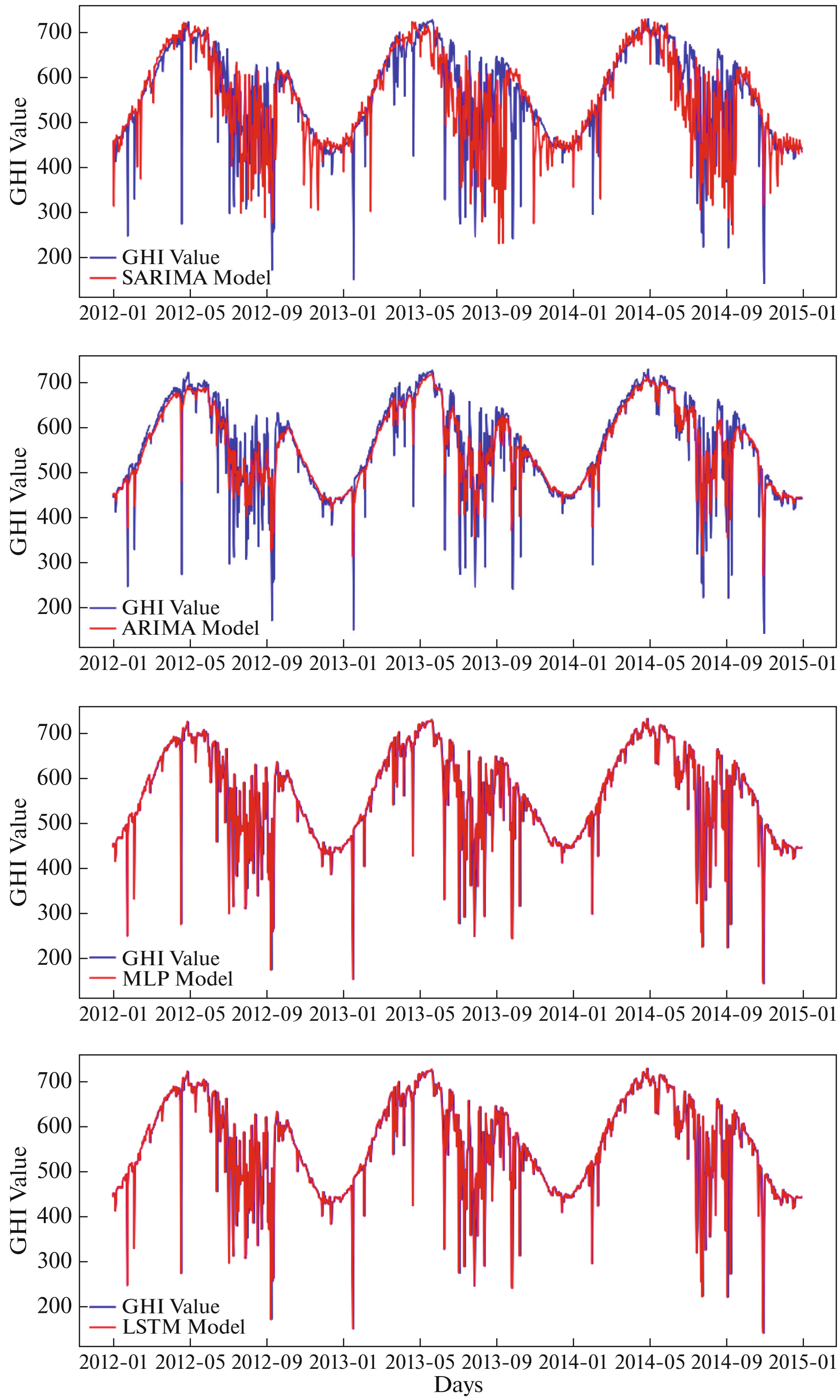


**Fig. 7.** Performance of the forecasting models for daily dataset from Gujarat.

*Concluding Remarks*

The renewable energy resources are regarded as the most pollution-free, abundant and safest resource of energy. India needs to boost the contribution of renewable energy sources in energy production to cater the increasing power demand in a sustainable

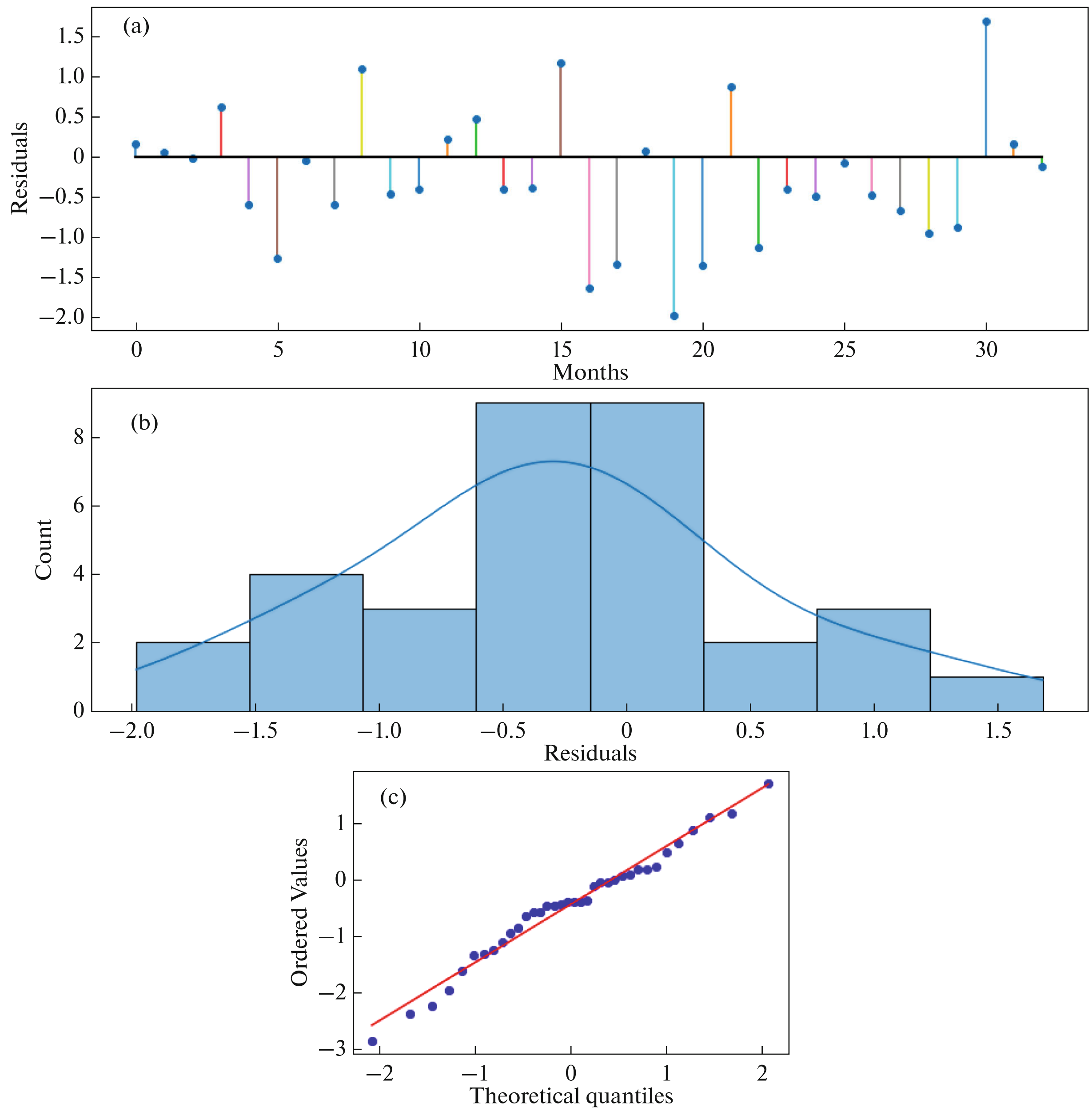


**Fig. 8.** (a) Residual plot, (b) histogram plot and (c) P-P plot of the standardised residuals of SARIMA model for monthly dataset from Rajasthan.

way. A stable forecasting model for solar energy generation is critical due to the involvement of several environmental factors, such as temperature, pressure, humidity and cloud formation. In this study, two time-series models, namely ARIMA and SARIMA along with two machine learning models, namely MLP and LSTM, are implemented for monthly, weekly and daily forecasting of GHI data at two different locations in India. The results bring out following key observations:

• Based on the results of the ADF (Augmented Dickey-Fuller) test, we conclude that the GHI time series is stationary.

• For long-term monthly forecasting, the SARIMA model consistently performs better for both the locations as this time series model can precisely capture the monthly seasonality in the dataset. However, for daily and weekly forecasts, it is hard to exactly recognise the seasonality pattern in the data through SARIMA model.

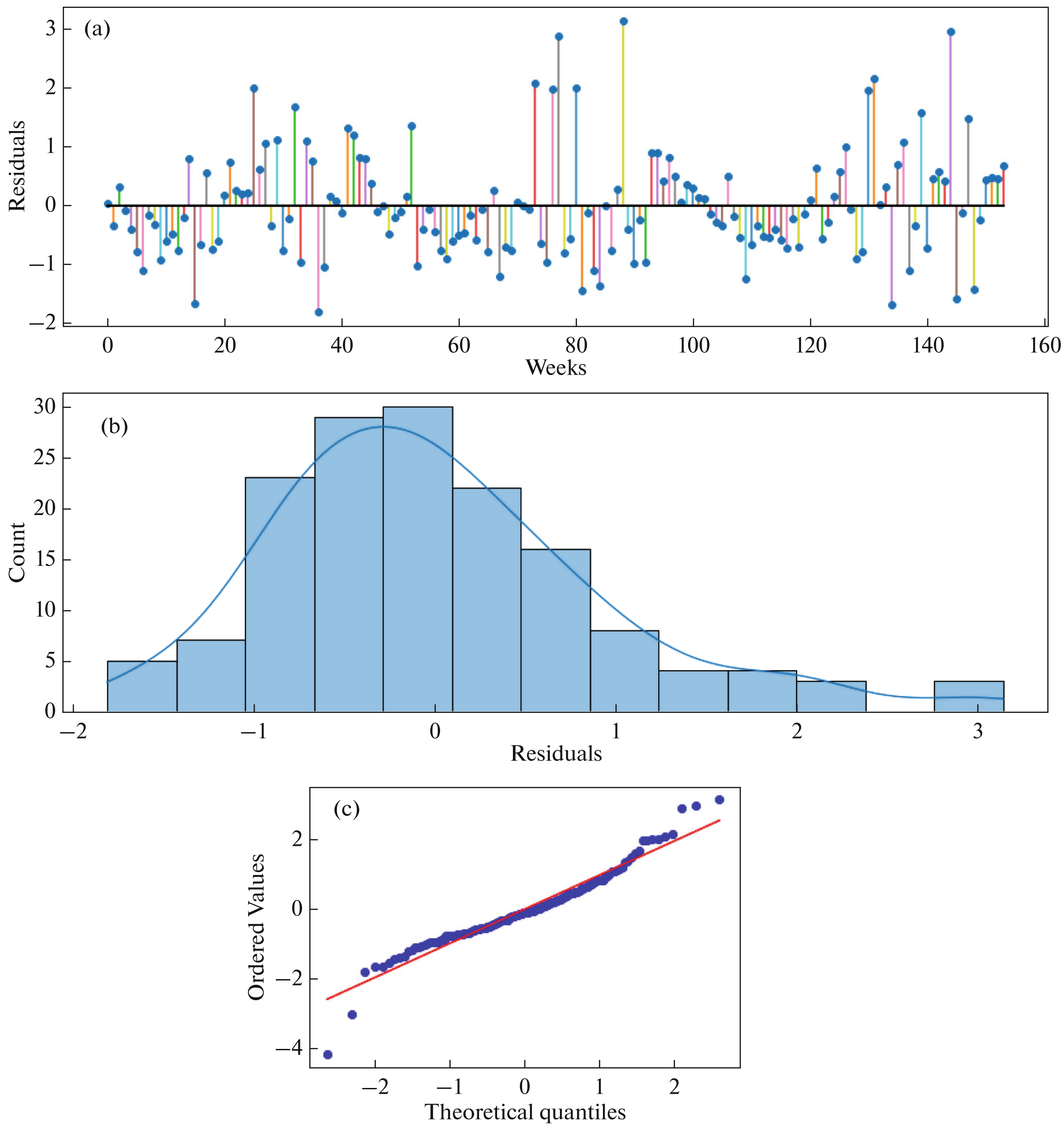


**Fig. 9.** (a) Residual plot, (b) histogram plot and (c) P-P plot of the standardised residuals of ARIMA-WS model for weekly dataset from Rajasthan.

• The SARIMA model could not fully recognise the seasonality pattern in the daily and weekly data. However, the ARIMA model with sliding windows of 365 days and 52 weeks for daily and weekly datasets (based on observed yearly seasonal patterns) provides significant improvement in results.

• For short term daily forecasting, the machine learning models outperform time series models. The overall accuracy of MLP and LSTM models are comparable. Yet, the MLP is preferable as it has much lesser training time than the LSTM.

• As one moves from monthly to weekly to daily, the RMSE of the time-series model increases. This is quite reasonable as the daily dataset is often influenced by the nonlinear variability in weather patterns and high order of seasonal fluctuations [10, 13]. However, the training time and overall computational resources for machine learning models are much higher than the time series models.

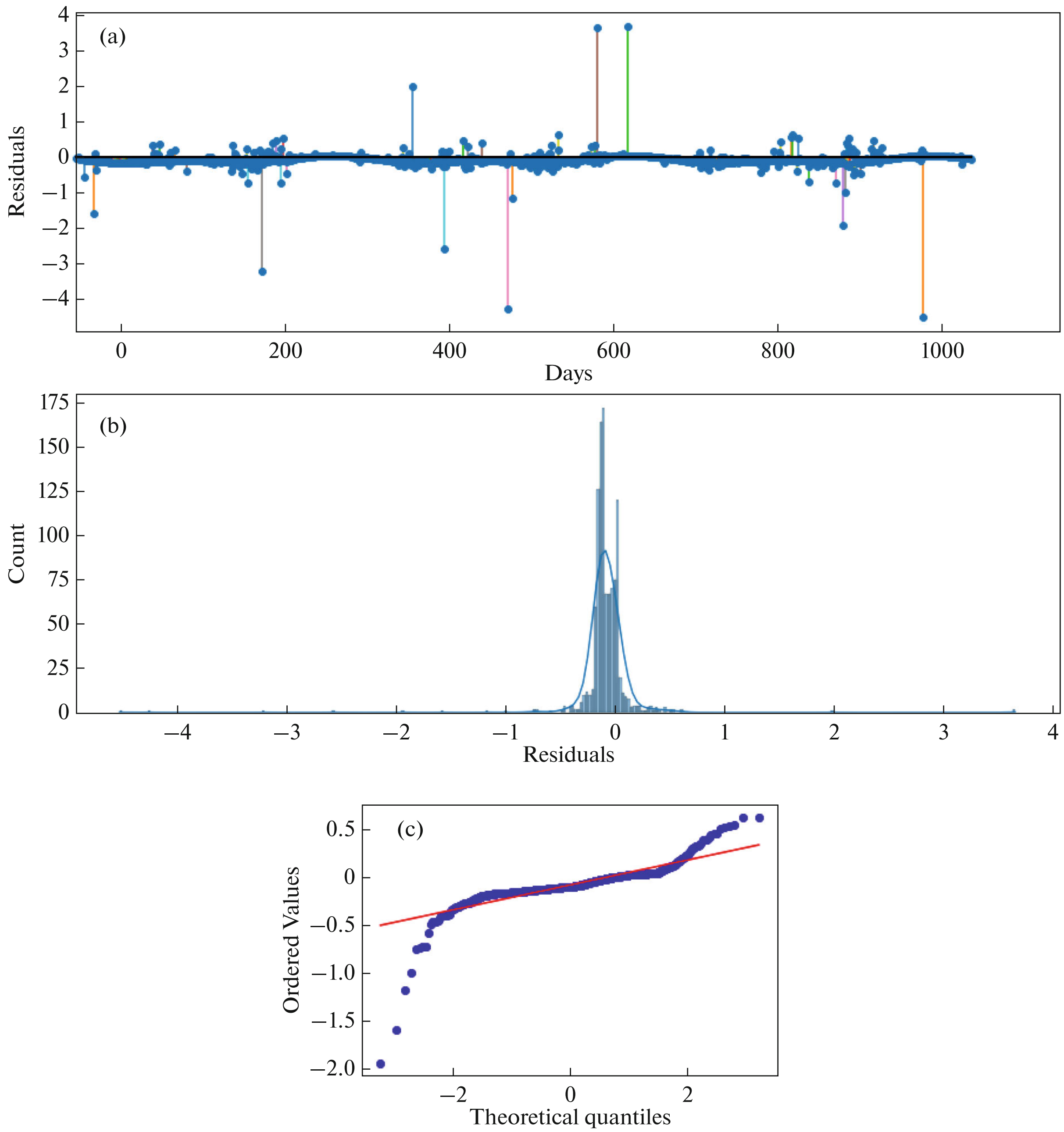


**Fig. 10.** (a) Residual plot, (b) histogram plot and (c) P-P plot of the standardised residuals of MLP model for daily dataset from Rajasthan.

In summary, the present study provides a clear guideline on the selection of forecasting models based on the desired time horizon.

## LIST OF ACRONYMS

| | |
|---|---|
| GHI | Global Horizontal Irradiation |
| SARIMA | Seasonal Autoregressive Integrated Moving Average |
| MLP | Multi-Layer Perceptron |
| LSTM | Long Short-Term Memory |
| MAPE | Mean Absolute Percentage Error |
| RMSE | Root Mean Square Error |
| DHI | Diffuse Horizontal Irradiance |
| DNI | Direct Normal Irradiance |
| ARIMA | Autoregressive Integrated Moving Average |
| ANN | Artificial Neural Network |

| | |
|---|---|
| ARMA | Autoregressive Moving Average |
| DBN | Deep Belief Network |
| RNN | Recurrent Neural Network |

## DATA AVAILABILITY STATEMENT

The dataset for the present study is publicly available in the National Solar Radiation Database (NSRDB) maintained by the US Department of Energy (https://nsrdb.nrel.gov/). The website was last accessed in July, 2021.

## ACKNOWLEDGMENTS

The authors acknowledge two anonymous reviewers for their valuable comments and suggestions. We sincerely thank Professor Gordon Reikard (USA Cellular) for his time-to-time advise on time series models. Partial funding was available from DST-SERB's MATRICS scheme (File No: MTR/2021/000458). The first author acknowledges the research fellowship from UGC, India (Ref. no. 1026/UGC-CSIR-June 2018).

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES

1. Government of India. Ministry of Power, Power sector at a glance all India, July 2021. https://powermin.gov.in/en/content/power-sector-glance-all-india.
2. Lu, T., Sherman, P., Chen, X., Chen, S., Lu, X., and McElroy, M., India's potential for integrating solar and on-and offshore wind power into its energy system, *Nat. Commun.*, 2020, vol. 11, pp. 1–10.
3. Voyant, C., Notton, G., Kalogirou, S., and Nivet, M., L., Paoli, C., Motte, F., and Fouilloy, A., Machine learning methods for solar radiation forecasting: A review, *Renewable Energy*, 2017, vol. 105, pp. 569–582.
4. Impram, S., Nese, S.V., and Oral, B., Challenges of renewable energy penetration on power system flexibility: A survey, *Energy Strategy Rev.*, 2020, vol. 31, p. 100539.
5. Gensler, A., Henze, J., Sick, B., and Raabe, N., Deep learning for solar power forecasting an approach using Auto Encoder and LSTM Neural Networks, 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 002858–002865.
6. Martın, L., Zarzalejo, L.F., Polo, J., Navarro, A., Marchante, R., and Cony, M., Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning, *Sol. Energy*, 2010, vol. 84, pp. 1772–1781.
7. Santhosh, M., Venkaiah, C., and Kumar, V., Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review, *Eng. Rep.*, 2020, vol. 2, p. e12178.
8. Reikard, G., Predicting solar radiation at high resolutions: A comparison of time series forecasts, *Sol. Energy*, 2009, vol. 83, pp. 342–349.
9. Reikard, G., Haupt, S.E., and Jensen, T., Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models, *Renewable Energy*, 2017, vol. 112, pp. 474–485.
10. Reikard, G. and Hansen, C., Forecasting solar irradiance at short horizons: Frequency and time domain models, *Renewable Energy*, 2019, vol. 135, pp. 1270–1290.
11. Yang, D., Jirutitijaroen, P., and Walsh, W.M., Hourly solar irradiance time series forecasting using cloud cover index, *Sol. Energy*, 2012, vol. 86, pp. 3531–3543.
12. Sahin, A., Kumbasar, T., Yesil, E., Doýdurka, M.F., and Karasakal, O., An approach to represent time series forecasting via fuzzy numbers, 2014 2nd Int. Conf. on Artificial Intelligence, Modelling and Simulation, 2014, pp. 51–56.
13. Alsharif, M.H., Younes, M.K., and Kim, J., Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea, *Symmetry*, 2019, vol. 11, p. 240.
14. Sharadga, H., Hajimirza, S., and Balog, R.S., Time series forecasting of solar power generation for large-scale photovoltaic plants, *Renewable Energy*, 2020, vol. 150, pp. 797–807.
15. Shi, J., Lee, W.J., Liu, Y., Yang, Y., and Wang, P., Forecasting power output of photovoltaic systems based on weather classification and Support Vector Machines, *IEEE Transactions on Industry Applications*, 2012, vol. 48, pp. 1064–1069.
16. Isaksson, E. and Karpe Conde, M., *Solar power forecasting with machine learning techniques, Degree Project in Mathematics*, KTH Royal Institute of Technology, Sweden, 2018.
17. Malakar, S., Goswami, S., Ganguli, B., Chakrabarti, A., Roy, S.S., Boopathi, K., and Rangaraj, A., Designing a long short-term network for short-term forecasting of global horizontal irradiance, *SN Appl. Sci.*, 2021, vol. 3, no. 4, pp. 1–15.
18. Yadav, A.K. and Chandel, S., Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model, *Renewable Energy*, 2015, vol. 75, pp. 675–693.
19. Adhikari R., and Agrawal, R.K., An introductory study on time series modelling and forecasting, arXiv:1302.6613, 2013.
20. Anderson, D.R., Sweeney, D.J., Williams, T.A., Camm, J.D., and Cochran, J.J., *Statistics for Business and Economics*, Boston, MA: Cengage Learning, 2016.
21. Nematchoua, M.K., Jose, O., and Afaifia, M., Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries, *Ecol. Inf.*, 2022, vol. 69, p. 101643.