Renewable Energy 83 (2015) 100–109

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Development of statistical time series models for solar power prediction

V. Prema*, K. Uma Rao

Electrical & Electronics Engineering, RVCE, Bangalore, India

A R T I C L E I N F O

Article history:
Received 30 July 2014
Accepted 14 March 2015
Available online 29 April 2015

Keywords:
Forecast
Prediction
Solar irradiance
Time series

A B S T R A C T

The increasing use of renewable energy sources necessitates accurate forecasting models for generation scheduling. Amongst the renewable sources, solar and wind have gained acceptance and are being increasingly used in distributed generation. The main problem with these sources is the dependence of their power output on natural environmental parameters at a given point of time. This paper proposes time series models for short-term prediction of solar irradiance from which solar power can be predicted. The predictions are done for 1 day ahead using different time-series models. For each model, these predicted values are compared with the actual values for the next day and graphs are plotted. Basic time-series models such as moving average and exponential smoothing were tested. The decomposition model is proposed, where the measured data is decomposed into seasonal and trend patterns and each of them predicted separately. The model was developed for different durations of data, to identify the best possible set of data. It is observed from the results that the prediction with decomposition model for 2 months data gave the best result with around 9.28% error.

1. Introduction

The focus on renewable energy sources is increasing tremendously, which motivates studies concerning the integration of renewable sources like wind and solar into existing energy systems. When the conventional sources are used for power generation, the generation is predominantly controlled by the machine ratings and generation capacity of the plant. Hence, there is a lesser need for short term generation forecasting. But in case of power generation from renewable energy sources, the generation is uncertain because the weather is erratic and the generation depends heavily on weather conditions. Thus there is a dire need to build forecast models for the generation in order to have better generation scheduling. Therefore with greater penetration of renewable sources in power generation, the focus is shifting towards generation forecasting [1–3]. Considering the advantages of solar energy as a sustainable energy, power prediction for photovoltaic installations is a decisive factor. Thus, development and research on solar power has been rising year by year. The predictions are used to optimize usage of the solar energy and provide reasonably accurate knowledge of the solar resource availability at any location [4].

Since the generation of power from solar energy is very erratic due to its heavy dependence on weather, seasonal changes, geographical location, time of the day, orientation and position of panel, etc., the forecasting methods may not give uniformly efficient results for all regions. A ubiquitously efficient forecast system, minimizing errors on the behavioural patterns of wind and solar energy has become a major subject matter of study for researchers across the globe [5]. Hence there is a need to critically examine the seasonality and the nature of the data to determine the models that can be satisfactorily used for prediction of solar power generation.

There are two ways to address the issue of solar prediction. One is by creating very complex models which best describe nature. Such models are used by Meteorological departments for weather forecasts. Computations are highly complex and a very powerful computer is used to solve the differential equations involved. The other method is to use statistical data and predict solar irradiance to a lower accuracy as compared to the former method but with less computational requirements. For the application of generation scheduling high accuracy is not demanded. Hence the second method is opted for in this paper and past data is used to predict the future using Time series models. This approach cannot be used to predict weather as accurately as the Meteorological department

* Corresponding author. Tel.: +91 9986041498.
E-mail address: premav@rvce.edu.in (V. Prema).

http://dx.doi.org/10.1016/j.renene.2015.03.038
0960-1481/

does, but is accurate enough to solve the problem of scheduling power from solar power plants.

Currently, the methods for solar generation forecasting are mostly based on Artificial Neural Networks (ANN) for about 32 h ahead. Complex models include multi-layered, feed forward, back-propagation Neural Networks designed using tools like Microsoft Excel, XL Miner and MATLAB [6,7].

A long-term predictive model for PV power output correction using neural networks and fuzzy approach was proposed [8]. An effective methodology to predict solar radiation was proposed in Ref. [9] where a generic model to find the hourly solar radiation at any time in a day, at any site by taking only one measurement in the morning has been developed. This model takes into account the difference in the air mass and the global solar radiation penetrations during the daytime hours. But as the term of the forecast increases, the error in the prediction substantially increases. Hence, efforts were made to predict solar radiation.

Prediction models for solar power generation based on ANN machine learning techniques have been developed and were found to be about 27% more accurate than regression models [10]. Several other forecasting methods such as AR, MA and ARMA have also been used. But the major models are based on MOS (Model Output Statistics), ANN and hybrid models that are a combination of ANN and time series models [11].

Many irradiance forecasting methods have been proposed. There are some outstanding forecasting models based on cloud motion analysis and numerical weather prediction (NWP) [12]. Statistical time series forecasting has also been the most popular technique for short time scales [13]. Linear time series analyses such random walk (RW), autoregressive (AR), moving average (MA), simple exponential smoothing (SES) method, are widely used for modelling and predicting solar irradiance data [14,15].

Considering the fact that solar and wind generation show typical behavioural patterns, with time-varying, seasonal and trend patterns in their generation, predictive mathematical models would serve as effective tools in forecasting the possible generation from these sources, based on past statistics. This would help the operators to be prepared for the load demand, without over-estimating or under estimating the generation capabilities of distributed generation sources. Available statistical analysis involves vast amount of data, in which is hidden the data relevant to predictive modelling. Therefore data mining forms an integral element of predictive modelling [16].

It is observed from literature survey that there is a need for improved models for solar irradiance prediction with more accuracy. The models need to capture extreme variations in the solar irradiance data that is typical of a tropical region like India. This paper explores the various time series techniques to implement solar prediction models. The methods used are moving average and exponential smoothing algorithms. The paper is organised as follows: Section 2 gives the insight of the data used for this work. Section 3 gives a detailed explanation of the various time series models. In Section 4 the results of the developed models are shown. The predicted irradiance is compared with the actual data and the errors are calculated for the developed models. The models are compared in Section 5.

2. Data used

The most important part of any statistical analysis is the collection and sampling of data. The data used for this paper is collected from the Bagalkot weather station, Karnataka, India, for the duration of 01/01/2012 to 30/05/2013. The data samples are measured for every minute in the given period. To reduce the bulk of data, the data is averaged over 10 min. Out of one year data of 2012, 9 months data is used for model development. The rest of the data is used for cross validation. The data contains the following parameters – time, air temperature, relative humidity, solar irradiance and battery voltage.

The solar irradiance being almost nil for 12 h in a day, 50% of the data points account for zero value. With these zero values, forecasted or predicted values are more scattered from the actual values for the regression model. Hence, the analysis is done by excluding zero values and the models are developed to predict the irradiance one day ahead. This information would be very helpful to the load dispatch centre which needs to match generation and load in slots of 15 min–30 min.

3. Time series models

A time series is a set of statistical data, usually collected at regular intervals. A time series is a sequence of observations on a variable measured at successive points in time or over successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular interval. The pattern of the data is an important factor in understanding how the time series has behaved in the past. If such behaviour can be expected to continue in the future, the past pattern can be used to select an appropriate forecasting method. Time series data occur naturally in many application areas. The methods of time series analysis predate those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarize time series data, fit low-dimensional models, and make forecasts [17]. Time series models for prediction of solar irradiance are developed using R software [18]. R is a free software programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

The main types of time series models include – parametric and non-parametric time series models [19]. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters. The non-parametric approach entertains the principle of ‘letting the data speak for themselves’ and avoids the difficulty of identifying an appropriate parametric model, including the non-linear functions and the error distributions. All the time series models used in this paper are parametric models.

3.1. Moving average models

Moving averages provide a simple method for smoothing past history data. There are several straight forward moving averages including simple moving average, double and weighted moving averages. In all cases the objective is to smooth past data to estimate the trend cycle component. Moving average is a fundamental building block in all the decomposition methods. Moving average describes the procedure of trend cycle. Each average is computed by dropping the oldest observation and including the next observation [20]. Simple moving average can be defined for any odd order. A moving average of order k (or k MA) where k is an odd integer is defined as the average consisting of an observation and the $m = (k - 1)/2$ points on either side [1]. The formula for moving average is as given by the equation (1).

$$F_t = \frac{1}{k} \sum_{j=-m}^{m} Y_{t+j} \tag{1}$$

where

F_t = present forecast value of the variable
Y_t = present actual value of the variable

Determining the appropriate length of a moving average is an important task in decomposition methods. One important point to be noted while using Moving Average method for forecasting is that the interval of Moving Average is to be chosen properly. Although a large interval would help in capturing the trend better, too large a value of interval can result in an erroneous forecasting model.

3.2. Exponential smoothing

An obvious extension to the moving average method is forecasting by weighted moving average. With simple moving average forecasts, the mean of the past k observations was used as forecasts. This implies equal weights (equal to 1/k) for all k points. However with forecasting, the most recent observations will usually provide the best guide as to the future, so a weighting scheme that has decreasing weights as the observations get older, need to be used.

A class of methods that imply exponentially decreasing weights as the observations gets older are called exponentially smoothing procedures. There is a variety of exponentially smoothing methods. They have a common property that recent values are given relatively more weight in forecasting than the older observations. In exponential smoothing there are one or more smoothing parameters to be determined explicitly and these choices determine the weights assigned to the observations [1].

3.2.1. Single exponential smoothing (SES)

Single exponential smoothing model provides an option of placing exponentially decreasing value of weights on past values of the variable in order to do the forecast, the immediately recent values having more weightage. The equation for single exponential smoothing forecasting model is given by equation (2)

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t \tag{2}$$

where

F_{t+1} = next forecast value of the variable
F_t = present forecast value of the variable
Y_t = present actual value of the variable
α is a constant between 0 and 1

From the equation (2), it can be seen that the forecast of the future value depends upon the present actual value and the forecast value of the present obtained from the past values. Clearly, this is an auto recursive process which can be represented as shown in equation (3)

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots + \alpha(1-\alpha)^{t-1}Y_1 + (1-\alpha)^t F_t \tag{3}$$

From equation (3), it can be seen that the weightage for the past values of the variable is exponentially decreasing. Hence this method is called the method of exponential smoothing. Here the immediate past value has more bearing on the forecast, when compared to others.

3.2.2. Double exponential smoothing (DES)

Single exponential smoothing method is mostly applicable for simple data with no trend or seasonal component. But when the data has any of these components, the double and triple exponential smoothing is considered for better modelling of the data. This method is also called as the Holt-Winter's trend and seasonality method (after the names of the inventors). For the solar irradiance time series data for 10 days with zero irradiance values, the Holt-Winter's trend and seasonality method is applied in R software. Seasonality is defined to be the tendency of time series data to exhibit behaviour that repeats itself every 'L' periods (where L is the season length in periods). Trend is defined to be the general increasing or decreasing nature of the time series data over a period of time. Double exponential smoothing is used when the data shows a trend. Exponential smoothing with a trend works much like simple smoothing except that two components must be updated each period – level and trend.

The specific formula for simple double exponential smoothing is given by equations (4) and (5). The equation (4) shows the smoothing for the level component – this component is assumed to be devoid of trend and seasonality. And equation (5) gives the smoothing of the trend component.

$$\text{Level}: \ L_t = \alpha Y_t + (1-\alpha)(L_{t-1} - b_{t-1}) \tag{4}$$

$$\text{Trend}: \ b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{5}$$

The forecast is obtained from both the level and trend components, given in equation (6)

$$\text{Forecast}: \ F_{t+m} = L_t + b_{tm} \tag{6}$$

where

α = level smoothing constant (lies between 0 and 1)
β = trend smoothing constant (lies between 0 and 1)
L_t = estimate of the level of the series at time t
Y_t = actual value of the series at time t
b_t = estimate of the slope of the series at time t
m = number of periods ahead to be forecast
F_{t+m} = forecast for m periods ahead

3.2.3. Triple exponential smoothing (TES)

Triple exponential smoothing method is used when the data shows trend and seasonality. To handle seasonality, we have to add a third parameter. We now introduce a third equation to take care of seasonality.

$$\text{Level}: \ L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{7}$$

$$\text{Trend}: \ b_t = \beta\ (L_t - L_{t-1}) + (1-\beta\ )b_{t-1} \tag{8}$$

$$\text{Seasonal}: \ S_t = \gamma(Y_{t-1} - L_t) + (1-\gamma)S_{t-s} \tag{9}$$

The forecast is obtained from the level, trend and seasonal components as in equation (10)

$$\text{Forecast}: \ F_{t+m} = L_t + b_{tm} + S_{t-s+m} \tag{10}$$

where

α = level smoothing constant (lies between 0 and 1)
β = trend smoothing constant (lies between 0 and 1)
γ = seasonal smoothing constant (lies between 0 and 1)
L_t = estimate of the level of the series at time t
Y_t = actual value of the series at time t
b_t = estimate of the slope of the series at time t
S_t = seasonal component
s = length of seasonality
m = number of periods ahead to be forecast
F_{t+m} = forecast for m periods ahead

3.2.4. Decomposition

Any time series data can be decomposed into several patterns to analyse how each pattern influences each data in the series. Time series data is mainly composed of seasonal pattern and trend pattern. There are two different decomposition models possible.

i) *Additive decomposition*: Here the total data is taken as the sum of the decomposed patterns

$$X_t = \text{seasonal}(S_t) + \text{Trend}(T_t) + \text{Random}$$

ii) *Multiplicative decomposition*: Here the given time-series data is treated as the product of the decomposed patterns

$$X_t = \text{seasonal}(S_t)^*\text{Trend}(T_t)^*\text{Random}$$

Additive decomposition is effective when the peak values of the seasonal data do not vary much. Multiplicative models are effective when the seasonal value changes over time. Both additive and multiplicative models were tested for the available data. It was found that the multiplicative model works well. The reason could be that the solar irradiance changes with the change in weather. Fig. 1 shows the decomposition of solar irradiance for the month of May 2013 using R.

Forecasting the decomposed seasonal and trend patterns separately to predict the solar irradiance worked very well in the case of solar irradiance. This analysis was done in Microsoft Excel using the package NumXL. The following procedure was used for the forecast.

The trend pattern was obtained by taking the centred moving average of the entire data. The interval of moving average is chosen as one day. Fig. 2 shows the trend pattern generated for May 2013. The de-trended series is calculated by dividing the actual data by the trend. To estimate the seasonal component, the day is divided into 10 min durations, giving 72 non-zero samples. The seasonal component for each 10 min period is estimated by averaging the de-trended values for that particular time period of each day, over a period of 1 month (If more data duration is considered like 2 months, 4 months, etc., the average is taken over the whole duration). The seasonal pattern for a day is thus obtained and this pattern is assumed to remain the same for all days of the duration. Fig. 3 shows the decomposed seasonal pattern for May 2013. The forecast for the seasonal component can easily be estimated by extending the seasonal pattern to the next day. The forecast for the de-seasonalised data is obtained by performing a simple linear regression for the de-seasonalised data.

The steps involved in developing the multiplicative decomposition model are listed below.

1. Obtain the trend pattern by calculating the centred moving average for the entire data.
2. Find out the de-trended pattern by dividing the total data by the trend series.
3. Seasonal component is calculated by averaging the de-trended values for that particular time of all the days
4. Calculate the de-seasonalised data by dividing the seasonal data from the actual data
5. Bring out a simple regression model for the de-seasonalised data
6. Forecast seasonal data by extending the data for the next day
7. Forecast de-seasonalised data by forecasting the regression model
8. Multiply both the forecasted data to get the forecast for the actual data

4. Forecasting models

The above mentioned forecasting methods were developed and tested for various durations of data to understand what amount of data gives the best forecast. Among the available data, four months data, Feb 2013 through May 2013, were considered for comparison of various models. Initially 30 days data for the month of May was considered to predict the solar irradiance for the 31st day of May. Then the month of April and May were considered for the forecast of the last day of May. Finally all the four months data were taken to predict the irradiance for 31st May. The results are compared.

The accuracy of forecasting was evaluated by estimating the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). Error is the difference between the actual solar irradiance and the predicted solar irradiance. MAE is calculated as the mean of the absolute values of the errors. MAPE is the mean of the percentage of absolute errors. MAPE is more easily interpretable.

4.1. Moving average model

The moving average models were tested for various values. For larger values of k the forecast values are not found to be accurate. For lesser values of odd intervals, the forecasted values are almost equal to actual values. Table 1 compares the predicted and actual values of irradiance for the selected period.

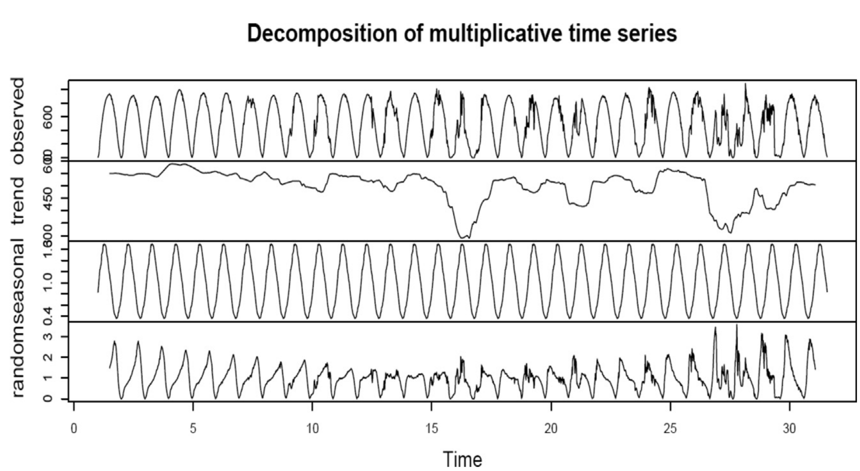

Fig. 1. Decomposition of solar irradiance of May 2013.

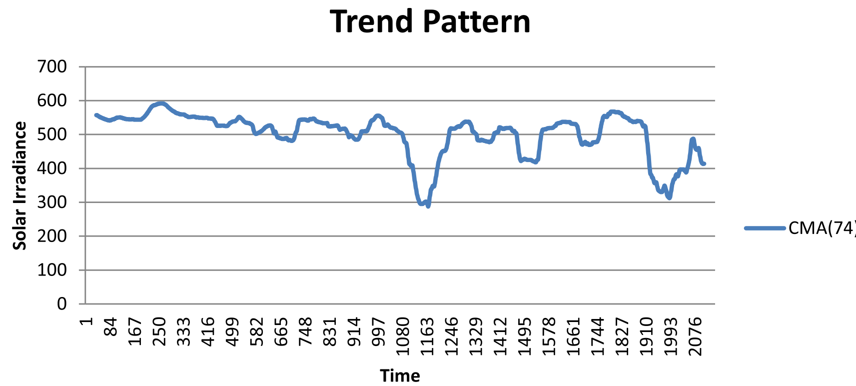

Fig. 2. Trend pattern for May 2013.

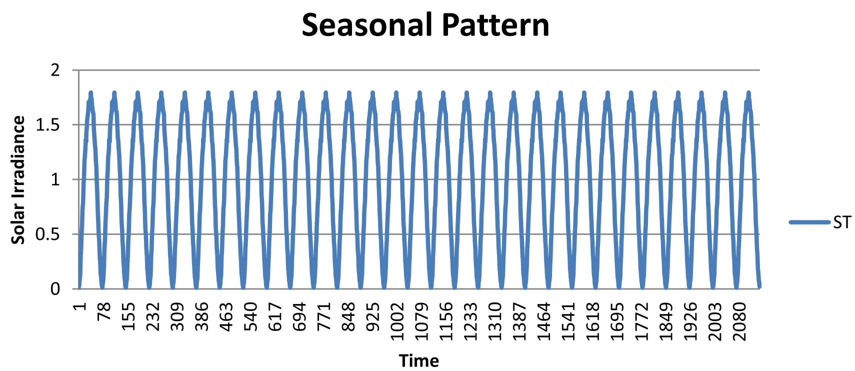

Fig. 3. Seasonal pattern for May 2013.

It can be observed from the Table 1 that for the lower order moving average models, though the predicted values are good for higher values of solar irradiance, when predicting lower values, tends to saturate towards a constant value. Due to this disadvantage, other models such as single and double exponential smoothing are to be tried.

4.2. Exponential smoothing models

Single and double exponential smoothing models were developed using R software. Fig. 4 shows the one-day forecast done for double exponential smoothing model with the values of smoothing parameters as $\alpha = 0.7631$ and $\beta = 0.2027$.

It can be observed from the figure that the forecast does not give a good result with a negative confidence interval. Thus triple exponential smoothing models were developed for different durations of data. The forecast for the last day of May was done using R software. Table 2 gives the best smoothing parameters for the developed models. The predicted value was compared with actual value and the errors were calculated. Figs. 5–7 give the comparison of actual solar irradiance with predicted values.

The predicted values of solar irradiance for the developed Triple Exponential Models were compared with the actual data available. The calculated values of errors are tabulated in Table 3.

It can be observed from the table that the error is least for the model with only one month data and maximum for models with 4 months data. This clearly indicates that the prediction is more accurate with the immediate one month data. The reason is that the climate pattern changes between months. By a close analysis of the data, it can be seen that there is a significant increase in the solar irradiance in the month of May as this month is the peak of summer in India. Thus we can say that with this model, where seasonal variations are very prominent, accuracy worsens if more statistical historical data is used.

Table 1
Results comparing moving average model values for different intervals.

Time (in 10 min)	Actual solar irradiance (W/m^2)	MA interval = 3	MA interval = 5	MA interval = 7	MA interval = 15
1	904.2	898.3667	889.12	875.7286	799.1467
2	898.3	901.4333	895.26	884.4429	818.86
3	906	902.8333	899.88	892.8429	834.5533
4	918.8	907.7	905.82	900.1571	849.1533
5	917.3	914.0333	908.92	905.0714	861.9933
6	911.6	915.9	910.4	908.2857	872.2467
7	893.2	907.3667	909.38	907.0571	879.6933
8	874.2	893	903.02	902.7714	884.4867
9	833.9	867.1	886.04	893.5714	885.56
10	857.3	855.1333	874.04	886.6143	886.8933
%error (max.)	–	**3.9813%**	**6.2525%**	**7.1557%**	**11.618%**

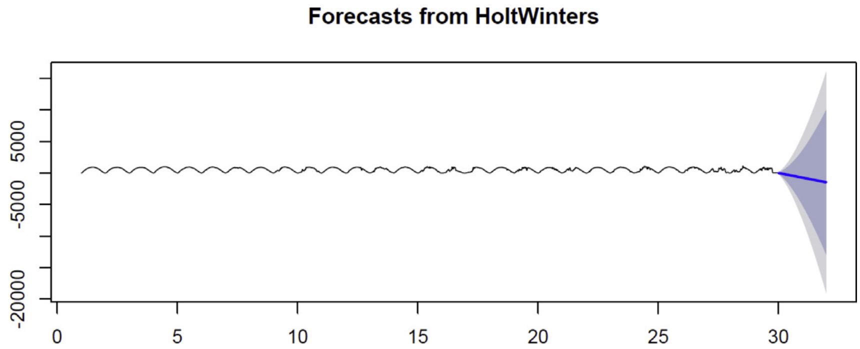

Fig. 4. Plot of the predicted values of solar irradiance for 1 day ahead using DES model.

Table 2
Smoothing parameters of the developed models.

Sl no	Model	Alpha	Beta	Gamma
1	TES for May 2013	0.9012228	0	0.22904
2	TES for April & May 2013	0.8119651	0	0.1034892
3	TES for 4 months data	0.7512265	0	0.03048048

4.3. *Decomposition models*

The decomposition models discussed in Section 3.2.4 were tested with different duration of data. The solar irradiance prediction was done for the last day of May 2013 by developing decomposition models for one month data, two months data (April and May) and 4 months data (February, March, April and May) separately. The predicted solar irradiance values are compared with the actual values. Mean Error (ME), Mean Absolute Error (MAE) and Mean Percentage Errors are calculated.

The predicted values of solar irradiance for the developed decomposition models were compared with the actual data available. Figs. 8–10 shows the comparison of predicted values with actual values of solar irradiance using decomposition model. The calculated values of errors are tabulated in Table 4.

It can be observed from the various models discussed above that the decomposition model reduces the error significantly. The best result is obtained with two months data. It can be seen that the error of 10.52% is lower than the errors for the other models found in literature [7] where the error reported is 13.56% This could be because the decomposition models give equal weightage to seasonal and trend patterns. Thus for further analysis, decomposition model is considered.

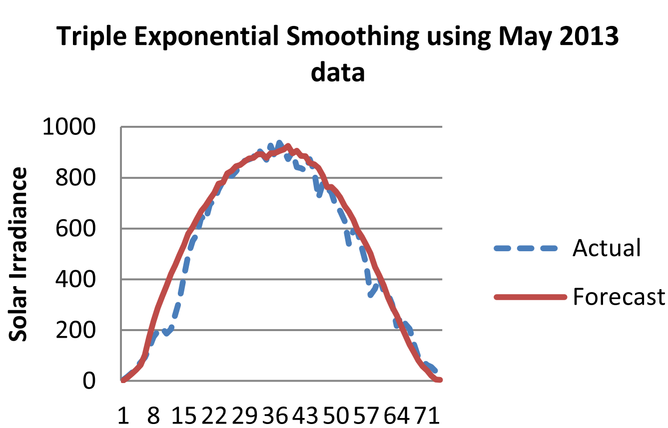

Fig. 5. Plot of the predicted values for 1 day ahead using TES model for May 2013.

5. Results and discussions

Decomposition model is analysed with different time periods of samples and for different forecasting intervals. Several cases were tested for different climatic conditions. A data set of solar irradiance for 1 year from 01/01/2012 to 31/12/2012 is considered for these case studies which are discussed below. Since the data size is huge, the data is averaged over 15 min intervals.

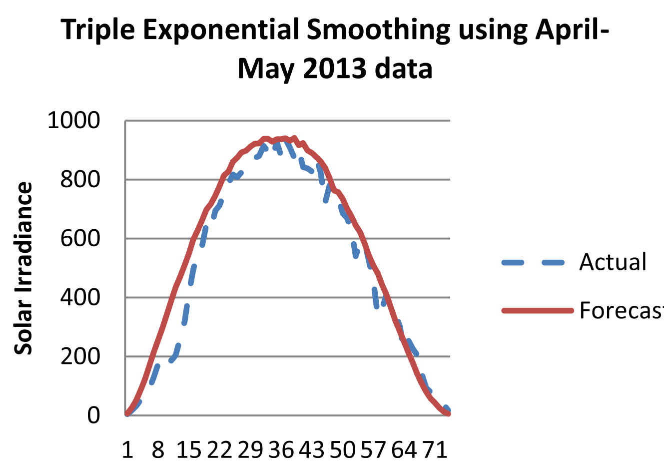

Fig. 6. Plot of the predicted values for 1 day ahead using TES model for April & May 2013.

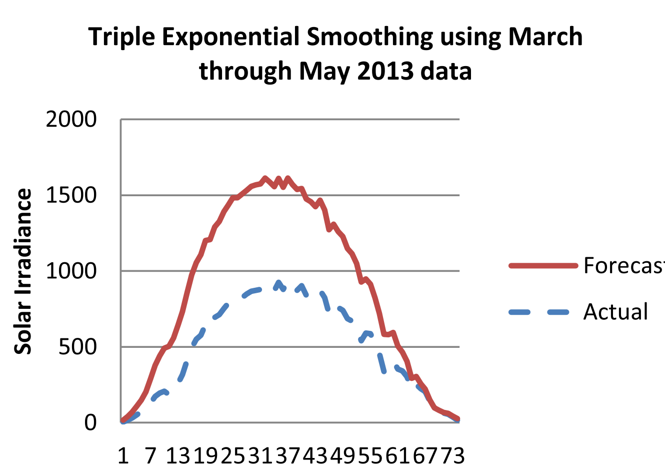

Fig. 7. Plot of the predicted values for 1 day ahead using TES model for March through May 2013.

Table 3
Comparison of errors of TES models.

Sl no	Model	Data period	MAE (W/m^2)	MAPE (%)
1	Triple exponential series	1 month (May)	45.28	18.42
2	Triple exponential series	2 month (April–May)	56.75	23.42
3	Triple exponential series	4 months (February through May)	144.86	38.94

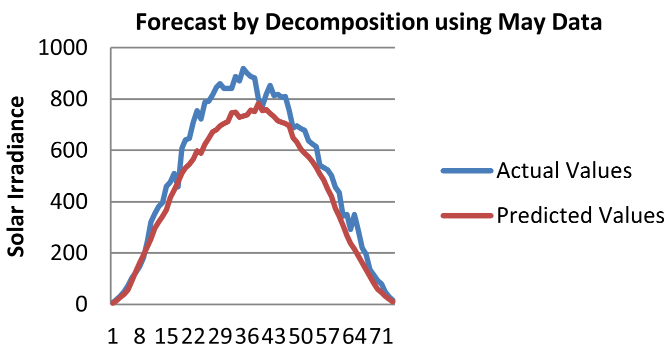

Fig. 8. Plot of the predicted values for 1 day ahead using decomposition model using May 2013 data.

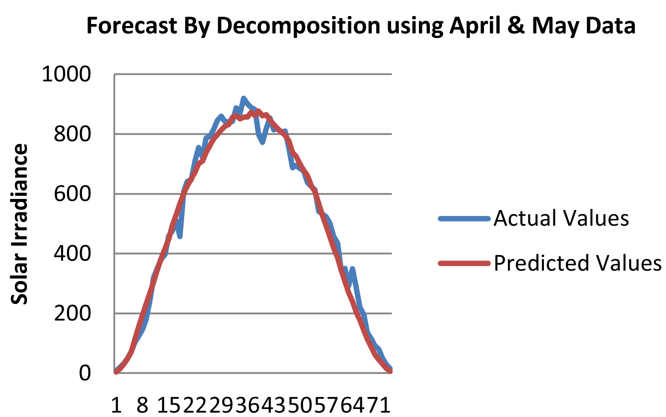

Fig. 9. Plot of the predicted values for 1 day ahead using decomposition model using April & May 2013 data.

5.1. Case study 1: to determine the ideal time period for model building

A main issue in model building is to determine the time duration of the historical data to be considered. In this analysis, various time durations of data for summer and winter seasons were tried to determine the ideal duration for the development of model.

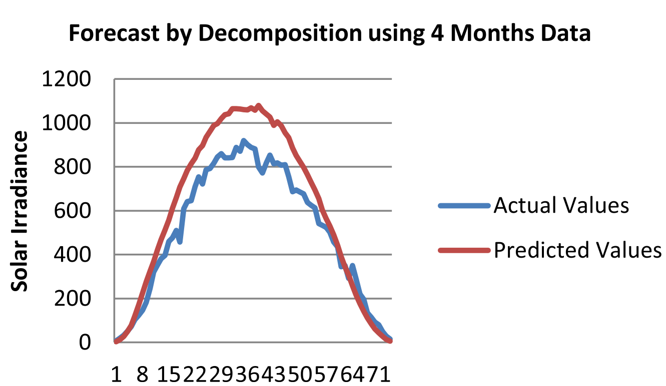

Fig. 10. Plot of the predicted values for 1 day ahead using decomposition model using February through May 2013 data.

5.1.1. Summer

The solar irradiance for one typical day in summer season is predicted. For the development of model, three time durations are considered. The irradiance for the last day of March 2012 is predicted by developing decomposition models for one month data, two months data (February and March) and 3 months data (January, February and March) separately. The prediction plots are shown in Fig. 11. The predicted solar irradiance values are compared with the actual values. The errors are calculated and tabulated in Table 5.

5.1.2. Winter

The solar irradiance for one typical day on summer season is predicted. For the development of model, three time durations are considered. The irradiance for the last day of September 2012 is predicted by developing decomposition models for one month data, two months data (August and September), 4 months data (June to September) and 9 months data (January to September) separately. The prediction plots are shown in Fig. 12. The predicted solar irradiance values are compared with the actual values. Mean Error (ME), Mean Absolute Error (MAE) and Mean Percentage Errors are calculated and tabulated in Table 6.

It can be observed for the above analysis that during summer, where the weather is clear and sunny, the percentage error is as less as 9.28%. But during winter, the variation in the irradiance is erratic and this causes an increase in error. It can also be observed that as the data duration is increased, the error is also increased. In case of winter a longer duration of 9 months also improved the prediction because in the geographical area considered, the month of October has a second peak in temperatures similar to that observed during March. However, longer durations were found to deteriorate the model performance in summer months. Different seasons were considered and it was observed that two month historical data is best suited for model development and is able to best capture the seasonal variations in tropical temperature.

5.2. Case study 2: extension of models to predict out samples

The objective of this case study is to see the extent of out-of-sample prediction which can be performed with a developed model. A decomposition model with January and February 2012 data is developed to predict irradiance for the first day of March, first week of March and the entire March month. The predicted data is compared with the actual irradiance the curves are shown in Figs. 13–15. The errors are tabulated in Table 7.

It can be observed that MAPE increases with increase in forecast horizon. The prediction can be extended for a week where the MAPE is 13.94%. As per Government of India, a deviation of 30% in the forecast is acceptable [21]. Hence the prediction horizon can be taken to be 1 month, which means that after a month the model has to be rebuilt with the new data set available.

Table 4
Comparison of errors of decomposition models.

Type of forecast	Data period	MAE (W/m^2)	MAPE (%)
Decomposition of data using excel	1 month (May)	75.43	17.15
Decomposition of data using excel	2 months (Apr & May)	27.75	10.52
Decomposition of data using excel	4 months (Feb through May)	110.17	23.96

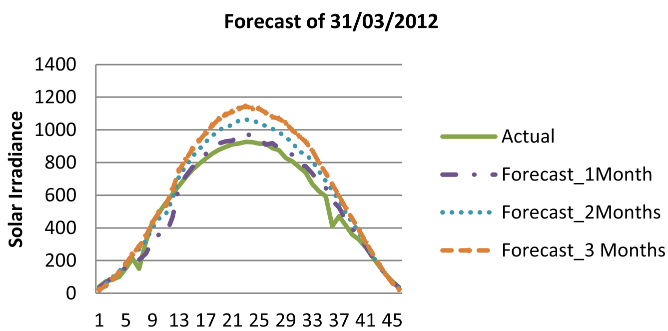

Fig. 11. Plot of the predicted values for 31 March 2012 using decomposition model with three different data sets.

5.3. *Case study 3: model performance for erratic cloudy weather*

This case study is intended to demonstrate the performance of the model for extremely erratic weather conditions. 31st July 2012 is chosen for this study as this day is cloudy and found to show the extreme variation in irradiance in the entire years' data. Fig. 16 shows the comparison of predicted to the actual irradiance. MAPE is calculated to be 25.3%, which is higher than the error for sunny days, but still is in the acceptable limits.

6. Comparison with existing models

A comparison of the developed models with other existing models becomes difficult as different groups use different error measures. Most of the papers in the literature quantify errors in terms of Root Mean Square Error (RMSE) which is not a parameter to be generalized as it mainly depends on the size and resolution of the system. An effort is done to compare the proposed model with some of the existing models which uses time series predictive models.

Two models are developed in Ref. [6], one using only past observations and the other using NWP data along with the historical data. The first model uses adaptive linear Auto Regression (AR) model and the second uses exogenous AR model. Here the RMSE is calculated only for the model, but no cross validation is done. In our work, each model is cross validated and MAPE is calculated for the predicted values.

In Ref. [9], a model is developed to estimate mean expected hourly solar radiation for any hour of the day based on only one measurement in the morning. The predicted values for Jan 17 and July 17 are compared with the actual data. The comparison is only through curve plots, but the errors are not quantified. In Ref. [22] two ANN models with Statistical Feature Parameters are developed. The forecast errors are calculated in comparison to the actual values. The MAPE for cloudy day goes up to a maximum of 81%, which is very high. In our models, the maximum MAPE for cloudy days is around 25%.

7. Conclusion

By using various techniques in Microsoft excel and R software, short term prediction of solar irradiance for one day ahead is carried out. The work started with the moving average model which is the simplest time series model. It was found that lower order models give good result, but for lower values of solar irradiance, the predicted irradiance gets saturated. Thus, exponential smoothing models and decomposition models were tested. The developed models can predict the solar irradiance of the next day. The multiplicative decomposition model was developed for different data duration and prediction done for different out-of-sample periods. It

Table 5
Comparison of errors of decomposition models during summer.

Data duration	MAE (W/m^2)	MAPE (%)
1 month (March 2012)	39.52	9.28
2 months (Feb & Mar 2012)	76.70	15.92
3 months (Jan–Mar 2012)	113.17	21.79

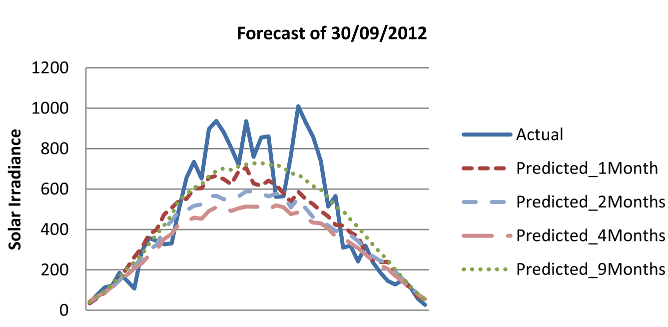

Fig. 12. Plot of the predicted values for 30 September 2012 using decomposition model with four different data sets.

Table 6
Comparison of errors of decomposition models during winter.

Data duration	MAE (W/m^2)	MAPE (%)
1 month (September 2012)	114.82	26.34
2 months (Aug & Sept 2012)	135.95	28.51
4 months (Jun–Sept 2012)	186.04	49.63
9 months (Jan–Sept 2012)	99.58	26.79

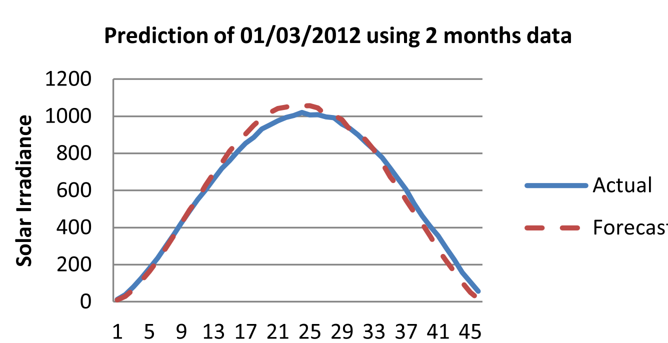

Fig. 13. Plot of the predicted values for 1st March 2012 using decomposition model.

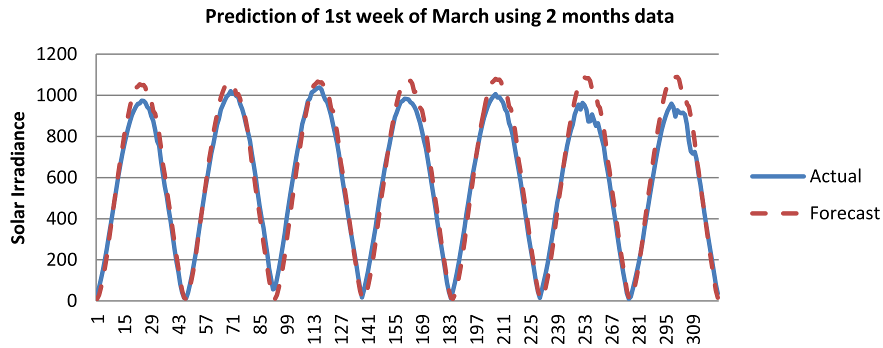

Fig. 14. Plot of the predicted values for the first week of March 2012 using decomposition model.

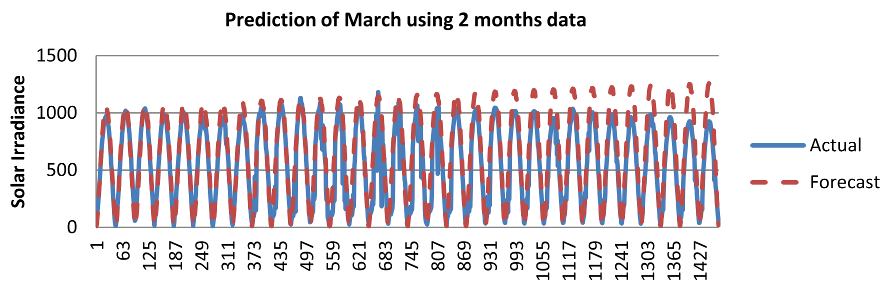

Fig. 15. Plot of the predicted values for March 2012 using decomposition model.

Table 7
Comparison of errors of decomposition models for different out of sample prediction.

Forecast horizon	MAE (W/m^2)	MAPE (%)
1 day	46.08	12.22
1 week	53.59	13.94
1 month	105.66	29.10

was found that the decomposition model works well with a maximum data duration of 2 months as in India, there are sudden changes in weather every 2 months. It was also found that the prediction can be extended upto a week without much increase in the error. The MAPE were calculated and decomposition model gave MAPE as less as 9.28%. The error is increased on cloudy days because of the sudden variations in the solar irradiance, but the error obtained was still far less than what is reported in the literature. If the data is viewed as a signal, the sudden variations can be treated as high frequency components. The performance can be improved if these high frequency components are captured. Work is being carried out by using wavelets to capture the high frequency components, which can be incorporated in the decomposition model to improve the prediction.

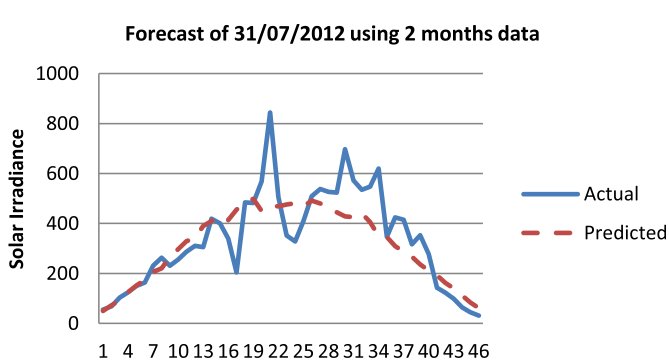

Fig. 16. Plot of the predicted values for 31st July 2012 using decomposition model.

Acknowledgement

The authors gratefully acknowledge the contributions of Dr. Suresh H Jangamshetti, Dept of EEE, Basaveshwar Engineering college, Bagalkot for providing the wind data without which this work would not have been possible.

References

[1] Makridakis Sypros, Wheelwright Steven C, Hyndman Rob J. Forecasting methods and applications. Wiley Student Edition. 3rd ed., ISBN 978-0-471-532330.
[2] Lim Pei Yi. Power management strategies for off-grid hybrid power systems [Ph.D thesis]. School of Electrical and Computer Engineering, Curtin University; Oct. 2011.
[3] Bacher Peder. Models for efficient integration of solar energy [Ph.D thesis]. Technical University of Denmark; 2012.
[4] Madsen Henrik, Nielsen Henrik Aalborg, Jonsson Tryggvi, Pinson Pierre, Bacher Peder. Forecasting of wind and solar power production [Research thesis]. Technical University of Denmark; Nov. 2011.
[5] Prema V, Uma Rao K. Predictive models for power management of a hybrid microgrid-a review. In: 2nd International Conference on Advances in Energy Conversion Technologies (ICAECT 2014). Manipal, India: Manipal Institute of Technology; 23-25 Jan. 2014. p. 7–12. ISBN: 978-1-4799-2206-2/14.
[6] Bacher Peder. Short-term solar power forecasting [Research thesis]. Technical University of Denmark; 2008.
[7] Mahamad Nurhashinnah, Kamal Muhamad, Mohammed Amin B. Power prediction analysis using artificial neural network in MS Excel. In: 7th International Conference on Latest Trends in Renewable Energy and Environmental Informatics, Malaysia; 2013. p. 54–8.
[8] Yona Atsushi, Senjyu Tomonobu, Funabashi Toshihisa, Kim Chul-Hwan. Determination method of insolation prediction with fuzzy and applying neural network for long-term ahead PV power output correction. IEEE Trans Sustain Energy Apr. 2013;4(2):527–33.
[9] Kaplanis S, Kaplani E. A model to predict expected mean and stochastic hourly global solar radiation I(h;nj) values. Elsevier Int J Renew Energy 2007;32:1414–25.
[10] Sharma Navin, Sharma Pranshu, Irwin David, Shenoy Prashant. Predicting solar generation from weather forecasts using machine learning. In: IEEE International Conference on Smart Grid Communications (Smart Grid Comm), Brussels, Belgium; 2011. p. 376–82.
[11] Espinar B, Aznarte JL, Girard R, Moussa AM, Kariniotakis G. Photovoltaic Forecasting: a state of the art. In: 5th European PV-Hybrid and Mini-Grid Conference-Tarragona, Spain; 2010. p. 65–70.

[12] Camacho Eduardo F, Samad Tariq, Garcia-Sanz Mario, Hiskens Ian. Control for renewable energy and smart grids. Int J Impact Control Technol 2011;6: 19–25.
[13] Dong Zibo, Yang Dazhi, Walsh Wilfred M, Reindl Thomas, Abrle Armin. Short-term solar irradiance forecasting using exponential smoothing state space model. Int J Energy 15 Jun. 2013;55:1104–13.
[14] Kleissl Jan. Solar energy forecasting and resource assessment. 1st ed. Academic Press; Jul. 2013. ISBN: 978-0-123-971777.
[15] Wu Ji, Chan Koeng. The prediction of monthly average solar radiation with TDNN and ARIMA. Int J Sol Energy May 2011;85(5):808–17.
[16] Mori H, Takahashi A. A data mining method for selecting input variables for forecasting model of global solar radiation. In: IEEE PES Transmission and Distribution Conference and Exposition (T&D), Orlando, Florida – USA; 2012. p. 1–6.
[17] Brockwell PJ, Davis RA. Time series: theory and methods. 2nd ed. Springer; 2013.
[18] Teetor Paul, Cookbook R: OReilly. ISBN:978-0-596-80915-7.
[19] Hamilton James D. Time series analysis. 1st ed. Princeton University Press; 1994. ISBN: 978-0-691-0-42891-6.
[20] Box George EP, Hunter J Stuart, Hunter William G. Statistics for Experimenters: design, discovery, and innovation. 2nd ed. John Wiley and sons Publications; 2005. 13 978-047 J -71813.
[21] 'Statement of objects and reasons', (terms and conditions for tariff determination from renewable energy sources) regulations. New Delhi: Central Electricity Regulatory Commission (CERC); 2012.
[22] Sharma N, Gummeson J, Irwin D, Shenoy P. Cloudy computing: leveraging weather forecasts in energy harvesting sensor systems. J Energies 2012;5. ISSN: 1996-1073:1355–70.