

# Comparative Study of Parameter Estimation Methods for Solar Irradiation Forecasting

Yashwant Kashyap, Ankit Bansal, Anil K. Sao, and Satyanarayan Patel

**Abstract**—In this paper, autoregressive moving average (ARMA) has modeling for solar irradiation forecast by combining two types of parameter estimation methods, Forgetting Factor (FF) and Kalman Filter (KF). For this purpose, the geographical location of length, width and average height than 76.75 e, 31.75 N and 1130.3 meters were used. Parameter with regard to the mean absolute error (MAE), root mean square error (RMSE), mean square error (MSE) and  $R^2$  estimator is compared. The result shows that the KF consists of high convergence rate to solve complex problems.

**Index Terms**—Solar irradiation, autoregressive moving average, forgetting factor, Kalman filter.

## I. INTRODUCTION

Today, researchers are focusing on the use of renewable energy sources, this is due to the fact that it will reduce air pollution and meet the current and future energy needs. However, the exact number of energy from these sources is unknown. This is due to the fact that the power generation is dependent on the number of parameters (for example, wind speed, insolation, waves and water plants, among others), which must be dissolved to leave this form of energy. In the production of renewable energy, is one of the most important resources of the incident solar radiation. It is depending on the geography of the earth (height, length, width and surface angle), Sun-Earth position hour, azimuth and inclination angle [1]. In India incident solar radiation varies from 44 to 77 per cent per kWh m<sup>2</sup> per day, the other varies from 68 to 82 percent under a clear sky days per year [2]. It is also affected by atmosphere and weather phenomena such as air mass, suspended particles, water vapor, aerosols, clouds, humidity, temperature, SO<sub>2</sub>, soot, CO<sub>2</sub>, aerosol Optical depth [3]. Therefore sunshine forecast for energy producers is necessary. Furthermore, it is very important issue for larger solar power stations Grid intergraded. In addition, helps forecast the consumer, energy planning and management company, which is advantageous from both an economic and technical point of view [4].

In this regard, several methods/models have been proposed to be radiation at the ground, the physics-based models,

moving average, Classic Car regression, auto regression and moving average, Markov Chain and Fuzzy logic includes predict, etc. Furthermore, an adaptive process such as Time Delay Neural Network (TDNN) has also shown that a reliable method for predicting the future evolution of the time series [5]. Other are his key questions radiation forecast weather and clouds. Therefore, it is very difficult to do a straight build forecast forecast models in one 24-hour cycle. One possible solution is the time series based model the exact radiation can evaluate the ground. Therefore, the solar radiation sequence is treated as a time series and mathematical models fit the random process underlying to predict the next values [6].

In the research community, Autoregressive Moving Average (ARMA) methods are widely used and popular time series models compared to other models (as mentioned above) [7]. The ARMA model is able to extract many regions, useful statistical properties and can easily take on the well-known box-Jenkins method [8]. In addition, these models are very flexible; therefore they can be used in various types of time series with different orders. Finally, it offers regular pervasive in individual phases (identification, estimation and diagnostic check) for a suitable model. In the ARMA model of one of the greatest difficulties need is the enormous amount of data. Moreover, this method requires an excessive agreement of knowledge and although it creates often acceptable results the individual results are determined on the level of research knowledge [9]. Recently, ARMA models based several studies have done in many areas, the current flow, supply chain management, business, earthquakes, land use, sales, products, includes transportation and weather forecasts. However, there is a scarcity of the use of this model in energy applications.

In this paper ARMA method is used to use the solar radiation that this connection be recursive least squares-based parameter estimation techniques two types of processes (FF KF) .In both methods predict cast together. In the first step aroused the method to estimate the model parameters and the best results with the measured data. The next step is used to define bias and mean error of the model that leads to accurate predictions.

## II. METHODOLOGY AND DATA

### A. Data Collection

The average value of extraterrestrial radiance is considered nearly about to 1360 Wm<sup>-2</sup>. It consists of two parameters: Diffuse Horizontal irradiance (DHI) and direct Normal Irradiance (DNI). These are used to calculate Global Horizontal Irradiance (GHI) as:

Manuscript received September 28, 2014; revised July 19, 2015.

Yashwant Kashyap and Satyanarayan Patel are with the School of Engineering, Indian Institute of Technology Mandi, 175005 India (tel.: 91-9805-911710; e-mail: yashwant.kashyap@gmail.com, satyou1990@gmail.com).

Ankit Bansal is with the Department of Mechanical and Industrial Engineering, 233 MIED, IIT Roorkee, 247667 India (e-mail: abansfme@iitr.ac.in).

Anil K. Sao is with the School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, India (e-mail: anil@iitmandi.ac.in).

$$\text{GHI} = \text{DHI} + \text{DNI} \times \cos(\Phi) \quad (1)$$

where  $\Phi$  of the solar zenith angle [10]. The GHI values are on the website of the Indian Ministry of Renewable Energy (MNRE) is available. Fig. 1 shows GHI time series database capable of IIT Mandi Himachal Pradesh India (longitude, latitude and altitude are averaging 76.75 E, 31.75 N and 1130.3 meters, respectively). This time zone is 5.5 at this location data and has 5000 hours of the year 2008, the time value in the form of 10 km (0.1°×0.1° resolution) spatial points were recorded collected [11].

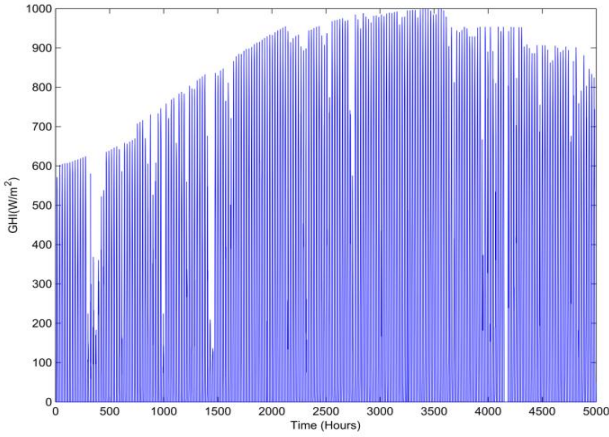


Fig. 1. GHI Time Series data collected for 5000 hours from 2008 as plotted.

### B. Time Series Analysis

The Autoregressive Moving Average (ARMA) model is used in various fields to perform temporal analysis after being introduced by Box-Jenkins [8]. In the ARMA, generally stochastic linear differential equations are applied to auto-correlation time series model. This data is used to understand and forecast behavior of the system. An ARMA model can be expressed as (generalized form):

$$f_1(S^t)f_2(S)(1-S^t)^N(1-S)^dO_M = f_3(S^t)f_4(S)B_M \quad (2)$$

where  $O_M$ ,  $B_M$ ,  $S$ ,  $d$  and  $N$  are discrete time observation process, random noise, and shift operator, seasonal and non-seasonal order respectively. The functions  $f_1(S)$  and  $f_2(S)$  are polynomials of finite order  $p$  and  $q$ . However,  $f_3(S)$  and  $f_4(S)$  are polynomials of finite order of  $R$  and  $r$ , respectively.

### C. Kalman Filter

The motion of extracted shadow features is tracked by a dynamic model and assimilated data extracted from successive frames with the help of KF. The tracking model, basically a state space model, forecasts the movement of the shadow at time  $t+1$  and the KF updates the forecast by assimilating features of the shadow extracted from the latest data. The equations of motion of shadow are written as:

$$x_{t+1} = x_t + v_t \times \Delta t \quad (3)$$

$$v_{t+1} = v_t \quad (4)$$

where  $x_t$  and  $x_{t+1}$  are positions and  $v_t$  and  $v_{t+1}$  are velocities of

the cloud at time  $t$  and  $t+1$ , respectively. There are four state variable used in this formulation namely:  $x$ ,  $y$ ,  $v_x$  and  $v_y$ . The state variables estimated in the previous time step are used as input for the current time step. In the notation of KF, the state-space model and the observation model are written as

$$\underline{x}_t = A \cdot \underline{x}_{t-1} + w_t \quad (5)$$

and

$$\underline{z}_t = H \cdot \underline{x}_t + v_t \quad (6)$$

where  $\underline{x}$  is the state vector and  $A$  is the state transition matrix,  $w_t$  and  $v_t$  are Gaussian white noise with zero mean:

$$p(w) \sim N(0, Q), p(v) \sim N(0, R) \quad (7)$$

where  $Q$  and  $R$  are covariance matrices of  $w_t$  and  $v_t$ , respectively.

The state vector  $\underline{x}_t$  of the present time  $t$  is estimated from the earlier predicted values and the latest observed value of  $\underline{z}_t$ . The KF works in two steps, the forecast step and the rectification (data assimilation) step. In the forecast step, the present state is projected using a dynamic model, to find a priori assessment or forecast of the state,  $\underline{x}_t^-$  as.

$$\underline{x}_t^- = A \cdot \underline{x}_{t-1} + w_t \quad (8)$$

In the rectification or the data assimilation step, the newly observed or measured values of the variables are taken and variances are determined to get a better posterior prediction  $\underline{x}_t^+$ . The  $\underline{x}_t^+$  is expressed as

$$\underline{x}_t^+ = \underline{x}_t^- + K_t(\underline{z}_t - H\underline{x}_t^-) \quad (9)$$

where  $K_t$  is the gain of the KF and is defined as

$$K_t = P_t^- H^T (H P_t^- H^T + R)^{-1} = \frac{P_t^- H^T}{H_t P_t^- H_t^T + R_t} \quad (10)$$

In Eq. (14), the  $P_t^-$  is the priori predicted error covariance and it is defined as

$$P_t^- = E[e_t^- e_t^{T}] \quad (11)$$

This on simplification gives

$$P_t^- = A \cdot P_{t-1} \cdot A^T + Q_{t-1} \quad (12)$$

where  $e_t^- = x_t - \hat{x}_t^-$  is the priori predicted error [12]. Furthermore, the posteriori predicted error covariance  $P_t^+$  is defined as

$$P_t^+ = E[e_t^+ e_t^{T}] \quad (13)$$

where  $e_t^+ = x_t - x_t^+$  is the posteriori predicted error and the expression for  $P_t^+$  after simplification reduces to

$$P_t^+ = (1 - K_t H) P_t^- \quad (14)$$

From Eq. (14), the observation error variance  $R$  and Kalman gain  $K_t$  are in opposite relation. The lesser the value of  $R$ , the more weightage is given to the measurements over the forecasted values from the state space model. On the other side, for large measurement error covariance, the measurements are given lesser weightage.

#### D. Forgetting Factor (FF)

The prediction model estimates multiple step recursive least square (RLS) with FF, which is described as:

$$\hat{\theta}_{(t)} = \hat{\theta}_{(t-1)} + G_t (\theta_{(t)} - \hat{\theta}_t) \quad (15)$$

where  $\hat{\theta}_{(t)}$  is estimated parameter at time  $t$ .  $\theta_{(t)}$  is measured data at time  $t$  and  $\hat{\theta}_t$  is forecasted data of  $\theta_{(t)}$  based on measured data up to  $t-1$ .  $G_t$  is gain of error of  $(\theta_{(t)} - \hat{\theta}_t)$  estimator, this is used to update the parameter estimation. This procedure tries to minimize the error term  $(\theta_{(t)} - \hat{\theta}_t)$ . Further it is estimated by:

$$\hat{\theta}_t = \varphi_t^T \hat{\theta}_{(t-1)} \quad (16)$$

where  $\varphi_t$  is gradient of estimated method result  $\hat{\theta}_{t/\theta}$  with respect to  $\varphi_t$ .

$$G_t = \delta_t \varphi_t \quad (17)$$

To find  $\delta_t$  called forgetting factor (FF), the subsequent purpose is to minimize FF at time  $t$ .

$$\delta_{(t)} = P_{(t)} = \frac{P_{t-1}}{\alpha + \varphi_t^T P_{t-1} \varphi_t} \quad (18)$$

$$P_t = \frac{1}{\varepsilon} \left( P_{t-1} - \frac{P_{t-1} \varphi_t \varphi_t^T P_{t-1}}{\varepsilon + \varphi_t^T P_{t-1} \varphi_t} \right) \quad (19)$$

$$\delta_t = \min \sum_{l=1}^t \varepsilon^{t-l} e_l^2 \quad (20)$$

The recursive process results differ because of several methods of selecting  $\delta_t$  and estimating  $\varphi_t$ . Hence any time regression parameter can be found out with respect to original values, which are found out using FF and RLS method. The main weightage of FF is on current data compared to previous data for parameter estimation [13].

### III. RESULTS AND DISCUSSION

In order to assess the capability of the model data are

divided into two sets, one set of inputs variables belongs to training processes and second data set for the validation. These validation sets also known as testing set, in the testing processes comprises of different model results. Finally, a model has been selected based on lowest forecasting error. The error can be estimated by number of ways such as RMSE, MAE, MBE and  $R^2$ . These methods of error estimation can be expressed as:

$$MAE = \frac{1}{m} \sum_{t=1}^m |P_t - \hat{P}_t| \quad (21)$$

$$MBE = \frac{1}{m} \sum_{t=1}^m P_t - \hat{P}_t \quad (22)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{t=1}^m (P_t - \hat{P}_t)^2} \quad (23)$$

$$R^2 = 1 - \frac{\sum_{t=1}^m (P_t - \hat{P}_t)^2}{\sum_{t=1}^m (P_t - \bar{P})^2} \quad (24)$$

where,  $P_t$  is measured at forecasted horizon  $t$ ,  $\hat{P}_t$  is forecasted value, and  $m$  is the total number of test samples. This time series data models need to follow a standard procedure for forecasting, using three basic processes identification assessment and validation checks [14].

In the ARMA model is the most important step is the appointment (d). The series is made stationary by differencing; however, a number of differentiating steps are needed to equal his order. Therefore, these time series plots are (normalized data) used to determine the non-stationary elements. This non-stationary elements by differentiation of the first, second or third order, which depends on the auto-correlation of the positive differentiation factor, as shown in Fig. 2 and Fig. 3 are removed. However, if new data shows positive autocorrelation then it requires first-order difference. The factor is zero or negative, after first difference he is on stationary series and there are no other requirements of differentiation. For an optimal series cases shows minimum standard deviation and corresponding optimal sequence.

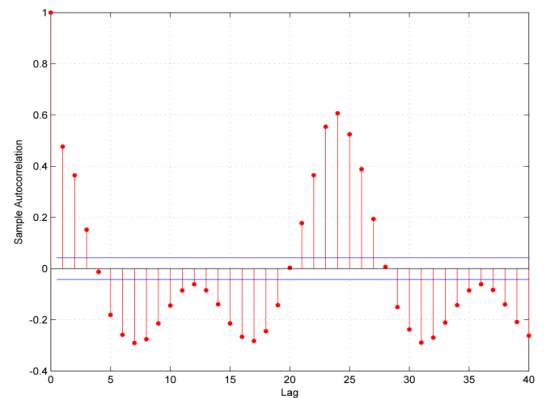


Fig. 2. Autocorrelation plot at first differentiation.

In order to further study, the next step is parameters estimation. In general 20-30% data is used for testing purpose during model estimation. For this purpose machine learning

tools in MATLAB is used. This machine learning tool can be minimizing the error for both FF and KF models. Optimum parameters are obtained by forming a number of trials shown in Table I. It includes minimum root mean square error (RMSE) and forecast vs. measure data variation at different sets of parameter for both the parameter estimation models.

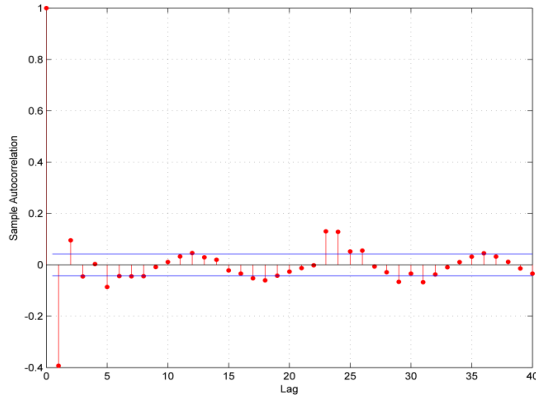


Fig. 3. Autocorrelation after the second differentiation.

Further study, Fig. 4 shows the residual plot of training data for FF and KF models under the time domain analysis. The residuals are used to find point to point difference of predicted and measured data. This difference is used to compare regression model accuracy and validation. In the validation processes, residual plot shows either increasing or decreasing trend. Hence, regression models are considered better for KF based on constant error variance, as shown in Fig. 4, the KF model has residuals not correlated with past inputs. This validation of correlation indicates that the model not fit properly between output and input during these parts of data. The goodness-of-fit for both the models are 0.9641 and 0.9942 for FF and KF respectively.

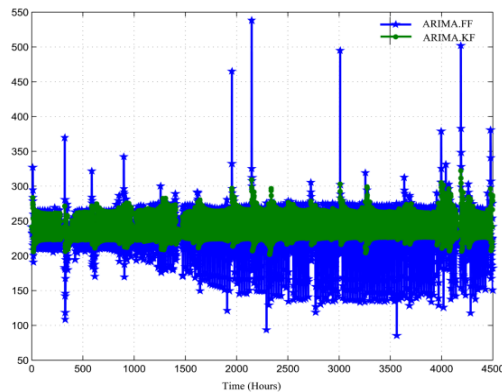


Fig. 4. Residual plot of training data for both models.

In the first step data are processed through a normalization of the time series, while hourly available time series are considered for probability plot. Probability graph with percent confidence limits are shown in Fig. 5. It shows accuracy curve fit for the time series data. Finally, data around the fitted curve line shows accuracy of the model. Fig. 5 shows the solar radiation property, wherein the x-axis and y-axis show the actual and expected data. Each of the two models FF and KF are safe limit of the curve fit measured to 95%. Fig. 5 shows that the KF model has better convergence compared to FF at higher point whereas the FF is in reverse order.

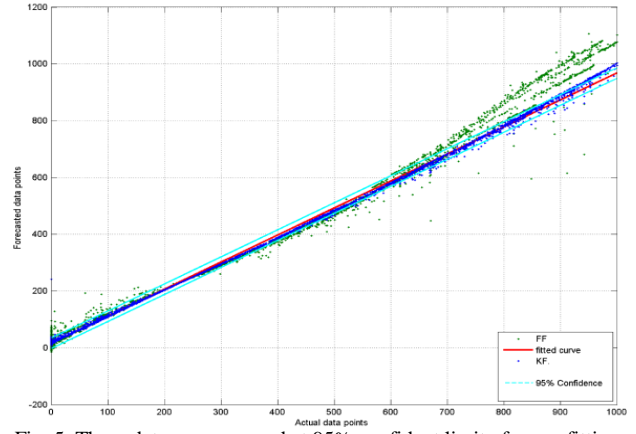


Fig. 5. These data are measured at 95% confident limit of curve fitting.

The smallest RMSE shows the method of estimation accuracy at the local level or small number of data set, while MAE shows global accuracy. This error is estimated for 500 hours in one step prediction. It includes both clear and cloudy day actual value and the forecast results. The FF model percent test results of "MAE", "MSE" and "RMSE" are 13.2426, 3.5512 and 18.8445, respectively. Likewise for KF model are 3.9922, 0.5767 and 7.5739. Therefore it can be concluded that the performance of the KF model better than FF model in the two cases, the training and testing. In the case of significant difference in model RMSE observed when an error of a model is 30% or higher. However, the difference of 3 to 4 percent then probably there is no significant difference. In this case it is about 11%, so the model performance would be somewhat important.

TABLE I: TRAINING AND TESTING ERROR

	Training (FF)	Training (KF)	Testing (FF)	Testing (KF)
MAE	4.83	2.31	13.24	3.99
MSE	1.03	0.16	3.55	0.57
RMSE	10.15	4.09	18.84	7.59
R <sup>2</sup>	0.98	0.99	0.96	0.99

In the evaluation process, it begins with the simplest model of the ARMA (1, 0, 0) for the data playback. Approximately 4500 hours previous data are used for the start of RLS with FF as a parameter estimator. The solar radiation prediction results for the year 2008 will be used for 500 hours. In this forecast FF than 0.8 is considered to 0.99. As in Fig. 6 shown, the smaller value of FF is less suitable model and prediction accuracy. A number of studies are underway in various FF values, the conclusion that higher range values show better result. Fig. 6 clearly shows that the predicted value varies (top and bottom points of land) during the morning, noon and night. Even given that FF model shows more difference when compared with actual values about 50-100 Wat clock with clear sky days. Similarly, predicted values for morning and evening are again less than the actual values. More specifically, however, follow KF real action during the entire period. Perform especially KF models and (marked with red circle) in the gloomy days, as shown in interval 4748-4800 and 4848-4900 hours.

The RMSE increases significantly hot season to cold season, that the difference between measured and predicted deviation is a result of weather instability during the cold season [15]. The estimated solar radiation error can be up to



70%, to achieve a result of the varying signal. However, may be affected radiation 30% variation of the ambient temperature, which is considered to be stable. This error can reach 30%, because it quantifies the climate change between the day and the day before. The error can be significantly reduced (61%) for short-term prediction approach based models [16]. In this connection, the ARMA models weather parameters for a day or between one day before and on the same day using the Kalman filter to be measured. This KF based predicted data show the effectiveness of this procedure in comparison to other methods as described by many authors [17].

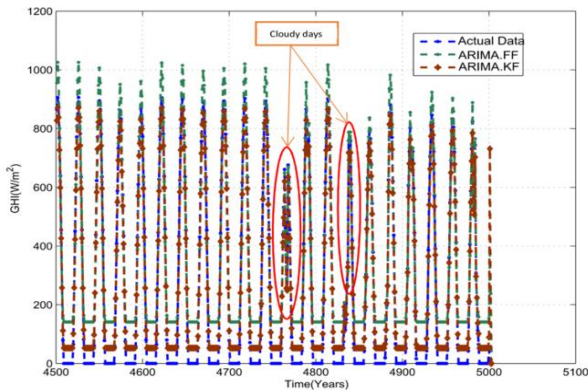


Fig. 6. The solar radiation forecasting results is used for year 2008 of 500 hours.

#### IV. CONCLUSIONS

In this work, an attempt was made to investigate the stochastic model of the solar radiation prediction. In this regard, collected Indian Sunny satellite global horizontal irradiance (GHI) data are used. Furthermore, two parameters evaluation techniques forgetting factor and Kalman filter are used based on the recursive least squares. These methods will be used in relation to estimate parameters of current estimation error are compared to the mean absolute error (MAE), the mean square error (RMSE), mean square error (MSE) estimator and  $R^2$  methods. It can be said that the parameters evaluation method has a great impact on reception accurately predicted values. Therefore, it can be summarized that KF is better when it is used in irradiation forecast.

#### REFERENCES

- [1] T. Ramachandra, R. Jain, and G. Krishnadas, "Hotspots of solar potential in India," *Renewable and Sustainable Energy Reviews*, vol. 15, pp. 3178-3186, 2011.
- [2] A. Kumar, K. Kumar, N. Kaushik, S. Sharma, and S. Mishra, "Renewable energy in India: Current status and future potentials," *Renewable and Sustainable Energy Reviews*, vol. 14, pp. 2434-2442, 2010.
- [3] V. Ramanathan, C. Chung, D. Kim, T. Bettge, L. Buja, J. Kiehl *et al.*, "Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle," in *Proc. the National Academy of Sciences of the United States of America*, 2005, vol. 102, pp. 5326-5333.
- [4] A. G. Tsikalakis and N. D. Hatziairgiou, "Centralized control for optimizing microgrids operation," in *Proc. 2011 IEEE Power and Energy Society General Meeting*, 2011, pp. 1-8.
- [5] R. Perez *et al.*, "Forecasting solar radiation — Preliminary evaluation of an approach based upon the national forecast database," *Solar Energy*, vol. 81, pp. 809-812, 2007.
- [6] A. C. Kyritsis, E. Tatakis, and N. Papanikolaou, "Optimum design of the current-source flyback inverter for decentralized grid-connected

photovoltaic systems," *IEEE Transactions on Energy Conversion*, vol. 23, pp. 281-293, 2008.

- [7] E. McKenzie, "General exponential smoothing and the equivalent ARMA process," *Journal of Forecasting*, vol. 3, pp. 333-344, 1984.
- [8] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng, "Model identification of ARIMA family using genetic algorithms," *Applied Mathematics and Computation*, vol. 164, pp. 885-912, 2005.
- [9] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, vol. 85, pp. 808-817, 2011.
- [10] R. H. Inman, H. T. Pedro, and C. F. Coimbra, "Solar forecasting methods for renewable energy integration," *Progress in Energy and Combustion Science*, vol. 39, pp. 535-576, 2013.
- [11] Nrel, *India Solar Resource Data: Hourly*, 2013.
- [12] X. Li, K. Wang, W. Wang, and Y. Li, "A multiple object tracking method using Kalman filter," in *Proc. IEEE International Conference on Information and Automation*, 2010, pp. 1862-1866.
- [13] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, pp. 1772-1783, 2009.
- [14] E. Lorenz *et al.*, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *J. of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, pp. 2-10, 2009.
- [15] C. Voyant, M. Muselli, C. Paoli, and M.-L. Nivet, "Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation," *Energy*, vol. 39, pp. 341-355, 2012.
- [16] C. Voyant *et al.*, "Multi-horizon solar radiation forecasting for Mediterranean locations using time series models," *Renewable and Sustainable Energy Reviews*, vol. 28, pp. 44-52, 2013.
- [17] B. P. Gibbs, *Advanced Kalman Filtering, Least-Squares and Modeling: A Practical Handbook*, John Wiley & Sons, 2011.



**Yashwant Kashyap** received the bachelor degree of engineering in electrical engineering from Guru Ghasidas University Bilaspur, India in 2003 and the M.Tech. degree in energy from Indian Institute of Technology Delhi, India in 2007. From 2012 onwards he has been enrolled as a PhD research scholar at Indian Institute of Technology Mandi, India. His present research interests include analysis on solar radiation forecasting, using artificial intelligent

technique.



**Ankit Bansal** was born in India on November 27, 1982. He received his bachelors of technology degree in mechanical engineering from the Institute of Technology, Banaras Hindu University, India in 2005, and PhD degree in mechanical engineering from the Pennsylvania State University, USA in 2011.

He served as an assistant professor of mechanical engineering at Indian Institute of Technology Mandi from 2012 to 2014. He is currently working as an assistant professor of mechanical engineering at Indian Institute of Technology Roorkee. His research interests include radiative heat transfer, solar energy, climate modeling, combustion and computational fluid dynamics.



**Anil Kumar Sao** has done his B.E. degree in electronics from GEC, Raipur in 1999. He did his MS and PhD degrees from CSE, IIT Madras in 2004 and 2010, respectively. He is working as an assistant professor in SCEE, IIT Mandi since 2010 to till date. His research interests are pattern recognition in the field of image and speech signals.



**Satyanarayan Patel** received a bachelor degree in mechanical engineering with honors from Rajasthan Technical University Kota, India in 2010 and the M.Tech. degree in energy engineering from Malaviya National Institute of Technology Jaipur, India in 2012. From 2012 onwards he has been enrolled as a PhD research scholar at Indian Institute of Technology Mandi, India. His present research interests include ferroelectric materials for energy storage and solid

state refrigeration.