IJDI-ERET

INTERNATIONAL JOURNAL OF DARSHAN INSTITUTE ON ENGINEERING RESEARCH AND EMERGING TECHNOLOGIES
Vol. 9, No. 2, 2020
www.ijdieret.in

(Research Article)

Solar Based Electrical Power Generation Forecasting Using Time Series Models

Rikinkumar B. Patel[1*], Mihir R. Patel[2], Dr. Nilaykumar A. Patel[3]

[1]PG Graduate Dept. of Electrical Engineering, CHARUSAT UNIVERSITY, Changa, Gujarat, India*
[2]Assistant Professor, Dept. of Electrical Engineering, CSPIT-CHARUSAT UNIVERSITY, Changa, Gujarat, India
[3]Associate Professor and Head, Dept. of Electrical Engineering, CSPIT-CHARUSAT UNIVERSITY, Changa, Gujarat, India

Abstract

The rapid growth of solar power (plant) installation in various countries can be beneficial to environment, but at the same time the interconnected solar panels to power grids can negatively impact the stability of power system. So, to maintain the stability of power system the accurate solar power generation forecasting is becomes essential. In this paper, we have applied three well-known time series models called Auto-regressive (AR), Auto-regressive Integrated Moving Average (ARIMA) and Seasonal Auto-regressive Integrated Moving Average (SARIMA) model for forecasting the solar based electrical power generation. Using this time series models the two different case studies are carried out i.e. the first case study will show the analysis for Elia dataset and the second case study will show the analysis for Sheffield solar dataset. In both the case studies one day ahead prediction is made and also, the predicted values are compared with real time measured values to test the effectiveness of applied time series model. The Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) techniques are used to calculate the accuracy of applied time series models.

Keywords: Solar power generation forecasting, Time Series, ARIMA, SARIMA, Solar generation historical dataset.

1. Introduction

Encouragement to reduce the pollution in the environment is increasing day by day. In an electrical power system, conventional plants play a significant role in the emission of various gases in the environment. Those emitted gases will cause the environment negatively. There is another way of electricity generation using renewable energy or non-conventional plants. Renewable energy is environmentally friendly and by adding more and more non-conventional plants one can help to reduce the pollution in a living environment. In renewable energy, solar power plays a major role in the generation of electricity after hydroelectric power.

Integrating a large number of solar power panels into the power grid will negatively impact the power system stability or make the system unstable. It impacts the power system because the nature of solar energy is uncertain (variation) or cannot give the same amount of solar energy for a whole day at a particular location. Solar energy is weather dependent and there are many factors that will affect the electricity generation from solar energy. Some of the main factors are irradiance insolation, ambient temperature, cell temperature, and internal and external efficiency of solar panels. The solar panel converts sun lights to in to the electricity, but due to uncertainty (variation) in sunlight, the generation of electricity will vary [1]. Because of this variation in a generation, the supply will fluctuate at the point of connection to the power system. Hence, this fluctuation results in making a healthy power system to unstable. A power system operator can maintain stability by knowing how much the solar panels will supply the electricity at the point of connection to a power system. Two foremost reasons to forecast solar power generation: (a) to maintain the stability of the power system and (b) to make effective use of generated power from solar energy. For meeting this foremost reason, the optimization of scheduled generators of conventional plants and storage batteries is carried out based on the results of forecasted solar power generation output and forecasted load demand [2]. There are two types of forecasting methods: (i) physical-based models and (ii) statistical models. In this paper, the statistical model's method is used to predict solar power generation. The statistical method directly predicts the solar power generation by using past output solar power generation data. Based on a time horizon the solar power is predicted from several hours ahead to several days ahead [3].

[*]Corresponding Author: e-mail: mihiurrpatel.ee@charusat.ac.in
Mobile- +91-8155953969
ISSN 2320-7590

The time series and machine learning are the most applied methods for doing solar power forecasting. In [4], the well-known statistical model called autoregressive integrated moving average has been applied to predict the daily total solar energy generation from a 10kW solar array. The dataset having both seasonality and non-stationarity. Using a differencing term from the model the dataset was transformed into stationarity. Akaike information criterion and residual sum of squares technique are used to select the optimal order of the ARIMA model. In [5], the very short-term PV generation prediction for multi-step ahead (20 minutes' resolution) was carried out using well define seasonal autoregressive integrated moving average model and persistence method. Using ACF and PACF the order of AR and MA terms has been selected. In [6], the four hours ahead forecasting was carried out using methods called SARIMA, SARIMAX, modified SARIMA, and ANN (Artificial Neural Network) for grid-connected PV plant in Greece. The prediction results were evaluated using Normalized Root Mean Square Error (NRMSE). In [7], the author has developed the new hybridized model called SARIMA-SVM which means that combinations of two well-known methods. This hybridized model has been used for forecasting the small scale 20 kWp grid-connected photovoltaic plant generation. The prediction results illustrate that the SARIMA-SVM has worked better than SARIMA and SVM models. In [8], the different methods called ARIMA, Persistent model, kNNs, ANNs, and ANNs optimized by Genetic Algorithms are applied to forecast the 1 MWp solar plant generation power. The prediction results were evaluated using Mean Absolute Error (MAE), Mean Bias Error (MBE), Root Mean Square Error, and NRMSE. In [9-11], the importance of forecasting in power systems and various time series and machine learning methods are explained which can be used for doing prediction in the power system.

The contribution of this paper is to implement the three well-defined time series models called AR, ARIMA, and SARIMA for the role of solar-based electrical power generation forecasting. For doing short-term solar power generation forecasting the two publicly available real-time solar power generation datasets are used which are (i) Elia real-time solar power generation dataset and (ii) Sheffield Solar real-time solar power generation dataset. The rest of the paper follows as in section 2 the two popular well define time series models have explained and also, this section will give understanding about AR, ARIMA, and SARIMA model. Section 3 includes the implementation of applied time series models and their achieved prediction results from two different case studies. Section 4 will summarize the paper in small.

2. Methodology

The two explained models in this section will give the understanding about different time series models or terms called AR, MA, ARMA, ARIMA, and SARIMA.

2.1 ARIMA Model: The ARIMA stands for Autoregressive Integrated Moving Average and as per the name suggest it is a combine of three terms called AR (Autoregressive), I (Integrated or differencing), and MA (Moving Average). Also, this model is denoted as ARIMA (p, d, q) and here the value of p, d, and q will be non-integer. The AR (p) term define as the output data depends linearly on its own past data and on a residual value. The AR term will predict future values using a linear combination of previous data of the dataset. The equation (1) show the mathematical representation of AR (p) term or model. From equation (1), $\mathrm{Y_t}$ describe that we want to forecast the value of 'Y' at time period 't', $\emptyset_1, \emptyset_2, \dots \emptyset_\mathrm{p}$ are the lag co-efficient up to order p, $\mathrm{y_{t-1}}, \mathrm{y_{t-2}}, \dots \mathrm{y_{t-p}}$ are the observed historical data, ε_t is residual value or white noise, and c is constant. The generalized equation of AR (p) term is shown in equation (2).

$$\mathrm{Y_t} = \mathrm{c} + \emptyset_1 * \mathrm{y_{t-1}} + \emptyset_2 * \mathrm{y_{t-2}} + \cdots + \emptyset_\mathrm{p} * \mathrm{y_{t-p}} + \varepsilon_\mathrm{t} \quad (1)$$

$$Y_t = c + \left(\sum_{i=1}^{p} \emptyset_i * y_{t-i}\right) \quad (2)$$

The MA (q) term define as the output value depends linearly on the present and different previous values of a residual term or white noise. The equation (3) show the mathematical representation of MA (q) term or model. From equation (3), $\mathrm{Z_t}$ describe that we want to forecast the value of 'Z' at time period 't', $\theta_1, \theta_2, \dots \theta_q$ are the parameters of the MA (q) model, $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots \varepsilon_{t-q}$ are the white noise or residual term, and μ is mean of the series. Majority of times the mean part in equation (3) is neglected. The generalized equation of MA (q) term is shown in equation (4).

$$\mathrm{Z_t} = \mu + \varepsilon_t + \theta_1 * \varepsilon_{t-1} + \theta_2 * \varepsilon_{t-2} + \cdots + \theta_q * \varepsilon_{t-q} \quad (3)$$

$$\mathrm{Z_t} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \quad (4)$$

The differencing term is making ARIMA model more valuable. The function of differencing is to make non-stationary time series into stationary time series and also, some non-stationary portion in time series into stationary. Because of differencing term, the ARIMA can deal with the non-stationary time series. By combining AR (p), MA (q), and differencing (d) term the final or generalized equation of ARIMA (p, d, q) model is written as shown in equation (5). From equation (5), the L is a lag operator. The model is explained by referring various resources [12-15].

$$\left(1 - \sum_{i=1}^{p} \emptyset_i * L^i\right) * (1 - L)^d * \, Y_t = \left(1 + \sum_{j=1}^{q} \theta_j * L^j\right) * \varepsilon_t \quad (5)$$

2.2 SARIMA Model: The time series contains three components which are trend, seasonal, and irregular components. The trend component will show long term movement in time series. The seasonal will show short term movement or repeating cycle in time series and the irregular component shows the random movement. Out of this components the ARIMA model has a problem to deal with seasonal data i.e. data which are repeating over time or time series with a repeating cycle. The Box and Jenkins time series model called Seasonal Autoregressive Integrated Moving Average (SARIMA) or Seasonal ARIMA model has ability to deal with seasonal data (seasonality) of univariate time series. The SARIMA is an extension of ARIMA model. The SARIMA has two parts i.e. non seasonal and seasonal part. In existing ARIMA model, three new hyper parameters are added to deal with seasonal part i.e. the added seasonal parameters are AR (P), Differencing (D), and MA (Q). This seasonal part in the model is very much similar to non-seasonal part in model and non-seasonal parameters in models are same as the ARIMA model. The generalized equation of SARIMA (p, d, q) * (P, D, Q) model is represented in equation (6).

$$\Phi_P(L^s) * \phi_p(L) * (1-L)^d * (1-L^s)^D * Y_t = \Theta_Q(L^s) * \theta_q(L) * \varepsilon_t \tag{6}$$

In equation (6), L is lag operator, the capitalized letter describes seasonal term and small letter describes non-seasonal term. In short the equation of SARIMA model is similar to ARIMA model, the difference is three more term is added to deal with seasonal data. The model is explained by referring various resources [12-15].

3. Implementation of Applied Time Series Models and Achieved Results

The two publicly available real time observed solar based power generation dataset of Belgium region which is uploaded by Elia system operator (Electric Service Company) [16] and Sheffield Solar (A Photovoltaic Industry) dataset [17] are used in this paper for testing the effectiveness of time series models. The analysis of time series models (prediction result and its mathematical evaluation) for two different cases i.e. The first case study in section 3.1 indicate the analysis for Elia dataset and the second case study in section 3.2 indicate the analysis for Sheffield solar dataset are as followed:

3.1 Elia Dataset Input to Time Series Models: For doing short term (one day ahead) solar power generation forecasting, the past one-month full day (both day and night measured values) historical data of Elia are fitted to all three applied time series models i.e. AR, ARIMA and SARIMA. The length of historical dataset is from 01-05-2020 to 31-05-2020 as shown in figure 1 [16]. The unit of observed values is in MW and time span is fifteen minutes i.e. each value is observed at every fifteen minutes. The observed data between 01-05-2020 and 30-05-2020 are fitted into applied time series models and using this fitted data the one day ahead prediction is made i.e. for 31-05-2020. The measured solar power generation values in MW are converted into GW to minimize the mathematical burden while fitting the data into models and also to improve the accuracy (by using input dataset of GW unit gives better prediction than using input dataset of MW unit). Because the fitted values are in GW the prediction results will be in GW unit. Further, the prediction values are converted back to MW unit for the evaluation of the models and comparison between actual values and forecasted values.

The comparison between real time (actual) solar power generation values by Elia and predicted values by applied time series models is shown in figure 2. Also, the mathematical evaluation between real time solar power generation values by Elia and forecasted values by applied time series models is shown in table 1. From figure 2 and table 1 it can be seen that AR (95) model is working better than ARIMA and SARIMA model with 147.27 MW measured error and 7.25% MAPE. The choice of order ‘p’ in AR model is made manually from p = 1 to 100. Out of this values of order ‘p’ the optimal order come out was p = 95 i.e. at p = 95 the AR model is giving better results than all other values of ‘p’. Similarly, the same concept of AR model is used for choosing the orders of ARIMA and SARIMA model. The ARIMA (10, 0, 4) model is not working accurately and giving poor mathematical results with 1016.15 MW measured error and 57.91% MAPE. But by adding some seasonality with the help of SARIMA (8, 0, 2) x (3, 0, 1, 48) model the prediction result become better than ARIMA model with 152.90 MW measured error and 12.49% MAPE.

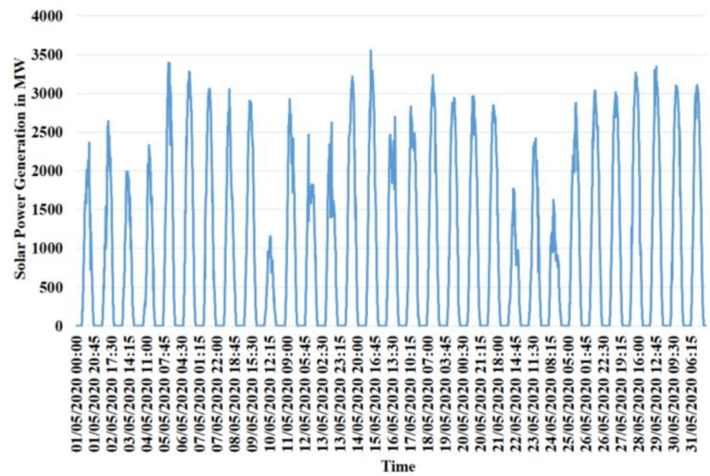

Figure 1. Historical solar power generation dataset of Elia from 01-05-2020 to 31-05-2020

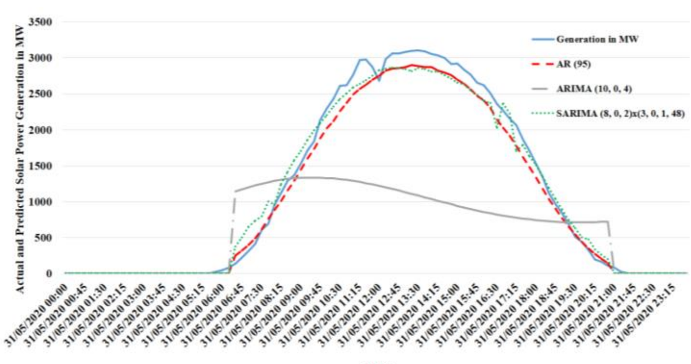

Figure 2. Comparison of AR, ARIMA and SARIMA prediction results with Elia observed solar power generation data in MW

Table 1. Time Series Models Evaluation for Elia Using RMSE and MAPE

Models	RMSE in MW	RMSE in GW	MAPE in %
AR (95)	147.27	0.147	7.25
ARIMA (10, 0, 4)	1016.15	1.016	57.91
SARIMA (8, 0, 2) x (3, 0, 1, 48)	152.90	0.152	12.49

3.2 Sheffield Solar Dataset Input to Time Series Models: In this case also we have applied AR, ARIMA and SARIMA model for doing one day ahead solar based power generation prediction. In this case the forecasting is carried out for Sheffield Solar dataset [17]. The length of historical dataset is from 01-03-2020 to 31-05-2020 as shown in figure 3. The unit of observed values are in MW and time span is thirty minutes i.e. each value is measured at every thirty minutes. This measured values in MW are converted into GW unit by dividing each values by 1000. Now, the values in GW unit between 01-03-2020 and 30-05-2020 are fitted into time series models for doing one day ahead i.e. for 31-05-2020 solar based power generation forecasting. Moreover, the fitted values are in GW so the prediction results will be in GW unit. Hence, the prediction results are converted back to MW unit for the evaluation of the models and the comparison between real values and predicted values.

The comparison between real time solar power generation values by Sheffield Solar and predicted values by three applied time series models is shown in figure 4. Also, the mathematical evaluation between real time solar power generation values by Sheffield Solar and forecasted values by three applied time series models is shown in table 2. From figure 4 and table 2 it can be seen that AR (45) model is working better than ARIMA and SARIMA model with 208.54 MW measured error and 8.56% MAPE. The choice of order 'p' in AR model is made manually from p = 1 to 96. Out of this values of order 'p' the optimal order come out was p = 45 i.e. at p = 45 the AR model is giving better prediction result than all other values of order 'p'. Likewise, the same concept of AR model is used for choosing the orders of ARIMA (p, d, q) and SARIMA (p, d, q) x (P, D, Q, m) model. As shown in figure 4 the ARIMA (4, 0, 1) model is working inaccurately and giving poor mathematical results with 2375.50 MW measured error and 94.06% MAPE. Same as in case A, by adding some seasonality with the help of SARIMA (6, 0, 1) x (4, 0, 1, 24) model the prediction result become better than ARIMA model with 314.60 MW measured error and 12.90% MAPE. It can be noted that the value of MAPE will not change irrespective to unit of actual and predicted values i.e. the MAPE will remain same whether we use values in MW unit or values in GW unit.

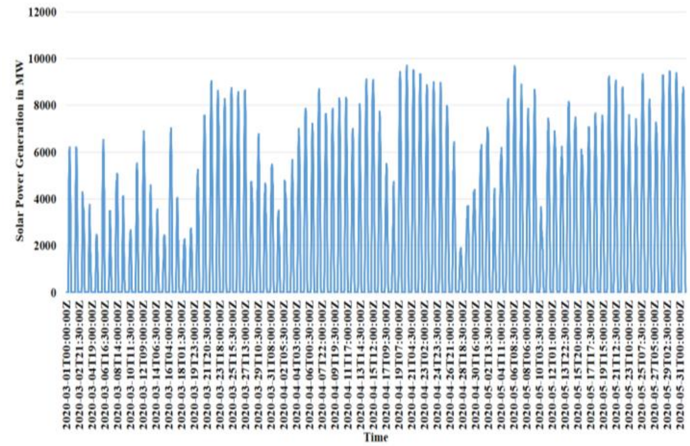

Figure 3. Historical solar power generation dataset of Sheffield Solar from 01-03-2020 to 31-05-2020

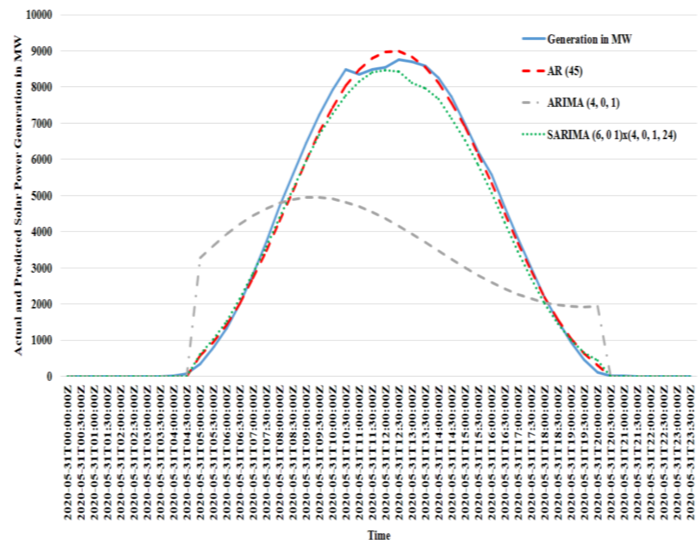

Figure 4. Comparison of AR, ARIMA and SARIMA prediction results with Sheffield Solar measured solar power generation data in MW

Table 2. Time Series Models Evaluation for Sheffield Solar Using RMSE and MAPE

Models	RMSE in MW	RMSE in GW	MAPE in %
AR (45)	208.54	0.208	8.56
ARIMA (4, 0, 1)	2375.50	2.375	94.06
SARIMA (6, 0, 1) x (4, 0, 1, 24)	314.60	0.314	12.90

4. Conclusion

The importance of integrating solar based power generation in power system and basic fundamentals of solar based electrical power generation forecasting are presented in introduction section (i.e. section 1). In this paper we have applied well defined three different time series models for doing solar based power generation forecasting and this models are explained in section 2. The achieved prediction results using time series models and their analysis are shown in section 3. The section 3 is conclude in a way as follow: For Elia dataset (Section 3.1) the model AR (95) is giving better one day-ahead solar power generation prediction result with approximately 92% (7.25 MAPE) accuracy and also, for Sheffield Solar dataset (Section 3.2) the model AR (45) is giving better one day-ahead solar power generation prediction result with approximately 91% (8.56 MAPE) accuracy than rest applied models i.e. ARIMA and SARIMA respectively.

References

1. Nor Azuana Ramli, Mohd Fairuz Abdul Hamid, Nurul Hanis Azhan, and Muhammad Alif As-siddiq Ishak, "Solar Power Generation Prediction by using k-Nearest Neighbor Method," in *AIP Conference Proceedings 2129*, 020116, July-2019.
2. Takeyoshi Kato, "Prediction of photovoltaic power generation output and network operation," in *ScienceDirect*, pp. 77-108, March-2016.
3. Bismark Singh and David Pozo, "A Guide to Solar Power Forecasting using ARMA Models," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1-4, November 2019.
4. Sharif Atique, Subrina Noureen, Vishwajit Roy, Vinitha Subburaj, Stephen Bayne and Joshua Macfie, "Forecasting of total daily solar energy generation using ARIMA: A case study," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, pp. 0114-0119, March-2019.
5. Kushwaha Vishal and Pindoriya Naran, "Very short-term solar PV generation forecast using SARIMA model: A case study," in *7th International Conference on Power Systems (ICPS)*, pp. 430-435, December 2017.
6. Stylianos I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou and A. G. Bakirtzis, "Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting," in *2016 IEEE International Energy Conference (ENERGYCON)*, Leuven, pp. 1-6, July 2016.
7. M. Bouzerdoum, A. Mellit, and A. Massi Pavan, "A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," in *Solar Energy*, Volume 98, pp. 226-235, December 2013.
8. Hugo T.C. Pedro, and Carlos F.M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," in *Solar Energy*, Volume 86, Issue 7, pp. 2017-2028, July 2012.
9. Mihir R. Patel, Ravi Patel, Dharmesh Dabhi, and Jignesh Patel, "Long Term Electrical Load Forecasting considering temperature effect using Multi-Layer Perceptron Neural Network and k-Nearest Neighbor algorithms," in *International Journal of Research in Electronics and Computer Engineering (IJRECE)*, Volume 7, Issue 2, pp. 823-827, April-June 2019.
10. Rikinkumar B. Patel, Mihir R. Patel, and Dr. Nilaykumar A. Patel, "Electrical load forecasting using machine learning methods, RNN and LSTM," in *Journal of Xidian University*, Volume 14, Issue 4, pp. 1376-1386, April 2020.
11. Rikin Patel, Mihir R. Patel, and Ravi V. Patel, "A Review: Introduction and Understanding of Load Forecasting," in *Journal of Applied Science and Computations (JASC)*, Volume 4, Issue 4, pp. 1449-1457, June 2019.
12. George Edward Pelham Box and Gwilym Jenkins, "Time Series Analysis Forecasting and Control," in *Holden-Day, Inc.*, USA, November 1990.
13. Alan Pankratz, "Forecasting with Univariate Box-Jenkins Models: Concepts and Cases," in *John Wiley & Sons, Inc.*, August 1983.
14. J. Lee, "Univariate time series modeling and forecasting (Box-Jenkins Method)", Econ 413, lecture 4.
15. Mihir R. Patel, Rikinkumar B. Patel, and Dr. Nilaykumar A. Patel, "Electrical Energy Demand Forecasting Using Time Series Approach", in *IJAST*, vol. 29, no. 3s, pp. 594 - 604, March 2020.
16. https://www.elia.be/en/grid-data/power-generation/solar-pv-power-generation-data
17. https://www.solar.sheffield.ac.uk/pvlive/

Biographical notes

Rikinkumar B. Patel has received M.Tech. in Electrical Power System from Charotar University of Science and Technology (CHARUSAT), Gujarat, India in 2020 and B.E. in Electrical Engineering from Gujarat Technological University (GTU), Gujarat, India in 2018. At present his area of interest includes Forecasting Techniques in Power System, Renewable Energy, Restructured and Deregulation of Power System.

Mihir R. Patel is working as an Assistant Professor in M & V Patel Department of Electrical Engineering, Chandubhai S. Patel Institute of Technology, CHARUSAT University, Gujarat, India since January, 2016. He has completed his B.E. Electrical from C. S. Patel Institute of Engineering and Technology in 2012 and M.Tech. in Electrical Power System from C. S. Patel Institute of Engineering and Technology in 2014. He was lecturer in Electrical Department at IIET-Dharmaj from June, 2013 to July, 2014 and Assistant professor & GTU coordinator at IIET-Dharmaj from August, 2014 to January, 2016. His research interest area includes Power System Stability, Renewable Energy, and Forecasting. He has published Eight research papers in international journal and one conference research paper. He is Member of Indian Smart Grid Forum working group. His other roles are Department coordinator of EOC (Equal Opportunity Cell), Central council member and Hub Member of HRDC (Human Resources Development Center) at CHARUSAT University.

Dr. Nilaykumar A. Patel has received his Ph.D. from Charotar University of Science and Technology (CHARUSAT), Gujarat, India and M.E. from Sardar Patel University, Vallabh Vidhyanagar, Gujarat, India, in Electrical Power System. He is Associate Professor and Head of Department M & V Patel Department of Electrical Engineering, Chandubhai S. Patel Institute of Technology, CHARUSAT University, Gujarat, India. His area of interest includes Power System, Electric Machines, FACTS, SSR.