

Chapter 7

Time Series Auto-Regressive Integrated Moving Average Model for Renewable Energy Forecasting



Sumanta Pasari and Aditya Shah

Abstract Due to the rapid pace of industrialization and growing demand for energy consumption, forecasting of renewable energy has become an inevitable focus of many recent studies. In this paper, our aim is to develop a univariate auto-regressive integrated moving average (ARIMA) model to forecast daily and monthly wind speed and temperature based on 15 years (2000–2014) of hourly data at Charanka Solar Park, Gujarat. To check the stationarity of time series, Dickey fuller test and rolling statistics plots are employed. Autocorrelation and partial autocorrelation plots are used to determine potential models, whereas Akaike information criterion (AIC) and Bayesian information criterion (BIC) are utilized to establish ARIMA (2, 1, 2) model. After rigorous training, model performance is validated using root means square (RMS) errors. The entire methodology is implemented in python using *pandas* for data exploration, and *stats* and *scikit-learn* libraries for model building and validation. On comparing results based on the log-likelihood, AIC and BIC values, we conclude that the ARIMA model provides better accuracy to the wind power forecasting as compared to solar power on the selected dataset.

Keywords Renewable energy · Forecasting · ARIMA

7.1 Introduction

The renewable energy sources like solar, wind, geothermal, ocean and biomass energy provide clean and replenishable energy, principally different from fossil fuels in terms of their diversity, abundance and potential to withstand energy shortage issues faced by the developing economies (Rather 2018; Kumar et al. 2010). Above all, these renewable sources support a more sustainable future by producing neither greenhouse

S. Pasari (✉) · A. Shah

Department of Mathematics, Birla Institute of Technology and Science, Pilani Campus, Pilani 333031, RJ, India

e-mail: sumanta.pasari@pilani.bits-pilani.ac.in; 86.sumanta@gmail.com

A. Shah

e-mail: f2014715p@alumni.bits-pilani.ac.in

© The Author(s) 2020

K. S. Sangwan and C. Herrmann (eds.), *Enhancing Future Skills and Entrepreneurship*, Sustainable Production, Life Cycle Engineering and Management, https://doi.org/10.1007/978-3-030-44248-4_7

gases which cause climate change nor carbon emissions (Kumar et al. 2010). There are different forms of renewable energy, depending on sunlight in a direct or indirect way. Renewable energy, by 2040, is projected to supply equal amount of electricity generation as obtained from coal and natural gas (Rather 2018; Kumar et al. 2010; Bhatia and Gupta 2018).

Solar energy is the direct conversion of sunlight (solar radiation) using solar panels or photovoltaic thermal solar collectors. It is the most abundant permanent energy resource on Earth. India, due to its geographical location, counts about 250–300 clear and sunny days in a year receiving about 4–7 kWh of solar radiation per square meter per day (Kumar et al. 2010). In particular, the western part of India experiences the highest amount of annual radiation energy. Therefore, the development of photovoltaic and solar panels is rapidly growing in India accounting for an installed capacity of 28.18 GW (as of March 2019) against an ambitious target of 100 GW by the end of 2022 (Rather 2018; Kumar et al. 2010; Bhatia and Gupta 2018). With substantial investment in solar electricity generation capacity, the cost of solar power is expected to fall about Rs. 1.90–2.30 per kWh by 2030 (Bhatia and Gupta 2018).

Wind energy, like solar power, is another abundant energy resource. However, the mechanism of wind generation is complex involving Earth's rotation, intensity of solar heat from sun, cooling effects of oceans, differential heating of the Earth's surface and the presence of other physical obstacles such as forest or mountains (Kumar et al. 2010). As a result, the wind speed at a location significantly varies over different seasons in a year. The wind power generation in India accounts for a total installed capacity of 36.63 GW (as of March 2019) against a target of 60 GW by 2022. Within a decade, the wind power cost will reduce to about Rs. 2.30–2.60 per kWh, while the storage cost will also come down by about 70% (Rather 2018; Kumar et al. 2010; Bhatia and Gupta 2018).

Apart from the solar and wind energy, biomass energy is a prominent source to meet the energy demand in developing countries. It contributes to about 32% of total primary energy, serving more than 70% of the country's population (Rather 2018; Kumar et al. 2010; Bhatia and Gupta 2018). Currently, there is about 5 GW capacity biomass powered plants against a national target of 10 GW installed biomass power by 2022 (Bhatia and Gupta 2018). Other sources of renewable energy include hydro-electricity and geothermal energy. While hydropower energy converts kinematic or potential energy of water into mechanical energy, geothermal energy is generated from the heat stored in the Earth, or the underground absorbed heat accumulation (Rather 2018; Kumar et al. 2010; Bhatia and Gupta 2018).

There have been several initiatives to forecast renewable energy variable resources, such as solar, wind and tidal using regressive models, artificial intelligence (AI) techniques, remote sensing models and numerical weather predictions (Inman et al. 2013; Lei et al. 2009; Kavasseri and Seetharaman 2009; Cadenas and Rivera 2010; Liu et al. 2012; Shukur and Lee 2015; Zhang et al. 2015; Wang et al. 2011; Tran 2013). Regressive models for day to year ahead forecasting include autoregressive (AR), moving average (MA), ARMA, ARIMA and fractional ARIMA (f-ARIMA) (Inman et al. 2013; Lei et al. 2009; Kavasseri and Seetharaman 2009).

Hybrid methods, such as ARIMA-ANN and ARIMA-Kalman are also used for wind speed prediction (Cadenas and Rivera 2010; Liu et al. 2012; Shukur and Lee 2015). In the present study, our objective is to develop ARIMA models for wind speed and solar energy (temperature) forecasting. Other concerns like cost-benefit analysis of new plants or technology development (Zhang et al. 2015; Wang et al. 2011) are not within the scope of this work.

7.2 Data Description

The data of Charanka Solar Park (23.95° N, 71.15° E) in Gujarat is procured from National Solar Radiation Database, maintained by National Renewable Energy Laboratory (NREL) (NREL homepage 2019). The obtained data comprises hourly data from year 2000 to year 2014 of the following variables: DHI (Diffuse Horizontal Irradiance), DNI (Direct Normal Irradiance), GHI (Global Horizontal Irradiance), clear-sky DHI, clear-sky DNI, clear-sky GHI, dew point, temperature, pressure, relative humidity, solar zenith angle, precipitable water, snow depth, wind direction and wind speed. For this study, we analyze the variables temperature and wind speed independently as univariate time series.

7.3 Methodology

Unlike traditional energy forecasting using probability distributions, here we develop a linear Autoregressive Integrated Moving Average (ARIMA) model that uses computer programming to provide reliable results with low computational complexity (Liu et al. 2012; Shukur and Lee 2015). The ARIMA model is a generalization of ARMA model comprising AR and MA parts. The AR part indicates that the changing variable regressed on its own lagged (i.e., prior) values, whereas the MA part incorporates the dependency between an observation and a residual error from a moving average model applied to prior observations. Here the wind speed and temperature hour-wise data are represented as individual univariate time series. Using the data of 2000–2013, results are first validated for the year 2014. Similarly, using the data till 2014, we can forecast for 2015 and so on. The method of training and validation for one-year ahead in itself is flexible for the available data.

As ARIMA models can be implemented only on stationary time series, the first task is to determine whether the time series corresponding to wind speed or temperature data is stationary or not. This in turn requires to verify whether the statistical properties such as mean, variance and autocorrelation are all constant over time (Tran 2013). A non-stationary time series may be converted to a stationary time series through differencing or detrending. The Dickey-Fuller (DF) test of stationarity and rolling (moving) statistics plots are often employed for this purpose (Tran 2013; Adhikari and Agrawal 2013). While the DF test is a statistical technique to

check stationarity, rolling statistics is more of a visual way to plot the mean and standard deviations over time. The DF test comprises a test statistic and some critical values corresponding to different confidence levels. To check the stationarity, the observed value of the test statistic is compared with the critical value to reject the null hypothesis (Tran 2013; Adhikari and Agrawal 2013). In brief, the regression model for the DF test uses the following equation:

$$y_t - y_{t-1} = (\rho - 1)y_{t-1} + u_t \quad (7.1)$$

where y_t is the variable of interest, t is the time index, ρ is a coefficient and u_t is the error term. A unit root is present if $\rho = 1$ and in that case, the time series becomes non-stationary. Thus, the null hypothesis for the test is $H_0: \rho = 1$ against the research hypothesis $H_1: \rho \neq 1$. In case of non-stationarity, the data may contain two factors: trend and seasonality. While trend is the variation in mean over time, seasonality is the variation between time frames (seasons). During the course of modeling, we estimate both quantities and eliminate them, if necessary (Tran 2013).

The basic technique to eliminate trend is to *log-transform* the series. This will result in penalizing larger values more than the smaller values affecting the trend accordingly. Then, to remove the trend, the noise is removed through smoothing, aggregation or polynomial fitting. This results in averaging out the trend which could be subtracted from the log-transformed time series. *Differencing* is another popular technique to eliminate (reduce) trend and seasonal effects by removing changes in the level of a time series. The third technique often used to reduce trend and seasonal components is to *decompose* the time series into various terms comprising the ones contributing to trend, seasonality, cyclic nature, irregularity and the residuals. Thus, decomposition method provides access to the residuals which are nothing but the time series after removing other components (Tran 2013; Adhikari and Agrawal 2013).

Renewable energy prediction using non-seasonal ARIMA (p, d, q) model is based on the non-negative integer parameters p, d and q where p is the order (number of time lags) of the AR model, d is the degree of differencing (for non-seasonal differences) and q is the order of the MA model (Tran 2013). For computing p and q values, two graphs, namely autocorrelation function (ACF) and partial autocorrelation function (PACF) are often employed (Tran 2013). The ACF is the metric to determine the correlation of a signal with its own copy delayed or separated by several time lags. Although ACF turns out to be an excellent tool in identifying the order of an MA (q) process, it is not very useful to identify the order in an AR (p) process. The PACF provides partial correlation of a stationary time series with its own prior values, and regressed the values of the time series at all shorter lags (Tran 2013; Adhikari and Agrawal 2013). The PACF is useful to identify the order p .

The above methodology is implemented in python using *pandas* for data exploration, and *stats* and *scikit-learn* libraries for model building and validation. The next section summarizes various results of the study.

7.4 Results and Conclusions

The rolling mean plots of temperature and wind speed data for a period of 365 days are provided in Fig. 7.1a, whereas the results of the DF test is summarized in Table 7.1. The decomposition of wind speed data into trend, seasonality and residuals is presented in Fig. 7.1b. Similar decomposition can be carried out for the temperature data.

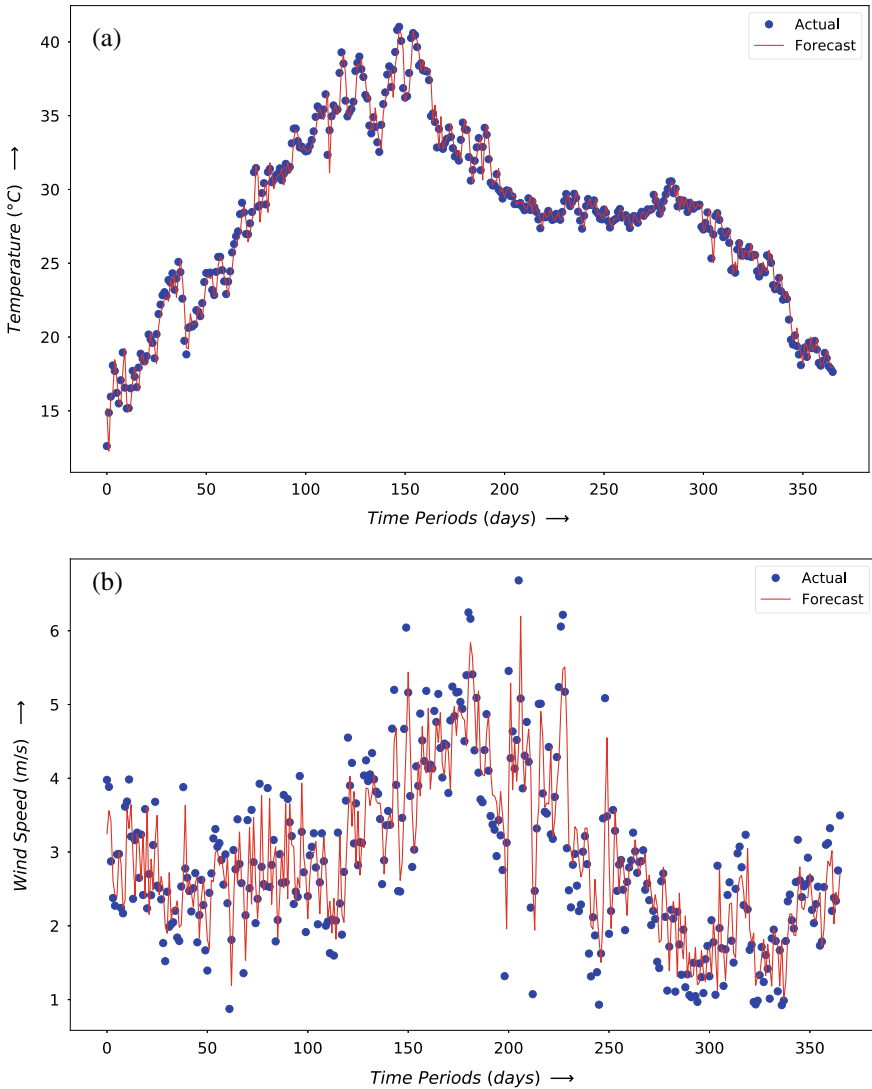


Fig. 7.1 Forecast versus actual data curve for **a** temperature and **b** wind speed data

Table 7.1 Results of Dickey-Fuller test

	Temperature	Wind speed
Test statistic	-6.96	-6.11
<i>p</i> -value	9.39e-10	9.40e-08
Number of lags	32	26
Number of observations	5442	5448
Critical value (1%)	-3.43	-3.43
Critical value (5%)	-2.86	-2.86
Critical value (10%)	-2.57	-2.57

It is observed from Table 7.1 that for both the series, the test statistic value falls in the rejection region of the null hypothesis. Therefore, we may consider both time series to be stationary. From the rolling mean/variation plots (available on request), it is also visually observed that both time series data are stationary. In addition, the ACF and PACF plots (available on request) suggest *p* and *q* values to be 2 for both temperature and wind speed data. Thus we use ARIMA (2, 1, 2) model to forecast the renewable energy data. From the results based on the log-likelihood, AIC and BIC, we observe that the linear univariate ARIMA model provides a better fit to the observed wind speed data in comparison to temperature data.

The classification model provides a validation accuracy of 83.24%. The root mean square (RMS) errors corresponding to observed and modeled temperature and wind speed data for the year 2014 are 0.893 and 0.659, respectively, indicating a reasonable match to the observed and modeled results (Fig. 7.1).

In summary, the present study has provided a scheme to model, forecast and validate temperature and wind speed data using a univariate linear ARIMA (2, 1, 2) model. The approach is generic, scalable and suitable to be applied for other renewable energy resources for planning, unit commitment analysis and integration of renewable energy to the main power grids. Here we have forecasted one-year ahead, though the method can be easily modified to two-year forecasting with more number on input data. The method is computationally inexpensive and provides good accuracy. In future, multivariate ARIMA models or Recurrent Neural Networks (RNN) could be further explored to improve the model accuracy.

References

- Adhikari R, Agrawal RK (2013) An introductory study on time series modeling and forecasting. Lambert Academic Publishing, pp 1–76
- Bhatia SC, Gupta RK (2018) Textbook of renewable energy. Woodhead Publishing, India, pp 1–334
- Cadenas E, Rivera W (2010) Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. *Renew Energy* 35(12):2732–2738
- Inman RH, Pedro HTC, Coimbra CFM (2013) Solar forecasting methods for renewable energy integration. *Progress Energy Combust Sci* 39(6):535–576

- Kavasseri RG, Seetharaman K (2009) Day-ahead wind speed forecasting using f-ARIMA MODELS. *Renew Energy* 34(5):1388–1393
- Kumar A, Kumar K, Kaushik N, Sharma S, Mishra S (2010) Renewable energy in India: current status and future potentials. *Renew Sustain Energy Rev* 14:2434–2442
- Lei M, Shiyang L, Chuanwen J, Hongling L, Yan Z (2009) A review on the forecasting of wind speed and generated power. *Renew Sustain Energy Rev* 13(4):915–920
- Liu H, Tian HQ, Li YF (2012) Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Appl Energy* 98:415–424
- NREL homepage. <https://www.nrel.gov/>, Accessed 24 June 2019
- Rather NUR (2018) Introduction to renewable energy technologies in India. Education Publishing, pp 1–140
- Shukur OB, Lee MH (2015) Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renew Energy* 76:637–647
- Tran VL (2013) stochastic models of solar radiation processes. Ph.D. Dissertation. University of Orleans, France
- Wang X, Guo P, Huang X (2011) A review of wind power forecasting models. *Energy Procedia* 12:770–778
- Zhang Y, Beaudin M, Taheri R, Zareipour H, Wood D (2015) Day ahead power output forecasting for small-scale solar photovoltaic electricity generators. *IEEE Trans Smart Grid* 6(5):2253–2262

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

