Chapter 16

Statistical Modeling of Solar Energy

Sumanta Pasari[✉] and Venkata Satya Siva Koundinya Nandigama[✉]

Abstract Renewable energy comprises solar, wind, tidal, biomass and geothermal energies. Use of renewable energy resources as a substitute for fossil fuels inevitably reduce environmental footprint. Therefore, integration of renewable energy to the power grid, smart grid planning and grid-storage preparations are some of the major concerns in all developing countries. However, unpredictability in renewable energy resources makes the situation challenging. In light of this, the present study aims to develop a solar energy forecasting model to estimate future energy supply for a smooth integration of solar energy to the current electric grids. A suite of eight probability models, namely exponential, gamma, normal, lognormal, logistic, log-logistic, Rayleigh and Weibull distributions are used. While the model parameters are estimated from the maximum likelihood estimation method, the performance of the candidate distributions is tested using three goodness of fit tests: Akaike information criterion, Chi-square criterion, and K-S minimum distance criterion. Based on the sample data obtained from the Charanka Solar Park, Gujarat, it is observed that the Weibull model provides the best representation to the observed solar radiations. The study concludes with the analysis of forecasted solar energy and its possible role in replacing thermal energy resources.

Keywords Solar energy · Probability distribution · Charanka solar park

16.1 Introduction

World energy resources can be broadly classified into three types: fossil fuels, renewable resources, and nuclear resources. Renewable energy is the energy collected from naturally replenished sources such as sunlight, wind, tides, biomass and geothermal

S. Pasari (✉) · V. S. S. K. Nandigama (Deceased)
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

V. S. S. K. Nandigama (Deceased)
e-mail: f2013108p@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_16

heat. It contributed to almost 20% to human's global energy consumption and 25% to global electricity generation in 2015 and 2016, respectively (REN 21 homepage 2019; Global energy homepage 2019). India is one of the largest renewable energy producing countries accounting for about 35% of the total installed power capacity in the electricity sector. The target by 2030, as stated in the Paris Agreement, is to achieve 40% of total India's electricity generation from non-fossil fuel sources (Global energy homepage 2019; Paris Agreement homepage 2019; Rather 2018). As a consequence, a large number of wind and solar energy plants are being installed in the country under the purview of the Ministry of New and Renewable Energy (MNRE) (Paris Agreement homepage 2019; Rather 2018).

Despite the installation of many renewable energy plants, their integration to the main power grids is crucial in harnessing renewable energy applications (REN 21 homepage 2019; Global energy homepage 2019; Paris Agreement homepage 2019; Rather 2018; Zhang et al. 2015). The unpredictability of renewable energy resources, such as wind speed and solar radiation makes integration difficult, as the current electric grids cannot operate unless there is a mutual balance between supply and demand (Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi 2011; Delucchi and Jacobson 2011; NREL homepage 2019). An imbalance may result in voltage fluctuations and even worse (NREL homepage. <https://www.nrel.gov/>. 2019). Other problems related to renewable energy sources include the unavailability of solar power at night during which the power consumption is at its peak and the lack of efficient energy storage systems to save the excess electricity production (Delucchi and Jacobson 2011; NREL homepage 2019). In addition, as renewable energy plants are usually located far away from the consumption location, transportation of power may cause unwanted transmission losses (Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi 2011; Delucchi and Jacobson 2011; NREL homepage 2019).

Several methods are employed for the forecasting of solar irradiation considering numerical weather prediction, artificial neural networks (ANN), linear and non-linear stochastic models, remote sensing based models and hybrid models (Ferrari et al. 2013; Zhang et al. 2015; Inman et al. 2013). Comparison of several autoregressive models (AR, ARMA, ARIMA) (Ferrari et al. 2013) and neural network based models such as Radial Basis Function Neural Networks (RBFNN), Least Square Support Vector Machine (LS-SVM), k-Nearest Neighbour (kNN), and Weighted kNN (WkNN) methods (Zhang et al. 2015) have been implemented as forecasting engines. Use of empirical probability models (Pasari 2015, 2018) could also be tried for energy forecasting. In summary, two main categories of studies have evolved, one focusing on the smart grid or grid energy storage technology and another aiming at forecasting of renewable energy (Rather 2018; Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi 2011; Delucchi and Jacobson 2011; NREL homepage 2019). The present study considers the latter issue and concentrates on the statistical modeling of solar power output at Charanka Solar Park, Gujarat. The aim is to select the best-fit probability distribution(s) among exponential, gamma, normal, lognormal, logistic, log-logistic, Rayleigh and Weibull models to forecast solar radiations.

16.2 Data Description

Solar radiation, the radiant energy emitted by the sun, is the primary data for the present analysis. When solar radiation enters into the Earth's atmosphere, a fraction of the radiation reaches directly to the surface. Such radiation is called beam or direct radiation. The remaining fraction may be scattered or absorbed by air molecules, clouds or aerosols. A part of such scattered radiation reaches the ground and is known as diffuse radiation. Another part of the direct radiation hitting the surface gets reflected and may reach upon another surface, such as solar collector or photovoltaic panel. Such radiation is called albedo. The sum of these three components is termed as global radiation (Rather 2018). The quantum of global irradiation collected per unit area is an important parameter for solar power forecast.

Direct Normal Irradiance (DNI) is the amount of solar radiation received per unit area by a surface that is always held perpendicular to the rays coming in a straight line from the direction of the sun at its current position in the sky. Diffuse Horizontal Irradiance (DHI), on the contrary, is the amount of radiation received per unit area by a surface that does not arrive on a direct path from the sun, but has been scattered by molecules and particles in the atmosphere and comes equally from all directions. Global Horizontal Irradiance (GHI) is the total amount of shortwave radiation received from above by a surface horizontal to the ground (Rather 2018). The GHI may be calculated from DNI and DHI as

$$GHI = DHI + DNI * \cos(\theta) \quad (16.1)$$

Where θ is the solar zenith angle (Rather 2018; Zhang et al. 2015).

The data of the Charanka Solar Power Park (23.95° N, 71.15° E) in Gujarat was procured from the National Solar Radiation Database of National Renewable Energy Laboratory (NREL) (NREL homepage 2019). It comprises hourly data of all the variables (e.g., DNI, DHI, GHI, and many others) affecting the solar irradiation from 2000 to 2014. It is observed that depending on the season, about 12 h of daily solar irradiation data (06:30–18:30 h) contain non-zero positive entries of DNI, DHI and GHI values. There are many zero values in the sample data indicating that the day did not start or the day had ended. To maintain consistency in the analysis, 08 h of daily data (09:30–16:30 h) is considered for modeling. With this filtering, yearly 2920 data points are obtained. All DNI, DHI and GHI data are fitted separately to identify the best-fit probability model(s) for solar power forecast.

It may be noted that the original datasets also contain information on temperature, pressure, relative humidity and precipitation among others, although those are not used in the present analysis.

16.3 Methodology and Results

On a temporal scale, solar power forecasting may be classified into now casting (forecasting up to a few hours in advance), short-term forecasting (forecasting up to a few days in advance) and long-term forecasting (forecasting months or years ahead) (Rather 2018; Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi 2011; Delucchi and Jacobson 2011). Depending upon the range of forecasts required, forecasting models have been developed accordingly incorporating parameters that are affecting solar radiation in the range (Zhang et al. 2015). Both short and long-term power forecasts have their specific applications. While system operators use short term forecasts in unit commitment analysis and determining reserve unit requirements, solar farm owners use such forecast for bidding strategy planning (in electricity markets) and dealing with voltage imbalance issues while integrating solar power supply to major thermal power distribution networks (Rather 2018; Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi 2011; Delucchi and Jacobson 2011; NREL homepage 2019). The long-term solar power forecasts are particularly important for smart city planning and negotiating contracts with financial entities or utilities (Zhang et al. 2015). Statistical approaches, as in this study, are preferred for long-term forecasts.

The methodology here comprises three major steps: probability model assumption, parameter estimation, and model validation. Based on some graphical representation of data, eight probability models are considered to fit DNI, DHI and GHI data separately.

Model parameters of the studied distributions are estimated from the classical maximum likelihood estimation (MLE) method, whereas the model selection is performed based on three goodness of fit tests, namely Akaike information criterion (AIC), Chi-square criterion and K-S minimum distance criterion. The AIC test is a simple modification from the log-likelihood scores and it accounts for the additional number of parameters in the competitive models.

The Kolmogorov–Smirnov (K–S) test, in contrast, is a non-parametric approach. The Chi-square test determines significant differences between the expected and observed frequencies in one or more categories (Ferrari et al. 2013; Zhang et al. 2015). The results of estimated parameters and selection scores corresponding to average GHI data are presented in Table 16.1.

It may be noted that each parameter in the studied distributions has its respective role (e.g., shape, scale, and location) (Pasari 2015, 2018). Moreover, like the results in Table 16.1 for GHI data, one may obtain results corresponding to DHI and DNI data using simple excel tools along with Matlab plots. It is observed that the Weibull model consistently provides the best representation for DHI, DNI and GHI data. The pictorial representation of the model fit for DHI, DNI and GHI data of year 2000 is illustrated in Figs. 16.1, 16.2 and 16.3.

With the above process of finding the best fit probability distribution, one can now analyze solar irradiation data for future estimation. As a secondary illustration, forecasting of solar irradiance may be carried out using a simple linear regression model (Rather 2018; Zhang et al. 2015; Su et al. 2012; Jacobson and Delucchi

Table 16.1 Estimated parameter values and model selection results for the GHI data

Distribution	Parameter values		Model selection		
			AIC	Chi-Square	K-S
Exponential	$\hat{\lambda}$ (scale)	468.01	334927.5	38992.2	0.3601
Gamma	$\hat{\alpha}$ (scale)	340.61	298718.7	2311.84	0.0757
	$\hat{\beta}$ (shape)	1.37			
Normal	$\hat{\mu}$ (location)	468.01	295710.6	421.84	0.0413
	$\hat{\sigma}$ (scale)	272.73			
Lognormal	$\hat{\mu}$ (log-scale)	5.74	301470.0	4349.08	0.0902
	$\hat{\sigma}$ (shape)	1.27			
Logistic	$\hat{\mu}$ (location)	476.86	296652.3	695.22	0.0467
	$\hat{s}$ (scale)	165.17			
Log-logistic	$\hat{\alpha}$ (scale)	5.98	299742.4	3751.30	0.0871
	$\hat{\beta}$ (shape)	0.59			
Rayleigh	$\hat{\alpha}$ (scale)	383.01	309226.2	10596.11	0.2297
Weibull	$\hat{\alpha}$ (scale)	505.43	295072.0	360.61	0.0348
	$\hat{\beta}$ (shape)	1.44

Fig. 16.1 Data fit of DNI values for the year 2000

Fig. 16.2 Data fit of DHI values for the year 2000

Fig. 16.3 Data fit of GHI values for the year 2000

Table 16.2 Regression analysis of total DHI during 2000–2014 for the month of March

Summary output: regression statistics		
	Coefficients	Standard error
Multiple R	0.2294	
R Square	0.0526	
Adjusted R-Square	0.0506	
Standard Error	245.8266	
Observations	465	
Multiple R	0.2294	
Intercept	2292.7576	22.8367
Variable	0.4306	0.08493

2011; Delucchi and Jacobson 2011). The regression analysis can well explain the relationship between criterion variables (dependent variables) and predictor variables (independent variables). Then those values are interpolated or extrapolated with the help of the relationship obtained by the regression model. In the present study, the analysis is performed with the help of MS-excel inbuilt statistical data analysis tool. The data is forecasted for 5 additional years from 2015 to 2019 month wise from the results of the linear regression analysis based on 2000–2014. Also, the solar irradiation is forecasted using the first 10 years of the sample data from 2000 to 2009 and the corresponding errors have been calculated. For a demonstration, the regression analysis for the month of March is provided in Table 16.2 below. Table 16.2 shows that the adjusted R-squared value is 0.0506, while the multiple R value is 0.2294. Similarly, regression analysis can be easily performed for all the months of the calendar. However, improvements in the results are necessary.

16.4 Summary and Conclusions

Statistical modeling of renewable energy plays a pivotal role in the future energy sector and therefore its importance can never be disregarded. In this work, first the best-fit probability model of solar irradiation data is identified using eight popular probability distributions. Then a linear regression analysis is carried out to forecast solar energy for the Charanka Solar Park, Gujarat. During the course of the study, the following important observations are noted:

- As the day progresses, the amount of DHI, DNI and GHI values increases till afternoon and then decreases. This is because the zenith angle gradually decreases to zero as the day advances to afternoon and then the zenith angle gradually increases as the day advances into evening followed by night.
- The best-fit distribution for a particular hour over the months remains consistent although with varying means. This may be attributed to the variation in the amount of solar irradiation received on account of the seasons in respective years.

- The standard deviations of the fitted distribution are very high. Even the MSE (Mean Squared Error) for the regression is also high probably due to the less amount of data points.

As a conclusion, the present work has provided a layout to develop a solar energy-forecasting model towards the endeavor of estimating future energy supply for a smooth integration of solar energy to the current electric grids. Results, based on the limited data, are preliminary and require further analysis for a stringent conclusion.

References

- Delucchi MA, Jacobson MZ (2011) Providing all global energy with wind, water, and solar power, part II: reliability, system and transmission costs, and policies. *Energy Policy* 39(3):1170–1190. <https://doi.org/10.1016/j.enpol.2010.11.045>
- Ferrari S, Lazzaroni M, Piuri V, Loredana C, Marco F (2013) Statistical models approach for solar radiation prediction. In: 2013 IEEE international instrumentation and measurement technology conference (I2MTC), pp 1734–1739. <https://doi.org/10.1109/i2mtc.2013.6555712>
- Global energy homepage. <https://www.iea.org>. Last accessed 24 June 2019
- REN 21 homepage. <http://www.ren21.net/gsr-2017/>. Last accessed 24 June 2019
- Inman RH, Pedro HTC, Coimbra CFM (2013) Solar forecasting methods for renewable energy integration. *Prog Energy Combust Sci* 39(6):535–576. <https://doi.org/10.1016/j.pecs.2013.06.002>
- Jacobson MZ, Delucchi MA (2011) Providing all global energy with wind, water, and solar power, part I: technologies, energy resources, quantities and areas of infrastructure, and materials. *Energy Policy* 39(3):1154–1169. <https://doi.org/10.1016/j.enpol.2010.11.040>
- NREL homepage. <https://www.nrel.gov/>. Last accessed 24 June 2019
- Paris Agreement homepage. <https://www4.unfccc.int/sites/submissions/indc/Submission%20Pages/submissions.aspx>. Last accessed 24 June 2019
- Pasari S (2015) Understanding himalayan tectonics from geodetic and stochastic modeling. PhD dissertation. Indian Institute of Technology Kanpur, India
- Pasari S (2018) Stochastic modeling of earthquake interoccurrence times in northwest Himalaya and adjoining regions. *Geomatics, Nat Hazards Risk* 9(1):568–588. <https://doi.org/10.1080/19475705.2018.1466730>
- Rather NUR (2018) Introduction to renewable energy technologies in India. Education Publishing, pp 1–140
- Su Y, Chan LC, Shu L, Tsui KL (2012) Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems. *Appl Energy* 93:319–326. <https://doi.org/10.1016/j.apenergy.2011.12.052>
- Zhang Y, Beaudin M, Taheri R, Zareipour H, Wood D (2015a) Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators. *IEEE Trans Smart Grid* 6(5):2253–2262. <https://doi.org/10.1109/tsg.2015.2397003>
- 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. <https://doi.org/10.1109/tsg.2015.2397003>

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.