Forecasting of Solar Irradiance using Probability Distributions for a PV System: A Case Study

Sudha Dukkipati *‡, V. Sankar*, P. Srinivasa Varma**

* Department of Electrical and Electronics Engineering, JNTUACE (A), Anantapur. 515002

** Department of Electrical and Electronics Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India. 522502

(dukkipati.sudha@gmail.com, vsankar.eee@jntua.ac.in, pinnivarma@kluniversity.in)

‡ Corresponding Author: Sudha Dukkipati, dukkipati.sudha@gmail.com, Tel: +9603152778

Received: 25.02.2019 Accepted: 20.03.2019

Abstract- Photo Voltaic (PV) generation is intermittent which cannot carry out the constant electric power. The amount of solar irradiance received for a particular area is one of the most important climatic conditions for forecasting PV generation. The solar irradiance data used for analysis is collected from Koneru Lakshmaiah Education Foundation (KLEF) and the data is collected for the year (2017-18). In this paper, four Probability Distribution Functions (PDFs) such as Normal, Weibull, Rayleigh, Log normal are used to investigate the best fit probability distribution of the collected solar irradiance data. Root Mean Square Error (RMSE) method is used as a goodness of fit test for identifying the best suited probability distribution function for the collected solar irradiance data. With this analysis, solar irradiance and PV power generation can be predicted for the next year. The efficacy of the proposed method is validated using MATLAB.

Keywords: Probability Distribution Function, Solar irradiance, Monte Carlo Simulation, Root Mean Square Error (RMSE).

1. Introduction

In the past years, the smart grid has pulled in increasing consideration in the power sector. The smart grid includes Advanced Metering Infrastructure (AMI), communication facilities, smart energy management methodologies, as well as distributed power generation using sustainable power sources [8]. Sustainable power sources are used to reduce the usage of fossil fuels and to reduce carbon footprint on environment. PV system is one of the quick-developing green power generations among the available renewable power sources [22]. A PV system generates electricity by direct transformation of solar energy into electric power. PV based power stations are very useful for the selection of substitution of conventional electrical energy generation as it is less unending and less pollutant. However, PV generation cannot carry out constant electrical power output. The PV output power is hard to anticipate precisely on the grounds that it emphatically connects with the atmosphere, surrounding temperature, geography and time [4]. Hence, a probabilistic model of PV output is required to reenact the genuine conduct of these stations.

Models that consider the stochastic behavior of PV power can be characterized using analytical methods [2], [7], [9], [12-14] [19], [21],Monte Carlo based methods [5], [10], [11], [15],[18] and Artificial Neural Networks. Unsymmetrical two-factor estimation method is used to predict the undetermined nature of the PV system was presented in [7]. A Monte Carlo based method was proposed in [15] for modeling PV based power generation in consideration of their dependency on other sustainable power sources, but not with atmospheric factors. In [23], a probabilistic power flow analysis was developed by modelling wind speed as Weibull PDF, load demand as Gaussian PDF, solar irradiance as single discrete PDF i.e. without categorizing into different seasons. In [3] solar irradiance was forecasted for a day using Multilayer Perceptron Model. But this method has difficulty in selection of best suited activation function for each model and selecting number of hidden nodes. This model is suitable for forecasting solar irradiance in sunny days than that of cloudy days. ANN-SFP model was proposed in [6] to forecast solar irradiance. İn [16] Back Propagation algorithm to forecast irradiance is presented. But it requires other metrological data i.e. mean temperature, mean humidity, mean rainfall also in addition to half-hourly solar irradiance as input data. Solar irradiance ramps were forecasted using image-based Integrated Solar Forecasting Platform [17]. But this method requires sky imaging data. Accuracy of cloud detection system need to be improved particularly for thin clouds. Analytical techniques represent the unpredictable PV system behavior which is completely dependent on meteorological parameters with the aid of mathematical model and forecast the solar irradiance by using mathematical analysis by considering some simplifications which causes some unrealistic nature. Monte Carlo based techniques are used so far to predict single probability model for solar irradiance or PV output power i.e. without considering weather effects which causes unrealistic PV output power. ANN based techniques require other meteorological parameters data i.e. temperature, humidity, rainfall, cloud cover also in addition to solar irradiance to forecast irradiance data for the next year which is always not possible to get all parameters. Moreover, forecasted solar irradiance data depend on the selection of initial parameters for training the ANN model. Another drawback of some of the existing ANN based techniques is forecasting of solar irradiance is effectively done in sunny days than that of in rainy days because of cloud cover [20].

From the literature, it is obsreved that, no work has been carried out to predict solar irradiance using probability distributions by considering different atmospheric conditions. In this paper, as an extension to fill the gaps in the existing literature, Inverse Transform based Monte Carlo simulation algorithm [1] is proposed by considering different weather conditions for forecasting solar irradiance data of next year. So, this suggested algorithm forecast the solar irradiance data without any assumptions and without the need of other meteorological parameters. The suggested algorithm is validated using MATLAB and the effects, resultant discussions show the efficiency of the proposed algorithm. This simulation results can be used to forecast PV power generation accurately.

This paper is arranged as follows: In Section II, different probability distribution functions, Monte Carlo Simulation (MCS) which were used to identify the best fit of solar irradiance data and Root Mean Square Error (RMSE) have been discussed. Flow chart for the proposed method has been presented in Section III. The Simulation results of solar irradiance data collected from KLEF and forecasted solar irradiance data with various PDFs are analyzed in Section IV. Conclusions are discussed in Section V.

2. Forecasting of Solar Irradiance

2.1. Monte Carlo Simulation

Monte Carlo Simulation (MCS) is an approach used for forecasting models. Monte Carlo methods are useful within the conditions where direct experimentation is not possible. Similarly, Monte Carlo techniques can also be applied to the mathematical problems which can't be solved via direct way, or where an instantaneous solution is simply too expensive or requires too much time. It is a treasured device when forecasting an unknown future. It is a method to simulate the actual process, random nature of the system and to forecast the future behavior of the system with the help of random variable. A random variable is one whose probable values are outcomes of a random phenomenon. It is not unusual that these outcomes rely upon some physical variables that aren't properly understood. It is described as a function that maps the outcomes of uncertain processes to numerical quantities, generally real numbers. The uniform random numbers generated between [0 1] can be converted to other non-uniform distributions using inverse transform method.

2.2. Probability Distribution Functions

It is impractical to report all the statistical distribution functions as there are many to be accommodated. Hence four distribution functions like., normal, Rayleigh, Weibull, log normal are used to identify the best fit.

Normal distribution:

The normal distribution is the significant and broadly used distribution in statistics. It is also referred to as "bell curve". The PDF of normal distribution can be expressed as

$$f(s) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(s-\mu)^2}{2\sigma^2}] \tag{1}$$

The parameter μ represents mean, σ represents standard deviation and s is the solar irradiance.

Rayleigh distribution:

The Rayleigh distribution is a continuous PDF and is extensively used in communications theory, physical sciences, in engineering, to measure the lifetime of an object. It is a special case of Weibull distribution.

The Rayleigh distribution's PDF is

$$f(s) = \frac{s}{b^2} \exp\left(-\frac{s^2}{2b^2}\right) \tag{2}$$

Where b is the parameter of Rayleigh distribution

Weibull distribution:

The PDF of Weibull distribution is

$$f(s) = 1 - \exp\left[-\left(\frac{s}{c}\right)^k\right] \tag{3}$$

where, s is the solar irradiance, k is the shape parameter and c is the scale parameter.

Log normal:

It is also a continuous PDF of a random variable whose logarithm is normally distributed. The PDF of Log Normal distribution is

$$f(s) = \frac{1}{s\beta\sqrt{2\pi}} exp\left\{-\frac{[\ln(s)-\lambda]^2}{2\beta^2}\right\} \tag{4}$$

where, λ represents mean, β represents standard deviation of the variable's natural logarithm.

2.3. Goodness of Fit:

The goodness of fit explicates how properly a probability distribution function suits a group of observations. Goodness of fit commonly summarize the variation between observed values and values predicted by the model. Such measures can be used in statistical hypothesis testing. Here RMSE is considered to determine the best PDF. RMSE is the standard deviation of the predicted errors which are also called as residuals.

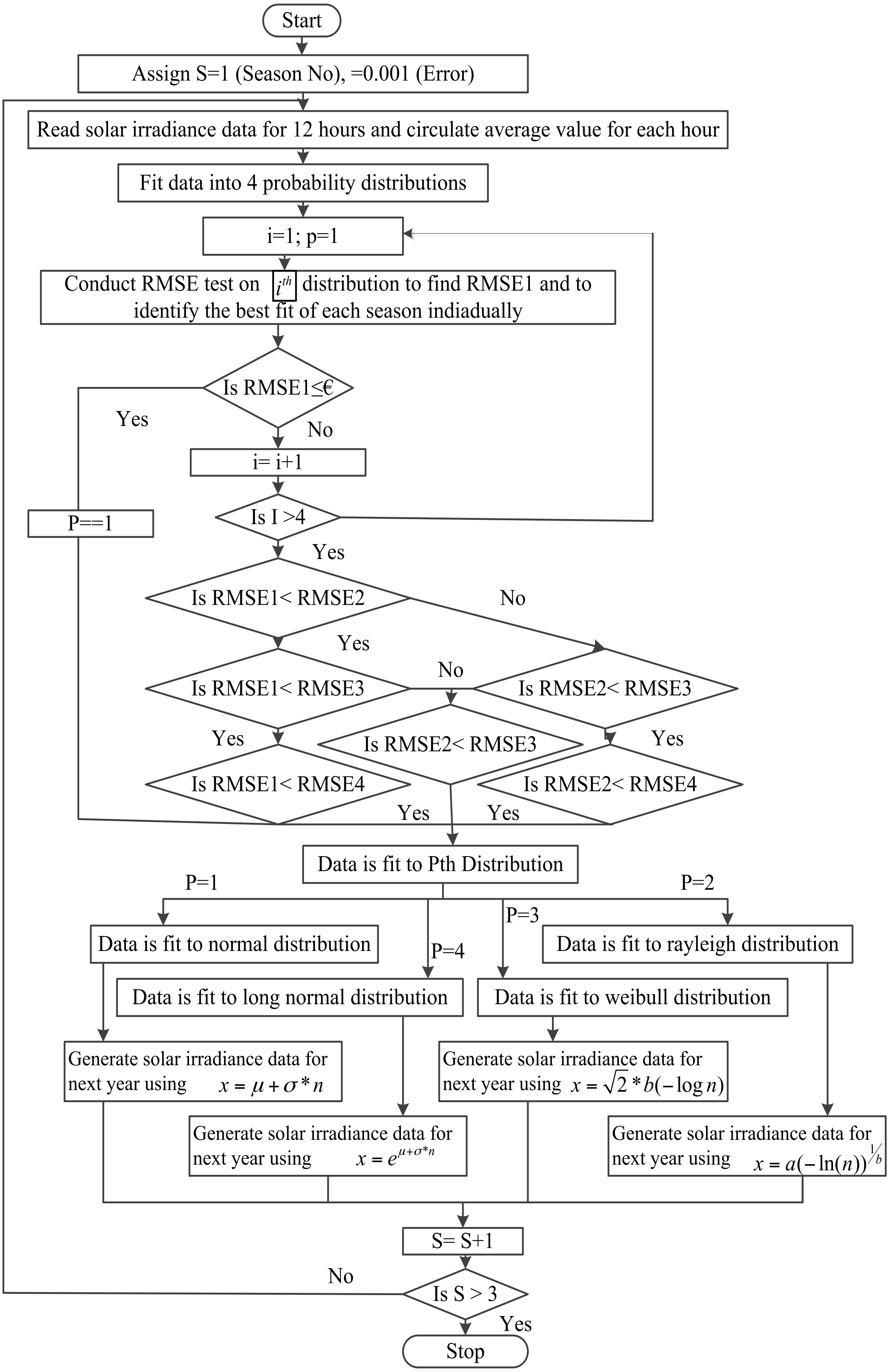


Fig. 1. Flowchart of proposed method

Table 1. Average hourly Solar Irradiance

Season/ Time	Solar Irradiance (Watt/m^2)												
	7:00	**8:00**	**9:00**	**10:00**	**11:00**	**12:00**	**13:00**	**14:00**	**15:00**	**16:00**	**17:00**	**18:00**	**19:00**
Summer	7.9	68.6	169	351.4	538.2	693.6	831	859	830	700	493.1	230.2	24.97
Rainy	3.3	86.6	114	252.2	390.4	487.7	567.2	581	539	442	336	179.6	47.47
Winter	26	94.1	222	401.1	577.5	694.8	739.1	687	536	301	77.63	0	0

Residuals are an indication of deviation of data points from the regression line. Hence it indicates the concentration of data points around the best fitted line

RMSE is calculated as

$$\text{RMSE}=\left[\frac{1}{n}\sum_{j=1}^{n}(S_j - S_{jc})^2\right]^{1/2} \qquad - \quad (5)$$

where, s_j are the observed values at time stage j, s_{jc} are the values computed from the probability distribution for the same stage, n is the number of data points. Smaller the deviations, better the fit is.

3. Flowchart for Proposed Method

The flow chart of the proposed method is shown in Fig. 1 and the steps are as follows:

Step 1: Assign Season No. S=1 (1 for summer, 2 for rainy and 3 for winter season) and ε = 0.001 (error)

Step 2: Collect the solar irradiance data of KLEF 160 kW plant from 7:00 A.M. to 7:00 P.M. for each season and calculate average value for each hour.

Step 3: Fit the collected solar irradiance data of each season to 4 probability distributions (Normal, Weibull, Lognormal, Rayleigh individually).

Step 4: Assign i=1 and p=1

Step 5: Conduct RMSE test to find deviation from the continuous probability distribution function for all the 4 distributions.

Step 6: Check whether RMSE of ith distribution is less than ε or not. If yes Move to Step 8, else Move to Step 6.

Step 7: Increment i and check whether i>4 or not. If yes Move to Step 7, else Move to Step 4.

Step 8: Find the least RMSE$_i$ (i=1 for Normal, 2 for Weibull, 3 for Rayleigh, 4 for Lognormal) by comparing the 4 Root Mean Square Errors to identify the best fit probability distribution function.

Step 9: Forecast solar irradiance data for the next year using Monte Carlo Simulation Technique for the best suited probability distribution.

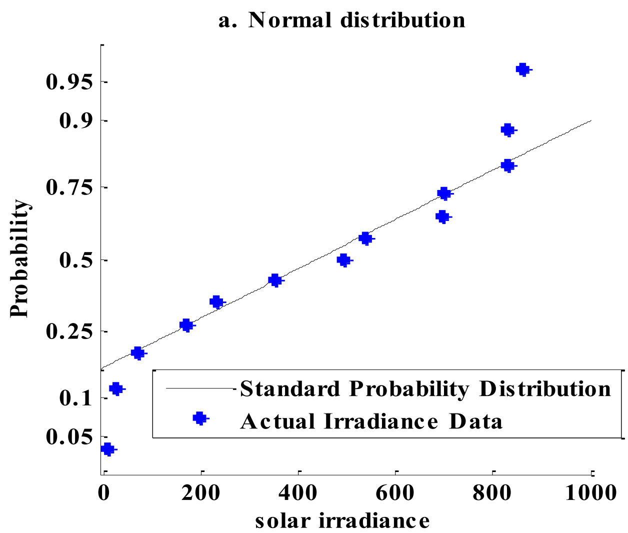
Step 10: Increase the Season No. and check whether all the 3 seasons are completed or not. If yes stop the procedure, else Move to Step 2.

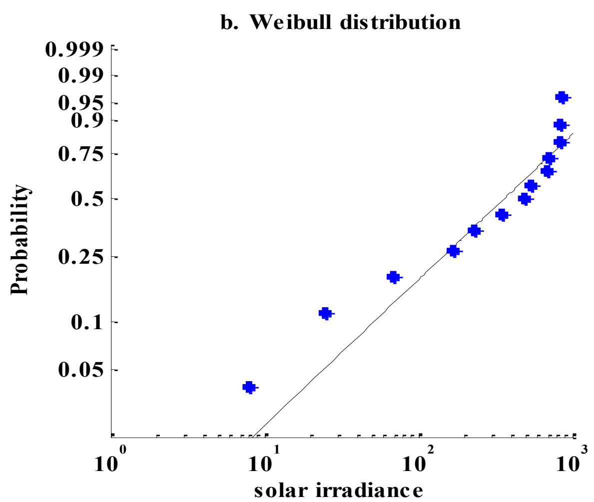
4. Simulation Results and Discussion

4.1. Solar Irradiance Probabilistic Model

The hourly solar irradiance data of 2016-17 for 12 hours is collected for summer, rainy and winter seasons from KLEF. The average hourly solar irradiance is calculated for these seasons and are furnished in Table 1. From Table. 1 it is noticed that highest solar irradiance occurred in summer season mid-day and solar irradiance in winter season is nearly zero because of early sunset.

The continuous Normal, Weibull, Rayleigh, Lognormal PDFs are compared with discrete actual PDFs of each season separately using MATLAB as shown in Figs.2 (a) to (d) for summer season, Figs. 3 (a) to (d) for rainy season and Figs. 4 (a) to (d) for winter season respectively. Continuous PDFs are represented with solid line and actual discrete PDFs are represented with dotted line. From Fig. 2, it is observed that solar irradiance of summer season from 200 W/m^2 to 800 W/m^2 is following Normal and Rayleigh distributions, i.e most of the solar irradiance data is following Normal and Rayleigh PDFs and the data is very much deviated from Weibull and Lognormal distributions for the initial values.





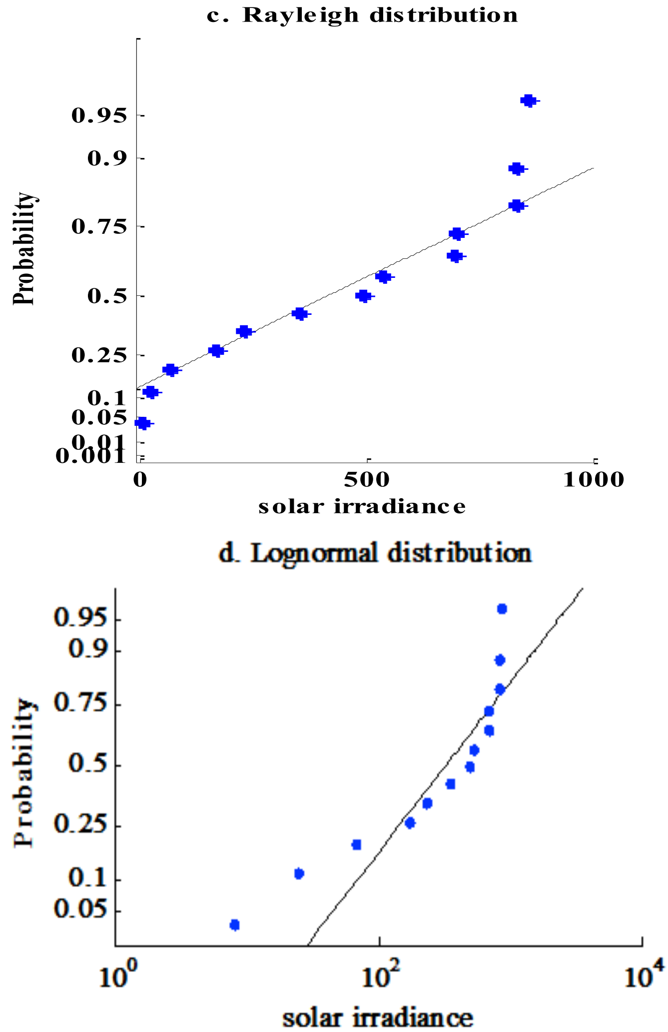
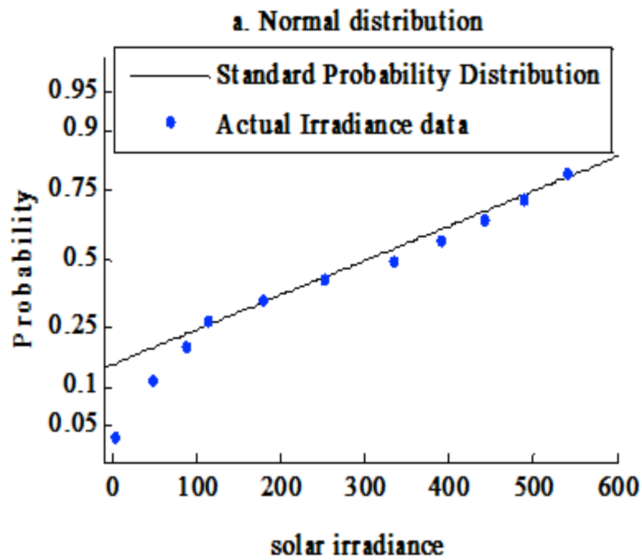


Fig.2. Probability Plots of Solar Irradiance for summer season

From Fig.3, it can be observed that solar irradiance of rainy season from 100 W/m^2 to 600 W/m^2 is following Normal and Rayleigh distributions and the data is very much deviated from Weibull and Lognormal distributions for the middle values of solar irradiance.



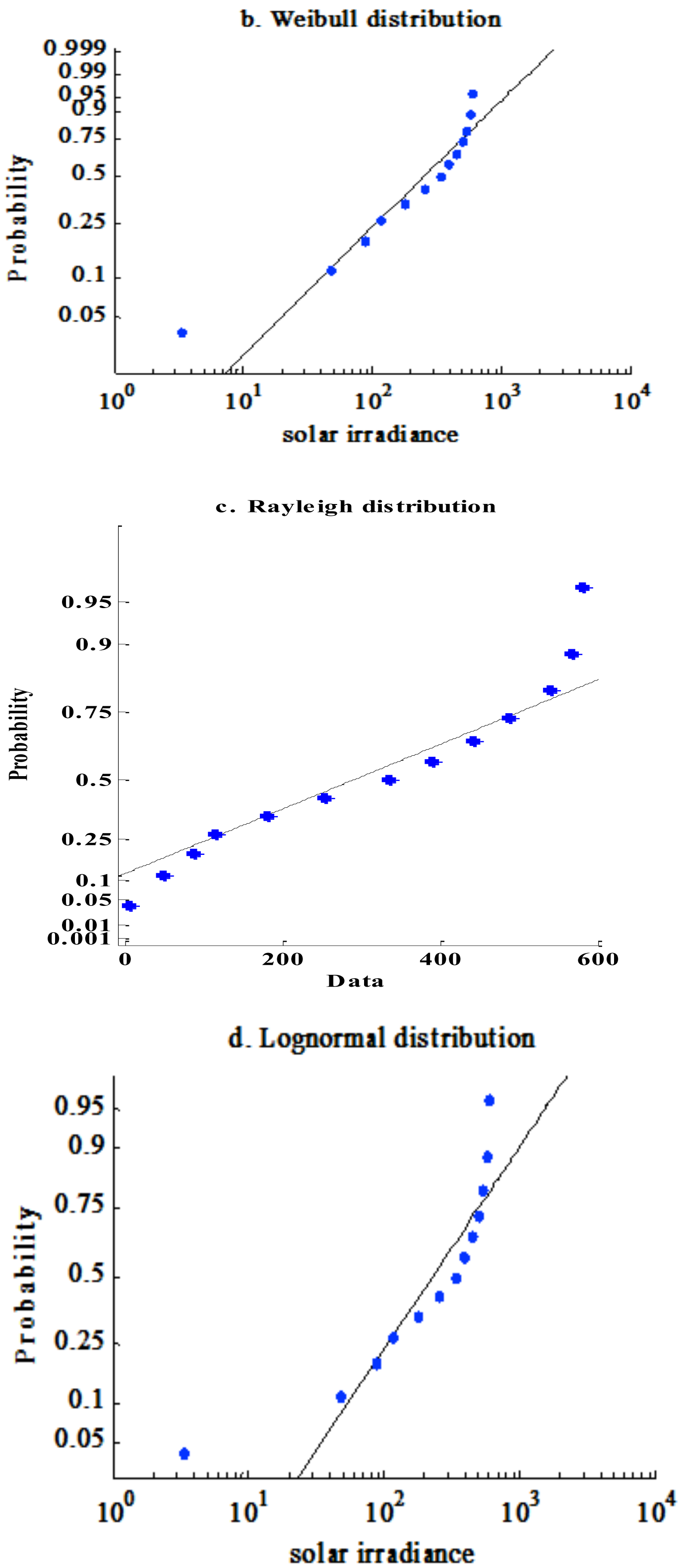
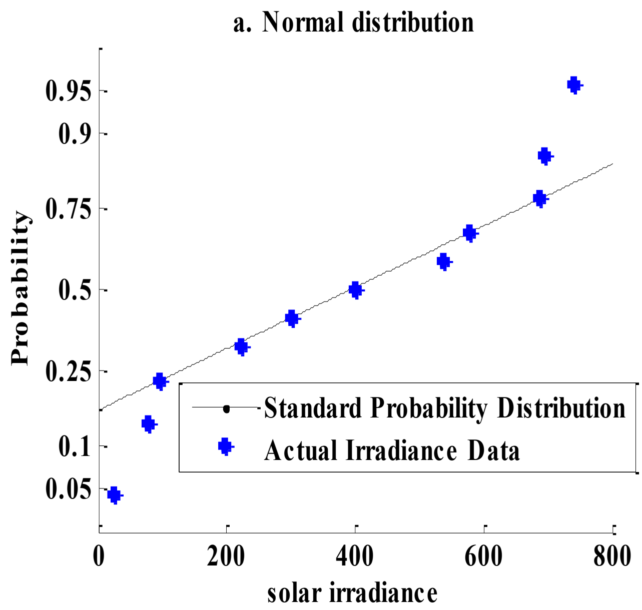
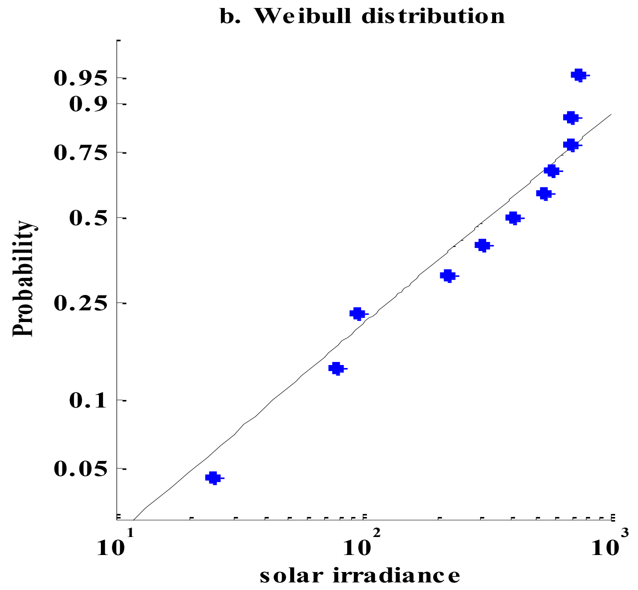
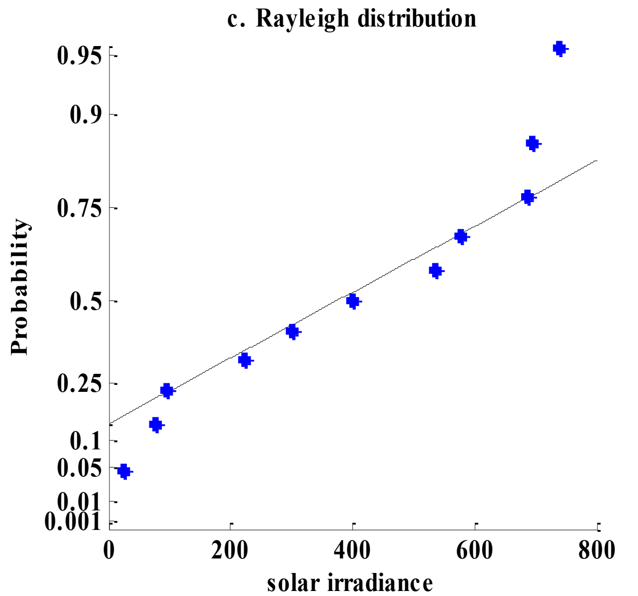


Fig. 3. Probability Plots of Solar Irradiance for rainy season

From Fig. 4 it can be seen that most of solar irradiance values of winter season are fitted for Normal and Rayleigh PDFs compared to Weibull and Lognormal PDFs. Actual Irradiance data follwoing the nature of Normal and Rayleigh Standard Probability Distributions and it is deviated more for Weibull and Lognormal PDFs.







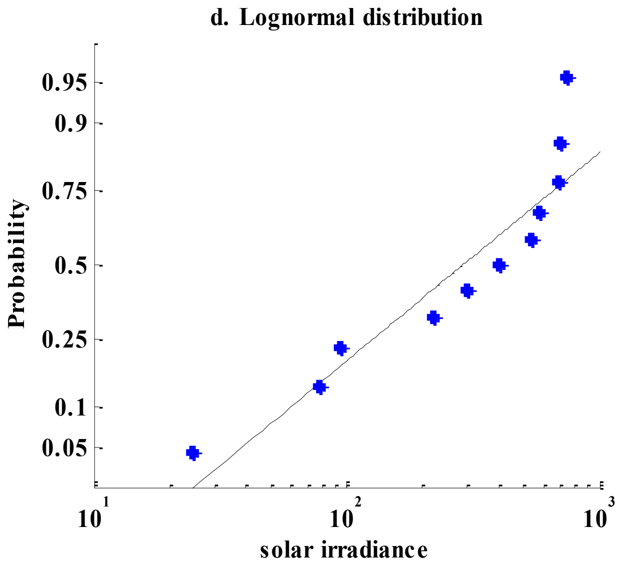


Figure 4. Probability Plots of Solar Irradiance for winter season

Hence, to determine the most suitable PDF of each season for forecasting solar irradiance for the next year, goodness of fit test is conducted for Normal, Weibull, Rayleigh and Log normal PDFs' plots by finding RMSE using Eq. (5). The RMSE values of each PDF can be observed as presented in Table 2.

From Table 2 it can be observed that RMSE is slightly less for Normal PDF in Summer and Rainy seasons and for Rayleigh PDF in winter season. Hence, it can be clearly concluded that best suited PDF for summer and rainy seasons is Normal and Rayleigh for winter season.

Table 2. Root Mean Square Error Values

Season	Root Mean Square Error			
	Normal	**Weibull**	**Rayleigh**	**Log normal**
Summer	0.0604	0.0926	0.0625	0.0825
Rainy	0.0591	0.0746	0.0601	0.0646
Winter	0.0771	0.0936	0.073	0.0849

Mean and Standard Deviation of Normal PDF are calculated for summer and rainy seasons. Similarly, parameter 'b' of Rayleigh distribution also calculated and is shown in Table-3.

Table 3. Parameters of Probabilistic model

Season	Probabilistic Model	Parameters
Summer	Normal	$\mu = 446.0$; $\sigma = 309.7$
Rainy	Normal	$\mu = 309.6$; $\sigma = 200.0$
Winter	Rayleigh	b = 306.2

4.2 Forecasted Solar Irradiance Data

A uniformly distributed random vector between 0 and 1 is generated. Solar irradiance random variable for every one-hour duration is generated for every random number using Monte Carlo Simulation for the best suited PDF using the parameter values listed in Table 3. So, the simulated solar irradiance is generated at each hour considering only day time for all the three seasons (i.e., a total of 4320 random variables are generated) and are compared with the continuous PDFs for three seasons.

The forecasted solar irradiance data obtained for summer, rainy and winter seasons have been obtained and compared with the actual data and are shown in Figs. 5 to 7 respectively. Solar irradiance data for summer season of next year is forecasted and from Fig.5, it is concluded that the forecasted data is following normal distribution. It is observed that probability of occurrence of solar irradiance >600 W/m^2 is more in summer season.

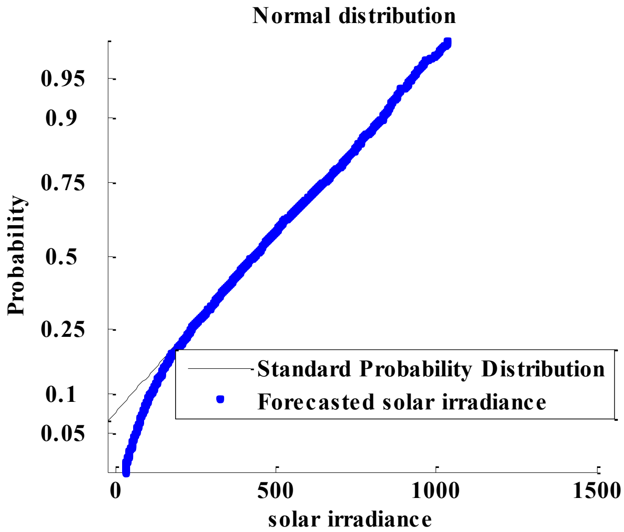


Fig.5. Probability plot of forecasted Solar irradiance for summer season

From Fig.6, it is inferred that forecasted solar irradiance for rainy season is same as that of Normal distribution and probability density of solar irradiance >800 W/m^2 is less when compared to that in summer season.

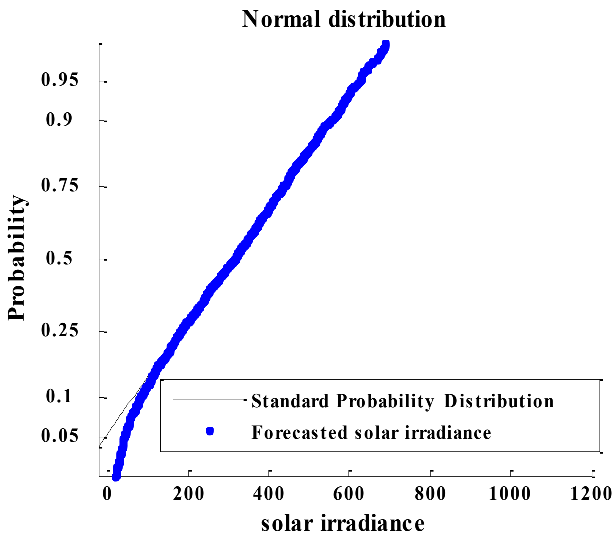


Fig. 6. Probability plot of forecasted Solar irradiance for rainy season

Forecasted Solar irradiance for winter season follow Rayleigh distribution as shown in Fig.7 and it is observed that there is a probability of occurrence of zero solar irradiance in winter season because of early sunset. It is verified with the 2016-17 solar irradiance average value data tabulated in Table 1.

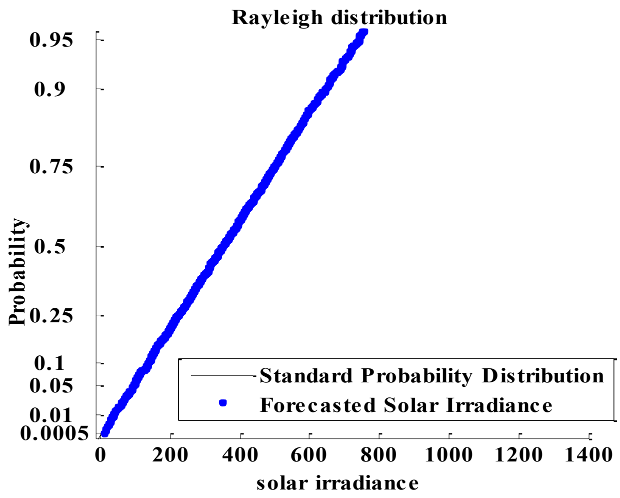


Fig. 7. Probability plot of forecasted Solar irradiance for winter season

5. Conclusions

In this paper, the most suitable probability distribution function for forecasting the solar irradiance of K L E F campus has been determined and the parameters of each fitted distribution have been calculated. From the results obtained by conducting RMSE test, it is concluded that normal distribution is the most suitable distribution function for summer and Rainy seasons and Rayleigh distribution is the most appropriate distribution function for winter season. From the forecasted solar irradiance plots, i.e. from Figs.5,6,7, it is observed that probability of occurrence of higher values of solar irradiance is more in summer season than that of in winter and rainy seasons.

From the results and analysis, it is observed that the proposed method of forecasting solar irradiance is very much useful with the following advantages; i) Prediction of realistic and accurate solar irradiance for each and every season individually by fitting data into different PDFs. ii) Improved accuracy without any assumptions and without considering other meteorological parameters.

References

[1] R. Billinton and R. N. Allan, *Reliability Evaluation of Engineering Systems: Concepts and Techniques*. Boston, MA: Springer US: Imprint: Springer, 1992.

[2] Y. M. Atwa, E. F. El-Saadany, M. M. A. Salama, and R. Seethapathy, "Optimal Renewable Resources Mix for Distribution System Energy Loss Minimization," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 360–370, Feb. 2010.

[3] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application

for performance prediction of a grid-connected PV plant at Trieste, Italy," *Sol. Energy*, vol. 84, no. 5, pp. 807–821, May 2010.

[4] F. J. Ruiz-Rodriguez, J. C. Hernández, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," in 2011 *International Conferene on Power Engineering, Energy and Electrical Drives*, Spain, May 2011, pp. 6-11.

[5] F. Attivissimo, A. Di Nisio, M. Savino, and M. Spadavecchia, "Uncertainty Analysis in Photovoltaic Cell Parameter Estimation," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 5, pp. 1334–1342, May 2012.

[6] F. Wang, Z. Mi, S. Su, and H. Zhao, "Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters," *Energies*, vol. 5, no. 5, pp. 1355–1370, May 2012.

[7] A. Soroudi, M. Aien, and M. Ehsan, "A Probabilistic Modeling of Photo Voltaic Modules and Wind Power Generation Impact on Distribution Networks," *IEEE Syst. J.*, vol. 6, no. 2, pp. 254–259, Jun. 2012.

[8] C. F. Calvillo, A. Sanchez, and J. Villar, "Distributed energy generation in smart cities," in 2013 *International Conference on Renewable Energy Research and Applications* (ICRERA), Madrid, Spain, 2013, pp. 161–166.

[9] R. Sayed, Y. G. Hegazy, and M. A. Mostafa, "Modeling of photovoltaic based power stations for reliability studies using Markov chains," in *2013 International Conference on Renewable Energy Research and Applications (ICRERA)*, Madrid, Spain, 2013, pp. 667–673.

[10] P. Siano and G. Mokryani, "Probabilistic Assessment of the Impact of Wind Energy Integration Into Distribution Networks," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4209–4217, Nov. 2013.

[11] Q. Zhao, P. Wang, L. Goel, and Y. Ding, "Evaluation of nodal reliability risk in a deregulated power system with photovoltaic power penetration," *IET Gener. Transm. Distrib.*, Dec. 2013.

[12] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu, "Chronological Probability Model of Photovoltaic Generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1077–1088, May 2014.

[13] S.Aguilar, R. C. Souza, and J. F. Pensanha, "Predicting probabilistic wind power generation using nonparametric techniques," in *2014 International Conference on Renewable Energy Research and Applications* (ICRERA), Milwaukee, WI, USA, 2014, pp. 709–712.

[14] K. Namrata, S. P. Sharma, and S. B. L. Seksena, "Determining regression constants for calculating global solar radiation at Jharkhand (India) region," in 2014 *International Conference on Renewable Energy Research and Applications (ICRERA)*, Milwaukee, WI, USA, 2014, pp. 795–797.

[15] A. Y. Abdelaziz, Y. G. Hegazy, W. El-Khattam, and M. M. Othman, "Optimal allocation of stochastically dependent renewable energy based distributed generators in unbalanced distribution networks," *Electr. Power Syst. Res.*, vol. 119, pp. 34–44, Feb. 2015.

[16] S.Watetakarn and S. Premrudeepreechacharn, "Forecasting of solar irradiance for solar power plants by artificial neural network," in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, Bangkok, Thailand, 2015, pp. 1–5.

[17] Y. Chu, H. T. C. Pedro, M. Li, and C. F. M. Coimbra, "Real-time forecasting of solar irradiance ramps with smart image processing," *Sol. Energy*, vol. 114, pp. 91–104, Apr. 2015.

[18] S. Basile, R. Burlon, and F. Morales, "Joint probability distributions for wind speed and direction. A case study in Sicily," in 2015 *International Conference on Renewable Energy Research and Applications* (ICRERA), Palermo, 2015, pp. 1591–1596.

[19] P. S. Loh, J. V. Chua, A. C. Tan, and C. I. Khaw, "Data-driven short-term forecasting of solar irradiance profile," *Energy Procedia*, vol. 143, pp. 572–578, Dec. 2017.

[20]A.Dolara, A.Gandelli, F.Grimaccia, S.Leva, M.Mussetta, "Weather-based Machine Learning Technique for Day-ahead Wind Power Forecasting Technique," in 2017 *International Conference on Renewable Energy Research and Applications (ICRERA)*, San Deigo, CA, USA, 2017, pp.206-209.

[21] X. G. Agoua, R. Girard, and G. Kariniotakis, "Probabilistic Model for Spatio-Temporal Photovoltaic Power Forecasting," *IEEE Trans. Sustain. Energy*, pp. 1–1, 2018.

[22] S.S. Alrwashdeh, "Assessment of Photovoltaic Energy Production at Different Locations in Jordan," *International Journal of Renewable Energy Research*, vol. 8, No.2, p.797-804, June 2018.

[23] G. E. Constante-Flores and M. S. Illindala, "Data-Driven Probabilistic Power Flow Analysis for a Distribution System With Renewable Energy Sources Using Monte Carlo Simulation," *IEEE Trans. Ind. Appl.*, vol. 55, no. 1, pp. 174–181, Jan. 2019.