

Investigation on Frequency Distribution of Global Radiation Using Different Probability Density Functions

Tian Pau Chang *

Department of Computer Science and Information Engineering, Nankai University of Technology, Nantou, Taiwan, R.O.C.

Abstract: The amount of daily irradiation received for a particular area is one of the most important meteorological parameters for many application fields. In this paper, the frequency distributions of global radiations are investigated using four kinds of probability density functions, i.e. the Weibull function, logistic function, normal function and lognormal function. The radiations observed at six meteorological stations in Taiwan are selected as sample data to be analyzed. To evaluate the performance of the probability functions both the Kolmogorov-Smirnov test and the root mean square errors are considered as judgment criteria. The results show that all the four probability functions are applicable for stations where weather conditions are relatively steady throughout the year as in Taichung and Tainan. While for stations revealing more dispersive distribution as in Hualien and Taitung, the lognormal function describes the frequency distribution quite better than other three functions. On the whole the lognormal function performs best followed by the normal function; the Weibull function widely used in other fields seems to be not appropriate in this case.

Keywords: global radiation; frequency distribution; probability density function; Kolmogorov-Smirnov test; root mean square error

1. Introduction

When solar radiation enters into the Earth's atmosphere, some of the incident energy will be scattered or absorbed by air molecules, clouds or aerosols. The radiation directly reaches the ground surface is called beam or direct radiation. The scattered radiation that reaches the ground is the diffusion radiation. A part of the radiation may reach upon a surface, such as solar collector or photovoltaic panel, after reflection from the ground is the albedo. The sum of these three components is called global radiation. The amount of daily global irradiation received for a given area is one of the most important meteorological pa-

rameters for many application fields, e.g. in solar designing, plant culturing, architecture, tourism, emigrating and so forth. As stated in Chang [1] and Shu et al. [2], the optimal tilt angle of a solar collector or photovoltaic panel in fine days is larger than that in cloudy days, because the amount of diffusion component, which is nearly an isotropic radiation, within global radiation gets weaker in fine days. As concluded by Gueymard [3-4] as well as Kudish and Ianetz [5] a concentrating system is more recommended and suitable for cloudless environment based on the consideration of economic cost. That is why the

* Corresponding author; e-mail: t118@nkut.edu.tw

study on the characterizations of global radiation has its importance.

To investigate the characterizations of solar radiation, some probability density functions have been proposed in literature to describe its frequency distribution, such as the normal function [5], the Boltzmann function [6-9], the gamma function [10], the logistic function [11] and so forth. Kudish and Ianetz [5] found that the frequency distribution for clear days is more approximate to normal distribution than that for cloudy days. Babu and Satya-murty [12] found that the distribution models available in literature are not applicable for all climates, and proposed new expressions to obtain a family of generalized distribution function. Assuncao et al. [13] analyzed the clearness index in Brazil to estimate sky condition based on five-minute global irradiation data. Therein the logistical and Weibull functions were used for clear and cloudy sky situations, respectively. Ettoumi et al. [14] stated that a linear combination of two beta distributions is found properly to fit the monthly frequency distributions of the hourly solar radiation data in Algeria. Moreover Jurado et al. [15] proposed a mixture of two normal distributions and Soubdhan et al. [16] used a mixture Dirichlet distribution to analyze solar radiation data. However a detailed comparison demonstrating how appropriate the probability density functions model the frequency distributions has never been found in literature.

Taiwan locates between the world's largest continent (Asia) and the largest ocean (Pacific). The Tropic of Cancer (23.5° N) divides roughly the island into two climates: the tropical monsoon climate in the south and the subtropical monsoon climate in the north. High temperature and humidity, massive rainfall and tropical cyclones in summer characterize the climate of Taiwan. According to Köppen's classification, the four climate types in Taiwan are a Monsoon and Trade-Wind Coastal Climate in the south; Mild, Humid Climate in the north; Wet-Dry

Tropical Climate in the west, and Temperate Rainy Climate with Dry Winter in mountain areas. The latitude and topography, ocean currents and monsoons are the main contributing factors [17]. In winter, when the north-eastern monsoon system is active, the north is constantly visited by drizzle while the south remains dry. In spring, the weather in the north is rainy. In summer, when the south-western monsoon comes into force, the eastern part of the island, which is situated on the leeward side of hills, experiences more sunny days. Generally, the southwestern part experiences more stable weather conditions throughout the year.

The situation of Taiwan has its uniqueness worth to be studied; the characterizations of solar radiation around here were analyzed previously in Chang [1]; but the frequency distributions of the radiation have not been touched in this article. In the present paper, the same radiation data as in Chang [1] is used, which is the global irradiation observed on the ground surface at six meteorological stations between 1990 and 1999, conducted by the Central Weather Bureau, i.e. the stations Taipei, Taichung, Tainan, Kaohsiung, Hualien and Taitung. Four kinds of probability density functions (pdf) available in statistics are applied to describe their frequency distributions, i.e. the Weibull, logistic, normal and lognormal functions. Among these functions, the lognormal function is applied for the first time to the field. The performance of the probability functions will be compared according to both the Kolmogorov-Smirnov test and the root mean square errors (RMSE).

2. Probability density functions

2.1. Weibull distribution

Two-parameter Weibull probability density function (pdf) has widely been applied in the field of renewable energy given as [18-20]:

$$f(x) = \frac{k}{c} \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right] \quad (1)$$

Where, x is the solar irradiation; k is the shape parameter, and c is the scale parameter. The Weibull cumulative distribution function (cdf) is given by:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{c}\right)^k\right] \quad (2)$$

Weibull shape and scale parameters can be estimated respectively by:

$$k = \left(\frac{\sigma}{\bar{x}}\right)^{-1.086} \quad (3)$$

$$c = \frac{\bar{x}}{\Gamma(1+1/k)} \quad (4)$$

where, \bar{x} and σ is the mean and standard deviation of data, respectively, and $\Gamma(\cdot)$ is the Gamma function.

2.2. Logistic distribution

Logistic function can be expressed as [13]:

$$g(x) = \frac{\exp[-(x - \bar{x})/\alpha]}{\alpha\{1 + \exp[-(x - \bar{x})/\alpha]\}^2} \quad (5)$$

The cumulative logistic function is given by:

$$G(x) = \frac{1}{1 + \exp[-(x - \bar{x})/\alpha]} \quad (6)$$

The curve of the logistic probability density function is symmetric to the mean of data (\bar{x}), where its frequency is a maximum. α is the scale parameter of logistic distribution

calculated by:

$$\alpha = \sqrt{3}\sigma / \pi \quad (7)$$

2.3. Normal distribution

The probability density function of normal distribution is given as:

$$h(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \bar{x})^2}{2\sigma^2}\right] \quad (8)$$

The cumulative normal distribution function is expressed as the following equation:

$$H(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x - \bar{x}}{\sigma\sqrt{2}}\right) \quad (9)$$

Where $\operatorname{erf}(\cdot)$ is the error function given by:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \quad (10)$$

2.4. Lognormal distribution

Lognormal distribution is a kind of probability density functions for any random variable whose logarithm is normally distributed; it can be expressed as below [21]:

$$r(x) = \frac{1}{x\beta\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x) - \lambda]^2}{2\beta^2}\right\} \quad (11)$$

Where, λ and β are the mean and standard deviation of the variable's natural logarithm, respectively.

The cumulative lognormal distribution function is given as:

$$R(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left[\frac{\ln(x) - \lambda}{\beta\sqrt{2}}\right] \quad (12)$$

3. Goodness of fit

To show how well a probability density function fit with observation data, two kinds of judgment criteria are employed in this study, the first one is the Kolmogorov-Smirnov test defined as the maximum error between two relevant cumulative distribution functions [22].

$$KS = \max|\phi_1(x) - \phi_2(x)| \quad (13)$$

Where $\phi_1(x)$ and $\phi_2(x)$ are the cumulative distribution functions for solar radiation not exceeding x in the theoretical and observed data set, respectively. The critical values for the Kolmogorov-Smirnov test at 95% and 99% confident level are respectively calculated by:

$$KS_{0.05} = \frac{1.36}{\sqrt{n}} \quad (14)$$

$$KS_{0.01} = \frac{1.63}{\sqrt{n}} \quad (15)$$

Where n is the number of data. Another judgment criterion of fit is the root mean square error (RMSE) defined as:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (y_i - y_{ic})^2 \right]^{1/2} \quad (16)$$

Where, y_i are the actual values at time stage i , y_{ic} are the values computed from correlation expression for the same stage, n is the number of data. Basically, the smaller the errors the better the fit is.

4. Results and discussion

Figure 1-6 show the annual frequency dis-

tribution of the global irradiation for six stations. Four probability density functions are merged into the plots; relevant curves of cumulative distribution functions are provided and referred to the right axis. Parameter values for each probability distribution are determined by using the corresponding mean and standard deviation of the global irradiation. Table 1 lists the annual statistical quantities of global irradiation for the six stations.

It is seen that the frequency distribution of observed radiation for a given area can appropriately reflect its weather condition. For example, in station Taipei, the distribution pattern of solar radiation reveals a relatively low level due to its frequently rainy days appeared in winter and spring seasons, as well as the attenuation effects of clouds and aerosols.

As for Taichung and Tainan, at which weather condition is steadier for the entire year, the distribution patterns of solar radiation more concentrate about their mean values accompanying with smaller percent coefficients of variation (about 20%). As a result all the four probability density functions proposed fit very well with the observations; all the max errors of cumulative distribution functions in the Kolmogorov-Smirnov test are below the critical value of 95% confident level. It implies that the solar potential energy is easier to be estimated for an area where the frequency distribution of solar radiation is more concentrated.

While for stations Hualien and Taitung, which locate at the leeward side of mountains when southwest monsoon prevails in summer and therefore receive stronger radiation, their distribution patterns are more dispersive than other four stations with percent coefficient of variation higher than 30% and with skewness coefficient 0.345 above; the Weibull, logistic and normal functions fail to model the observations. The max errors calculated from both the logistic and normal functions exceed even the critical value of 99% confident level. But the lognormal function firstly used in the field

describes still very well the observations.

As for root mean square errors, they exhibit similar trends with the max errors in the Kolmogorov-Smirnov test; the lognormal function presents smaller errors in fitting the distributions. On the whole the lognormal function gives the best description to the global radiations followed by the normal function. Note that the Weibull function widely used in other fields seems not applicable in the case.

5. Conclusion

In this study, the frequency distributions of global irradiations observed in Taiwan had been investigated using four kinds of probability density functions. The lognormal function was applied for the first time to the field. To evaluate how well a probability density function fits with the observation data, both the Kolmogorov-Smirnov test and root mean square errors were considered as judgment criteria. The study on the distribution of solar radiation would offer engineers with useful knowledge regarding the applications of solar energy. The conclusions can be summarized as follows:

(a) All the four probability density functions

would be applicable for stations as Taichung and Tainan where weather conditions are steadier throughout the year.

(b) While for data revealing more dispersive distribution as in stations Hualien and Taitung, the lognormal function describes the frequency distribution quite better than other three functions.

(c) Root mean square errors present similar result with the max errors in the Kolmogorov-Smirnov test while doing performance comparison of the probability density functions.

(d) Overall the lognormal function performs best followed by the normal function; the Weibull function commonly applied in other fields seems to be not appropriate in this case.

Acknowledgments

The author would deeply appreciate the Central Weather Bureau for providing irradiation data and thank Dr. Huang MW, researcher of the Institute of Earth Sciences, Academia Sinica, Taiwan, for his treasure suggestions. This work was supported partly by the National Science Council under contract NSC99-2221-E-252-011.

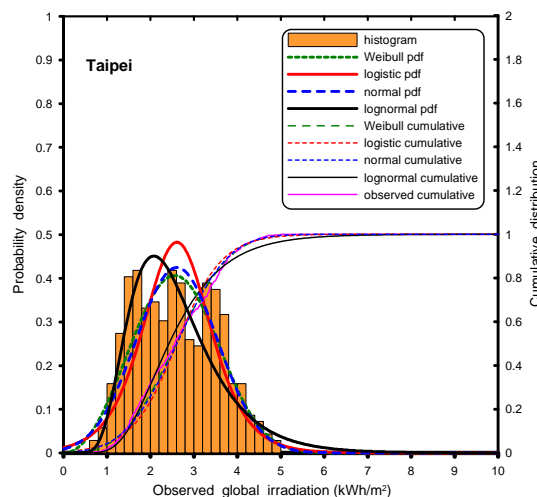


Figure 1. Annual frequency distribution of global irradiation for Taipei

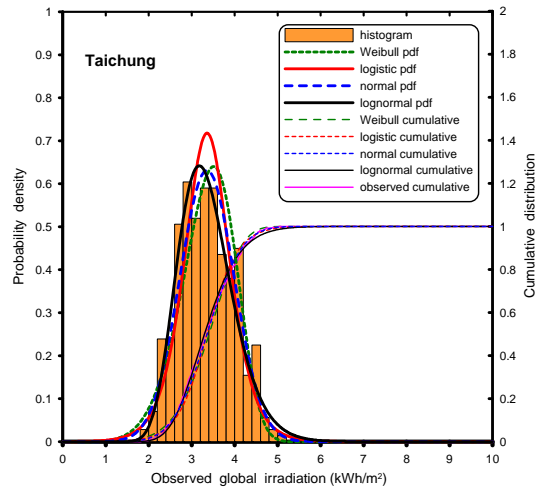


Figure 2. Annual frequency distribution of global irradiation for Taichung

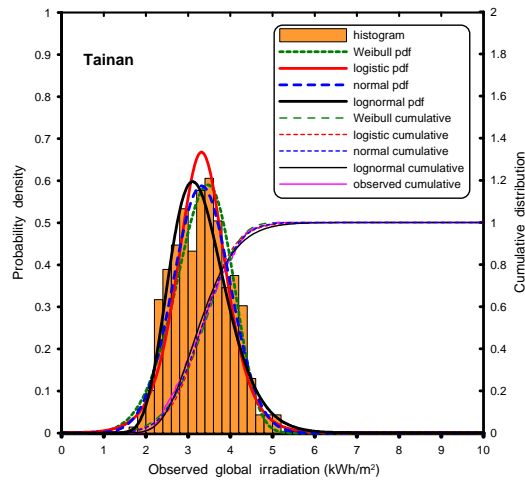


Figure 3. Annual frequency distribution of global irradiation for Tainan

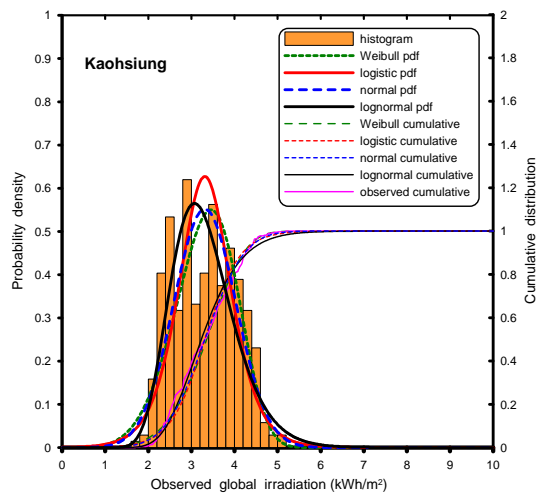


Figure 4. Annual frequency distribution of global irradiation for Kaohsiung

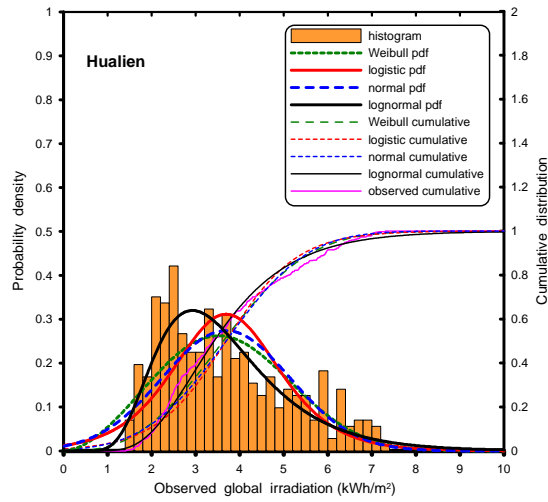


Figure 5. Annual frequency distribution of global irradiation for Hualien

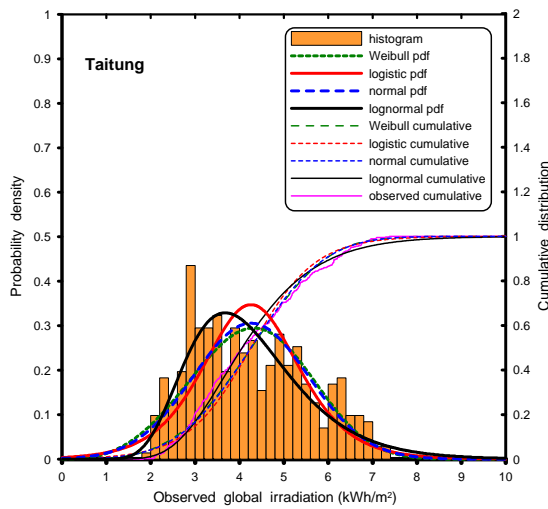


Figure 6. Annual frequency distribution of global irradiation for Taitung

Table 1. Annual statistical quantities of global irradiation (in kWh/m²) for 6 stations studied

Quantities	Taipei	Taichung	Tainan	Kaohsiung	Hualien	Taitung
Mean	2.609	3.357	3.317	3.311	3.690	4.258
Standard deviation	0.940	0.632	0.679	0.723	1.455	1.305
Percent coefficient of variation (%)	36.02	18.83	20.46	21.84	39.44	30.65
Skewness coefficient	0.186	0.169	0.144	0.125	0.615	0.345
Kurtosis coefficient	2.152	2.440	2.411	2.074	2.382	2.148
Weibull shape parameter	3.033	6.197	5.660	5.273	2.741	3.630
Weibull scale parameter	2.920	3.612	3.588	3.595	4.147	4.724
Logistic scale parameter	0.5180	0.3484	0.3743	0.3987	0.8024	0.7196
Max error (Weibull)	0.052	0.070	0.061	0.077^a	0.077	0.075
Max error (logistic)	0.088	0.064	0.064	0.088	0.116	0.100
Max error (normal)	0.066	0.043	0.042	0.068	0.094	0.079
Max error (lognormal)	0.070	0.034	0.047	0.062	0.068	0.059
Critical value (95%)	0.071	0.071	0.071	0.071	0.071	0.071
Critical value (99%)	0.085	0.085	0.085	0.085	0.085	0.085
RMSE (Weibull)	0.0286	0.0247	0.0459	0.0345	0.0625	0.0573
RMSE (logistic)	0.0210	0.0192	0.0305	0.0299	0.0681	0.0602
RMSE (normal)	0.0154	0.0269	0.0364	0.0258	0.0613	0.0422
RMSE (lognormal)	0.0136	0.0132	0.0210	0.0204	0.0428	0.0316

^a Boldface entries indicate that they exceed the critical value at 95% confident level in the Kolmogorov- Smirnov test

References

- [1] Chang, T. P. 2009. Performance evaluation for solar collectors in Taiwan. *Energy*, 34: 32-40.
- [2] Shu, N., Kameda, N., Kishida, Y., and Sonoda, H. 2006. Experimental and theoretical study on the optimal tilt angle of photovoltaic panels. *Journal of Asian Architecture and Building Engineering*, 5: 399-405.
- [3] Gueymard, C. A. 2003. Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment. *Solar Energy*, 74, 5: 355-379.
- [4] Gueymard, C. A. 2003. Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high-quality measurements. *Solar Energy*, 74, 5: 381-395.
- [5] Kudish, A. I. and Ianetz, A. 1996. Analysis of daily clearness index, global and beam radiation for Beer Sheva, Israel: Partition according to day type and statistical analysis. *Energy Conversion and Management*, 37, 4: 405-416.
- [6] Bendt, P., Collares-Pereira, M., and Rabl, A. 1981. The frequency distribution of daily insolation values. *Solar Energy*, 27: 1-5.
- [7] Saunier, G. Y., Reddy, T. A., and Kumar, S. 1987. A monthly probability distribution function of daily global irradiation values appropriate for both tropical and temperate locations. *Solar Energy*, 38: 169-177.
- [8] Akuffo, F. O. and Brew-Hammond, A. 1993. The frequency distribution of daily global irradiation at Kumasi. *Solar Energy*, 50: 145-154.
- [9] Tovar, J., Pozo-Vazquez, D., Batlles, J., Lopez, G., and Munoz-Vicente, D. 2004. Proposal of a function for modeling the hourly frequency distributions of photo-synthetically active radiation. *Theoretically*

- cal and Applied Climatology*, 79: 71-79.
- [10] Hollands, K. G. T., and Huget, R. G. 1983. A probability density function for the clearness index with applications. *Solar Energy*, 30: 235-253.
- [11] Tovar, J., Olmo, F. J., and Alados-Arboledas, L. 1998. One-minute global irradiance probability density distributions conditioned to the optical air mass. *Solar Energy*, 62: 387-393.
- [12] Babu, K. S. and Satyamurty, V. V. 2001. Frequency distribution of daily clearness indices through generalized parameters. *Solar Energy*, 70: 35-43.
- [13] Assuncao, H. F., Escobedo, J. F., and Oliveira, A. P. 2007. A new algorithm to estimate sky condition based on 5 minutes-averaged values of clearness index and relative optical air mass. *Theoretical and Applied Climatology*, 90: 235-248.
- [14] Ettoumi, F. Y., Mefti, A., Adane, A., and Bouroubi, M. Y. 2002. Statistical analysis of solar measurements in Algeria using beta distributions. *Renewable Energy*, 26: 47-67.
- [15] Jurado, M., Caridad, J. M., and Ruiz, V. 1995. Statistical distribution of the clearness index with radiation data integrated over five minute intervals. *Solar Energy*, 55: 469-473.
- [16] Soubdhan, T., Emilion, R., and Calif, R. 2009. Classification of daily solar radiation distributions using a mixture of Dirichlet distributions. *Solar Energy*, 83: 1056-1063.
- [17] Liu, C. M. 1996. Climate and weather in Taiwan. Formosa Folkways Publishing Company, Taipei.
- [18] Chang, T. P. 2010. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy* in press, doi: 10.1016/j.apenergy. 2010.06.018.
- [19] Carta, J. A., Ramirez, P., and Velazquez, S. 2009. A review of wind speed probability distributions used in wind energy analysis Case studies in the Canary Islands. *Renewable and Sustainable Energy Reviews*, 13: 933-955.
- [20] Ucar, A. and Balo, F. 2009. Investigation of wind characteristics and assessment of wind-generation potentiality in Uludag-Bursa, Turkey. *Applied Energy*, 86: 333-339.
- [21] Hogg, R. V., and Craig, A. T. 1988. Introduction to Mathematical Statistics. Fourth edition, Jwang Yuan Publishing Company, Taipei.
- [22] Sulaiman, M. Y., Akaak, A. M., Wahab, M. A., Zakaria, A., Sulaiman, Z. A., and Suradi, J. 2002. Wind characteristics of Oman. *Energy*, 27: 35-46.