

Review Article

Statistical modeling of the daily global solar radiation in Queensland, Australia

Mohammed Mohammed El Genidy

*Department of Mathematics and Computer Science, Faculty of Science,
Port Said University, Port Said City, Port Said Governorate, 42522 Egypt*

Received: 2 March 2018; Revised: 23 June 2018; Accepted: 20 August 2018

Abstract

Daily global solar radiation (DGSR) is a renewable energy source that cannot be depleted. Many researchers have used the sample method to collect the dataset of DGSR to obtain the statistical characteristics of the population. However, since this is difficult, descriptive statistics and statistical inference became the primary objective of obtaining the characteristics of the statistical population. In this study, a statistical model combined with generalized extreme value (GEV) distribution is proposed to represent the dataset of DGSR. The moments method, the Kolmogorov-Smirnov test, and the properties of GEV were performed to estimate the parameters of GEV and check the validity of the estimated parameters with the actual dataset of DGSR. Nonlinear regression and multiple nonlinear regression of GEV with its corresponding days during a year were produced. Eventually, a flowchart was designed to obtain the close probability distribution of the DGSR.

Keywords: generalized extreme value distribution, moments method, Kolmogorov-Smirnov test, multiple nonlinear regression

1. Introduction

Many researchers are interested in solar energy and its importance. It is known that solar energy is generated by radiation from nuclear reactions within the sun and it is considered as one of the largest sources of light and heat that is distributed on the earth according to the proximity to the equator. Cooling and heating systems are key prominent examples of the solar energy in the distillation and purification of drinking water and in reducing the costs that individuals pay to electricity companies.

The generalized extreme value (GEV) is the most widely applied statistical distribution for climate modeling, whereby the daily global solar radiation (DGSR) is one of its apt applications. The three-parameter GEV distribution was also proved to be efficient in describing rainfall, wind speeds, annual floods, wave heights, and snow depths to name but a few ecological phenomena.

Nonetheless, researchers have faced serious difficulties in obtaining a prediction equation of the DGSR. They have noticed that the sampling method turns into statistical inference methods that rely on their ability to make statistical decisions. Thus, they need to select an appropriate statistical model for the terminal objective of the significant study. Here lies the importance of the estimation methods.

In particular, estimation methods are used in mathematical statistics for measuring the parameters of a statistical population. Among these methods, there are the statistical estimations, which include the method of moments and the maximum likelihood function. In addition, the estimation problems were studied to propose a pattern of similarity-based clustering algorithm and to maintain its application in solar radiation evaluation. Along with the Pearson R model, Bhardwaj *et al.* (2013) used the hidden Markov model for the extraction of shape-based clusters from meteorological parameters. The generalized fuzzy model was then applied to accurately evaluate solar radiation. Emad and Adam (2013) believed that designing a forecasting model that applies artificial neural networks was meant to evaluate the monthly average DGSR in Qena in upper Egypt whereby good concordance between the evaluated and measured data of global solar irradiation was recorded.

*Corresponding author
Email address: drmmg2016@yahoo.com

Remarkably, Hahne *et al.* (2014) examined a comparison between nonlinear and linear regression methods for simultaneous, independent, and proportional myoelectric control of wrist motions with two degrees of freedom offline with electro-myographic signals that were gained from ten healthy subjects and one person who had a congenital upper limb deficiency.

Empirical models were created to detect the correlation of the mean DGSR and the air temperature data in six Algerian cities. As recorded by Mecibah, Boukelia, Tahtah, and Gairaa (2014), the results supported the results of previous studies and showed that sunshine based models were commonly more accurate than air temperature based models. Furthermore, the two regression models that were applied to the prepared generalized models could predict the monthly mean global solar radiation in other Algerian locations.

On the other hand, linear, quadratic, and cubic curve estimations were constructed to evaluate the global solar radiation in the eastern Mediterranean region in such a manner that regression models were monthly and annually updated with the MINITAB program. In addition, Teke and Yildirim (2014) compared monthly and general models by statistical test methods. Sanchez-Lorenzo, Calbó, and Wild (2013) checked the expansion of a new dataset of the surface solar radiation in Spain from the records of global solar radiation beginning in the early 1980s. Linares-Rodriguez, Ruiz-Arias, Pozo-Vazquez, and Tovar-Pescador (2013) suggested an optimized artificial neural network ensemble model to evaluate the DGSR in large regions by applying clear-sky assessments and satellite images as input data to reinforce the same results of current models even on cloudy days. The Bristow-Campbell model evaluated the DGSR evaluation in the Tibetan Plateau while a tentative method analyzed the DGSR. Pan, Wu, Dai, and Liu (2013) concluded that the standardized Bristow-Campbell fit the Tibetan-Plateau as it produced moderately accurate global solar radiation assessments. Besharat, Dehghan, and Faghhih (2013) chronologically compiled and surveyed the inclusive global solar radiation model, then sorted it out into four classes on the basis of meteorological parameters as model information.

In previous studies, Bang and Shin (2016) made a comparison study on different common methods for non-crossing multiple linear quantile regression to describe practical instruction of their applications. Moghaddasi, Bazzazi, and Aalianvari (2016) introduced an adaptive neuro-fuzzy inference system to anticipate the ground inflow rate into the Amir Kabir tunnel in Iran, whereby a sample of 110 datasets containing most of the influential parameters on ground inflow rate was set to develop the ground inflow rate forecasting model.

Backbreak is an unwanted phenomenon of blasting that causes the instability of mine walls, inefficiency of drilling, and unstable machinery. Still, Faradonbeh, Monjezi, and Armaghani (2015) created a new artificial intelligence known as the genetic programming to forecast backbreak.

Due to the large number of bird types in the world, it is difficult to predict the maximum number of migratory bird types during a limited number of migration years. However, El Genidy (2017) believed that a multiple nonlinear regression model could accurately perform the job.

In the current study, the moments method and Kolmogorov-Smirnov test were applied to the dataset of DGSR in

Queensland, Australia, to estimate the parameters of the GEV distribution. Also, the multiple nonlinear regression model of GEV was used to predict the amount of DGSR on any day of a year.

2. Materials and Methods

2.1 Daily global solar radiation (DGSR)

DGSR is the sum of solar radiation during a day represented on a horizontal surface. Common values of the DGSR prove to range from 1 to 35 MJ/m² (megajoules per square meter) where the values are usually higher in the clear sun during the summer and lower through the winter or on very cloudy days.

2.2 Dataset

In this study, a dataset of DGSR during 274 days in 2016 in Queensland, Australia was derived by the Australian Bureau of Meteorology, 2016 (Figure 1). Figure 1 shows the daily global radiation during 2016, which was measured by (MJ/m²) in the city of Queensland, Australia.

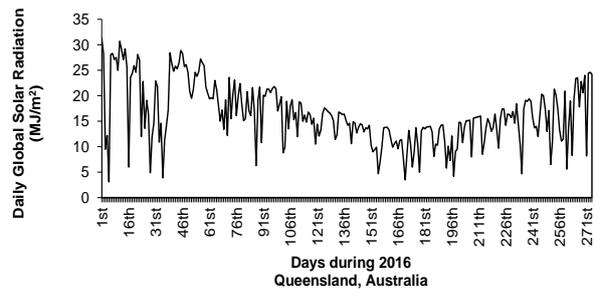


Figure 1. Daily global solar radiation in Queensland, Australia during 2016.

2.3 Software

The software used in this study included Mathematica 4, version number 4.0.1.0 (Wolfram Research, Inc. 100 Trade Center Drive Champaign, IL USA) and SPSS 16.0 (IBM SPSS software, New York, USA).

2.4 Generalized extreme value distribution

Suppose that X is a continuous random variable representing the dataset of DGSR, then the probability density function of GEV is defined as

$$f(x) = \frac{1}{\theta} g(x)^{\alpha+1} e^{-g(x)} ; x \in \begin{cases} (-\infty, \lambda - \frac{\theta}{\alpha}] & \text{if } \alpha < 0 \\ (-\infty, \infty) & \text{if } \alpha = 0, \\ [\lambda - \frac{\theta}{\alpha}, \infty) & \text{if } \alpha > 0 \end{cases} \quad (1)$$

where $\theta > 0, \lambda, \alpha \in \mathbb{R}$; \mathbb{R} is the set of real numbers,

$$g(x) = \begin{cases} \left(1 + \frac{\alpha(x-\lambda)}{\theta}\right)^{\frac{-1}{\alpha}}, & \text{if } \alpha \neq 0 \\ e^{-\frac{(x-\lambda)}{\theta}}, & \text{if } \alpha = 0 \end{cases}, \quad (2)$$

where the cumulative distribution function is

$$F(x) = e^{-g(x)}, \quad (3)$$

such that α is the shape parameter, θ is the scale parameter, and λ is the location parameter.

2.5 Moments method

Likewise, the moment generating function is defined as

$$M(t) = E(e^{tx}) = \int_{-\infty}^a e^{tx} f(x) dx, \text{ where } a = \lambda - \frac{\theta}{\alpha} \text{ and } \alpha < 0,$$

and thus

$$M(t) = e^{\left(\lambda - \frac{\theta}{\alpha}\right)t} \left[1 + \sum_{n=1}^{\infty} \frac{(\theta t)^n}{n! \alpha^n} \Gamma(1 - n\alpha) \right]. \quad (4)$$

The first moment around zero reads:

$$M'(0) = \text{Mean} = -\frac{\theta}{\alpha} + \lambda + \frac{\theta\Gamma(1-\alpha)}{\alpha}, \quad (5)$$

whereas the second moment around zero is

$$M''(0) = \left(-\frac{\theta}{\alpha} + \lambda\right)^2 + \frac{\theta^2\Gamma(1-2\alpha)}{\alpha^2} + \frac{2\theta\left(-\frac{\theta}{\alpha} + \lambda\right)\Gamma(1-\alpha)}{\alpha}. \quad (6)$$

Therefore, the significant equations will be

$$\text{Variance} = M''(0) - (M'(0))^2 = \frac{\theta^2(\Gamma(1-2\alpha) - (\Gamma(1-\alpha))^2)}{\alpha^2}. \quad (7)$$

$$\text{Median} = \lambda + \frac{\theta\left((\text{Ln}(2))^{-\alpha} - 1\right)}{\alpha}. \quad (8)$$

2.6 Multiple nonlinear regression

Estimating the multiple nonlinear regression of $F^*(x, y)$ on $F(x)$ and $F(y)$ is performed by the software as follows:

<< Statistical `Nonlinear Fit`

data = { { $F(x_1), F(y_1), F(x_1, y_1)$ }, { $F(x_2), F(y_2), F(x_2, y_2)$ }, ..., { $F(x_n), F(y_n), F(x_n, y_n)$ } }

Nonlinear Fit

[data, $F^*(x,y) = a [F(x) F(y)]^3 + b [F(x) F(y)]^2 + c F(x) F(y) + d$, { x, y }, { a, b, c, d }]

$F^*(x,y) = a [F(x) F(y)]^3 + b [F(x) F(y)]^2 + c F(x) F(y) + d$, where a, b, c , and d satisfy that the model

$F^*(x,y) = a [F(x) F(y)]^3 + b [F(x) F(y)]^2 + c F(x) F(y) + d$, best fit the $(F(x), F(y), F^*(x, y))$ data quadrant in the least-squares sense, Mathematica software is applied to carry out this process.

3. Results and Discussion

This section starts by juxtaposing the relevant results of the significant previous studies with the results of the current study in terms of (a) estimation methods and (b) distribution models of the dataset. Then hopefully, focusing on the points of contribution foregrounds the significance of the current study in terms of (a) the Kolmogrov-Smirnov test for goodness of fit and (b) multiple nonlinear regression of $F(x, k)$.

3.1 Estimation methods

El Damsesy, El Genidy, and El Gazar (2014) made use of the maximum likelihood function to estimate the reliability and failure rate of the electronic system with the help of the mixture Lindley distribution. Similarly, Hosking, Wallis, and Wood (1985) applied the method of probability weighted moments to estimate the GEV distribution.

On the other hand, in the current study, the moments method and the properties of the GEV distribution were applied to accurately produce a nonlinear regression of the DGSR in Queensland, Australia during 2016. Undoubtedly, these methods emphasize the accuracy of estimations, and support the validity of their corresponding predictions.

3.2 Distribution models of dataset

Hasan, Radi, and Kassim (2012) presented the modeling of extreme temperature, using the GEV distribution in a case study of Penang, Malaysia. Smith (1989) based his study on the extreme value analysis of environmental time series, which was an application to trend detection in ground level ozone. Joe (1994) managed to apply the multivariate extreme value distribution, along with some other applications, to environmental data. Sousa, Martins, Alvim-Ferraz, and Pereira (2007) revealed how multiple linear regression and artificial neural networks, based on principal components, could predict ozone concentrations. Later, in a case study of Hong Kong, Gong, and Ordieres-Meré (2016) performed a study on the prediction of the daily maximum ozone threshold exceedances by preprocessing and ensembling artificial intelligence techniques.

However, the current study adopted the GEV distribution to represent the dataset of DGSR in Queensland, Australia, 2016. The dataset was then mathematically treated until significant results were obtained. The statistics of the dataset of DGSR and the daily maximum temperature during 2016 (Figure 1, a-d) are listed in Table 1, where Q1, Q2, Q3, and Q4 are first quartile, second quartile, third quartile, and fourth quartile respectively. Hence, from Equation 5, Equation

Table 1. Descriptive statistics for both the daily global solar radiation and the daily maximum temperature during 2016.

Statistics	Solar 2016	Temp 2016	Statistics	Solar 2016	Temp 2016
Average	16.3365	22.69964	Skewness	0.257266	-0.11614
Variance	34.87582	28.99106	Kurtosis	-0.34372	-0.60405
SD ^a	5.905575	5.384335	Q1	12.4	18.4
Min	3.1	8.9	Q2	15.6	23.2
Max	31.3	36	Q3	20.1	26.9
Mode	14.6	26.7	Q4	31.3	36
Median	15.6	23.2	AD ^b	4.713485	4.569724

a. SD indicates standard deviation
 b. AD indicates average deviation

7, and Equation 8, we can extract the mean, median, variance, and standard deviation as follows:

$$\text{Mean} = \lambda + \theta \left(\frac{\Gamma(1-\alpha)-1}{\alpha} \right) = 16.3365, \tag{9}$$

$$\text{Median} = \lambda + \frac{\theta \left((\ln 2)^{-\alpha} - 1 \right)}{\alpha} = 15.6, \tag{10}$$

$$\text{Variance} = \frac{\theta^2 \left(\Gamma(1-2\alpha) - (\Gamma(1-\alpha))^2 \right)}{\alpha^2} = 34.87582, \tag{11}$$

$$\text{Standard Deviation} = \frac{\theta}{\alpha} \sqrt{\Gamma(1-2\alpha) - (\Gamma(1-\alpha))^2} = 5.905575. \tag{12}$$

In a similar vein, from Equation 9 and Equation 10, θ^2 reads.

$$\theta^2 \cong \frac{0.541769 \alpha^2}{[\Gamma(1-\alpha) - (\ln 2)^{-\alpha}]^2}. \tag{13}$$

Then,

$$\frac{\Gamma(1-2\alpha) - (\Gamma(1-\alpha))^2}{[\Gamma(1-\alpha) - (\ln 2)^{-\alpha}]^2} \cong 64.37. \tag{14}$$

Thus,

$$\alpha = -0.07481199, \lambda = 13.781, \theta = 5.03146.$$

The graphical representations of $F(x)$ and $f(x)$ are shown in Figure 2 and Figure 3, respectively. In particular, Figure 2 presents the relationship between the values of the DGSR "x" and its cumulative distribution function "F(x)". Correspondingly, Figure 3 foregrounds the relationship between the values of the DGSR "x" and its density function $f(x)$.

Moreover, the mean and standard deviation can be separately obtained from $f(x)$ in the following manner.

$$\text{Mean} = E(x) = \int_{3.1}^{31.3} x f(x) dx = \int_{3.1}^{31.3} \frac{1}{\theta} g(x)^{\alpha+1} e^{-g(x)} dx = 15.7243 \cong 16$$

$$E(x^2) = \int_{3.1}^{31.3} x^2 f(x) dx = \int_{3.1}^{31.3} \frac{1}{\theta} g(x)^{\alpha+1} e^{-g(x)} dx = 280.377$$

$$\text{Standard Deviation} = \sqrt{E(x^2) - (E(x))^2} \cong 5.8$$

Here, the equation is as follows:

$$g(x) = \left(1 + \frac{\alpha(x-\lambda)}{\theta} \right)^{\frac{-1}{\alpha}} \text{ and}$$

$$\alpha = -0.07481199, \lambda = 13.781, \theta = 5.03146.$$

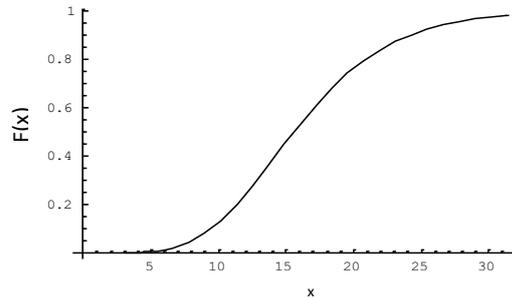


Figure 2. Cumulative distribution function.

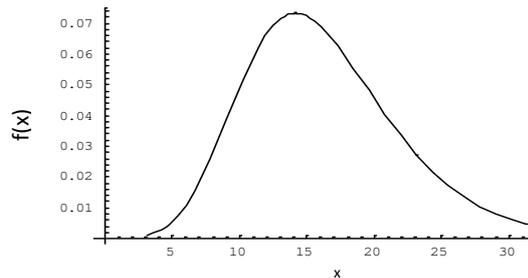


Figure 3. Probability density function $f(x)$ of GEV.

The value of median can be also extracted from $F(x)$ in this way:

$$F(x) = e^{-g(x)} = 0.50$$

$$\frac{1}{e^{(1+0.0148688(-13.781+x))^{13.3668}}} = 0.50$$

$$\text{Median} = x = 15.6506.$$

As a result, with the use of the new model $F(x)$ of GEV, the values of mean, standard, and median deviations are extracted, and they prove close to the real dataset DGSR.

With regard to computations of quartiles Q_n and percentiles P_n , they are spelled out in accordance with the theoretical model and the real dataset (Tables 2 and 3).

A comparison between the values computed by the real dataset and those computed by the theoretical model is tabulated in Table 4.

3.3 Kolmogrov-Smirnov test for goodness of fit

This test is to compare the experimental cumulative frequency $S_n(x)$ with the cumulative distribution function of the DGSR. A stepwise experimental cumulative frequency function $S_n(x)$ is defined as follows:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ \frac{k}{n} & x_k \leq x < x_{k+1} \\ 1 & x \geq x_n \end{cases}$$

where x_1, x_2, \dots, x_n are the observed values of the order dataset X and n is the sample size. Figure 4 shows the step-function plot of $S_n(x)$ with $F(x)$. The maximum differences $D_{x(n)}$ is defined as follows:

$D_{x(n)} = \max_x |F(x) - S_n(x)|$, where D_n^α is the critical value at significance level α .

The simple linear regression estimation of $F(x)$ will be:

$$F^*(x) = 0.0037x - 0.0013, \text{ where } R^2 = 0.9966.$$

Figure 4 stresses the relationship between the empirical cumulative frequency $S_n(x)$ and the theoretical $F(x)$ such that the coefficient of the determination confirms the accuracy of the estimate. Hence, the Kolmogrov-Smirnov test helps reach the results arranged in Table 5. In this regard, the GEV distribution is an acceptable model for the dataset of DGSR at the significance levels of $\alpha = 0.20, 0.10, 0.05,$ and 0.01 . It can be used to create a prediction model $F^*(x, k)$ for the joint distribution function $F(x, k)$ as a multiple nonlinear regression on $F(x)$ and cumulative distribution function $S(k)$ of all days of a year.

3.4 Using the Anderson-Darling test to compare the fit of GEV and Weibull distributions with the actual dataset DGSR

The parameters of the Weibull distribution are estimated by the SPSS program. The resulting values of the scale and shape parameters are equal to 17.798 and 3.108, respectively, whereby the adjusted R square is 0.996. On the other hand, GEV and Weibull distributions are considered extreme distributions, the reason for which the Anderson-Darling test for goodness of fit is meant to determine the most appropriate distributions for the actual dataset DGSR.

Table 2. Computations of the values of the daily global solar radiation X , base on the theoretical model and the real dataset at their quartile values.

N	F(Q _n)	X (Theoretical)	X (Dataset)
1	0.25	12.1575	12.4
2	0.50	15.6506	15.6
3	0.75	20.3511	20.1
4	0.96≈1	31.9629	31.3

Table 3. Computations the values of daily global solar radiation X by theoretical model and real dataset at their percentile values.

n	F(P _n)	X (Theoretical)	X (Dataset)
1	0.10	9.71285	9.4300
2	0.15	10.6351	10.500
3	0.20	11.4287	11.260
4	0.25	12.1575	12.400
5	0.30	12.8535	13.500
6	0.35	13.5368	13.855
7	0.40	14.2223	14.400
8	0.45	14.9227	14.985
9	0.50	15.6506	15.600
10	0.55	16.4198	16.300
11	0.60	17.2471	16.900
12	0.65	18.1546	17.890
13	0.70	19.1734	19.110
14	0.75	20.3511	20.100
15	0.80	21.7676	21.340
16	0.85	23.5734	23.405
17	0.90	26.1126	24.770
18	0.95	30.5159	27.100
19	0.96≈1	31.9630	27.548

Table 4. Comparison the produced values of quartiles and percentiles from the real dataset and from the theoretical model of the daily solar radiation at significance level $\alpha = 0.01$.

Statistics measures	Daily solar radiation	Correlation p-value	Test of normality Shapiro-Wilk p-value	ANOVA test p-value	Decision
Quartiles	Real Dataset	1	0.547	0.977	There are no significance differences
	Theoretical Model		0.573		
Percentiles	Real Dataset	0.991	0.398	0.787	There are no significance differences
	Theoretical Model		0.124		

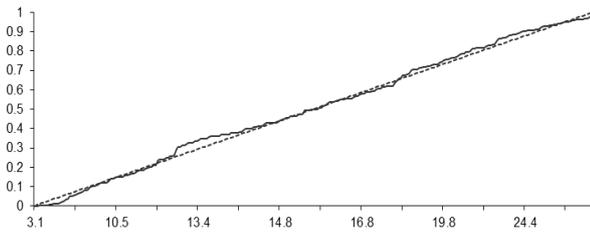


Figure 4. Empirical cumulative frequency $S_n(x)$ versus theoretical $F(x)$.

Table 5. Computations of D_n^α , $D_{n(x)}$ with the decisions of accepting the generalized extreme value distribution for the dataset X at significance level $\alpha = 0.20, 0.10, 0.05, 0.01$

$D_n^\alpha, n = 274$	$D_{n(x)} = 0.048$	Decision
$D_{274}^{0.20} = 0.065$	$D_{n(x)} < D_{274}^{0.20}$	Accept
$D_{274}^{0.10} = 0.074$	$D_{n(x)} < D_{274}^{0.10}$	Accept
$D_{274}^{0.05} = 0.082$	$D_{n(x)} < D_{274}^{0.05}$	Accept
$D_{274}^{0.01} = 0.098$	$D_{n(x)} < D_{274}^{0.01}$	Accept

Let N be the size of dataset DGSR, and $F(x)$ be the cumulative distribution function, then the Anderson-Darling test statistic (A-D Statistic) is defined as follows:

$$A = -N - \frac{1}{N} \sum_{i=1}^N (2i - 1) [\text{Ln}(F(x_i)) + \text{Ln}(1 - F(x_{N-i+1}))],$$

where $i = 1, 2, \dots, N$.

As for the extreme distributions, such as Weibull and GEV, the adjusted statistic of A will be defined as:

$$A^* = A \left(1.0 + \frac{0.2}{\sqrt{N}} \right).$$

On applying the equations of A and A^* on Weibull and GEV distributions, A-D Statistic of Weibull distribution proves equal to 1.73325 and its adjusted statistic is 1.754196. Yet, the A-D Statistic of the GEV distribution is 1.073301 and its adjusted statistic is equal to 1.086269. As a result, the adjusted statistic of GEV distribution is the least, and GEV distribution is more suitable than Weibull distribution for the actual DGSR dataset.

3.5 Multiple nonlinear regression of $F(x, k)$

The multiple nonlinear regression $F^*(x, k)$ on $F(x)$ and $S(k)$ is rendered in the following form of multiple nonlinear regression model.

$$F^*(x, k) = a [F(x) S(k)]^3 + b [F(x) S(k)]^2 + c F(x) S(k) + d. \tag{15}$$

Estimating the parameters $a, b, c,$ and d can change the equation this way:

$$F^*(x, k) = -3.574 \cdot 10^{-8} [F(x) S(k)]^3 + 4.469 \cdot 10^{-8} [F(x) S(k)]^2$$

$$+ F(x) S(k) + 1.126 \cdot 10^{-9} \approx F(x) S(k), \tag{16}$$

where $k = 1, 2, \dots, 274$.

Figure 5 is a flowchart, summarizing the comparison method of all of the different distribution functions, whereby researchers can use any other distribution in the comparison process, in addition to the GEV distribution which was applied in the current study. Also, $F^*(x, k) = (k/n)$ ($0.0037x - 0.0013$), where $k = 1, 2, \dots, 274$ and $n = 274$, thus $F^*(x)$ and $F^*(x, k)$ can be predicted by the values of the DGSR x and during k days in a year. Accordingly, $F^*(x, k)$ will reach its maximum value at $(x, k) = (24.6, 273)$, and its minimum at $(x, k) = (3.1, 5)$.

3.6 Validity of the prediction model $F^*(x, k)$

To validate the prediction model $F^*(x, k)$, researchers should (a) calculate the values of both $F(x, k)$ and $F^*(x, k)$, where x is the value of DGSR, and (b) use their values in SPSS program, then (c) apply the homogeneity-of-variances test and the Mann-Whitney's test for $F(x, k)$ and $F^*(x, k)$. In this regard, the p-value of the homogeneity test is equal to 1, whereas the p-value of the Mann-Whitney's test is 0.941. Thus, $F(x, k)$ and $F^*(x, k)$ are equal and homogenous.

4. Conclusions

In this study, the GEV distribution and the DGSR were introduced. The moments method, test fitting of quartiles and percentiles were performed to estimate the parameters of generalized extreme value distribution. Furthermore, the Kolmogorov-Smirnov test was used to fit the dataset with the mentioned distribution after estimating their parameters. Non-linear regression of the cumulative distribution functions for the DGSR were thus derived and the multiple nonlinear regression of the joint distribution function of the DGSR, with its corresponding days, were obtained and checked vis-à-vis the real dataset.

Significant Statement

The statistical method in this study provides accurate results of the daily global solar radiation, the point that is highly beneficial to applications of environmental sciences. It also helps uncover the critical areas of the daily global solar radiation. Now, researchers can work out a new statistical method of sustainable energy sources and possibly of other new distributions. Moreover, the results of the current study enable us to better understand the causes of high temperatures in the atmosphere, and whether it is due to environmental pollution or to increased solar radiation.

Acknowledgements

I would like to thank the Australian Bureau of Meteorology, 2016, for providing online climate data of Queensland, Australia. I also thank the contributing authors in this field of study.

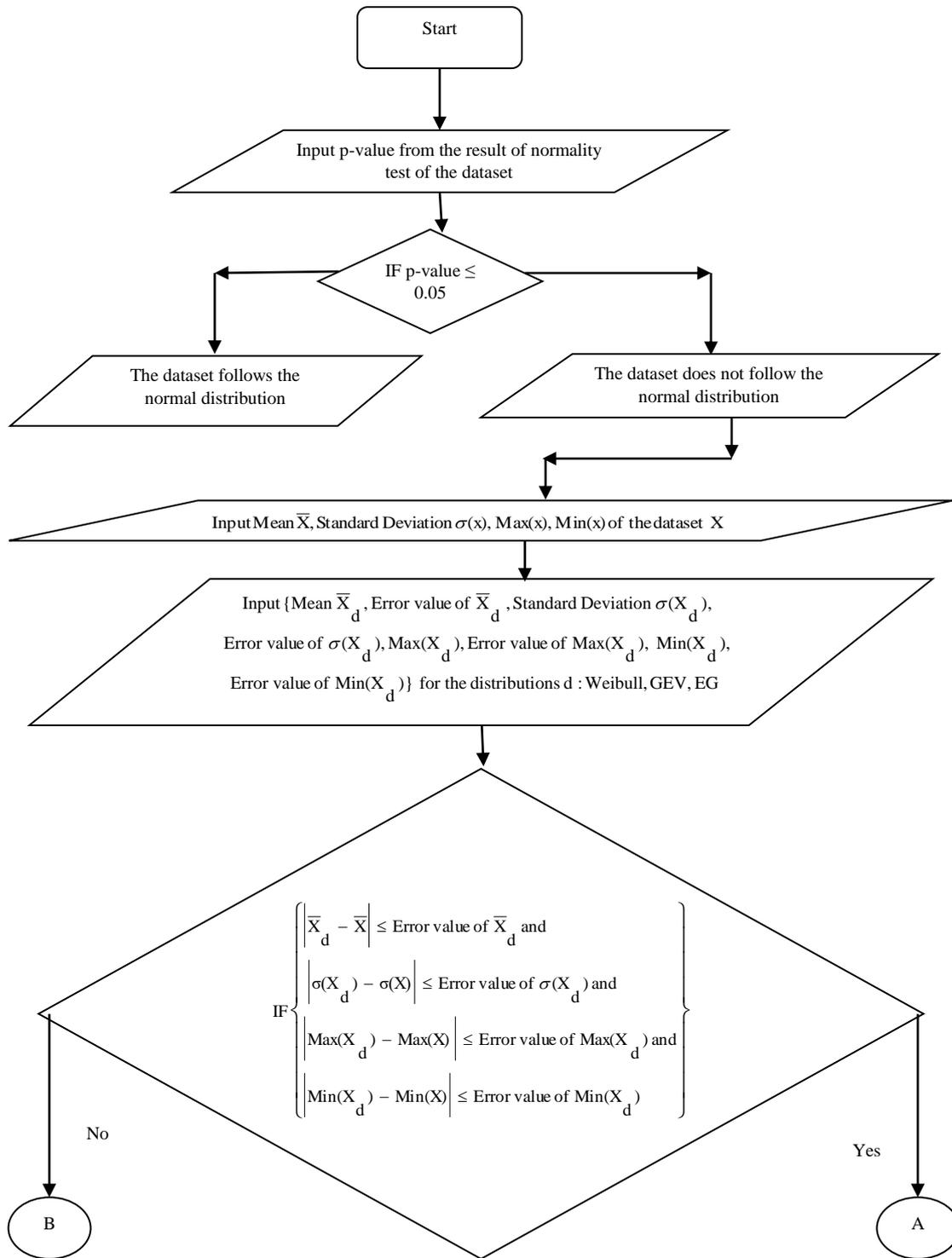


Figure 5. Flowchart comparison of different distributions.

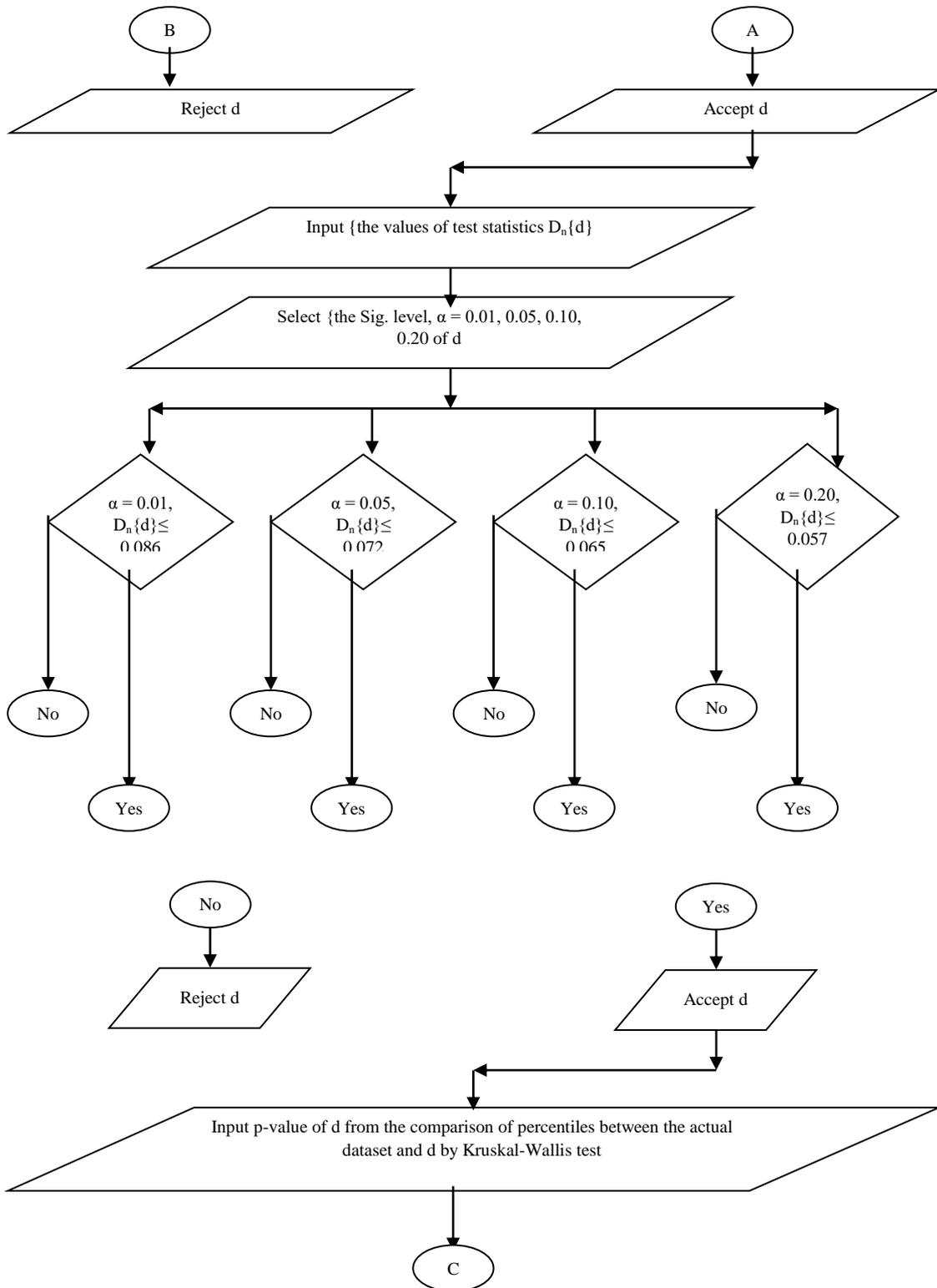


Figure 5. Continued

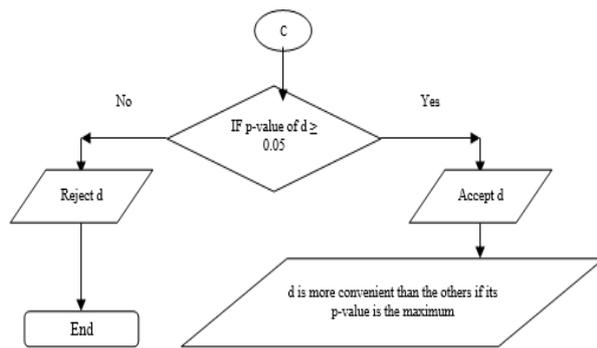


Figure 5. Continued

References

- Bang, S., & Shin, S. J. (2016). A comparison study of multiple linear quantile regression using non-crossing constraints. *Korean Journal of Applied Statistics*, 29(5), 773-786. doi:10.5351/KJAS.2016.29.5.773
- Besharat, F., Dehghan, A. A., & Faghih, A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, 21(C), 798-821. doi:10.1016/j.rser.2012.12.043
- Bhardwaj, S., Sharma, V., Srivastava, S., Sastry, O. S., Bandyopadhyay, B., Chandel S. S., & Gupta, J. R. P. (2013). Estimation of solar radiation using combination of hidden Markov model and generalized fuzzy model. *Solar energy*, 93, 43-54. doi:10.1016/j.soler.2013.03.020
- El Damsesy, M. A., El Genidy, M. M., & El Gazar, A. M. (2014). Estimation of a system performance in Pareto distribution with two independent random variables. *Journal of Applied Sciences*, 14(21), 2854-2856. doi:10.3923/jas.2014.2854.2856
- El Genidy, M. M. (2017). Multiple nonlinear regression model for the maximum number of migratory bird types during migration years. *Communications in Statistics-Theory and Methods*, 46(16), 7969-7975. doi:10.1080/03610926.2016.1167909
- Emad, A. A., & Adam, M. E. (2013). Estimate of global solar radiation by using artificial neural network in Qena, Upper Egypt. *Journal of Clean energy technologies*, 1(2), 148-150. Retrieved From <http://pdfs.semanticscholar.org/ffee/8ab23640ed76b5916be6d07166f08aed24ee.pdf>
- Faradonbeh, R. S., Monjezi, M., & Armaghani, D. J. (2015). Genetic programming and nonlinear multiple regression techniques to predict backbreak in blasting operation. *Engineering with computers*, 32(1), 123-133. doi:10.1007/s00366-015-0404-3
- Gong, B., & Ordieres-Meré, J. (2016). Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environmental Modelling and Software*, 84, 290-303. doi:10.1016/j.envsoft.2016.06.020
- Hahne, J. M., Bießmann, F., Jaing, N., Rehbaum, H., Farina, D., Meinecke, F. C., & Parra, L. C. (2014). Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(2), 269-279. doi:10.1109/TNSRE.2014.2305520
- Hasan, H. B., Ahmad Radi, N. F. B., & Kassim, S. B. (2012). Modeling of extreme temperature using generalized extreme value distribution a case study of Penang. *Proceedings of the World Congress on Engineering I*, 181-186. Retrieved from http://www.iaeng.org/publication/WCE2012/WCE2012_pp181-186.pdf
- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *American Society for Quality*, 27(3), 251-261. Retrieved from <http://hydrology.agu.org/wp-content/uploads/sites/19/2016/09/Hosking1985.pdf>
- Joe, H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, 22(1), 47-64. doi:10.2307/3315822
- Linares-Rodriguez, A., Ruiz-Arias, J. A., Pozo-Vazquez, D., & Tovar-Pescador, J. (2013). An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. *Energy*, 61, 636-645. doi:10.1016/j.energy.2013.09.008
- Mecibah, M. S., Boukelia, T. E., Tahtah, R., & Gairaa, K. (2014). Introducing the best model for estimation of the monthly mean daily global solar radiation on a horizontal surface (Case study: Algeria). *Renewable and Sustainable Energy Reviews*, 36(C), 194-202. doi:10.1016/j.rser.2014.04.054
- Moghaddasi, M. R., Bazzazi, A. A., & Aalianvari, A. (2016). Prediction of groundwater inflow rate using non linear multiple regression and ANFIS models: A case study of Amirkabir tunnel in Iran. *Conference 2-4 November: International Black Sea Mining and Tunneling Symposium*, 134-141.
- Pan, T., Wu, S., Dai, E., & Liu, Y. (2013). Estimating the daily global solar radiation spatial distribution from diurnal temperature ranges over the Tibetan Plateau in China. *Applied Energy*, 107, 384-393. doi:10.1016/j.apenergy.2013.02.053
- Sanchez-Lorenzo, A., Calbó, J., & Wild, M. (2013). Global and diffuse solar radiation in Spain: Building a homogeneous dataset and assessing their trends. *Global and Planetary Change*, 100, 343-352. doi:10.1016/j.gloplacha.2012.11.010
- Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4(4), 367-377. doi:10.1214/ss/1177012406
- Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C. M., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling and Software*, 22(1), 97-103. doi:10.1016/j.envsoft.2005.12.002
- Teke, A., & Yildirim, H. B. (2014). Estimating the monthly global solar radiation for eastern Mediterranean region. *Energy Conversion and Management*, 87, 628-635. doi:10.1016/j.enconman.2014.07.052