Research Article

A New High Accurate Estimation Method for Evaluating the Daily Solar Energy by Nested Percentiles Algorithm

1Mohammed Mohammed El Genidy and 2Doaa Abd El-Shafi Abd El-Rahman

1Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt
2Department of Basic Science, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

Abstract

Background and Objective: Owing to the wide use of Weibull distribution in solar energy analysis, researchers are constantly looking for easy and applicable methods for estimating its parameters with the smallest margin of error. This study was aimed to find a new algorithm that computes the Weibull parameters easily and accurately. Materials and Methods: Weibull distribution was performed to model the daily solar energy and corresponding maximum temperature in Queensland, Australia over a year. A new method, called nested percentiles algorithm was suggested for estimating the parameters. Anderson-Darling test was applied to measure the fitness of the suggested estimator. Results: Nested percentile algorithm gave high significant results as indicated by Anderson-Darling test. The parameters estimated by the new algorithm provided the best fit of the used datasets over than other results produced by EasyFit program. Conclusion: Nested Percentile Algorithm made two-parameter Weibull distribution more accurate than three-parameter distributions such as; modified Weibull distribution. It also gave more accurate results than EasyFit program.

Key words: Solar energy, Weibull distribution, nested percentiles algorithm, Anderson-darling test

Received: March 10, 2019  Accepted: May 07, 2019  Published: 


Corresponding Author: Mohammed Mohammed El Genidy, Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Port Said, Egypt Tel: +201124774443/+20663657602 Fax: +20663657601

Copyright: © 2019 Mohammed Mohammed El Genidy and Doaa Abd El-Shafi Abd El-Rahman. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.
INTRODUCTION

Daily Solar Energy (DSE) is considered one of the most important resources of renewable energy. It has a variety of irreplaceable applications, such as electricity generation. Australian government has set a goal of meeting about 20% of their needs of electricity using DSE by 2020. Taking into consideration that Queensland, Australia provides about 4% of electricity using DSE. The present study derived a statistical modeling for the amount of DSE and Daily Maximum Temperature (DMT) in Queensland, Australia during 2015. Weibull distribution with two parameters was considered in this study to fit the dataset and a new algorithm is performed to estimate the parameters. Two-parameter Weibull Distribution (WD) is the most frequently used in renewable energy application. This is because of its flexibility and ease of use. The wide use of WD gives greater significance to the estimation of its parameters. It has been used to model wind energy in a number of studies.

Four methods were carried out to determine Weibull parameters using dataset of daily wind speed in Jiwani town over 10 years in a named study. The methods were Maximum Likelihood (ML), Moments, empirical and energy pattern factor methods. The ML method led to the best result. Some other researchers applied 7 methods for estimating the parameters of WD. Namely Graphical, ML, Energy Pattern Factor, Moments, Empirical, Modified ML and Equivalent Energy methods, where the dataset of wind speed in the northeast region of Brazil was used.

A comparison of 5 numerical methods in terms of estimation were made by analyzing data of wind speed measurements in the district of Kousseri for 28 years. These methods were the Graphical, ML, Modified ML, Empirical and Energy Pattern factor methods. The comparison showed that energy pattern factor method gave the most accurate results. Carrillo et al. analyzed data of wind speed in Galicia using WD and part density energy method was presented for parameters estimation.

A previous study presented modeling of the amount of pollutants in the Ozone Layer using Generalized Extreme Value Distribution. The method of Moments was used to estimate the parameters and its results were compared with other software results, namely EasyFit program. It was found that the results of Moments method were more accurate. A different study dealt with Least Squares and ML methods using two-parameter in WD. Still, the two parameters of WD were adopted with estimation method of ML by El Genidy and Hebeshy. Later on, Quartiles-Moments method was proposed with Exponentiated Gumbel Maximum distribution using data of the solar radiation and maximum temperature. Furthermore, ML and uniformly minimum variance unbiased methods were introduced to estimate the parameters of Lindley distribution.

Because of the wide use of WD in data analysis, the estimation of Weibull parameters has received a great deal of attention. The aim of this study was to suggest a new high accurate estimator of parameters called Nested Percentiles Algorithm (NPA). It was performed to model datasets of DSE and DMT in Queensland, Australia over a year.

MATERIALS AND METHODS

In this study, a new method called NPA was produced to estimate the amount of solar energy. This method was characterized by high accuracy and ease in estimating the parameters of WD more than those that were used in the previous studies.

Software programs, EasyFit and Mathematica were carried out to fulfill the research requirements and extract the numerical results. Also, Anderson-Darling test (AD) was applied to measure the fitness of the results with the actual data.

Dataset: The selected dataset in this study included 2 sets, one is DSE in Queensland, Australia for 365 days (Fig. 1a). The other one is of DMT in the same area during the same period (Fig. 1b). They were recorded by the Bureau of Meteorology in Australian government, during 2015 and published online in the website of http://www.bom.gov.au/climate/data/index.shtml.

Figure 1a-b presents the values of DSE and DMT during the 365 days of 2015 in Queensland, Australia.

Daily solar energy: The DSE is the total amount of the solar energy that reaches the earth’s surface per day and it's measured by Mega joules per square meter (MJ m⁻²). Typically, its value varies from 1-35 MJ m⁻² (Fig. 1a).

Daily maximum temperature: The highest temperature recorded in a certain period of time (24 h), is the daily maximum temperature, where the measure unit is degrees Celsius (Fig. 1b).

Methodology: A new algorithm to estimate WD's parameters called Nested Percentile Algorithm was used in this study. Firstly, the range of the cumulative function, [0, 1] was divided into 100 small equal intervals. Then, percentiles function was used to find the value of DSE, x at which
Fig. 1(a-b): Dataset of (a) DSE and (b) DMT during 2015 at Queensland, Australia

DSE: Daily solar energy, DMT: Daily maximum temperature

F(x_i) = 0.01, F(x_i) = 0.02, F(x_i) = 0.03,..., F(x_100) = 1. That means 98 equations in terms of α and λ were obtained. Every 2 equations, P_i = F(x_i) and P_i = F(x_i), for all i = 1-49, were solved together to get 49 approximated values of α and λ. Note that, P_i is the front percentile of order i, P_i is the last percentile of order li, while x_i and x_i are the values of DSE at fi and li, respectively. Similarly, the solutions of the 2 equations of DMT, F(y_i) and F(y_i) were obtained.

Weibull distribution: Consider x be a random variable of DSE in Queensland, Australia for a year 2015 and has a WD. The Probability Density Function (PDF) for WD is:

\[ f(x) = \frac{\alpha x^{\alpha-1} e^{-\left(\frac{x}{\lambda}\right)^\alpha}}{\lambda^\alpha} \quad ; \quad x \geq 0, \lambda > 0, \alpha > 0 \quad (1) \]

The Cumulative Distribution Function (CDF) for WD is:

\[ F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^\alpha} \quad (2) \]

where, \( \alpha \) is a shape parameter and \( \lambda \) is a scale parameter.

Let CDFs at any two points of DSE dataset \( x_i \) and \( x_i \) are:

\[ P_n = 1 - e^{-\left(\frac{x_i}{\lambda_n}\right)^\alpha_n} \quad \text{and} \quad P_n = 1 - e^{-\left(\frac{x_i}{\lambda_n}\right)^\alpha_n} \quad \text{such that} \quad P_n = 1 - P_n \]

where, \( i = 1, 2, 3,... \) N/2 and N is an even number. Then:

\[ \alpha_i = \frac{\ln(\ln(P_n / \ln(1 - P_n)))}{\ln(x_i / x_i)} \quad (3) \]

And:

\[ \lambda_i = x_i / [-\ln(P_n)]^{\alpha_i} \quad (4) \]

Iterate the above steps for each pair of equations that satisfy \( P_n = 1 - P_n \). Thus, N/2 different values for each \( \alpha \) and \( \lambda \) were obtained. Then, the averages are calculated to get the final value of the two parameters:

\[ \alpha = \frac{\sum_{i=1}^{N/2} \alpha_i}{N/2} \quad (5) \]

\[ \lambda = \frac{\sum_{i=1}^{N/2} \lambda_i}{N/2} \quad (6) \]
\[
\lambda_i = \frac{\sum_{i=1}^{N/2} \lambda_{ni}}{N/2}
\]  

Similarly for DMT:

\[
\alpha_i = \frac{\ln(\ln(P_i / \ln(1 - P_i))]}{\ln(y_i / y_2)}
\]

\[
\lambda_i = y_i / (1 - P_i)^{\alpha_i}
\]

\[
\alpha_j = \frac{\sum_{i=1}^{N/2} \alpha_{ni}}{N/2}
\]

\[
\lambda_j = \frac{\sum_{i=1}^{N/2} \lambda_{ni}}{N/2}
\]

Algorithm:

Begin
Rem "N is an even number"
Read N
m = N/2
For i = 1 to m-1
Read x_i, y_i, y_2, P_i
\[
\alpha_{ni} = \frac{\ln(\ln(P_i / \ln(1 - P_i))]}{\ln(x_i / y_2)}
\]
\[
\lambda_{ni} = x_i / (1 - P_i)^{\alpha_{ni}}
\]
\[
\lambda_{ni} = y_i / (1 - P_i)^{\alpha_{ni}}
\]
S1 = S1 + \alpha_{ni}
S2 = S2 + \lambda_{ni}
S3 = S3 + \alpha_{ni}
S4 = S4 + \lambda_{ni}
Next i
\[
\alpha_i = S1 / m
\]
\[
\lambda_i = S2 / m
\]
\[
\alpha_j = S3 / m
\]
\[
\lambda_j = S4 / m
\]
Print \(\alpha\), \(\lambda\), \(\alpha\), \(\lambda\)
End

Steps of NPA are shown in the Fig. 2 to solve 98 percentiles equations which were divided into 49 forward equations and similarly 49 backward equations. The Algorithm was applied on the two datasets of DSE (x) and DMT (y), given that F(x) and F(y) are the CDF of WD and P_n and P_b represent the front percentiles and last percentiles, respectively.

Test of fitting the distribution: The AD test was performed to fit the datasets with WD approximated by NPA. It's defined as:

\[
A^2 = \sum_{i=1}^{n-1} \left( \frac{2}{n-1} \right) \left[ \ln(F_i(x)) + \ln\left(1 - F_i(x_{n-i})\right) \right] / n - n
\]

The adjusted AD test statistic of WD is given by:

\[
A' = A^2\left(1 + \frac{0.3}{n}\right)
\]

Software: The following programs have been carried out on the used dataset of this study:

- EasyFit professional, version 5.5 (February, 2010), Math Wave Technologies (http://www.mathwave.com)
- Mathematica 8, version 8.0.1 (March 2011), Wolfram Mathematica (http://www.wolfram.com)

RESULTS

Parameters estimation method: Applying NPA on the two datasets, DSE and DMT, estimated values of parameters of WD were obtained accurately in Table 1 and 2.

Calculation process expressed in Table 1, it demonstrated the estimation of the parameters of WD for the dataset DSE.
Table 1: Numerical values of $\alpha_s$ and $\lambda_s$ of DSE by NPA

<table>
<thead>
<tr>
<th>$P_s = F(y_s)$</th>
<th>$x_s$</th>
<th>$y_s$</th>
<th>$\alpha_s$</th>
<th>$\lambda_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00001</td>
<td>2.4018</td>
<td>32.0996</td>
<td>5.3831</td>
<td>20.3877</td>
</tr>
<tr>
<td>0.0001</td>
<td>2.4182</td>
<td>32.0964</td>
<td>4.4207</td>
<td>19.4235</td>
</tr>
<tr>
<td>0.001</td>
<td>2.582</td>
<td>32.0636</td>
<td>3.5091</td>
<td>18.4851</td>
</tr>
<tr>
<td>0.01</td>
<td>3.792</td>
<td>31.672</td>
<td>2.8868</td>
<td>18.6605</td>
</tr>
<tr>
<td>0.02</td>
<td>4.336</td>
<td>31.072</td>
<td>2.6740</td>
<td>18.6563</td>
</tr>
<tr>
<td>0.03</td>
<td>5.852</td>
<td>30.516</td>
<td>2.8738</td>
<td>19.7208</td>
</tr>
<tr>
<td>0.04</td>
<td>6.556</td>
<td>30.088</td>
<td>2.8663</td>
<td>20.0109</td>
</tr>
<tr>
<td>0.05</td>
<td>7.54</td>
<td>29.28</td>
<td>2.9980</td>
<td>20.3064</td>
</tr>
<tr>
<td>0.06</td>
<td>8.052</td>
<td>28.732</td>
<td>3.0060</td>
<td>20.3541</td>
</tr>
<tr>
<td>0.07</td>
<td>8.796</td>
<td>28.452</td>
<td>3.0677</td>
<td>20.6848</td>
</tr>
</tbody>
</table>

Average: $\alpha_s = 3.009 \quad \lambda_s = 20.5364$

DSE: Daily solar energy, NPA: Nested percentiles algorithm

Table 2: Numerical values of $\alpha_s$ and $\lambda_s$ of DMT by NPA

<table>
<thead>
<tr>
<th>$P_s = F(y_s)$</th>
<th>$y_s$</th>
<th>$\alpha_s$</th>
<th>$\lambda_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00001</td>
<td>8.0116</td>
<td>34.9653</td>
<td>26.3981</td>
</tr>
<tr>
<td>0.0001</td>
<td>8.1165</td>
<td>34.7979</td>
<td>25.7417</td>
</tr>
<tr>
<td>0.001</td>
<td>9.1648</td>
<td>34.6749</td>
<td>25.5271</td>
</tr>
<tr>
<td>0.01</td>
<td>12.528</td>
<td>33.372</td>
<td>6.254</td>
</tr>
<tr>
<td>0.02</td>
<td>13.356</td>
<td>32.772</td>
<td>5.8667</td>
</tr>
<tr>
<td>0.03</td>
<td>13.692</td>
<td>31.724</td>
<td>5.6482</td>
</tr>
<tr>
<td>0.04</td>
<td>14.712</td>
<td>31.144</td>
<td>5.8237</td>
</tr>
<tr>
<td>0.05</td>
<td>15.02</td>
<td>30.8</td>
<td>5.6638</td>
</tr>
<tr>
<td>0.06</td>
<td>15.268</td>
<td>30.264</td>
<td>5.5788</td>
</tr>
<tr>
<td>0.07</td>
<td>15.448</td>
<td>29.9</td>
<td>5.4533</td>
</tr>
</tbody>
</table>

Average: $\alpha_s = 4.9855 \quad \lambda_s = 25.1932$

DMT: Daily maximum temperature, NPA: Nested percentiles algorithm

Table 3: Estimated Weibull parameters of DSE and DMT

<table>
<thead>
<tr>
<th>Dataset</th>
<th>EasyFit program</th>
<th>NPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSE</td>
<td>$\alpha_s = 2.8292$</td>
<td>$\alpha_s = 3.00929$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_s = 20.599$</td>
<td>$\lambda_s = 20.53638$</td>
</tr>
<tr>
<td>DMT</td>
<td>$\alpha_s = 5.2706$</td>
<td>$\alpha_s = 4.9855$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_s = 24.919$</td>
<td>$\lambda_s = 25.1932$</td>
</tr>
</tbody>
</table>

using NPA. Hence, the final estimated value of the shape parameter $\alpha_s$ was 3.009 and 20.5364 for the scale parameter $\lambda_s$.

Similarly, Table 2 presented the numerical results of the estimated parameters of WD for the dataset DMT using NPA. As illustrated, the final parameters were 4.9855 for the shape parameter $\alpha_s$ and 25.1932 for the scale parameter $\lambda_s$.

EasyFit program was implemented to obtain approximated values of parameters of WD for each set and results of NPA were compared with software results. Clearly, in Table 3, the results of NPA are close to EasyFit program’s results.

Table 3 showed the parameters of WD that were produced by EasyFit program and the others estimated by NPA for the datasets of DSE and DMT.

Validation of results: The NPA’s results were checked out by 2 ways, graphically and by using Hypothesis test as following:

Graphically: The CDF of WD in Eq. 2 with the parameters created by NPA is almost identical to CDF, of the actual dataset as shown in Fig. 3a-f.

Figure 3 is clarified the accuracy of NPA. As it presented the following graphs:

- Figure 3a is the CDF of the actual dataset of DSE
- Figure 3b represents the CDF of WD with the parameters of $\alpha_s = 3.009$ and $\lambda_s = 20.5364$, which were estimated by NPA for DSE dataset
- Figure 3c shows the CDF of WD obtained by EasyFit program for DSE with the parameters of $\alpha_s = 2.8292$ and $\lambda_s = 20.599$
- Figure 3d is the CDF of the actual dataset of DMT
- Figure 3e represents the CDF of WD with the parameters $\alpha_s = 4.9855$ and $\lambda_s = 25.1932$, which calculated using NPA for DMT dataset
- Figure 3f shows the CDF of WD obtained by EasyFit program for DMT with the parameters of $\alpha_s = 5.2706$ and $\lambda_s = 24.919$

So, Fig. 3 showed that the CDFs of actual datasets are almost identical to the CDFs of WD created by NPA and so is the CDFs resulted by EasyFit program.

Hypothesis test: According to AD test, NPA is a suitable method for estimation of parameters. The AD test for each dataset was defined as:

$H_0$: The dataset follows the estimated WD by NPA

$H_1$: The dataset doesn’t follow the estimated WD by NPA
Since the critical value of AD test for WD at significance level 0.01 is 1.959. Then, for DSE, A* for the results generated by NPA is less than the critical value as shown in Table 4, which means $H_0$ is accepted (i.e., the dataset of DSE follows WD with the estimated parameters using NPA). In addition to that, it is less than A* for the results given by EasyFit program. This also means NPA gives more accurate results than EasyFit program. Thus, NPA relocated WD from the 13th rank to 4th rank for DSE and also for DMT from the 10th rank to 6th rank.

Table 4 illustrated the accuracy of NPA over EasyFit program through the AD test statistic values, A* and their ranks, where the applied distribution is WD.


Table 4: AD test values and their ranks of NPA and EasyFit program

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Significance level</th>
<th>Critical value</th>
<th>A*</th>
<th>State</th>
<th>Rank</th>
<th>A*</th>
<th>State</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSE</td>
<td>0.01</td>
<td>1.959</td>
<td>0.86012</td>
<td>Accepted</td>
<td>13</td>
<td>0.64308</td>
<td>Accepted</td>
<td>4</td>
</tr>
<tr>
<td>DMT</td>
<td>0.01</td>
<td>1.959</td>
<td>1.8486</td>
<td>Accepted</td>
<td>10</td>
<td>1.47249</td>
<td>Accepted</td>
<td>6</td>
</tr>
</tbody>
</table>

AD: Anderson-darling test, A*: Results generated by NPA

**DISCUSSION**

In this study, high-accurate parameters of WD were estimated using NPA, for DSE and DMT in Queensland, Australia in 2015. The graphical representations of the CDF as well as AD test were derived to validate the results that were found very satisfactory.

In previous studies, some researchers presented a new modified Weibull distribution (MWD) in which the estimation of its three parameters were obtained by ML method\(^1\). The shape, scale and location parameters were 0.023, 0.062 and 0.356, respectively. Others also estimated the three parameters of MWD using ML estimation side by side with least squares estimation, where the parameters were 0.03065 for the shape parameter, 0.05 for the scale parameter and 0.219493 for the location parameter\(^2\). When MWD was applied to the datasets of DSE and DMT, which were used in this study, by EasyFit program, the three parameters were 3.4547, 22.712 and -2.1041 for DSE and 4.072, 20.036, 4.8369 for DMT. Thus, the statistics values of AD test A* were 0.7541 and 1.594, respectively, but, in this study, NPA gave a less value of A*, which was equal to 0.64308 for DSE and 1.47249 for DMT. Therefore, WD with NPA estimator was more accuracy than MWD with ML method.

The WD was estimated using ML, approximate ML and Bayes methods to fit two datasets, investigated already\(^3\). The first set fit WD with 5.505 for the shape parameter and 214.131 for the scale parameter while the second one fit WD with 5.049 and 424.574 for the shape and the scale parameters, respectively. On the other hand, in the present study, EasyFit program generated WD with parameters of 2.8292 and 20.599 for DSE and 5.2706 and 24.919 for DMT. The resulted WD fit the used datasets with A* of 0.86012 and 1.8486 for DSE and DMT, respectively, which were still larger than A* produced by NPA estimator. That means WD with NPA estimator was more accuracy over than the same distribution that was estimated by other methods.

Previously, a study discussed the estimation of generalized Gamma distribution (GGD) parameters with ML method to model a dataset of carbon fibers\(^4\). The parameters were k = 4.0735, \(\alpha = 3.34592\) and \(\beta = 3.09225\). Also, Gamma Distribution (GD) was performed to fit a dataset of lifetime of the series system\(^5\). It had shape parameter of 0.972 and scale parameter of 0.111. Moreover, the parameters of GGD were k = 0.9554, \(\alpha = 6.9919\) and \(\beta = 2.378\) for DSE with a rank of 29th and k = 0.98743, \(\alpha = 19.643\) and \(\beta = 1.1262\) for DMT with a rank of 25th. In addition, GD of DES had 7.6954 and 2.378 for the shape and the scale parameters, respectively with a rank of 30th. The same distribution had 20.412 and 1.1262 for DMT’s parameters with a rank of 28th. On the other side, NPA along with WD had ranks of 4th for DSE and 6th for DMT. That means NPA made WD have precedence in ranking compared to the mentioned distributions.

Gualandi and Toscani\(^6\) dealt with lognormal distribution and resulted in parameters of 4.9 and 1.2 for mean and variance, respectively. This distribution had mean of 0.44071 and variance of 2.8252 for DSE with 37th rank compared to WD estimated by NPA which ranked 4th according to EasyFit program. Its parameters were 0.23604 and of 3.1085 with rank of 29th for DMT, while WD estimated by NPA ranked 6th. It is clear that the new suggested technique, NPA, was really much better. A study analyzed data of air pollutants in a traffic-congested area in India using four-parameter Burr distribution and ML estimation\(^7\). The resulted parameters were k = 0.3471, \(\alpha = 21.03\), \(\beta = 16.279\) and \(\gamma = 0.91619\). Another one also estimated the four parameters of Burr distribution by using ML method and gave parameters\(^8\) of k = 1141.10, \(\alpha = 1.51\), \(\beta = 293.76\) and \(\gamma = 0.07\). Meanwhile, DSE gave Burr’s parameters of k = 24.312, \(\alpha = 3.7747\), \(\beta = 56.408\) and \(\gamma = -3.6668\) which ranked 7th. For the same dataset, WD that was estimated by EasyFit program ranked 13th while WD given by NPA ranked 4th. Overall, NPA improved the rank of WD significantly with respect to Burr distribution. And so it was with DMT.

As a recommendation, NPA can be applied on probability distributions with three or four parameters such as generalized extreme value and Burr distributions with different datasets. Its algorithm can be programmed to generalize its usage in the statistical analysis.
CONCLUSION

In this study, a new parameter estimator, called Nested Percentiles Algorithm was created to get the values of daily solar energy and maximum temperature. This estimator proved its accuracy over other estimation methods such as; Maximum Likelihood, Moments and EasyFit program. Moreover, it made two-parameter Weibull distribution more accurate than many four-parameter distributions such as Burr distribution which saves time and effort.

SIGNIFICANCE STATEMENT

Modeling and predicting the amount of solar energy is an important issue as it has become the most important alternative source to fuel. Since the goodness of fit relies mainly on the estimation method of parameters. In this study, NPA was performed in order to enable the researchers to get high accurate results. This algorithm can be applied to data in several science fields and can be programmed by following the mentioned algorithm.

ACKNOWLEDGMENT

It is a pleasure to thank Australian Bureau of Meteorology for providing online climate data of Queensland, Australia. In addition, I thank the contributing authors in this field of study.

REFERENCES