Fitting Wind Speed to a 3-Parameter Distribution Using Maximum Likelihood Technique

Okumu Otieno Kevin, Troon John Benedict, Samuel Muthiga Ngaga

Department of Mathematics and Physical Sciences, Maasai Mara University, Narok, Kenya

Email address:

kevinotieno15@gmail.com (O. O. Kevin), troonbenedict@gmail.com (T. J. Benedict), muthiga@mmarau.ac.ke (S. M. Ngaga)

Abstract: Kenya is one of the countries in the world with a good quantity of wind. This makes the country to work on technologies that can help in harnessing the wind with a vision of achieving a total capacity of 2GW of wind energy by 2030. The objective of this research is to find the best three-parameter wind speed distribution for examining wind speed using the maximum likelihood fitting technique. To achieve the objective, the study used hourly wind speed data collected for a period of three years (2016 – 2018) from five sites within Narok County. The study examines the best distributions that the data fits and then conducted a suitability test of the distributions using the Kolmogorov-Smirnov test. The distribution parameters were fitted using maximum likelihood technique and model comparison test conducted using Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values with the decision rule that the best distribution relies on the distribution with the smaller AIC and BIC values. The research showed that the best distribution is the gamma distribution with the shape parameter of 2.071773, scale parameter of 1.120855, and threshold parameter of 0.1174. A conclusion that gamma distribution is the best three-parameter distribution for examining the Narok country wind speed data.

Keywords: Maximum Likelihood Estimation, Wind Speed, Weibull, Gamma, Lognormal, AIC, BIC

1. Introduction

Distribution characteristics refer to the wind speed parameters like the mean, variance, standard deviation, and covariance. These parameters vary from place to place depending on various factors like the length of data observed, the site of an experiment, and time of observing the wind speed data among others. Therefore, there is a need to study the variations of these parameters for any specific site before installing the wind plant. This is only possible if there is an approved statistical distribution that has been examined and recommended for the site or region of interest. Many researchers have come out to model wind speed using different statistical distributions with 2-parameter and 3-parameter namely: log-normal distribution, Weibull distribution, gamma distribution, Rayleigh distribution, and Erlang distribution among others in most parts of the world with a good quantity of wind [4, 17].

In Kenya, the past reviews indicate that the researcher has only considered Weibull distribution with 2-parameter to be the best for study the wind speed leaving behind the 3-parameters distribution yet the 3-parameter may be best than the 2- parameter distribution for studying the wind speed for this region [4]. The best 2-parameter and 3-parameter distributions used and recommended by a good number of researchers are Weibull distribution, gamma distribution, and lognormal distribution [12]. The 2-parameter distributions involve scale and shape parameter where the scale guides on how windy the region is (statistically meaning the distribution of the wind) and the shape parameter guides on how peaked the region is (statistically indicating the most frequently expected wind speed) and the 3-parameter include the third parameter called threshold parameter which now helps in understanding the minimum expected wind speed in the region [12].

In Kenya, Narok is one of the regions with plenty of wind making it one of the places with the potential of generating more wind energy [17], and this why the wind speed data for the study was collected in Narok county. To assist the wind sector to achieve their target of 2GW of wind energy by 2030,

the stakeholders need to understand the wind speed characteristics of this region. This is because the wind speed is the most significant factor in the installation process of any wind plant. Hence there is a need to have complete information about wind speed characteristics. This can only be possible by having a recommended statistical distribution for examining the wind speed data. There are existing statistical distributions for studying wind speed data for a different specific region but the problem is that among the existing wind speed distributions there is no underlying 3-parameter distribution for studying the wind speed of Kenya. Therefore, leaving the wind industry and other investors with incomplete knowledge on the suitable distribution for examining wind speed data leading to less interest since they are lacking a control tool for their study. Therefore, this research fills this gap by analyzing the hourly wind speed data, using the maximum likelihood by fitting the three-parameter statistical distributions of Weibull, log-normal, and gamma, to enables us to choose the best distribution among the three that fits the data perfectly to help for studying wind speed.

The three-parameter distributions used are represented as follows;

1.1. Weibull Distribution with 3 Parameters

The Weibull statistical distribution with three parameters is given by [1, 2, 6, 18, 26, 27];

$$f(u) = \left(\frac{b}{p}\right)\left(\frac{u-w}{p}\right)^{b-1} exp\left[-\left(\frac{u-w}{p}\right)\right]^b \ (b, u > 0, p > 1) \ (1)$$

and the cumulative density function is given as below

$$F(u) = 1 - exp\left[-\left(\frac{u-w}{p}\right)^b\right]$$

Where:
u is the wind speed
b is the shape parameter
p is the scale parameter measured in m/s
w is the thresh-hold parameter

1.2. Lognormal Distribution with 3 Parameters

This distribution has three parameters namely scale parameter, shape parameter, and thresh-hold parameter also known as location parameter. The probability density function and the cumulative density function are given by the below equations [3, 14, 19, 29],

$$f(p) = \frac{1}{(p-y) \, k\sqrt{2\pi}} \, exp\left[-\left(\frac{\ln(p-y)-v}{2k}\right)\right]^2 \quad (2)$$

The cumulative function is given as;

$$F(p) = \frac{1}{k\sqrt{2\pi}} \int_{-\infty}^{p} \frac{1}{(p-y)} \, exp\left[-\left(\frac{\ln(p-y)-v}{2k}\right)\right]^2$$

Where:
v > 0 is the scale parameter.
k > 0 is the shape parameter.

y is the thresh-hold parameter, also referred to as the location parameter.
p ≥ is the wind speed

1.3. Gamma Distribution with 3 Parameters

From the past studies [11, 13, 20, 28], the gamma function is given as follows;

$$f(y, z, q, t) = \frac{y t^{z-1}}{\Gamma(z/t) q^z} \, exp\left[-\left(\frac{y}{q}\right)^t\right] \ (z, y, q, t > 0) \ (3)$$

Where:
q is the scale parameter.
z is shape parameters
t is the thresh-hold parameter
The Γ is defined by

$$\Gamma(v) = \int_{0}^{\infty} y^{v-1} \, exp^{-y} dy, v > 0$$

2. Method

2.1. Maximum Likelihood Estimation Method (MLE)

According to wind sector prospect Kenya [21], maximum likelihood estimation techniques are more precise than the method of moments estimation, least-square estimation, and also graphical methods and can be applied in many problems since it has a strong intuitive appeal and it yields a reasonable estimator. He also stated that the maximum likelihood method is widely used because it is more precise especially when dealing with large samples since it yields an excellent estimator when the sample is large. Maximum likelihood function $\widehat{M}$ of M is a solution to the maximization problem given as [9, 10, 12, 15, 16, 22-25].

$$\widehat{M} = argMax \, Ln(M: x1, x2, \ldots \ldots, xN) \quad (4)$$

Where x1,...., xN represents the wind speed observations. Under suitable regularity conditions, the first-order condition is given as

$$\frac{\partial Ln(M: x1, \ldots, xN)}{\partial M} = -N + \frac{1}{M}\left(\sum_{i=1}^{N} xi\right) \quad (5)$$

These conditions are generally called the likelihood or log-likelihood equations. The first derivative or gradient of a condition (log-likelihood) solved at the point $\widehat{M}$ satisfies the following equation [10, 16, 22-25]

$$\frac{\partial Ln(M: x1, \ldots, xN)}{\partial M} = \frac{\partial Ln(\widehat{M}: x1, \ldots, xN)}{\partial M} = 0 \quad (6)$$

The log-likelihood equation that corresponds to a linear or non-linear system of P equations with P unknown parameters M1,... MP is given by

$$\frac{\partial Ln(M: x1, \ldots, xN)}{\partial M} = \left(\frac{\partial Ln(M: x1, \ldots, xN)}{\partial M1}, \ldots, \frac{\partial Ln(M: x1, \ldots, xN)}{\partial Mp}\right) = (0, \ldots, 0) \ (7)$$

Maximum likelihood is a recommended technique for many distributions because it uses the values of the distribution's

parameters that makes the data more likely than any other parameters. This is achieved by maximizing the likelihood function of the parameters given the data. Some good features of maximum likelihood estimators are that they are asymptotically unbiased since the bias tends to zero as the sample size increases and also they are asymptotically efficient since they achieve the Cramer-Rao lower bound as sample size approaches ∞ and lastly they are asymptotically normal [16, 21].

The third parameter called the threshold parameter is also known as the location parameter which determines where to shift the 3-parameter density function along the X-axis. The threshold parameter locates the distribution along the time scale and has the same units as the distribution's variable units. This third parameter is used to try to fit the data point into a straight line when the initial data do not fall on a straight line [12, 16]. After obtaining the threshold parameter, we subtract it from the original data and obtain a new data set which is then used to estimate the other two parameters (shape and scale parameters). Since the threshold parameter value is not constant, we use the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to estimate the threshold parameter. The threshold value will the lowest AIC and BIC values will be considered to be efficient and precise for further analysis. It will be subtracted from the original data set and the resulting data set will then be used for estimating the scale and shape parameters for both Weibull, Lognormal, and Gamma 3-parameter distributions using the same maximum likelihood estimates obtained for the 2-parameters.

The shape and scale for the 3-parameters are estimated as follows;

2.2. Weibull Distribution Estimates

Where is the shape parameter and p the scale parameter, the estimate for $\hat{p}$ is given as

$$\hat{p} = -\left[\left((\hat{b}-1)\sum_{i=1}^{n}\ln(u_i)\right) - \left(\sum_{i=1}^{n}u_i^{\hat{b}}\right)\right]\Big/n \quad (8)$$

The estimate for $\hat{b}$ is obtained as shown below

$$\frac{n}{b} + \frac{\sum_{i=1}^{n}\ln(u_i)}{p} - \frac{1}{p}\sum_{i=1}^{n}u_i^b\ln(u_i) = 0$$

$$\frac{n}{\hat{b}} + \frac{\sum_{i=1}^{n}\ln(u_i)}{\hat{p}} = \frac{1}{\hat{p}}\sum_{i=1}^{n}u_i^{\hat{b}}\ln(u_i)$$

$$\hat{b} = \sqrt{\frac{n}{\dfrac{\sum_{i=1}^{n}\ln(u_i)-1}{\sum_{i=1}^{n}\ln(u_i)\Big/n}[\ln(u_i)]}} \quad (9)$$

Equations (8) and (9) can be solved simultaneously for $\hat{b}$ which also obtains $\hat{p}$ subsequently [1, 7, 8, 12].

2.3. Lognormal Distribution Estimates

For the k shape parameter and v scale parameter from equation (2), their estimates are obtained as shown [2, 5, 14].

$$\hat{v} = \frac{\sum_{i=1}^{n}ln(p_i)}{n}$$

$$\hat{k} = \sqrt{\frac{\sum_{i=1}^{n}\left(ln(p_i) - \frac{\sum_{i=1}^{n}ln(p_i)}{n}\right)^2}{n}} \quad (10)$$

2.4. Gamma Distribution Estimates

Using z as the shape parameter and q as the scale parameter in equation (3), their respective estimates are obtained by [11, 20];

$$\hat{z} = \ln(q) + \frac{\Gamma'(\hat{z})}{\Gamma(\hat{z})} = \frac{\sum_{i=1}^{n}\ln y_i}{n}$$

$$\hat{q} = \frac{\bar{y}}{\hat{z}} = \frac{\sum_{i=1}^{n}y_i}{\sum_{i=1}^{n}\ln(y_i)} \quad (11)$$

2.5. Test of Goodness of Fit

After analyzing the data using the three statistical distribution, it is important to verify the suitability and the accuracy of the distribution by performing the goodness of fit test using Kolmogorov-Smirnov statistics which simply tell how good the data followed any specific distributions among the three distribution namely gamma, Weibull and log-normal distribution to help have the most precise and reliable distribution [3, 18, 19]. And the goodness test criteria will be examined using Akaike's Information Criterion and Bayesian Information Criterion.

2.5.1. Kolmogorov-Smirnov Test

This is a two-sample test with the advantage that it does not depend mostly on the underlying cumulative distribution function being tested and also applies only to continuous distributions which in this case is applicable since we are only investigating the continuous statistical distributions [19]. It is calculated as,

$$D^* = \max(|F_1(t) - F_2(t)|) \quad (12)$$

Where:
$F_1(t)$ is the proportion of t1 values less than or equal to t
$F_2(t)$ is the proportion of t2 values less than or equal to t
H_0: The data follows a specified distribution
H_1: The data do no follow the specified distribution
The smaller the test statistic the better the fit.

2.5.2. Akaike's Information Criterion (AIC)

The Akaike's Information Criterion is calculated as shown below

$$AIC = -2logL(p) + 2w \quad (13)$$

Where log L (P) defines the value of the maximized log-likelihood objective function for a model with w parameters. A smaller AIC value represents a better fit.

2.5.3. Bayesian Information Criterion (BIC)

The Bayesian Information Criterion is calculated as below

$$BIC = -2logL(p) + wlogM \qquad (14)$$

Where log L (P) represents the values of the maximized log-likelihood objective function for a model with w parameters fit M data points. A smaller Bayesian Information Criterion value indicates a better fit (best model for fitting the data)

2.6. Model/Distribution Comparison Test

This was done using the comparison between the Akaike's Information Criterion and the Bayesian Information Criterion for the two distributions whereby the distribution with the smallest Akaike's and Bayesian Information Criterion values will be picked as the best.

3. Results

3.1. Maximum Likelihood Analysis: Threshold Estimation

The three parameters are the threshold, shape, and scale parameters. To find the precise threshold parameter, we investigate different threshold values using AIC and BIC values. This is after finding the original threshold value and then performing several iterations to get the precise threshold value since it is not a constant value as shown in table 1.

From table 1, it can be observed that gamma is giving smaller AIC and BIC values under all threshold values. Therefore, gamma is the best distribution. The threshold value used for the three-parameter analysis is 0.1174. The reason for picking 0.1174 is because from the analysis of the originally collected wind data using the Minitab statistical package it is identified that gamma has a threshold value of

0.1174 m/s as shown in Figure 1.

Table 1. *Threshold values.*

Threshold value	Distribution	AIC	BIC
0.1174	Weibull	190003.8	190021.9
	Gamma	189803	189821.1
	Log-normal	195803	195821.1
0.1175	Weibull	190002.9	190021
	Gamma	189803.9	189822
	Log-normal	195817.8	195835.9
0.1180	Weibull	189999.1	190017.2
	Gamma	189809.1	189827.2
	Log-normal	195898.9	195917.1
0.1185	Weibull	189996.1	190014.3
	Gamma	189816	189834.1
	Log-normal	195997.8	196015.9
0.1190	Weibull	189994.7	190012.9
	Gamma	189826	189844.1
	Log-normal	196129.8	196147.9
0.1195	Weibull	189997.1	190015.2
	Gamma	189843.4	189816.5
	Log-normal	196347	196365.2
0.1199	Weibull	190012.5	190030.7
	Gamma	189884.1	189902.2
	Log-normal	196869	196887.1

Table 2. *Summary statistics.*

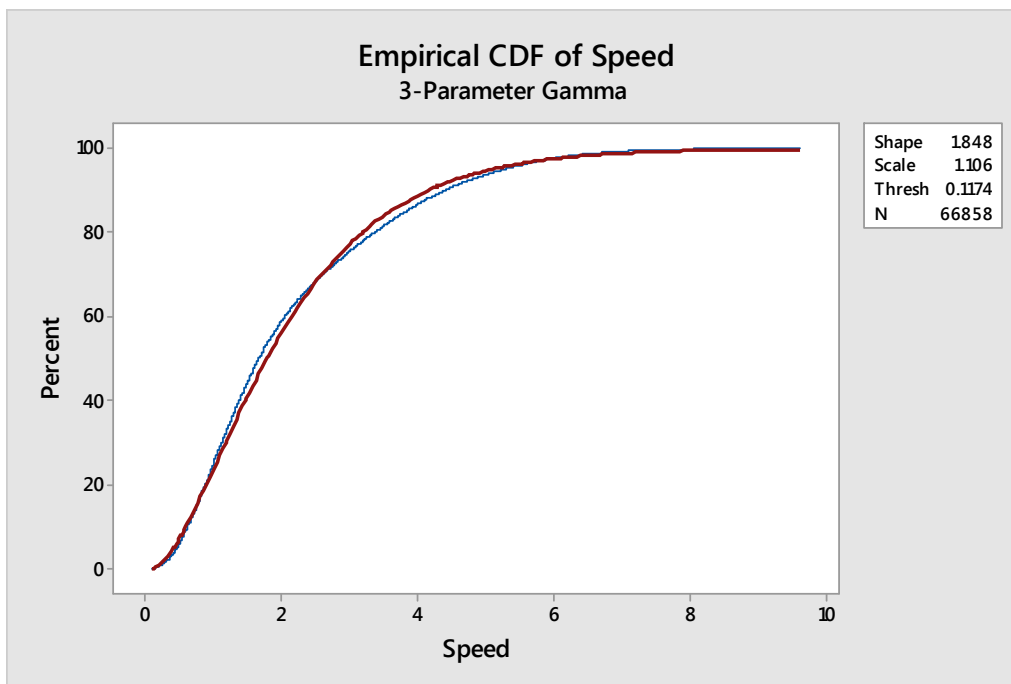|	
Min value	0.0026
Max value	5.2326
Mean	1.8484
Median	1.5026
Estimated std	1.24065
Estimated kurtosis	2.809401
Estimated skewness	0.8433485

Figure 1. *Threshold value.*

3.2. Summary Statistics Using the Threshold Value of 0.1174

Using this threshold value, we have the following summary statistics.

3.3. Test of Goodness of Fit

To confirm that the data follows either of the three distributions namely Weibull, gamma, and log-normal a goodness test of statistics was performed using Kolmogorov-Smirnov, and results are given in table 3. From table 3, it can be seen that data follows the Weibull and gamma distributions and lognormal distributions since their statistics values are less than the critical value of 0.136, making us not to reject the null hypothesis stating that the data follows either of the three distributions. The data fit gamma distributions best compared to the other two distributions since gamma has the smaller test statistic value (0.033614).

Table 3. *Test of Goodness of fit analysis.*

Statistic	Weibull	Log-normal	Gamma
Kolmogorov-Smirnov	0.042349	0.049434	0.033614

With the threshold parameter as 0.1174, the other two parameters namely shape and scale parameters are given in table 4.

3.4. Scale and Shape Estimated Parameters

Table 4. *Parameters.*

Distribution	Parameter	Estimate
Weibull	Shape	1.539360
	Scale	2.059215
Gamma	Shape	2.071773
	Scale	1.120855
Log-normal	Shape	0.359998
	Scale	0.788256

3.5. Best 3-parameter Distribution

Using the maximum likelihood method, we conclude that for the three-parameter distributions, the gamma three-parameter distribution is the best for fitting the data since it has the smallest Akaike's Information Criterion value of 189803, Bayesian Information Criterion value of 189821.1, and the smallest Kolmogorov-Smirnov test value of 0.033614. The three parameters are threshold (0.1174), shape (2.071773), and scale (1.120855) as indicated in table 5.

Table 5. *Best distribution.*

Distribution	Criteria	Estimate
Gamma	AIC	189803
	BIC	189821.1
	Parameters	
	Threshold	0.1174
	Shape	2.071773
	Scale	1.120855

4. Conclusion

From the analysis, we can conclude that using the maximum likelihood method gamma distribution with three parameters is the best distribution for fitting the Narok region wind speed data since it yields lower AIC and BIC values. The distribution is given as:

$$f(y, z, q, t) = \frac{0.1174 y^{2.071773-1}}{\Gamma(2.071773/0.1174)1.120855^{2.071773}} exp\left[-\left(\frac{y}{1.120855}\right)^{0.1174}\right] \quad (15)$$

Where gamma function in this case is treated as a continuous function depending on the wind speed data.

5. Recommendation

We recommend that for the investors or wind industry interested in studying and/or predicting the Narok wind speed, they should use the gamma distribution since it will give the best wind speed probabilities compared to the other form of distributions. We also recommend to researchers, investors, and wind industries to apply this type of gamma distribution in the examination of wind speed distribution in the other regions/parts of the country and also in other parts of the world.

References

[1] Akyuz, E., and Gamgam, H., (2017). Statistical analysis of wind speed data with Weibull and gamma distributions.

[2] Ayodele, R., Adisa, A., Munda, L., and Agee, T., (2012). Statistical analysis of wind speed and wind power potential of Port Elizabeth using Weibull parameters. *Tshwane University of Technology, Pretoria, South Africa.*

[3] Azami, Z., Khadijah, S., Mahir, A., and Sopian, K., (2009). Wind speed analysis in the east coast of Malaysia. *European journal of scientific research.* Vol. 2.

[4] Barasa, M., (2013). Wind regime analysis and reserve estimation in Kenya.

[5] Brenda, F. G., (2009). Parameter estimation for the lognormal distribution. *Brigham Young University.*

[6] Celik, H., and Yilmaz, V., (2008). A statistical approach to estimate the wind speed distribution: the case study of Gelubolu region. Pp 122-132.

[7] Chu, Y. K., and Ke, J. C., (2012). Computation approaches for parameter estimation of Weibull distribution. *Mathematical and computational applications.* Vol. 17. pp 39-47.

[8] Gungor A. and Eskin, N., (2008). The characteristics that define wind as an energy source.

[9] Gupta, R., and Biswas, A., (2010). Wind data analysis of Silchar (Assam India) by Rayleigh and Weibull methods. *Journal of mechanical engineering research.* Vol. 2, pp 10-24.

[10] Hurlin, C., (2013). Maximum likelihood estimation. *Advanced econometrics. The University of Orleans.*

[11] Johnson, W., Donna, V., and Smith, L., (2010). Comparison of estimators for parameters of gamma distributions with left-truncated samples.

[12] Maleki, F., and Deiri, E., (2007). Methods of estimation for three-parameter reflected Weibull distribution.

[13] Mert, I., and Karakus, C., (2015). A statistical analysis of wind speed using Burr, generalized gamma, and Weibull distribution in Antakya, Turkey. *Turkish Journal of electrical engineering and computer science.*

[14] Oludhe, C., (1987). Statistical characteristics of wind power in Kenya. The *University of Nairobi.*

[15] Rahayu, A., Purhadi., Sutikno., and Prastyo, D. D., (2020). Multivariate gamma regression: parameter estimation, hypothesis testing, and applications.

[16] Rambachan, A., (2018). Maximum likelihood estimates and Minimum distance estimates.

[17] Solar and wind energy resource assessment. (2008). Kenya country report.

[18] Sukkiramathi, K., Seshaiah, C., and Indhumathy, D., (2014). A study of Weibull distribution to analyze the wind speed at Jogimatti in India. Vol. 01. pp 189-193.

[19] Ulgen, K., and Hepbasli, A., (2002). Determination of Weibull parameters for wind energy analysis of Izmir, Turkey.

[20] Ulgen, K., Genc, A., Hepbasli, A., and Oturanc, G., (2009). Assessment of wind characteristics for energy generation.

[21] Wind sector prospectus Kenya. (2013). Wind energy data analysis and development program.

[22] Zheng, S., (2018). Maximum likelihood estimation. *Statistical theory II.*

[23] Rigby, R. S., Mikis, H., Gillian, D., and Fernade. (2019). Maximum likelihood estimation. 10.1201/9780429298547-11.

[24] Denuit, Michel & Hainaut, Donatien & Trufin, Julien. (2019). Maximum Likelihood Estimation. 10.1007/978-3-030-25820-7_3.

[25] Magnus, Jan. (2019). Maximum likelihood estimation. 346-366. 10.1002/9781119541219.ch15.

[26] Clement, Nadia & Lasky, Ronald. (2020). Weibull Distribution and Analysis: 2019. 1-5. 10.23919/PanPacific48324.2020.9059313.

[27] Kizilersu, Ayse & Kreer, Markus & Thomas, Anthony. (2018). The Weibull distribution. Significance. 15. 10-11. 10.1111/j.1740-9713.2018.01123.x.

[28] Dagpunar, John. (2019). The gamma distribution. Significance. 16. 10-11. 10.1111/j.1740-9713.2019.01226.x.

[29] Hart, P. (2017). Lognormal Distribution. 10.1057/978-1-349-95121-5_1089-2.