

---

# Fitting Wind Speed to a Two Parameter Distribution Model Using Maximum Likelihood Estimation Method

Okumu Otieno Kevin<sup>1</sup>, Edgar Otumba<sup>2</sup>, Alilah Anekeya David<sup>3</sup>, John Matuya<sup>1</sup>

<sup>1</sup>Department of Mathematics and Physical Sciences, Maasai Mara University, Narok, Kenya

<sup>2</sup>Department of Statistics and Actuarial Sciences, Maseno University, Kisumu, Kenya

<sup>3</sup>Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

## Email address:

Kevinotieno15@gmail.com (O. O. Kevin), eotumba@maseno.ac.ke (E. Otumba), aliladavid2010@gmail.com (A. A. David),

jmatuya@mmarau.ac.ke (J. Matuya)

## To cite this article:

Okumu Otieno Kevin, Edgar Otumba, Alilah Anekeya David, John Matuya. Fitting Wind Speed to a Two Parameter Distribution Model Using Maximum Likelihood Estimation Method. *International Journal of Statistical Distributions and Applications*.

Vol. 6, No. 3, 2020, pp. 57-64. doi: 10.11648/j.ijstd.20200603.13

**Received:** September 13, 2020; **Accepted:** September 27, 2020; **Published:** October 13, 2020

---

**Abstract:** Kenya is among the countries that are continuously investing in wind energy to meet her electricity demand. Kenya is working towards its vision 2030 of achieving a total of 2GW of energy from wind industry. To achieve this, there is a need that all the relevant data on wind characteristics must be available. The purpose of this study is, therefore, to find the most efficient two-parameter model for fitting wind speed distribution for Narok County in Kenya, using the maximum likelihood method. Hourly wind speed data collected for a period of three years (2016 to 2018) from five sites within Narok County was used. Each of the distribution's parameters was estimated and then a suitability test of the parameters was conducted using the goodness of fit test statistics, Kolmogorov-Smirnov, and Anderson-Darling. An efficiency test was determined using the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values, with the best decision taken based on the distribution having a smaller value of AIC and BIC. The results showed that the best distributions were the gamma distribution with the shape parameter of 2.47634 and scale parameter of 1.25991, implying that gamma distribution was the best distribution for modeling Narok County wind speed data.

**Keywords:** Maximum Likelihood Estimation, Wind Speed, Weibull, Gamma, Lognormal

---

## 1. Introduction

Wind speed 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 the experiment, and the time of observing the wind speed data among others. Therefore, there is a need to study the variations of these parameters for any given specific geographical 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 types of research have come out to model wind speed using different statistical distributions like 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 but not

showing great interest in Kenya even if Kenya is stated as one of the countries in Africa with the best wind [4, 16].

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 [16]. To achieve this, the investors and the wind industry needs 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 distribution for examining Narok county wind speed characteristics. Therefore, leaving the wind industry and other investors with incomplete

knowledge on the suitable distribution for examining Narok region wind data leading 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 of Narok county using the maximum likelihood by fitting the two-parameter statistical distributions of Weibull, log-normal and gamma, to enable us to choose the suitable distribution among the three that fits the data perfectly and efficiently to help for studying wind speed characteristics of Narok county.

The two-parameter distributions used are represented as follows.

### 1.1. Weibull Distribution

A study carried by [1, 2, 6, 17] on the Weibull distribution to analyze wind speed concluded that the Weibull distribution function is the best in estimating the parameters of wind speed. The Weibull distribution model applied by the study is given by;

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

Where:

$f(u)$  is the probability of observing wind speed.

$u$  is the wind speed.

$b$  is the shape factor (parameter) which has no unit but range from 1.5 to 3.0 for most wind conditions

$p$  is the value in the unit of wind speed called the Weibull scale parameter in m/s.

### 1.2. Lognormal Distribution

In studies by [3, 13, 18] they did wind speed analysis and one of the statistical distributions they used in examining the wind data was the log-normal statistical model with parameters  $v$  and  $k$ . The log-normal density function with the two parameters is given by:

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

Where:

$p$  is the log-normal random variable

$\ln(p)$  is the normal random variable

$v$  is the mean for a normal random variable

$k$  is the standard deviation for the normal random variable

### 1.3. Gamma Distribution

The probability density function of gamma random variable  $y$  in combination with two parameters  $z$  and  $q$  representing the shape and scale parameters respectively is given by [10, 12, 19].

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

The  $\Gamma$  is defined by

$$\Gamma(v) = \int_0^{\infty} y^{v-1} \exp^{-y} dy, \quad v > 0$$

Where:

$z$  is the shape parameter

$q$  is the scale parameter

$y$  are the random variables (wind speed)

## 2. Methods

### 2.1. Maximum Likelihood Estimation Method (MLE)

According to [21], the maximum likelihood estimation method 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 estimation 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. According to [9, 15] it is stated that a maximum likelihood function let us say  $\hat{M}$  of  $M$  is a solution to the maximization problem given as.

$$\hat{M} = \text{argMax } Ln(M : x_1, x_2, \dots, x_N) \quad (4)$$

Where  $X_1, \dots, X_N$  represents the wind speed observations. Under suitable regularity conditions, the first-order condition is given as

$$\frac{\partial Ln(M : x_1, \dots, x_N)}{\partial M} = -N + \frac{1}{M} \left( \sum_{i=1}^N x_i \right) \quad (5)$$

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  $\hat{M}$  satisfies the following equation

$$\frac{\partial Ln(M : x_1, \dots, x_N)}{\partial M} = \frac{\partial Ln(\hat{M} : x_1, \dots, x_N)}{\partial M} = 0 \quad (6)$$

The log-likelihood equation that corresponds to a linear or non-linear system of  $P$  equations with  $P$  unknown parameters  $M_1, \dots, M_P$  is given by

$$\frac{\partial Ln(M : x_1, \dots, x_N)}{\partial M} = \left( \frac{\partial Ln(M : x_1, \dots, x_N)}{\partial M_1}, \dots, \frac{\partial Ln(M : x_1, \dots, x_N)}{\partial M_P} \right) = (0, \dots, 0) \quad (7)$$

Maximum likelihood estimation is a recommended technique for many distributions because it uses the values of the distribution's parameters that make the data more likely

than any other parameter. 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  $\infty$  and lastly they are asymptotically normal.

**2.2. Maximum Likelihood Estimation for Weibull Distribution**

For the two-parameter distributions, the shape parameter is dimensionless and shows how peak the site under examination is and the scale parameter is to show how windy the site under examination is. This research used the Weibull two-parameter distribution for the wind speed analysis which is given in equation (1), [1, 11].

According to [7], the two constants, shape and scale parameters are positive, the scale parameter scales u variable (wind speed variable) and the shape parameter decides the shape of the rate function,  $R(u) = \left(\frac{b}{p}\right)\left(\frac{u}{p}\right)^{b-1}$ . If the shape parameter b, is less than 1, then the rate is decreasing with u. Whereas if the shape parameter is greater than 1, then the rate is increasing with u and if the shape parameter = 1, then the rate is said to be constant, and in this case, the Weibull distribution is said to be the exponential distribution.

Suppose that  $u_1, u_2, \dots, u_n$  are independent and identically distributed Weibull random variables representing the wind speed with a probability density function f(u) given in the equation (1) where the two parameters are assumed to be unknown. To estimate the parameters using the maximum likelihood method, the likelihood function of  $u_1, u_2, \dots, u_n$  can be formulated from equation (1) as

$$L(p, b) = \prod_{i=1}^n f(u_i) = \left(\frac{b^n}{p^n}\right) \left(\prod_{i=1}^n \frac{u_i}{p}\right)^{b-1} \exp\left(-\sum_{i=1}^n \frac{u_i}{p}\right) \quad (8)$$

By taking the natural logarithm transformation, we have the equation

$$\ln L(p, b) = n \ln b - n \ln p + (b-1) \frac{\sum_{i=1}^n \ln(u_i)}{p} + \left(-\sum_{i=1}^n \frac{u_i}{p}\right) \quad (9)$$

$$\ln L(p, b) = n \ln b - n \ln p + \frac{(b-1)}{p} \sum_{i=1}^n \ln(u_i) + \left(-\frac{1}{p} \sum_{i=1}^n u_i\right)$$

Differentiating  $\ln L(p, b)$  with respect to p, we obtain

$$\frac{\partial}{\partial p} \ln L(p, b) = -\frac{n}{p} - \frac{1}{2}(b-1) \frac{\sum_{i=1}^n \ln(u_i)}{p} + \frac{1}{2} \sum_{i=1}^n \frac{u_i}{p} \quad (10)$$

Differentiating  $\ln L(p, b)$  with respect to b, we obtain

$$\frac{\partial}{\partial b} \ln L(p, b) = \frac{n}{b} + \frac{\sum_{i=1}^n \ln(u_i)}{p} - \frac{1}{p} \sum_{i=1}^n u_i^b \ln(u_i) \quad (11)$$

Equating equations (10) and (11) to zero gives the maximum likelihood estimates ( $\hat{p}, \hat{b}$ ) of (p, b). The estimate for  $\hat{p}$  is as shown

$$\begin{aligned} -\frac{n}{p} - \frac{1}{2}(b-1) \frac{\sum_{i=1}^n \ln(u_i)}{p} + \frac{1}{2} \sum_{i=1}^n \frac{u_i}{p} &= 0 \\ \frac{n}{p} &= \frac{1}{2}(b-1) \frac{\sum_{i=1}^n \ln(u_i)}{p} + \frac{1}{2} \sum_{i=1}^n \frac{u_i}{p} \\ \frac{n}{p} &= -\frac{1}{2} \left[ (b-1) \frac{\sum_{i=1}^n \ln(u_i)}{p} - \frac{\sum_{i=1}^n u_i}{p} \right] \\ \frac{n}{p} p^2 &= - \left[ (b-1) \sum_{i=1}^n \ln(u_i) - \sum_{i=1}^n u_i \right] \\ \hat{p} &= \frac{- \left[ (b-1) \sum_{i=1}^n \ln(u_i) - \sum_{i=1}^n u_i \right]}{n} \quad (12) \end{aligned}$$

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

$$\begin{aligned} \frac{n}{b} + \frac{\sum_{i=1}^n \ln(u_i)}{p} - \frac{1}{p} \sum_{i=1}^n u_i^b \ln(u_i) &= 0 \\ \frac{n}{b} + \frac{\sum_{i=1}^n \ln(u_i)}{p} &= \frac{1}{p} \sum_{i=1}^n u_i^b \ln(u_i) \quad (13) \end{aligned}$$

Equations (3.10) and (3.11) can be solved simultaneously for  $\hat{b}$  which also obtains  $\hat{p}$  subsequently.

**2.3. Maximum Likelihood Estimation for Lognormal Distribution**

According to [5], the density function for the two-parameter lognormal distribution is given as in equation (2), [2, 13].

To compute the maximum likelihood, we obtain the likelihood function first for equation (2). The likelihood function of lognormal distribution for a series of  $p_{is}$  ( $i = 1, 2, \dots, n$ ) is derived by taking the product of the probability density of the individual  $p_{is}$  given as below.

$$L(v, k^2) = \prod_{i=1}^n f(p)$$

$$= \prod_{i=1}^n \left( (2\pi k^2)^{-1/2} p_i^{-1} \exp \left[ \frac{-(\ln(p_i) - v)^2}{2k^2} \right] \right) \Rightarrow \hat{v} = \frac{\sum_{i=1}^n \ln(p_i)}{n} \tag{16}$$

$$= (2\pi k^2)^{-n/2} \prod_{i=1}^n p_i^{-1} \exp \left[ \sum_{i=1}^n \frac{-(\ln(p_i) - v)^2}{2k^2} \right] \tag{14}$$

We then derive the likelihood function by taking the natural logarithm

$$\begin{aligned} \ln L(v, k^2) &= \ln \left( (2\pi k^2)^{-n/2} \prod_{i=1}^n p_i^{-1} \exp \left[ \sum_{i=1}^n \frac{-(\ln(p_i) - v)^2}{2k^2} \right] \right) \\ &= -\frac{n}{2} \ln(2\pi k^2) - \sum_{i=1}^n \ln(p_i) - \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{2k^2} \\ &= -\frac{n}{2} \ln(2\pi k^2) - \sum_{i=1}^n \ln(p_i) - \frac{\sum_{i=1}^n [\ln(p_i)^2 - 2\ln(p_i)v + v^2]}{2k^2} \\ &= -\frac{n}{2} \ln(2\pi k^2) - \sum_{i=1}^n \ln(p_i) - \frac{\sum_{i=1}^n \ln(p_i)^2}{2k^2} + \frac{\sum_{i=1}^n 2\ln(p_i)v}{2k^2} - \frac{\sum_{i=1}^n v^2}{2k^2} \\ &= -\frac{n}{2} \ln(2\pi k^2) - \sum_{i=1}^n \ln(p_i) - \frac{\sum_{i=1}^n \ln(p_i)^2}{2k^2} + \frac{\sum_{i=1}^n \ln(p_i)v}{k^2} - \frac{nv^2}{2k^2} \tag{15} \end{aligned}$$

We find  $\hat{v}$  and  $\hat{k}^2$  which maximizes  $\ln L(v, k^2)$ . To find this, we differentiate equation (15) with respect to  $v$  and  $k^2$  by setting the equation equal to 0: with respect to  $v$ , we obtain

$$\begin{aligned} \frac{\partial}{\partial v} \ln L(v, k^2) &= \frac{\sum_{i=1}^n \ln(p_i)}{k^2} - \frac{2nv}{2k^2} = 0 \\ \Rightarrow \frac{\sum_{i=1}^n \ln(p_i)}{k^2} &= \frac{nv}{k^2} \\ \Rightarrow nv &= \sum_{i=1}^n \ln(p_i) \end{aligned}$$

With respect to  $\hat{k}^2$ , we obtain,

$$\begin{aligned} \frac{\partial}{\partial k^2} \ln L(v, k^2) &= -\frac{n}{2} \frac{1}{k^2} - \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{2} (-k^{-2})^2 \\ &= -\frac{n}{2k^2} + \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{2(k^2)^2} = 0 \\ \Rightarrow \frac{n}{2k^2} &= \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{2k^4} \\ \Rightarrow n &= \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{k^2} \\ \Rightarrow k^2 &= \frac{\sum_{i=1}^n (\ln(p_i) - v)^2}{n} \\ \Rightarrow k^2 &= \frac{\left( \sum_{i=1}^n \ln(p_i) - \frac{\sum_{i=1}^n \ln(p_i)}{n} \right)^2}{n} \tag{17} \end{aligned}$$

**2.4. Maximum Likelihood Estimation for Gamma Distribution**

In this section, we considered also a gamma distribution with shape parameter and scale parameter since it is the distribution that is widely used in real-life data sets. The probability density function of gamma random variable  $y$  in combination with two parameters  $z$  and  $q$  representing the shape and scale parameters respectively is given in equation (3), [10, 19].

For maximum likelihood estimation, we first get the likelihood function which is given by [14]

$$L(z, q) = \prod_{i=1}^n f(y_i, z, q) = \prod_{i=1}^n \frac{y_i^{z-1} \exp \left( -\frac{y_i}{q} \right)}{\Gamma(z) q^z} \tag{18}$$

The log of the likelihood function is given by

$$\begin{aligned} \ln L(z, q) &= \sum_{i=1}^n \ln \left[ \frac{y_i^{z-1} \exp\left(-\frac{y_i}{q}\right)}{\Gamma(z) q^z} \right] \\ &= \sum_{i=1}^n \left[ \ln \left( \frac{1}{\Gamma(z) q^z} \right) + \ln \left( y_i^{z-1} \exp\left(-\frac{y_i}{q}\right) \right) \right] \\ &= \sum_{i=1}^n \ln \left( \frac{1}{\Gamma(z) q^z} \right) + \sum_{i=1}^n \ln \left( y_i^{z-1} \exp\left(-\frac{y_i}{q}\right) \right) \end{aligned}$$

Since,  $\ln(q^z) = z\ln(q)$ , we obtain

$$\begin{aligned} &= -\sum_{i=1}^n \left[ \ln(\Gamma(z)) + z\ln(q) \right] + \sum_{i=1}^n \left[ (z-1)\ln(y_i) - \frac{y_i}{q} \right] \\ &= -n \left[ \ln(\Gamma(z)) + z\ln(q) \right] + (z-1) \sum_{i=1}^n \ln(y_i) - \frac{1}{q} \sum_{i=1}^n y_i \end{aligned} \tag{19}$$

To find the maximum likelihood estimates for  $\hat{z}$  and  $\hat{q}$  for  $z$  and  $q$ , we equate equation (19) to zero and then find out the partial derivatives with respect to  $\hat{z}$  and  $\hat{q}$  respectively.

$$\begin{aligned} \frac{\partial}{\partial z} \ln L(z, q) &= -n \left[ \ln(q) + \frac{\Gamma'(z)}{\Gamma(z)} \right] + \sum_{i=1}^n \ln(y_i) = 0 \\ \Rightarrow \hat{z} &= \ln(q) + \frac{\Gamma'(z)}{\Gamma(z)} = \frac{\sum_{i=1}^n \ln(y_i)}{n} \end{aligned} \tag{20}$$

With respect to  $\hat{q}$  and setting the equation equal to 0: we get,

$$\begin{aligned} \frac{\partial}{\partial q} \ln L(z, q) &= -n \frac{z}{q} + \frac{1}{q^z} \sum_{i=1}^n y_i = 0 \\ &= \frac{nz}{q} = \frac{1}{q^z} \sum_{i=1}^n y_i \Rightarrow \frac{z}{q} = \frac{1}{q^z} \left( \frac{\sum_{i=1}^n y_i}{n} \right) \Rightarrow \frac{z}{q} = \frac{1}{q^z} \bar{y} \\ &\Rightarrow \hat{q} = \frac{\bar{y}}{\hat{z}} \end{aligned} \tag{21}$$

**2.5. Goodness of Fit Test**

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 which simply tells you how good the distributions fit the data. Several tests will be used to analyze the suitability of the three distribution namely gamma, Weibull, and log-normal distribution to help have the most precise and reliable distribution. The goodness of fit test was examined using the following metrics namely the Kolmogorov-Smirnov test and Anderson-Darling test. And the goodness test criteria will be examined using Akaike’s Information Criterion and Bayesian Information Criterion. 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 vis applicable since we are only investigating the continuous statistical distributions. It is calculated as,

$$D^* = \max(|F_1(t) - F_2(t)|) \tag{22}$$

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 not follow the specified distribution
- The smaller the test statistic the better the fit.

**2.5.1. Akaike’s Information Criterion (AIC)**

The Akaike’s Information Criterion is calculated as

$$AIC = -2 \log L(p) + 2w \tag{23}$$

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.

**2.5.2. Bayesian Information Criterion (BIC)**

The Bayesian Information Criterion is calculated as

$$BIC = -2 \log L(p) + w \log M \tag{24}$$

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. Efficiency Test**

An estimator is said to be more efficient than another estimator if it is more reliable and precise for the same sample size. For the research to achieve part of its specific objectives, there is a need to understand how efficiency test is carried out. This will be 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. Descriptive Analysis for Two-parameter Distribution

The minimum speed in the data is 0.12 m/s and the maximum speed is 5.35 m/s with the mean speed of 1.9658 m/s and the estimated standard deviation of 1.2407 m/s. The estimated kurtosis and skewness are 2.8094 and 0.8433. Figure 1 shows the data distribution for 63778 observations free from outliers.

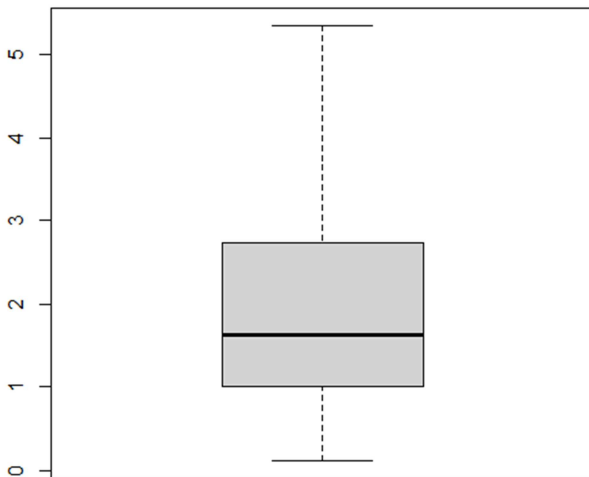


Figure 1. Box plot.

Table 1 shows the summary statistics.

Table 1. Summary statistics.

Min value	0.12
Max value	5.35
Mean	1.965777
Median	1.62
Estimated std	1.24065
Estimated kurtosis	2.809401
Estimated skewness	0.8433485

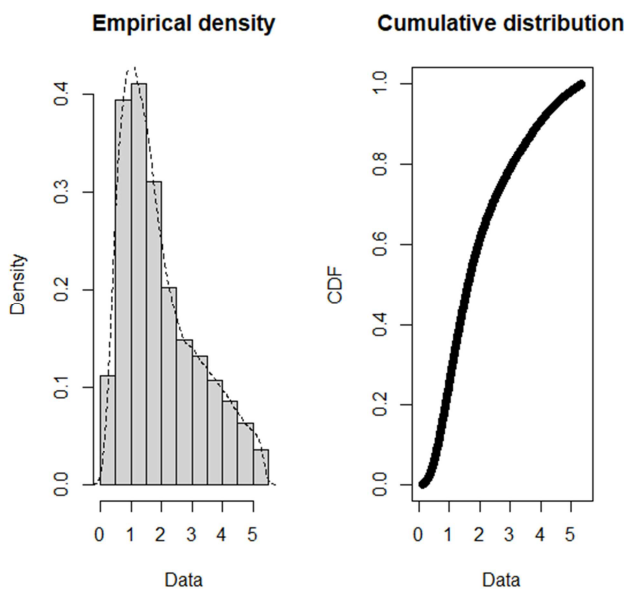


Figure 2. Histogram and cumulative curve.

From Figure 2, it can be observed that the distribution of the data is positively skewed since the peak of the data is towards the left and the right tail is longer. This shows that the data is not perfectly symmetrical since the skewness is not equal to or close to zero. From the cumulative distribution, it can be seen that the probability of having a wind speed of less than 4 m/s is almost 0.8. Meaning that the observed wind speeds above 4 m/s are less compared to those below 4 m/s.

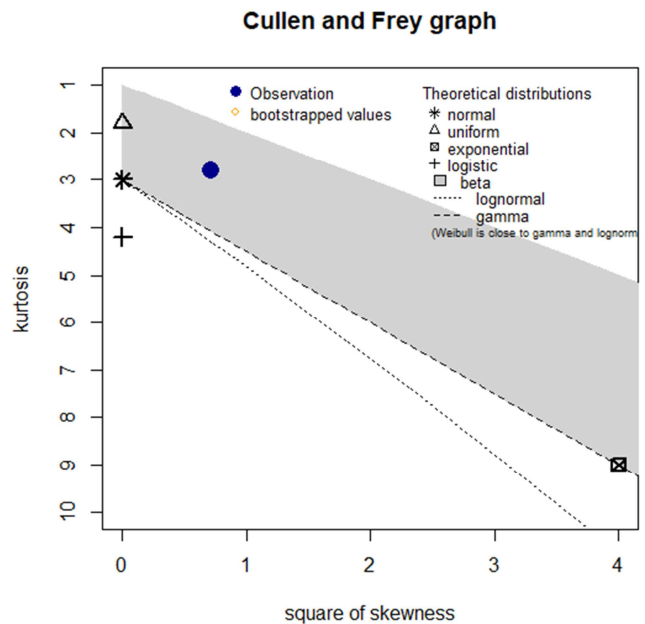


Figure 3. Data explanation.

Figure 3, is used to show a scatter plot of kurtosis and skewness. This graph helps in understanding the best possible distribution or distributions that are or are fitting the data. From the graph, we can observe that the plot is can be estimated at around a kurtosis of 2.8 and a square of skewness of around 0.7 (skewness = 0.8). With a skewness of 0.8 and kurtosis of 2.8, we can conclude that the normal distribution cannot fit the data best since normal distribution requires that kurtosis = 3 and skewness = 0. The uniform distribution is not also the best distribution for fitting this data since the observed difference between the scatter plot of kurtosis and square of skewness and that of the uniform distribution is not that close (for a uniform distribution needs a kurtosis value of 1.8 and skewness value of 0). For the logistic distribution, we can say that it is also not the best for the data since logistic distribution always has a kurtosis of 4.2 and skewness value = 0 and from the graph, it can be observed that the logistic plot is not closer to the data plot. For the exponential distribution, we can observe that its point is far away from the data point, this is because exponential distribution is expected to have a kurtosis of 9 and skewness of 2 compared to the data point skewness of 0.7 therefore exponential distribution is not the best for the data. From the graph, it can be seen that beta distribution can fit the data but this distribution cannot be applied to the data since beta distribution is a family of

continuous probability distributions defined on the interval of [20, 1] which is not the case with the collected data for this research. From the graph, log-normal and gamma distributions can fit the data best because they appear to be close to the data points and well distributed. Weibull distribution is also another good distribution for fitting the data since from the graph it is said that Weibull is close to gamma and log-normal.

**3.2. Two Parameter Estimates**

Table 2 shows each distribution with its estimated parameters, standard error of the parameters, and their correlations coefficients.

*Table 2. Parameter estimates.*

Distribution	Parameter	Estimate	Std Error	Correlation
Weibull	Shape	1.669565	0.005108	0.32456
	Scale	2.210888	0.005544	
Gamma	Shape	2.47634	0.013041	0.902297
	Scale/rate	1.25991	0.007354	
Log-normal	Shape	0.460632	0.002730	-5.753411xe-11
	Scale	0.689522	0.001931	

From table 2, we can see that Weibull's two parameters namely shape and scale parameters have a weak positive correlation. Gamma parameters are also positively correlated with a strong positive correlation of 0.9. The log-normal two parameters show a weak negative correlation. It is very important to examine if the estimated values of the parameters are useful in predicting wind speed. This can be done by investigating the significance of each of the parameters under each distribution. This is achievable by applying the t-test with the test statistic given as

$$t^* = \frac{\text{Estimator} - \text{Parameter}}{\text{Standard Error}} \tag{25}$$

The appropriate hypothesis test about the adequacy is given as  $H_0$ : Estimator = 0 versus  $H_1$ : Estimator  $\neq$  0. We reject  $H_0$  if the t statistic value is greater than t (table) value. In this case, we use 1.96 as the t table value since our sample size is above 1000 and we assume that the parameter value = 0.

*Table 3. Significance test for parameters.*

Distribution	Parameter	Estimate	Std Error	t stat	t value
Weibull	Shape	1.669565	0.005108	326.853	1.96
	Scale	2.210888	0.005544	398.7893	
Gamma	Shape	2.47634	0.013041	189.8888	1.96
	Scale/rate	1.25991	0.007354	171.3231	
Log-normal	Shape	0.460632	0.002730	168.7297	1.96
	Scale	0.689522	0.001931	357.0803	

From table 3, all the t statistic values are greater than the table value (1.96) therefore we reject the null hypothesis and conclude that all the distributions and the data are useful in predicting the wind speed of the Narok region since the estimated parameters are all not equal to zero.

To get the best distribution among the three, we can use

either graphical analysis or goodness of fit analysis

**3.3. Test of Goodness of Fit**

The discussion here will help us to the best distribution that can be applied to study the wind speed data of Narok county. The conclusion will depend on the values of Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model with the smaller value for both the AIC and BIC is considered the best distribution for the study.

*Table 4. Goodness of fit statistics.*

Statistics	Weibull	Log-normal	Gamma
Kolmogorov-Smirnov	0.05372	0.038854	0.036181
Anderson-Darling	365.758263	189.673705	211.647056

Table 4 shows that one or more distribution (s) can be used to fit the Narok region wind speed data since the data follows all the three distributions. This is evidenced from the first three statistics namely Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling test. This is because these three statistics are used to investigate if the applied data follows a certain specified distribution. Since the two statistics, Anderson-Darling and Cramer-von Mises are the refinements of the Kolmogorov-Smirnov (K-S) test, we can use either of them to make a decision. Therefore, using the Kolmogorov-Smirnov statistic it can be concluded that the data follow gamma distribution and log-normal distribution since the critical values for Kolmogorov-Smirnov is 0.0430. Because gamma and lognormal has the smaller KS statistic values of 0.036181 and 0.038854 compared to the KS critical value, we fail to reject the null hypothesis of the other two distributions.

*Table 5. Goodness of fit criteria.*

Criteria	Weibull	Lognormal	Gamma
AIC	191777.5	192340.2	190407.2
BIC	191795.7	192358.4	190425.3

From table 5, the AIC and BIC values in the table, it can be observed that the gamma distribution has the lower values for both the AIC (190407.2) and BIC (190425.3) tests. This is a clear indication that gamma distribution is the best among the three distributions for fitting the Narok wind speed data with the shape parameter of 2.47634 and scale parameter of 1.25991.

**4. Conclusion**

From the analysis, we can conclude that using maximum likelihood estimation method, gamma distribution with two parameters is the best distribution for fitting wind speed data since it yields lower AIC and BIC values. The scale parameter which shows how windy the site under study is, is estimated by the value 1.25991 and the shape parameter how peak the site is, is estimated by the value 2.47634. Therefore, the gamma distribution is the efficient distribution for studying how windy and/or peaked the region or site under examination is. The distribution is given as:

$$f(y) = \frac{y^{2.47634-1} \exp\left(-\frac{y}{1.25991}\right)}{\Gamma(2.47634) 1.25991^{2.47634}}, (y > 0) \quad (26)$$

## 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 the gamma distribution in the examination of wind speed distribution in the other regions/parts of the country with good wind quantity like Loitoktok, Marsabit, etc 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 on 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 the 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] Hurlin, C., (2013). Maximum likelihood estimation. *Advanced econometrics. The University of Orleans*.
- [10] Johnson, W., Donna, V., and Smith, L., (2010). Comparison of estimators for parameters of gamma distributions with left-truncated samples.
- [11] Maleki, F., and Deiri, E., (2007). Methods of estimation for three-parameter reflected Weibull distribution.
- [12] 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*.
- [13] Oludhe, C., (1987). Statistical characteristics of wind power in Kenya. *The University of Nairobi*.
- [14] Rahayu, A., Puhadi., Sutikno., and Prastyo, D. D., (2020). Multivariate gamma regression: parameter estimation, hypothesis testing, and applications.
- [15] Rambachan, A., (2018). Maximum likelihood estimates and Minimum distance estimate.
- [16] Solar and wind energy resource assessment. (2008). Kenya country report.
- [17] Sukkiramathi, K., Sessaiah, C., and Indhumathy, D., (2014). A study of Weibull distribution to analyze the wind speed at Jogimatti in India. Vol. 01. pp 189-193.
- [18] Ulgen, K., and Hepbasli, A., (2002). Determination of Weibull parameters for wind energy analysis of Izmir, Turkey.
- [19] Ulgen, K., Genc, A., Hepbasli, A., and Oturanc, G., (2009). Assessment of wind characteristics for energy generation.
- [20] Wind sector prospectus Kenya. (2013). Wind energy data analysis and development program.
- [21] Zheng, S., (2018). Maximum likelihood estimation. *Statistical theory II*.