FITTING WIND SPEED TO PROBABILITY DISTRIBUTIONS

BY

OTIENO O. KEVIN

A THESIS SUBMITTED TO THE SCHOOL OF PURE, APPLIED AND HEALTH SCIENCES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS MAASAI MARA UNIVERSITY

JULY 2021

DECLARATION

This research project is my original work and has been submitted in the award of any degree or other awards in any other University. This work will be only submitted by me as a thesis for the award of my master's degree at Maasai Mara University. No part of this research should be reproduced without my consent or that of Maasai Mara University.

OTIENO OKUMU KEVIN

SM04/JP/MN/6535/2018 Signature.....Date.....

This research thesis has been submitted for examination with my approval as the University supervisor.

DR. OTUMBA EDGER

Department of Mathematics and Physical Sciences

Maasai Mara University

P.O Box 861 - 20500, Narok - Kenya.

Signature.....Date.....

DR. ALILA DAVID

Masinde Muliro University of Science and Technology

P.O Box 190 - 50100, Kakamega - Kenya.

Signature.....Date.....

DR. JOHN MATUYA

Department of Mathematics and Physical Sciences

Maasai Mara University

P.O Box 861 - 20500, Narok - Kenya.

Signature.....Date.....

ACKNOWLEDGEMENT

I would like to take this chance to thank the Almighty God for his guidance and support throughout my studies and during my undertaking of this thesis. I would also like to acknowledge my supervisors, Dr. E. Otumba, Dr. D. Alila and Dr. J. Matuya for their guidance in the writing of this thesis. I would also like to appreciate all those individuals who aided me in one way or the other in the completion of this thesis.

ABSTRACT

Many researchers have fitted wind speed data to different probability distributions in the world. In Kenya, it is only Weibull distribution with two parameter which have been used to fit the wind speed data. Although, the other distributions like exponential, gamma, normal, log-normal can be the best and more efficient for fitting the wind data and for predicting wind speeds compared to Weibull distribution. Also, Minimum Distance Estimation (MDE) fitting technique is not commonly applied in fitting the two parameters (2-p) and three parameters (3-p) distributions yet it is stated as a better alternative to Maximum Likelihood Estimation (MLE) fitting technique which is considered as the most efficient fitting technique. To achieve this, the study aimed at fitting wind data to a distribution using (MLE) and MDE techniques to help us find the best and efficient probability distribution and most efficient fitting technique. The study used wind speed data from five sites in Narok county namely; Irbaan primary, Imortott primary, Mara conservancy, Oldrkesi and Maasai Mara University. The wind speed probability distributions that the data fits best was examined using the Cullen and Frey graph and a suitability test on the models done using Kolmogorov-Smirnov statistical test of goodness of fit. The wind speed data were fitted to the recommended distributions using MLE and MDE techniques. The best distribution was identified using Akaike's Information Criteria (AIC) and Bayesian Information criteria (BIC). The efficient method or technique and the efficient distribution was investigated using relative efficiency. The results showed that maximum likelihood method is the best and efficient technique for fitting the 2-p distributions and the 3-p distributions. For the comparison of the distributions for the 2-p and 3-p distributions, gamma distribution emerged as the best in all cases under MLE and MDE techniques. Gamma with 3-p distribution gave lower AIC and BIC values hence concluded as the best distribution. The efficiency test showing that gamma distribution with 3-p is more efficient than gamma distribution with 2-p, and also showed that MLE is more efficient than MDE fitting technique. The study concluded that gamma distribution with 3-p is the best and efficient distribution for fitting wind speed data with the three parameters given as; threshold parameter of 0.1174, shape parameter of 2.071773 and scale parameter of 1.120855.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

2-P: Two parameter

3-P: Three parameters

AIC: Akaike's Information Criterion

BIC: Bayesian Information criterion

LSE: Least Square Estimation

MDE: Minimum Distance Estimation

MLE: Maximum Likelihood Estimation

MOM: Method of Moments

CHAPTER ONE

INTRODUCTION

In this chapter, fitting techniques and some of the probability distributions that have been used in the area of natural wind speed analysis have been discussed. The problem that leads to this research and further outline on how the objectives will be worked upon to help in solving the research problem have also been discussed.

1.1 BACKGROUND INFORMATION

Many researchers has fitted several distributions for predicting wind speeds. Some of the fitted distributions are Weibull, gamma, log-normal, Rayleigh, hazard Weibull function and Erlang among others (Azami et al., 2009; Lawan et al.,2015; Otieno et al., 2014). In Kenya, several research groups have applied the wind speed distribution with two parameters in studying wind regime analysis and reserve estimation and in analysis of wind behavior in Juja site respectively (Barasa, 2013; Otieno, 2011). From the study by Otieno (2011) using Weibull distribution with two parameters it was found that the Weibull parameters which are scale and shape are used to show how the site is windy and the variability of the site in terms of peakedness. This wind speed distributions are examined both for two and three parameters. The two parameters are scale and shape parameters while the three parameters distribution has scale, shape and threshold. From the several distributions, the most commonly used distributions from the past studies are Weibull distribution, gamma distribution and log-normal distribution for both two parameters and three parameters distributions.

It has been found that, the researchers who studied wind speed from different parts of the world using probability distributions applied different fitting techniques for estimating the model

parameters. According to Sultan (2008) some of the common techniques applied on the analysis process are Maximum Likelihood Estimation (MLE), Method of Moments estimation (MOM), Least Square Estimation method (LSE), mean wind speed and standard deviation (point estimation) method and Minimum Distance Estimation technique. According to Salma and Abdelali (2019) compared MLE, LSE and MOM and found that MLE method gives goods estimates compared to MOM and LSE.

From the past reviews related to this study, the articles related to Kenya on the wind speed distribution studies major on Weibull distribution with two parameters and applied the technique of point estimation to estimate the mean and the variance of the distributions. Kenya has good wind speed data which can be investigated using other distributions like gamma and log-normal distribution without limiting to Weibull distribution. This is because, the other distributions can be the best and more efficient for fitting the wind data in our region and yet we are limiting to Weibull distribution for studying and predicting the wind speed. Also, there is need to try other fitting techniques like MDE to examine if it can give us precise estimates compared to MLE method, point estimation method or MOM.

According to Mumford (1997) study, it is stated that MLE under certain conditions is consistent, and have asymptotic properties of efficiency, unbiasedness, sufficient and normality if they exist. However, while this technique enjoy many asymptotic properties, it is less desirable under some scenarios like when the sample size is small or moderate and when the distributions under study have more number of parameters to be estimated like Weibull, gamma or a mixture of distributions. The researcher further explains that even if MLE is efficient and hold many asymptotic properties, it still has some problematic assumptions such as; the user should know the correct family of the distribution, the sample should contain no significant outliers, the sample needs to accurately

represent the population and the desirable properties must hold for small and moderate sample sizes.

According to Mumford (1997) and Woodward et al. (1982), a promising alternative to maximum likelihood estimation techniques is the minimum distance estimation technique because it is considered less sensitive to the four problematic assumption of maximum likelihood estimation. Thus, it is referred to as robust estimation technique implying that it attempts to protect against minor deviations from the underlying assumptions. The concept of minimum distance estimate is that better estimates will be obtained by fitting a distribution to a sample data.

Maximum likelihood estimation is considered the best with many researchers who compared it with other methods like method of moments, least square estimation method, graphical method and mean wind speed and standard deviation method among other techniques. Also, it is confirmed from the past literature that the promising alternative to maximum likelihood estimation is minimum distance estimation technique. This strongly leads to the reason why we are fitting wind speed to a distribution using maximum likelihood and minimum distance techniques with an aim of getting the best and efficient techniques and distribution.

To get the best and efficient fitting technique and distribution, good data for conducting the study is required, thus, the study uses data from Narok county meteorological firm which was collected from five difference sites within the county. The sites are; Irbaan primary, Imortott primary, Mara conservancy, Maasai mara university and Olderkesi primary. For the uniformity of the data, the study opted to use the hourly wind speed data collected from January 2016 to December 2018. The good distribution of these sites within the county makes the research viable for the entire county. With the data, the study assumed that the distributions under study are unknown and will use the

Cullen and Frey graph to know the distribution that the data fits before proceeding to the application of the two fitting techniques.

1.2 PROBLEM STATEMENT

Based on the past studies, it has been established that several authors have fitted wind speed data to various statistical distributions that can assist in predicting the chances of getting particular wind speeds. This process of fitting has been conducted using MLE method ignoring the method of MDE. In Kenya, the studies have not compared the distributions to determine the best and efficient distribution in predicting the chances of obtain particular wind speeds since it is mostly Weibull distribution with 2-p that have been applied in studying the wind speed behavior. Also, the studies have neither applied MDE technique to fit wind speed to a probability distributions with 2-p or 3-p nor compare technique of MLE and MDE used in parameter estimation for the distributions with 2-p and 3-p in order to determine the best and most efficient technique in estimation of parameter for such distributions in Kenya. This study will fill the gap by fitting wind speed data to a probability distribution by use of MLE and MDE methods to see whether MDE can give precise estimates as MLE and comparing the distribution to get the efficient distribution.

1.3 OBJECTIVES

1.3.1 Main Objective:

The main objective of this research is to statistically fit the wind speed data to probability distributions using MLE and MDE methods to give the best and efficient distribution.

1.3.2 Specific Objectives:

The specific objectives of the study were:

1. To fit wind speed data to probability distributions using MLE method;

2. To fit wind speed data to probability distributions using MDE method;
3. To compare Maximum likelihood estimation and Minimum distance estimation methods to obtain the best and efficient distribution and the efficient fitting technique.

1.4 SIGNIFICANCE OF THE STUDY

This research will be of great significance to the Kenyan Government since the efficiency of the best distribution will help to play a bigger role in the wind power generation projects within the County and the nation as it will be used as a potential tool for estimating the expected wind power of a region before installing the wind plant.

Increase in human population demands for speedy advancement in the industrialization sector in the country and also demand for more electricity for domestic use both in rural and in urban areas. Currently, Kenya is experiencing energy shortage both at industrial level and for domestic use. Thus, vision 2030 under Energy sector, Kenya targets to generate a total of 2GW capacity of wind energy by 2030 (Solar and Wind energy resource assessment 2008; Wind solar prospectus Kenya, 2002). This calls for good understanding of wind speed characteristics which needs a well-fitting distribution to be used as a control tool to help in investigating these characteristics.

The study added a value to the field of statistics by fitting wind speed to a probability distribution with 2-p and 3-p using MDE and by comparing the efficiency between MLE and MDE for fitting 2-p and 3-p probability distributions. Hence, recommending on the best fitting techniques to be applied by future researchers on 2-p and 3-p probability distributions.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter discusses the key probability distributions that has been so far used in the world to fit wind speed data, the fitting techniques applied in fitting the wind speed data to probability distributions, and the comparative studies on the fitting techniques and/or probability distributions.

2.2 WIND SPEED DISTRIBUTIONS WITH 2-P

This section majors on the discussion of the past researches which employed the statistical distributions with two parameters namely the shape parameter and scale parameter. The distributions are discussed as follows;

2.2.1 Weibull distribution with 2-p

The 2-p Weibull distribution has been widely used by many researchers from different countries within the world. It has been applied in the analysis of wind speed behavior in Malaysia, Turkey, USA and even Kenya. From Gungor and Eskin, (2008) study on the characteristics that defines wind as an energy source in Turkey, the researcher applied the general form of Weibull distribution function which is a two parameter function for wind speed and concluded that even though data time periods vary among the different researchers, 2-p Weibull is the best distribution to be used for examining wind speed before the installation of wind energy plant. Another study carried by Sukkiramathi et al. (2014) on the Weibull distribution to analyse wind speed in India reached a conclusion that, Weibull distribution function is the best in estimating the parameters of wind in India. The Weibull distribution model applied by these researchers is given by;

$$f(u)=\left(\frac{b}{p}\right)\left(\frac{u}{p}\right)^{b-1}\exp\left[-\left(\frac{u}{p}\right)\right]^{b},(b,u>0:p>1) \quad (2.1)$$

Where:

f(u) is the probability density function of 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.

These two significant parameters namely: shape and scale are estimated using the exponential techniques and are closely related to the mean value of the wind speed (Sukkiramathi et al., 2014)

2.2.2 Rayleigh model with 2-P

According to Ulgen and Heplasi (2002) study on the assessment of wind characteristics in Turkey, the author stated that this is a special case of Weibull distribution with the shape factor/parameter value is 2.0. A study by Lawan et al. (2015) on statistical modelling of long term wind speed data in Malaysia used different statistical models including the Rayleigh model with the cumulative distribution function given by:

$$f(u)=\frac{2u}{p^{2}}\exp-\left[\frac{u}{p}\right]^{2} \quad (2.2)$$

Where:

f(u) is the probability density function of wind speed

u is the wind speed.

p is the Weibull scale parameter in m/s.

2.2.3 Log-normal distribution with 2-P

According to Azami et al. (2009) study on wind speed analysis in the east coast of Malaysia 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 (Azami et al., 2009). The log-normal density function with the two parameters is given by:

$$f(p) = \frac{1}{k\sqrt{2p\pi}} \exp \frac{(\ln p - v)^2}{2k^2} \quad (2.3)$$

Where:

p is the log-normal random variable

ln(p) is the normal random variable

v is the mean for normal random variable

k is the standard deviation for the normal random variable

The log-normal distribution is defined with great reference to normal distribution. A random variable is said to be normally distributed if the logarithm of the random variable is also normally distributed. Log-normal is mostly used to model continuous data sets especially when the distribution is believed to be skewed. Sometimes, v is also the median of normal random variable.

2.2.4 Gamma distribution with 2-P

Apart from the other discussed distributions, gamma distributions is widely used among the best known distributions for wind speed. This is because choosing the most appropriate distribution of wind speed that fits the wind data of a given location is sometimes difficult. Being that the distribution to be used is essential in the first stage of wind generation, some researchers consider it important to fit the different distributions to the data to help them come up with the suitable distribution for a given location. A study Azami et al. (2009) fitted the same data in the east coast of Malaysia using the gamma model and Weibull model. According to the authors, the gamma distribution is a two parameter family of continuous probability function. It has a shape parameter and a scale parameter which are estimated using exponential approach. The probability density function of gamma distribution can be expressed in terms of shape parameter and scale parameter as shown in the equations below (Azami et al., 2009).

The applied gamma distribution with three parameters is given as follows;

$$f(y,z,q)=\frac{y^{z-1}}{\Gamma(z)q^{z}}\exp\left(-\frac{y}{q}\right),(y,z,q>0) \tag{2.4}$$

Where:

$$\Gamma(v)=\int_{0}^{\infty}y^{v-1}\exp^{-y}dy,(v>0) \tag{2.5}$$

And:

z is the shape parameter,

q is the scale parameter,

y is the wind speed.

According to Azami et al. (2009) study on wind analysis at the coast of Malaysia using the gamma distribution, log-normal distribution and the Weibull distribution, it was found that gamma distribution and Weibull distribution are the best for fitting the wind data. This was evidenced from the test of goodness of fit (Kolmogorov-Smirnov). Also, Azami et al. (2009) did another research on fitting of statistical distributions to wind speed data in Malaysia using log-normal and Weibull distributions and found out from the test of goodness of fit that Weibull distribution is better than log-normal for fitting that data. Another study by Lawan et al. (2015) on statistical modelling of long term wind speed data and they used a number of statistical distributions like Erlang model, Weibull distribution, Rayleigh distribution, log-normal distribution and gamma distribution and from the test of goodness of fit, gamma distribution and log-normal distributions were found to fit the data well compared to the other models used. According to Akyuz and Gamgam (2017) study on statistical analysis of wind speed data with gamma, Weibull and log-normal distributions in Bitlis in Turkey recommended that gamma distributions is the best for that region since it yields lower value from the test of fit compared to log-normal and Weibull distributions.

The study observed that these five distributions with 2-p namely Weibull, Rayleigh, log-normal, Gamma and Erlang are the mostly used statistical distributions in examining the wind data of a given area or location of interest. There are other distributions like the hybrid Weibull distribution, exponential distribution, generalized pareto distribution, generalized extreme values model, Nakagami distribution and normal distribution but the disadvantage is that they are not of great interest to the wind industry and to the investors in studying the wind speed variations like the five distributions discussed above. From these related studies, only Weibull with 2-p has been applied

to fit Kenya wind speed data. This leaves a gap because there is no study clarifying that Weibull with 2-p is the best of all other distributions with two parameters for investigating the wind speed behavior in Kenya or any part of Kenya. This gap will be filled by this research since the intends to investigate a good number of probability distributions namely; Weibull, exponential, normal, gamma and log-normal with different parameters and make conclusion on the best and efficient probability distribution to be used in fitting wind speed data in Kenya.

2.3 WIND SPEED DISTRIBUTIONS WITH 3-P

This section discusses the statistical distributions with the parameters which have been used in the past with various statistical researchers. The three parameters are the scale parameter, shape parameter and thresh-hold parameter.

2.3.1 Weibull distribution with 3-p

The number of articles in relation to three parameters Weibull distribution are not as many as two parameter distribution. Azami et al. (2009) study on fitting of statistical distribution to wind speed data in Malaysia using the three-parameter distribution was given as follows;

$$f(u)=\left(\frac{b}{p}\right)\left(\frac{u-w}{p}\right)^{b-1}\exp\left[-\left(\frac{u-w}{p}\right)\right]^{b}\ (b,u>0,p>1):(u\neq w) \tag{2.6}$$

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

2.3.2 Log-normal distribution with 3-p

Study by Oludhe (1987) on characteristics of wind power in Kenya used a number of mathematical formulas including a log-normal distribution with three parameters. Also, Azami et al. (2009) applied the same statistical distribution for fitting Malaysia wind data. The probability density distribution function was defined as follows;

$$f(p)=\frac{1}{(p-y)k\sqrt{2\pi}}\exp\left[-\left(\frac{\ln(p-y)-v}{2k}\right)\right]^2,(v,k>0:p\geq 1);(p\neq y) \tag{2.7}$$

Where:

v is the scale parameter.

k is the shape parameter.

y is the thresh-hold parameter, also referred to the location parameter.

p is the wind speed.

2.3.3 Gamma distribution with 3-p

A study by Louzada et al. (2016) used the three parameter generalized probability density function to conduct his research on the comparison of estimation methods for generalized gamma distribution. The same statistical distribution was used to carry statistical analysis of wind speed data in Turkey (Lu et al., 2002; Mert & Karakus, 2015). The applied gamma distribution with three parameters is given as follows;

$$f(y,z,q,t)=\frac{y t^{z-1}}{\Gamma(z/t)q^{z}}\exp\left[-\left(\frac{y}{q}\right)^{t}\right](z,y,q,t>0) \tag{2.8}$$

Where:

z is the shape parameter,

q is the scale parameter,

z, t are shape parameter.

From the three parameters distribution, there is no articles providing 3-p probability distribution for investigating the wind speed of Kenya or any part of Kenya using wind speed data. Therefore, there is no article advising that a three parameter distribution is the best for examining wind speed characteristics in Kenya, leaving a gap which will be filled by this research.

2.4 COMPARATIVE STUDIES OF THE DISTRIBUTIONS

Some past studies have compared few distributions:

A study by Sukkiramathi (2014) in India investigated five distributions namely; Weibull, mixture gamma and Weibull, mixture Normal and Weibull using the method of moments to estimate the parameters.This study concluded that Weibull distribution fits well compared to the other mixture of the distributions.

A study by Azami et al. (2009) applied the maximum likelihood estimation principle, graphical analysis and goodness of fit test to investigate the best distributions with two parameters for Weibull, gamma and log-normal distributions, it concluded that Weibull and gamma distributions with two parameters fits the wind speed data used best compared to the log-normal distribution.

Study conducted in Malaysia using Miri wind speed data compared a number of distributions namely; Erlang, Rayleigh, gamma, log-normal and Weibull distributions all with two parameters. The researcher applied goodness of fit tests (Kolmogorov-Smirnov, Anderson-Darling, & Chi-square tests). These tests leads to a conclusion that gamma and log-normal distributions fits the observed data well compared to the other distributions applied (Lawan et al., 2015)

A study by Mahyoub (2006) in Taiz-Yemen used Weibull and Rayleigh two parameter distributions in the study and applied chi-square and correlation coefficient, R^2, and Root Mean Square Error (RMSE) analysis which leads to conclusion that Weibull distribution gives better performance than Rayleigh distribution.

Study conducted using Tridad hourly wind speed to examine different distributions with two parameters namely; gamma, Weibull, Rayleigh, normal and Nakagami, and three parameters distributions namely; Birnbaum-Saunders, generalized extreme value, and generalized pareto distributions. The study used maximum likelihood estimation method to estimate the parameters, graphical analysis and test of goodness of fit statistics (chi-square and Kolmogorov-Smirnov) was used to access the distribution and used Akaike's Information Criteria and Bayesian Information Criteria to made decision on the best distribution. This research found that Rayleigh distribution showed better results as Weibull distribution performed poorly (Dookie et al., 2018).

A study by Mert and Karakus, (2015) in Turkey compared 4-parameter Burr, 2-parameter Weibull and 3-parameter generalized distributions using maximum likelihood estimation method and graphical analysis. The study considered Burr distribution over Weibull and generalized gamma distributions.

From the comparative studies, it can be observed that the researchers are not more interested in comparing the 3-p distributions among themselves. It can also be observed that there is no model comparison study conducted in Kenya to give the best model both for 2-p and 3-p distributions. Lastly, it can be observed that from the methods, the researchers are not concentrating on investigating the efficiency of the distributions.

2.5 DATA FITTING TECHNIQUES

This section discusses the various fitting techniques or methods used by different authors who have been fitting data to the different distributions with two or three parameters. Some of the key methods which are discussed are MLE, LSE, MOM, Mean Wind Speed and standard deviation, and MDE.

From the past articles, we found out that most of the authors used MLE, LSE, MOM and Mean wind speed and standard deviation.

Some of the comparative studies on the fitting techniques are discussed below;

A study by Saleh et al. (2012) assessed different methods used to estimate Weibull parameters namely; mean wind speed and standard deviation, maximum likelihood method, method of moments, the commonly used graphical method, modified maximum likelihood method and Power density method. The study concluded that mean wind speed and maximum likelihood estimation methods are the best methods for estimating the Weibull distribution parameters for the purpose of wind speed analysis.

A study by Johnson et al. (2010) compared maximum likelihood estimation technique and method of moment technique on gamma distribution and reached a conclusion that maximum likelihood estimation is superior to the method of moments recommending that researchers should use maximum likelihood estimation technique.

A study by Yilmaz and Celik (2008) and Salma and Abdelali (2018), compared the three estimation methods; maximum likelihood estimation (MLE), method of moments (MOM) and least square estimation (LSE) method. Both articles reached a conclusion that maximum likelihood method gives good estimates compared to MOM and LSE. According to Yilmaz and Celik (2008), it is

further explained that least square method gives more accurate results compared to method of moments.

A study by Woodward et al. (1982), which compared minimum distance and maximum likelihood techniques using mixture of asymmetric distributions concluded that minimum distance method is more robust than maximum likelihood estimation in that it is less sensitive to symmetric departures from the existing normality assumption of component distributions. Another researcher applied minimum distance method and maximum likelihood method on the study aimed at estimating 3-parameters for Weibull distribution and concluded that minimum distance methods showed significant improvements over the maximum likelihood estimation method meaning that it is superior (Sultan, 2008).

A study conducted on robust parameter estimation for mixed Weibull distribution with seven parameters used minimum distance estimation and maximum likelihood estimation and came to conclusion that minimum distance estimation gives better estimates compared to maximum likelihood estimation technique (Mumford, 1997).

From the assessment of related studies, we found that researchers are not applying the minimum distance technique in fitting the single 2-p or 3-p distributions. Secondly, from the comparison it was found that most of the researchers only compared maximum likelihood and minimum distance on mixed distribution but not comparing for single distribution like Weibull, gamma, log-normal and Rayleigh with 2-p or 3-p.

2.6 RESEARCH GAP

From the past studies, it was found that most researchers from Kenya were interested only in Weibull distribution with 2-parameters. Also, the researchers did not see the essence of

investigating other distribution like gamma and log-normal with 2-p and 3-p so as to give the best distribution for fitting the wind speed data of Kenya.

Secondly, since the few studies which compared MDE with MLE on mixed distributions concluded that minimum distance is more robust than maximum likelihood estimation technique, there is need to examine if its robustness can still be seen in single distributions with 2-parameters and 3-parameters. Also, it was found that many researchers are not applying this MDE technique on fitting the distribution as majority prefer fitting the distribution using MLE, graphical method, LSE, MOM and mean wind speed and standard deviation method.

Lastly, from the related literature, it was found that most researchers who performed comparative studies on more than one distributions gave the best fitting distribution only and made their conclusion at that point without examining if the distribution is efficient or not.

In conclusion therefore, this study intends to fill in the study gap by fitting wind speed to a probability distribution using MLE and MDE techniques to get the best wind speed distribution for the dataset and also by assessing the efficiency between the two fitting techniques namely MLE and MDE techniques.

CHAPTER THREE

METHODS

3.1 INTRODUCTION

This chapter describes the type of data used in this research and the various methods and formulas' that are employed to fit the data. The data is used to estimate either 2-p and/or 3-p for the probability distribution(s) and test for the goodness of fit for the distributions. The probability distributions namely; gamma distribution, exponential distribution, normal distribution, log-normal distribution, and Weibull distribution among others are examined and then the distribution(s) which the data fits best is/are then used in further analysis to help come up with the best distribution. The best data fitting probability distribution(s) is/are all discussed here with their respective parameter estimation techniques using MLE technique and MDE technique.

3.2 DATA

This research used secondary data collected by Narok meteorological site on the five sites well scattered within Narok namely Maasai Mara University, Mara conservancy, Olderkesi primary school, Imortott primary school and Irbaaan primary. The data was collected hourly for a constant level of 10 meters for period of three years from 2016 to 2018. Data available at: www.tahmo.org/climate.data.

3.21 Data description

From the analysis of 66858 observations, the mean wind speed is 2.1617 m/s with the standard deviation of 1.5124m/s and a median value of 1.69 m/s. The box plot for the distribution of the hourly wind data is shown in the Figure 3.1.

From Figure 3.1, it can be seen that the data contains some extreme values of the wind speed (outliers) hence there was need to remove all the outliers before proceeding with further analysis. After removing all the outliers in the data we remained with 63778 observations out of 66858 observations. Meaning that we lost 3080 observations as outliers (extreme wind speed observations as per the level of the research.

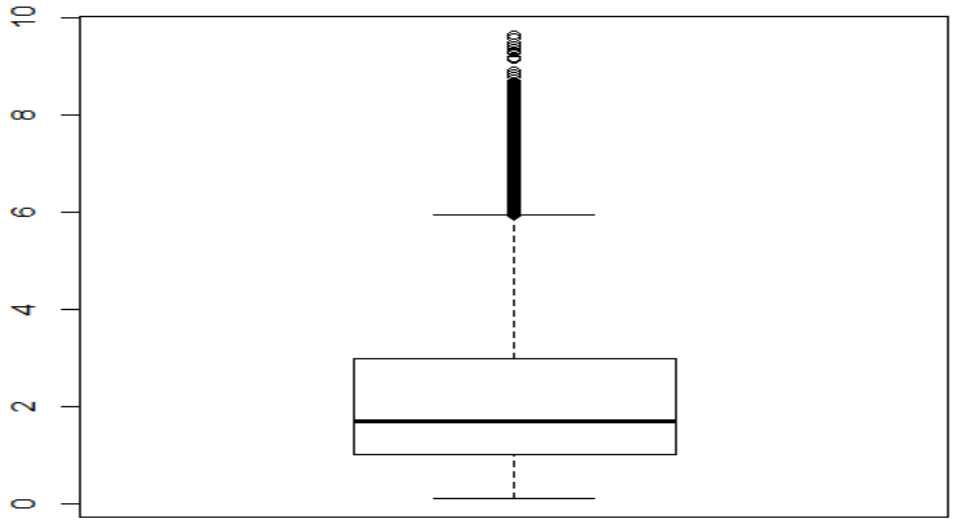

Figure 3.1: Box plot of the original data

The data distribution for 63778 observations free from outliers as shown in Figure 3.2 and the descriptive statistics shown in Table 3.1.

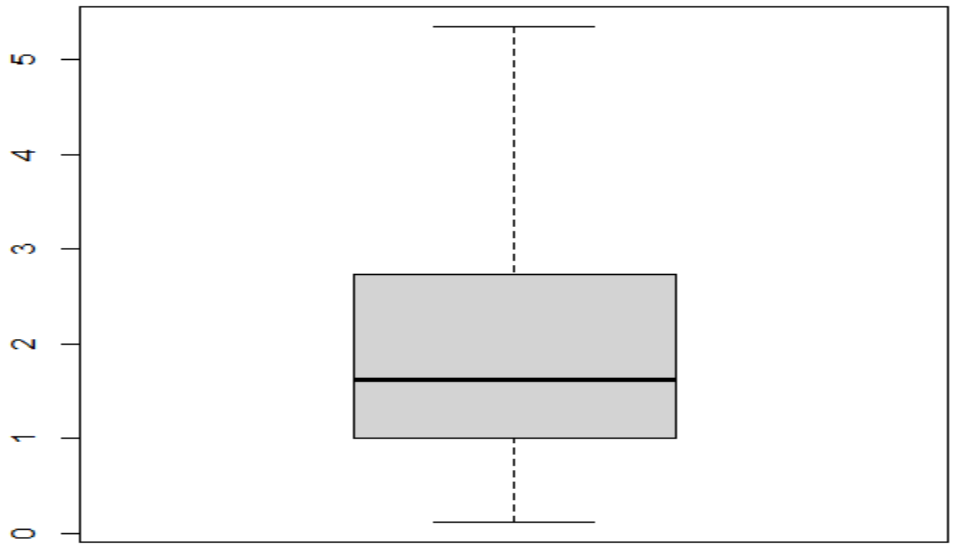

Figure 3.2: Box plot after removing outliers

Table 3.1: Summary Statistics after Removing the Outliers

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

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 is 2.8094 and 0.8433.

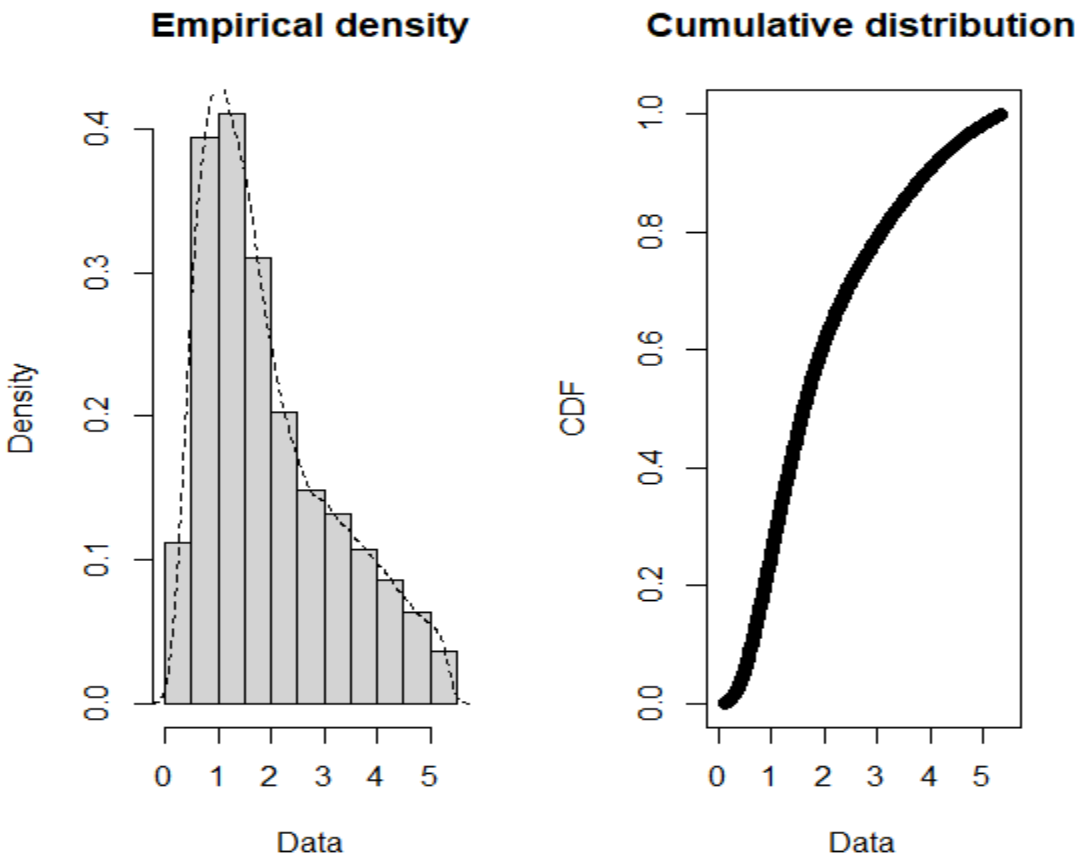

Figure 3.3: Histogram and Cumulative distribution graph for wind

Figure 3.3 show 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. Some of the known positively skewed distributions are exponential distribution, log-normal distribution, Weibull distribution and gamma distribution.

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 since from the curve the probability of expecting wind speeds of around 5m/s and 6m/s and above 6m/s is low. Meaning that the observed wind speeds above 4 m/s are less compared to those below 4 m/s.

The Fig. 3.4 is used to show a skater plot of kurtosis and skewness. This graph helps in understanding the best possible distribution or distributions that is 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 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 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 have 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 [0, 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.

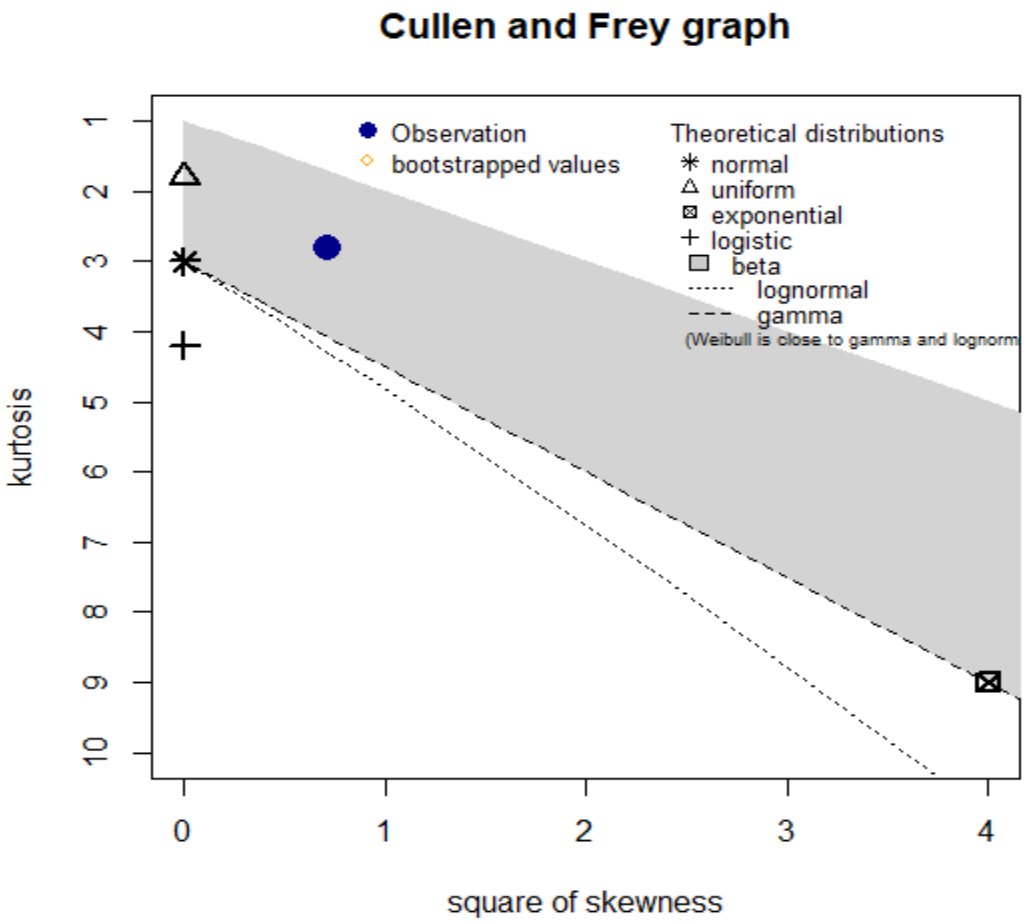

Figure 3.4: Cullen and Frey graph for wind speed data

3.2.2 Data description after subtracting the threshold value from the original data

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 raw data, the threshold values are given that Weibull has a threshold value of 0.1195 m/s, gamma a threshold value of 0.1174 m/s and log-normal a threshold of -0.1107 m/s as shown in Table 3.2

Table 3.2: Threshold values

Distribution	Threshold value
Weibull	0.1195
Gamma	0.1174
Log-normal	-0.1107

To have the best threshold value, the study uses 0.1174 as the first value and then perform iteration up to 0.1195 using an interval of 0.0005 as discussed in chapter four. The study did not use the threshold value for log-normal because it is negative (-0.1107), and the value of wind speed cannot be negative since threshold parameters takes the same value and unit as the wind speed variable. The threshold value was picked based on the AIC and BIC results. AIC and BIC are discussed on the next sub topic under test of goodness of fit.

After subtracting the threshold value, the summary statistics is as represented in the table 3.3

Table 3.3: Summary statistics

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

3.3 TEST OF GOODNESS OF FIT

After analyzing the data using Figure 3.4, it is important to verify the suitability and the accuracy of the distribution by performing the test of goodness of fit to tell us how good the data fits the distribution(s). The test of goodness of fit statistic was examined using Kolmogorov-Smirnov test. And also, for the three parameter distribution fitting, the decision on best threshold value is reached

using the test of goodness of fit criteria examined using AIC and BIC. The test of goodness of fit criteria is applied in investigating the efficiency between the distributions under study.

3.3.1 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 (Lawan et al., 2015). It is calculated as;

$$D^* = \max\left(\left|F_1(t) - F_2(t)\right|\right) \tag{3.1}$$

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.

The hypothesis is stated as;

H_0 : The data follows Weibull distribution.

H_1 : The data do no follow Weibull distribution.

Or

H_0: The data follows Gamma distribution.

H_1 : The data do no follow Gamma distribution.

Or

H_0: The data follows Log-normal distribution.

H_1 : The data do no follow Log-normal distribution.

The decision rule: The study rejects H_0 if K-S values is greater than or equals to K-S critical value, otherwise, the study do not reject H_0 and also smaller the test statistic the better the fit.

3.4 COMPARISON CRITERIA

3.4.1 Akaike's Information Criterion (AIC)

The Akaike's Information Criterion is calculated as;

$$AIC = -2\log L(P) + 2w \quad \textbf{(3.2)}$$

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.

3.4.2 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion is calculated as below

$$BIC = -2logL(p) + wlogM \quad (3.3)$$

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

3.5 MAXIMUM LIKELIHOOD ESTIMATION METHOD (MLE)

According to Zheng, (2018), maximum likelihood method can be applied in many problems since it has a strong intuitive appeal and it yield a precise estimator. He also stated that the maximum likelihood method is widely used because it is more precise especially when dealing with large sample size since it yields accurate estimator for such samples.

According to Hurlin (2013), maximum likelihood lets say $\hat{M}$ of M is a solution to the maximization problem given as

$$\widehat{\boldsymbol{M}} = argMax\ Ln(M: x1, x2, \dots\dots, xN) \tag{3.4}$$

Where X1,....,XN represents the wind speed observations. Under suitable regularity conditions, the first order condition is given as

$$\frac{\partial Ln(M : x1,.....,xN)}{\partial M} = -N + \frac{1}{M}\left(\sum_{i=1}^{N} xi\right) \tag{3.5}$$

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

$$\frac{\partial Ln(M : x1,....,xN)}{\partial M} = \frac{\partial Ln\left(\hat{M} : x1,.....,xN\right)}{\partial M} = 0 \tag{3.6}$$

The log-likelihood equation that corresponds to linear or non-linear system of P equations with P unknown parameters M1, M2, ...MP is given by;

$$\frac{\partial Ln(M : x1,....,xN)}{\partial M} = \left(\frac{\partial Ln(M : x1,...,xN)}{\partial M1},...,\frac{\partial Ln(M : x1,...,xN)}{\partial Mp}\right) = (0,...,0) \tag{3.7}$$

MLE is a recommended technique for many distributions because it uses the values of the distribution 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 is 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 (Gupta & Biswas,2010; Hurlin, 2013).

For the two parameter distributions, the shape parameter is dimensionless and shows how the wind speed of site under examination peaked and the scale parameter is to show how windy the site under examination is (spread of the wind speed). Increasing the variation/spread of the wind speed (scale parameter) reduces the peak of the site (shape parameter) and vice versa

3.5.1 MLE for Weibull distribution with 2-P

This study used the Weibull two parameter distribution for the wind speed analysis which is given as (Ayodele et al., 2012)

$$f(u)=\left(\frac{b}{p}\right)\left(\frac{u}{p}\right)^{b-1}\exp\left[-\left(\frac{u}{p}\right)\right]^{b},(b,u>0:p>1) \tag{3.8}$$

According to Chu and Ke, (2012), the two constants, shape and scale parameters are positive constants, the scale parameter is scale to the u variable (wind speed variable) and the shape parameter decides shape of the rate function;

$$1-f(u)=\left(\frac{b}{p}\right)\left(\frac{u}{p}\right)^{b-1}$$

If the shape parameter b, is less than 1, then the rate is decreasing with u. Whereas if 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 $u_1, u_2, \ldots\ldots, u_n$ are independent and identically distributed Weibull random variables representing the wind speed with a probability density function f(u) given in the equation (3.5) where the two parameters are assumed to be unknown. To estimate the parameters using maximum likelihood method, the likelihood function of $u_1, u_2, \ldots\ldots, u_n$ can be formulated from equation (3.5) as shown in equation (3.9).

The product of the constants are not performed (introducing the general summation including the constants is not necessary to avoid interference with the rate function in the Weibull distribution). The aim is to understand how the Weibull random variable u (wind speed) is scaled or shaped therefore, there is no need of summing the power constant (Chu & Ke, 2012).

$$L(p,b) = \prod_{i=1}^{n} f(u_i) = \left(\frac{b^n}{p^n}\right)\left(\prod_{i=1}^{n}\frac{u_i}{p}\right)^{b-1} \exp\left(-\sum_{i=1}^{n}\frac{u_i}{p}\right)^{b} \tag{3.9}$$

By taking the natural logarithm transformation, we have the equation

$$\ln L(p,b) = n\ln b - n\ln p + (b-1)\frac{\sum_{i=1}^{n}\ln(u_i)}{p} + \left(-\sum_{i=1}^{n}\frac{u_i}{p}\right)^{b} \tag{3.10}$$

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

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

$$\frac{\partial}{\partial p}\ln L(p,b) = -\frac{n}{p} - \frac{1}{p^2}(b-1)\sum_{i=1}^{n}\ln(u_i) + \frac{1}{p^2}\sum_{i=1}^{n}u_i^{b} \tag{3.11}$$

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

$$\frac{\partial}{\partial b}\ln L(p,b)=\frac{n}{b}+\frac{\sum_{i=1}^{n}\ln(u_i)}{p}-\frac{1}{p}\sum_{i=1}^{n}u_i^{b}\ln(u_i) \tag{3.12}$$

Equating equations (3.11) and (3.12) to zero gives the maximum likelihood estimates $(\hat{p},\hat{b})$ of (p, b). The estimate for $\hat{p}$ is as shown

$$-\frac{n}{p}-\frac{1}{p^{2}}(b-1)\sum_{i=1}^{n}\ln(u_i)+\frac{1}{p^{2}}\sum_{i=1}^{n}u_i^{b}=0 \tag{3.13}$$

$$\frac{n}{p}=\frac{1}{p^{2}}(b-1)\sum_{i=1}^{n}\ln(u_i)+\frac{1}{p^{2}}\sum_{i=1}^{n}u_i^{b} \tag{3.14}$$

$$\frac{n}{p}=-\frac{1}{p^{2}}\left[(b-1)\sum_{i=1}^{n}\ln(u_i)-\sum_{i=1}^{n}u_i^{b}\right] \tag{3.15}$$

$$\frac{n}{p}p^{2}=-\left[(b-1)\sum_{i=1}^{n}\ln(u_i)-\sum_{i=1}^{n}u_i^{b}\right] \tag{3.16}$$

$$\hat{p}=-\left[(b-1)\sum_{i=1}^{n}\ln(u_i)-\sum_{i=1}^{n}u_i^{b}\right]\Big/n \tag{3.17}$$

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

$$\frac{n}{b}+\frac{\sum_{i=1}^{n}\ln(u_i)}{p}-\frac{1}{p}\sum_{i=1}^{n}u_i^{b}\ln(u_i)=0 \tag{3.18}$$

$$\frac{n}{b}+\frac{\sum_{i=1}^{n}\ln(u_i)}{p}=\frac{1}{p}\sum_{i=1}^{n}u_i^{b}\ln(u_i) \tag{3.19}$$

Further solution to find $\hat{b}$ is given as;

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

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

Substituting $\hat{p}$ from equation (3.17), gives;

$$\frac{n}{\hat{b}} = \frac{\sum_{i=1}^{n} u_i^{\hat{b}} \ln(u_i) - \sum_{i=1}^{n} \ln(u_i)}{-\left[\left(\left(\hat{b}-1\right)\sum_{i=1}^{n} \ln(u_i)\right) - \left(\sum_{i=1}^{n} u_i^{\hat{b}}\right)\right] \Big/ n} \tag{3.20}$$

Introduce logarithm to eliminate the power $\hat{b}$ in equation (3.20)

$$\frac{n}{\hat{b}} = \frac{\sum_{i=1}^{n} \ln u_i^{\hat{b}} \ln(u_i) - \sum_{i=1}^{n} \ln(u_i)}{-\left[\left(\left(\hat{b}-1\right)\sum_{i=1}^{n} \ln(u_i)\right) - \left(\sum_{i=1}^{n} \ln u_i^{\hat{b}}\right)\right] \Big/ n} \tag{3.21}$$

$$\frac{n}{\hat{b}} = \frac{\sum_{i=1}^{n} \hat{b} \ln(u_i) \ln(u_i) - \sum_{i=1}^{n} \ln(u_i)}{-\left[\left(\left(\hat{b}-1\right)\sum_{i=1}^{n} \ln(u_i)\right) - \left(\sum_{i=1}^{n} \hat{b} \ln(u_i)\right)\right] \Big/ n} \tag{3.22}$$

Factorizing equation (3.22) gives;

$$\frac{n}{\hat{b}} = \frac{\hat{b}\ln(u_i)\left[\sum_{i=1}^{n}\ln(u_i)-1\right]}{-\left(\hat{b}-1\right)+\hat{b}\left[\sum_{i=1}^{n}\ln(u_i)\right]/n} \tag{3.23}$$

$$\frac{n}{\hat{b}^2\ln(u_i)} = \frac{\left[\sum_{i=1}^{n}\ln(u_i)-1\right]}{\left[\sum_{i=1}^{n}\ln(u_i)\right]/n} \tag{3.24}$$

$$\frac{n}{\hat{b}^2} = \left(\frac{\left(\sum_{i=1}^{n}\ln(u_i)-1\right)}{\left(\sum_{i=1}^{n}\ln(u_i)\right)/n}\right)\bullet\left(\ln(u_i)\right) \tag{3.25}$$

$$\frac{1}{\hat{b}^2} = \frac{\left[\left(\frac{\left(\sum_{i=1}^{n}\ln(u_i)-1\right)}{\left(\sum_{i=1}^{n}\ln(u_i)\right)/n}\right)\bullet\left(\ln(u_i)\right)\right]}{n} \tag{3.26}$$

$$\hat{b}^2 = \frac{n}{\left[\left(\frac{\left(\sum_{i=1}^{n}\ln(u_i)-1\right)}{\left(\sum_{i=1}^{n}\ln(u_i)\right)/n}\right)\bullet\left(\ln(u_i)\right)\right]} \tag{3.27}$$

$$\hat{b} = \sqrt{\frac{n}{\left[\left(\frac{\left(\sum_{i=1}^{n}\ln(u_i)-1\right)}{\left(\sum_{i=1}^{n}\ln(u_i)\right)/n}\right)\bullet\left(\ln(u_i)\right)\right]}} \tag{3.28}$$

3.5.2 MLE for Lognormal distribution with 2-P

According to Brenda (2009), the density function for the two-parameter log-normal distribution with two parameters v and k given as:

$$f(p) = \frac{1}{k\sqrt{2p\pi}}\exp\frac{(\ln p - v)^2}{2k^2} \tag{3.29}$$

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

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

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

$$= \left(2\pi k^2\right)^{-n/2}\prod_{i=1}^{n} p_i^{-1}\exp\left[\sum_{i=1}^{n}\frac{-\left(\ln(p_i)-v\right)^2}{2k^2}\right] \tag{3.30}$$

We then derive the likelihood function by taking the natural logarithm

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

$$=-\frac{n}{2}\ln\left(2\pi k^2\right)-\sum_{i=1}^{n}\ln\left(p_i\right)-\frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right)-v\right)^2}{2k^2}$$

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

$$=-\frac{n}{2}\ln\left(2\pi k^2\right)-\sum_{i=1}^{n}\ln\left(p_i\right)-\frac{\sum_{i=1}^{n}\ln\left(p_i\right)^2}{2k^2}+\frac{\sum_{i=1}^{n}2\ln\left(p_i\right)v}{2k^2}-\frac{\sum_{i=1}^{n}v^2}{2k^2}$$

$$=-\frac{n}{2}\ln\left(2\pi k^2\right)-\sum_{i=1}^{n}\ln\left(p_i\right)-\frac{\sum_{i=1}^{n}\ln\left(p_i\right)^2}{2k^2}+\frac{\sum_{i=1}^{n}\ln\left(p_i\right)v}{k^2}-\frac{nv^2}{2k^2} \qquad (3.31)$$

To find $\hat{v}$ and $\hat{k}^2$, maximize $lnL(v,k^2)$. To find this, we differentiate equation (3.31) with respect to v and k^2 by setting the equation equal to 0: with respect to v, to obtain

$$\frac{\partial}{\partial v}\ln L\left(v,k^2\right)=\frac{\sum_{i=1}^{n}\ln\left(p_i\right)}{k^2}-\frac{2nv}{2k^2}=0$$

$$\Rightarrow\frac{\sum_{i=1}^{n}\ln\left(p_i\right)}{k^2}=\frac{nv}{k^2}$$

$$\Rightarrow nv = \sum_{i=1}^{n} \ln\left(p_i\right)$$

$$\Rightarrow \hat{v} = \frac{\sum_{i=1}^{n} \ln\left(p_i\right)}{n} \tag{3.32}$$

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

$$\frac{\partial}{\partial k^2} \ln L\left(v, k^2\right) = -\frac{n}{2}\frac{1}{k^2} - \frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right) - v\right)^2}{2}\left(-k^2\right)^2$$

$$= -\frac{n}{2k^2} + \frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right) - v\right)^2}{2\left(k^2\right)^2} = 0$$

$$\Rightarrow \frac{n}{2k^2} = \frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right) - v\right)^2}{2k^4}$$

$$\Rightarrow n = \frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right) - v\right)^2}{k^2}$$

$$\Rightarrow k^2 = \frac{\sum_{i=1}^{n}\left(\ln\left(p_i\right) - v\right)^2}{n}$$

$$\Rightarrow k^2 = \frac{\sum_{i=1}^{n}\left(\ln(p_i) - \frac{\sum_{i=1}^{n}\ln(p_i)}{n}\right)^2}{n} \tag{3.33}$$

3.5.3 MLE for Gamma distribution with 2-P

In this section, we considered also a gamma distribution with shape parameter and scale parameter since it is the distribution which 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 by (Azami et al., 2009; Lawan et al., 2015; Sukkiramathi et al., 2014).

$$f(y,z,q) = \frac{y^{z-1}}{\Gamma(z)q^z}\exp\left(-\frac{y}{q}\right),(y,z,q>0) \tag{3.34}$$

Where:

$$\Gamma(v) = \int_0^\infty y^{v-1}\exp^{-y}dy,(v>0) \tag{3.35}$$

For maximum likelihood estimation, we first get the likelihood function which is given by (Rahayu et al., 2020):

$$L(z,q) = \prod_{i=1}^{n} f(y,z,q) = \prod_{i=1}^{n}\frac{y_i^{z-1}\exp\left(-\frac{y_i}{q}\right)}{\Gamma(z)q^z} \tag{3.36}$$

The log of the likelihood function is given by

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

$$= \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\Gamma(z) q^z}\right) + \ln\left(y^{z-1} \exp\left(-\frac{y}{q}\right)\right) \right]$$

$$= \sum_{i=1}^{n} \ln\left(\frac{1}{\Gamma(z) q^z}\right) + \sum_{i=1}^{n} \ln\left(y^{z-1} \exp\left(-\frac{y}{q}\right)\right)$$

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

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

Therefore, we have;

$$= -n\left[\ln(\Gamma(z)) + z \ln(q)\right] + (z-1)\sum_{i=1}^{n} \ln(y_i) - \frac{1}{q}\sum_{i=1}^{n} y_i \qquad (3.37)$$

To find the maximum likelihood estimates for $\hat{z}$ and $\hat{q}$ for z and q, we equate equation (3.37) to zero and then find out the partial derivatives with respect to $\hat{z}$ and $\hat{q}$ respectively.

$$\frac{\partial}{\partial z} \ln L(z,q) = -n\left[\ln(q) + \frac{\Gamma'(z)}{\Gamma(z)} \right] + \sum_{i=1}^{n} \ln(y_i) = 0$$

$$\Rightarrow \hat{z} = \ln(q) + \frac{\Gamma'(z)}{\Gamma(z)} = \frac{\sum_{i=1}^{n} \ln(y_i)}{n} \qquad (3.38)$$

Differentiating with respect to $\hat{q}$ and setting the equation equal to 0: to get;

$$\frac{\partial}{\partial q}\ln L(z,q) = -n\frac{z}{q} + \frac{1}{q^{z}}\sum_{i=1}^{n} y_i = 0$$

$$= \frac{nz}{q} = \frac{1}{q^{z}}\sum_{i=1}^{n} y_i \Rightarrow \frac{z}{q} = \frac{1}{q^{z}}\left(\sum_{i=1}^{n} y_i \Big/ n\right) \Rightarrow \frac{z}{q} = \frac{1}{q^{z}}\bar{y}$$

$$\Rightarrow \hat{q} = \frac{\bar{y}}{\hat{z}} \tag{3.39}$$

Therefore, we have;

$$\hat{q} = \frac{\sum_{i=1}^{n} y_i \Big/ n}{\sum_{i=1}^{n} \ln(y_i) \Big/ n} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} \ln(y_i)} \tag{3.40}$$

3.6 MLE FOR 3-P PROBABILITY DISTRIBUTIONS

The three parameters are shape, scale and threshold parameter. The three parameter distributions applied on this study are Weibull, gamma and log-normal and are formulated as given in equations (2.7), (2.8) and (2.9) respectively.

The third parameter called threshold parameter is also known as the location parameter which determines where to shift the 3-p density function along the X-axis. The threshold parameter locates the distribution along the time scale and has same units as the distribution 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. It was therefore used to transform the data set to fit or resemble the

hypothesized distribution better. After obtaining the threshold parameter, it is subtracted 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, the AIC and BIC are used to estimate the threshold parameter. The threshold value with the lowest AIC and BIC values was considered to be the efficient and precise for further analysis. It was subtracted from the original data set and the resulting data set was then used for estimating the scale and shape parameters for both Weibull, Log-normal and Gamma probability distributions with 3-p using the same maximum likelihood estimates obtained for the 2-p under each of the three distributions.

3.7 MINIMUM DISTANCE ESTIMATION (MDE)

According to Rambachan (2018), the minimum distance method reduces the computational complexity since it omits the jacobian element which is usually present in the likelihood function. The method of minimum distance estimation depends on the test statistics of Anderson-Darling (AD) test (Lucen`o, 2006; Macdonald, 1971). The expression for Anderson-Darling based on minimum distance estimation is formulated as follows;

$$AD = A_n^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left(\log\left(y_{i:n}\right) + \log\left(1 - y_{(n+1-i)}\right)\right) \tag{3.41}$$

For the 2-p estimation for Weibull, gamma and log-normal using Minimum Distance Estimation method, the study uses the Anderson-Darling Minimum Distance estimator.

3.7.1 Anderson-Darling Estimation method

The MDE technique is based on the application of Anderson-darling statistics and is defined as Anderson-Darling estimator (ADE). A study developed this test as an alternative to statistical test

to be used significantly to examine sample distribution departure from normality (Anderson & Darling, 1952; D'Agostino & Stephens, 1986). By applying the Anderson-Darling test statistics, we can obtain the Anderson-Darling estimates $\widehat{M}_{ADE}$ and $\widehat{K}_{ADE}$ representing the scale and shape parameter estimates respectively for the three distributions from the following equation

$$A(m,k) = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left(\log F\left(y_{i:n} \setminus m,k\right) + \log V\left(y_{(n+1-i)} \setminus m,k\right)\right) \tag{3.42}$$

The estimates $\widehat{M}_{ADE}$ and $\widehat{K}_{ADE}$ are obtained by minimizing equation (3.41) with respect to m and k. Similarly, these estimates can be obtained from the solution of the following non-linear equations (Sanku & Menezes, 2019).

$$\sum_{i=1}^{n}(2i-1)\left[\frac{\Delta_1\left(y_{i:n} \setminus m,k\right)}{F\left(y_{i:n} \setminus m,k\right)} - \frac{\Delta_1\left(y_{(n+1-i)} \setminus m,k\right)}{V\left(y_{(n+1-i)} \setminus m,k\right)}\right] = 0 \tag{3.43}$$

$$\sum_{i=1}^{n}(2i-1)\left[\frac{\Delta_2\left(y_{i:n} \setminus m,k\right)}{F\left(y_{i:n} \setminus m,k\right)} - \frac{\Delta_2\left(y_{(n+1-i)} \setminus m,k\right)}{V\left(y_{(n+1-i)} \setminus m,k\right)}\right] = 0 \tag{3.44}$$

Where $\Delta_1(|m,k)$ and $\Delta_2(|m,k)$ in equation (3.43) and (3.44) are given as;

$$\Delta_1\left(y_{i:n} \setminus m,k\right) = \frac{\partial}{\partial m}F\left(y_{i:n} \setminus m,k\right) = \frac{\exp^{ky_i} - 1}{\left(\exp^{ky_i} - 1 + m\right)^2}$$

$$\Delta_2\left(y_{i:n} \setminus m,k\right) = \frac{\partial}{\partial k}F\left(y_{i:n} \setminus m,k\right) = \frac{ky\exp^{ky_i} - 1}{\left(\exp^{ky_i} - 1 + m\right)^2}$$

For the 3-p estimation, after getting the threshold value we apply the same Anderson-Darling estimation technique to get the other two parameters namely; shape and scale parameters.

3.8 TEST FOR EFFICIENCY BETWEEN MLE AND MDE TECHNIQUES

3.8.1 Comparison of the probability distributions

This was performed using the comparison between the Akaike's Information Criterion and the Bayesian Information Criterion for the two distributions where by the distribution with the smallest Akaike's and Bayesian Information Criterion values will be picked as the best. The Akaike's Information Criteria and the Bayesian Information Criteria are described in equation (3.2) and equation (3.3) respectively.

3.8.2 Test for Efficiency

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 need to understand how efficiency test is carried out. This was achieved using relative efficiency test. According to Dookie et al. (2018) study it is said that the method of MLE is popularly applied because its estimators are generally asymptotically consistent and unbiased. From the study conducted by Galvao and Wang (2015), it is concluded that MDE method is also unbiased estimation method. Since the two studies concluded that both MLE and MDE are unbiased, the relative efficiency test is given as follows;

$$Relative\ Efficency\ (R.E) = \frac{MSE(MLE)}{MSE(MDE)} = \frac{MSE(\hat{X}_1)}{MSE(\hat{X}_2)} = \frac{Var\ (\hat{X}_1)}{Var(\hat{X}_2)} \tag{3.45}$$

Where $\hat{X}_1$ and $\hat{X}_2$ are the estimators under MLE and MDE respectively.

If the ratio is less than 1, implies that MLE is more efficient (have smaller mean square error) and the estimator are therefore unbiased, sufficient and consistent. If the is greater than 1, then it indicates that MDE is more efficient meaning that it has small mean square error and therefore its estimates are unbiased, consistent and sufficient.

The relative efficiency for Weibull distribution, gamma distribution and log-normal distribution are as given in Table 3.4

Table 3.4: Relative Efficiency formulas

Distribution	Relative Efficiency
Gamma	$\frac{z_1 q_2^2}{z_2 q_1^2}$
Lognormal	$\frac{\left[\exp k_1^2 - 1\right] \exp\left(2v_1 - k_1^2\right)}{\left[\exp k_2^2 - 1\right] \exp\left(2v_2 - k_2^2\right)}$
Weibull	$\frac{p_2^{\frac{2}{b_2}}}{p_1^{2/b_1}}$

For the three parameter distribution, the relative efficiency formulations was the same as the formulations for two parameter distributions because the threshold value is the constant for both MLE and MDE techniques and if the threshold is introduced in the ratio, it will cancel itself making the relative efficiency formula for 3-p to fall back to the relative efficiency formula for 2-p.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 INTRODUCTION

Under this chapter, we present the results and findings obtained from the fitting of hourly wind data collected from the five sites in Narok County. The results are presented per objective that is; fitting using MLE, fitting using MDE and the efficiency test on the two fitting techniques. The analysis enabled us to understand the best 2-p or 3-p probability distribution for the three distributions under examination (Weibull, log-normal and gamma), that the wind data fitted.

4.2 FITTING WIND SPEED DATA TO PROBABILITY DISTRIBUTIONS USING MLE

4.2.1 2-p probability distributions analysis

a. Graphical analysis

Under this section, the study used histogram, QQ plots, PP lots and DCF graph to investigate the best distribution among the three distribution namely Weibull, Gamma and Log-normal.

From the histogram show it can observed that the level of peakedness for the frequency polygons are not uniform with log-normal appearing to be more peaked followed by gamma distribution and then Weibull even if both of them are peaked to the left. Between the wind speed of 2 m/s and 3.5 m/s we can observe that gamma and Weibull distributions are over estimating the values. From 4 m/s all the distributions under estimate the data. Therefore, from the histogram with frequency polygon it can be concluded that log-normal distribution is the best for predicting the wind speed data.

From Figure 4.1, the QQ plots it can be seen that for the wind speed less than 3 m/s all the three distributions have almost matching results but above 4 m/s there is bigger deviations which cannot be explained probabilistically with log-normal distribution recording more deviation that the other two. Therefore, from the QQ plot analysis, gamma and Weibull distributions shows better results.

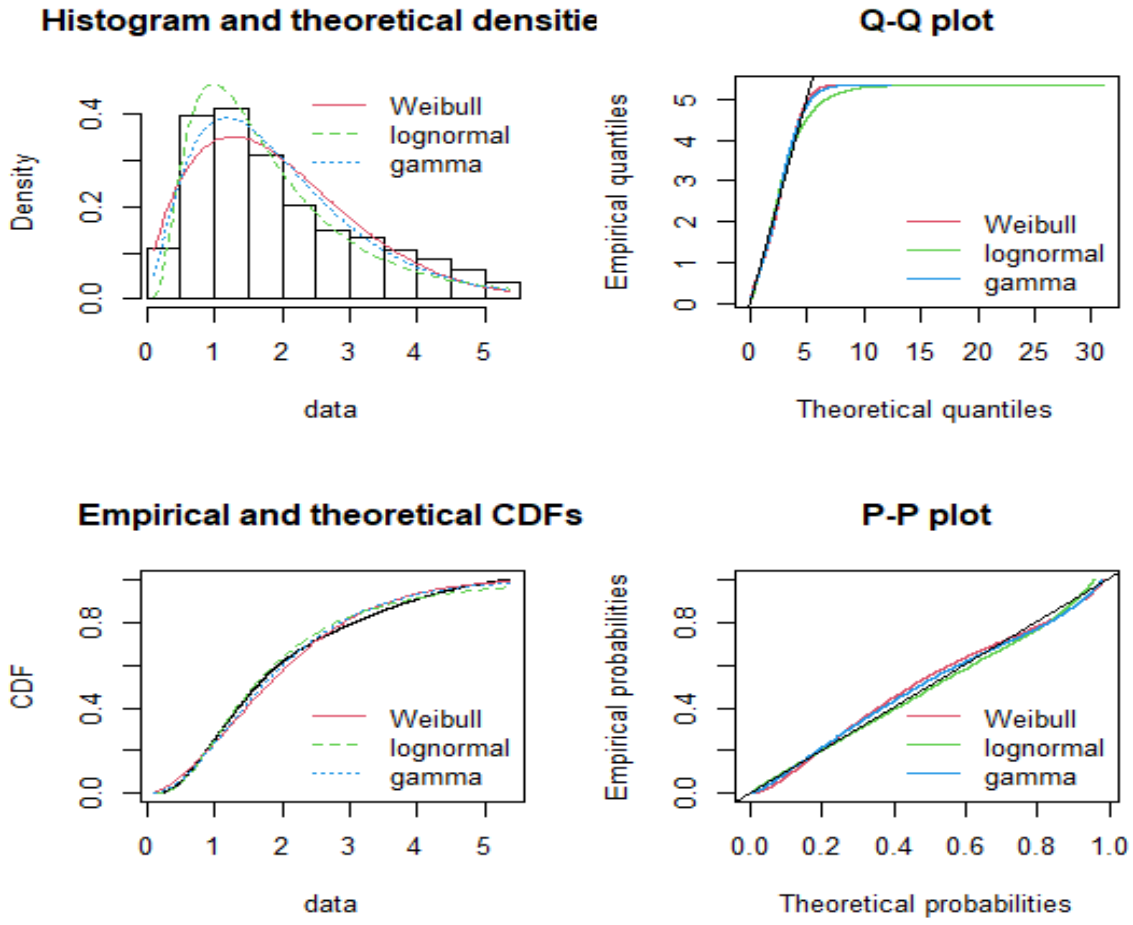

Figure 4.1: Graphical outputs for wind speed data

From Figure 4.1, the PP plot show more deviations between the best line of fit for the data and the three distribution though log-normal distribution seems to be more close to the best line than gamma and Weibull distributions.

Figure4.1, from the CDF curve it can observed that the three distributions are almost rimming up to 80 percent (probability of 0.8). From the graph, it can be seen that log-normal distribution looks better than gamma and Weibull distributions since it appears to be more closer to the best curve for the speeds below 4 m/s than the other two distributions.\\ By graphical presentation it can be said that log-normal is the best distribution for studying this wind speed data since the histogram, QQ plots and the CDF graph recommends the log-normal distribution over Weibull and gamma distributions.

b. 2-P probability distribution statistical analysis

Table 4.1 shows each distribution with its estimated parameters, standard error of the parameters and their correlations coefficients.

Table 4.1: Parameter Estimation

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

From Table 4.1 it can be explained that Weibull two parameters namely shape and scale parameters have weak positive correlation. Gamma parameters are also positively correlated with a strong positive correlation of 0.9. The log-normal two parameters shows a weak negative correlation. It is very important to examine if the estimated values of the parameters are useful in predicting the wind speed. This can be done by investigating the significance of each of the parameter under each distribution. This is achievable by applying the t test with the test statistic given as;

$$t^* = \frac{Estimator - Parameter}{Standard\ Error} = \frac{\beta - 0}{S.E} \quad (4.1)$$

The appropriate hypothesis test about the adequacy is given as;

H_0 : Estimator = 0 ($\beta = 0$)

H_1 : Estimator≠ 0 ($\beta \neq 0$).

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 4.2: t test for parameters

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

From Table 4.2, 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 is useful in predicting the wind speed since the estimated parameters are all not equal to zero. The t critical value used is 1.96 because for a sample size more than 100, the t critical value is 1.96 at 5 percent level of significance.

c. 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. The conclusion will depend on the values of AIC and BIC. The model with the smaller value for both the AIC and BIC is considered the best distribution for the study.

Table 4.3: Test of goodness of fit using MLE for 2-p

Statistics	Weibull	Log-normal	Gamma
Kolmogorov-Smirnov	0.05372	0.038854	0.036181
Criteria			
AIC	191777.5	192340.2	190407.2
BIC	191795.7	192358.4	190425.3

From Table 4.3, it is clear that the data fits all the distributions, Weibull, gamma and log-normal distributions. This is because Kolmogorov-Smirnov is used to investigate if the applied data follows a certain specified distribution. Therefore, using the Kolmogorov-Smirnov statistic it can be concluded that the data follows all the distributions because they have small statistical test values compared to the critical value of 0.136 hence, we fail to reject the null hypothesis. The critical value for Kolmogorov-Smirnov at 5 percent significance level is given by $\frac{1.36}{\sqrt{n}}$ for maximum value of n=100, any test with n above 100 use the n at 100 critical values which equals to 0.136.

From 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 wind speed data.

Therefore, using the MLE method for fitting 2-p probability distributions, it can be concluded that gamma distribution is the best distribution for studying this wind speed data.

4.22 3-P probability distribution analysis

a. Graphical analysis

From Figure 4.2, under histogram, log-normal is highly peaked followed by gamma then lastly Weibull distribution. But log-normal shows high level of under estimation than gamma distribution. In this case gamma is considered the best. For QQ plot, Weibull shows better results

compared to gamma and log-normal distributions since it can be used to estimate speeds up to 4.5 m/s.

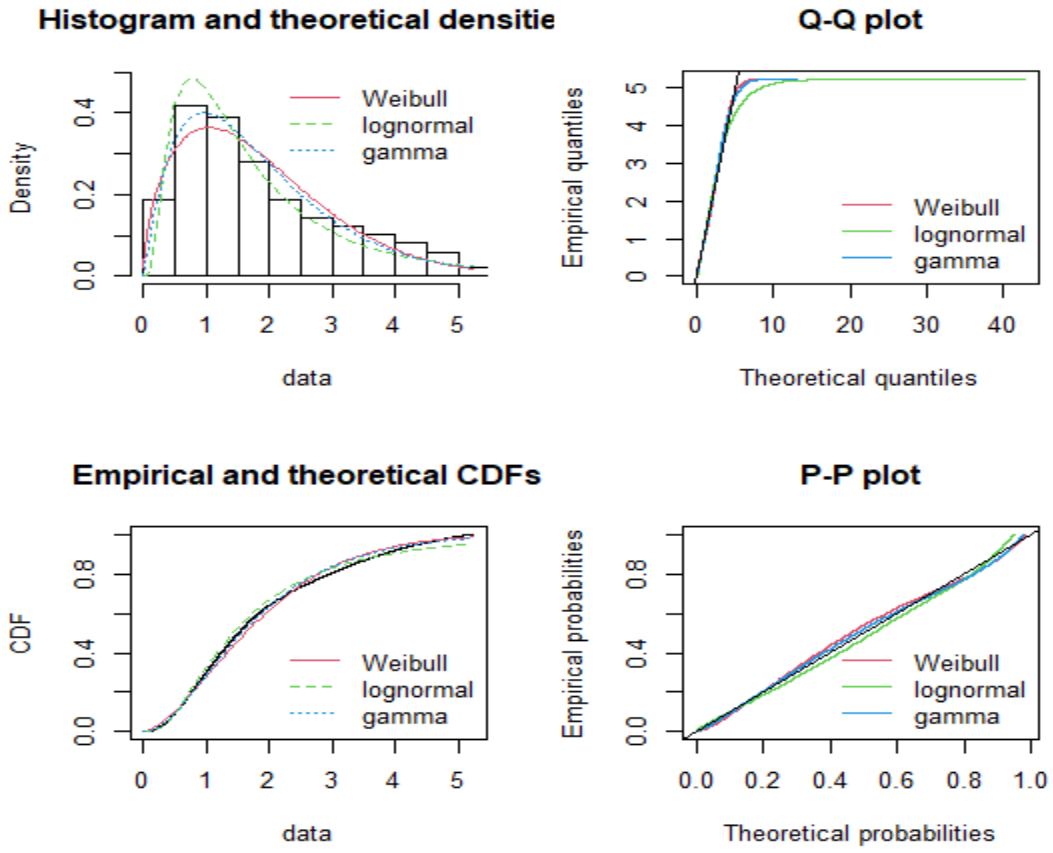

Figure 4.2: Graphical outputs for wind speed after subtracting threshold value.

From Figure 4.2, for the CDF graph, gamma distribution appears the best for probabilistically estimating wind speeds below 4 m/s followed by log-normal and lastly Weibull distribution. The case with P-P plot is also similar since gamma looks better for estimating almost 80 percent of the wind speed data. Its deviation from the best line is not big compared to the other two distributions, log-normal and Weibull.

From the graphical presentation, gamma three parameter distribution is considered the best because out of the four graphical presentations three of them displays gamma as the best for fitting

this data. Therefore, gamma three parameter distribution is the best using the MLE method since both the results from graphical analysis and statistical analysis proved that gamma is the best distribution.

b. Statistical analysis

Under this section, we checked on different level of thresholds to help us understand at what threshold value a certain distribution is at its optimal point. The threshold value in this case should not be greater than or equal to the minimum speed value (0.12 m/s) since our fitted value was the difference between the observations and the threshold value. For the research, the threshold value ranges from 0.1174 m/s to 0.1195 m/s as shown in the Table 4.4.

Table 4.4: Determination of threshold value

Threshold value	Distribution	AIC	BIC
0.1170	Weibull	190007.3	190025.4
	Gamma	189799.8	189817.9
	Log-normal	195747.6	195765.8
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

From Table 4.4, it can observed that as the threshold value increases, the AIC and BIC for gamma and log-normal distributions increases while for Weibull distribution the AIC and BIC value is showing slight change in the trend since they are almost rotating at nearly the same value. From the table it can be seen that in almost all aspects gamma distribution is reporting lower AIC and BIC values with the lowest AIC and BIC value observed under the threshold value of 0.1174 with AIC value of 189803 and BIC value of 189821.1.

Since speed cannot be negative as per the threshold for log-normal and also that gamma appeared to be having smaller AIC and BIC values for all the tested threshold values, the study opted to use the threshold value for gamma (0.1174 m/s) as the threshold value for the rest of the analysis under maximum likelihood method.

To confirm that the data follows these three distributions namely Weibull, gamma and log-normal a goodness test of statistics was performed using Kolmogorov-Smirnov, the results are given in the following table. From Table 4.5, it can be seen that data follows all the three distributions but it best follows the Weibull log-normal and gamma distribution since all the three statistic tests values are less compared to the Kolmogorov-Smirnov critical value of 0.136.

Table 4.5: K-S Statistics

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 4.6

Table 4.6: Parameters for 3-P probability distributions

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

Using MLE, we conclude that for the three parameter distributions, gamma three parameter distribution is the best for fitting the data since it has the smallest AIC value of 189803, BIC 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 4.7

Table 4.7: Best distribution under MLE

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

4.3 FITTING WIND SPEED DATA TO PROBABILITY DISTRIBUTION USING MDE

4.3.1 2-P probability distributions analysis

a. Graphical analysis

Under this section, various graphical techniques are discussed namely QQ plots, PP plots, CDF distribution and combined histogram with frequency polygons.

From Figure 4.4 under histogram, it can be seen that the log-normal distribution is more peaked than the gamma and the Weibull distribution even though all of them are positively skewed. From the frequency polygons, it can be seen that the distribution of log-normal have less cases of under

estimation and over estimation of the data hence log-normal is considered in this case the best distribution.

By considering the Q-Q plots in Figure 4.4, it can be observed that all the three distributions are almost estimating the same values from the minimum speed to a speed of 4 m/s. Above speed of 4 m/s, deviations can be observed but Weibull distribution seems to deviate less from the best line of fit. Therefore if we are to make our verdict using Q-Q plots then Weibull distribution can be marked as the best distribution for studying the data.

From the Cumulative density function graph, it can be observed that the speeds above 4 m/s cannot be probabilistically predicted accurately due to the deviations experienced but it can be seen from the curves that log-normal is much closer to the curve representing the observed data than Weibull and gamma distributions (for speeds below 4 m/s).

For the P-P plots, the deviations are observed but log-normal looks more close to the best line of fit with less deviations compared to Weibull and gamma distributions.

From the graphical analysis it can be seen that most of the plots/graphs like histogram, PP plots and CDF graph displays log-normal as the best distribution for fitting the data

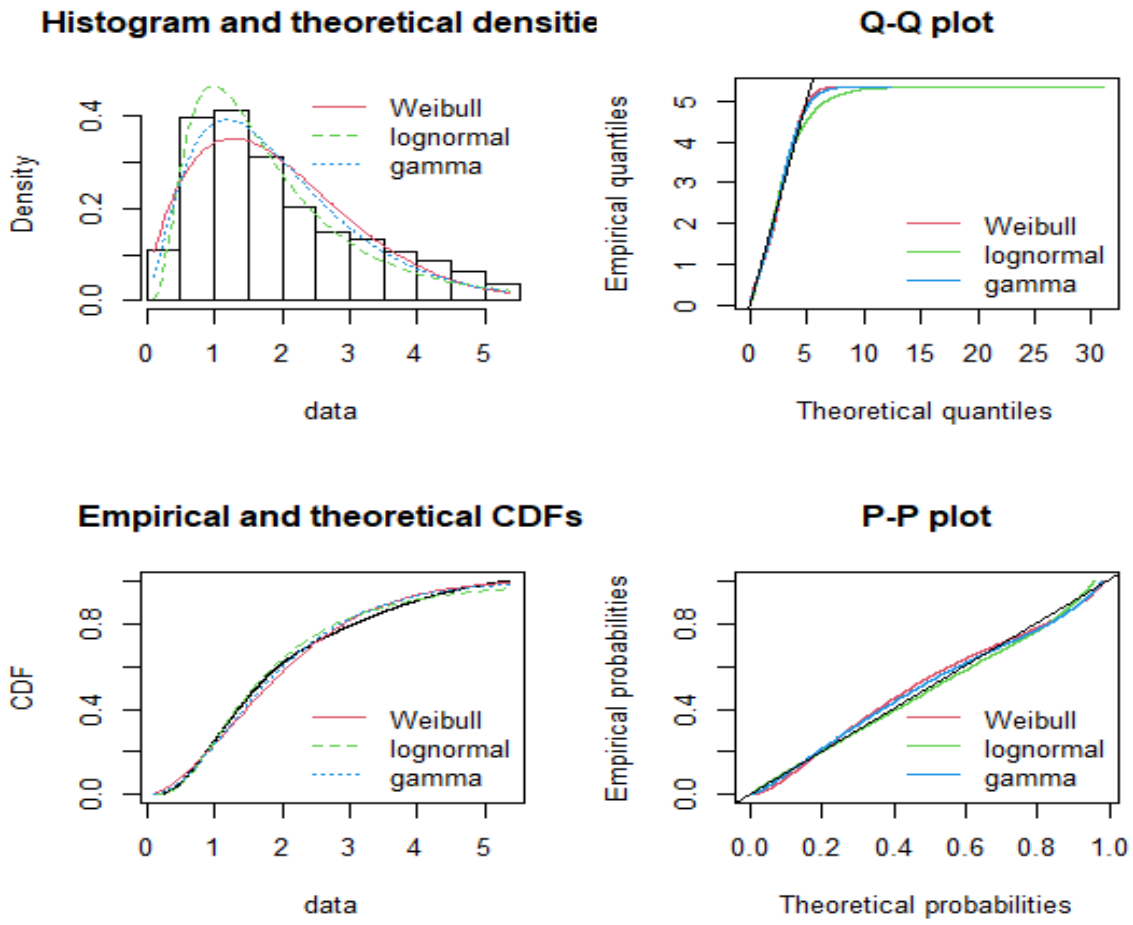

Figure 4.3: Graphical outputs for wind speed data

b. Test of goodness of fit analysis

By using the minimum distance technique method, the estimated parameters for the three distributions in examination are given as in the table 4.8.

Table 4.8: Estimated parameters for 2-P distributions using MDE

Distribution	Parameter	Estimate
Weibull	Shape Scale	1.502943 2.172747
Gamma	Shape Scale	2.107526 1.062046
Log-normal	Shape Scale	0.490416 0.68618

Under this technique there is need to identify the best distribution for fitting the data under study. Similar to the first case, both graphical analysis and goodness of fit analysis is discussed under this section.

Table 4.9 show the goodness test of fit criteria and goodness test of fit statistics. The goodness of test statistic is used to test if the distribution can fit the data under study while the goodness of fit criteria is used to understand the best distribution for the data.

Table 4.9: Test of goodness of fit using MDE for 2-P

Statistics	Weibull	Gamma	Log-normal
Kolmogorov-Smirnov	0.051438	0.031019	0.041869
Criteria			
AIC	192915.1	191315.6	192463.5
BIC	192933.3	191333.7	192481.6

Using the Kolmogorov-Smirnov statistic it can be confirmed that the data followed all the three distributions namely Weibull, gamma and log-normal since all of them give statistics test values lower than the critical value (0.136), this lead to a decision of not rejecting the null hypothesis for all the three distributions. Gamma fits the data best because from the AIC (191315.6) and BIC (191333.7) it is clearly evidenced that the distribution with smaller values is gamma distribution and therefore as per the decision rule it is considered the best of the three distributions for fitting the wind speed data.

Goodness of fit test is considered to be more accurate method for identifying the best distribution compared to the graphical methods because graphical methods are not most precise figures when it comes to estimating the value of each distribution as seen with goodness of fit test. Therefore it is concluded that by using the distance technique the gamma distribution is the best distribution for fitting the wind speed data and examining its characteristics with the AIC value of 191315.6 and BIC value of 191333.7.

4.3.2 3-P probability distributions analysis

a. Graphical analysis.

It is also important to understand the distribution of the wind data under these three distributions of interest hence it is very important to also study some graphical distributions.

From Figure 4.4 on the histogram, we can see that even though log-normal is more peaked than gamma and Weibull, it shows larger deviations compared to gamma and Weibull. Since gamma is more peaked than Weibull and its deviations are not much compared to log-normal, it can be picked as the best distribution for fitting this data.

For the P-P plot on Figure 4.4, gamma shows slight deviation which is almost uniform from the line of best fit hence from the PP lot distribution gamma is the best.

From Figure 4.4 on the QQ plot, all the three distributions are fitting the data well. The deviations from the line of best fit is observed to start at speed above 3.5 m/s with Weibull showing less deviation. Therefore using QQ plot Weibull is observed to be the best distribution for the study.

For the CDFs graph, gamma shows less deviations from the best line for almost 80 percent of the wind speed data compared to log-normal and Weibull distributions hence it is termed the best distribution since it can be used to probabilistically examine 80 percent of the data.

From the graphical analysis it can be concluded that gamma distribution is the best for studying this regions data since from the four graphs displayed three of them exposed gamma as the best distribution (histogram, PP plot and CDFs graph)

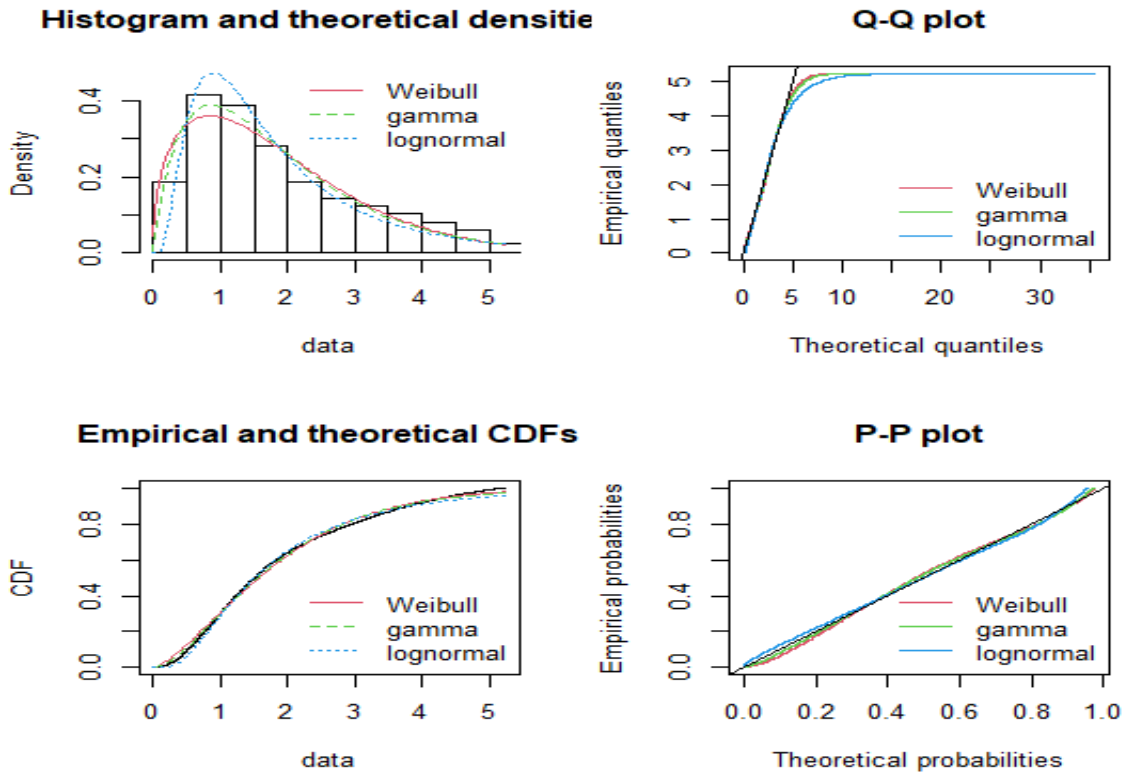

Figure 4.4: Graphical output after subtracting the threshold value

b. Statistical analysis

In this section we looked at both statistical analysis and the graphical representation of each distribution. First, there is need to investigate how the AIC and BIC behave under different value for the threshold parameter. This was investigated using the Table 4.10.

Table 4.10: Determination of Threshold value for MDE

Threshold value	Distribution	AIC	BIC
0.1170	Weibull	190791	190809.1
	Gamma	190226.4	190244.6
	Log-normal	196818.8	196837
0.1174	Weibull	190785.2	190803.3
	Gamma	190227.2	190245.3
	Log-normal	196888.9	196907
0.1180	Weibull	190777.2	190795.3
	Gamma	190228.9	190247
	Log-normal	197010.2	197028.3
0.1185	Weibull	190771.5	190789.6
	Gamma	190232	190250.1
	Log-normal	197133.1	197151.3
0.1190	Weibull	190766.7	190784.9
	Gamma	190237.3	190255.4
	Log-normal	197296.9	197315
0.1195	Weibull	190764.9	190783
	Gamma	190248.9	190267.1
	Log-normal	197561.8	197580
0.1199	Weibull	190773.6	190791.8
	Gamma	190279.	190298
	Log-normal	198192.7	198210.8

From Table 4.10 generated using the MDE method, it can be seen that gamma distribution has the smaller AIC and BIC value under all tested threshold values. This makes gamma the best distribution therefore we chose to use the threshold of 0.1174 for our analysis. This is because from the analysis using the original data, it was found that gamma has a threshold value of 0.1174. Therefore gamma is the best with the AIC value of 190227.2 and BIC value of 190245.3.

Using the threshold value of 0.1174, the other two parameters for all the three distributions are given in Table 4.11.

Table 4.11: Scale and shape parameters for 3-P distributions

Distribution	Parameter	Estimate
Weibull	Shape Scale	1.4115 2.038586
Gamma	Shape Scale	1.864567 0.993702
Log-normal	Shape Scale	0.4091521 0.732395

To be sure that our data followed this three specific distributions, we performed a statistical tests using the goodness of fit statistics. The goodness of fit statistics applied are summarized as shown in the Table 4.12.

Table 4.12: K-S statistics for 3-P distributions

Statistics	Weibull	Gamma	Log-normal
Kolmogorov-Smirnov	0.038007	0.028086	0.044478

From Table 4.12 it can be confirmed that all the three distributions that were recommended under Cullen and Frey graph, are still found to be statistically significant from the analysis using Kolmogorov-Smirnov test. This is clearly supported by the fact that all the Kolmogorov-Smirnov statistics values are less than the critical value 0.136, accepting the null hypothesis.

Therefore, using statistical analysis it can be summarized that gamma three parameter distribution is the best among the three distributions for studying the wind speed data.

Using the minimum distance estimation method, we therefore conclude that gamma three parameter distribution is the best with the following characteristics in table 4.13.

Table 4.13: Best distribution using MDE

Distribution	Criteria	Estimate
Gamma	AIC	190227.2
	BIC	190245.3
	Parameters	
	Threshold	0.1174
	Shape	1.864567
	Scale	0.993702

4.4 TEST OF EFFICIENCY AND COMPARISON OF THE DISTRIBUTIONS

4.4.1 Comparison of the probability distributions

The precision of this methods is based on the decision rule that the best method is the one which gives smaller AIC and BIC values.

By comparing the maximum likelihood method and the minimum distance method, we choose the method with the smaller AIC and BIC value for their best distributions which will be termed as the best distribution for studying the wind speed data.

From the Table 4.14, it can be observed that maximum likelihood estimation method yields smaller values for both AIC and BIC for both the two parameter distribution and the three parameter distribution. This leads us to a conclusion that the maximum likelihood estimation is the best estimation technique and the best distribution is gamma in both 2-parameter and 3-parameter distributions.

Lastly to know the best distribution between the two parameter and the three parameter we again compare their AIC and BIC value under maximum likelihood method since it is the best estimation technique. The decision rule relies on the distribution with the smaller AIC and BIC value. Since in both cases of two and three parameter distribution analysis we have gamma as the best

distribution, we now compare the AIC and BIC for this two gamma distributions. The comparison is as given in Table 4.15.

Table 4.14: Model comparison

Method	Parameters	Distribution	Criteria	Value
Maximum likelihood method	Two-parameter	Weibull	AIC	191777.5
			BIC	191795.7
		Log-normal	AIC	192340.2
			BIC	192358.4
		Gamma	AIC	190407.2
			BIC	190425.3
Minimum Distance method	Two-parameter	Weibull	AIC	192915.1
			BIC	192933.3
		Log-normal	AIC	192463.5
			BIC	192481.6
		Gamma	AIC	191315
			BIC	191333.7
Maximum likelihood method	Three parameters	Weibull	AIC	190003.8
			BIC	190021.9
		Log-normal	AIC	195803
			BIC	195821.1
		Gamma	AIC	189803
			BIC	189821.1
Minimum Distance method	Three parameters	Weibull	AIC	190785.2
			BIC	190803.3
		Log-normal	AIC	196888.9
			BIC	196907
		Gamma	AIC	190227.2
			BIC	190245.3

Table 4.15: Best distributions for 2-P and 3-P

Distribution	Criteria	Value
Gamma two parameter	AIC	190407.2
	BIC	190425.3
Gamma three parameter	AIC	189803
	BIC	189821.1

From Table 4.15 gamma distribution with 3 parameters has smaller AIC and BIC value compared to gamma distribution with 2-p. Therefore, gamma distribution with 3-p is the best distribution for examining wind speed data. This distribution has the following defined parameters given in Table 4.16.

Table 4.16: Best distribution estimates

Distribution	Parameter	Estimate
Gamma	Threshold	0.1174
	Shape	2.071773
	Scale	1.120855

The gamma three parameter distribution is given as follows;

$$f(y,z,q,t)=\frac{y t^{z-1}}{\Gamma(z/t)q^{z}}\exp\left[-\left(\frac{y}{q}\right)^{t}\right](z,y,q,t>0) \qquad (4.2)$$

Where:

q is the scale parameter.

z is shape parameters

t is the thresh-hold parameter

Therefore, the gamma distribution to be fitted will be as follows after inserting the estimated parameters in the equation.

$$f(y,z,q,t)=\frac{y 0.1174^{2.071773-1}}{\Gamma(2.071773/0.1174)1.120855^{2.071773}}\exp\left[-\left(\frac{y}{1.120855}\right)^{0.1174}\right] \qquad (4.3)$$

Where;

y>0 is the hourly wind speed data,

And Γ is a continuous gamma function given a;

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

The shape parameter shows the peakedness meaning that it represents the expected most frequent wind speed.

The scale parameter tells us how the region under study is windy, meaning that it help in knowing how the distribution of wind speed is expected to spread.

Threshold parameter assist in understanding the expected minimum wind speed value for the region of interest.

4.4.2 Relative Efficiency

The efficiency test is assisting us to judge the most efficient techniques between the two techniques namely; MLE and MDE fitting method. This test is also used to conclude on the efficient distribution. The efficiency of the distributions will be investigate for only the best 2-p distribution under the two fitting technique and for the best 3-p distribution under the two fitting techniques. For both techniques, the best distribution was gamma for 2-p and 3-p analysis. This means that the study used the relative efficiency formula for gamma distribution given by;

$$Gamma = \frac{z_1 q_2^2}{z_2 q_1^2},$$

Where; z is the shape and q is the scale parameter.

Table 4.17: Efficiency test for estimation techniques.

Best distribution	Technique	Parameter	Estimate	R. Efficiency
Gamma 2-P	MLE	Shape	2.47634	0.8349
		Scale	1.25991	
	MDE	Shape	2.107526	
		Scale	1.062046	
Gamma 3-P	MLE	Shape	2.071773	0.8733
		Scale	1.120855	
	MDE	Shape	1.864567	
		Scale	0.993702	

From the results in Table 4.17, the relative efficiencies are 0.8348 and 0.8733 respectively for the best 2-parameter and 3- parameter distributions under MLE and MDE techniques or methods. Because the relative efficiencies are both less than 1, we conclude that MLE is more efficient than MDE and therefore, its shape and scale estimates are unbiased, sufficient and consistent.

There is also need to examine the efficiency between the best two distributions obtained in the study for 2-p and 3-p fitting. Maximum likelihood estimation fitting method obtained gamma distribution as the best in 2-p and 3-p analysis for fitting the wind speed data. Table 4.18 show the relative efficiency between the best two distributions given by MLE

$$R.E = \frac{Var(Gamma(2-P))}{Var(Gamma(3-P))}$$

Table 4.18: Efficiency test for the best 2-P and 3-P distributions

Distribution	Technique	Parameter	Estimate	R. Efficiency
Gamma 2-P	MLE	Shape	2.47634	1.0571
		Scale	1.25991	
Gamma 3-P	MLE	Shape	2.071773	
		Scale	1.120855	

From Table 4.18, the relative efficiency value is 1.0571 which is greater than 1, indicating that gamma distribution with 3-p is more efficient compared to gamma distribution with 2-p. This leads

us to a conclusion that gamma with 3-p is the best distribution for fitting this wind speed data, it is still the efficient distribution for fitting the wind speed data because its estimated parameters are confirmed to be consistent, sufficient and unbiased.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 INTRODUCTION

The data was composed of 66859 hourly wind speed observations. Out of these observations, only 63778 observations were used in the analysis meaning that 3080 observations were left out as outliers. These outliers are as a result of extreme wind speed observations caused by wind gusts. The wind gusts is a brief and unexpected increase in the speed of wind followed by lull. They are known to be caused by solar heating of the ground, turbulence due to friction and/or wind shear (change in wind over distance which can be change in wind direction, speed or both). In the case of solar heating of the ground, it happens when the ground is heated on sunny days and is due to rising air currents that can generate a thermal warm air that rises, with air from above sinking to replace the rising thermal warm air. This descending air cause wind gusts. In the case of friction, gusts are generated when wind blows around trees, buildings or other forms of obstacles.

From the analysis of hourly wind speed data using the MLE and MDE methods on the two parameter distributions of Weibull, gamma and log-normal, based on the method of analysis the study found that MLE was the best method of fitting the two parameter distributions. This was because it gave the smaller AIC and BIC value compared to the MDE method. Among the three distributions under examination, gamma distribution gives the smaller AIC and BIC values (190407.2, 190425.3) making it to be the best two parameter distribution for fitting this hourly wind speed data.

For the three parameter analysis using the same decision rule that the method or/and distribution with the smaller AIC and BIC values is the best method for fitting the data to the distribution. The

study reached a conclusion that MLE method was the best technique for fitting the three parameter distributions. From the distributions, gamma three parameter was picked as the best for studying and predicting wind speeds in Kenya since it had a smaller AIC of 189803 and a smaller BIC value of 189821.1.

Lastly for the efficiency test on the two methods namely MLE and MDE methods on the two best distributions namely; gamma two parameter and gamma three parameter distributions, the study found that MLE method was more precise than the distance method this was clearly supported by the evidence that the the relative efficiency ratio gives a value less than 1 for the techniques for both 2-parameters and 3-parameters analysis. The gamma distribution with three parameters is found to be more efficient compared to the gamma distribution with 2-parameters. Therefore gamma three parameter distribution was concluded to the best and efficient distribution for fitting the wind speed data in Kenya.

5.2 CONCLUSION

In conclusion, and based upon the objectives stated in chapter 1, we found that gamma distribution with 3-p is best under both MLE and MDE fitting techniques. This is because it yields lower AIC and BIC values compared to Weibull and Log-normal distributions. Also, from the analysis, using relative efficiency we can reach a conclusion that MLE is the efficient technique/method for fitting the wind speed data to a probability distributions. We also conclude that gamma distribution with three parameters is the best and efficient distribution for fitting wind speed data based on comparison of the distribution analysis and relative efficiency test.

5.3 RECOMMENDATION

From the study, it is recommended that future researchers should incorporate the gamma distribution with three parameters in their analysis of wind speed characteristics since it will give the best and efficient wind speed probabilities compared to the other form of distributions.

It is also recommended that for further work regarding the study of the distributions, MLE method can be used to estimate the parameters since it gives precise estimates.

The study also recommend to researchers, to use other datasets from other parts of the world to examine if the Minimum Distance Estimation techniques can give efficient estimates compared to other fitting techniques like Method of Moments and Least Square Estimation.

REFERENCES

Akyuz, E., & Gamgam, H., (2017). Statistical analysis of wind speed data with Weibull and gamma distributions: *European journal of scientific research*. 6, 131-146

Anderson, T.W., & Darling, D.A., (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes: *The Annals of mathematical statistics*. 23, 193-212.

Ayodele, R., Adisa, A., Munda, L., & 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*. 5, 88-107.

Azami, Z., Khadijah, S., Mahir, A., & Sopian, K. (2009). Wind speed analysis in east coast of Malaysia: *European journal of scientific research*. 32, 208-215.

Barasa, M., (2013). Wind regime analysis and reserve estimation in Kenya. Available at: http://www.secheresse.info/spip.php?rubrique2931

Brenda, F.G. (2009). Parameter estimation for log-normal distribution: *Brigham Young University*. 11, 121-133.

Celik, H., & Yilmaz, V., (2008). A statistical approach to estimate the wind speed distribution: *The case study of Gelubolu region*. 2, 122-132.

Christopher M. B., (2006). Pattern Recognition and Machine Learning (Information Science and Statistics: *Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006*. ISBN 0387310738.

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

D'Agostino, R., & Stephens, M., (1986). Goodness of fit techniques. Available at: http://www.hep.uniovi.es/sscruz.

Dookie, I., Rocke, S., Singh, A., & Craig, J.R. (2018). Evaluating wind speed probability distribution models with novel goodness of fit metric: A Trinidad and Tobacco case study. *International journal of energy and environmental engineering*.7, 33-59.

Galvao, F.A., & Wang, L., (2015). Efficient minimum distance estimator for quantile regression fixed effects panel data: *Journal of multivariate analysis*. 16, 211-243. DOI:10.1093/jjfinec/nbx016

Gungor A. & Eskin, N. (2008). The characteristics that defines wind as an energy source. Available at: https://www.semanticscholar.org/paper.

Gupta, R., & Biswas, A. (2010). Wind data analysis of Silchar (Assam India) by Rayleigh and Weibull methods: Journal of mechanical engineering research. 2, 10-24.

Hurlin, C., (2013). Maximum likelihood estimation: Advanced econometrics, University of Orleans. Available at: https://www.univ-orleans.fr/deg/masters.

Johnson, W., Donna, V., & Smith, L., (2010). Comparison of estimators for parameters of gamma distributions with left-truncated samples. *Journal of Statistics Education*. Available at: www.amstat.org/publications/jse/v23n1/johnson.pdf.

Lawan, S.M., Abidin, W.A.W.Z., Chai, W.Y., Baharum, A., & Masri, T., (2015). Statistical modelling of long-term wind speed data: *American journal of computer science and information technology*. 13, 79-121.

Louzada, F., Ramos, P.L., & Gleici, S.C.P. (2016). Different estimation procedures for the parameters of the extended exponential geometric distribution for medical data: *Computational and mathematical methods in medicine*. 5, 55-74.

Lu, L., Yang, H., & Burnett, J. (2002). Investigation on wind power potential on Hong Kong islands-an analysis of wind power and wind turbine characteristics: *Renewable energy*, 1-12.

Lucen`o, A. Fitting the generalized pareto distribution to data using maximum goodness of fit estimators 2006: *Computational statistics and data analysis*. 51, 904-917.

Macdonald, P.D.M. (1971). Comment on, 'an estimation procedure for mixtures of distributions by Choi and Bulgren' : *Journal of the royal statistical society*. 33, 326-329.

Mahyoub, H., (2006). Statistical analysis of wind speed data and an assessment of wind energy: Potential in Taiz-Yemen. 2.

Maleki, F., & Deiri, E., (2007). Methods of estimation for three parameter reflected Weibull distribution. Available at: https://www.semanticscholar.org/paper.

Mert, I., & 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*. 4, 17-33.

Mumford, A.D., (1997). Robust parameter estimation for mixed Weibull (Seven parameters) including the method of maximum likelihood and the method of minimum distance: *Department of air force, Air force institute of technology*.

DOI: 10.13140/RG.2.1.2868.6566

Oludhe, C., (1987). Statistical characteristics of wind power in Kenya: *University of Nairobi*. Available at: http://erepository.uonbi.ac.ke/bitstream.

Otieno, C. S., (2011). Analysis of wind speed based on Weibull model and data correlation for wind pattern description for a selected site in Juja, Kenya.

DOI: 10.20944/preprints201110.0256.v1

Otieno, F., Gaston, S., Kabende, E., Nkunda, F., & Ndeda, H., (2014). Wind power potential in Kigali and western provinces of Rwanda: *Asia pacific journal of energy and environment.* 1, 189-199.

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

DOI: 10.3390/sym12050813.

Rambachan, A., (2018). Maximum likelihood estimates and Minimum distance estimate. *International Journal of Statistical Distributions and Applications*. 6(3), 57-64

Saleh, H., Abou, A. S., & Abdel-Hady, S., (2012). Assessment of different methods used to estimate Weibull distribution parameters for wind speeds in Zafarana wind farm, Suez gulf, Egypt.

Salma, O.B., & Abdelali A.E., (2018). Comparing maximum likelihood, least square and method of moments for Tas distribution: *Journal of Humanities and Applies science.* 11, 1-19

Sanku, D., Menezes, A.F.B., & Mazucheli, J., (2019). Comparison of estimation methods for unit-Gamma distribution: *Journal of data science*. 17, 768-801.

Solar and wind energy resource assessment. (2008). Kenya country report. 1-68.

Sukkiramathi, K., Seshaiah, C., & Indhumathy, D., (2014). A study of Weibull distribution to analyse the wind speed at Jogimatti in India: 1, 189-193.

Sultan, M.M.A., (2008). A data driven parameter estimation for the three parameter Weibull population from censored samples: *Mathematical and computational applications*. 13, 129 – 136.

Ulgen, K., & Hepbasli, A., (2002). Determination of Weibull parameters for wind energy analysis of Izmir, Turkey. *Journal of Energy in Southern Africa*. 23(2), 30-38

Ulgen, K., Genc, A., Hepbasli, A., & Oturanc, G., (2009). Assessment of wind characteristics for energy generation. DOI: http://dx.doi.org/10.17159/2413-3051/2012/v23i2a3152.

Wind sector prospectus Kenya, (2013). Wind energy data analysis and development program. DOI: 10.1016/j.rser.2013.12.061

Woodward, A.W., William, R. S., Lindsey, H., & Gray, H. L., (1982). Comparison of minimum distance and maximum likelihood technique for proportion estimation: *Department of statistics, Southern Methodist University*. Available at: https://www.tandfonline.com. 590 – 598.

Zheng, S., (2018). Maximum likelihood estimation: Statistical theory II. *Journal of modern applied statistical methods*. 17, 1-17.

APPENDIX

Appendix 1

R codes

```
data=read.csv(file.choose(),header=TRUE)#Importing data
attach(data)
mean(Speed)
sd(Speed)
median(Speed)
boxplot(Speed)
outlier=function(x){
Q1<-quantile(x,probs=0.25)
Q2<-quantile(x,probs=0.75)
iqr<-IQR(x)
low=Q1-(1.5*iqr)
up=Q2+(1.5*iqr)
x2<-subset(x,x>=low )
x3<-subset(x2,x2<=up)
return(x3)
}
Speed=outlier(Speed)
boxplot(Speed)
library("fitdistrplus")
plotdist(Speed, histo=TRUE, demp=TRUE)
descdist(Speed, boot=1000)
dist1=fitdist(Speed, "weibull")
summary(dist1)
dist2=fitdist(Speed, "gamma")
summary(dist2)
dist3=fitdist(Speed, "lnorm")
summary(dist3)
par(mfrow = c(2, 2))
```

```
plot.legend <- c("Weibull", "lognormal", "gamma")
denscomp(list(dist1,dist3,dist2), legendtext = plot.legend)
qqcomp(list(dist1,dist3,dist2), legendtext = plot.legend)
cdfcomp(list(dist1,dist3,dist2), legendtext = plot.legend)
ppcomp(list(dist1,dist3,dist2), legendtext = plot.legend)
gofstat(list(dist1,dist3,dist2),fitnames=c("weibull","lnorm","gamma"))
mod4=fitdist(Speed, "weibull",method="mge", gof="ADR")
summary(mod4)
mod5=fitdist(Speed, "gamma",method="mge", gof="ADR")
summary(mod5)
mod6=fitdist(Speed, "lnorm",method="mge", gof="ADR")
summary(mod6)
par(mfrow = c(2, 2))
plot.legend <- c("Weibull", "gamma","lognormal")
denscomp(list(mod4,mod5,mod6), legendtext = plot.legend)
qqcomp(list(mod4,mod5,mod6), legendtext = plot.legend)
cdfcomp(list(mod4,mod5,mod6), legendtext = plot.legend)
ppcomp(list(mod4,mod5,mod6), legendtext = plot.legend)
gofstat(list(mod4,mod5,mod6),fitnames=c("weibull","gamma","lnorm"))
#Three parameter distribution fitting
minimum=min(Speed)
minimum
Speed2=Speed-0.1195
plotdist(Speed2, histo=TRUE, demp=TRUE)
descdist(Speed2, boot=1000)
dist12=fitdist(Speed2, "weibull")
summary(dist12)
dist22=fitdist(Speed2, "gamma")
summary(dist22)
dist32=fitdist(Speed2, "lnorm")
```

```
summary(dist32)
par(mfrow = c(2, 2))
plot.legend <- c("Weibull", "lognormal", "gamma")
denscomp(list(dist12,dist32,dist22), legendtext = plot.legend)
qqcomp(list(dist12,dist32,dist22), legendtext = plot.legend)
cdfcomp(list(dist12,dist32,dist22), legendtext = plot.legend)
ppcomp(list(dist12,dist32,dist22), legendtext = plot.legend)
gofstat(list(dist12,dist32,dist22),fitnames=c("weibull","lnorm","gamma"))
#Distance Method
mod42=fitdist(Speed2, "weibull",method="mge", gof="ADR")
summary(mod42)
mod52=fitdist(Speed2, "gamma",method="mge", gof="ADR")
summary(mod52)
mod62=fitdist(Speed2, "lnorm",method="mge", gof="ADR")
summary(mod62)
par(mfrow = c(2, 2))
plot.legend <- c("Weibull", "gamma", "lognormal")
denscomp(list(mod42,mod52,mod62), legendtext = plot.legend)
qqcomp(list(mod42,mod52,mod62), legendtext = plot.legend)
cdfcomp(list(mod42,mod52,mod62), legendtext = plot.legend)
ppcomp(list(mod42,mod52,mod62), legendtext = plot.legend)
```