ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

EVALUATION OF PROBABILITY DISTRIBUTION FUNCTIONS FOR WIND SPEED ANALYSIS: A CASE STUDY OF ADDIS ABABA

A Thesis in Structural Engineering

By: Abdu Lelamo Ebrahim

September 2020

Addis Ababa, Ethiopia

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

The undersigned have examined the thesis entitled 'Evaluation of Probability Distribution Functions for Wind Speed Analysis: A Case Study of Addis Ababa' presented by **Abdu Lelamo Ebrahim**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

Adil Zekaria (Dr. -Ing.)		
Advisor	Signature	Date

Esayas G/ Youhannes (Dr. -Ing.)		
Internal Examiner	Signature	Date

Girma Z/Yohannes (Dr. -Ing.)		
External Examiner	Signature	Date

Chairman	Signature	Date

UNDERTAKING

This is to Certify that the Thesis entitled "*Evaluation of Probability Distribution Functions for Wind Speed Analysis: A Case Study of Addis Ababa"* Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Structural Engineering is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/ referred.

Signature: ____________

Undertaker's Full Name: Abdu Lelamo Ebrahim

Place of Undertaking: Addis Ababa, Ethiopia

Date: September, 2020

ABSTRACT

Selection of best fit probability distribution model to the wind data sets is very important for reducing uncertainties in the extreme wind speed modeling at a given site. This study presents extreme wind speed analyses by using two approaches: The Block Maxima and Peaks Over Threshold, and compare of the results of both methods for the data from Addis Ababa; Bole wind speed recording station. Wind data collected include wind speed and wind direction, recorded for 63 years (1954-2016) with three-hours' time interval. Seven Two-parameter, five Three-parameter, four Four-parameter, and Five-parameter Wakeby distributions are fitted to the Block Maximum data series, and Generalized Pareto distribution is fitted to the Peaks Over Threshold series. Three parameter estimation techniques were considered for estimating parameters involved with these distributions namely, Maximum Likelihood, L-moments, and Methods of Moments. The best fit models to the data are selected by examining Probability-Probability Plots and four goodness-of-fit statistics: Root Mean Square Error, Coefficient of Determination, Kolmogorov-Smirnov, and Cramer-VonMises, at 95 percent confidence level. The L-Moments estimation method has performed better for calculating the parameters of most of the distributions while the Method of Moments is the preferred method for obtaining the parameters of the JohnsonSB and Kumaraswamy distributions. The results showed that the JohnsonSB distribution gives a best fit to the block maximum series. The Peaks Over Threshold method with 2 peaks per year gave better results than the Block Maximum method; as a result, Peaks Over Threshold method is recommended for design. Wind direction analysis along with a wind rose chart for the study area is also provided. Analysis showed that most of the winds come from the East and East-Southeast direction with the maximum magnitude of 3.60- 5.70 m/s. Finally, the selected distribution model is used for forecasting the extreme wind speeds for return periods of 5, 10, 20, 50 and 100 years.

Key Words: Block Maximum, Extreme Wind Speed, Goodness-of-Fit, L-Moments, Maximum Likelihood, Peaks Over Threshold, Probability Distribution Model.

ACKNOWLEDGMENTS

All praise is due to **ALLAH**. I praise Him, seek His help and ask His forgiveness. I thank Him for giving me the strength and patience to pass various obstacles happened to my life and come up to the accomplishment of this thesis. Without His will I would not be able to do anything at all. I glorify Him for His infinite mercy throughout my life; ask Him to grant me His love and the love of those who loved Him.

This Master's Thesis has been carried out at the school of civil and environmental engineering at Addis Ababa University Institute of Technology (AAiT) during the academic year 2019/20 - 20/21. A number of people deserve thanks for their support and help. It is therefore my greatest pleasure to express my gratitude to them all in this acknowledgement.

First of all, I would like to thank my thesis advisor, Dr.-Ing. Adil Zekaria who gave me help and time to assist in the completion of this thesis. I also gratefully acknowledged him for providing me the wind data.

I would like to thank Sadik Muzeyn (PhD candidate and lecturer at AAiT) for offering helpful suggestions and for providing illustrative materials for this thesis.

The Scholarship awarded to me by Ethiopian Roads Authority/ERA under the program of postgraduate education is gratefully acknowledged.

I am grateful to SNNPR Siltie Zone East Azernet Berbere Woreda for their unconditional support throughout the whole period of my graduate study.

Finally, I would like to express special thanks to my family, who always have supported me through life and during my studies, especially my beloved brother Aman Lelamo, thank you for standing beside me.

Dedication

To my family

TABLE OF CONTENTS

ABSTRACT 2

ACKNOWLEDGMENTS 3

DEDICATION 4

TABLE OF CONTENTS 5

LIST OF TABLES 9

LIST OF FIGURES 10

LIST OF ABBREVIATIONS AND ACRONYMS 11

CHAPTER 1 INTRODUCTION 13

1.1 General background 13
1.2 Statement of the problem 15
1.3 Research questions 16
1.4 Objectives of the study 16
1.4.1 Main objective 16
1.4.2 Specific objectives 17
1.5 Scope of the study and limitations 17
1.6 Ethical considerations 17
1.7 Organization of the study 17

CHAPTER 2 LITERATURE REVIEW 19

2.1 Chapter introduction 19
2.2 Overview of wind speed distribution models 19
2.3 Approaches to estimation of extreme wind speed 21
2.3.1 Block Maxima (BM) approach 21
2.3.2 Limitations of BM approach 21
2.3.3 Peaks-Over-Threshold (POT) approach 21
2.3.4 Limitations of POT approach 22

2.3.5 Independent and identically distributed data (i.i.d.) 22
2.3.6 Independence criteria 23
2.3.7 Threshold selection techniques 23
2.4 Comparison of BM and POT approaches 24
2.5 Overview of parameter estimation methods 24
2.6 Goodness-of-Fit (GoF) tests 25
CHAPTER 3 RESEARCH METHODOLOGY 27
3.1 Introduction 27
3.2 Site selection 29
3.3 Description of the study site 30
3.4 Data quality assessment of the station 30
3.5 Actual wind speed observations and check for missing records 31
3.6 A procedure followed for extracting the maxima 31
3.7 Check for outliers 32
3.7.1 Graphical way of checking outliers using time series plots 32
3.7.2 Statistical way of checking outliers using Grubb's Test 33
3.8 Extreme value analysis methods 34
3.9 General description of some possible distribution models for BM analysis 34
3.10 Parameter fitting methods 38
3.10.1 Methods of Moments Estimators (MOMEs) 38
3.10.2 Probability Weighted Moments (PWMs) equations 39
3.10.3 L-Moments Estimators (L-MEs) 39
3.10.4 Maximum Likelihood Estimators (MLEs) 40
3.10.5 The numerical iterative methods used 40
3.11 Fitting techniques used for distribution models 40
3.12 Peaks over a Threshold (POT) approach 43
3.12.1 Threshold selection approaches followed in this thesis 44

3.12.2 Quantile estimation using GPD 44

3.13 Initial visualization and model choice 44

3.13.1 Probability-Probability Plots (P-P Plots) 44

3.13.2 Plotting position formula used in this thesis 45

3.14 Quantitative Goodness of Fit (GoF) tests 45

3.14.1 Coefficient of Determination (R^2) test 45

3.14.2 Root Mean Square Error (RMSE) test 46

3.14.3 Kolmogorov-Smirnov (K-S) test 46

3.14.4 Cramer-VonMises (C-VM) test 47

3.15 Return period and forecasting 47

CHAPTER 4 RESULTS AND DISCUSSION 48

4.1 Introduction 48

4.2 Graphical display of original and outlier corrected yearly maximum data 48

4.3 Wind direction analysis 49

4.4 Fitted results of distributions to outlier corrected BM data 50

4.4.1 Goodness of fit assessment using a probability–probability (P–P) plots 52

4.4.2 Goodness of Fit (GoF) test-based analysis 54

4.5 Selection of the best fitting distributions 56

4.6 Analysis results of POT series 59

4.7 Best fitting POT series 60

4.8 Comparison of analysis results between POT and BM series 62

4.9 Quantile estimates based on BM approach 62

4.10 Quantile estimates based on POT approach 63

CHAPTER 5 CONCLUSIONS AND RECCOMENDATIONS 65

5.1 Conclusions 65

5.2 Recommendations 67

5.3 Limitations 67

REFERENCES **68**

APPENDIX A **72**

APPENDIX B **73**

APPENDIX C **74**

APPENDIX D **80**

LIST OF TABLES

Table 3-1. Wind dataset acquired from NMAE for Addis Ababa-Bole station. 30
Table 3-2. Missing data points for a number of months in the data set. 31
Table 3-3. Two-parameter distribution models CDF, PDF, Quantile, and Range. 35
Table 3-4. Three-parameter distribution models CDF, PDF, Quantile, and Range. 36
Table 3-5. Four and Five-parameter distribution models CDF, PDF, Quantile, and Range. 37
Table 3-6. L-Moments and L-Moment ratios. 39
Table 3-7. Fitting techniques used for Two-parameter distribution models. 41
Table 3-8. Fitting techniques used for Three-parameter distribution models. 42
Table 3-9. Fitting techniques used for Four and Five-parameter distribution models. 43
Table 4-1. Wind speed frequency according to wind direction at the study site 49
Table 4-2. Fitted Parameters of two parameter distributions. 51
Table 4-3. Fitted Parameters of three parameter distributions. 51
Table 4-4. Fitted Parameters of Four and Five parameter distributions. 52
Table 4-5. Results of GoF tests for Two-parameter distributions. 54
Table 4-6. Results of GoF tests for Three-parameter distributions. 55
Table 4-7. Results of GoF tests for Four-parameter distributions. 56
Table 4-8. Ranking of Two-parameter distributions based on the GoF statistics. 57
Table 4-9. Ranking of Three-parameter distributions based on the GoF statistics. 58
Table 4-10. Ranking of Four-parameter distributions based on the GoF statistics. 58
Table 4-11. Total ranking of all distributions based the BM data. 59
Table 4-12. Results for the POT method and parameters estimated with the MLE, L-ME. . 60
Table 4-13. GoF results of different POT values for selection of best POT level. 61
Table 4-14. Comparison of GoF results of best fit BM data and selected POT values......... 62
Table 4-15. Return levels (in m/s) for all distribution models based on BM approach. 63
Table 4-16. Return levels (in m/s) based on the POT 2 series. 63
Table A-5-1. Yearly maximum wind speed for Addis Ababa at Bole recording station. 72
Table B-5-2. Some descriptive statistics of the data used for analysis in this thesis............ 73
Table C-5-3. All outliers detected from daily maximum data. 74

LIST OF FIGURES

Figure 3-1. Methodology flow chart of the research approach. 28

Figure 3-2. Stations distribution over Ethiopia as updated on December 2012 [41]. 29

Figure 3-3. Time series plot of daily maximum wind speed at 1970 (left) and 1994 (right) 32

Figure 3-4. Outlier plot of yearly maximum outlier corrected data. 33

Figure 4-1. Time series plot of yearly maximum original and outlier corrected data. 48

Figure 4-2. Wind rose chart for observed wind speed data at Addis Ababa- Bole site. 50

Figure 4-3. P-P plots of Two-Parameter distributions to BM data. 53

Figure 4-4. P-P plots of Three-Parameter distributions to BM data. 53

Figure 4-5. P-P plot of Four-Parameter distributions to BM data. 53

Figure 4-6. P-P plots of all candidate POT values. 61

Figure C-5-1. Graphical display of all outliers detected from the data. 79

LIST OF ABBREVIATIONS AND ACRONYMS

AAiT	Addis Ababa Institute of Technology
AM	Annual Maximum
BM	Block Maximum
CDF	Cumulative Distribution Function
Crit.	Critical
Dr,-Ing.	Doctor Engineer
EM	Empirical Method
Eq.	Equation
ERA	Ethiopian Roads Authority
EV	Extreme Value
EVT	Extreme Value Theory
EW	Exponential Weibull
Exp	Exponential
G	Grubb’s test
G.C.	Gregorian Calender
GAM	Gamma distribution
GAM3	Three parameter gamma distribution
GEV	Generalized Extreme Value Distribution
GLD	Generalized Lambda Distribution
GMM	Generalized Methods of Moments
GoF	Goodness of Fit
GP	Generalized Pareto
GPD	Generalized Pareto Distribution
GUM	Gumbel distribution
IW	Inverse Weibull
JSB	Johnson Bounded distribution
KAP	Kappa distribution
K-S	KolmogorovSmirnov goodness of fit statistics
KUM	Kumaraswamy distribution
LL	Log-Logistic distribution
L-M	L-Moments
L-MEs	L-Moments Estimators
LN	Lognormal distribution
LN3	Three parameter Lognormal distribution
LSE	Least Squares Estimation
Max	The maximum value in the set of data
Min	The minimum value in the set of data
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
MLEs	Maximum Likelihood Estimators
MML	Modefied Maximum Likelihood
MOEL	Marshall-Olkin Extended Lindley

MOM	Method of Moments
MOMEs	Methods of Moments Estimators
MSc	Master of Science
MSE	Mean Square Error
Nak	Nakagami distribution
NMAE	National Meteorological Agency of Ethiopia
P3	Pearson type III distribution
PDF	Probability Density Function
phD	Doctor of Philosophy
POT	Peaks Over a Threshold
PWM	Probability Weighted Moment
PWMs	Probability weigted Moments
R^2	Coefficient of determination
RAY	Rayleigh distribution
REC	Reciprocal distribution
RMSE	Root Mean Square Error
S	Standard deviation
SNNPR	South Nations Nationalities and People's Region
T	Return period or recurrence interval
UAE	United Arab Emirates
W^2	Cramer-von Mises goodness of fit statistics
WAK	Wakeby distribution
WEI	Two parameter Weibull distribution
WEI3	Three parameter Weibull distribution
WMO	World Meteorological Organization
$V(ref)$	Fundamental basic reference wind velocity
erf^{-1}	Inverse of error function
erf	Error function
n	Number of record year
Γ	Gamma function
$\Gamma_{(.)}$	Lower incomplete gamma function
$\Phi(.)$	Normal distribution of N(0;1)
$\Phi^{-1}(.)$	Inverse of $\Phi(.)$
σ	Scale parameter
β, β_1, β_2	Shape parameters
μ	Location parameter
ξ	Location parameter
ζ	JohnsonSB distribution parameter
π	Mathematical constant 3.14159…….
x_T	Value of *x* for return period T
γ_3	Coefficient of Skewness

CHAPTER 1 INTRODUCTION

1.1 General background

The construction industry in Ethiopia, especially in the capital Addis Ababa has been on fast growing mode which plays an important role in the country development. Currently, large and complex building projects have been constructed due to the demands of the public and private sectors for different purposes. An important part in the design and analysis of these structures is the knowledge of statistical analyses of the wind speed that, directly or indirectly, may cause the structures to fail [1].

Wind is a natural element of weather; and normal wind is beneficial for life. High wind speeds which is too much and cause a threat to the integrity of civil engineering structures is categorized as extreme wind speed. An accurate estimation of the occurrence of extreme wind speeds and the reliable forecasting can help mitigate extreme wind hazards due to human error and minimize their related losses; wrong choice may severely underestimate the design wind speed, thus leading to infrastructure failures; and on the other hand, overestimation can also be a possibility has negative consequences in terms of infrastructure cost [2]. Therefore, the need for analyzing extremes is often to find an optimum balance between adopting high safety and preventing major damage to structures from extreme events that are likely to occur during the design life time of the structure.

Codes and standards on wind action usually assign a reference value of the basic wind speed which represents the characteristic 10 minutes mean wind velocity, irrespective of wind direction and time of year, at 10 m above ground level, in open country terrain with low vegetation and isolated obstacles, having a yearly exceedance probability of 0.02, which is equivalent to a mean return period of 50 years. Typically, buildings are designed to resist a strong wind with a very long return period, such as 50 years or more. The design wind speed is determined from historical records using extreme value theory to predict future extreme wind speeds [3].

According to the new Ethiopian Standards ES EN, 1991:2015 [4], the extreme wind speed $V(ref)$ is a basic parameter for wind speed classification and therefore strongly related to design of buildings and civil engineering works. The parameter $V(ref)$ termed as basic reference design value is defined as the extreme 10-min average wind speed with a return period of 50 years. In general, $V(ref)$ has to be determined statistically on the basis of on-site measurements.

Designers typically rely upon either codified values or results of wind climate analyses for the wind speeds to be used in design; both of which depend upon accurate recordings of long-term wind records from meteorological sites. Therefore, it is desirable to have a mathematical probabilistic distribution model that makes it possible to describe or predict the observed value of some characteristic of interest [5]. The choice of a model follows a procedure with a number of steps, starting with the critical analysis of historical data, the selection of extreme value analysis approaches to examine the magnitude of a wind, and finally, evaluating the adequacy between a wind speed sample and a distribution type to inform the selection. Corresponding to the expected distribution, the estimation of parameters is done based on wind speed data to be modeled. A better fit model can be used to forecast wind speed several periods ahead.

According to the Extreme Value Theory (EVT), there are two commonly used approaches for estimating extreme values: Block Maxima (BM) and Peaks Over a Threshold (POT) approaches. In the BM approach, the data are divided in to blocks of size n (with n reasonably large, usually 1 year), and extracting only the maximum observation of each block regardless of whether the second largest event in a block exceeds the largest events of other blocks; and selected probability distribution model (usually the generalized extreme value (GEV)) is applied. The main drawback of this method is the limited number of values (maxima), especially in short time series. In the second approach, called Peaks Over Threshold (POT), all peak values that exceed a certain lower bound level is selected and then commonly the Generalized Pareto Distribution (GPD) is applied to describe the set of values which exceed this threshold. Major difficulties in using the POT method are assuring the independence of the data series and choosing an appropriate threshold value [6-8].

In the literatures, many probability distribution models were applied for determination of extreme wind speed frequency, but none of the distribution models is accepted as a universal distribution model to describe the wind speed frequency at any site. In this thesis, seventeen commonly used probability distribution models are investigated to obtain more accurate estimates, namely: Gumbel (GUM), 2 and 3 parameter Weibull, (WEI & WEI3), Generalized Extreme Value (GEV), Generalized Pareto Distribution (GPD), Normal (N), 2 and 3 parameter Log-Normal (LN & LN3), 2 and 3 parameter Gamma (GAM & GAM3), 2 parameter Rayleigh (RAY), Reciprocal (REC), JohnsonSB (JSB), Kumaraswamy (KUM), Generalized Lambda Distribution (GLD), Kappa (KAP), and Wakeby (WAK). The extreme wind speed estimates are based on Block (yearly) Maximum and Peaks Over Threshold series, which are extracted from observed daily wind speed data, followed by development of necessary equations and software programs to estimate the parameters of the models. Three parameter estimation methods - Methods of Moments (MOM), L-Moments Estimation (L-ME), and Maximum Likelihood Estimation (MLE) are used.

This work is intended to provide a starting point for future wind map development throughout the Addis Ababa with future data reducing estimate uncertainty and values presented in this thesis could be useful for general understanding of the wind climate in Addis Ababa, and possible calibration of a mathematical models for other towns/cities in Ethiopia.

1.2 Statement of the problem

In the previous, Fikadie Alamirew [9] has analyzed reference basic wind speed for Addis Ababa using Addis Ababa observatory stations data of 31 years (1985-2015). Only Methods of Moments (MOM) were used for parameter estimation. However, instead of using the same method of estimation in different probability distribution functions, using the best method that matches the condition provide calculated wind speed to be more accurate [10]. Furthermore, in the data some of the wind speed records are missing (approximately 29.03% missing). Additionally, the goodness of fit test of the fitted distribution are not assessed in the study.

The present thesis enriched this previous study with new data set Addis Ababa-Bole, and with new procedures up-to-date with the state of the art. The new dataset provides a much more suitable basis for the assessment of the use of such data for extreme value analysis and to develop appropriate techniques for the estimation of extreme wind speed over the Addis Ababa.

1.3 Research questions

- Are wind speeds at Addis Ababa better modeled by the Weibull distribution than by other different probability distribution models? and, if not, which probability distribution model should be fitted to the estimation of extreme design wind speed for Addis Ababa?

- How to test the fitting performance of probability distribution models, and how is it to be decided whether or not the fit is acceptable?

- What is the value of basic reference wind speed; $V(ref)$ to be used for design of buildings and civil engineering work in Addis Ababa?

1.4 Objectives of the study

1.4.1 Main objective

The main objective of this research is to develop more reliable probabilistic extreme wind speed estimation model for use in design of buildings and structures in Addis Ababa using observational data obtained from Addis Ababa-Bole wind speed recording station.

1.4.2 Specific objectives

Particularly, the study has the following sub-objectives.

- Assess and fit various probability distribution models in order to investigate their efficiency and capability in modeling wind speed.
- Fit parameters of the chosen distributions, and check that the resulting functions does, indeed, provide a reasonable fit using graphical and numerical goodness-of-fit tests.
- Compare the BM and the POT series methods and obtain the required design values for given return periods of exceedance.

1.5 Scope of the study and limitations

The scope of the research covers extreme value analysis of wind speed data obtained from only Addis Ababa-Bole station, wind speed record from Addis Ababa observatory stations is not considered. In this study, the investigation is based only on the statistics of the wind speed and do not include structural response to wind loads.

1.6 Ethical considerations

The data are taken as already undergone quality checks at the National Meteorological Agency of Ethiopia (NMAE), and there was no reason to expect that gross errors would be present.

1.7 Organization of the study

This thesis is structured as follows:

Chapter 1 presents the general background of the study, statement of the problem, research questions, general and specific objectives of the study, scope and limitation, and ethical consideration of the research.

Review on the wind speed distribution models, statistical analysis approaches used in wind speed frequency analysis, such as, Block Maximum and Peaks Over Threshold; their comparison, advantages and limitations, common procedure of parameter fitting of probability distribution models with goodness-of-fit assessment from various studies and recommendations are presented in Chapter 2.

Description of the study site and data, theoretical description of commonly used extreme wind speed estimation probability distribution models, commonly used approaches of wind speed probability distributions parameter fitting, best-fit selection and method used are provided in Chapter 3.

Chapter 4 of the thesis is devoted to the results of the study; here, a comparison of the performance of selected analytic distributions and an illustration of the fitting results of the best distribution models suited for the estimation of extreme wind speed events for Addis Ababa is presented in detail.

Conclusions drawn from the study along with recommendations for future studies are finally presented in chapter 5.

CHAPTER 2 LITERATURE REVIEW

2.1 Chapter introduction

Selecting proper distribution functions with better goodness-of-fit is of great significance for extreme wind speed estimation. Several published works around the globe are focused on the applications of extreme values in wind speed and most are focused on the identification of the best fitting probability distribution model for the data [11]. Section 2.2 discusses overview of wind speed probability distribution models, several literatures which are related to the topic are extensively reviewed. Section 2.3 presents approach to estimation of extreme wind speed, BM, limitation of BM, POT, and, limitation of POT. Section 2.4 presents threshold selection techniques, comparison of BM and POT approaches, overview of probability distribution models parameter fitting, Goodness-of-fit (GoF) tests are presented in sections, 2.4 to 2.5, respectively. Finally, over view of Goodness-of-Fit (GoF) tests are presented in section 2.6.

2.2 Overview of wind speed distribution models

Weibull distributions have received particular considerations in many research publications of wind speed analysis, because it was found to fit a wide collection of wind data. For instance, Natei Ermias Benti et al. [12] investigated the wind power potential of Ambo area, a city located in West Shewa Zone of Oromia Region, West of Addis Ababa, Ethiopia using Weibull distribution. Wais [13] employed the Weibull distribution function to describe the available wind energy for three different turbine locations in Poland. Baseer, Meyera [14] utilized the Weibull distribution to determine the wind characteristics of seven locations in Jubail, Saudi Arabia. Keyhani, Ghasemi-Varnamkhasti [15] assessed the wind energy potential in Tehran, capital of Iran with a meteorological model and used the Weibull distribution to study wind speed characteristics. Despite common usage for modeling wind speed, Weibull distribution may not be always appropriate for representing the all wind regimes [16]. Some authors have noted that, in certain cases, the Weibull distribution failed to represent the wind speed data while another distribution may better fit the data.

For instance, Soukissian [17] utilized JohnsonSB, Kappa and Wakeby distributions for describing the empirical distribution of offshore wind speed, and found that they had better performance than the Weibull distribution. Brano, Orioli [18] applied seven probability density functions (Weibull, Rayleigh, Lognormal, Gamma, Inverse Gaussian, Pearson type V and Burr) for studying the wind speed characteristics recorded in the urban area of Palermo, in the south of Italy, at four weather stations. They found that the Burr probability density function was the most reliable statistical distribution.

Akgül and Şenoğlu [19], compared the fitting performances of ten wind speed distributions, namely, Rayleigh, Inverse Weibull (IW), Burr Type III, Extreme Value (EV), Gamma, Inverse Gamma (IG), Marshall–Olkin extended Lindley (MOEL), GEV, and Exponentiated Weibull (EW) distributions to determine the distribution which provides the best fit for different wind regimes located on Turkey's Aegean coast. They found out that, the Weibull distribution remains incapable of modeling all wind regimes. Alavi, Mohammadi [20] examined eight different probability functions; Exponential (EXP), Weibull (WEI), Gamma (GAM), Lognormal (LN), Inverse Gaussian (IG), Loglogistic (LL), Nakagami (NAK), and Generalized extreme value (GEV) for estimating wind speed distribution at five stations, distributed in the east and southeast of Iran. Based on the analysis, the Nakagami distribution qualified as best models to two stations and ranks third to fifth in the remaining stations.

T.B.M.J. Ouarda, C. Charron [21] examined thirteen different models (parametric models, mixture models and one non-parametric model using the kernel density concept) that are commonly employed for extreme wind speed frequency analysis. These models were applied to nine datasets recorded from different stations in the United Arab Emirates (UAE). Results of this study showed that: among the one-component parametric distributions, the Kappa (KAP) and Generalized Gamma Distribution (GLD) provides generally the best fit to the wind speed data for all stations, and the Weibull was identified as the best 2-parameter distribution and performs better than some 3-parameter distributions such as the Generalized Extreme Value (GEV) and Three- parameter Lognormal (LN3).

2.3 Approaches to estimation of extreme wind speed

Caires [22] treated two approaches of extreme wind analysis; the first approach was concerned with estimating the best model for the annual maximum of daily wind speed, while the second was concerned with estimating the model of the annual peak wind speed. These two approaches are referred to as the block maxima (BM) and the peaks over threshold (POT).

A brief discussion of each of these are as follows:

2.3.1 Block Maxima (BM) approach

The block maxima approach in extreme value theory (EVT), consists of dividing the observation period into non over lapping periods of equal size (usually to 1 year), and defined by the maximum peak wind speed of each block [1]. Many researchers often use the block maxima for the statistical analysis of extreme winds. For instance, Dombry [23] used the block maxima approach to determine wind speed. G. Anastasiades and P. E. McSharry [24] applied the BM method with block size of 1 year for the estimation of 50-year design values of wind speed.

2.3.2 Limitations of BM approach

BM sample is defined by the maximum peak wind speed of each block. However, the approach of extracting a single maximum value of wind speed from each block of historical data obviously has limitations in that there may be many relatively high wind speed during any block, and only one value from all these extremes is being considered, i.e., blocks that have several high wind speeds would still only return a single value for that block, similarly a block where all wind speeds are relatively small would still return a single maximum value. Another limitation of the BM series is that the sample is usually rather small [25].

2.3.3 Peaks-Over-Threshold (POT) approach

An alternative approach to the BM, which makes use only of the data of above a certain truncation level to extreme wind prediction is the POT approach [7]. POT approach uses larger data samples than the BM approach, since wind speeds other than the largest BM speeds can also be included in the data sample.

This generally leads to a larger number of extreme wind speeds that exceed an optimal threshold being available for analysis, and this in turn to more precise estimates for return levels and return periods. Additionally, ensuring large sample minimize sampling errors (see [6]. Unlike BM approach, the number of extreme values will not be fixed; we could have blocks with no extreme values and we could have other blocks with multiple extreme values. Simiu and Heckert [5] have applied this method to estimate the extreme wind speed in the United States to provide reference for design winds.

2.3.4 Limitations of POT approach

Major limitations of the POT approaches are assuring the independence of the data series and choosing an appropriate threshold value. In contrast, the BM approach, based on the selection of the largest wind speed for each block (yearly) of the record, naturally leads to wind speed events that are generally identically distributed [22]. The two approaches are illustrated in the following example; adopted from Simiu and Yeo [6].

Assume that in Year 1 the largest speed is 36 m/s and the second largest speed is 34 m/s, and that in Year 2 the largest speed is 43 m/s and the second and third largest speeds are 35 m/s and 31 m/s, respectively. If a threshold of 32 m/s is chosen, the speeds during Years 1 and 2 included in the sample are 43 m/s, 36 m/s, 35 m/s, and 34 m/s (four speeds). In the BM (for blocks of 1 year) approach only two speeds are included in the sample: 36 m/s (Year 1) and 43 m/s (Year 2). If the threshold is very high, the advantage of a larger sample size is lost. For example, if the threshold were 40 m/s, only one speed: 43 m/s would be included in the two-year sample. If the threshold were very low, the sample would include non-extreme wind speeds; this would result in incorrect (biased) estimates of the extreme wind speeds.

A general review on extreme wind speed analysis approaches can be found in [26].

2.3.5 Independent and identically distributed data (i.i.d.)

Records are said to be independent when observing the current observation doesn't have previous memory or provide insight about the value of the next observation, or any other observations in the sample i.e., there is no connection between the observations.

Identically distributed implies one probability distribution should adequately model all values observed in a sample. Consequently, a dataset should not contain trends because they indicate that one probability distribution does not describe all the data.

2.3.6 Independence criteria

One of the main steps of the POT approaches is consideration of the independence criteria. A wide range of values of the time interval to guarantee the independency between consecutive peaks can be found in the literature.

For instance, Simiu and Heckert [27] considered the time intervals between the successive peak wind speed to be at least 3 days to secure the required independence. Similarly, Caires [22] recommended a time interval of 2-4 days between successive peaks to maintain the condition of independency. Guedes Soares C. and M. G. Scotto [28] used a time interval of 20 days in order to guarantee serial correlation. Viselli, Forristall [29] compared a time window of 4, 8, 12, and 16 days for POT analysis, and generalized that a larger time window is preferred between successive peaks to maintain the independence criteria.

2.3.7 Threshold selection techniques

The choice of the threshold represents a balance between bias and variance: choosing too low a threshold results in lower variance, but the bias increases, while too high a threshold will generate small samples that are not enough to estimate the model, leading to high variance and lower bias [22]. Various approaches can be found in the literature to determine the threshold. [25]. Alaswed and Mohamed [30] applied a widely used graphical threshold selection techniques, Mean Residual Life (MRL) plots and Threshold Choice (TC) plots to select the best threshold in fitting the Generalized Pareto Distribution (GPD). Ragan and Manuel [31] and Toft, Naess [32] write that thresholds greater than mean of the data plus 1.4 or 1.9 times the standard deviation or higher can be selected with the condition that there be enough data points to avoid sampling error. Another alternative approach is to fix the average number of times the threshold is exceeded over a given block [25]. Cunnane [33] recommend choosing on average at least 1.65 peaks per year to achieve an advantage over the BM (yearly) series approaches.

2.4 Comparison of BM and POT approaches

The relative merits of POT and BM have been discussed in several papers. Madsen, Rasmussen [34] compared the two approaches and concluded that with respect to accuracy of T-year return value estimation, the POT is in general more efficient for positive shape parameter, whereas the BM (annual) model is preferable for negative shape parameter (GPD expression according to here), again with the number of exceedances larger than the number of blocks.

Caires [1], applied the two methods and mentioned that with POT samples having an average of two or more observations per block, the estimates are more accurate than the corresponding BM estimates, and with more than 200 years of data the accuracies of the two approaches are similar, based on several estimators including the PWM and ML estimators. Bucher and Zhou [35] mentioned that for independent and identically distributed (i.i.d.) samples there is a general consensus among researchers in extreme value statistics that the POT method produces more efficient estimators than the BM method. The main reason is that all large observations are used for the calculation of POT estimators, while BM estimators may miss some large observations falling into the same block.

From all these studies, the following two conclusions can be reached: First, POT is more efficient than BM in many situations for a number of exceedances larger than the number of blocks. Secondly, POT and BM often have comparable performances for large sample sizes. Generally, the optimal number of exceedances should be higher than the optimal number of blocks [36].

2.5 Overview of parameter estimation methods

There are several numerical methods available for the estimation of the parameters of extreme wind speed distribution models, and new approaches are continuously introduced. However, “there is not a single and definite answer in the question which is the ‘‘best’’ estimation method in the case of Extreme Value Analysis (EVA)” [37].

Some of them, for instance, the method of Moments (MOM) and the method of probability weighted Moments (PWM), give explicit expressions for the parameter estimates, but others, such as the maximum likelihood (ML) method, require numerically solving nonlinear systems of equations in order to determine the parameter estimates. Fikadie Alamirew [9] applied method of Moments with Log- Normal (LN) distribution on data of Addis Ababa observatory stations. Natei Ermias Benti et al. [12], analyzed extreme wind speed data by applying Weibull (WEI) distribution with five different estimation methods, such as: Modified Maximum Likelihood (MML), Methods of Moments (MOM).

Soukissian and Tsalis [37], analyzed extreme wind speed data by applying GEV distribution with nine different estimation methods such as MOM, MLE, L-ME, ordinary entropy and quantile least squares methods. T.B.M.J. Ouarda, C. Charron [21] estimated the parameters of thirteen distributions in which the estimation of parameters was done by MOM, Generalized Method of Moment (GMM), MLE, L-ME, and Least squares (LS). Arslan, Bulut [10] made a comparison between the estimation of Weibull (WEI) distribution by MOM, PWM, and MLE methods, and write that the L-ME method yields better results with small sample sizes, and as the sample size increases the ML method proves to be more efficient.

Moreover, Morgan, Lackner [38], dealt with MLE method as a good estimator with various distributions, because they usually yield lower mean square errors (MSE) associated with model parameter estimates than MOM estimators for the large samples. Overall, MLE is considered a common and widely used estimator, even though it is quite hardly applied compared to L-ME and MOM. Compared to conventional Moments, L-Moments are more robust to the existence of outliers in the data [39].

2.6 Goodness-of-Fit (GoF) tests

The GoF describes consistency of a data with a particular distribution model. A variety of criteria and techniques could be applied to evaluate the suitability of a probability distribution for describing a set of data.

Statistical GoF tests as well as graphical display such as probability plots are effective way to determine whether the fitted distributions are consistent with the given set of observations. Each method has advantages and disadvantages, so the appropriate method is determined based primarily on the type of evaluation desired as well as the nature of the problem. For instance, the Kolmogorov-Smirnov (K-S) test gives more weight to the body of probability distributions and relatively flat at the tails.

On the other hand, Cramer- VonMises (W^2) and the Anderson Darling (A-D) tests can be a more accurate tests when the tails of the selected theoretical distribution are the focus of the analysis, as with extreme wind speed analysis (see [11]). Additionally, a large number of studies have been published in scientific literature that use the correlation coefficient (R^2) and Root Mean Square Error (RMSE) to check the performance of probability distributions [19]. D.K. Kidmo, R. Danwe [40] applied the Chi-Square test (χ^2), Correlation Coefficient (R^2), Root Mean Square Error (RMSE) and Kolmogorov-Smirnov test (K-S) to analyze the wind speed data for the District of Garoua, Cameroon using the two-parameter Weibull distribution model to select the most accurate and efficient methods that are adequate for the specific wind data, collected in the district of Garoua.

CHAPTER 3 RESEARCH METHODOLOGY

3.1 Introduction

In this chapter, distribution models which are commonly used in extreme wind speed analysis are presented, the determination of the parameter of distributions was carried out using the method of Maximum Likelihood Estimation (MLE), L-moments Estimation (L-ME), and Methods of Moments (MOM). Selection of the best-fit distribution was accomplished using, the CDF P–P plots, and classical goodness-of-fit statistics: Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Kolmogorov–Smirnov (K-S), and, Cramer-VonMises (C-VM). The whole procedure used in this study is as shown in Figure 1.

For calculating the parameters by using MOM; the sample mean $\overline{x}_i$, standard deviation S and coefficient of skewness, γ_3 are determined by the following equations:

$$\overline{x}_i = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (3\text{-}1)$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x}_i)^2} \qquad (3\text{-}2)$$

$$\gamma_3 = \frac{n\sum_{i=1}^{n}(x_i - \bar{x}_i)^3}{(n-1)(n-2)S^3} \qquad (3\text{-}3)$$

Where, n is the number of observations in the data set, i is the observation number and x_i is the observation data.

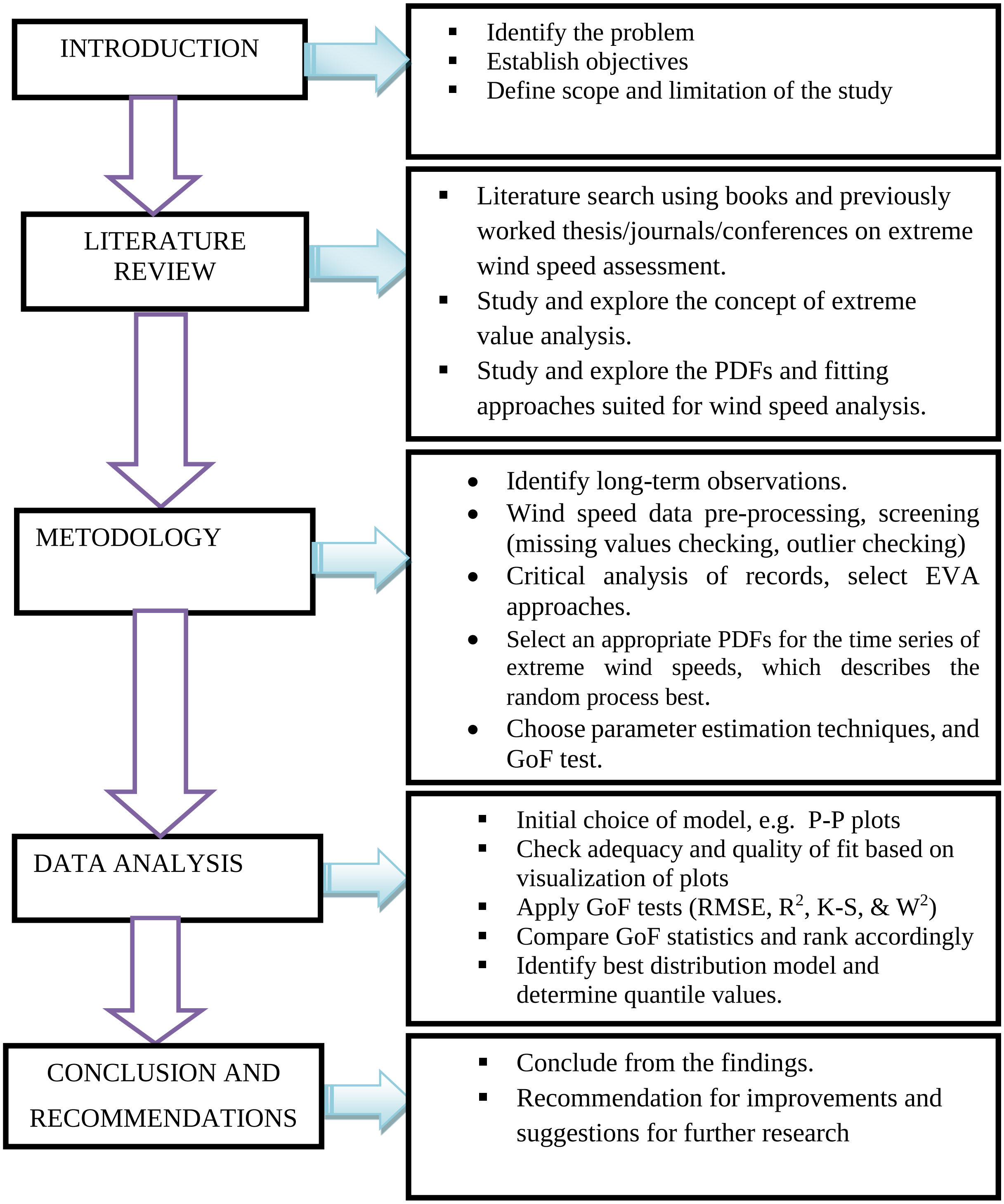


Figure 3-1. Methodology flow chart of the research approach.

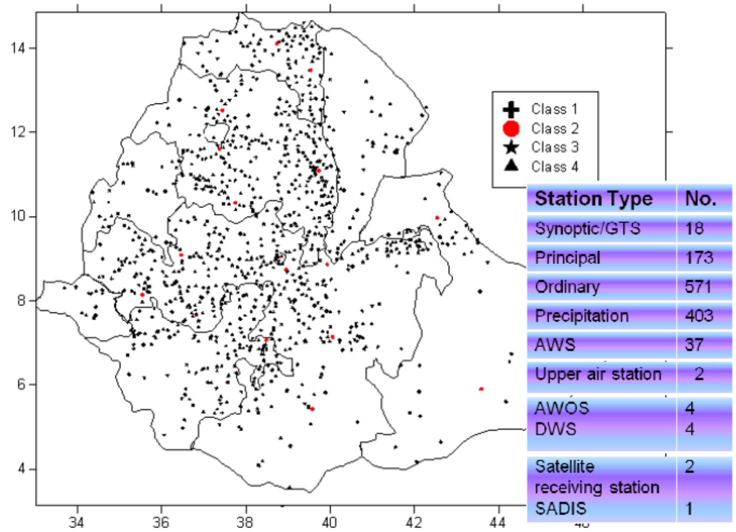


Figure 3-2. Stations distribution over Ethiopia as updated on December 2012 [41].

3.2 Site selection

National Meteorological Agency of Ethiopia (NMAE) is responsible for conducting a measurement, control, and storage of meteorological data of the areas all over the country. There are many wind speed recording stations throughout Ethiopia, see Figure 3-2 [41]. In Addis Ababa there are two currently functioning wind speed recording stations [41]: Addis Ababa; Bole and Addis Ababa Observatory stations, which contain records of wind speeds and directions. Bole station is selected for analysis in this study. The station was selected for its long records and completeness. Furthermore, the station represents somewhat the standard wind climate for wind recording. The wind records at Addis Ababa observatory stations, are highly affected by missing data (approximately 27.27% missing). Furthermore, the periods of records are shorter than Bole station. For these reasons, such wind speed datasets are rejected.

3.3 Description of the study site

The station used in this thesis is located in the eastern part of Addis Ababa at latitude 38.75°N, longitude 9.0333°E, and 2354 meters of elevation from sea level. The analyzed data in this work were obtained from NMAE containing daily observations of wind speed and wind direction taken at intervals of 3hrs. for every day of the month throughout the year that cover a time period from 1954-2016 (see Table 3-1). Units of measurement are m/s. The station has the longest and most reliable wind speed observation records in the city. Thus, it is the obvious reference for the planning, engineering design, and management of the proposed civil engineering structures in the city.

Table 3-1. Wind dataset acquired from NMAE for Addis Ababa-Bole station.

Year	Number of Records per day
1954-1963	3
1964-1968,1971,1973-1974, 1976-1978	5
1969-1970,1972, 1975	3
1979	5 (1-4th month), and 3 (5-12th month)
1980-1984	3
1985	5 (1-10th month), and 3 (11-12th month)
1986	3 (1-4th month), and 5 (10-12th month)
1987-2001	7
2002	8
2003-2012	5
2015	5 (1-3rd month), and 4 (12th month)
2016	5

3.4 Data quality assessment of the station

It is important to point out that the dataset was processed as it was received from NMAE. Therefore, quality analysis of the dataset was not attempted.

3.5 Actual wind speed observations and check for missing records

The data, as mentioned before is available from the NMAE. These extensive time series are 63 years long and ranges from 01-Sep-1954 to 31-Aug-2016. The time series has a total of 2 years of missing values (2013 and 2014). As the percentage of yearly missing values is small (< 4%), the effect of the missing values is practically negligible. Moreover, there are no changes in the measurement location during the sampling period (see [42]). However, there are missing data points for a number of months (see Table 3-2). Additionally, there are also some missing data for a number of days. The total number of measurements is 98,560 out of a maximum possible number of 106,430 (approximately 7.4% missing data). In the data set analyzed here, records may contain up to 66.7% of monthly missing values (year 2015).

Table 3-2. Missing data points for a number of months in the data set.

Year	1957	1963	1986	1989	1998	1999	2005	2011	2012	2015	2016
Months with missing data	1^{st}	1^{st}	5^{th} -8^{th}	1^{st}	3^{rd}	8^{th}-11^{th}	5^{th}	4^{th}	12^{th}	4^{th} -11th	1^{st} -4th

3.6 A procedure followed for extracting the maxima

Depending on the number of daily observations (3, 4, 5, 7 or 8) are used to derive the daily maximum wind speed. Again, from the daily maximum observations yearly maximum are extracted to create new time series that are utilized to conduct Block Maximum (BM) analysis. The block maxima are extracted irrespective of the years' missing values percentage, but; if the extracted maxima are less than annual maximum of those years with full record, then it is deleted from the time series and assumed as missing values. For Peaks Over Threshold (POT) analysis, Two-peaks, Three-peaks, Four-peaks, Six-peaks, and Twelve-peaks per year are selected.

3.7 Check for outliers

An outlier is an observation that lies an unusual distance from other observations. In any time, series data, outliers may or may not exist. In this thesis statistical (using Grubb's test) and graphical (using time series plot) of summarizing the data was done and unusual/extreme values were detected and corrected.

3.7.1 Graphical way of checking outliers using time series plots

Outliers are checked before and after extracting the BM series. No outliers were detected when extracting the BM series without checking each year for outlier. However, considering each year, the maximum wind speed recorded in the station is 40 m/s at day 2, month 10, 1970 at 12:00 p.m. The average daily maximum wind speed on that year was approximately 12.36 m/s without this measurement and 12.38 m/s with this measurement. Similarly, a speed of 40 m/s was recorded as the maximum wind speed in 1994 at 21:00 p.m. on month 6, day 15, followed by a measurement of 12 m/s at 12:00 p.m. The average wind speed on that year without this measurement was approximately 5.98 m/s and 6.07 with this measurement. The occurrence of a large wind speeds in these years is not confirmed by wind speed records at other consecutive years. Plots of the data for these years are given in Figure 3-3). These two wind speeds are outlying and were replaced with the second maximum values of that year and other outliers are retained (see Appendix C for all outliers detected from the data set used for analysis in this work).

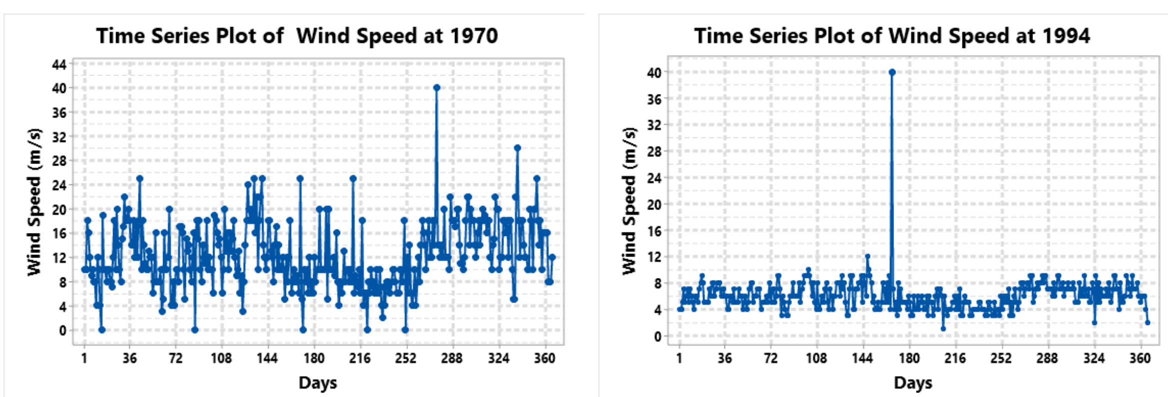


Figure 3-3. Time series plot of daily maximum wind speed at 1970 (left) and 1994 (right)

3.7.2 Statistical way of checking outliers using Grubb's Test

The general formula for two - sided Grubb's test statistic is defined as:

$$G = \frac{\max|x_i - \bar{x}_i|}{S} \quad (3\text{-}4)$$

Where x_i is the element of the dataset, $\bar{x}_i$ and S denoting the mean and standard deviation of the sample data, respectively.

The hypothesis of no outliers is rejected at significance level α if:

$$G > \frac{n-1}{\sqrt{n}}\sqrt{\frac{{t_{n-2,p}}^2}{n-2+{t_{n-2,p}}^2}} \quad (3\text{-}5)$$

Where, $t_{n-2,p}$ is the p-quantile of the t-distribution with $v = n - 2$ degrees of freedom, $p = 1 - \frac{\alpha}{2n}$, α is significance level. $\alpha = 0.05$ for 5% significance level (95% confidence interval).

After correcting the data, using Eq. 3-4, the Grubb's formula gives, $G = 2.164$. Excel "$T.INV$" function is used to calculate p-quantile of t-distribution with n-2 degrees of freedom, then (Eq. 3-5) is used to determine the critical value; it is found to be 3.206. Since, $3.206 > 2.164$ it implies no outlier. Additionally, from the timeseries plot (Figure 3-4), it can be seen that no outlier is detected (at the 95% confidence level).

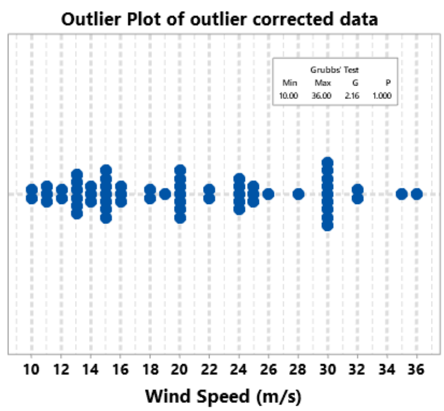


Figure 3-4. Outlier plot of yearly maximum outlier corrected data.

3.8 Extreme value analysis methods

Two methods were considered here to provide alternative ways of analyzing the data and are compared in this study. The first method fits a chosen candidate distribution models to the data based on yearly maximum and the second method fits the Generalized Pareto Distribution (GPD) to specified peaks over a threshold value, as described in chapter 2.

3.9 General description of some possible distribution models for BM analysis

As mentioned in chapter 1, there is no evidence for selecting a particular probability distribution function for extreme wind speed analysis. As a result, a number of probability distribution functions are used in the literatures. Some of the most common wind speed probability distribution models used in previous studies have been fitted to the wind speed data in this thesis.

Table 3-3. Two-parameter distribution models CDF, PDF, Quantile, and Range.

Model		CDF, PDF, and Quantile function	Range
GUM	CDF	$F(x_i) = exp\left[-exp\left(-\left(\frac{x_i-\mu}{\alpha}\right)\right)\right]$	$-\infty < x_i < \infty$ $\alpha > 0$
	PDF	$f(x_i) = \frac{1}{\alpha} exp\left[-\left(\frac{x_i-\mu}{\alpha}\right) - exp\left(-\frac{x_i-\mu}{\alpha}\right)\right]$	
	Quantile	$x(F) = \mu - \alpha\, ln(-\, ln\, F)$	
WEI	CDF	$F(x) = 1 - exp\left[-\left(\frac{x_i}{\alpha}\right)^{\beta}\right]$	$x_i > 0$ $\alpha > 0$ $\beta > 0$
	PDF	$f(x) = \frac{\beta}{\alpha}\left(\frac{x_i}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x_i}{\alpha}\right)^{\beta}\right]$	
	Quantile	$x(F) = \alpha(-\, ln(1-F))^{1/\beta}$	
N	CDF	$F(x_i) = \frac{1}{2}\left[1 + erf\left(\frac{x_i-\mu}{\sigma\sqrt{2}}\right)\right]$	$-\infty < x_i, \mu < \infty$ $\alpha > 0$
	PDF	$f(x_i) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x_i-\mu)^2\right]$	
	Quantile	$x(F) = \mu + \sigma\sqrt{2}\left(erf^{-1}(2F-1)\right)$	
LN	CDF	$F(x_i) = \frac{1}{2} + \frac{1}{2} erf\left(\frac{ln\, x_i-\mu}{\sigma\sqrt{2}}\right)$	$\sigma > 0$
	PDF	$f(x_i) = \frac{1}{x_i\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(ln(x_i)-\mu)^2\right]$	
	Quantile	$x(F) = exp\left[\mu + \sigma\sqrt{2}\left(erf^{-1}(2F-1)\right)\right]$	
GAM	CDF	$F(x_i) = \frac{\Gamma_{\left(\beta, \frac{x_i}{\alpha}\right)}}{\Gamma(\beta)}$	$\alpha, \beta, x_i > 0$
	PDF	$f(x_i) = \frac{1}{\alpha^{\beta}\Gamma(\beta)} {x_i}^{\beta-1} exp\left(-\frac{x_i}{\alpha}\right)$	
	Quantile	Using excel, $x(F) = GAMMA.INV(F, \beta, \alpha)$	
RAY	CDF	$F(x_i) = 1 - exp\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\alpha}\right)^2\right)$	$x_i > \mu$ $\alpha > 0$
	PDF	$f(x_i) = \frac{x_i-\mu}{\alpha^2} exp\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\alpha}\right)^2\right)$	
	Quantile	$x(F) = \mu + \alpha\sqrt{-2\, ln(1-F)}$	
REC	CDF	$F(x_i) = \frac{ln(x_i)-ln(a)}{ln(b)-ln(a)}$	$0 < a < b$ $a \le x_i \le b$
	PDF	$f(x_i) = \frac{1}{x_i[ln(b)-ln(a)]}$	
	Quantile	$x(F) = exp[ln(a) + F(ln(b) - ln(a))]$	

Table 3-4. Three-parameter distribution models CDF, PDF, Quantile, and Range.

Model	CDF, PDF, and Quantile function		Range
WEI3	CDF	$F(x) = 1 - exp\left[-\left(\frac{x_i-\mu}{\alpha}\right)^{\beta}\right]$	$x_i > \mu$ $\alpha > 0$ $\beta > 0$
	PDF	$f(x) = \frac{\beta}{\alpha}\left(\frac{x_i-\mu}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x_i-\mu}{\alpha}\right)^{\beta}\right]$	
	Quantile	$x(F) = \mu + \alpha(-\ln(1-F))^{1/\beta}$	
GEV	CDF	$F(x_i) = exp\left[-\left(1 - k\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{k}}\right]$	$-\infty < x_i \le \mu + \alpha/k$ $if\ \ k > 0;$ $-\infty < x_i < \infty$ $if\ k = 0;$ $\mu + \alpha/k \le x_i < \infty$ $if\ k < 0.$
	PDF	$f(x_i) = \frac{1}{\alpha}\left[1 - k\left(\frac{x_i-\mu}{\alpha}\right)\right]^{\frac{1}{k}-1} exp\left[-\left(1 - k\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{k}}\right]$	
	Quantile	$x(F) = \mu + \frac{\alpha}{k}[1 - (-\ln F)^{k}]$	
GPD	CDF	$F(x) = 1 - \left[1 - k\frac{(x_i-\mu)}{\alpha}\right]^{\frac{1}{k}}$	$\mu \le x_i \ \le \mu + \ \alpha/k$ $if\ k \ > 0;$ $\mu \le x_i \ < \infty$ $ifk \ \le 0$
	PDF	$f(x) = \frac{1}{\alpha}\left[1 - k\frac{(x_i-\mu)}{\alpha}\right]^{\frac{1}{k}-1}$	
	Quantile	$x(F) = \mu + \frac{\alpha}{k}[1 - (1-F)^{k}]$	
LN3	CDF	$F(x_i) = \Phi\left(\frac{\ln(x_i-\xi)-\mu}{\sigma}\right)$ Excel: $NORM.S.DIST\left(\frac{ln(x_i-\xi)-\mu}{\sigma}, 1\right)$ is used	$x_i > \xi$ $\alpha > 0$ $\beta > 0$
	PDF	$f(x_i) = \frac{1}{(x_i-\xi)\sigma\sqrt{2\pi}} exp\left[-\frac{(ln(x_i-\xi)-\mu)^2}{2\sigma^2}\right]$	
	Quantile	$x(F) = \xi + exp\left[\mu + \sigma\left(\Phi^{-1}(F)\right)\right]$	
GAM3	CDF	$F(x_i) = \frac{\Gamma\left(\beta, \frac{(x_i-\mu)}{\alpha}\right)}{\Gamma(\beta)}$ "$zipfR$" R package is used $library(zipfR) \rightarrow \frac{Igamma(\beta,z)}{\Gamma(\beta)};\ z = \left(\beta, \left(\frac{x_i-\mu}{\alpha}\right)\right)$	$\mu \le x_i < \infty$ $\alpha > 0$ $\beta > 0$
	PDF	$f(x_i) = \frac{(x_i-\mu)^{\beta-1}}{\alpha^{\beta}\Gamma(\beta)} exp\left(-\frac{(x_i-\mu)}{\alpha}\right)$	
	Quantile	Explicit analytical form is not available	

Table 3-5. Four and Five-parameter distribution models CDF, PDF, Quantile, and Range.

Model	CDF, PDF, and Quantile function		Range
JSB	CDF	$F(x_i) = \Phi\left(\gamma + \delta\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right)\right)$	$\zeta < x_i < \zeta + \eta$, $-\infty < \zeta < \infty$, $-\infty < \gamma < \infty$, $\delta, \eta > 0$
	PDF	$f(x_i) = \frac{\delta\eta}{\sqrt{2\pi}(x_i-\zeta)(\zeta+\eta-x_i)} \cdot exp\left[-\frac{1}{2}\left(\gamma+\delta\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right)\right)^2\right]$	
	Quantile	$x(F) = \zeta + \eta\left(1+exp\left(\frac{\gamma-\Phi^{-1}(F)}{\delta}\right)\right)^{-1}$	
KUM	CDF	$F(x_i) = 1 - \left[1-\left(\frac{x_i-a}{b-a}\right)^{\beta_1}\right]^{\beta_2}$	$a < b$ $a \le x_i \le b$
	PDF	$f(x_i) = \frac{\alpha\beta\left(\frac{x_i-a}{b-a}\right)^{\beta_1-1}\left[1-\left(\frac{x_i-a}{b-a}\right)^{\beta_1}\right]^{\beta_2-1}}{b-a}$	
	Quantile	$x(F) = a + (b-a)\left[1-(1-F)^{1/\beta}\right]^{1/\alpha}$	
GLD	CDF	Explicit analytical form is not available Numerically solved using "gld" R package	$0 \le F \le 1$
	PDF	Explicit analytical form is not available Numerically solved using "gld" R package	
	Quantile	$x(F) = \xi + \alpha[F^{\kappa} - (1-F)^{h}]$	
KAP	CDF	$F(x_i) = \left[1-h\left(1-\kappa\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{\kappa}}\right]^{\frac{1}{h}}$	$\mu + \frac{\alpha}{k}[1-h^{-k}] \le x_i$ $if\ k > 0;$ $\mu + \frac{\alpha}{k} \le x_i$ $if\ h \le 0, k < 0$ $x_i \le \mu + a/k$ $if\ k > 0$
	PDF	$f(x_i) = \frac{1}{\alpha}\left(1-\kappa\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{\kappa}}[F]^{1-h}$	
	Quantile	$x(F) = \mu + \frac{\alpha}{\kappa}\left(1-\left(\frac{1-F^{h}}{h}\right)^{\kappa}\right)$	
WAK	CDF	Explicit analytical form is not available	$\xi \le x_i < \infty$ $if\ \delta \ge 0, and\ \gamma > 0$ $\xi \le x_i \le \xi + \alpha/\beta - \gamma/\delta$ $if \delta < 0\ or\ \gamma = 0$
	PDF	Explicit analytical form is not available	
	Quantile	$x(F) = \xi + \frac{\alpha}{\beta}\left[1-\left(1-F(x_i)\right)^{\beta}\right] - \frac{\gamma}{\delta}\left[1-(1-F(x_i))^{-\delta}\right]$	

Where, CDF is Cumulative Distribution Function, PDF is Probability Distribution Function (derivative of CDF with respect to x_i), μ, ξ are location parameters, α, σ are scale parameters, $\beta, \beta_1, \beta_2, h, k, \kappa$ are shape parameters, a is lower bound of the data, b is upper bound of the data, GUM is Gumbel, WEI is Two-parameter Weibull, WEI3 is Three-parameter Weibull, GEV is Generalized Extreme Value, GPD is Generalized Pareto Distribution, N is Two-parameter Normal, LN3 is Three-parameter Log-Normal, GAM is Two -parameter Gamma

GAM3 is Three-parameter Gamma, RAY is Two-parameter Rayleigh, REC is Reciprocal, JSB is Johnson Bounded, KUM is Kumaraswamy, GLD is Generalized Lambda Distribution KAP is Kappa6, and WAK is Wakeby. $F = 1 - \frac{1}{T}$, T is return period, erf is error function, erf^{-1} is inverse of error function, $\Phi(.)$ is normal distribution of N(0;1), $\Phi^{-1}(.)$ is inverse of $\Phi(.)$, and, $\Gamma_{(.)}$ is lower incomplete gamma. Furthermore, λ, is commonly used to designate L-moments; here, the designation used for GLD by [43] is adopted to avoid confusion.

The error function erf is defined as:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp(-t^2)dt \quad (3\text{-}6)$$

Excel "erf" function is used here.

Using Maclaurin series $erf^{-1}(x)$ can be estimated as:

$$erf^{-1}(x) \approx \frac{\sqrt{\pi}}{2}x + \frac{1}{3}\left(\frac{\sqrt{\pi}}{2}x\right)^3 + \frac{7}{30}\left(\frac{\sqrt{\pi}}{2}x\right)^5 + \frac{37}{630}\left(\frac{\sqrt{\pi}}{2}x\right)^7 + \frac{773}{22680}\left(\frac{\sqrt{\pi}}{2}x\right)^9 \quad (3\text{-}7)$$

3.10 Parameter fitting methods

Many methods can be used for estimating the parameters of distributions presented in Tables (3-5), extensive details of some of the methods are already available in the literature, (e.g. Singh [44], K. H. Hamed and A. R. Rao [45]). In this work, the MOM, L-ME, and, MLE were used.

3.10.1 Methods of Moments Estimators (MOMEs)

The MOMEs are formed by equating the sample moments with theoretical moments of the distribution function, CDF. Let $x_1, x_2, \dots, x_n$ be a sample of observations from a population with the CDF $F(x_i; \theta_1, \theta_2, \dots, \theta_p)$, where θ_1, $\theta_2, \dots, \theta_p$ are unknown parameters to be estimated based on the sample. The MOM estimators can be obtained by solving the following system of equations for $\theta_1, \theta_2, \dots, \theta_p$.

$$\begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i = E(x_1) \\ \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 = E({x_1}^2) \\ \quad\quad . \\ \quad\quad . \\ \frac{1}{n}\sum_{i=1}^{n} {x_i}^p = E({x_1}^p) \end{cases} \tag{3-8}$$

Where, $E({x_1}^p) = \int_{-\infty}^{\infty} x^p\, f\big(x; \theta_1, \theta_2, \dots, \theta_p\big)dx, \;\; k = 1, 2, \dots, p.$

3.10.2 Probability Weighted Moments (PWMs) equations

PWMs are needed for the calculation of L-Moments. The data first must be arranged in ascending order, and then apply Eq. (3-9) and (Table 3-6) [46].

$$\begin{cases} \beta_o = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \beta_1 = \frac{1}{n}\sum_{i=2}^{n} \frac{(i-1)}{(n-1)} x_i \\ \beta_2 = \frac{1}{n}\sum_{i=3}^{n} \frac{(i-1)(i-2)}{(n-1)(n-2)} x_i \\ In\ general,\ \beta_r = \frac{1}{n}\sum_{i=r+1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_i \end{cases} \tag{3-9}$$

Where, n is the sample size, x_i is the wind speed value, and i is the rank of the value in ascending order.

3.10.3 L-Moments Estimators (L-MEs)

Unbiased sample estimates of the PWMs, for any distribution can be computed from Table 3-6:

Table 3-6. L-Moments and L-Moment ratios.

L-moments	L-moment ratios
$\lambda_1 = \beta_0$ $\lambda_2 = 2\beta_1 - \beta_0$ $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$ $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$	$\tau = \frac{\lambda_2}{\lambda_1}(L - CV)$ $\tau_3 = \frac{\lambda_3}{\lambda_2}(L - Skewness)$
In general, $\lambda_{r+1} = \sum_{k=0}^{r} p_{r,k}\beta_k,\;\; p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$	$\tau_4 = \frac{\lambda_4}{\lambda_2}(L - Kurtosis)$

3.10.4 Maximum Likelihood Estimators (MLEs)

As the name suggests, the MLEs tries to find values of the distribution parameters that maximize the likelihood function. For a given samples $x_1, x_2, \ldots, x_n$, the function defined by equation 10 is known as the likelihood function [11].

$$L(\Theta/x_i) = \prod_{i=1}^{n} f(x_i; \Theta) \quad (3\text{-}10)$$

Maximizing $\mathrm{L}(\Theta/\mathrm{x_i})$ is equivalent to maximizing $ln\mathrm{L}(\Theta/\mathrm{x_i})$ because ln is a monotonic increasing function.

The log-likelihood function is given by:

$$lnL(\Theta/x_i) = \prod_{i=1}^{n} ln\, f\,(x_i; \Theta) \quad (3\text{-}11)$$

Then the MLEs are the values of the Θ that maximize the log-likelihood with respect to Θ and are obtained by solving the system of equations using equation 12.

$$\frac{\partial lnL(\Theta/x_i)}{\partial \theta_i} = 0,\; i = 1,2,\ldots n \quad (3\text{-}12)$$

3.10.5 The numerical iterative methods used

It is usual for approximations to the Maximum Likelihood Estimators (MLEs) to be calculated iteratively. In this thesis, the MLEs of the selected distribution models are estimated using numerical iteration method excel "*solver*" and "*goal seek*" functions. R programing software is also extensively used. The R programming software is a free software and can be downloaded at http://www.r-project.org/.

3.11 Fitting techniques used for distribution models

As mentioned earlier, the parameters of distribution models can be found by a number of ways. Tables (3-7 to 3-9) present techniques used for parameter estimation of models selected for analysis in this work.

Table 3-7. Fitting techniques used for Two-parameter distribution models.

Model	Fitting Techniques		
	MLEs	L-MEs	MOMEs
GUM	$\alpha - \overline{x}_i + \frac{\sum_{i=1}^{n} x_i\, exp\left(\frac{x_i}{\alpha}\right)}{\sum_{i=1}^{n} exp\left(\frac{x_i}{\alpha}\right)} = 0$ $\mu = \alpha\ ln\left[\frac{n}{\sum_{i=1}^{n} exp\left(-\frac{x_i}{\alpha}\right)}\right]$ Initial: α=$\frac{\sqrt{6}\, S}{\pi}$	$\alpha = \frac{\lambda_2}{ln(2)}$ $\mu = \lambda_1 - 0.5772\ \alpha$	$\alpha = \frac{\sqrt{6}\, S}{\pi}$ $\mu = \overline{x}_i - 0.5772\ \alpha$
WEI	$\beta - \left[\frac{\sum_{i=1}^{n} x_i{}^{\beta}\, ln\, x_i}{\sum_{i=1}^{n} x_i{}^{\beta}} - \frac{ln\, x_i}{n}\right]^{-1} = 0$ $\alpha = \left(\frac{\sum_{i=1}^{n} x_i{}^{\beta}}{n}\right).$ Initial: $\beta = 2$	$\beta = -\frac{ln(2)}{ln\left(1-\frac{\lambda_2}{\lambda_1}\right)}$ $\alpha = \frac{\lambda_1}{\Gamma(1+1/\beta)}$	$\beta = \left(\frac{0.9874\, \overline{x}_i}{S}\right)^{1.0983}$ $\alpha = \frac{\overline{x}_i}{\Gamma(1+1/\beta)}$
N	$\mu = \overline{x}_i$ $\sigma = S$	$\mu = \lambda_1$ $\sigma = \sqrt{\pi}\lambda_2$	$\mu = \overline{x}_i$ $\sigma = \mathrm{S}$
LN	$\mu = \overline{lnx_i}$ $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(ln(x_i)-\mu)^2}{n}}$	$\mu = ln(\lambda_1) - \frac{\sigma^2}{2}$ $\sigma = 2erf^{-1}\left(\frac{\lambda_2}{\lambda_1}\right)$	$\mu = \overline{lnx_i} - \frac{\sigma^2}{2}$ $\sigma = \sqrt{ln\left[\frac{S^2}{\overline{x}_i{}^2} + 1\right]}$
GAM	$\alpha\beta = \overline{x}_i$ $n\, ln(\alpha) + n\psi(\beta) = \sum_{i=1}^{n} ln(x_i)$	$\beta - \frac{\lambda_1 \Gamma\left(\beta+\frac{1}{2}\right)}{\lambda_2\sqrt{\pi}\Gamma(\beta)} = 0$ $\alpha = \frac{\lambda_1}{\beta}$. Initial: $\beta = \beta_{MOM}$	$\overline{x}_i = \alpha\beta$ $S = \alpha\sqrt{\beta}$
RAY	$\frac{2n\sum_{i=1}^{n}(x_i-\mu)}{\sum_{i=1}^{n}(x_i-\mu)^2} - \sum_{i=1}^{n}(x_i - \mu)^{-1} = 0$ $\alpha = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}(xi - \mu)^2}.$ Initial: $\mu = \mu_{MOM}$	$\alpha = \sqrt{(7.420984\lambda_2{}^2)}$ $\mu = \lambda_1 - 3.414214\lambda_2$	$\alpha = 1.5264\, S$ $\mu = \bar{x}_i - 1.253314\alpha$
REC	$a =$ Minimum $value\ of\ the\ sample$ $b =$ M$aximum\ value\ of\ the\ sample$	$a = exp(\lambda_1 - 3\lambda_2)$ $b = exp(\lambda_1 + 3\lambda_2)$ (after log transform)	$a + b = \frac{2\left(S^2+\overline{x}_i{}^2\right)}{\overline{x}_i}$

Where, x_i is the wind speed, $\overline{x}_i$ is the mean of the data, $\overline{lnx_i}$ is the mean of the log transformed data, S is standard deviation of the data, n is the number of samples, λ_1, λ_2 are L-moments defined in Table 3-6, and $\psi(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$ is digamma function, can be determined by using "*digamma*" R function, MLEs is the maximum likelihood estimators, L-MEs is L-Moments Estimators, and MOMEs is Method of Moments Estimators.

Table 3-8. Fitting techniques used for Three-parameter distribution models.

Model	Fitting Techniques	Estimators
WEI3	MLE	See appendix D. "*weibullR*" R package is used.
	L-ME	$\tau_3 = \frac{1-\frac{3}{2^{\frac{1}{\beta}}}+\frac{2}{3^{\frac{1}{\beta}}}}{1-\frac{1}{2^{\frac{1}{\beta}}}}, \alpha = \lambda_2\left[\Gamma\left(1+\frac{1}{\beta}\right)\left(1-\frac{1}{2^{\frac{1}{\beta}}}\right)\right]^{-1} \quad \mu = \lambda_1 - \alpha\Gamma\left(1+\frac{1}{\beta}\right)$ Initial: $\beta = 2.$ Excel "*goal seek*" function is used to determine β.
	MOM	$\gamma_3 = \frac{\Gamma(1+3/\beta)-3\Gamma(1+2/\beta)\Gamma(1+1/\beta)+2\Gamma^2(1+1/\beta)}{[\Gamma(1+2/\beta)-\Gamma^2(1+1/\beta)]^{3/2}}$; Initial, $\beta = 2$ $\alpha = \sqrt{\frac{S}{\left[\Gamma\left(1+\frac{2}{\beta}\right)-\Gamma^2\left(1+\frac{1}{\beta}\right)\right]}}, \mu = \overline{x}_i - \alpha\Gamma(1+1/\beta)$
GEV	MLE	See appendix D. "*EnvStats*" R package is used.
	L-ME	$k \approx 7.8590c + 2.9554c^2, c = \frac{2}{3+\tau_3} - \frac{ln\,2}{ln\,3}$ $\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)}, \ \mu = \lambda_1 - \frac{\alpha}{k}\left(1-\Gamma(1+k)\right)$
	MOM	$\gamma_3 = \frac{\beta}{\lvert\beta\rvert}\frac{-\Gamma(1+3\beta)+3\Gamma(1+\beta)\Gamma(1+2\beta)-2\Gamma^3(1+\beta)}{[\Gamma(1+2\beta)-\Gamma^2(1+\beta)]^{3/2}}$, Initial, $\beta = 2$ $\alpha = \sqrt{\frac{\beta^2 S^2}{[\Gamma(1+2\beta)-\Gamma^2(1+\beta)]}}, \ \mu = \bar{x}_i + \frac{\alpha}{\beta}[1-\Gamma(1-\beta)]$
GPD	MLE	See appendix D. "*ismev*" R package is used.
	L-ME	$k = \frac{1-3\tau_3}{1+\tau_3}, \quad \alpha = (1+k)(2+k)\lambda_2, \quad \mu = \lambda_1 - (2+k)\lambda_2$
	MOM	$\gamma_3 = \frac{2(1-k)(1+2k)^{1/2}}{(1+3k)}, \ \alpha = S\sqrt{(1+k)^2(1+2k)}, \ \mu = \overline{x}_i - \frac{\alpha}{1+k}$
LN3	MLE	$\mu = \frac{1}{n}\sum_{i=1}^{n} ln(x_i - \xi), \sigma = \frac{1}{n}\sum_{i=1}^{n}[ln(x_i-\xi)-\mu]$ $\sum_{i=1}^{n}(x_i-\xi)^{-1}(\mu-\sigma^2) = \sum_{i=1}^{n}\frac{ln(x_i-\xi)}{(x_i-\xi)}$ "*EnvStats*" R package is used to estimate parameters
	L-ME	Detail derivations are already available in [46]. Here, "*lmomco*" R package is used.
	MOM	$\xi = \overline{x}_i - \frac{S}{z}, \sigma = \sqrt{ln(z^2+1)}, \mu = ln\left(\frac{S}{z}\right) - \frac{1}{2}ln(z^2+1)$ $z = 1 - \left[\frac{-\gamma_3+\sqrt{{\gamma_3}^2+4}}{2}\right]^{2/3}\left(\left[\frac{-\gamma_3+\sqrt{{\gamma_3}^2+4}}{2}\right]^{1/3}\right)^{-1}$
GAM3	MLE	Detail derivations are already given in [45]. Here,"*PearsonDS*" R package is used.
	L-ME	$\beta = \begin{cases} \frac{1+0.2906z}{z+0.1882z^2+0.0442z^3}, & z = 3\pi{\tau_3}^2, 0 < \lvert\tau_3\rvert < \frac{1}{3} \\ \frac{0.36067z-0.59567z^2+0.25361z^3}{1-2.78861z+2.56096z^2-0.77045z^3}, & z = 1-\lvert\tau_3\rvert, \frac{1}{3} \le \lvert\tau_3\rvert < 1 \end{cases}$ $\alpha = \frac{\sqrt{\pi}\Gamma(\beta)\lambda_2}{\Gamma\left(\beta+\frac{1}{2}\right)}, \mu = \lambda_1 - \alpha\beta$
	MOM	$\beta = \left(\frac{2}{\gamma_3}\right)^2, \ \alpha = \frac{S\gamma_3}{2}, \ \mu = \bar{x}_i - \alpha\beta$

Table 3-9. Fitting techniques used for Four and Five-parameter distribution models.

Model	Fitting Technique	Approach
GLD	L-ME	$\tau_3 = \frac{\kappa(\kappa-1)(h+3)(h+2)(h+1) - [h(h-1)(\kappa+3)(\kappa+2)(\kappa+1)]}{(\kappa+3)(h+3)[\kappa(h+2)(h+1) + h(\kappa+2)(\kappa+1)]}$ $\alpha = \frac{\lambda_2}{\left(\frac{\kappa}{(\kappa+2)(\kappa+1)} + \frac{h}{(h+2)(h+1)}\right)}$, $\xi = \lambda_1 - \alpha\left(\frac{1}{\kappa+1} - \frac{1}{h+1}\right)$ Different pairing of κ and h can be a solution for the first equation. See ([46],[43]). Here, "*lmomco*" R package is used.
JSB	MLE	See appendix D. "*extDist*" R package is used.
	MOM	"*EasyFit*" Software is used.
KUM	MLE	"*EasyFit*" Software is used.
	L-ME	$\beta_2 = \frac{\lambda_1 \Gamma\left(\frac{1}{\beta_1}+1+\beta_2\right)}{\Gamma\left(\frac{1}{\beta_1}+1\right)\Gamma(\beta_2)}$ $\beta_1 = \frac{\left(\frac{\lambda_1}{\Gamma\left(\frac{1}{\beta_1}+1\right)\Gamma(\beta_2)}\right)\Gamma(\beta_2)\left(\frac{\Gamma(\beta_2)\Gamma\left(\frac{1}{\beta_1}+1+2\beta_2\right)\Gamma\left(\frac{1}{\beta_1}+1+\beta_2\right)}{\Gamma\left(\frac{1}{\beta_1}+1+2\beta_2\right)}\right)}{\lambda_2}$ (after transforming in to $\frac{x_i - min}{max - min}$)
	MOM	$\overline{z} = \frac{\beta_2 \Gamma\left(\frac{1}{\beta_1}+1\right)\Gamma(\beta_2)}{\Gamma\left(\frac{1}{\beta_1}+1+\beta_2\right)}$, ${S^2}_z = \beta_2 \Gamma(\beta_2)\left[\frac{\Gamma\left(\frac{2}{\beta_1}+1\right)}{\Gamma\left(\frac{2}{\beta_1}+1+\beta_2\right)} - (\overline{z})^2\right]$, a= $min(x_i)$, b= $max(x_i)$, $z = \frac{x_i - min}{max - min}$. Excel “solver” function is used.
KAP	L-ME	Detail derivations are already available in [46]. Here, "*lmomco*" R package is used.
WAK	L-ME	Detail derivations are already available in [46]. Here, "*lmomco*" R package is used.

3.12 Peaks over a Threshold (POT) approach

For the POT approach, the time window in days and the extreme wind speed threshold, u, in meters per second that give the best fit are provided. Generalized Pareto distributions (GPD) are used for modelling the magnitudes of exceedances above a threshold. The 50-year design wind speed - a critical design parameter for building structures per the new Ethiopian Standards [4]: Guide for Actions on Structures - Part 1-4: General actions - Wind actions are also provided.

3.12.1 Threshold selection approaches followed in this thesis

Two, three, four, six, and twelve peak wind speeds per year were selected for POT analysis. If the maxima of two adjacent periods are less than half a period apart, the smaller of the two maxima were replaced by the next smaller value in the respective period which is at least half a block apart from the larger maximum. If the data point is within this duration, the point is considered to be serially correlated and discarded. The next peak that exceeded the threshold and does not lie within one half of time window is then selected. A data sample is thus obtained in which adjacent data are one block apart on the average and never less than half a block apart. Then, the POT analyses were made for selected threshold values and the GPD is finally fitted.

3.12.2 Quantile estimation using GPD

To the extent that gust speed exceedances follow a GPD distribution, we can estimate the GPD quantiles. The expression for the number of exceedances in *T* yrs. is:

$$x(T) = \mu + \frac{\alpha}{k}[1 - (\lambda T)^{-k}] \qquad (3\text{-}13)$$

Where, μ is location parameter, α is scale parameter, k is shape parameter, T is return period and, λ is the average number of exceedances per year ($\frac{m}{n}$), *m* is the size of the threshold values and *n* is the number of years for which data is available.

3.13 Initial visualization and model choice

The acceptance or rejection of a fitted distribution is usually based on visual inspection and formal goodness of fit tests. Two procedures are used for checking the normality: graphical method using probability- probability plots (P-P plots) and formal normality tests.

3.13.1 Probability-Probability Plots (P-P Plots)

Let $F(x_i)$ be the estimated CDF from a random sample $x_1, \ldots, x_n$. Then a plot of $F(x_i)$ against p_i (which is defined in equation 16) is known as the P-P plot. The PP plot helps in assessing how close the plot of $F(x_i)$ against p_i.

The pattern of points in the plot is used to compare the two distributions. If the two distributions being compared are identical, the P–P plot follows the 45° line $y = x$ [11].

$$p_i = \frac{i-\alpha}{n+1-2\alpha}; i = 1,2,\ldots,n \tag{3-14}$$

Where, i is the rank of the observed value x_i (x_i is ascending order), i = 1, 2, 3, ..., n, n is the total number of observed values and α is a constant.

3.13.2 Plotting position formula used in this thesis

There is no agreement among researchers related to the plotting position of the empirical distribution of data. This problem influences in the graphical representation and it can lead into different visual display of the goodness of fit of a distribution model. A general formula to estimate the probability of empirical data is given by Eq. 3-16. Here, a plotting position of the non-exceedance probability p_i is computed for each distribution using the Weibull plotting position formula ($\alpha = 0$ of Eq. 3-14) that yields approximately unbiased quantiles for a wide range of distributions (see [47]).

3.14 Quantitative Goodness of Fit (GoF) tests

There are a number of GoF tests used by various literatures to quantify the GoF, to accept or reject distributions, and to choose between varies fitted distributions. In this thesis, the most well-known tests; namely, the Correlation Coefficient (R^2), the Root Mean Square Error (RMSE), the Kolmogorov-Smirnov (K-S), and the Cramer-VonMises (W^2), were used to check the adequacy of the tested distribution functions for both BM and POT analysis. The purpose of applying these different GoF tests is to look for possible consistencies among the tests and to make a meaningful judgment about better-fit probability models. These tests were also used to compare the MLE, the L-ME and the MOM for estimation of the distribution parameters. Critical values for different significance levels can be found in the literatures.

3.14.1 Coefficient of Determination (R^2) test

R^2 measures the relationship between two variables explained by a linear model. The value of R^2 closer to 1 implies a better fit [20].

The R^2 statistic is calculated as follows:

$$R^2 = \frac{\sum_{i=1}^{n}\left[\left(F(x_i)-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n+1}\right)\right)^2\right]}{\sum_{i=1}^{n}\left[\left(F(x_i)-\frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n+1}\right)\right)^2\right]+\sum_{i=1}^{n}\left[\left(\frac{i}{n+1}-F(x_i)\right)^2\right]} \qquad (3\text{-}15)$$

Where, i represents the ordered dataset, and $F(x_i)$ represents the CDF of the examined theoretical probability models.

3.14.2 Root Mean Square Error (RMSE) test

RMSE is a well-known indicator for the error metric, can be calculated by the square root of the average of the squares of the errors [19]. The RMSE statistic is calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(F(x_i)-\frac{i}{n+1}\right)^2} \qquad (3\text{-}16)$$

Where x_i are the ordered actual observed data and $F(x_i)$ are the CDF of candidate distribution. The lower value of RMSE implies better fit.

3.14.3 Kolmogorov-Smirnov (K-S) test

The Kolmogorov-Smirnov test is given by

$$D = Max\left[F(x_i)-\frac{i-1}{n}, \frac{i}{n}-F(x_i)\right] \qquad (3\text{-}17)$$

Where, $F(x_i)$ is the CDF of the specified distribution.

At the 95% confidence level the critical value is approximately given by [48]:

$$D_{crit,0.05} = \frac{1.3581}{\sqrt{n}}, n > 35 \qquad (3\text{-}18)$$

If calculated value is less than critical value; accept null hypothesis, and if calculated value is greater than table value; reject null hypothesis.

3.14.4 Cramer-VonMises (C-VM) test

The C-VM test statistics can be written as:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x_i) - \frac{i-0.5}{n}\right]^2 \qquad (3\text{-}19)$$

Where *i* are the rank of observed data, *n* is the total number of samples, and $F(x_i)$ is the CDF of candidate distribution. If the test statistic W^2 exceeds the critical value of 0.220 at a 95% confidence level $(\alpha = 0.05)$, the test hypothesis can be rejected. The model with the smaller values of W^2 statistics can be preferred.

3.15 Return period and forecasting

Subsequent to the estimation of the distribution parameters, the return levels should be estimated. One of the main objectives of this thesis is the estimation of the once per T year return value, the value which is exceeded on average once every T year.

The "Once in T Years" value of wind speed is defined as the value of wind speed which has a probability of $\frac{1}{T}$ of being exceeded in any one year is called the "Return Period" or the "Mean Recurrence Interval" [49].

For the extreme wind speed data, parameters of the selected distributions were estimated. Based on the estimate of these parameters, return levels for the forecast periods of 5, 10, 20, 50, and 100 years are determined for each of the distributions. For POT methods the return period T, expressed in years, for a specified level of exceedance p, is computed as:

$$T = \frac{1}{\lambda(1-p)} \qquad (3\text{-}20)$$

Where, λ is the mean number of extreme events per year and *P* the probability of non-exceedance for a single extreme wind speed. For annual maximum data the parameter λ is set to one.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Introduction

This chapter analyses the wind speed data recorded from Addis Ababa-Bole wind speed recording station. The collected data from the station were tabulated and analyzed. Estimates of the extreme wind speed were made based on the Block Maximum (BM) and Peaks Over Threshold (POT) approaches. For the Block Maximum method, the data sampling that give the best fit as determined from a Methods of Moments, Maximum Likelihood Estimation, and L-Moments Estimation is given, and for the POT method, MLE and L-ME were applied.

4.2 Graphical display of original and outlier corrected yearly maximum data

The following graphs show the time series plots of the original data (left) and outlier corrected data (right) for the study area. From graphs in can be seen that wind speeds are increasing from 1960 until it reaches its peak 35 m/s in 1969 and declines from its peak 36 m/s in1982 to 11 m/s in 1993. Generally, high wind speeds were observed from the years 1966-1985 and relatively low wind speeds were observed from 1986-2016.

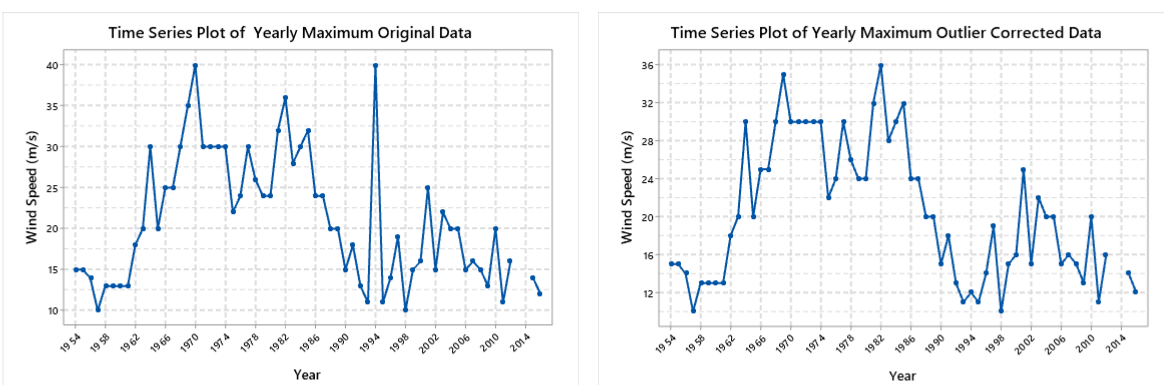


Figure 4-1. Time series plot of yearly maximum original and outlier corrected data.

4.3 Wind direction analysis

Wind speed and wind direction observation recorded from the measurement station have been arranged into a bin-sector of 24, shown in Table 4-1. Analysis suggests that the most frequent wind direction falls between the sectors 82.5˚–97.5˚, (which correspond to East) and 112.5-127.5˚, (which correspond to East-Southeast) with frequencies of 9.372% and 9.739%, respectively. Lower wind speeds appear from 7.5˚ to 22.5˚ with an average frequency of about 1.176%. Graphical representation of Table 4-1 in the form of a wind rose chart is presented in Figure 4-3.

Table 4-1. Wind speed frequency according to wind direction at the study site

Directions	0.5 - 2.1	2.1 - 3.6	3.6 - 5.7	5.7 - 8.8	8.8 - 11.1	≥11.1	Total (%)
352.5 - 7.5	0.75492	0.46878	0.49922	0.25063	0.05885	0.02638	2.05879
7.5 - 22.5	0.4627	0.16641	0.27904	0.19076	0.03653	0.04059	1.17602
22.5 - 37.5	0.48197	0.24352	0.33282	0.20497	0.05682	0.03754	1.35765
37.5 - 52.5	0.63621	0.4008	0.70317	0.48299	0.10553	0.14104	2.46974
52.5 - 67.5	0.54083	0.46574	0.95076	0.59866	0.15626	0.12582	2.83807
67.5 - 82.5	0.97308	1.05223	1.83759	1.40939	0.39674	0.38355	6.05258
82.5 - 97.5	0.95685	1.21457	2.5154	2.52757	0.91423	1.24299	**9.37161**
97.5 - 112.5	0.63621	0.77522	1.90354	2.80357	1.00048	2.06082	**9.17983**
112.5 - 127.5	0.68491	0.95887	2.15316	3.19321	1.39316	1.35562	**9.73892**
127.5 - 142.5	0.78638	0.83711	1.65799	1.86296	0.7255	0.5875	6.45744
142.5 - 157.5	0.4079	0.47081	0.89394	0.90916	0.28411	0.23541	3.20132
157.5 - 172.5	0.74072	0.75289	1.29372	0.94873	0.25874	0.21004	4.20484
172.5 - 187.5	0.71535	0.85335	1.25212	0.72042	0.25671	0.16438	3.96233
187.5 - 202.5	0.53981	0.57228	0.80769	0.46168	0.1796	0.11364	2.6747
202.5 - 217.5	0.32571	0.41399	0.59156	0.38254	0.15322	0.0832	1.95022
217.5 - 232.5	0.41501	0.41501	0.63316	0.38558	0.1177	0.06494	2.03139
232.5 - 247.5	0.31658	0.30948	0.47792	0.32774	0.08523	0.06392	1.58088
247.5 - 262.5	0.4769	0.56315	0.66563	0.34195	0.10857	0.05987	2.21607
262.5 - 277.5	0.59156	0.63824	0.75898	0.4282	0.13191	0.05581	2.60469
277.5 - 292.5	0.66056	0.51749	0.56416	0.37239	0.09639	0.04566	2.25665
292.5 - 307.5	0.63316	0.41095	0.47284	0.3318	0.09437	0.04363	1.98675
307.5 - 322.5	1.15877	0.60982	0.71332	0.58344	0.09132	0.05581	3.21248
322.5 - 337.5	0.77826	0.43834	0.52561	0.35311	0.08929	0.05276	2.23737
337.5 - 352.5	1.1953	0.60272	0.67984	0.41602	0.10045	0.0345	3.02883
Sub-Total	15.8696	14.1518	**23.1632**	20.4875	6.89172	7.28542	87.8492
Calms (Zeros)							12.1508
Total							100

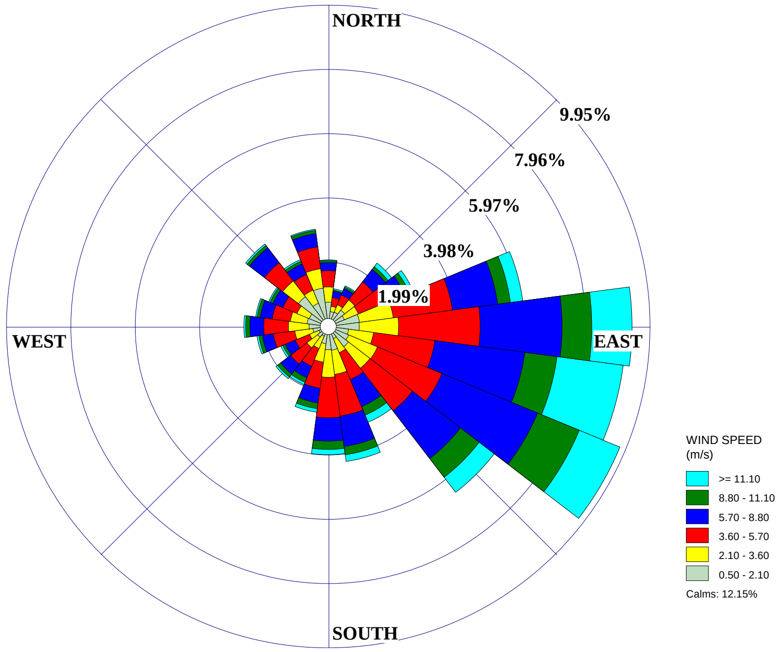


Figure 4-2. Wind rose chart for observed wind speed data at Addis Ababa- Bole site.

4.4 Fitted results of distributions to outlier corrected BM data

After correcting the outliers, the candidate distributions for the annual maximum of daily maximum wind speed data are fitted. Table 4-2 to Table 4-3 show the fitted parameters for each distribution with BM series obtained by using the MOM, L-ME, and MLE.

Table 4-2. Fitted Parameters of two parameter distributions.

Distribution	Parameters	Fitting Methods		
		MLE	L-ME	MOM
GUM	α	5.823465	5.956833	5.629972
	μ	16.961192	16.938676	17.127341
WEI	α	22.850621	22.798561	22.791550
	β	3.115415	3.061144	3.081788
N	σ	7.220720	7.318394	7.220720
	μ	20.377049	20.377049	20.377049
LN	σ	0.357141	0.363099	0.343930
	μ	2.951563	2.948489	2.955265
GAM	α	2.509788	2.717432	2.558702
	β	8.119045	7.498641	7.963823
RAY	α	10.301418	11.247904	11.021705
	μ	7.690270	6.279892	6.563390
REC	a	10	10.245105	10.270803
	b	36	35.742003	35.600700

Table 4-3. Fitted Parameters of three parameter distributions.

Distribution	Fitting Method	Parameters		
WEI3	MLE	α = 11.810708	μ = 9.621609	β =1.445972
	L-ME	α =15.800805	μ =6.371842	β =1.984408
	MOM	α =18.367539	μ =4.094266	β =2.402291
GEV	MLE	α =5.874189	μ =17.020945	k = 0.019046
	L-ME	α =6.410044	μ =17.182073	k =0.085855
	MOM	α =6.649852	μ =15.314225	k =0.158440
GPD	MLE	α =18.313740	μ =10.000000	k =0.691408
	L-ME	α =16.907649	μ =9.705994	k =0.584440
	MOM	α =17.418996	μ =9.592988	k =0.615254
LN3	MLE	σ = 0.487611	μ =2.636002	ξ = 4.749079
	L-ME	σ =0.238025	μ =3.402156	ξ =-10.514524
	MOM	σ =0.129987	μ =4.004595	ξ =-34.937888
GAM3	MLE	α =6.048131	μ =9.394281	β = 1.815895
	L-ME	α =2.630093	μ =-0.633756	β =7.988617
	MOM	α =1.458002	μ =-15.383385	β =24.527007

NB: The shape parameter, k values of GEV is larger than zero which indicates Weibull type.

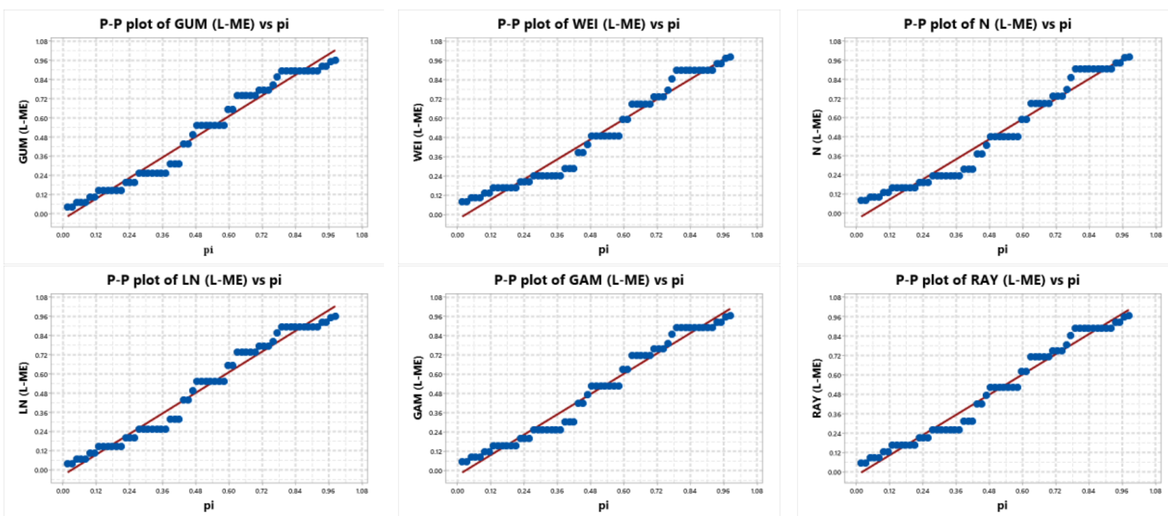
Table 4-4. Fitted Parameters of Four and Five parameter distributions.

Model	Method of Fitting	Parameters			
GLD	L-ME	ξ = 22.791766	α =12.510719	κ = 2.022204	h =0.908775
JSB	MLE	γ = 0.401438	δ = 0.661121	η =27.220694	ζ = 9.641914
	MOM	γ =0.374072	δ =0.601511	η = 25.953810	ζ =10.077756
KAP	L-ME	μ = 0.343299	α = 32.776248	κ =0.938532	h =1.456189
KUM	MLE	β_1 =0.986761	β_2 = 1.536900	a =10	b =36.730879
	L-ME	β_1 =0.851989	β_2 =1.266727	a =10	b =36
	MOM	β_1 =0.849487	β_2 =1.262921	a =10	b =36
WAK	L-ME	Unsuccessful			

Wakeby distributions fitting is not successful (imaginary result was obtained for β). Therefore, Generalized Pareto distribution (GPD) was fitted instead (see [45] and [46] for details).

4.4.1 Goodness of fit assessment using a probability–probability (P–P) plots

The P–P plots drawn for the data series related to the annual maximum wind speed for the selected theoretical distributions are given in (Figure 4-3) to (Figure 4-5).



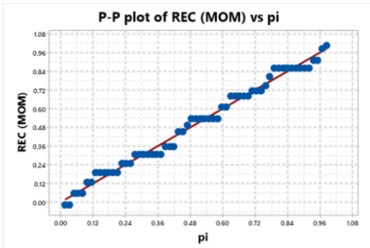


Figure 4-3. P-P plots of Two-Parameter distributions to BM data.

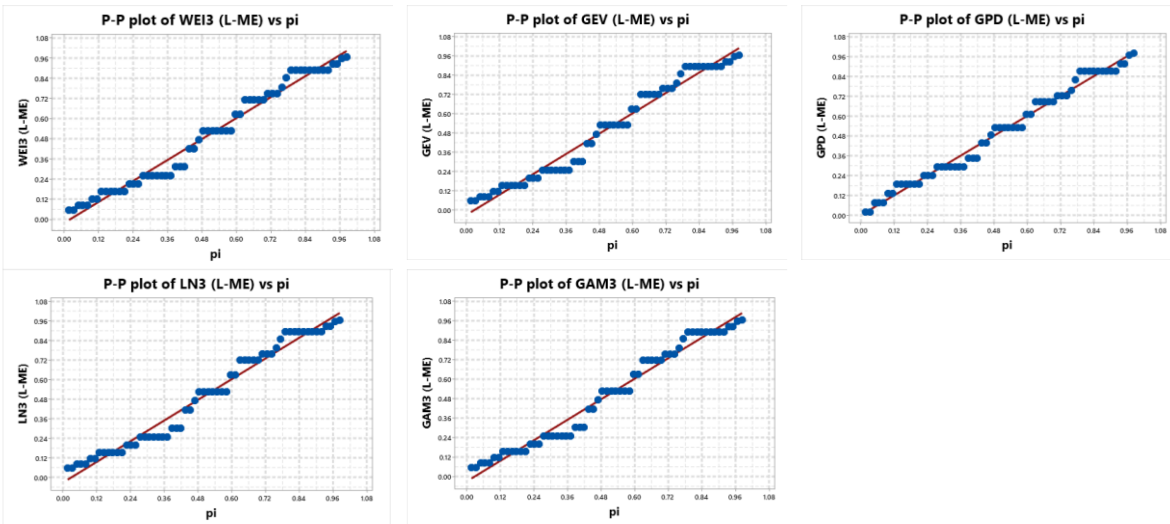


Figure 4-4. P-P plots of Three-Parameter distributions to BM data.

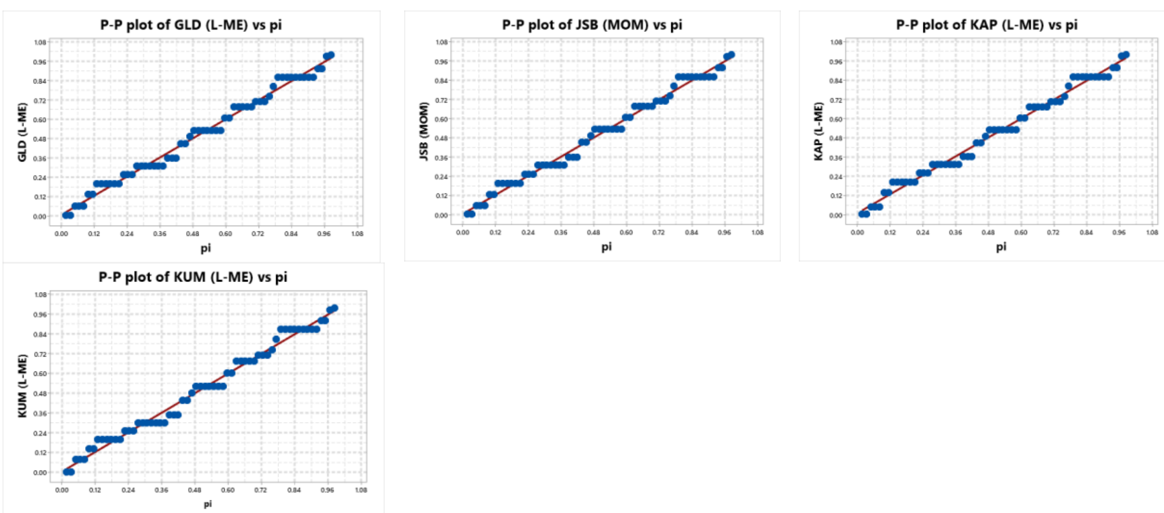


Figure 4-5. P-P plot of Four-Parameter distributions to BM data.

It can be seen from Figs. (4-4) – (4-6), that there is no common representation pattern of all the distributions. The Reciprocal (REC) distribution indicates small gaps between the fitted or theoretical line and the simulated values among the two parameter distributions. From all the figures, it was found that, the GUM and WEI yielded weak-fit results. The trend of weak fitting of the two-parameter distributions can be expected as they are less flexible than three and four-parameter distributions. The Four-parameter distributions matches the data well, with the right tail falling near the $45°$ line among all distributions. Therefore, the observed data series were taken to be the true distribution, and the quantiles estimated from the observed data were also assumed to be the true quantiles of the theoretical distributions. However, to select the best fit probability model not only the graphical observation, but also the numerical test is also important.

4.4.2 Goodness of Fit (GoF) test-based analysis

The GOF tests statistics given by equations (15-19) were applied to the BM series data. The results of the tests for the case when the parameters were estimated with the MLE, L-ME, and, MOM are shown in Tables (4-5) to (4-7). When the parameters were estimated with the L-ME, the results of the test statistics were generally better than those in the case of the MLE and MOM for most of the selected distributions. Generally, it is observed that, the more parameters a distribution has, the better it will fit to the data.

Table 4-5. Results of GoF tests for Two-parameter distributions.

Model	Parameter Fitting Approach	GoF Test Statistics			
		RMSE	R^2	W^2 (Crit.=0.220)	K-S ($Crit.=\frac{1.3581}{\sqrt{n}}$=0.173887)
GUM	MLE	0.057358	0.967686	0.188783	0.130556
	L-ME	0.054384	0.970116	0.170691	0.126640
	MOM	0.063309	0.962597	0.228673	0.144617
WEI	MLE	0.058869	0.963848	0.205636	0.145542
	L-ME	0.056292	0.966245	0.188995	0.139248
	MOM	0.057052	0.965625	0.193602	0.140780
N	MLE	0.063262	0.959531	0.236371	0.154032
	L-ME	0.061796	0.960706	0.226539	0.151339
	MOM	0.063262	0.959531	0.236371	0.154032

LN	MLE	0.055561	0.969329	0.177165	0.129379
	L-ME	0.053293	0.971156	0.163790	0.123123
	MOM	0.061232	0.964542	0.213794	0.140916
GAM	MLE	0.056752	0.968016	0.185521	0.138345
	L-ME	0.051211	0.972517	0.153126	0.126646
	MOM	0.055334	0.969196	0.176812	0.135500
RAY	MLE	0.062984	0.962694	0.226660	0.154484
	L-ME	0.047026	0.975998	0.130517	0.117484
	MOM	0.049425	0.974267	0.142297	0.123101
REC	MLE	0.033021	0.986167	0.073158	0.090069
	L-ME	0.032382	0.987327	0.066651	0.075838
	MOM	0.032589	0.987297	0.066633	0.075416

Table 4-6. Results of GoF tests for Three-parameter distributions.

Model	Parameter Fitting Approach	GoF Test Statistics			
		RMSE	R^2	W^2 (Crit.=0.220)	K-S ($Crit. = \frac{1.3581}{\sqrt{n}}$ =0.173887)
WEI3	MLE	0.048618	0.975629	0.135219	0.133853
	L-ME	0.046906	0.976122	0.129802	0.128436
	MOM	0.051825	0.971420	0.158532	0.128872
GEV	MLE	0.057730	0.967369	0.191161	0.132674
	L-ME	0.052550	0.971298	0.161086	0.130129
	MOM	0.086643	0.916460	0.460835	0.170440
GPD	MLE	0.043821	0.978531	0.114441	0.115964
	L-ME	0.034295	0.985975	0.073570	0.086751
	MOM	0.034600	0.985660	0.075268	0.087385
LN3	MLE	0.054204	0.970372	0.169233	0.122247
	L-ME	0.052825	0.971051	0.162731	0.130192
	MOM	0.057768	0.966135	0.194958	0.141693
GAM3	MLE	0.048722	0.974755	0.138701	0.118322
	L-ME	0.051398	0.972311	0.154348	0.127314
	MOM	0.057156	0.966704	0.190962	0.140486

Table 4-7. Results of GoF tests for Four-parameter distributions.

Model	Parameter Fitting Approach	GoF Test Statistics			
		RMSE	R^2	W^2 (Crit.=0.220)	K-S $(Crit. = \frac{1.3581}{\sqrt{n}}$ =0.173887)
GLD	L-ME	0.031803	0.987574	0.063065	0.081571
JSB	MLE	0.032129	0.987683	0.064561	0.081829
	MOM	0.031314	0.988046	0.063077	0.078010
KAP	L-ME	0.032506	0.986930	0.069048	0.088882
KUM	MLE	0.037742	0.984080	0.083574	0.098692
	L-ME	0.033659	0.986191	0.072665	0.082663
	MOM	0.033641	0.986178	0.072755	0.082522

4.5 Selection of the best fitting distributions

(Table 4-8) to (Table 4-10) shows the statistical summary results for selected distributions, the best-fit results of RMSE, R^2, W^2 and K-S results. The best-fit result is taken as the smallest goodness-of-fit result among RMSE, W^2 and K-S, and the largest goodness-of-fit result for R^2. For selecting the best-fit distribution a ranking system was made, with rank 1 being the best, 2 the second best and so on. The distribution type with the lowest sum of the four rankings is selected as the best fit distribution.

The REC distribution gives the best results in all of the applied tests in the case of Two-parameter distributions when the parameters were estimated with the L-ME, the RAY (L-ME) came second and the GAM (L-ME) third. All the classical GoF test statistics for the K–S, tests were acceptable for the estimation of extreme wind speed based on the test hypothesis at a 95% significance level and the W^2 statistics for GUM (MOM), N, and RAY (MLE) were rejected. Poor test results were obtained with the N (MOM). In the case of Three- parameter distributions, the GPD (L-ME), WEI3 (L-ME), and GAM3 (MLE), distributions came first, second and third, respectively.

For all distributions GoF test statistics for the K-S, and W^2 test were acceptable except the GEV (MOM) in which the W^2 statistic were rejected. The least performance was observed in the GEV (MOM).

Finally, in the case of four- parameter distributions JSB (MOM) comes first in three of the test statistics (RMSE, R^2, and K-S), and second in the W^2 statistics, GLD (L-ME) comes second, and KUM (MOM) third. KUM (MLE) performs least. All the tested Four-parameter distributions could not be rejected for the K-S, and W^2 with the chosen significance level of 0.05. Generally, it is observed that, the JSB (MOM), GLD (L-ME), and REC (L-ME), respectively are the best fitting among all the distributions. The N (MOM), GEV (MOM), and GUM (MOM) distribution models had the least quality of performance. Table 4-11, presents the summary of ranking of all the distribution models.

Table 4-8. Ranking of Two-parameter distributions based on the GoF statistics.

Model	Fitting Approach	GoF test statistics								
		RMSE		R^2		W^2		K-S		Sum
		Statistics	Rank	Statistics	Rank	Statistics	Rank	Statistics	Rank	
REC	L-ME	0.032382	1	0.987327	1	0.066651	2	0.075838	2	**6**
REC	MOM	0.032589	2	0.987297	2	0.066633	1	0.075416	1	**6**
REC	MLE	0.033021	3	0.986167	3	0.073158	3	0.090069	3	**12**
RAY	L-ME	0.047026	4	0.975998	4	0.130517	4	0.117484	4	**16**
RAY	MOM	0.049425	5	0.974267	5	0.142297	5	0.123101	5	**20**
GAM	L-ME	0.051211	6	0.972517	6	0.153126	6	0.126646	8	**26**
LN	L-ME	0.053293	7	0.971156	7	0.163790	7	0.123123	6	**27**
GUM	L-ME	0.054384	8	0.970116	8	0.170691	8	0.126640	7	**31**
GAM	MOM	0.055334	9	0.969196	10	0.176812	9	0.135500	11	**39**
LN	MLE	0.055561	10	0.969329	9	0.177165	10	0.129379	9	**38**
WEI	L-ME	0.056292	11	0.966245	13	0.188995	13	0.139248	13	**50**
GAM	MLE	0.056752	12	0.968016	11	0.185521	11	0.138345	12	**46**
WEI	MOM	0.057052	13	0.965625	14	0.193602	14	0.140780	14	**55**
GUM	MLE	0.057358	14	0.967686	12	0.188783	12	0.130556	10	**48**
WEI	MLE	0.058869	15	0.963848	16	0.205636	15	0.145542	17	**63**
LN	MOM	0.061232	16	0.964542	15	0.213794	16	0.140916	15	**62**
N	L-ME	0.061796	17	0.960706	19	0.226539	17	0.151339	18	**71**
RAY	MLE	0.062984	18	0.962694	17	0.226660	18	0.154484	21	**74**
N	MLE	0.063262	19	0.959531	20	0.236371	20	0.154032	19	**78**
N	MOM	0.063262	20	0.959531	21	0.236371	21	0.154032	20	**82**
GUM	MOM	0.063309	21	0.962597	18	0.228673	19	0.144617	16	**74**

Table 4-9. Ranking of Three-parameter distributions based on the GoF statistics.

Model	Fitting Approach	GoF test statistics								
		RMSE		R^2		W^2		K-S		Sum
		Statistics	Rank	Statistics	Rank	Statistics	Rank	Statistics	Rank	
GPD	L-ME	0.034295	1	0.985975	1	0.073570	1	0.086751	1	**4**
GPD	MOM	0.034600	2	0.985660	2	0.075268	2	0.087385	2	**8**
GPD	MLE	0.043821	3	0.978531	3	0.114441	3	0.115964	3	**12**
WEI3	L-ME	0.046906	4	0.976122	4	0.129802	4	0.128436	7	**19**
WEI3	MLE	0.048618	5	0.975629	5	0.135219	5	0.133853	12	**27**
GAM3	MLE	0.048722	6	0.974755	6	0.138701	6	0.118322	4	**22**
GAM3	L-ME	0.051398	7	0.972311	7	0.154348	7	0.127314	6	**27**
WEI3	MOM	0.051825	8	0.971420	8	0.158532	8	0.128872	8	**32**
GEV	L-ME	0.052550	9	0.971298	9	0.161086	9	0.130129	9	**36**
LN3	L-ME	0.052825	10	0.971051	10	0.162731	10	0.130192	10	**40**
LN3	MLE	0.054204	11	0.970372	11	0.169233	11	0.122247	5	**38**
GAM3	MOM	0.057156	12	0.966704	13	0.190962	12	0.140486	13	**50**
GEV	MLE	0.057730	13	0.967369	12	0.191161	13	0.132674	11	**49**
LN3	MOM	0.057768	14	0.966135	14	0.194958	14	0.141693	14	**56**
GEV	MOM	0.086643	15	0.916460	15	0.460835	15	0.170440	15	**60**

Table 4-10. Ranking of Four-parameter distributions based on the GoF statistics.

Model	Fitting Approach	GoF Test Statistics								
		RMSE		R^2		W^2		K-S		Sum
		Statistics	Rank	Statistics	Rank	Statistics	Rank	Statistics	Rank	
JSB	MOM	0.031314	1	0.988046	1	0.063077	2	0.0780102	1	**5**
GLD	L-ME	0.031803	2	0.987574	3	0.063065	1	0.081571	2	**8**
JSB	MLE	0.032129	3	0.987683	2	0.064561	3	0.0818293	3	**11**
KAP	L-ME	0.032506	4	0.98693	6	0.069048	4	0.0888824	6	**20**
KUM	MOM	0.033641	5	0.986178	4	0.072755	6	0.0825222	4	**19**
KUM	L-ME	0.033659	6	0.986191	5	0.072665	5	0.0826632	5	**21**
KUM	MLE	0.037742	7	0.98408	7	0.083574	7	0.0986916	7	**28**

Table 4-11. Total ranking of all distributions based the BM data.

Model	Fitting Approach	GoF test statistics									
		RMSE		R^2		W^2		K-S		Sum	Total Rank
		Statistics	Rank	Statistics	Rank	Statistics	Rank	Statistics	Rank		
JSB	MOM	0.031314	1	0.988046	1	0.063077	2	0.07801	2	6	**1**
GLD	L-ME	0.031803	2	0.987574	2	0.063065	1	0.081571	3	8	**2**
REC	L-ME	0.032382	3	0.987327	3	0.066651	3	0.075838	1	10	**3**
KAP	L-ME	0.032506	4	0.98693	4	0.069048	4	0.088882	6	18	**4**
KUM	MOM	0.033641	5	0.986178	5	0.072755	5	0.082522	4	19	**5**
GPD	L-ME	0.034295	6	0.985975	6	0.07357	6	0.086751	5	23	**6**
RAY	L-ME	0.047026	8	0.975998	8	0.130517	8	0.117484	7	31	**7**
WEI3	L-ME	0.046906	7	0.976122	7	0.129802	7	0.128436	12	33	**8**
GAM3	MLE	0.048722	9	0.974755	9	0.138701	9	0.118322	8	35	**9**
GAM	L-ME	0.051211	10	0.972517	10	0.153126	10	0.126646	11	41	**10**
GEV	L-ME	0.05255	11	0.971298	11	0.161086	11	0.130129	13	46	**11**
LN	L-ME	0.053293	13	0.971156	12	0.16379	13	0.123123	9	47	**12**
LN3	L-ME	0.052825	12	0.971051	13	0.162731	12	0.130192	14	51	**13**
GUM	L-ME	0.054384	14	0.970116	14	0.170691	14	0.12664	10	52	**14**
WEI	L-ME	0.056292	15	0.966245	15	0.188995	15	0.139248	15	60	**15**
N	L-ME	0.061796	16	0.960706	16	0.226539	16	0.151339	16	64	**16**

4.6 Analysis results of POT series

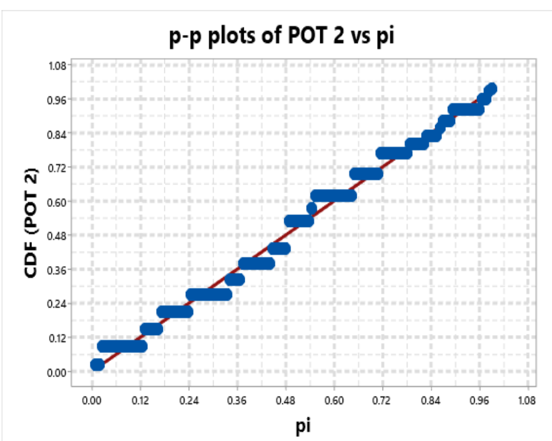
The POT analyses were made for different threshold values. Five possible choices for threshold values were applied to the data for the period 1954–2016. The RMSE, R^2, C-VM and K-S tests were used to find out which of these possible threshold levels gave the best fit to the POT sample. The results of the tests for different threshold values for the POT series are given in Table 4-13. It is observed that the best threshold of the one considered would be POT 2 (L-ME), which determines a lowest (RMSE, W^2, and K-S) and highest R^2 statistics. Therefore, POT 2 were selected for POT analysis. Lechner, S. D. Leigh [50] write that, the variation in sign of the shape parameter, negative or zero shape parameters make the function unbounded which is physically incorrect since natural wind speeds have a finite maximum. Positive values make the function converge to a maximum. Note: in this work the sign of the shape parameter is based on authors such as John D. Holmes [7] and Palutikof J. P., B. B. Brabson [26] in the GPD expression of Table 3-4.

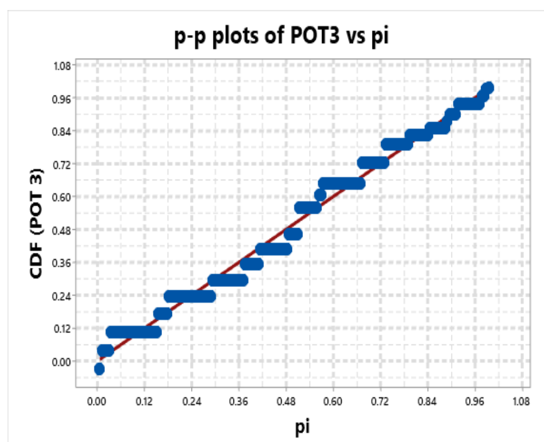
Table 4-12. Results for the POT method and parameters estimated with the MLE, L-ME.

Number POT samples	Sample Size	Fitting Approach	GPD Parameters		
			μ	α	k
POT 2	122	MLE	9	14.747316	0.526287
		L-ME	8.697529	15.020236	0.529732
POT 3	180	MLE	8	14.951174	0.519900
		L-ME	8.541571	14.011312	0.492756
POT 4	240	MLE	8	13.985763	0.481908
		L-ME	7.87140	13.953246	0.487775
POT 6	358	MLE	7	13.163850	0.440658
		L-ME	7.779610	12.114650	0.409145
POT 12	705	MLE	5	13.596126	0.430524
		L-ME	6.963383	10.341531	0.317844

4.7 Best fitting POT series

In this study, numerical: Root Mean Square Error (RMSE), Correlation Coefficient (R^2), Cramer- VonMises (W^2) and Kolmogorov- Smirnov (K-S) tests, and graphical P-P plots (Figure 4-6) were calculated to find the best fitting POT series. All five POT values (POT2, POT3, POT4, POT6, and POT12) show that the POT 2 series are the best fitting POT series among all POT candidates.





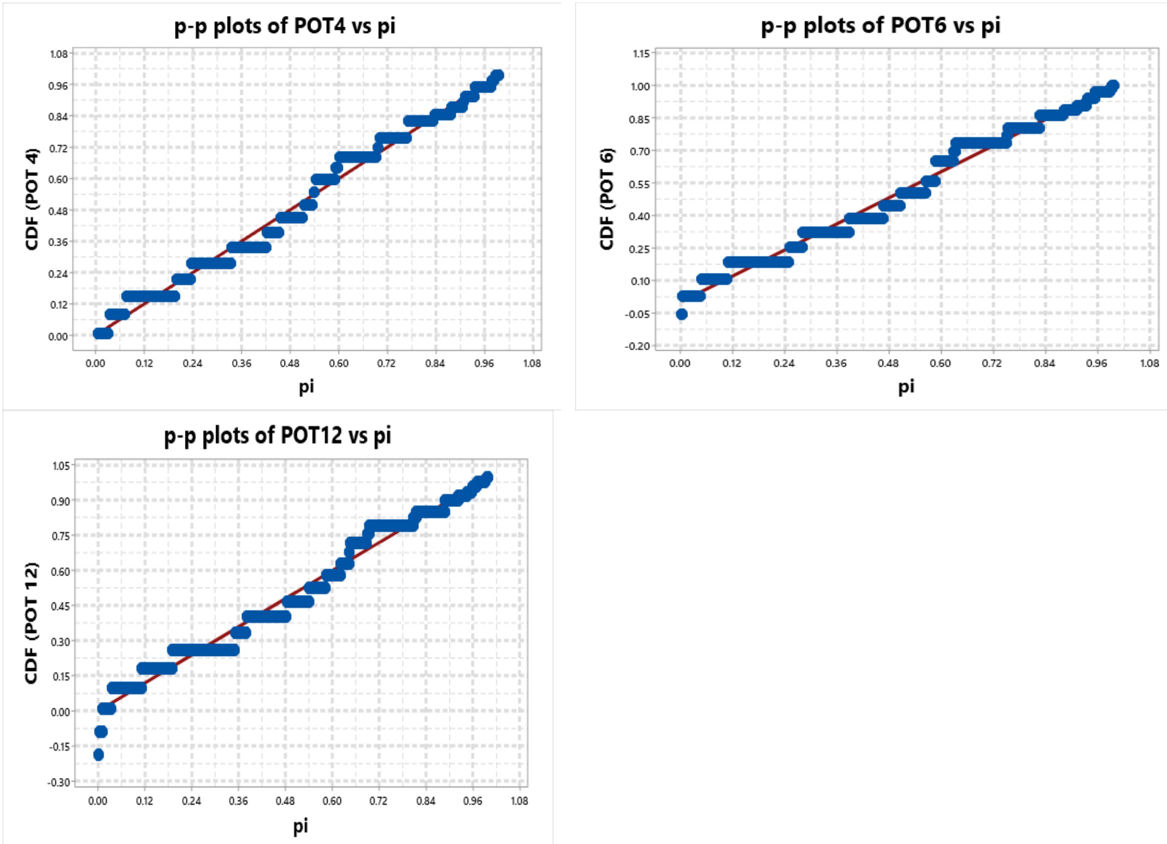


Figure 4-6. P-P plots of all candidate POT values.

Table 4-13. GoF results of different POT values for selection of best POT level.

Number of POT	Fitting Approach	RMSE	R^2	W^2	K-S
POT 2	MLE	0.029980	0.989777	0.106811	0.082341
POT 2	L-ME	0.027465	0.991024	0.092832	0.069028
POT 3	MLE	0.032965	0.986113	0.092113	0.068325
POT 3	L-ME	0.031140	0.988593	0.119096	0.074103
POT 4	MLE	0.035748	0.985182	0.306283	0.096287
POT 4	L-ME	0.034776	0.985863	0.290457	0.087769
POT 6	MLE	0.040262	0.978813	0.590262	0.107360
POT 6	L-ME	0.036392	0.984617	0.474010	0.098718
POT 12	MLE	0.066766	0.961466	3.166142	0.174308
POT 12	L-ME	0.041669	0.989253	1.221056	0.203851

4.8 Comparison of analysis results between POT and BM series

The comparison between BM and POT series was based on daily maximum wind speed values. Table 23 presents the sample size and GoF results of the BM and POT series. The POT 2 sample give better results in three of the four applied statistical tests than the BM data. Therefore, POT 2 was chosen as optimal POT series for the data in this work.

Table 4-14. Comparison of GoF results of best fit BM data and selected POT values

	Data Series					
	BM	POT 2	POT 3	POT 4	POT 6	POT 12
Sample size	61	122	180	240	358	705
RMSE	0.031314	0.027465	0.031140	0.034776	0.036392	0.041669
R^2	0.988046	0.991024	0.988593	0.985863	0.984617	0.989253
W^2 (Crit. =0.220)	0.063077	0.092832	0.119096	0.290457	0.47401	1.221056
K-S (Critical value in bracket)	0.07801 (0.173887)	0.068627 (0.122957)	0.074103 (0.101227)	0.087769 (0.087665)	0.098718 (0.071778)	0.203851 (0.051149)

4.9 Quantile estimates based on BM approach

It has been found that the JSB, GLD, REC, and KAP distribution models gives a very good fit to the annual maximum observations so that estimations of the maximum wind speed could be given for any duration of return periods. Using quantile function given in (Table 3-3) to (Table 3-5) , estimates of 5, 10, 20, 50, and 100 years return levels are obtained. It can be seen that, the observed patterns of estimated quantiles obtained from the first five best fitted distributions (Table 4-15) are quite similar.

Table 4-15. Return levels (in m/s) for all distribution models based on BM approach.

Distribution	Best Fitting Method	Return period (years)				
		5	10	20	50	100
JSB	MOM	28	31	33	35	35
GLD	L-ME	28	31	33	34	35
REC	L-ME	28	32	34	35	35
KAP	L-ME	28	31	33	34	35
KUM	MOM	28	31	33	35	35
GPD	L-ME	27	31	34	36	37
RAY	L-ME	26	30	34	38	40
WEI3	L-ME	26	30	34	38	40
GAM3	MLE	26	31	36	43	47
GAM	L-ME	26	30	34	38	42
LN	L-ME	26	30	35	40	44
GEV	L-ME	26	30	34	38	42
LN3	MLE	26	30	34	38	42
GUM	L-ME	26	30	35	40	44
WEI	L-ME	27	30	33	36	38
N	L-ME	27	30	32	35	37

4.10 Quantile estimates based on POT approach

The equation presented in equation 4-13 is applied here to forecast the return values for a return periods of 5, 10, 20, 50, and, 100 years.

Table 4-16. Return levels (in m/s) based on the POT 2 series.

Return period, years	Quantile estimation using MLE			Quantile estimation using L-ME		
	μ	α	k	μ	α	k
	9	14.74732	0.526287	8.697529	15.020236	0.529732
5	29			29		
10	31			31		
20	33			33		
50	**35**			**35**		
100	35			36		

The return level with return period of *T* years for the POT models are given in Table 4-15. For the following reasons the threshold selection is successful: the observed patterns of estimated quantiles obtained from MLE and L-ME generally show little variation between adjacent return periods, and the estimated GoFs, is satisfactorily low.

Generally, from Tables 4-15 and 44-16, it can be seen that the estimated extreme wind speed quantile values from the best fit BM series and best fit POT series are quite similar. Therefore, 35m/s is the estimated return value for 50 years return period. However, attention should be given for estimating 100 years return value; the first five best fit distributions performs poor for forecasting long return periods.

CHAPTER 5 CONCLUSIONS AND RECCOMENDATIONS

5.1 Conclusions

This thesis presents key wind design parameters required in the design of civil engineering structures in several possible locations of Addis Ababa. These design parameters were estimated based on wind speed data from Bole recording station for the period 1954–2016. Extreme wind speed frequency analysis was carried out with the Block Maxima (BM) and Peaks Over a Threshold (POT) series. Extreme wind speeds and related goodness-of-fit (GoF), return periods and associated peak wind speeds for design purpose have been calculated.

The wind speed distribution was modeled by seventeen commonly used probability distribution functions to the BM data, with the parameters determined by three parameter estimation methods, Methods of Moments (MOM), L-moments Estimation (L-ME), and, Maximum Likelihood Estimations (MLE). The graphical procedures using probability-probability (P-P) plots and numerical criteria using Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Cramer- VonMises (W^2), and Kolmogorov- Smirnov (K-S) goodness-of-fit test statistics were applied to investigate the descriptive and predictive abilities of each distribution.

One of the main limitations concerning the Peaks Over Threshold approach is the selection of the threshold value. Two-peaks per year (POT 2), Three-peaks per year (POT 3), Four-peaks per year (POT 4), Six-peaks per year (POT 6), and Twelve-peaks per year (POT 12) threshold values were applied and compared to the data used in this thesis. The Generalized Pareto Distribution (GPD) was used for modelling the POT series. Four statistical goodness of fit tests (RMSE, R^2, W^2, and K-S) were applied to find which threshold value gave the best fit to the data in addition to P-P plots. The overall objectives of this study, as set out in chapter one has been achieved. Based on the results carried out within the scope stated, several conclusions can be drawn, which may help to improve the extreme wind speed estimation and provide a better understanding on the actual subject area.

Generally, this thesis has yielded the following main conclusions;

- Wind direction trends were analyzed and summarized in the form of a wind rose chart, and the results showed that majority of the wind comes from the East (E) and the East-Southeast (E-SE) directions.
- The MOM, L-ME, and, MLE, showed different performances for different probability distributions. The MOM gives better results for JSB and KUM distributions, in the case of Reciprocal (REC) distribution, MOM and L-ME give the same results in the three of the four GoF tests. The L-ME generally showed better results than the MOM and the MLE for most of distributions.
- All four and most of three parameter distributions could not be rejected by the K-S and W^2 tests with a confidence interval of 95% for BM series.
- The JSB distribution gave the best fit to the BM data with parameters estimated using the MOM. The GEV (MOM) distribution gave a poor fit to the BM sample. Generally, it is observed that, the more parameters a distribution has, the better it will fit to the data.
- The POT sample with two peaks per year gave better results than the best fit of the BM series; JSB (MOM). The L-ME estimates better in searching for the optimal solution for POT series.
- The results from these two methods were compared and the POT method was found to perform better than BM method. The predictions made by POT method are therefore recommended for design over the BM method. Consequently, the required design values of **V(ref) = 35 m/s** should be adopted to predict the return level with return period of 50 years.

5.2 Recommendations

- Based on the identified distribution function, the extreme wind speeds were estimated for specific return periods. Overall, the current work provides a reasonable and comprehensive way to estimate extreme wind speeds. However, in order to gain precise predictions, similar analysis should also be performed for all wind speed recording stations in Addis Ababa.

- Since this study focused on extreme wind speeds irrespective of their direction, it would be interesting to study the extreme wind speeds based on their direction and compare the results so that more definitive conclusions may be reached that were not included in this study.

- A final decision on whether to include the outliers corrected in this study for the analysis might require further investigation of whether they are due to measurement or other man-made error, or are accurate observations perhaps arising from a typical meteorological phenomenon. The missing values in 2013 and 2014 should also be investigated.

5.3 Limitations

One of the challenges faced during this research was to get many stations to use for analysis. Because relying on a single meteorological site may cause uncertainties.

REFERENCES

1. Caires, S., *Extreme Wind Statistics for the Hydraulic Boundary Conditions for the Dutch Primary Water Defences*. SBW-Belastingen: Phase 2 of Subproject 'Wind Modeling', 2009.

2. WMO, *Guidelines on Analysis of Extremes in a Changing Climate in Support of Informed Decisions for Adaptation*. 2009: Geneva, Switzerland.

3. L.C. Pagnini and G. Solari, *Preliminary elements for an innovative wind map of Italy*, in *EACWE 5*. 2009: Florence, Italy. p. 1-12.

4. Ethiopian Standards, *ES EN1991:2015, Part 1-4 Actions on Structures - Part 1-4 : General actions - Wind actions*. 2015, Addis Ababa: Minstry of Construction.

5. Zaharim, A., et al. *Analyzing Malaysian Wind Speed Data Using Statistical Distribution*. in *Proceedings of the 4th IASME / WSEAS International Conference on Energy & Environment*. 2009. University of Cambridge.

6. Simiu, E. and D. Yeo, *Wind Effects on Structures: Fundamentals and Appilications to Design*. 2019, USA: John Wiley & Sons Ltd.

7. John D. Holmes, *Wind Loading of Structures*. Third ed. 2015, Boca Raton: CRC Press, Taylor & Francis Group, LLC.

8. Vagenas, C., C. Anagnostopoulou, and K. Tolika, *Climatic Study of the Marine Surface Wind Field over the Greek Seas with the Use of a High Resolution RCM Focusing on Extreme Winds.* climate, 2017. **5**(25): p. 1-19.

9. Fikadie Alamirew, *Rererence Basic Wind Speed for Selected Towns/Cities in Ethiopia*, in *School of Civil and Environmental Engineering*. 2016, Addis Ababa institute of Technology (AAiT): Addis Ababa.

10. Arslan, T., Y.M. Bulut, and A.A. Yavuz, *Comparative study of numerical methods for determining Weibull parameters for wind energy potential.* Renewable and Sustainable Energy Reviews, 2014. **40**: p. 820–825.

11. S. S. Eljabri, *New Statistical Models for Extreme Values*, in *School of Mathematics*. 2013, Faculty of Engineering and Physical Sciences: University of Manchester.

12. Natei Ermias Benti et al., *Evaluations of Wind Speed Distribution and Wind Power Potential over Ethiopia (A case of Ambo).* Journal of physical and chemical sciences V5I4, 2017.

13. Wais, P., *Two and three-parameter Weibull distribution in available wind power analysis.* Renewable Energy, 2017. **103**: p. 15–29.

14. Baseer, M.A., et al., *Wind power characteristics of seven data collection sites in Jubail, Saudi Arabia using Weibull parameters.* Renewable Energy, 2017. **102**: p. 35–49.

15. Keyhani, A., et al., *An assessment of wind energy potential as a power generation source in the capital of Iran, Tehran.* Energy, 2010. **35**: p. 188–201.

16. Torrielli, A., *Long term simulation and reliability analysis of structures subjected to the wind action*. 2011.

17. Soukissian, T., *Use of multi-parameter distributions for offshore wind speed modeling: The Johnson SB distribution.* Applied Energy, 2013. **111**: p. 982–1000.

18. Brano, V.L., et al., *Quality of wind speed fitting distributions for the urban area of Palermo, Italy.* Renewable Energy 2011. **36**: p. 1026-1039.

19. Akgül, F.G. and B. Şenoğlu, *Comparison of wind speed distributions: a case study for Aegean coast of Turkey.* Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2019.

20. Alavi, O., K. Mohammadi, and A. Mostafaeipour, *Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran.* Energy Conversion and Management 2016. **119**: p. 101–108.

21. T.B.M.J. Ouarda, et al., *Probability distributions of wind speed in the UAE.* Energy Conversion and Management 2015. **93**: p. 414–434.

22. Caires, S., *A Comparative Simulation Study of the Annual Maxima and the Peaks-Over-Threshold Methods.* Journal of Offshore Mechanics and Arctic Engineering, 2016. **138**.

23. Dombry, C., *Maximum likelihood estimators for the extreme value index based on the block maxima method.* arXiv:1301.5611, 2013.

24. G. Anastasiades and P. E. McSharry, *Extreme value analysis for estimating 50 year return wind speeds from reanalysis data.* Wind Energy, 2014. **17**: p. 1231–1245.

25. Bezak, N., M. Brilly, and M. Šraj, *Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis.* Hydrological Sciences Journal – Journal des Sciences Hydrologiques, 2014. **59**(5): p. 959-977.

26. Palutikof J. P., et al., *A Review of Methods to Calculate Extreme Wind Speeds.* Meteorological Applications, 1999. **6**(2): p. 119-132.

27. Simiu, E. and N.A. Heckert, *Extreme Wind Distribution Tails: A "Peaks Over Threshold" Approach.* Journal of Structural Engineering ,ASCE, 1996. **122**(5): p. 539-547.

28. Guedes Soares C. and M. G. Scotto, *Application of the r largest order statistics for long-term predictions of significant wave height.* Coastal Engineering, 2004. **51**: p. 387– 394.

29. Viselli, A.M., et al., *Estimation of extreme wave and wind design parameters for offshore wind turbines in the Gulf of Maine using a POT method.* Ocean Engineering 2015. **104**: p. 649–658.

30. Alaswed, H.A. and M.A. Mohamed, *Threshold Selection Using Graphical Methods.* Journal of Pure & Applied Sciences, 2018. **18**(1): p. 1-5.

31. Ragan, P. and L. Manuel, *Statistical Extrapolation Methods for Estimating Wind Turbine Extreme Loads.* Journal of Solar Energy Engineering 2008. **130**.

32. Toft, H.S., et al., *Response load extrapolation for wind turbines during operation based on average conditional exceedance rates*. Wind Energy, 2011. **14**: p. 749–766.

33. Cunnane, C., *A particular comparison of annual maxima and partial duration series methods of flood frequency prediction.* Journal of Hydrology, 1973. **18**(4): p. 257- 271.

34. Madsen, H., P.F. Rasmussen, and D. Rosbjerg, *Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling.* Water resources research, 1997. **33**(4): p. 747-757.

35. Bucher, A. and C. Zhou, *A horse racing between the block maxima method and the peak–over–threshold approach.* http://arxiv.org/abs/1807.00282v1, 2018.

36. Ferreira, A. and L. deHaan, *On the Block Maxima Method in Extreme Value Theory: PWM Estimators.* The Annals of Statistics, 2015. **43**(1): p. 276–298.

37. Soukissian, T.H. and C. Tsalis, *The effect of the generalized extreme value distribution parameter estimation methods in extreme wind speed prediction.* Natural Hazards, 2015. **78**: p. 1777-1809.

38. Morgan, E.C., et al., *Probability distributions for offshore wind speeds.* Energy Conversion and Management, 2011. **52**: p. 15–26.

39. Teimouri, M., et al., *Comparison of estimation methods for the weibull distribution.* A Journal of Theoretical and Applied Statistics, 2013. **47**(1): p. 93–109

40. D.K. Kidmo, et al., *Statistical analysis of wind speed distribution based on six Weibull Methods for wind power evaluation in Garoua, Cameroon.* Revue des Energies Renouvelables, 2015. **18**(1): p. 105-125.

41. M. Lemma, *Climate Data observation, climate monitoring and long range forecast methods in Ethiopia*, in *WMO Workshop on Climate Monitoring including the implementation of Climate Watch Systems in RA-I with focus on Eastern and Southern Africa*. 2013: Pretoria, South Africa, April 15 - 18.

42. S. M. Papalexiou and D. Koutsoyiannis, *Battle of extreme value distributions: A global survey on extreme daily rainfall.* Water Resources Research, 2013. **49**: p. 187–201.

43. W. H. Asquith, *L-moments and TL-moments of the generalized lambda distribution.* Computational Statistics & Data Analysis, 2007. **51**: p. 4484–4496.

44. Singh, V.P., *Entropy-Based Parameter Estimation in Hydrology*. 1998, U.S.A.: Springer Science+Business Media Dordrecht.

45. K. H. Hamed and A. R. Rao, *Flood Frequency Analysis*. 2000, USA: CRC Press.

46. J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis: An Approach Based on L-Moments.* . 1997, Cambridge, UK: Cambridge University Press.

47. L. Makkonen, *Bringing Closure to the Plotting Position Controversy.* Communications in Statistics - Theory and Methods, 2008. **37**(3): p. 460-467.

48. Paul Chiou, Weiwen Miao, and T. C. Ho, *The Annual Maximum Wind Speed at Pisa Airport in Italy.* International Journal of Environmental Science and Development, 2012. **3**(3): p. 1-5.

49. Lawson, T., *Building Aerodynamics*. 2001, London: Imperial College Press.

50. Lechner, J.A., S. D. Leigh, and E. Simiu, *Recent approaches to extreme value estimation with application to wind speeds. Part I: The Pickands method.* . Journal of Wind Engineering and Industrial Aerodynamics, 1992. **41**: p. 509–519.

51. H. Rinne, *The Weibull Distribution: A Handbook*. 2009, Boca Raton,Germany: Chapman & Hall/CRC.

52. H. N. Phien and T. J. Ajirajah, *Fitting the SB Curve By the Method of Maximum Likelihood.* Journal of Hydrology, 1984. **67**: p. 67-75.

Appendix A

Table A-5-1. Yearly maximum wind speed for Addis Ababa at Bole recording station.

Year	Wind Speed (m/s)	Year	Wind Speed	Year	Wind Speed (m/s)
1954	15	1975	22	1996	14
1955	15	1976	24	1997	19
1956	14	1977	30	1998	10
1957	10	1978	26	1999	15
1958	13	1979	24	2000	16
1959	13	1980	24	2001	25
1960	13	1981	32	2002	15
1961	13	1982	36	2003	22
1962	18	1983	28	2004	20
1963	20	1984	30	2005	20
1964	30	1985	32	2006	15
1965	20	1986	24	2007	16
1966	25	1987	24	2008	15
1967	25	1988	20	2009	13
1968	30	1989	20	2010	20
1969	35	1990	15	2011	11
1970	40	1991	18	2012	16
1971	25	1992	13	2013	Missing
1972	30	1993	11	2014	Missing
1973	30	1994	40	2015	14
1974	30	1995	11	2016	12

Appendix B

Table B-5-2. Some descriptive statistics of the data used for analysis in this thesis.

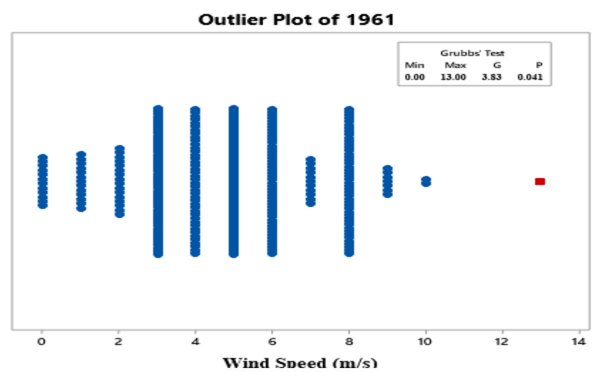
Year	n	Mean	S	Min	Max	Year	n	Mean	S	Min	Max
1954	366	5.044	2.819	0	15	1985	365	13.551	5.336	4	32
1955	365	5.433	2.886	0	15	1986	212	11.92	4.606	5	24
1956	366	5.443	2.612	0	14	1987	365	7.603	2.335	3	24
1957	332	5.53	2.176	0	10	1988	366	7.639	2.57	3	20
1958	364	5.341	2.949	0	13	1989	334	7.799	2.561	3	20
1959	365	4.764	2.576	0	13	1990	365	7.071	2.296	3	15
1960	362	5.577	2.26	0	13	1991	361	7.313	2.421	3	18
1961	364	4.758	2.154	0	13	1992	366	6.145	1.986	2	13
1962	363	5.154	1.998	0	18	1993	365	5.6055	1.8151	0	11
1963	333	9.835	3.624	0	20	1994	365	6.071	2.493	1	12
1964	366	12.609	4.831	0	30	1995	364	5.706	1.7873	2	11
1965	365	9.564	4.726	0	20	1996	366	6.273	2.208	2	14
1966	365	9.241	4.445	1	25	1997	365	8.047	2.505	3	19
1967	364	11.495	4.928	0	25	1998	333	5.9279	1.7735	2	10
1968	366	12.281	4.882	2	30	1999	273	7.813	2.357	3	15
1969	365	13.123	5.719	2	35	2000	366	6.811	2.531	0	16
1970	365	12.384	5.435	0	30	2001	365	7.225	2.685	3	25
1971	365	12.258	5.112	0	30	2002	365	7.458	2.378	3	15
1972	366	12.773	5.336	0	30	2003	365	7.619	2.827	2	22
1973	365	13.145	5.231	0	30	2004	366	7.525	2.888	2	20
1974	344	13.113	4.996	0	30	2005	334	7.323	2.421	3	20
1975	365	12.885	3.974	6	22	2006	365	7.2	2.203	3	15
1976	366	8.38	4.247	0	24	2007	365	7.266	2.416	3	16
1977	365	13.047	4.463	5	30	2008	366	7.596	2.504	3	15
1978	365	14.025	4.437	4	26	2009	359	6.362	1.931	2	13
1979	365	12.307	4.208	4	24	2010	359	5.5209	1.6575	2	20
1980	366	12.385	4.307	4	24	2011	335	5.89	1.845	2	11
1981	365	13.227	4.745	4	32	2012	334	6.063	1.915	2	16
1982	365	13.326	5.014	4	36	2015	121	6.669	1.938	2	14
1983	365	13.189	4.312	4	28	2016	245	5.327	1.826	0	12
1984	366	15.068	5.057	4	30						

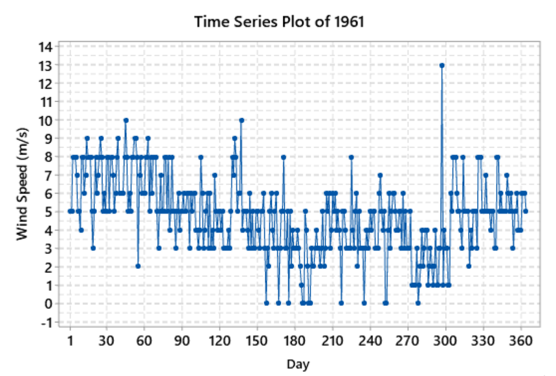
n =Number of records per year, S = Standard deviation.

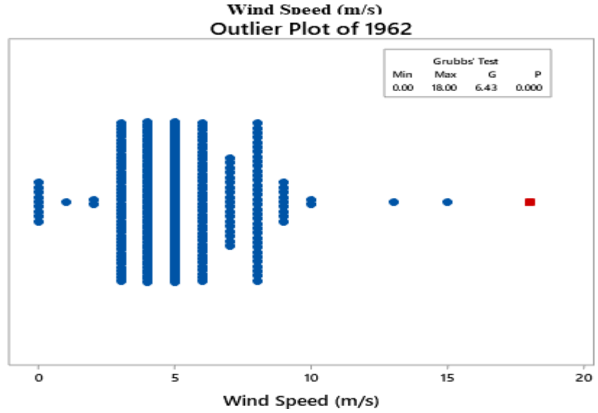
Appendix C

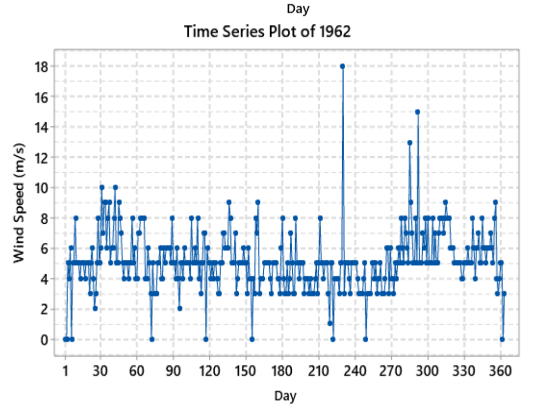
Table C-5-3. All outliers detected from daily maximum data.

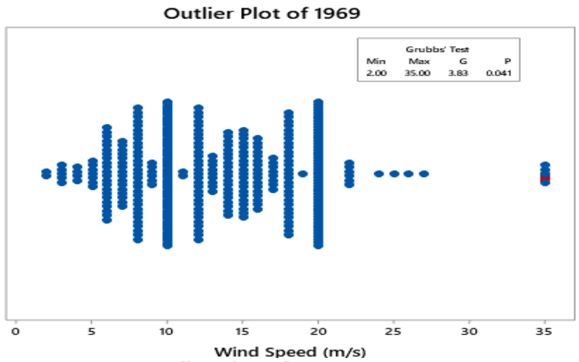
Year	Outlier	Year	Outlier
1961	13	1991	18
1962	18	1994	40
1969	35	1997	19
1970	40	2001	25
1977	30	2003	22
1981	32	2004	20
1982	36	2005	20
1987	24	2010	20
1988	20	2012	16
1989	20	2015	14

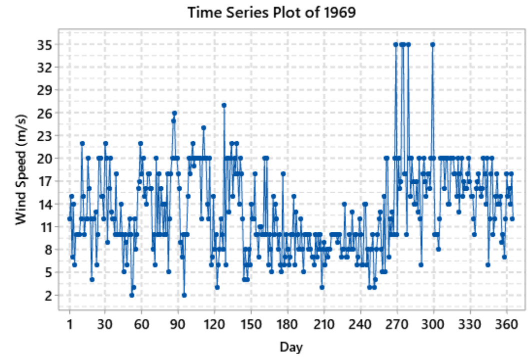


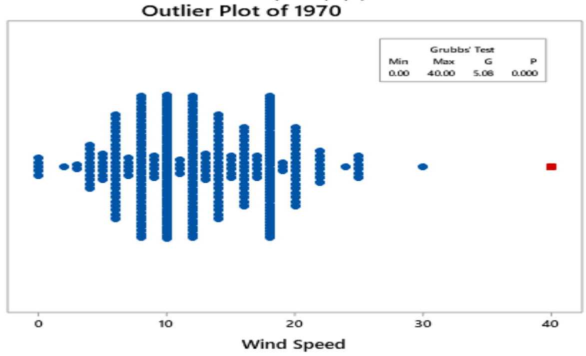


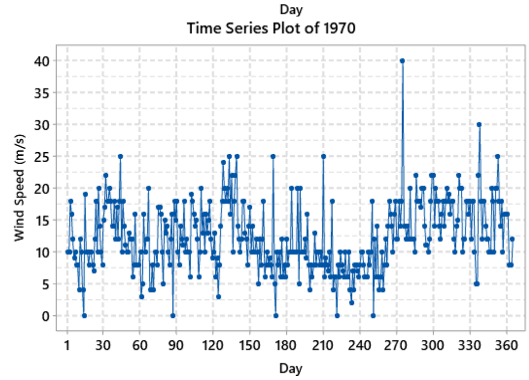


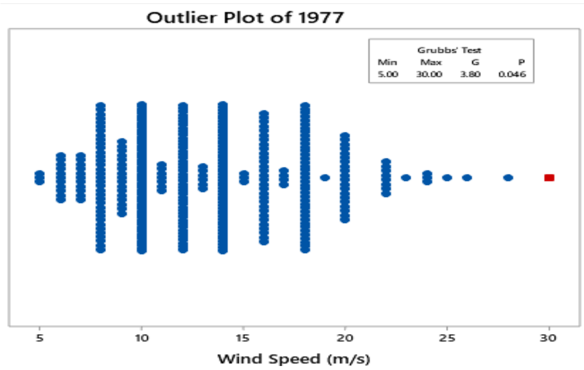


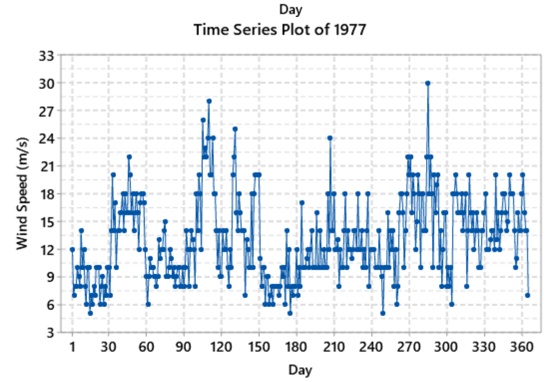

Outlier Plot of 1969
Grubbs' Test
Min Max G P
2.00 35.00 3.83 0.041
Wind Speed (m/s)

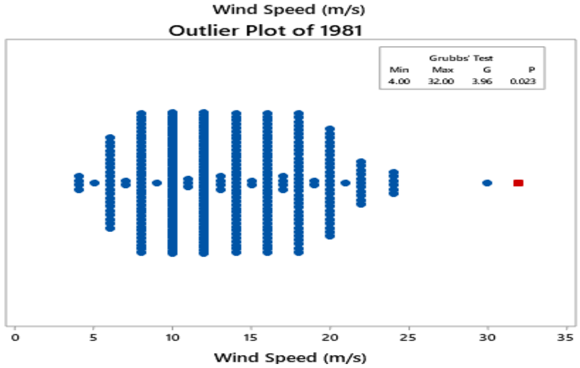

Time Series Plot of 1969
Wind Speed (m/s)
Day

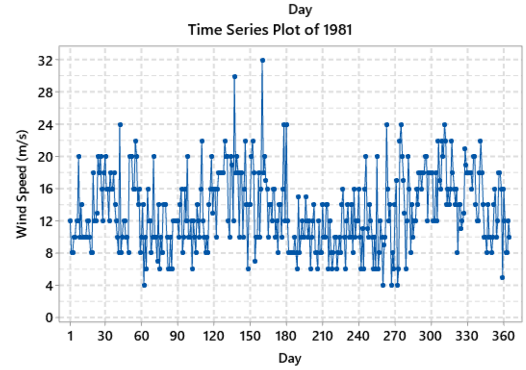

Outlier Plot of 1970
Grubbs' Test
Min Max G P
0.00 40.00 5.08 0.000
Wind Speed

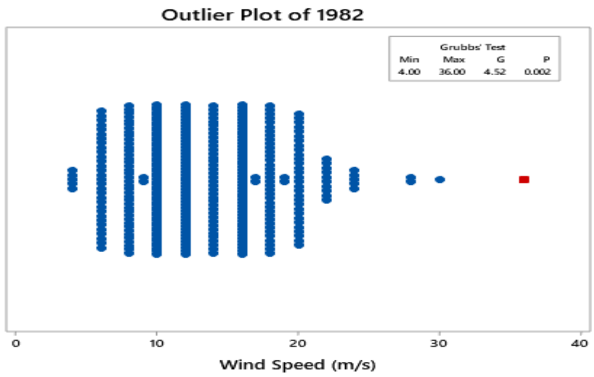

Time Series Plot of 1970
Wind Speed (m/s)
Day

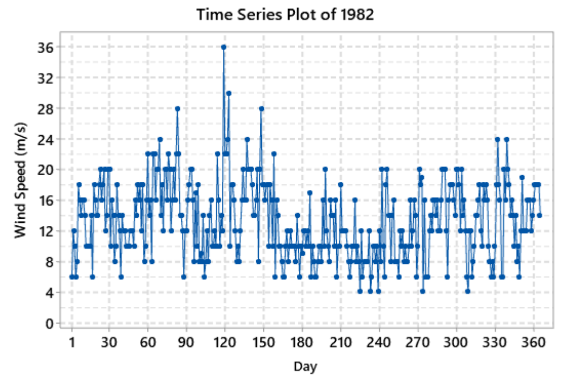

Outlier Plot of 1977
Grubbs' Test
Min Max G P
5.00 30.00 3.80 0.046
Wind Speed (m/s)

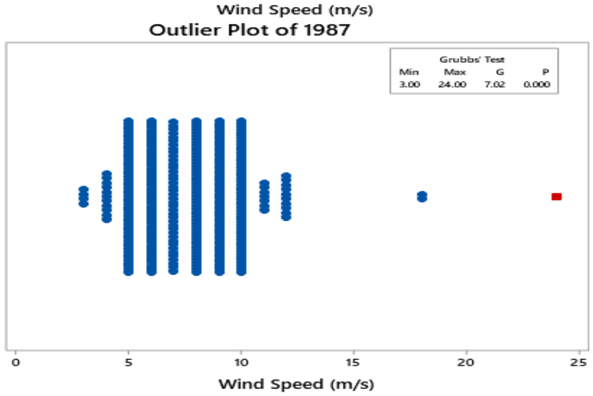

Time Series Plot of 1977
Wind Speed (m/s)
Day

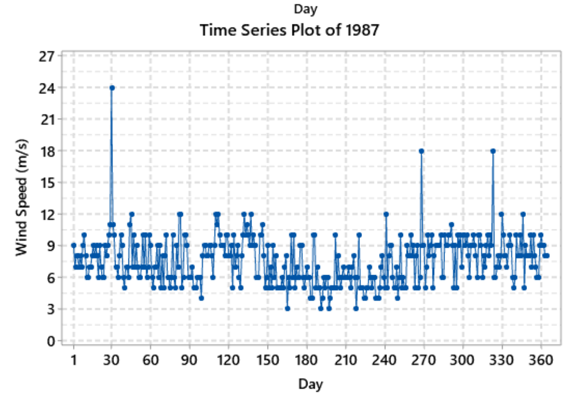

Outlier Plot of 1981
Grubbs' Test
Min Max G P
4.00 32.00 3.96 0.023
Wind Speed (m/s)

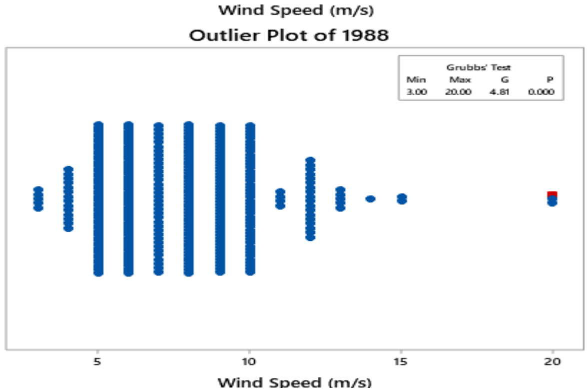

Time Series Plot of 1981
Wind Speed (m/s)
Day

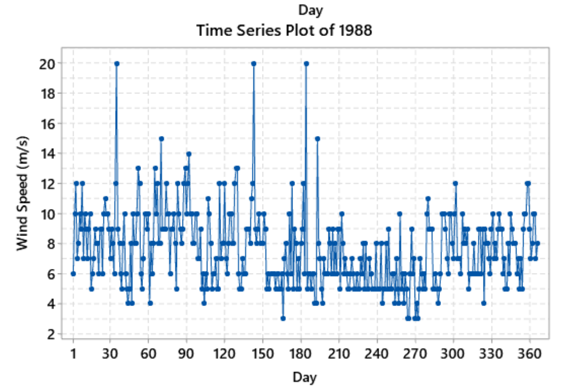

Outlier Plot of 1982
Grubbs' Test
Min Max G P
4.00 36.00 4.52 0.002
Wind Speed (m/s)

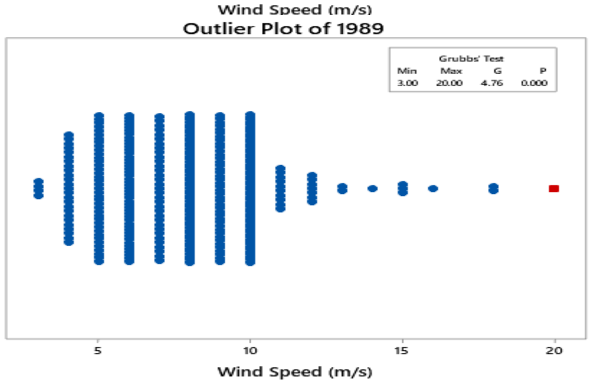

Time Series Plot of 1982
Wind Speed (m/s)
Day

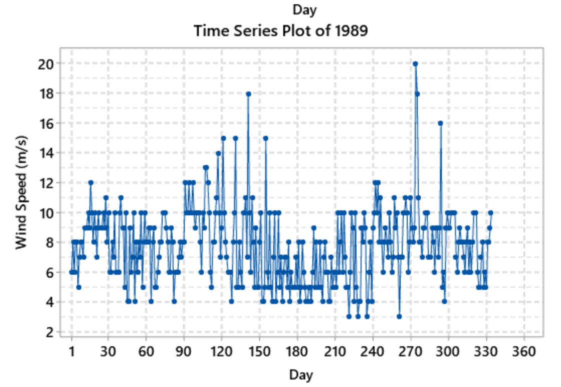

Outlier Plot of 1987
Grubbs' Test
Min Max G P
3.00 24.00 7.02 0.000
Wind Speed (m/s)

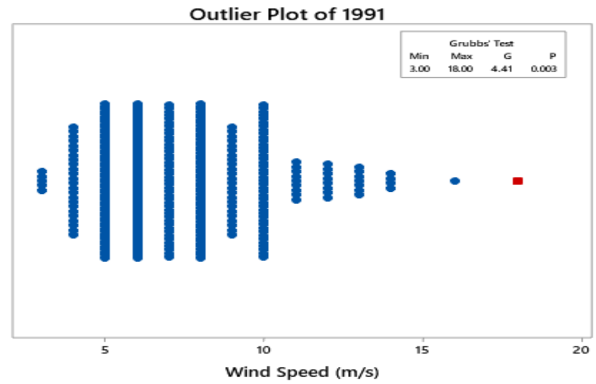

Time Series Plot of 1987
Wind Speed (m/s)
Day

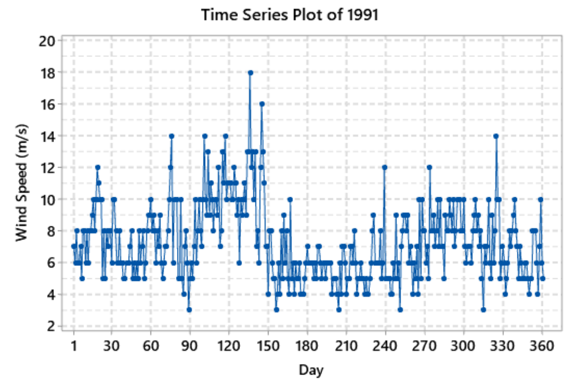

Outlier Plot of 1988
Grubbs' Test
Min Max G P
3.00 20.00 4.81 0.000
Wind Speed (m/s)

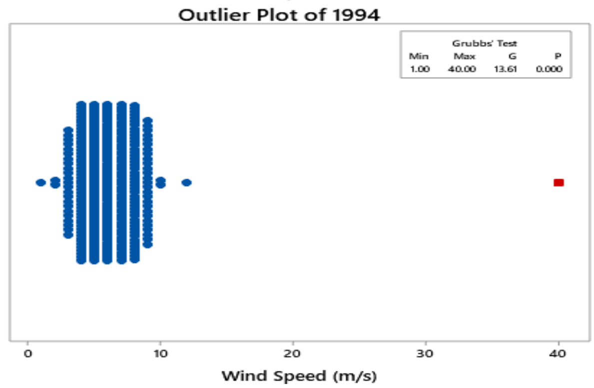

Time Series Plot of 1988
Wind Speed (m/s)
Day

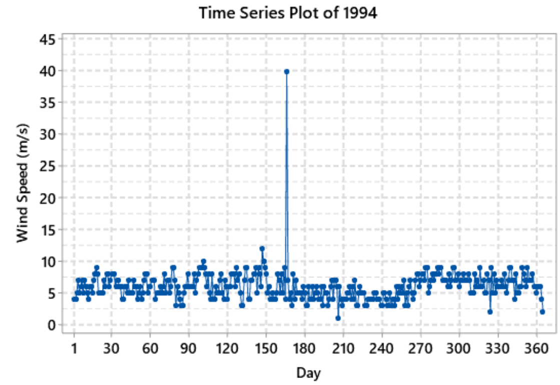

Outlier Plot of 1989
Grubbs' Test
Min Max G P
3.00 20.00 4.76 0.000
Wind Speed (m/s)

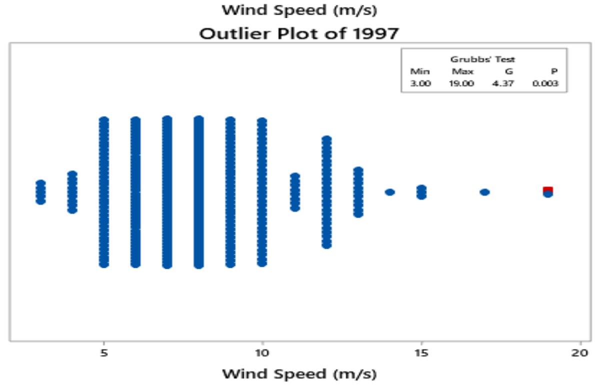

Time Series Plot of 1989
Wind Speed (m/s)
Day

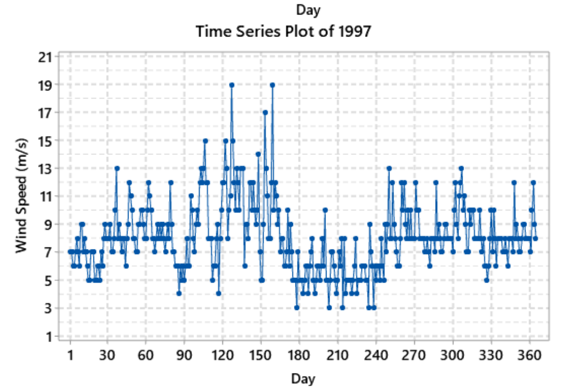

Outlier Plot of 1991
Grubbs' Test
Min Max G P
3.00 18.00 4.41 0.003
Wind Speed (m/s)

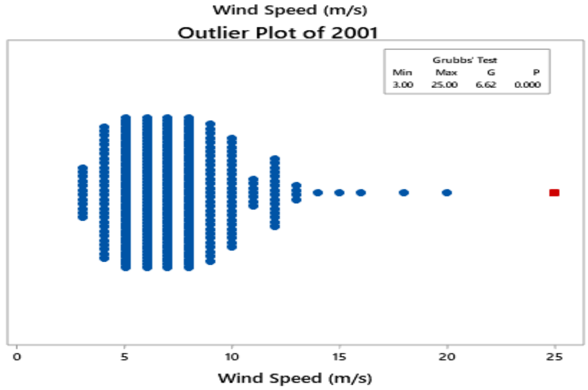

Time Series Plot of 1991
Wind Speed (m/s)
Day

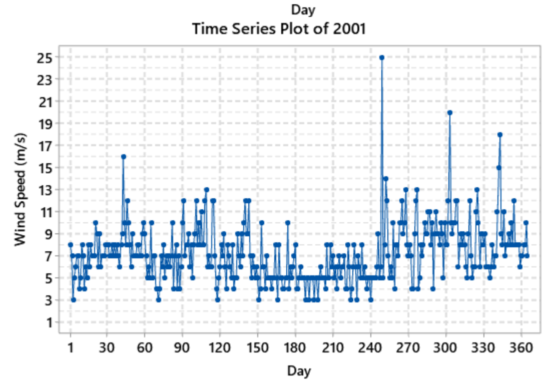

Outlier Plot of 1994
Grubbs' Test
Min Max G P
1.00 40.00 13.61 0.000
Wind Speed (m/s)

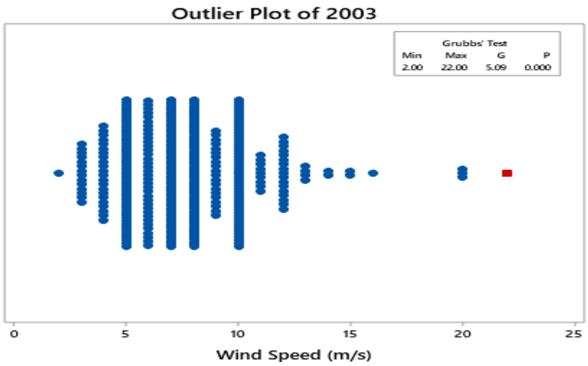

Time Series Plot of 1994
Wind Speed (m/s)
Day

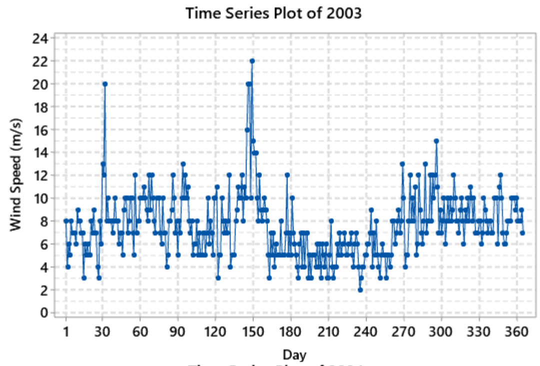

Outlier Plot of 1997
Grubbs' Test
Min Max G P
3.00 19.00 4.37 0.003
Wind Speed (m/s)

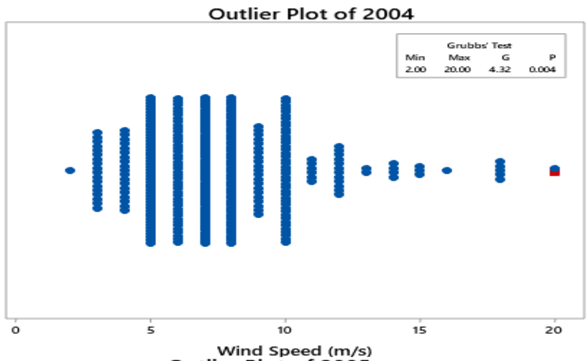

Time Series Plot of 1997
Wind Speed (m/s)
Day

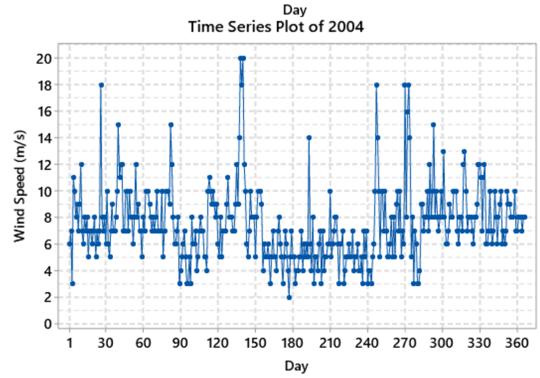

Outlier Plot of 2001
Grubbs' Test
Min Max G P
3.00 25.00 6.62 0.000
Wind Speed (m/s)

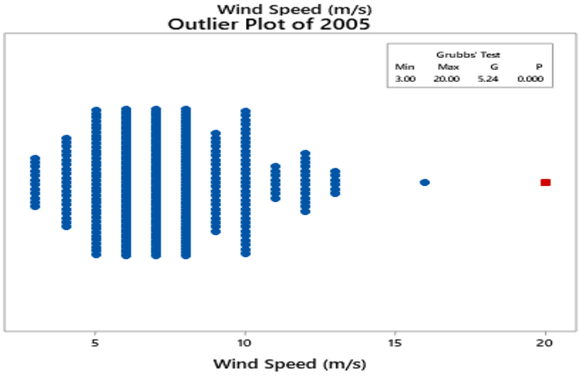

Time Series Plot of 2001
Wind Speed (m/s)
Day

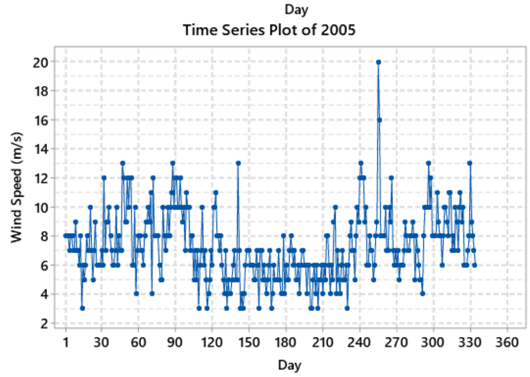

Outlier Plot of 2003
Grubbs' Test
Min Max G P
2.00 22.00 5.09 0.000
Wind Speed (m/s)

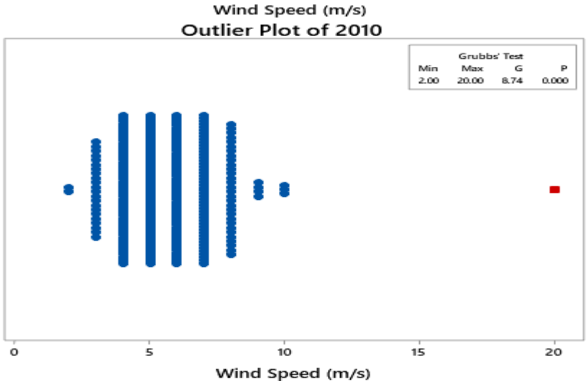

Time Series Plot of 2003
Wind Speed (m/s)
Day

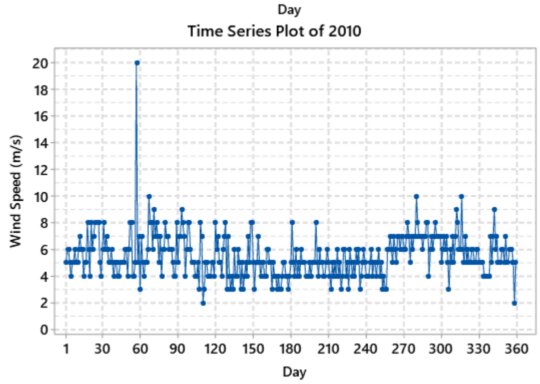

Outlier Plot of 2004
Grubbs' Test
Min Max G P
2.00 20.00 4.32 0.004
Wind Speed (m/s)


Time Series Plot of 2004
Wind Speed (m/s)
Day


Outlier Plot of 2005
Grubbs' Test
Min Max G P
3.00 20.00 5.24 0.000
Wind Speed (m/s)


Time Series Plot of 2005
Wind Speed (m/s)
Day


Outlier Plot of 2010
Grubbs' Test
Min Max G P
2.00 20.00 8.74 0.000
Wind Speed (m/s)


Time Series Plot of 2010
Wind Speed (m/s)
Day

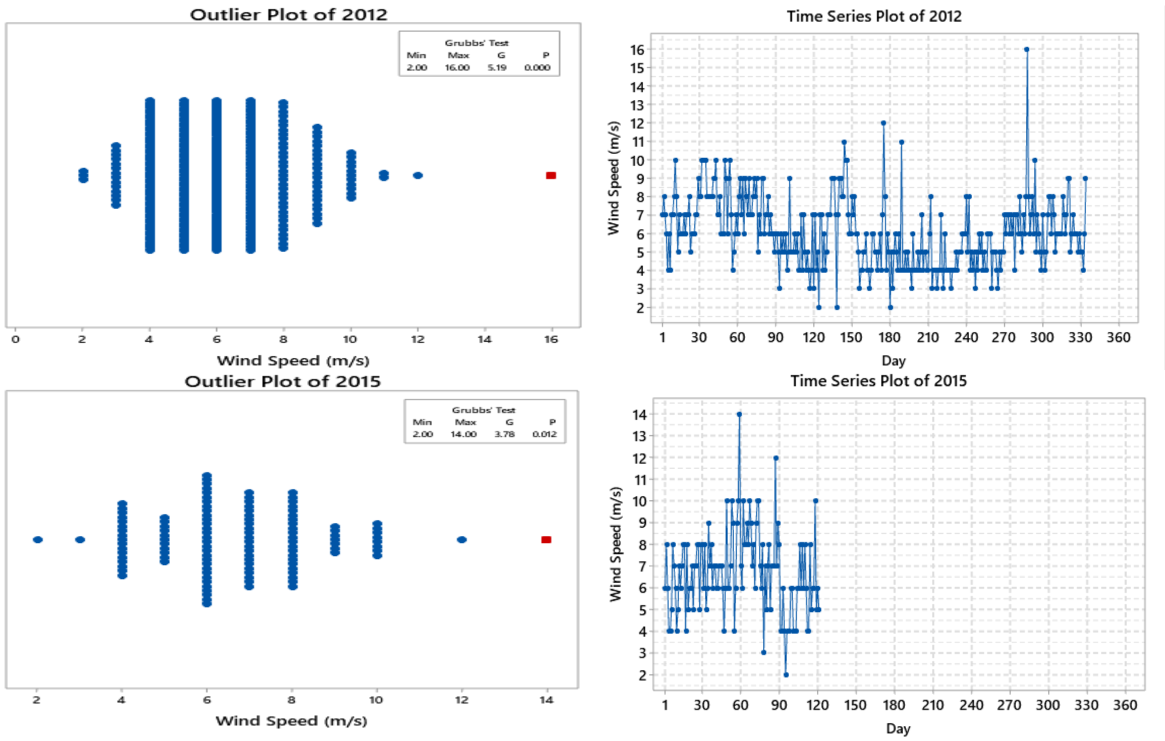


Figure C-5-1. Graphical display of all outliers detected from the data.

Appendix D

Maximum Likelihood Estimators (MLEs) of some distributions

Three-parameter Weibull distribution (WEI3)

The WEI3 CDF & PDF are described as [51]:

$$F(x) = 1 - exp\left[-\left(\frac{x_i-\mu}{\alpha}\right)^{\beta}\right]; x_i \leq \mu; \quad \text{(D-1)}$$

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x_i-\mu}{\alpha}\right)^{\beta-1} exp\left[-\left(\frac{x_i-\mu}{\alpha}\right)^{\beta}\right]; x_i < \mu \quad \text{(D-2)}$$

Where, α, β and μ are defined as scale parameter, shape parameter and location parameter, respectively.

The Maximum Likelihood Estimators (MLEs) of WEI3 can be determined by solving simultaneously the following equations.

$$\frac{n}{\beta} + \sum_{i=1}^{n} ln\left(\frac{x_i-\mu}{\alpha}\right) - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\alpha}\right)^{\beta} ln\left(\frac{x_i-\mu}{\alpha}\right) = 0 \quad \text{(D-3a)}$$

$$-\frac{n\beta}{\alpha} + \frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\alpha}\right)^{\beta} = 0 \Rightarrow \alpha = \left[\frac{\sum_{i=1}^{n}(x_i-\mu)^{\beta}}{n}\right]^{\frac{1}{\beta}} \quad \text{(D-3b)}$$

$$-(\beta-1)\sum_{i=1}^{n}(x_i-\mu)^{-1} + \frac{\beta}{\alpha}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\alpha}\right)^{\beta-1} = 0 \quad \text{(D-3c)}$$

Substituting (D-3b) into (D-3a) and into (D-3c) and rearranging, we can arrive at:

$$\frac{\beta-1}{\beta}\sum_{i=1}^{n}(x_i-\mu)^{-1} - n\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^{\beta-1}}{\sum_{i=1}^{n}(x_i-\mu)^{\beta}}\right) = 0 \quad \text{(D-4)}$$

$$\frac{1}{\beta} + \frac{1}{n}\sum_{i=1}^{n} ln(x_i-\mu) - \left(\frac{\sum_{i=1}^{n}(x_i-\mu)^{\beta} ln(x_i-\mu)}{\sum_{i=1}^{n}(x_i-\mu)^{\beta}}\right) = 0 \quad \text{(D-5)}$$

These equations must be solved numerically to determine the parameters. Fortunately, there are many softwares available which can solve these nonlinear equations, such as "*WeibullR*" R package.

Generalized Extreme Value Distributions (GEV)

The generalized extreme value distributions combine the Gumbel, Fréchet, and Weibull distributions into one single family of models. The CDF & PDF of the GEV are of the form [45]:

$$F(x) = exp\left\{-\left[1-\beta\left(\frac{x_i-\mu}{\alpha}\right)\right]^{\frac{1}{\beta}}\right\} \quad \text{(D-6)}$$

$$f(x) = \frac{1}{\alpha}\left[1-\beta\left(\frac{x_i-\mu}{\alpha}\right)\right]^{\frac{1}{\beta}-1} exp-\left[1-\beta\left(\frac{x_i-\mu}{\alpha}\right)\right]^{1/\beta} \quad \text{(D-7)}$$

Where, μ, α, and β are location, scale and shape parameters, respectively.

Gumbel distribution is obtained when β is zero, Fréchet distribution is obtained when β is negative, and Weibull distribution is obtained when β is positive.

The MLEs of GEV are obtained by solving:

$$\frac{1}{\alpha}\sum_{i=1}^{n}\left[\frac{1-\beta-\left(1-\frac{\beta}{\alpha}(x_i-\mu)\right)^{1/\beta}}{1-\frac{\beta}{\alpha}(x_i-\mu)}\right] = 0 \quad \text{(D-8)}$$

$$-\frac{n}{\alpha}+\frac{1}{\alpha}\sum_{i=1}^{n}\left[\frac{(x_i-\mu)}{\alpha}\left(\frac{1-\beta-\left(1-\frac{\beta}{\alpha}(x_i-\mu)\right)^{1/\beta}}{1-\frac{\beta}{\alpha}(x_i-\mu)}\right)\right] = 0 \quad \text{(D-9)}$$

$$-\frac{1}{\beta^2}\sum_{i=1}^{n}\left\{ln\left(1-\frac{\beta}{\alpha}(x_i-\mu)\right)\left[1-\beta-\left(1-\frac{\beta}{\alpha}(x_i-\mu)\right)^{\frac{1}{\beta}}\right]+\left(\frac{\beta}{\alpha}(x_i-\mu)\right)\frac{1-\beta-\left(1-\frac{\beta}{\alpha}(x_i-\mu)\right)^{1/\beta}}{1-\frac{\beta}{\alpha}(x_i-\mu)}\right\} = 0 \quad \text{(D-10)}$$

Again here, closed form of equations (D-8) to (D-10) are not available. However, these equations can easily be solved with the help of different R packages such as, "$envstats$" R package.

Generalized Pareto Distribution (GPD)

GPD can be expressed as [45]:

$$F(x_i) = 1 - \left(1 - \beta\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{\beta}} \quad \text{(D-11)}$$

$$f(x_i) = \frac{1}{\alpha}\left(1 - \beta\left(\frac{x_i-\mu}{\alpha}\right)\right)^{\frac{1}{\beta}-1} \quad \text{(D-12)}$$

Where, x_i is the wind speed, μ is a location parameter, α is a scale parameter, β is a shape parameter, $F(x_i)$ is the CDF, and $f(x_i)$ is the PDF.

A maximum likelihood estimators for GPD can be obtained by solving (see [45]):

$$\mu = Lower\ bound\ of the\ data \quad \text{(D-13a)}$$

$$\frac{1}{\alpha}\left(\frac{1}{k} - 1\right)\sum_{i=1}^{n}\frac{1}{\left(1-k\frac{x_i-\mu}{\alpha}\right)} - \frac{n}{\alpha k} = 0 \quad \text{(D-14)}$$

$$\frac{n}{k}\left(\frac{1}{k} - 1\right) - \frac{1}{k^2}\sum_{i=1}^{n} ln\left(1 - k\frac{x_i-\mu}{\alpha}\right) - \frac{1}{k}\left(\frac{1}{k} - 1\right)\sum_{i=1}^{n}\frac{1}{\left(1-k\frac{x_i-\mu}{\alpha}\right)} = 0 \quad \text{(D-15)}$$

Equations D-13b and D-13c can be numerically solved to obtain the values of α and β using various R packages, such as, "*ismev*", "*evir*", "*POT*".

Johnson Bounded (JSB) distribution

The CDF and PDF of the JohnsonSB distribution are given by:

$$F(x_i) = \Phi\left(\gamma + \delta\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right)\right) \quad \text{(D-16)}$$

$$f(x_i) = \frac{\delta\lambda}{\sqrt{2\pi}(x_i-\zeta)(\zeta+\eta-x_i)} exp\left[-\frac{1}{2}\left(\gamma + \delta\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right)\right)^2\right] \quad \text{(D-17)}$$

Where, γ, δ, η and ζ are the four parameters; *exp* is the exponential function; and *ln* represents the natural logarithmic function.

The MLEs of JSB can be obtained by the solving the following system of equations (detail derivations are given in [52] :

$$\gamma + \frac{\delta}{n}\sum_{i=1}^{n} ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right) = 0 \qquad \text{(D-16a)}$$

$$-\frac{n}{\delta} + \gamma\sum_{i=1}^{n} ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right) + \delta\sum_{i=1}^{n} ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right)^2 = 0 \qquad \text{(D-16b)}$$

$$(\lambda + 2\zeta + \gamma\delta\eta)\sum_{i=1}^{n}(x_i-\zeta)\,(\zeta+\eta-x_i)^{-1} - 2\sum_{i=1}^{n} x_i[(x_i-\zeta)(\zeta+\eta-x_i)^{-1}] +$$

$$\delta^2\eta\sum_{i=1}^{n}(x_i-\zeta)(\zeta+\eta-x_i)^{-1}\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right) = 0 \qquad \text{(D-16c)}$$

$$n + \eta(\gamma\delta - 1)\sum_{i=1}^{n}(\zeta+\eta-x_i)^{-1} + \delta^2\eta\sum_{i=1}^{n}(\zeta+\eta-x_i)^{-1}\, ln\left(\frac{x_i-\zeta}{\zeta+\eta-x_i}\right) = 0 \qquad \text{(D-16d)}$$

In order to obtain the MLEs of JSB, equations (D-16a) to (D-16d) must be solved numerically "$ExtDist$" R package can do the job.