

UNIVERSITY OF GHANA



AN EXTREME VALUE ANALYSIS OF THE SEA LEVEL AT AXIM

BY

ENOCK OPOKU

(10244275)

**THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF MPhil
STATISTICS DEGREE**



JULY 2019

DECLARATION

Candidate's Declaration

I, Enock Opoku hereby declare that this thesis is my original work except for references to the works of other people, which I have acknowledged duly. I also declare that this work has not been submitted either wholly or in part to any university for an academic degree.

Signature.....

Date:

ENOCK OPOKU

(10244275)

Certified by:

Signature.....

Date:

DR RICHARD MINKAH

(Principal Supervisor)

Signature.....

Date:

DR LOUIS ASIEDU

(Co-Supervisor)

ABSTRACT

Assessing the probability of the sea level rising because of heavy rains and tidal waves is an important issue to engineers and coastal development planners. In the case of Axim, the sea level rise leads to flooding, loss of lives and the destruction of properties in communities such as Brawie, upper and lower Axim. This has led to the commencement of a sea defence project in Axim to protect the community from floods and further destruction of properties. The data used in this study were hourly sea level data from the Axim sea spanning the period January 1980 to January 2019. In this study, we used Extreme Value Theory (EVT) to estimate the exceedance probabilities and return periods of high sea levels that can result in flooding and its associated effects. The study began with an assessment of the domain of attraction of the tails of the Axim sea level data. The Generalised Pareto distribution (GPD) was used to fit the excess data above a chosen high threshold. The Probability Weighted Moment (PWM) and Maximum Likelihood (ML) methods were used to estimate the shape parameter γ of the GPD. The study showed that the tail distribution of the Axim sea level data is in the Weibull domain of attraction, that is, it has a negative shape parameter ($\gamma < 0$). Also, the study revealed that the probability of the sea exceeding the maximum observed data of 1.83 meters (above mean sea level) is 0.0031. Finally, the study concluded that based on theory and data at hand, there is a negligible chance of the Axim sea rising above 2 meters.

DEDICATION

I dedicate the work to God Almighty, whose guidance and protection have taken me this far.

To my parents and siblings who have been my source of inspiration.

Finally, this work is dedicated to all my lecturers and friends who enriched my knowledge.

ACKNOWLEDGEMENT

I thank the Almighty God for the divine guidance throughout my period of study and for the successful completion of this work.

My profound gratitude goes to Dr. Richard Minkah and Dr. Louis Asiedu, my principal supervisor and co-supervisor respectively for their immense support and guidance throughout the duration of this work. I say thank you to them and may God bless them abundantly.

My next sincere gratitude goes to Professor Kwasi Appeaning-Addo of the Institute for Environment and Sanitation Studies (IESS) and Mr. Philip-Neri Jayson-Quashigah, PhD student at the Marine Science department for providing me with the data needed for this work.

I extend this gratitude to Nana Otuo Acheampong for his support during my MPhil studies.

I am extremely grateful to my family whose prayers and support have taken me this far.

Finally, I am grateful to all my friends who in one way or the other helped in the successful completion of this work.

LIST OF TABLES

Table 4.1: Summary statistics of Axim Sea level	35
Table 4.3: Simulated critical points for the test statistic G_m^* for choice of GPd models	40
Table 4.4: Marohn (2000) One-Sided Test Result.....	41
Table 4.5: Final ML estimates for GPd parameters	42
Table 4.6: Likelihood Ratio Test	42
Table 4.7: Simulated critical values of the Kolmogorov-Smirnov statistic adapted to the Exponential distribution with unknown parameters	44
Table 4.8: Cramer-Von Mises and Anderson-Darling Test Values.....	45
Table 4.9: Simulated critical values of the Cramer-von Mises (normal style) and Anderson-Darling (bold) statistics adapted to the GPd with unknown parameters	45
Table 4.10: Final ML and PWM estimates for GPd parameters	47
Table 4.11: Estimated Exceedance probabilities for selected sea levels	51
Table 4.12: Return Period estimate for GPd fitted Model	51

LIST OF FIGURES

Figure 4.1: Histogram showing distribution of sea level data above 1.6 meters.....	35
Figure 4.2: Mean Excess plot showing selected threshold value of 1.77 meters	36
Figure 4.3: Exponential QQ plot for dataset	37
Figure 4.4: Exponential QQ plot for excesses above threshold	38
Figure 4.5: Diagnostic plots for GPd fit using ML estimates.....	47
Figure 4.6: Diagnostic plots for GPd fit using PWM estimates.	48

TABLE OF CONTENTS	
DECLARATION.....	ii
DEDICATION.....	iv
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
CHAPTER ONE.....	1
1.1 BACKGROUND OF STUDY	1
1.2 PROBLEM STATEMENT.....	2
1.3 OBJECTIVES OF THE STUDY.....	3
1.4 SIGNIFICANCE OF STUDY.....	4
1.5 SCOPE AND CONTRIBUTION OF STUDY.....	4
1.6 ORGANISATION OF STUDY	4
1.7 LIMITATIONS OF THE STUDY	5
CHAPTER TWO.....	6
2.1 DEVELOPMENT OF EXTREME VALUE THEORY (EVT)	6
2.2 FIELDS OF APPLICATIONS OF EVT.....	7
2.2.1 APPLICATIONS IN HYDROLOGY.....	8
2.2.2 APPLICATION OF EVT IN OTHER FIELDS.....	9
2.3 MODELS UNDER EXTREME VALUE THEORY	11
CHAPTER THREE.....	14
EXTREME VALUE ANALYSIS.....	14
3.1 DISTRIBUTION OF THE SAMPLE MAXIMA.....	14
3.2 EXTREMAL TYPES THEOREM.....	15
3.3. THE PEAKS OVER THRESHOLD (POT) METHOD.....	18
3.3.1 MAXIMUM LIKELIHOOD ESTIMATION.....	22
3.3.2. PROBABILITY WEIGHTED MOMENTS ESTIMATION	23
3.3.3 ESTIMATION OF OTHER PARAMETERS OF EXTREME EVENTS	25
3.3.4 CONFIDENCE INTERVALS FOR THE EXTREME VALUE INDEX.....	26
3.3.5 STATISTICAL CHOICE OF GPD MODELS	27
3.3.5.1 GOMES AND VAN MONFORT (1986) TEST	28
3.3.5.2 MAROHN (2000) GPD TEST.....	29
3.3.5.3 LIKELIHOOD RATIO TEST.....	30
3.3.6 THE CHOICE OF THE THRESHOLD	31

3.3.7 GOODNESS OF FIT TEST	32
3.3.7.1 KOLMOGOROV-SMIRNOV	32
3.3.7.2 CRAMER-VON MISES AND ANDERSON-DARLING TESTS	33
CHAPTER FOUR	34
DATA ANALYSIS	34
4.1 PRELIMINARY STATISTICAL ANALYSIS	34
4.1.1 DATA	34
4.1.2 DESCRIPTIVE STATISTICS	34
4.1.3 MEAN EXCESS PLOT	36
4.1.4 EXPONENTIAL QQ PLOT	37
4.2 TEST TO DETERMINE DOMAIN-OF ATTRACTION	38
4.2.1 GOMES AND VAN MONFORT (1986) TEST	39
4.2.2 MAROHN (2000) TEST	41
4.2.3 THE LIKELIHOOD RATIO TEST (LRT)	41
4.2.4 GOODNESS-OF-FIT TEST	43
4.2.4.1 KOLMOGOROV-SMIRNOV TEST	43
4.2.4.2 THE "ANDERSON-DARLING AND CRAMER-VON MISES" TESTS	44
4.3: FITTING THE GPD MODEL AND ESTIMATION OF EXTREME EVENTS	46
4.3.1 ESTIMATION OF PARAMETERS	46
4.3.2 DIAGNOSTIC PLOTS	47
4.3.3 CONFIDENCE INTERVAL ESTIMATION	49
4.3.4 ESTIMATION OF UPPER ENDPOINT	50
4.3.5: EXCEEDANCE PROBABILITY	51
4.3.6: RETURN LEVEL ESTIMATE	51
CHAPTER FIVE	52
CONCLUSIONS AND RECOMMENDATIONS	52
5.1 CONCLUSIONS	52
5.2 RECOMMENDATIONS	55
REFERENCES	56
APPENDICES	66

CHAPTER ONE

Climate change and its related sea-level rise are expected to significantly affect susceptible coastal regions. In many parts of the world, floods, as a result of sea-level rise, massive rainstorm and tidal waves have resulted in the loss of lives and destruction of properties. Estimation of probable maximum flood discharges due to sea-level rise with a specific return period is essential for the design of hydraulic structures.

Since the focus of this thesis is to model sea level, a statistical and theoretical framework that deals with sampling extremes need to be employed, hence the use of Extreme Value Theory.

1.1 BACKGROUND OF STUDY

The global average sea level between 1961 and 2003 rose at an average rate of around 1.8 mm/year compared to 1993 to 2003 average rate of about 3.1 mm per annum (IPCC, 2007). Global temperature as a result of climate change contributed highly to the rise in the level of the sea. The rise in the sea-level is expected to be dominant for centuries, and this will probably result in flooding of coastal areas (Kebede et al., 2012). Areas to be flooded and the rate of sea erosion is dependent on the beach topography, the kind of geological material, and human factors (Appeaning-Addo, 2013).

According to Boko et al. (2007), an estimated 40% of the population in West Africa is in the coastal region and expected to grow to say 50 million in 2020. This coastal erosion associated disasters along the coast significantly affect the economy of the coastal regions and their ecosystems.

In the work of Peeler (2007) and Wax (2007), Nigeria's sea level is expected to rise to 0.2m compared 0.4m of Bay of Bengal leaving 700,000 persons and about 10 million displaced

respectively in Nigeria and Bangladesh. According to Appeaning-Addo (2009), Ghana's sea level is rising at a rate of approximately 2 mm per annum, which corresponds to the global estimate.

According to Olsthoorn et al. (2008), the fragmentation of the West-Antarctic Ice Sheet (WAIS) is as a result of climate change, which contributed estimated to contribute a 5m rise in sea level (Lythe et al., 2001). This disintegration of the WAIS is the result of the severe displacement of communities along the coastal regions and eventually contributing to severe economic and ecological damage especially around the West African coastal communities (Boko et al., 2007).

This thesis focuses on sea-level rise and its associated effect on the coastal communities in Axim. More specifically, the focus will be on the use of EVT to assess the possibility of the sea rising above a selected level and the probability of it leading to flooding in surrounding communities.

1.2 PROBLEM STATEMENT

Flooding has become a global concern Ghana is not an exception. In the coastal areas of Ghana, flooding is largely attributed to sea level rise. The continuous level of sea level rise in the coastal areas of Ghana has prompted local authorities to put sea defensive measures in place to curb the situation. Sea defensive walls were erected; however, strong tidal waves continue to cause flooding. By May 2016, strong tidal waves had for the fourth time in the year displaced hundreds of residents and destroyed valuable properties.

These happenings are common to the people of Axim in the Western region of Ghana. Notable landmarks in the area have been completely washed away by the sea. The situation which has led to families seeking for shelter elsewhere is a source of worry to the indigenes.

The sea is severely cutting into the Axim Township and washing away valuable properties which may lead to the collapse of the town which serves as a tourist site. Tidal waves continue to submerge coastal communities leading to the destruction of properties worth thousands of Ghana Cedis due to rising sea levels. Due to increased global warming with an associated increase in sea level, communities along the shorelines of West African coast like Ghana is expected to experience perennial flooding and eventual submerging of some communities and cities.

Assessing the probability of a very high sea level when it comes to the construction of any coastal defence project (Sea wall, revetment etc.) is an important issue that cannot be ignored. When it comes to the case of the Axim Sea Defence wall, very high sea level above the maximum height of the sea defence wall will lead to flooding and destruction of properties. Though flooding along the coast is caused by other factors aside high sea level, this thesis will look at the effect of rising sea level on the Axim community. Taking all these into consideration, EVT offers a firm mathematical basis to analyse extreme cases, in this case, high sea level.

In this study, EVT is used to analyse the effect of very high sea levels on the Axim community.

1.3 OBJECTIVES OF THE STUDY

The study seeks to measure the effect of very high sea level on the Axim community.

Specifically, focus of this thesis is to use EVT to analyse the Axim sea data to determine:

1. whether the sea level can rise above 2 meters.
2. The underlying distribution of the Axim sea data
3. the probability of the sea exceeding a given height in meters
4. the 100-year return level of the sea.

1.4 SIGNIFICANCE OF STUDY

This study will enable engineers and coastal development planners know how high a proposed coastal defence wall to be constructed to prevent flooding in communities along the coast. The study will also contribute to the existing literature on the applications of EVT on sea levels.

1.5 SCOPE AND CONTRIBUTION OF STUDY

The study was conducted on the Axim sea with emphasis on the hourly sea height above mean sea level at any point in time. The data used in the study consisted of hourly sea level of the Axim sea between the periods January 1980 to January 2019.

1.6 ORGANISATION OF STUDY

This study comprises of five chapters. The rest of the chapters are organised as follows: Chapter 2 looks at some related literature in different dimensions. The third chapter of the study looks at the parametric approach to estimating parameters in the Generalized Pareto (GP) distributions.

The chapter also looks at the methods to estimating other extreme events. In chapter 4, the Axim sea level data is analysed by fitting a GP distribution to the excesses to answer the objectives to the study. Estimation of other extreme events are also performed.

In chapter 5, results from chapter 4 are summarized and conclusions drawn. Recommendations for future studies are also presented.

1.7 LIMITATIONS OF THE STUDY

The data used for the study did not enable us to consider other factors that may lead to flooding in the Axim municipality. Factors such as wind waves, shoreline elevation of the Axim beach, gravitational force of attraction between the moon and the sea etc. The study only considered the Axim sea level.

Also, all efforts to get information on the new defence wall being constructed in Axim proved futile. The engineers were not ready to provide information that would have helped the researcher predict the number of years it will take for the sea to rise above the maximum height of the seawall.

CHAPTER TWO

LITERATURE REVIEW

This section examines both theoretical and empirical literature on Extreme Value Theory (EVT). The chapter consists of a careful review of the Development of Extreme Value Theory (EVT) since its inception. This chapter also sheds light on some of the applications of EVT. Finally, the chapter ends by looking at some models as developed by pioneers in the field.

2.1 DEVELOPMENT OF EXTREME VALUE THEORY (EVT)

EVT is a branch of statistics that deals with techniques for modelling and estimation of rare events. Unlike most of traditional statistical analysis that deal with the centre of the of the underlying distribution, EVT enables us to restrict attention to the behaviour of the tails of the distribution function. Engineers, hydrologists and theoretical probability were the first group of people to develop interest in the EVT.

Work on the development of the Extreme Value framework may be dated back to the year 1709 when Nicolas Bernoulli discussed the mean largest distance from the origin when n points lie at random on a straight line of length t (Johnson et al., 1995). According to Kinnison (1985), Fourier while working on the normal distribution a century later also reported that a probability of 1 in 50,000 is assigned to an observation that is more than three times the square root of two standard deviations from the mean.

Later in 1922, Bortkiewicz (1922) was the literature that triggered the systematic development of the EVT. The paper was on the distribution range in random sample coming from a normal

distribution. According to Kotz and Saralees (2000), this paper clearly introduced the notion of distribution of large numbers for the first time.

Misses (1923) and Dodd (1923) all continued with concepts relating to the law of large numbers.

The breakthrough occurred in 1928 when the literature Fisher and Tippet (1928) was published.

Beirlant et al., (2004) indicated that this important outcome realised by Fisher and Tippet in 1928 on the possible limit laws of the sample maximum brought to the fore that EVT was different from the central limit theorem. Kotz and Saralees (2000) revealed that in 1943, Gnedenko presented a solid foundation for EVT and came up with necessary and sufficient conditions for the weak convergence of extreme order statistics. Kotz and Saralees (2000) further stated that Gnedenko (1943) consolidated and established the ideas into the basic assumption in EVT known as the Extreme Value condition.

According to Beirlant et al., (2004), a doctoral dissertation presented by de Haan in 1970 on the Regular Variation and its application to the weak convergence of sample extremes has also contributed immensely to the theoretical development of EVT.

We will explore further this result and its applications in the next chapter.

2.2 FIELDS OF APPLICATIONS OF EVT

The statistical analysis of extreme values is employed in numerous disciplines; including hydrology, meteorology, engineering, finance, economics, reinsurance, and telecommunication, sport science, environmental problems, demography, Oceanography, etc.

In this section, the researcher looks at the application of EVT in hydrology and other fields

2.2.1 APPLICATIONS IN HYDROLOGY

In Hydrology EVT is applied in areas such as high sea levels, dam bursts and flood threats. Unusually strong wind, high waves and the excessive river levels can be predicted by an environmentalist with the help of EVT. All these scenarios are extreme cases that put human life at risk.

Annual maximum sea levels data recorded at Port Pirie from 1923 to 1987 was analysed by Coles (2001) using EVT. EVT was also applied to similar data collected from Fremantle by Coles (2001). The hydrologists, pioneers in the application of EVT are mostly interested in estimating the level that will be exceeded once on average in t years known as the t -years flood discharge. Hydrologists prefer the Generalised Extreme Value (GEV) distribution whenever the data to be used for statistical analysis is on annual maximum discharges. On the other hand, the Generalised Pareto distribution (GPD) is used whenever the analysis is based on data that exceeds a chosen high threshold.

Extensive work has been done in the field of hydrology with abundant literature to select from whenever a researcher decides to work in the field. The GPD model was used to analyse detailed data coming from the river Nidd in England, looking at seasonality and serial dependence of data. Hosking et al. (1985) and Hosking & Wallis (1987) also applied the GEV model to the same river Nidd, but this time on 35 annual maxima floods data. Dekkers & de Haan (1989) also used the idea of exceedance probability and return period of floods recommended that the height of sea defences be predicted correctly with the chance of the sea dike being exceeded is small and predefined.

Bivariate EVT was employed by de Haan & de Ronder (1998) to model wind and sea data in the Netherlands. Barão & Tawn (1999) used the same technique on sea level data coming from two

east coast sites in the United Kingdom. Tawn (1992) on his own employed EVT in analysing hourly sea level data. In the end, he concluded that because of an astronomical tidal component, the series could not be considered as a stationary sequence.

Extensive literature on the application of EVT to rainfall data can be found in Coles & Tawn (1996), when they worked on extremes of spatially aggregated rainfall over fixed durations. Smith (1989) and Küchenhoff & Thamerus (1996) all used EVT on ground-level ozone data.

The applications of EVT highlighted above shows that hydrologists were really the pioneers in the application of EVT to data. In the next section, the researcher brings to light other areas where the EVT has already been successfully implemented.

2.2.2 APPLICATION OF EVT IN OTHER FIELDS.

Sports was not left out of the numerous fields EVT is being applied. EVT is gaining popularity in sports science where a research was conducted by Coles (2001) spanning the period 1972 to 1992 to determine for the women's 1500-meter race, the fastest race time for every year. On swimming, EVT models were applied by Adam & Tawn (2012) to the Olympic 400 meters freestyle data for the period 1924 to 2004 to determine the time of a gold medallist. Strand & Boes (1998) as well as Barão & Tawn (1999) all applied EVT on athletics data.

Insurance is another field where EVT has been applied extensively. This is not surprising since insurance companies deal with chance events. An example is when McNeil & Saladin (1997) applied the EVT on the Danish data to assess fire insurance losses. Rootzen & Tajvidi (1997) with the help of the EVT assessed the nature of the Swedish windstorm insurance claims. The reader is

referred to Beirlant et al. (1994) and Mikosch (1997) as some studies that explains into details how EVT is applied in the field of insurance.

Computation of tail risk measures with their associated confidence intervals was done by Gilli & Kellezi (2006) using EVT. They also applied the EVT to numerous major indices on the stock market. Since the EVT is used to study rare events, it is applied in the field of Risk Management. The importance of EVT to the risk management officer is highlighted and explained by Danielsson and de Vries (1997), McNeil (1998 and 1999), Embrechts et al. (1998, 1999) and Embrechts (1999).

In finance, Gencay & Faruk (2004), Jockovic (2012) and Marimoutou et al. (2009) employed EVT in one way or the other to estimate and forecast using the Value-at-Risk (VaR). Gencay & Faruk (2004) also opined that estimates for the VaR computed using EVT are more accurate at higher quantiles. Human longevity is also estimated with the help of EVT models. EVT models were applied by Watts et al. (2013) to determine the upper tail of the distribution of human lifespans. Blanchet et al. (2009) also applied EVT to snowfall data.

Aside the numerous areas of application of EVT highlighted by the researcher, other areas include engineering for strength of materials (Harter, 1978); Extreme risk in futures of contracts traded Cotter (2005); earthquake size distribution (Kagan, 1997); returns of electricity demand in the UK (Chan and Nadarajah, 2015); city-sizes, corrosion analysis, exploitation of diamond deposits, demography, geology and meteorology among others.

2.3 MODELS UNDER EXTREME VALUE THEORY

The Classical extreme value theory assumes independence and constant distribution through time.

The Classical theorem applies only to sequences of independent and identically distributed (iid) random variables, however, it also holds when the hypotheses are relaxed moderately; for instance, when there exists a relatively weak statistical dependence between the random variables (Méndez & Menéndez, 2006). According to Méndez & Menéndez (2006), the theorem can be employed in practice, if every year can be conceived to be composed of many small “sampling” intervals (e.g. hours), such that successive values of the process are appropriately identically distributed and show “relatively weak” statistical dependence.

Hawkes et al. (2008) stated that, standard methodologies employed in extreme value analysis should consist of adopting asymptotic models to describe the stochastic variations of the process. D’Onofrio et al. (1999) employed the Generalized Extreme Value (GEV) distribution to estimate the return periods of extreme water levels. Haung et al. (2008) also employed the GEV model to analyze annual maximum water levels on the coast of the United States.

What all these studies have in common is that, all the authors assumed independence and constant distribution through time, but in the context of environmental processes, the assumption of constant distribution through time restrict the findings in the form of the data used in the study. The moment there is a change in the characteristics of the data over time; such findings will no longer be valid (Coles, 2001). Environmental processes are usually non-stationary; non-stationary processes have characteristics that change systematically through time. Non-stationarity is often apparent in environmental processes because of seasonal effects, perhaps due to different climate patterns in different months or in a form of trends, possibly due to long-term climate changes (Coles, 2001). In modeling extreme non-stationary processes, the general assumptions of

independence and constant distribution through time cannot be applied (El Adlouni et al., 2007) For the non-stationary approach, the parameters of the distribution functions are replaced with time-dependent parameters, so that the results of the extreme value analysis also vary with time (Katz et al., 2002).

According to Mudersbach & Jensen (2010), the non-stationary problem can be solved by de-trending the data, to apply the classical extreme value analysis, however, the non-stationary approach has the benefit of enabling the extrapolation of the results up to future time horizons.

Northrop & Jonathan (2010) mentioned that, in the application of EVT to environmental data, it is common for extremes of variables of interest to be non-stationary, which changes systematically in space, time or with the values of covariates.

Coles (2001) studied the annual maximum sea-level data at Fremantle where the location parameter was modeled as a linear time-dependent model.

Mudersbach & Jensen (2010) applied the non-stationary statistical model to annual maximum water levels from 1849 - 2007 at the German North Sea gauge at Cuxhaven using linear and exponential time models for the location and scale parameters of the GEV. The results were compared with stationary methods and it was found that, the non-stationary GEV approach is suitable for determining coastal design water levels.

Katz et al. (2002) employed a non-stationary GEV model and recommended a linear model for the location and a log-transformed model for the scale parameter but with fixed shape parameter.

Hundecha et al. (2008) also implemented a non-stationary extreme value analysis to estimate quantiles of extreme wind speed and their changes against time. It was applied to 10-m wind speed

data from the North American Regional Reanalysis (NARR) data set and data from some selected stations of environment of Canada.

Zhang et al. (2004) compared different methods of detecting trends in extreme values and concluded that methods based on modeling trends in the parameter of the distribution of the extremes are powerful methods of detecting statistically significant trends in the extremes (as cited by Hudecha et al., 2008). Trambly et al. (2013) analyzed heavy rainfall events from the Southern France using both the Non-stationary model and the classical stationary model and concluded that the non-stationary model is better.

Watts et al. (2013) employed the non-stationary Generalized Extreme Value and Generalized Pareto distributions in modelling life span data.

Bezak et al. (2014) compared the Peak-Over-Threshold method and the Annual maxima method and concluded that, the Peak-Over-Threshold method yields better results than the Annual maxima method.

CHAPTER THREE

EXTREME VALUE ANALYSIS

A theoretical scheme used in analysing the nature of sample extremes, for instance, the sample minimum or maximum is known as the Extreme Value Theory (EVT). The exact or limiting distribution function (d.f) of these order statistics are used to judge its behaviour. In this chapter, the researcher brings to light the behaviour of the sample maximum as opined by the pioneers in the field.

3.1 DISTRIBUTION OF THE SAMPLE MAXIMA

Consider an independent and identically distributed (iid) samples X_1, X_2, \dots, X_n . Define a new random variable as

$$M_n = X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$$

Then the explicit distribution of the sample maxima (M_n) derived all $X \in \mathfrak{R}$ as

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) = \prod_{j=1}^n P(X_j \leq x) = \mathcal{F}^n(x)$$

Theorem 3.1: Let \mathcal{F} be the underlying d.f. of a sequence of r.v.'s and $x^{\mathcal{F}}$ its right endpoint, i.e., $x^{\mathcal{F}} = \sup\{x : \mathcal{F}(x) < 1\}$, which may be infinite. Then

$$M_n \xrightarrow{p} x^{\mathcal{F}}$$

as $n \rightarrow \infty$, where \xrightarrow{p} means convergence in probability.

Therefore, M_n possesses a degenerate distribution that is asymptotic. In order to make deductions however, there is the need to obtain a non-degenerate reaction for M_n . Like the central limit theorem, some form of standardization is needed. The central limit theorem is based on the nature of the asymptotic reaction of partial sums

$$X_1, X_1 + X_2, \dots, \sum_{i=1}^n X_{i,\dots}, \text{ as } n \rightarrow \infty$$

Theorem 3.2

Considering a sequence of i.i.d. r.v.'s, X_1, X_2, \dots, X_n with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$.

Therefore, as $n \rightarrow \infty$,

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{1})$$

In order to obtain a suitable limiting distribution that is non-degenerate for the sample maxima, an identical theorem is needed: that is, there is the need for standardisation sequences $a_n > 0$ and b_n real so that as $n \rightarrow \infty$

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \tag{3.1}$$

with G non-degenerate.

3.2 EXTREMAL TYPES THEOREM

All the possible limiting distribution functions that may appear on (3.1) are called the extreme value distributions. The difficulty here is how to determine the kind of d.f. G will represent. The following concepts will help us determine the kind of d.f. G will represent.

Definition 3.1 (maximum domain of attraction)

A d.f. \mathcal{F} belongs to the maximum domain of attraction of G , if there are sequences $\{a_n > 0\}$ and $\{b_n\}$ real such that

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) = \lim_{n \rightarrow \infty} \mathcal{F}^n(a_n x + b_n) = G(x) \quad (3.2)$$

For every stability point x of G and can be written as $\mathcal{F} \in \mathcal{D}(G)$.

Definition 3.2 (distribution functions of the same type)

Two d.f.'s \mathcal{F}_1 and \mathcal{F}_2 are said to be of the same type if there exist constants $a > 0$ and $b \in \mathfrak{R}$ such that

$$\mathcal{F}_2(ax + b) = \mathcal{F}_1(x) \quad (3.3)$$

It means that \mathcal{F}_1 and \mathcal{F}_2 are equal, except for the scale and location parameters.

Fisher & Tippett (1928) were able to deal with difficulty associated with determining the extreme value distributions, which was finalised by Gnedenko (1943) and later finetuned by de Haan (1970). These pioneers in the field of EVT showed that if (3.2) stands, one out of just three types will represent the limiting distribution of G . This is stated formally in the theorem below:

Theorem 3.3 (Fisher and Tippett, 1928, Gnedenko, 1943)

If $\mathcal{F} \in \mathcal{D}(G)$, the limiting d.f. G of the sample maximum, satisfactorily standardized, will be similar in type as one of the distributions below:

$$(I) : G_{\alpha}(x) = \exp\left(-\exp\left(-\frac{x-b}{a}\right)\right), \quad x \in \mathfrak{R}$$

$$(II) : G_{\alpha}(x) = \begin{cases} 0, & \text{if } x \leq b, \\ \exp\left(-\left(\frac{x-b}{a}\right)^{-\alpha}\right), & \text{if } x > b, \quad \alpha > 0. \end{cases}$$

$$(III) : G_{\alpha}(x) = \begin{cases} \exp\left(-\left(-\frac{x-b}{a}\right)^{\alpha}\right), & \text{if } x < b, \quad \alpha > 0 \\ 1, & \text{if } x \geq b. \end{cases}$$

for all $a > 0$ and $b \in \mathfrak{R}$.

where the shape parameter α of (II) and (III) defines the tail's nature of the fundamental d.f. \mathcal{F} .

The class of the limiting distributions (I), (II) and (III) are known as the Gumbel, Fréchet and Weibull types, respectively.

Jenkinson (1955) attained a depiction for the three types and called it the Generalised Extreme Value (GEV) distribution given by

$$G_{\gamma}(x) = \begin{cases} \exp\left(-\left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right), & 1 + \gamma \frac{x-\mu}{\sigma} > 0, \quad \gamma \neq 0 \\ \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), & x \in \mathfrak{R}, \quad \gamma = 0 \end{cases} \quad (3.4)$$

where γ is often referred to as the tail index or the extreme value index (EVI).

- i. For $\gamma = 0$, G_γ and (I) are of the same type.
- ii. For $\gamma > 0$ and taking $\gamma = \frac{1}{\alpha}$, G_γ and (II) are of the same type.
- iii. For $\gamma < 0$ and taking $\gamma = -\frac{1}{\alpha}$, G_γ and (III) are of the same type.

Also,

(i) For $\gamma = 0$, the distribution function belongs to the Gumbel. These are class of distributions that possess exponential-like tails. Examples include the Gaussian, Gamma, Lognormal, Exponential etc.

(ii) For positive γ , the distribution function resides in the Fréchet domain. They possess tails that decay polynomially and have infinite upper endpoint.

(iii) For negative γ , the distribution function belongs to the Weibull domain. They consist of light-tailed distributions with finite upper endpoint.

Extensive work has been done by pioneers in the field of EVT on the block maxima approach of fitting a GEVD to a set of data. The reader is referred to Coles, 2001; Embrechts et al., 1997; de Haan and Ferreira, 2006 and Beirlant et al., 2004.

3.3. THE PEAKS OVER THRESHOLD (POT) METHOD

In choosing an appropriate observation within the EVT structure, two basic approaches exist.

These are the Block Maxima (Gumbel's method) and the Peaks-Over-Threshold (POT) approaches. The Block Maxima approach selects the maximum observed value within consecutive blocks, where the length of each block is predefined whereas the POT approach concentrates on observed values that are more than a high deterministic threshold.

The POT approach concentrates on observed values that are more than a high deterministic threshold and fitting a suitable parametric model to the excesses above that threshold. The whole focus is then shifted to the observations that are above the fixed threshold, making sure that there are enough data above this threshold. Since the focus is on a fraction of the sample that exceed the threshold, there must exist an acceptable conditional distribution for the excesses above the threshold. The Maximum Likelihood (ML) or the Probability Weighted Moment (PWM) method is used when to estimate parameters when fitting such a distribution to data under this approach.

Let (X_1, \dots, X_n) be a random sample having $X \sim \mathcal{F}$ and high fixed threshold u , less than the right endpoint of the support \mathcal{F} . If we let

$$N_u = \{i : X_i > u, i = 1, \dots, n\}$$

Represent the number of observations among (X_1, \dots, X_n) exceeding the threshold u . The random variable N_u is binomially distributed, that is $N_u \sim \mathcal{B}(n, 1 - \mathcal{F}(u))$. The exceedances over the fixed threshold u defined by

$$\{W_i\}_{i=1}^{N_u} = \{X_i : X_i > u, i = 1, \dots, N_u\} \quad (3.5)$$

If an observation is more than the selected threshold u , then an exceedance has occurred. Hence, the exceedances are denoted by the random variable (r.v) $X|X > u$. The conditional d.f. of $X|X > u$ is represented as

$$\mathcal{F}_{X|X>u}(x) = P(X \leq x|X > u) = \frac{\mathcal{F}(x) - \mathcal{F}(u)}{\bar{\mathcal{F}}(u)}, \text{ for } x \geq u. \quad (3.6)$$

For a sufficiently high threshold u , theorem 3.5 showed that under standard conditions, the conditional distribution above may be approximated by the Generalized Pareto distribution. That is, they showed that the GPD is the limiting distribution of properly normalized exceedances.

Definition 3.4 (Generalized Pareto distribution)

The Generalized Pareto distribution is characterised by

$$H_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, & 1 + \gamma x > 0, x \geq 0, \\ 1 - \exp(-x), & x \geq 0, \end{cases} \quad \begin{array}{l} \text{if } \gamma \neq 0 \\ \text{if } \gamma = 0 \end{array} \quad (3.7)$$

Theorem 3.4

$\mathcal{F} \in \mathcal{D}(G_\gamma)$ if and only if $\mathcal{F}_{X|X>u}(x) \simeq H_\gamma(x|u, \sigma_u)$, where

$$H_\gamma(x|u, \sigma_u) = \begin{cases} 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)^{-\frac{1}{\gamma}}, & 1 + \gamma \frac{x-u}{\sigma_u} > 0, x \geq u, \\ 1 - \exp\left(-\frac{x-u}{\sigma_u}\right), & x \geq u, \end{cases} \quad \begin{array}{l} \text{if } \gamma \neq 0 \\ \text{if } \gamma = 0 \end{array} \quad (3.8)$$

The results by the Pickands-Balkema-de Haan Theorem (Balkema and de Haan, 1974 and Pickands III, 1975) can be summarised as below:

Theorem 3.5 (Pickands-Balkema-de Haan Theorem) $\mathcal{F} \in \mathcal{D}(G_\gamma)$ if and only if

$$\lim_{u \rightarrow x^{\mathcal{F}}} |\mathcal{F}_{X|X>u}(x) - H_\gamma(x|u, \sigma_u)| = 0, \quad (3.9)$$

for some GPD with shape, location and scale parameters γ, u and $\sigma_u > 0$ respectively (Vicente, 2012). If (3.9) holds, we say that \mathcal{F} belongs to the POT-domain of attraction of the GPD, H_γ (Vicente, 2012).

It is possible to work with the excesses instead of the exceedances. We can denote the excesses by the r.v. $Y = X - u$. Hence, theorem 3.4 and 3.5 can be rewritten as:

Theorem 3.6

$\mathcal{F} \in \mathcal{D}(G_\gamma)$ if and only if $\mathcal{F}_{Y|Y>0}(y) \simeq H_\gamma(y|0, \sigma_u)$, where

$$H_\gamma(y|0, \sigma_u) = \begin{cases} 1 - \left(1 + \gamma \frac{y}{\sigma_u}\right)^{-\frac{1}{\gamma}}, & 1 + \gamma \frac{y}{\sigma_u} > 0, y \geq 0, & \text{if } \gamma \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & y \geq 0, & \text{if } \gamma = 0 \end{cases} \quad (3.10)$$

Theorem 3.7 (Pickands-Balkema-de Haan Theorem)

$\mathcal{F} \in \mathcal{D}(G_\gamma)$ if and only if

$$\lim_{u \rightarrow x^{\mathcal{F}}} |\mathcal{F}_{Y|Y>0}(y) - H_\gamma(y|0, \sigma_u)| = 0,$$

for some GPD with shape and scale parameters γ and $\sigma_u > 0$, respectively (Vicente, 2012). In this case, the GPD is the limit distribution of the scaled excesses (Vicente, 2012).

3.3.1 MAXIMUM LIKELIHOOD ESTIMATION

Consider the original random sample (X_1, \dots, X_n) of a random variable (r.v.) X , with $X \sim \mathcal{F}$. Given a threshold value u , and letting m represent the number of exceedances of the original sample.

The m excesses obtained can be represented by

$$Y_j = X_i - u | X_i > u, \text{ for } i = 1, \dots, n \text{ and } j = 1, \dots, m.$$

For $\gamma \neq 0$, a given random sample (y_1, \dots, y_m) of the r.v Y with GPD has a log-likelihood function given by

$$\ell(\gamma, \sigma_u | y_1, \dots, y_m) = -m \log \sigma_u - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log \left(1 + \frac{\gamma y_i}{\sigma_u}\right), \quad (3.11)$$

where $1 + \frac{\gamma y_i}{\sigma_u} > 0$, $j = 1, \dots, m$.

For $\gamma = 0$, the log-likelihood function becomes:

$$\ell(0, \sigma_u | y_1, \dots, y_m) = -m \log \sigma_u - \frac{1}{\sigma_u} \sum_{i=1}^m y_i$$

Most at times, we favour a reparameterization of the log-likelihood function in (3.11) for computational purposes. Defining $\tau = \frac{\gamma}{\sigma_u}$, the log-likelihood function may be rewritten as

$$\ell(\gamma, \sigma_u | y_1, \dots, y_m) = -m \log \left(\frac{\gamma}{\tau}\right) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log \left(1 + \frac{\gamma y_i}{\sigma_u}\right),$$

$$\ell(\gamma, \tau | y_1, \dots, y_m) = -m(\log \gamma - \log \tau) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log(1 + \tau y_i),$$

$$\ell(\gamma, \tau | y_1, \dots, y_m) = -m \log \gamma + m \log \tau - \left(\frac{1}{\gamma} + 1 \right) \sum_{i=1}^m \log(1 + \tau y_i),$$

where $1 + \tau y_i > 0$, for $i = 1, \dots, m$.

The ML estimator $(\hat{\gamma}, \hat{\tau})$ of the parameters (γ, τ) follows then from

$$\frac{1}{\hat{\tau}} - \left(\frac{1}{\hat{\gamma}} + 1 \right) \frac{1}{m} \sum_{i=1}^m \frac{y_i}{1 + \hat{\tau} y_i} = 0, \quad (3.12)$$

where $\hat{\gamma} = \frac{1}{m} \sum_{i=1}^m \log(1 + \hat{\tau} y_i)$

The main reason for the introduction of this reparameterization by Davison (1984) was to get $\hat{\gamma}$ explicitly expressed as a function of $\hat{\tau}$. This can be attained through (3.12) but after replacing

$\hat{\gamma}$ with $\frac{1}{m} \sum_{i=1}^m \log(1 + \hat{\tau} y_i)$. When $\gamma = 0$, we have the exponential distribution, yielding $\hat{\sigma}_u = \bar{Y}$.

3.3.2. PROBABILITY WEIGHTED MOMENTS ESTIMATION

Despite the flexibility and easy adaptability to models, the ML method is less efficient for small samples. To address this drawback, Hosking and Wallis (1987) proposed the use of the PWM estimator. The PWM of a r.v. X with d.f. \mathcal{F} are given by the quantities

$$M_{p,r,s} = E\{X^p (F(X))^r (1 - F(X))^s\},$$

for $p, r, s \in \mathbb{R}$.

Using the GPD and considering $M_{p,r,s}$ with $p = 1, r = 0$ and $s = 0, 1, \dots$, yields

$$M_{1,0,s} = \frac{\sigma_u}{(s+1)(s+1-\gamma)}, \text{ for } \gamma < 1. \quad (3.13)$$

As for the case of fitting a GEVD, in the Block Maxima method, we can replace $M_{1,0,s}$ by its empirical counterpart

$$\widehat{M}_{1,0,s} = \frac{1}{m} \sum_{i=1}^m \left(\prod_{j=1}^s \frac{m-i-j+1}{m-j} \right) Y_{i:m}$$

and solving (3.13), for $s = 0, 1$ with respect to γ and σ and, we obtain the PWM estimators

$$\widehat{\gamma} = 2 - \frac{\widehat{M}_{1,0,0}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}}$$

and

$$\widehat{\sigma}_u = \frac{2\widehat{M}_{1,0,0}\widehat{M}_{1,0,1}}{\widehat{M}_{1,0,0} - 2\widehat{M}_{1,0,1}}$$

Note that, in the GPD case, the $r - th$ moments only exist for $\gamma < 1/r$.

As for the ML estimation, the PWM estimators are asymptotically Normal distributed. Considering the details, we refer to Hosking and Wallis (1987). They showed that, for the GPD with shape parameter in the range $0 \leq \gamma \leq 0.4$ and specially for small samples, the PWM estimators perform better than the ML estimators, since they exhibit small dispersion. The difference is less pronounced as the sample size increases. They also noted that the traditional Method of Moments is preferable when $\gamma < 0$. Nevertheless, the PWM estimation has some problems: on one hand, for $\gamma \geq 1$, PWM estimators do not exist and on the other hand, we can obtain estimates that can be inconsistent with the data, in the sense that some of the observations may fall above the estimate of the right endpoint, $x^{\mathcal{F}}$.

3.3.3 ESTIMATION OF OTHER PARAMETERS OF EXTREME EVENTS

Defining, once again, y_{1-p} as the extreme quantile of order $(1 - p)$ of the GPD underlying the excesses Y , with p sufficiently small, we can obtain estimates of extreme quantiles inverting the GPD given by (3.10), yielding

$$y_{1-p} = H_\gamma^{-1}(1 - p|0, \sigma_u) = \begin{cases} \frac{\sigma_u}{\gamma}(p^{-\gamma} - 1), & \text{if } \gamma \neq 0, \\ -\sigma_u \log p, & \text{if } \gamma = 0, \end{cases} \quad (3.14)$$

and replacing (γ, σ_u) by its ML or PWM estimator.

If $\gamma < 0$, the right endpoint of the GPD is finite and is given by

$$y_1 = -\frac{\sigma_u}{\gamma}$$

which can be estimated replacing again (γ, σ_u) by its ML or PWM estimator.

One key point is that: provided that, under the parametric methodology, $\mathcal{F}_{Y|Y>0}(y) \simeq H_\gamma(y|0, \sigma_u)$, the quantiles estimates obtained from (3.14) are the estimated quantiles of the d.f. $\mathcal{F}_{Y|Y>0}(y)$. However, if we want to estimate the extreme quantiles of the original and unknown d.f. \mathcal{F} , associated with the r.v. X , we can use the identity (3.6). If $Y = X - u$, we have

$$\mathcal{F}_{Y|Y>0}(y) = P(Y \leq y|Y > 0)$$

which evaluates to

$$\hat{\mathcal{F}}(x) = \frac{m}{n} \left(1 - H_{\hat{\gamma}}(x - u|0, \hat{\sigma}_u) \right). \quad (3.15)$$

where $\frac{m}{n}$ is the sample frequency of observations that exceed the threshold u in the original sample (X_1, \dots, X_n) and $H_{\hat{\gamma}}$ is obtained by replacing the GPD parameters with their ML or PWM estimators.

Defining \mathcal{X}_{1-p} as the extreme quantile of order $(1-p)$ of the d.f. \mathcal{F} underlying the r.v. X , i.e., a quantity such that $\mathcal{F}(\mathcal{X}_{1-p}) = 1-p$, with p sufficiently small, we can estimate these quantiles, for instance for $\gamma \neq 0$, using (3.15):

$$\hat{\mathcal{X}}_{1-p} = \hat{U}\left(\frac{1}{p}\right) = \hat{\mathcal{F}}^{-1}(1-p) = u + \frac{\hat{\sigma}_u}{\hat{\gamma}} \left[\left(\frac{np}{m}\right)^{-\hat{\gamma}} - 1 \right],$$

and, for $\gamma < 0$, an estimate for the right endpoint of the d.f. \mathcal{F} is given by

$$\hat{x}^{\mathcal{F}} = \hat{U}(\infty) = u - \frac{\hat{\sigma}_u}{\gamma}, \quad (3.16)$$

3.3.4 CONFIDENCE INTERVALS FOR THE EXTREME VALUE INDEX

A better alternative approach to the ‘‘delta method’’ for constructing confidence interval is the profile likelihood proposed by (Beirlant et al. 2004). Therefore, the $100(1-\alpha)\%$ CI for γ is

$$CI_{\gamma} = \left\{ \gamma : \log \mathcal{L}_p(\gamma) \geq \log \mathcal{L}_p(\hat{\gamma}) - \frac{\chi_{(1)}^2(1-\alpha)}{2} \right\}. \quad (3.17)$$

A similar approach is employed to construct the CI for other parameters of the model.

3.3.5 STATISTICAL CHOICE OF GPD MODELS

In the POT method, the following hypotheses test is mainly considered, for the same reasons seen in the case of the GEVD fitting:

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma \neq 0 .$$

The exponential distribution function is prioritised in this test for modelling the excesses above a significantly large threshold. The test procedures for this test can be found in literature such as Gomes & Van Monfort (1986), Reiss & Thomas (2007), Marohn (2000), and Vicente (2012). The problem of goodness-of-fit for the GPD model has been studied by Choulakian and Stephens (2001). Lilliefors (1969) presented the special case of the Kolmogorov-Smirnov test, applied to the Exponential distribution with unknown parameters.

Since we will be interested in determining the domain of attraction of the underlying distribution function, we discuss three of the tests that will enable us to achieve our purpose. We will be interested in tests (3.18) and (3.19).

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma \neq 0 \tag{3.18}$$

and

$$H_0 : \gamma = 0 \text{ vs } H_1 : \gamma < 0 \tag{3.19}$$

These statistical tests will be performed with the exceedances obtained from the available data since we are in the POT context.

If we let (W_1, \dots, W_m) be the m exceedances over a non-random threshold u , as defined in (3.5), extracted from the available random sample (X_1, \dots, X_n) .

3.3.5.1 GOMES AND VAN MONFORT (1986) TEST

The first test we will look at is the Gomes and van Monfort test. The test statistic used to test (3.19) is given by

$$G_m = \frac{W_{m:m}}{W_{[m/2]+1:m}} \quad (3.20)$$

Under the validity of H_0 , we have

$$G_m^* = \log 2 G_m - \log m \xrightarrow{d} \Lambda$$

as $m \rightarrow \infty$. H_0 is rejected at the asymptotic level α if $G_m^* \leq \mathcal{G}_\alpha$, where \mathcal{G}_α represents the standard Gumbel ε – quantile.

The associated p – value based on the decision rule can be obtained by

$$p(G_m^*) = \Lambda(G_m^*).$$

3.3.5.2 MAROHN (2000) GPD TEST

Mahron (2010) proposed a GPD test procedure based on the sample coefficient of variation for testing (3.18) and (3.19)

The test statistics is given by

$$T_m = \frac{1}{2} \left(\frac{S_W^2}{(\bar{W} - u)^2} - 1 \right), \quad (3.21)$$

where $S_W^2 = \frac{1}{m} \sum_{i=1}^m (W_i - \bar{W})^2$ is the sample variance.

As $m \rightarrow \infty$, we have

$$T_m^* = \sqrt{m} T_m \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{1})$$

under H_0 . At an asymptotic level of α , H_0 is rejected if

$$|T_m^*| \geq z_{1-\alpha/2} \text{ or if } T_m^* \leq z_\alpha,$$

For the tests in (3.18) and (3.19) respectively. The p – values associated with the tests are obtained with

$$p(T_m^*) = 2 - 2\phi(|T_m^*|)$$

or

$$p(T_m^*) = \phi(T_m^*)$$

Using simulation studies, Mahron (2000) indicated that when working with small and moderate sample sizes, the test statistic for (3.18) is biased with very poor power. The result showed that the test leads to reasonable results only for large sample sizes, that is $m \geq 500$.

3.3.5.3 LIKELIHOOD RATIO TEST

If we let $\ell(\gamma, u, \sigma_u | w_1, \dots, w_m)$ be the respective unrestricted log-likelihood function, where

$\ell(0, u, \sigma_u | w_1, \dots, w_m)$ denotes the restricted log-likelihood function, which corresponds to the Exponential case. The likelihood ratio test statistic is given by

$$\mathbf{L} = -2 \left(\ell(0, u, \hat{\sigma}_{u, H_0} | w_1, \dots, w_m) - \ell(\hat{\gamma}_{H_\gamma}, u, \hat{\sigma}_{u, H_\gamma} | w_1, \dots, w_m) \right)$$

with $\hat{\sigma}_{u, H_0}$ and $(\hat{\gamma}_{H_\gamma}, \hat{\sigma}_{u, H_\gamma})$ denoting the ML estimators for H_0 and H_γ models respectively. As $m \rightarrow \infty$, we have

$$\mathbf{L} \xrightarrow{d} \chi_{(1)}^2$$

under H_0 . To achieve a higher accuracy in the χ^2 -approximation, Reiss and Thomas (2007) recommend the Bartlett correction, yielding the statistic

$$\mathbf{L}^* = \frac{\mathbf{L}}{1 + 4/m} \xrightarrow{d} \chi_{(1)}^2 \quad (3.22)$$

as $m \rightarrow \infty$. For the test in (3.18), at the asymptotic size α , H_0 is rejected if

$$\mathbf{L}^* \geq \chi_{1, 1-\alpha}^2$$

where $\chi_{1, p}^2$ stands for the $\chi_{(1)}^2$ p -quantile.

The corresponding p-value can be calculated as follows:

$$p(\mathbf{L}^*) = 1 - \chi_{(1)}^2(\mathbf{L}^*)$$

3.3.6 THE CHOICE OF THE THRESHOLD

The choice of the threshold u is still an unsolved problem and in the literature of the POT method, not so much attention has been given to this issue. The choice of the threshold is not straightforward; indeed, a compromise must be found between high values of u , where the bias of the estimators are smaller, and low values of u , where the variance is smaller.

Davison and Smith (1990) suggest the use of the Mean Excess function. In the GPD case, the mean excess function is given by:

$$e(u) = E(X - u | X > u) = E(Y | Y > 0) = \frac{\sigma_u + \gamma u}{1 - \gamma}, \quad \text{for } \gamma < 1. \quad (3.23)$$

If the GPD assumption is valid, the plot of $e(u)$ versus u , called Mean Excess plot (or shortly ME-plot), should follow a straight line. In practice, based on a sample of size n , (x_1, x_2, \dots, x_n) , $e(u)$ is estimated by its empirical counterpart, the sample Mean Excess function:

$$\hat{e}_n(u) = \frac{\sum_{i=1}^n x_i I_{]u, \infty[}(x_i)}{\sum_{i=1}^n I_{]u, \infty[}(x_i)} - u, \quad \text{where } I_{]u, \infty[}(x_i) = \begin{cases} 1, & \text{if } x_i \in]u, \infty[, \\ 0, & \text{if } x_i \in]-\infty, u[. \end{cases} \quad (3.24)$$

To view this function, we generally construct the sample ME-plot

$$\{(X_{n-k:n}, \hat{e}_n(X_{n-k:n})) : 1 \leq k \leq n - 1\},$$

where $X_{n-k:n}$ denotes the $(k + 1) - th$ largest observation and where $\hat{e}_n(u)$ may be rewritten as

$$\hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^k x_{n-j+1:n} - x_{n-k:n}. \quad (3.25)$$

If the data support a GPD over a high threshold, we would expect the sample ME-plot to become linear in view of (3.18). At least, this is the ideal situation. But even for data that are genuinely

GP-distributed the sample ME-plot is seldom perfectly linear, particularly toward the right-hand end, where we are averaging a small number of large excesses. In fact, we often omit the final few points from consideration, as they can severely distort the plot. Consequently, the threshold u is chosen at the point to the right of which a rough linear pattern appears in the plot.

Another procedure consists in choosing one of the sample points as a threshold, i.e. $u = X_{n-k:n}$ $k = 1, \dots, n - 1$. With such a random threshold, we work then with the $k + 1$ top order statistics associated to the whole sample of size n , $X_{n:n}, X_{n-1:n}, \dots, X_{n-k:n}$.

3.3.7 GOODNESS OF FIT TEST

These are tests performed to check the goodness of fit of a GPD model that has been adopted.

Three goodness of fit tests that will be considered are the Kolmogorov-Smirnov test, Cramer-von Mises test and the Anderson-Darling goodness-of-fit tests.

3.3.7.1 KOLMOGOROV-SMIRNOV

Lilliefors (1969) studied the Kolmogorov-Smirnov test in the context of the Exponential distribution with unknown parameters. Since the Exponential distribution is embodied in the GPD distribution when

$\gamma = 0$, we can use the procedures described in his work to check the goodness-of-fit of the Exponential distribution to the r.v. Y . The Kolmogorov-Smirnov statistic for the null hypothesis of Exponential model is given by

$$D_m = \max_{1 \leq i \leq m} \left(\left| 1 - \exp\left(-\frac{Y_{i:m}}{\hat{\sigma}_u}\right) - \frac{i}{m} \right|, \left| 1 - \exp\left(-\frac{Y_{i:m}}{\hat{\sigma}_u}\right) - \frac{i-1}{m} \right| \right), \quad (3.36)$$

where $\hat{\sigma}_u$ stands for the ML estimator of σ_u for the exponential model.

3.3.7.2 CRAMER-VON MISES AND ANDERSON-DARLING TESTS

The Cramer-von Mises and Anderson-Darling tests were treated extensively by Choulakian and Stephens (2001). The null hypothesis for these tests postulates a GPD with unknown parameters.

The two test statistics for the GPD are expressed below:

Cramer-von Mises

$$W_m^2 = \sum_{i=1}^m \left(H_{\hat{\gamma}} \left(Y_{i:m} | \hat{\sigma}_{u, H_{\gamma}} \right) - \frac{2i-1}{2m} \right)^2 + \frac{1}{12m}, \quad (3.37)$$

Anderson-Darling

$$A_m^2 = -m - \frac{1}{m} \sum_{i=1}^m \left\{ (2i-1) \log \left(H_{\hat{\gamma}} \left(Y_{i:m} | \hat{\sigma}_{u, H_{\gamma}} \right) \right) + (2m+1-2i) \log \left(1 - H_{\hat{\gamma}} \left(Y_{i:m} | \hat{\sigma}_{u, H_{\gamma}} \right) \right) \right\} \quad (3.38)$$

where $\hat{\gamma}, \hat{\sigma}_{u, H_{\gamma}}$ represent the ML estimators for the GPD, H_{γ} .

CHAPTER FOUR

DATA ANALYSIS

In this section, the data is analysed using EVT, that is the POT methodologies described in the preceding chapter to answer the research objectives in chapter 1. The rest of this chapter is divided into three sections. These are the preliminary statistical analysis section, the choice of extreme value model section and the parametric estimation of extreme events section.

4.1 PRELIMINARY STATISTICAL ANALYSIS

This section is made up of description of the data used in the analysis, summary statistics of the data and the exponential QQ plot that will give us an idea about the domain of attraction of the underlying distribution function. The mean excess plot will be used in selecting the threshold for the GPD.

4.1.1 DATA

The data used for the study were hourly sea level data from the Axim sea in the Nzema East municipality of the Western Region of Ghana, for the period spanning January 1980 to January 2019. Due to the nature of the research (interest in rising sea level), Sea heights exceeding 1.6 meters were used in the analysis. In all, 6643 data points was used.

4.1.2 DESCRIPTIVE STATISTICS

Table 4.1 shows the summary statistics of the Axim sea data. From the table we are see that the maximum sea level recorded over the 39-year period was 1.83 meters above mean sea level. The third quartile value of 1.7 meters means that approximately 25% of the data set fell above 1.7

meters. Also, a standard deviation value of 0.0482 means the variability within the dataset wasn't that much. The other descriptive statistics are reported in table 4.1.

Table 4.1: Summary statistics of Axim Sea level

Statistic	Value
Minimum	1.6100
1st quartile	1.6300
Median	1.6500
3rd quartile	1.7000
Mean	1.6659
Maximum	1.8300
Standard deviation	0.0482
Skewness	0.8832

Figure 4.1 below gives a pictorial display about the distribution of the Axim sea level data. The diagram shows that the data is positively skewed and not normally distributed. The value for skewness from Table 4.1 confirms this assertion. This means the data is fit for extreme value analysis.

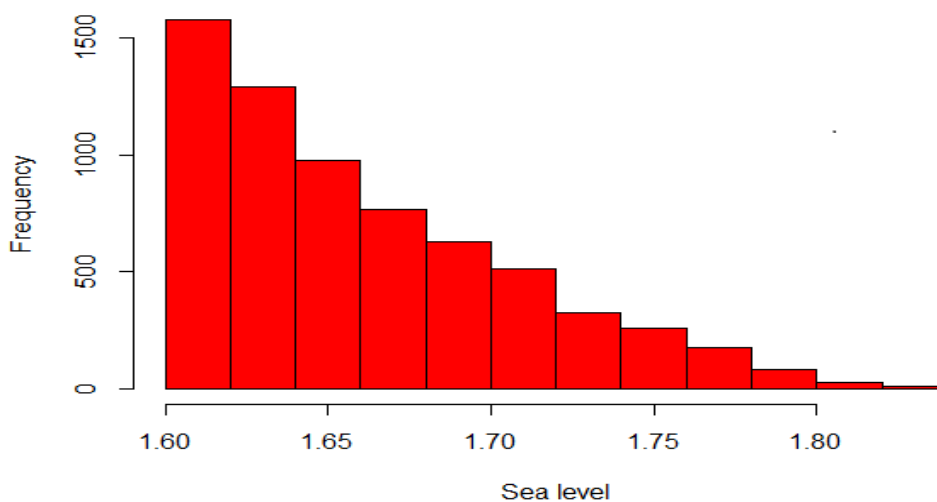


Figure 4.1: Histogram showing distribution of sea level data above 1.6 meters

4.1.3 MEAN EXCESS PLOT

Since we are using the POT method, there is the need to select a threshold above which we fit the GPD distribution to the excesses. In using the mean excess plot for threshold selection, the plot needs to be linear above the threshold. However, in selecting your threshold when fitting a GPD, there need to be compromise between precision and bias. Low threshold allows for more data (low variance) whereas high threshold allows for small data (low bias). From the mean excess plot, a threshold value of 1.77 meters is selected. The selection of 1.77 meters as the threshold leads to 215 exceedances, representing a proportion of 0.0324 (3.24%).

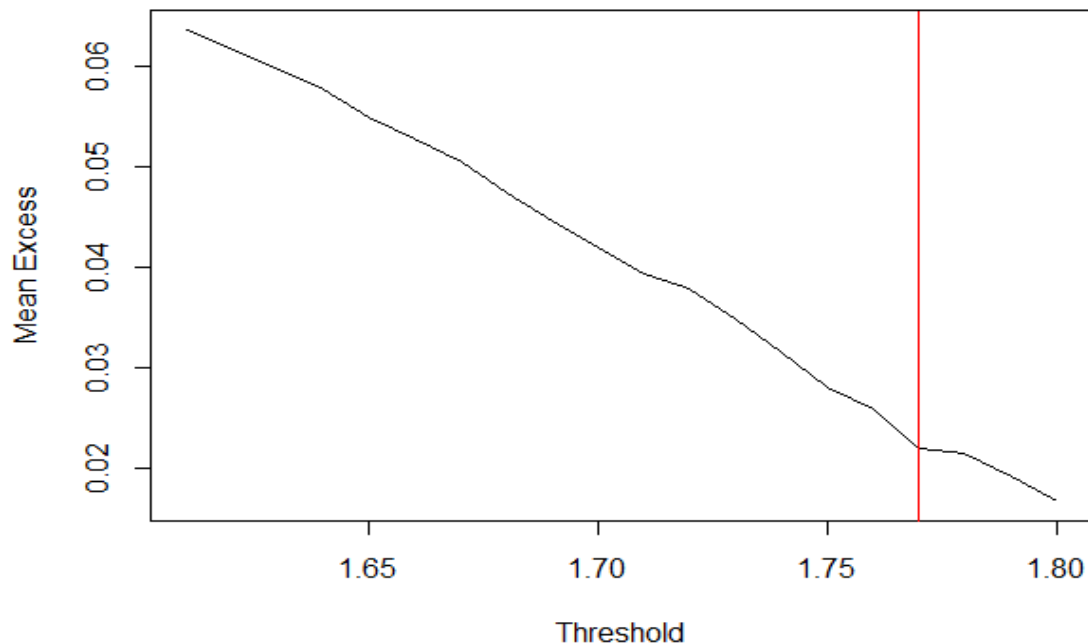


Figure 4.2: Mean Excess plot showing selected threshold value of 1.77 meters

4.1.4 EXPONENTIAL QQ PLOT

Determining the domain of attraction of the excess above the chosen threshold is of great importance in EVT modelling. Since we are using the GPD, we know from the preceding sections that the GPD becomes an Exponential distribution whenever $\gamma = 0$. As such, the researcher fit an exponential model to the excess data above the selected threshold of 1.77 meters. Through this, we can determine the extent to which a GPD can be used to model these excesses for $\gamma = 0$. The exponential QQ plot will be used to determine how good the exponential model fit the excess data above our selected threshold.

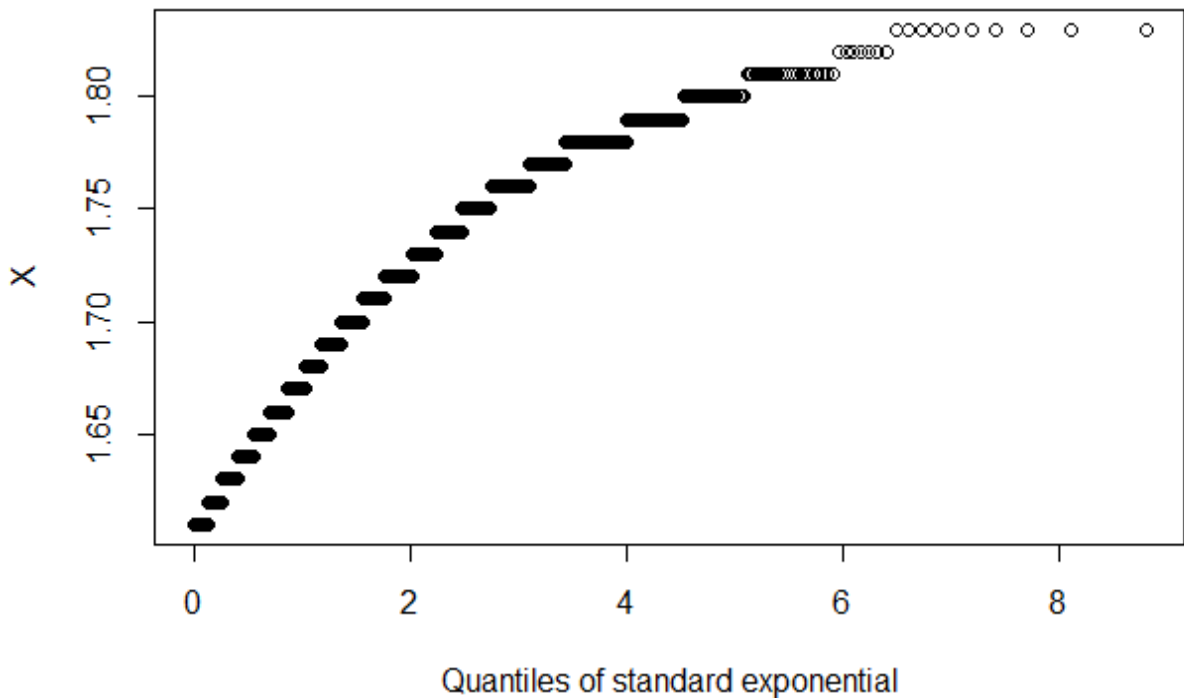


Figure 4.3: Exponential QQ plot for dataset

From the exponential Q-Q plot in figure (4.3), we realized that the sample path follows a concave pattern but not a linear trend. The same pattern was observed in figure (4.4) for the QQ plot for the excesses above the threshold value of 1.77 meters.

This shows that the exponential model is not a suitable parametric model to be fitted to the excesses Y .

Beirlant et al. (2004) indicated that a concave pattern of the exponential QQ plot suggests that the distribution function \mathcal{F} has a lighter right tail than expected from an Exponential distribution. This leads us to a possible GPD with $\gamma < 0$, that is, a Beta-type right tail distribution.

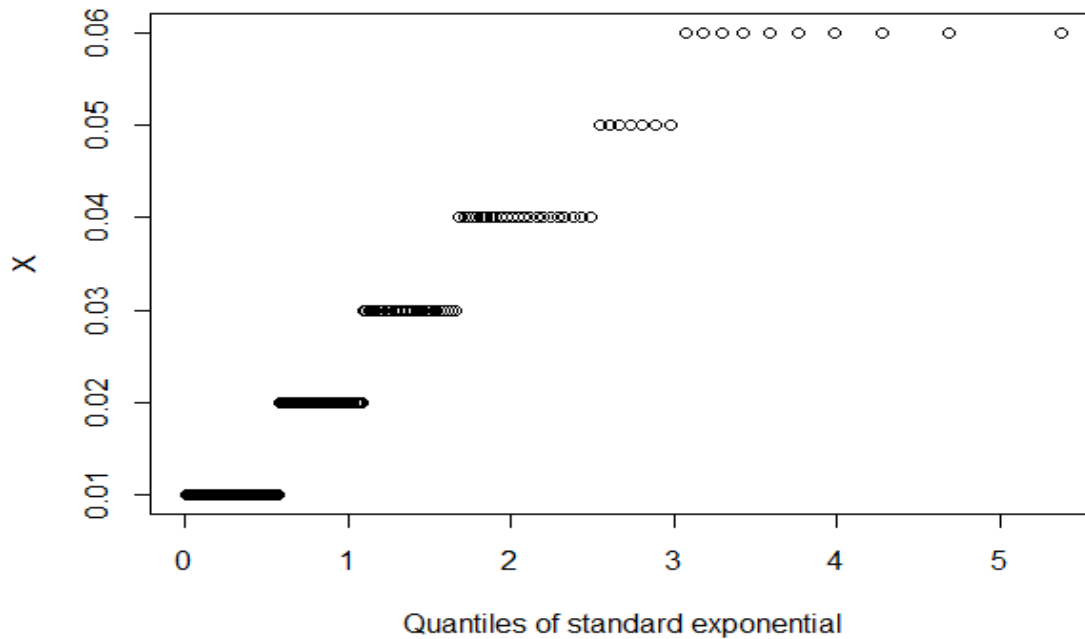


Figure 4.4: Exponential QQ plot for excesses above threshold

4.2 TEST TO DETERMINE DOMAIN-OF ATTRACTION

The initial analysis performed using the POT procedure pointed to a possible GPD with $\gamma < 0$ as the appropriate distribution to model the excesses above the chosen threshold. However, since the QQ plot is not free from some personal judgement, there is the need to perform some unbiased

tests to satisfy our self that the random variable Y can really be modeled with a GPD where $\gamma < 0$.

Three of these tests which will be used in this thesis were mentioned in the preceding chapter and are going to be applied to the sea level data.

Therefore, let X be the r.v that represents the sea heights used in the study. We denote by $Y = X - u$ the r.v representing excesses above the fixed threshold of u , in this case $u = 1.77$.

Since we are using the parametric approach, the assumption is that Y is GP- distributed

i. e. $Y \sim H_\gamma$ where H_γ is defined in chapter3. From the preceding chapter, we saw that we are particularly interested in testing (3.18) and (3.19).

With the POT approach, these numerical tests will be conducted with the exceedances acquired from the sea level.

If we let (W_1, \dots, W_m) stand for the m exceedances over a fixed threshold u , as defined in (3.5), removed from the available random sample (X_1, \dots, X_n) .

4.2.1 GOMES AND VAN MONFORT (1986) TEST

The Gomes and Van Monfort (1986) test is performed on the one-sided test (3.19). The test statistic for this test is given by (3.20). The test procedure as explained in chapter 3 leads to the rejection of H_o at a significant level of α if $G_m^* \leq \mathcal{G}_\alpha$, where \mathcal{G}_ε represents the standard Gumbel $\varepsilon - quantile$. Table 4.2 reports on the results obtained for the test. From the results obtained, H_o is rejected since the reported test value is highly significant at any asymptotic level of α . This means the appropriate parametric model needed to model the excesses Y is not the exponential model based on this test procedure. Hence the choice of a GPD with $\gamma < 0$ is maintained.

Table 4.2: Gomes and van Monfort (1986) Test

Statistic	Value
Gm	1.0055
Gm*	-4.6737
p-value	3.103e- 47

The decision concerning H_0 was taken based on the distribution of the test statistic at significant level of α . However, simulated critical points results suggested by Gomes & van Monfort (1986) for small and moderate sample sizes could have been used. Table 4.3 shows these critical points for sample sizes of $m = 20, m = 100$ and $m = 250$ at $\alpha = 0.05$ and $\alpha = 0.1$ significant levels respectively, where $x \downarrow$ represents estimated results smaller than x .

Table 4.3: Simulated rejection points for the test statistic Gm* for choosing GPD model

Statistic	m	Rejected region for $H_1: \gamma < 0$	
		0.1	0.05
Gm*	20	-0.89↓	-1.19↓
	100	-0.94↓	-1.21↓
	250	-0.86↓	-1.13↓
	∞	-0.83↓	-1.09↓

Gomes & van Monfort (1986)

Since the table does not include the sample size of $m = 215$, the researcher worked with the size $m = 250$ since it is the nearest available sample size. The observed value of $G_m^* = -4.6737$ is much smaller than any of the tabled rejection points. Hence, we maintain our decision of rejecting H_0 and state that based on the Gomes & Van Monfort (1986) test, the excesses Y is modelled by a GPD in the Weibull family.

4.2.2 MAROHN (2000) TEST

The second test we will consider is the one proposed by Marohn (2000) which is established on the sample coefficient of variation for analysing (3.18) and (3.19). As mentioned in chapter 3, Mahron (2000) used simulation studies and arrived at the conclusion that when working with small and moderate sample sizes, the test statistic for (3.18) is biased with very poor probability of not committing a type II error. The result showed that the test leads to an acceptable outcome only for big samples, that is $m \geq 500$. Since the sample size is 215 exceedances, the test given by (3.19) will be performed. The test statistic for the Mahron (2000) test is given by (3.21). Table 4.4 reports on the results obtained for the test for our fixed threshold of $u = 1.77$ meters.

Table 4.4: Marohn (2000) One-Sided Test Result

Statistic	Value
Tm	-0.5
Tm*	-7.3314
p-value	1.1385e-13

Therefore, at a 5% level of significance we reject the H_0 in favour of the alternative, implying a GPD model in the Weibull family.

4.2.3 THE LIKELIHOOD RATIO TEST (LRT)

The last GPD test to be considered in the thesis is the Likelihood Ratio Test (LRT). The likelihood Ratio Test (LRT) is applied to the sample exceedances to test (3.18). Like the two previous tests, the test procedure is presented in the preceding chapter. The test statistic for the LRT is given by (3.22). Table 4.5 reports on the final estimate of the GPD parameters using the ML approach.

Table 4.5: Final ML estimates for GPD parameters

Model	$\hat{\gamma}$ (shape)	$\hat{\sigma}_u$ (scale)
Exponential	0	0.0221
GPD	-0.4877	0.0326

With the final ML estimates in hand, the researcher can now perform the LRT using (3.22). Table 4.6 reports on the results obtained for the test.

Table 4.6: Likelihood Ratio Test

Statistic	Value
l	41.8318
l*	41.0678
p-value	1.4704e-10

Therefore, at a 5% level of significance we reject the H_0 in favour of the alternative, implying a GPD model in the Weibull family.

All the tests performed lead to the rejection of the Exponential model as an appropriate parametric model to be fitted to the excesses above our fixed threshold of $u = 1.77$ meters. These three objective tests agree with the preliminary subjective test performed using the Exponential QQ plot. We therefore conclude that the excesses above our fixed threshold $u = 1.77$ meters is best modeled by a GPd in the Weibull family.

4.2.4 GOODNESS-OF-FIT TEST

Since we adopted a GPD having a negative tail index based on the tests in the previous section, we can now perform some tests to determine how well the GPD model fit the data. Three tests for the GPD domain discussed in chapter 3 will be considered. These are the Kolmogorov-Smirnov, the Cramer-von Mises and the Anderson Darling tests.

4.2.4.1 KOLMOGOROV-SMIRNOV TEST

The Kolmogorov test was studied in the conditions of the exponential distribution without known parameters by Lilliefors (1969). Hence how well the exponential distribution fit the r.v Y (excesses above our fixed threshold) can be checked, since the GPD represents an exponential model when $\gamma = 0$. The null hypothesis for this test states that the fit is an exponential model. The test statistic for this test is given by (3.36) and the test procedure is explained in the preceding chapter.

With $\hat{\sigma}_u$ given in table 4.5, the Kolmogorov-Smirnov statistic value obtained with the help of the R software is

Kolmogorov-Smirnov Statistic: 0.9059

The value of the test statistic is matched with the rejection values proposed by Lilliefors (1969), where the null hypothesis is rejected if the reported Kolmogorov-Smirnov statistic is greater than its corresponding critical point. Table 4.7 shows some transcribed rejection values.

Table 4.7: Rejection values of the Kolmogorov-Smirnov statistic transformed to the Exponential distribution with parameters unknown

Statistic	m	Significance level for Dm		
		0.1	0.05	0.01
Dm	5	0.406	0.442	0.504
	10	0.295	0.325	0.38
	15	0.244	0.269	0.315
	20	0.212	0.234	0.278
	30	0.174	0.192	0.226
	>30	$0.96/\sqrt{m}$	$1.06/\sqrt{m}$	$1.25/\sqrt{m}$

Lilliefors (1969)

Since the sample size of 215 is greater than 30, our critical values are given by

$$\frac{0.96}{\sqrt{215}} = 0.0655, \quad \frac{1.06}{\sqrt{215}} = 0.0723, \quad \text{and} \quad \frac{1.25}{\sqrt{215}} = 0.0852$$

for $\alpha = 0.1$, $\alpha = 0.05$ and $\alpha = 0.01$ respectively.

Therefore, because the observed Kolmogorov-Smirnov statistic of 0.9059 is more than any of the rejection values, the null hypothesis is rejected.

4.2.4.2 THE ANDERSON-DARLING AND CRAMER-VON MISES TESTS

The null hypothesis for these two tests states that the distribution function is a GPD with unknown parameters. The test statistic for the two tests is given by (3.38) and (3.37) respectively. The values of $\hat{\gamma}$ and $\hat{\sigma}_{u,H\gamma}$ are reported in table 4.5 and the corresponding observed results for the tests are reported in table 4.8.

Table 4.8: Cramer-Von Mises and Anderson-Darling Test Values

Statistic	Value
Cramer-Von Mises	0.0493
Anderson-Darling	0.9259

The values for these tests are matched against the rejection values that Choulakian and Stephens (2001) proposed. Part of the tabled rejection values are reported in the table below.

Table 4.9: Rejection Values for the Anderson-Darling (bold) and Cramer-von Mises (normal style) statistics.

γ	Upper-Tail Significant Points		
	0.1	0.05	0.01
0.9	0.094	0.115	0.165
0.9	0.641	0.771	1.086
0.5	0.101	0.124	0.179
0.5	0.685	0.83	1.18
0.1	0.116	0.144	0.21
0.1	0.766	0.935	1.348
0	0.124	0.153	0.224
0	0.796	0.974	1.409
-0.1	0.129	0.16	0.236
-0.1	0.831	1.02	1.481
-0.5	0.174	0.222	0.338
-0.5	1.061	1.321	1.958

Choulakian & Stephens (2001)

The rejection points were acquired by simulating values of γ between -0.5 and 0.9 inclusive. Choulakian and Stephens (2001) indicated that these critical values can be used with superior precision for $m \geq 25$.

The table is entered at $\hat{\gamma} = -0.5$ since the estimate for γ is -0.4877 as reported in table 4.5. Therefore, at a 5% level of significance, we reject H_0 for both tests in favour of the alternative.

All three tests conducted are consistent, indicating that a GPD model in the Weibull family is an appropriate parameter distribution that fits the excesses.

4.3: FITTING THE GPD MODEL AND ESTIMATION OF EXTREME EVENTS

Parameter estimation and various inferences can now be performed once the underlying distribution function is known. In fitting the GPD model, there is the need to check whether to model the Axim Sea level as stationary or non-stationary. There is therefore the need to perform a test to confirm the stationarity or non-stationarity nature of the model.

4.3.1 ESTIMATION OF PARAMETERS

The preliminary visualization investigation and all the three domain of attraction tests conducted on the data led to the selection of a GPD having a negative tail index, that is $\gamma < 0$ as a suitable parametric distribution to model the excesses Y above the non-random threshold $u = 1.77$ meters.

In this section, the performance of the Probability Weighted Moment (PWM) estimates and the Maximum Likelihood (ML) estimates on the sample is compared. This was necessary because of the notion that the PWM estimators work well than the ML estimates in a small sample situation. The question is, what value denotes a small sample?

The estimation results for the GPD parameters using both the PWM and ML estimation procedures are presented in table 4.10.

Table 4.10: Final PWM and ML estimates for GPD model

Estimation Method	$\hat{\gamma}$ (shape)	$\hat{\sigma}_u$ (scale)
ML	-0.4877	0.0326
PWM	-0.4833	0.0324

The values recorded for the scale and shape parameters indicates there isn't much difference in the model parameters by the two estimation procedures. The PWM method also settled on a light right tail for the d.f \mathcal{F} . The ML and PWM estimate for the scale parameter ($\hat{\sigma}_u$) are very low, pointing to a smaller dispersion in the sea level excesses.

4.3.2 DIAGNOSTIC PLOTS

Figures (4.5) and (4.6) shows the quality of the GPD model using both ML and PWM fits, respectively.

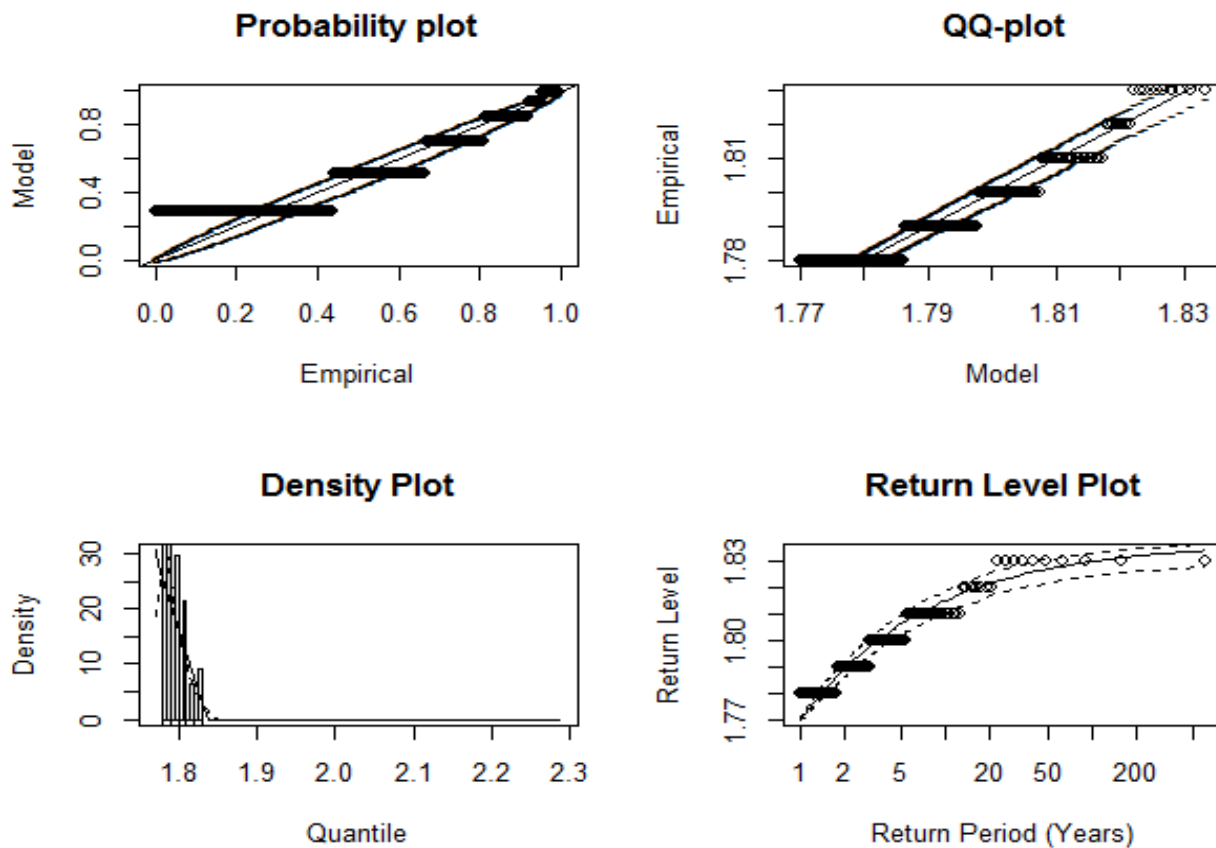


Figure 4.5: Diagnostic plots for GPD fit using ML estimates

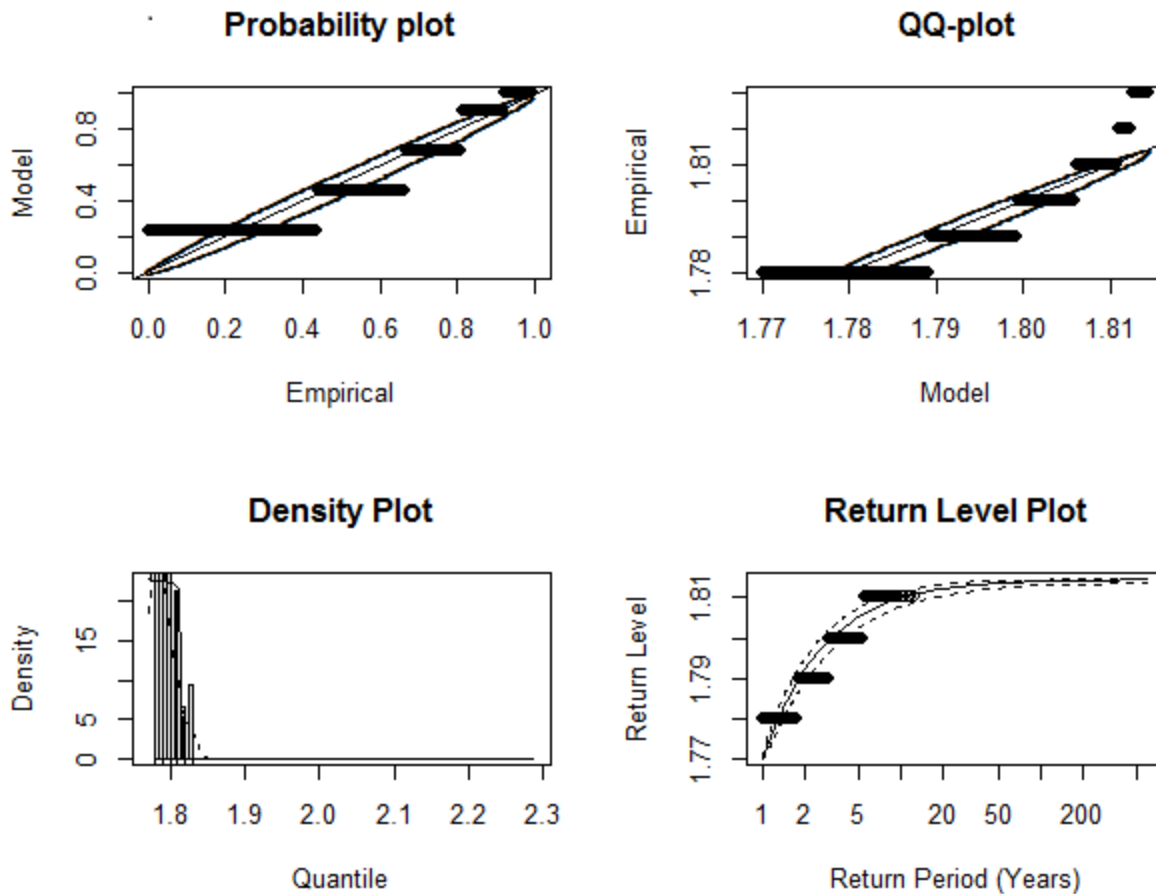


Figure 4.6: Diagnostic plots for GPD fit using PWM estimates.

All the three diagnostic plots for both estimation methods shows a satisfying GPD fit by both methods. The first two plots (probability and quantile plots) shows a near linear pattern.

The density plot also shows a good fit of the GPD by both methods since there is not much difference between the kernel density (or histogram) and the fitted density. This means based on our sample size of 215 excesses, both the ML and PWM estimation methods fits a GPD model to the excesses above the fixed threshold of 1.77 meters better.

4.3.3 CONFIDENCE INTERVAL ESTIMATION

A confidence interval (CI) can now be obtained for the shape parameter estimate from the ML procedure using the profile likelihood method. The ML method is not applicable for $\gamma \leq -1$ and as the profile likelihood confidence interval procedure is based on likelihoods method, one need to check the value of the estimated shape parameter before finding the profile likelihood CI.

Since our ML estimate for γ is $-0.4877 (> -1)$, we can find our profile likelihood CI without any problem. The profile log-likelihood function for γ with the confidence interval is shown in figure (4.7). The plot establishes the presence of the profile log-likelihood function.

Hence, the estimate of γ at a fixed threshold $u = 1.77 \text{ meters}$ is -0.4877 and the 95% profile likelihood confidence interval is $[-0.6112, -0.3642]$

Since the CI for the EVI does not include zero but encompass negative values, we conclude that the fit is a GPD with $\gamma < 0$.

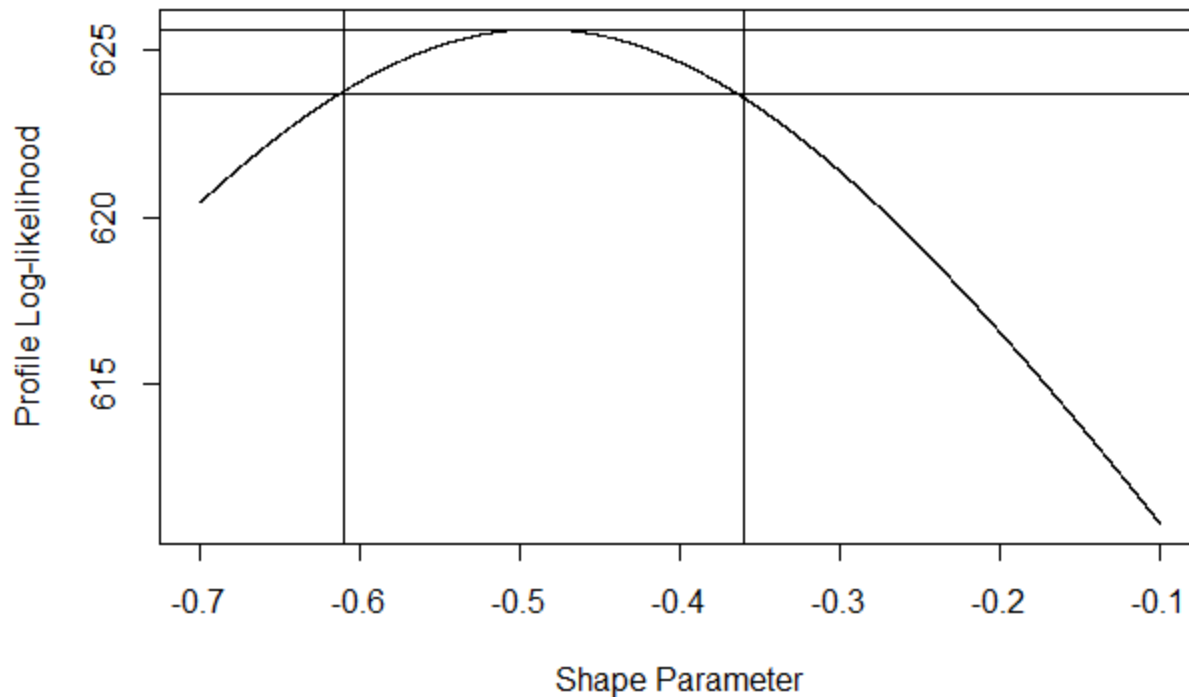


Figure 4.7: A 95% confidence interval for the GPD shape parameter based on profile-likelihood

4.3.4 ESTIMATION OF UPPER ENDPOINT

Since a GPD with $\gamma < 0$ was adopted, it means an estimate for the upper endpoint can be obtained.

This value is the maximum height we expect the Axim sea to attain based on our model.

The right endpoint calculated using the ML estimation method gave a value of 1.8370 meters.

By this method, it means the maximum upper bound estimated for the sea height is 1.8370 meters.

The ML estimate of the right endpoint is very close to the sample maximum of 1.83 meters, leaving practically no space for increases in the sea level.

This value is based on theory and might be different for different method and the data at hand.

4.3.5: EXCEEDANCE PROBABILITY

We estimate the exceedance probabilities for some selected sea heights close to the upper end-point estimate. Table 4.11 shows the selected heights with their corresponding probability of exceedance. We can see that the probability of exceeding the selected heights is very low.

Table 4.11: Estimated Exceedance probabilities for selected sea levels

Sea Level (meters)	Exceedance Probability
1.8278	5.4526e-03
1.8300	3.0991e-03
1.8304	2.6954e-03
1.8323	1.3393e-03
1.8339	5.6916e-04
1.8348	2.7671e-04

4.3.6: RETURN LEVEL ESTIMATE

In this section, we estimate the 100-year return period for the sea heights. The 100-year return period is the sea level that can be observed on average once every 100 years. Table 4.12 shows the return periods along with their return levels. The 100-year return period for the sea height has a return level of 1.8348 meters. this means we expect the Axim sea height to exceed 1.8348 meters once in 100 years. This sea level is close to the estimated upper endpoint of 1.837 meters but more than more than the currently observed maximum of 1.83 meters.

Table 4.12: Return Period estimate for GPD fitted Model

Return Period (years)	Return Level (meters)
5	1.8278
10	1.8304
20	1.8323
50	1.8339
100	1.8348

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

This study was performed to estimate how high the sea level of the Axim sea can rise based on available sample data, with the idea that an excess sea level rise will lead to flooding in communities along the coast. Extreme value Theory was used to determine how high the sea can rise above the current maximum measurement.

5.1 CONCLUSIONS

The study was made up of a general objective and four specific objectives. The general objective of the study is to assess the effect of very high sea level on the Axim community. The findings and conclusions drawn that answers the research objectives are presented in this section. Since the study was to find out increases that will lead to flooding, sea levels above 1.6m were considered in the study.

The first objective of the study was to determine the domain of attraction of the Axim sea level data. The peaks-over-threshold (POT) approach was used to model a Generalised Pareto distribution (GPD) to the excess sea level data above the fixed threshold value of 1.77 meters. Parameter estimation techniques used in the study were the Maximum Likelihood (ML) and the Probability Weighted Moment (PWM) methods. The exponential QQ plot was used as a visualisation tool to test for the domain of attraction of the underlying distribution function. The QQ plot was concave in nature, showing that we are in the Weibull domain, i.e. $\gamma < 0$.

In addition, three domain of attraction tests were performed to check the domain of attraction of the underlying distribution function and to confirm or deny the outcome from the QQ plot. The tests performed were the Gomes and van Monfort (1986) test, the Marohn (2000) and the

Likelihood Ratio Test (LRT). The outcome from all three tests show that the underlying distribution function of the data belongs to the Weibull domain of attraction.

The ML and PWM estimation procedures also reported negative EVI's ($\gamma < 0$) for the GPD model. The negativity nature of the EVI means tail distribution of the Axim sea level is in the Weibull domain of attraction.

The two parameter estimation procedures (ML and PWM) reported almost the same values for the scale and shape parameter for the fitted GPD model. The EVI value obtained for the ML method was a little greater than -0.5 (i.e.-0.4877) and that for the PWM is -0.4833. The value obtained for the EVI estimate using ML indicates that usual asymptotic properties were satisfied and that the estimators are regular. Smith (1985) showed that when $\gamma > -\frac{1}{2}$, the ML estimators are regular, and the estimators satisfy the standard asymptotic properties.

The diagnostic plots for the two methods all showed a good fit of the model.

Three diagnostic tests were also performed with the ML parameter estimates using the exponential model and the GPD model. The tests performed were the Kolmogorov-Smirnov test, the Cramer-Von Mises test and the Anderson-Darling goodness-of-fit tests. The results of the tests also led us to the adoption of a GPD with a negative EVI ($\gamma < 0$).

Once the underlying distribution of a Gpd with negative EVI ($\gamma < 0$) was adopted, it meant in theory, the maximum level we expect the sea to attain can be estimated. From the analysis carried out, a maximum value of 1.837 meters was obtained. This means in theory the sea level can rise to a maximum height of 1.837 meters above mean sea level but cannot exceed this value.

Knowing the upper endpoint of our underlying distribution function \mathcal{F} , exceedance probabilities for some selected sea levels were estimated. These probabilities provide coastal development planners with information on the chances of the sea rising above these selected sea levels. The study showed that the probability of the sea rising above the maximum observed value within the study period of 1.83m is 0.003099 (3.099e-3)

Lastly, extreme quantiles were estimated by determining the return period of the Axim sea level. The study showed that the 100-year return levels is 1.8348 meters. This means we expect the sea level to rise above 1.8348 meters once every 100 years.

In conclusion, the Axim sea level is modelled with a GPD having a negative tail index ($\gamma < 0$), and model is fitted better with estimates obtained with both the ML and PWM estimation method as compared to the PWM estimation method.

5.2 RECOMMENDATIONS

- Since it is possible for the sea level to rise above the current maximum, coastal development planners can consider increasing the height of any defence wall to accommodate the increases.
- Decisions to scale up or replicate similar intervention must consider the critical issues of transfer of impacts to neighbouring communities, costs as well as impacts on socio-economic activities and biodiversity conservation.

5.3 AREAS FOR FUTURE STUDIES

- Multivariate extreme value analysis. Including other factors that affect sea level rise and flooding such as shoreline change, wind waves, sand winning etc. can be considered for further investigation into this problem.
- Future researchers can employ non-stationary extreme value analysis on the sea level data.

REFERENCES

- Adam, M. B. & Tawn, J. A. (2012). Bivariate Extreme Analysis of Olympic Swimming Data. *Journal of Statistical Theory and Practice*, 6(3), 510 – 523.
- doi: 10.1080/15598608.2012.695702.
- Appeaning-Addo, K. (2013). Assessing Coastal Vulnerability Index to Climate Change: the case of Accra - Ghana. *Journal of Coastal Research* (Special Issue), 1892 - 1897.
- Balkema, A. A. & de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2 (5), 792-804.
- Barão, M.I. & Tawn, J.A. (1999). Extremal Analysis of Short Series with Outliers: Sea-Levels and Athletic Records. *Applied Statistics*, 48 (4), 469-487.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, England.
- Bezak, N., Brilly, M. & Sraj, M. (2014). Comparison between the Peaks-Over-Threshold Method and Annual Maximum Method for Flood Frequency Analysis. *Hydrological Sciences Journal*, 59(5), 959 – 977.
- Blanchet, J., Marty, C. & Lehning, M. (2009). Extreme Value Statistics of Snowfall in the Swiss Alpine region. *Water Resources Research*, 45, 1 – 12.

- Boko, M., Niang, I., Nyong, A., Vogel, C., Githeko, A., Medany, M. et al., (2007), 'Africa. Climate change 2007: Impacts, adaptation and vulnerability', in M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden & C.E. Hanson (eds.), contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 433–467, Cambridge University Press, Cambridge.
- Bortkiewicz, L., von (1922). Variationsbreite und mittlerer Fehler. Sitzungsberichte der Berliner Mathematischen Gesellschaft 21, 3-11. (in Johnson et al. 1995)
- Chan, S. & Nadarajah, S. (2015). Extreme Value Analysis of Electricity Demand in the UK. *Applied Economics Letters*. 22(15), 1246 – 1251.
- Choulakian, V. & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43(4), 478-484.
- Coles, S. (2001). An Introduction to Statistical Modelling of Extreme Values. Great Britain: Springer-Verlag, London.
- Coles, S.G. & Tawn, J.A. (1996). Modelling Extremes of the Areal Rainfall Process. *Journal of the Royal Statistical Society B*, 58(2), 329-347.
- Cotter, J. (2005). Extreme Risk in Futures Contracts. *Applied Economics Letters*, 12(8), 489 – 492. doi: 10.1080/13504850500109816.
- Danielsson, J. & de Vries, C. (1997). Tail index and quantile estimation with very High frequency data, *Journal of Empirical Finance*, 4, 241- 257.

- Davison, A. C. (1984). *Statistical Extremes and Applications*. D. Reidel, Dordrecht, Holland,
Ch. Modelling excesses over high thresholds, pp. 461-482.
- Davison, A.C. & Smith, R.L. (1990). Models for Exceedances over High Thresholds.
Journal of the Royal Statistical Society B, 52 (3), 393-442.
- de Haan, L. (1970). On Regular Variation and its Applications to the Weak Convergence
of Sample Extremes. Mathematical Centre Tract, 32, Amsterdam.
- de Haan, L. & de Ronde, J. (1998). Sea and Wind: Multivariate Extremes at Work.
Extremes, 1 (1), 7-45.
- de Haan, L. & Ferreira, A. (2006). *Extreme Value Theory - An Introduction*. Springer, New
York.
- Dekkers, A.L.M. & de Haan, L. (1989). On the Estimation of the Extreme-Value Index
and Large Quantile Estimation. *The Annals of Statistics*, 17 (4), 1795 - 1832.
- D'Onofrio, E. E., Fiore, M. E. M. & Romero, I. S. (1999). Return Periods of Extreme Water
Levels Estimated for some vulnerable areas in Buenos Aires. *Continental Shelf Research*,
19, 1681 – 1693.
- Dodd, E.L. (1923). The Greatest and Least Variate under General Laws of Error.
Transactions of the American Mathematical Society, 25, 525-539. (in Johnson et al.1995)
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R. & Bobee, B. (2007). Generalised
Maximum Likelihood Estimators for the Non-Stationary Generalised Extreme Value
Model. *Water Resources Research*, 43. doi: 10.1029/2005WR004545.

- Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance, 1, Springer, Berlin.
- Embrechts, P., McNeil, A. & Straumann, D. (1999). Correlation and dependency in risk management: properties and pitfalls, preprint, ETH Zurich.
- Embrechts, P., Resnick, S. & Samorodnitsky, G. (1998), Living on the edge, *RISK Magazine*, 11(1), 96-100.
- Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- Frechet, M. (1927). Sur la loi de Probabilité de l'Écart Maximum. *Annales de la Société Polonaise de Mathématique, Cracovie* 6, 93-116. (in Johnson et al. 1995)
- Gencay, R. & Selcuk, F. (2004). Extreme Value Theory and Value-at-Risk: Relative Performance in Emerging Markets. *International Journal of Forecasting*, 20, 287-303.
- Gilli, M. & Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27(1), 1 – 23.
- Gnedenko, B. (1943). “On the limit distribution of the maximum term of a random series”. *Annals of Mathematics*, 44 (3), 423-453.
- Gomes, M. I. & van Monfort, M. A. J. (1986). Exponentiality versus Generalized Pareto, quick tests. In: *Proc. III Internat. Conf. Statistical Climatology*. pp. 185-195.
- Harter, H.L. (1978). A Bibliography of EVT. *International Statistical Review*, 46, 279-306.

- Hawkes, P. J., Ganzalez - Marco, D., Sanchez – Arcilla, A. & Prinos, P. (2008). Best Practice for the Estimation of Extremes: A Review. *Journal of Hydraulic Research*, 46(S2), 324-332. doi: 10.1080/00221686.2008.9521965.
- Hosking, J. R. M. & Wallis, J. R. (1987). Parameter and quantile estimation for the generalised Pareto distribution. *Technometrics*, 29, 339-349.
- Huang, W., Xu, S. & Nnaji, S. (2008). Evaluation of GEV Model for Frequency Analysis of Annual Maximum Water levels in the Coast of United States. *Ocean Engineering*, 35, 1132–1147.
- Hundecha, Y., St - Hilaire, A., Ouarda, T. B. M. J. & El Adlouni, S. (2008). A Non-stationary Extreme Value Analysis for the Assessment of Changes in Extreme Annual Wind Speed over the Gulf of St. Lawrence, Canada. *Journal of Applied Meteorology and Climatology*, 47, 2745 - 2759. doi: 10.1175/2008JAMC1665.1.
- Ibe, A. C. & Quelennac, R. E. (1989). Methodology for Assessment and Control of Coastal Erosion in West Africa and Central Africa UNEP Regional Sea Reports and Studies, 107, New York: United National Environmental Programme.
- Intergovernmental Panel on Climate Change (IPCC). (1995). ‘The science of climate change’, contribution of Working Group I to the Second Assessment Report of the IPCC, 572, Cambridge University Press, Cambridge.

Intergovernmental Panel on Climate Change (IPCC). (2007). 'Observations of climate change', in Core Writing Team, et al. (eds.), *Climate change 2007: Synthesis report, contribution of Working Groups, I, II, and III to the Fourth Assessment Report of the IPCC*, IPCC, Geneva.

Intergovernmental Panel on Climate Change (IPCC). (2007). Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In C. C. Report (Ed.), (pp. 315-356): Intergovernmental Panel of Climate Change.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158-171.

Jockovic, J. (2012). Quantile Estimation for the GPD with Application to Finance. *Yugoslav Journal of Operations Research*, 22(2), 297 - 311.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*, 2(2), Wiley, New York.

Kagan, Y.Y. (1997). Earthquake Size Distribution and Earthquake Insurance. *Communication in Statistics – Stochastic Models*, 13(4), 775-797.

- Katz, W., Richard, Parlange, B., Marc & Naveau, P. (2002). Statistics of Extremes in Hydrology. *Advances in Water Resources*, 25, 1287 - 1304.
- Kebede, A.S., Nicholls, R.J., Hanson, S. & Mokrech, M. (2012). 'Impacts of climate change and sea-level rise: A preliminary case study of Mombasa, Kenya', *Journal of Coastal Research*, 28(1A), 8–19.
- Khalique, M., Ouarda, T., Ondo, J. C., Gachon, P. & Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of hydrology*, 329(3), 534-552.
- Kinnison, R.R. (1985). Applied Extreme Value Statistics. MacMillan, New York.
- Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A. & Rominger, D. (2009). Testing exponentiality versus Pareto distribution via likelihood ratio. *Commun. Stat.- Simul. C.* 38 (1), 118-139.
- Küchenhoff, H. & Thamerus, M. (1996). Extreme Value Analysis of Munich Air Pollution Data. *Environmental and Ecological Statistics*, 3(2), 127-141.
- Lilliefors, H. W. (1969). On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *Journal of American Statistical Association*, 64 (325), 387-389.
- Marimoutou, V., Reggad, B. & Trabelsi, A. (2009). Extreme Value Theory and Value at Risk: Application to Oil Market. *Energy Economics*, 30, 519 - 530.

- Marohn, F. (2000). Testing extreme value models. *Extremes*, 3(4), 363-384.
- McNeil, A. (1997). 'Estimating the tails of loss severity distributions using EVT',
ASTIN Bulletin, 27, 117-137.
- McNeil, A. (1998). 'History repeating', *Risk*, 11(1), 99.
- McNeil, A. J. (1999). Extreme value theory for risk managers. British Bankers' Association,
Internal Modelling and CAD II: Qualifying and Quantifying Risk within a Financial
Institution, 93-113. RISK Books, London.
- Mikosch, T. (1997). Heavy-Tailed Modelling in Insurance. *Communication in
Statistics– Stochastic Models*, 13 (4), 799-815.
- Mendez, F., J. & Menendez, M. (2006). Analysing Monthly Extreme Sea levels with a Time-
Dependent GEV Model. *Journal of Atmospheric and Oceanic Technology*, 24, 894 - 910.
- Mudersbach, C. & Jensen, J. (2010). Nonstationary Extreme Value analysis of Annual Maximum
Water Levels for Designing Coastal Structure on the German North Sea Coastline.
Journal of Flood Risk Management, 3, 52 - 62.
- Northrop, P. & Jonathan, P. (2011). Threshold Modelling of Spatially Dependent Non-stationary
Extremes with Application to Hurricane-induced Wave Heights. *Environmetrics*, 22,
799 - 809.
- Peeler, K. (2007). 'Nigeria in the dilemma of climate change', Country report, viewed
23 June 2012, from [http://www.kas.de/proj/home/pub/33/2/dokument_id-
11468/index.html](http://www.kas.de/proj/home/pub/33/2/dokument_id-11468/index.html)

- Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annual Statistics*, 3 (1), 119-131.
- Reiss, R., D. & Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 3rd Edition. Birkhauser Verlag, Basel-Boston-Berlin.
- Rootzén, H. & Tajvidi, N. (1997). Extreme Value Statistics and Wind Storm Losses: A Case Study. *Scandinavian Actuarial Journal*, 70-94.
- Strand, M. & Boes, D. (1998). Modelling Road Racing Times of Competitive recreational Runners Using EVT. *The American Statistician*, 52 (3), 205-210.
- Tawn, J. A. (1992). Estimating Probabilities of Extreme Sea-Levels. *Applied Statistics*, 41 (1), 77-93.
- Tippet, L.H.C. (1925). On the Extreme Individuals and the Range of Samples Taken from a Normal Distribution. *Biometrika*. 17(3-4), 364-387. (in Kinnison, 1985)
- Tramblay, Y., Neppel, L., Carreau, J. & Najib, K. (2013). Non-stationary Frequency Analysis of Heavy Rainfall Events in Southern France. *Hydrological Sciences Journal*, 58(2), 280-294.
- Vicente, S. L. G. (2012). Extreme value theory: an application to sports. PhD thesis, University of Lisbon, Portugal.
- Von Mises, R. (1923). "About the variation width of a series of observations." Meeting reports of the Berlin Mathematical Society, 22, 3-8. (in Johnson et al.1995).

- von Mises, R. (1936). La distribution de la plus grande den values. Reprinted in Selected Papers Volume II, Amer. Math. Soc., Providence, R.I., 271-294.
- Watts, K. A., Dupuis, D. J. & Jones, B. L. (2013). An Extreme Value Analysis of Advance Ade Mortality Data. *North American Actuarial Journal*, 10(4), 162 - 178.
- Wax, E. (2007). 'In flood-prone Bangladesh, a future that floats.' the Washington Post.
- Zhang, Q., Xu, C.Y., Chen, Y., David & Liu, C. L. (2009). Extreme Value Analysis of Annual Maximum Water Levels in the Pearl River Delta, China. *Front Earth Sci. China*, 3(2).

APPENDICES

APPENDIX I

R codes used in the research to analyse the Axim sea level data.

Libraries used

```
library(fBasics)
library(evir)
library(ReIns)
library(evd)
library(fitdistrplus)
library(fExtremes)
library(POT)
library(ismev)
```

Preliminary Statistical Analysis

Summary Statistics

```
mydata<-data$Height[data$Height>1.60]
basicStats(mydata)
```

Histogram for data

```
hist(mydata,col="red",freq= FALSE)
```

Mean Excess Plot

```
meplot(mydata,type="l",omit=3, labels=TRUE,main="Mean excess
plot")
abline(v=1.77,col="red")
```

Exponential QQ Plot

```
excess<-mydata[which(mydata>1.77)]-1.77
ExpQQ(excess, plot=TRUE, type="l", main="")
ExpQQ
```

Estimating parameters of GPD

ML estimation

```
Qgpd<-((1-ie/(me+1))^( -ggpdopt$maximum)-1)/(ggpdopt$maximum)
linereggpd<-lm(excess~Qgpd-1)
a_exp<-as.vector(1/linereggpd$coefficients)
fexp<-fitdist(excess,"exp",start=list(rate=a_exp))
aexp<-as.vector(1/fexp$estimate)
fgpd<-gpd.fit(mydata,threshold=1.77,show=F)
shapgpd<-as.vector(fgpd$mle[2])
scalgpd<-as.vector(fgpd$mle[1])
cat(" [1] Exponential ML estimates","\n"," sigma_u=",aexp,"\n",
     "[2] Gpd ML estimates","\n"," gamma=",shapgpd,"
     sigma_u=",scalgpd,"\n").
```

PWM estimation

```
fgpdpwm<-gpdFit(mydata,threshold=1.77,method = "pwm")
shapgpdpwm<-as.vector(fgpdpwm$par.ests[2])
scalgpdpwm<-as.vector(fgpdpwm$par.ests[1])
cat(" [1] Gpd PWM estimates","\n","gamma=",shapgpdpwm,"
     sigma_u=",scalgpdpwm,"\n")
```

GPD fit diagnosis

ML fit diagnosis

```
fitml<-fitgpd(mydata,threshold=1.77)
par(mfrow=c(2,2))
plot(fitml,which=1)
par(mfrow=c(1,1))
```

PWM fit diagnosis

```
fitpwm<-fitgpd(mydata,threshold=1.77,est="pwmu")
par(mfrow=c(2,2))
plot(fitpwm,which=1:4)
par(mfrow=c(1,1))
```

Test for Domain of Attraction

Gomes and Van Monfort (1986) Test

```
excess<-mydata[which(mydata >1.77)] -1.77
me<-length(excess)
exc<-mydata[which(mydata>1.77)]
Gm<-exc[me]/exc[floor(me/2)+1]
Gmstar <-log(2)*Gm-log(me)
pvalueGmstar<-pgumbel(Gmstar)
cat("[1] g_m=",Gm," g_m*=",Gmstar," p-value=",pvalueGmstar,"\n")
```

Marohn (2000) Test

```
T_m<-0.5*((var(exc)*(me-1)/me)/(mean(exc)-1.77)^2-1)
T_mstar<-sqrt(me)*T_m
pvaluniTmstar<-pnorm(T_mstar)
cat("[1] One-sided Test","\n","t_m=",T_m," t_m*=",T_mstar," p-
value=",pvaluniTmstar,"\n")
```

Goodness-of-fit Test

Kolmogorov-Smirnov test

```
KS<-max(max(abs(pexp(excess,rate=1/aexp)-ie/me),
abs(pexp(excess,rate=1/aexp)-(ie-1)/me)))
cat("Kolmogorov-Smirnov statistic: ",KS,"\n")
```

Cramer-von Mises and Anderson Darling

```
pGP<-function(x,g,a){1-(1+g*x/a)^(-1/g)}
CVM<-sum((pGP(excess,shapgpd,scalgpd)-(2*ie-
1)/(2*me))^2)+1/(12*me)
AD<--me-1/me*sum((2*ie-1)*log(pGP(excess,shapgpd,scalgpd))+
(2*me+1-2*ie)*log(1-pGP(excess,shapgpd,scalgpd)))
cat(" Cramer-von Mises statistic: ",CVM,"\n","Anderson-Darling
statistic: ",AD,"\n")
```

Profile confidence Interval

```
pot<-fitgpd(mydata,threshold=1.77,"mle" )
gpd.pfshape(pot, c(-0.7, -0.1), nrang = 5000)
gpd.pfscale(pot, c(0, 0.1))
gpd.pfrl(pot, c(1.7, 1.90))
```

Profile Confidence Interval

```
fgpd<-gpd.fit(mydata,threshold=1.77,show=F)
gpd.profxi(fgpd,xlow=-0.7,xup=-.1,nint=8000)
abline(v=-0.36)
```

Estimation of upper Endpoint

```
xF_potml<-1.77-scalgpd/shapgpd
xF_potpwm<-1.77-scalgpdpwm/shapgpdpwm
cat(" [1] Maximum Likelihood: x^F=",xF_potml,"\n","[2]
Probability Weighted Moments: x^F=",xF_potpwm ,"\n")
```

Exceedance Probability

Above 1.83 meters

```
m<-length(mydata)
exceedmaxpotml<-me/m*(1-pGP(0.06,shapgpd,scalgpd))
exceedmaxpotpwm<-me/m*(1-pGP(0.06,shapgpdpwm,scalgpdpwm))
cat(" [1] Maximum Likelihood:
P(X>1.83)=",exceedmaxpotml,"\n","[2] Probability Weighted
Moments: P(X>1.83)=",exceedmaxpotpwm ,"\n")
```

Above 2 meters

```
m<-length(mydata)
exceedmaxpotml<-me/m*(1-pGP(0.23,shapgpd,scalgpd))
exceedmaxpotpwm<-me/m*(1-pGP(0.23,shapgpdpwm,scalgpdpwm))
cat(" [1] Maximum Likelihood:
P(X>2.0)=",exceedmaxpotml,"\n","[2] Probability Weighted
Moments: P(X>2.0)=",exceedmaxpotpwm ,"\n")
```

APPENDIX II

DATA

A sample of the data used in the research has been added for easy reference by researchers. The rest of the data can be accessed by writing to the email opokuenock63@yahoo.com.

HOURLY SEA LEVEL OF THE AXIM SEA, JANUARY 1980.

TIME	1ST	2ND	3RD	4TH	5TH	6TH	7TH	8TH	9TH	10TH
12:00 AM	0.46	0.58	0.74	0.93	1.1	1.23	1.3	1.32	1.29	1.2
1:00 AM	0.35	0.37	0.47	0.62	0.79	0.97	1.11	1.2	1.24	1.22
2:00 AM	0.38	0.31	0.32	0.39	0.53	0.7	0.87	1.01	1.12	1.18
3:00 AM	0.55	0.41	0.33	0.31	0.37	0.49	0.64	0.81	0.96	1.08
4:00 AM	0.81	0.63	0.48	0.39	0.36	0.39	0.49	0.63	0.79	0.95
5:00 AM	1.09	0.91	0.74	0.59	0.49	0.44	0.46	0.54	0.66	0.81
6:00 AM	1.34	1.2	1.04	0.87	0.72	0.61	0.55	0.55	0.61	0.72
7:00 AM	1.49	1.43	1.31	1.16	1	0.85	0.74	0.66	0.65	0.68
8:00 AM	1.52	1.55	1.5	1.4	1.27	1.12	0.97	0.85	0.76	0.73
9:00 AM	1.42	1.53	1.57	1.54	1.46	1.34	1.2	1.06	0.94	0.84
10:00 AM	1.2	1.38	1.5	1.56	1.55	1.48	1.38	1.26	1.13	0.99
11:00 AM	0.95	1.14	1.31	1.44	1.51	1.52	1.48	1.4	1.29	1.16
12:00 PM	0.74	0.88	1.05	1.22	1.36	1.44	1.47	1.46	1.4	1.31
1:00 PM	0.62	0.69	0.81	0.97	1.13	1.27	1.37	1.42	1.43	1.4
2:00 PM	0.65	0.62	0.66	0.76	0.9	1.05	1.19	1.3	1.38	1.42
3:00 PM	0.79	0.69	0.64	0.66	0.73	0.85	1	1.14	1.26	1.36
4:00 PM	1.02	0.87	0.75	0.69	0.68	0.74	0.83	0.96	1.1	1.24
5:00 PM	1.27	1.11	0.96	0.83	0.75	0.73	0.75	0.83	0.95	1.09
6:00 PM	1.49	1.36	1.2	1.05	0.92	0.82	0.77	0.78	0.84	0.94
7:00 PM	1.63	1.54	1.42	1.28	1.13	0.99	0.88	0.81	0.79	0.83
8:00 PM	1.63	1.62	1.56	1.46	1.32	1.18	1.03	0.91	0.82	0.78
9:00 PM	1.48	1.57	1.59	1.55	1.46	1.34	1.19	1.04	0.9	0.79
10:00 PM	1.21	1.37	1.48	1.52	1.5	1.42	1.31	1.17	1.01	0.86
11:00 PM	0.88	1.07	1.24	1.36	1.42	1.42	1.36	1.26	1.12	0.96

HOURLY SEA LEVEL OF THE AXIM SEA, JANUARY 1980.

TIME	11TH	12TH	13TH	14TH	15TH	16TH	17TH	18TH	19TH	20TH
12:00 AM	1.07	0.91	0.73	0.56	0.43	0.38	0.44	0.62	0.87	1.14
1:00 AM	1.15	1.03	0.86	0.67	0.48	0.33	0.27	0.33	0.5	0.76
2:00 AM	1.19	1.13	1.02	0.85	0.64	0.43	0.27	0.2	0.25	0.43
3:00 AM	1.16	1.19	1.15	1.05	0.87	0.65	0.43	0.25	0.17	0.23
4:00 AM	1.08	1.18	1.23	1.22	1.12	0.94	0.71	0.47	0.28	0.2
5:00 AM	0.97	1.12	1.24	1.32	1.31	1.22	1.04	0.79	0.55	0.36
6:00 AM	0.86	1.02	1.18	1.33	1.42	1.43	1.34	1.15	0.9	0.65
7:00 AM	0.77	0.9	1.07	1.25	1.42	1.53	1.55	1.46	1.27	1.01
8:00 AM	0.74	0.81	0.93	1.1	1.3	1.49	1.62	1.65	1.56	1.36
9:00 AM	0.78	0.76	0.81	0.93	1.11	1.33	1.54	1.68	1.72	1.62
10:00 AM	0.87	0.79	0.75	0.78	0.89	1.08	1.32	1.55	1.7	1.73
11:00 AM	1.02	0.88	0.77	0.7	0.72	0.83	1.03	1.28	1.52	1.68
12:00 PM	1.18	1.03	0.87	0.73	0.64	0.64	0.75	0.96	1.22	1.46
1:00 PM	1.32	1.19	1.03	0.85	0.68	0.57	0.56	0.67	0.88	1.15
2:00 PM	1.41	1.34	1.22	1.04	0.84	0.65	0.53	0.51	0.61	0.83
3:00 PM	1.42	1.44	1.39	1.26	1.07	0.85	0.64	0.51	0.48	0.59
4:00 PM	1.36	1.45	1.49	1.45	1.32	1.12	0.89	0.66	0.52	0.49
5:00 PM	1.24	1.39	1.5	1.56	1.53	1.4	1.19	0.94	0.71	0.56
6:00 PM	1.09	1.25	1.42	1.56	1.63	1.61	1.47	1.25	0.99	0.76
7:00 PM	0.93	1.07	1.25	1.44	1.61	1.69	1.67	1.53	1.3	1.04
8:00 PM	0.8	0.89	1.03	1.23	1.45	1.63	1.72	1.7	1.55	1.32
9:00 PM	0.73	0.73	0.81	0.97	1.18	1.41	1.61	1.72	1.69	1.54
10:00 PM	0.73	0.65	0.64	0.71	0.87	1.09	1.35	1.56	1.67	1.64
11:00 PM	0.8	0.65	0.54	0.51	0.58	0.74	0.99	1.25	1.47	1.58

HOURLY SEA LEVEL OF THE AXIM SEA, JANUARY 1980.

TIME	21ST	22ND	23RD	24TH	25TH	26TH	27TH	28TH	29TH	30TH	31ST
12:00 AM	1.37	1.48	1.45	1.3	1.08	0.85	0.66	0.53	0.46	0.47	0.54
1:00 AM	1.04	1.27	1.37	1.35	1.21	1.01	0.8	0.61	0.48	0.4	0.38
2:00 AM	0.69	0.97	1.19	1.29	1.27	1.15	0.97	0.78	0.6	0.46	0.36
3:00 AM	0.41	0.67	0.94	1.15	1.25	1.24	1.13	0.97	0.8	0.63	0.49
4:00 AM	0.26	0.43	0.69	0.95	1.15	1.26	1.25	1.17	1.03	0.87	0.71
5:00 AM	0.28	0.34	0.51	0.75	1	1.19	1.3	1.31	1.24	1.13	0.99
6:00 AM	0.47	0.4	0.45	0.61	0.84	1.07	1.25	1.36	1.39	1.35	1.25
7:00 AM	0.77	0.59	0.52	0.57	0.72	0.92	1.14	1.32	1.43	1.47	1.45
8:00 AM	1.11	0.87	0.69	0.62	0.67	0.8	0.99	1.19	1.36	1.48	1.54
9:00 AM	1.42	1.17	0.94	0.77	0.7	0.73	0.85	1.01	1.19	1.36	1.49
10:00 AM	1.64	1.44	1.2	0.98	0.81	0.74	0.76	0.85	0.99	1.15	1.32
11:00 AM	1.71	1.62	1.43	1.2	0.98	0.83	0.75	0.74	0.81	0.92	1.07
12:00 PM	1.63	1.66	1.57	1.39	1.17	0.97	0.82	0.73	0.7	0.74	0.83
1:00 PM	1.4	1.56	1.6	1.51	1.35	1.15	0.96	0.8	0.7	0.66	0.67
2:00 PM	1.09	1.34	1.5	1.54	1.47	1.32	1.13	0.96	0.81	0.69	0.63
3:00 PM	0.8	1.06	1.3	1.46	1.5	1.44	1.31	1.15	0.99	0.84	0.71
4:00 PM	0.59	0.8	1.05	1.28	1.43	1.49	1.44	1.34	1.2	1.05	0.9
5:00 PM	0.53	0.63	0.83	1.07	1.29	1.44	1.5	1.47	1.39	1.27	1.13
6:00 PM	0.61	0.58	0.68	0.86	1.09	1.3	1.44	1.52	1.52	1.46	1.36
7:00 PM	0.8	0.66	0.63	0.72	0.89	1.1	1.3	1.45	1.53	1.55	1.52
8:00 PM	1.05	0.83	0.69	0.66	0.73	0.89	1.08	1.27	1.42	1.53	1.57
9:00 PM	1.3	1.04	0.82	0.68	0.65	0.71	0.85	1.02	1.21	1.37	1.48
10:00 PM	1.48	1.25	1	0.78	0.65	0.61	0.65	0.77	0.93	1.1	1.27
11:00 PM	1.55	1.4	1.17	0.93	0.73	0.59	0.54	0.57	0.66	0.8	0.97