UNIVERSITY FOR DEVELOPMENT STUDIES

MODELLING EXTREME TEMPERATURE BEHAVIOUR IN UPPER EAST REGION, GHANA.

MORI KOJO WILSON

Thesis submitted to the Department of Statistics, Faculty of Mathematical Sciences, of the University for Development Studies, in Partial Fulfillment of the Requirement for the Award of Master of Science Degree in Applied Statistics

2016

UNIVERSITY FOR DEVELOPMENT STUDIES

MODELLING EXTREME TEMPERATURE BEHAVIOUR IN UPPER EAST REGION, GHANA.

BY

MORI KOJO WILSON (B. Ed. Mathematics)

(UDS/MAS/0033/13)

Thesis submitted to the Department of Statistics, Faculty of Mathematical Sciences University for Development Studies, in Partial Fulfillment of the Requirement for the Award of Master of Science Degree in Applied Statistics

OCTOBER, 2016

DECLARATION

Student

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere:

Candidate's Signature:…………………………….. Date:…………….

Name:…………………………………………………………………………….

Supervisor

I hereby declare that the preparation and presentation of the thesis was supervised in accordance with the guidelines on supervision of thesis laid down by the University for Development Studies:

Supervisor's Signature:………………………….. Date:…………….

Name: Dr. Albert Luguterah

ABSTRACT

The impacts of extremely high temperatures on plants, human beings and animals' health have been studied in several parts of the world. However, extreme events are uncommon and have only attracted attention recently. In this study, extreme temperature behaviour was modelled through the application of extreme value theory using maximum monthly temperatures over a 32 years period. Data on monthly maximum temperature from the Upper East Region were modelled using generalized extreme value (GEV) and generalized Pareto distributions (GPD) models. A trend analysis revealed that the maximum temperature returns followed a log-quadratic trend model. The results revealed that the GEV model was better in modelling extreme temperature behaviour because it had the least AIC and BIC values of -1003.6050 and -991.7600 respectively. Two comparative tests, namely, Anderson-Darling and Kolmogorov-Smirnov confirmed the GEV model to be adequate for the data.

Diagnostic checks of the two models using probability-probability (PP) plot, quantile-quantile (QQ) plot, return level plot and mean residual life plot revealed that the GEV fitted the data well. Return periods of 2, 20 and 100 years also revealed return levels of 0.000015, 0.106150 and 0.142920 respectively. It was therefore concluded that maximum temperature returns showed an increasing trend for long return periods.

Based on the outcome of the study, it is recommended that farmers are educated to undertake measures that will help conserve the environment while they engage in farming activities. Also ministries responsible for land and natural resources must design educative programmes on radio and other networks to sensitize the population on climate change treats and how they can combat while they cultivate the land.

ACKNOWLEDGEMENT

This research was able to reach its final stage due to numerous contributions and support in diverse forms from many persons. First of all my sincere thanks go to the Almighty God for divine support in this study. Secondly my profound gratitude goes to my supervisor Rev. Dr. Albert Luguterah for his fatherly guidance. This work would never have been realized without the huge support and guidance of my co-supervisor, Mr Nasiru Suleman. I owe him profound gratitude for the constructive criticism, professional guidance and inspiration offered me throughout my writing of this thesis.

I am also very much grateful to the entire postgraduate students of 2013 and 2014 especially Dokurugu Emmanuel who supported me in diverse ways to complete this work.

I also sincerely thank my wife, Vivian and my entire family and friends for their immeasurable support to me since I started this study.

DEDICATION

This work is dedicated to my beloved mother, Madam Leticia Mori who has encouraged me since childhood to rise to the sky no matter the obstructions

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AD	Anderson Darling
ADF	Augmented Dickey Fuller
AIC	Akaike Information Criteria
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criteria
CAIC	Consistent Akaike Information Criteria
CSM	Cerebrospinal Meningitis
COV	Covariance
DF	Dickey Fuller
EPA	Environmental Protection Agency
EVT	Extreme Value Theory
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GCM	Global Climate Model
GEV	Generalized Extreme Value
GLM	Generalized Linear Models

GPD	Generalized Pareto Distribution
GSS	Ghana Statistical Service
IPCC	Intergovernmental Panel on Climate Change
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
KS	Kolmogorov Smirnov
MK	Mann Kendall
NADMO	National Disaster Management Organization
PACF	Partial Autocorrelation Function
POT	Peaks Over Threshold
PP	Point Process
PP	Probability-Probability
QQ	Quantile-Quantile
RSS	Residual Sum of Squares
UER	Upper East Region
UNDP	United Nation Development Programme
WHO	World Health Organization

CHAPTER ONE

INTRODUCTION

1.1Background of the Study

Recently, issues relating to global warming have become very popular, taking the centre stage in many policy discussions both locally and internationally. The importance of this statement relies on the fact that as temperatures rise, the variability of climate will increase, and consequently an increase in natural disasters due to temperature extremes. The issue of global warming and climate change affect both developed and developing countries alike. These surely have negative consequences for our development. The increase in global temperatures has been attributed to both natural phenomena (e.g. volcanic eruptions) and human activities (e.g. bush burning, industrialization and urbanization). Extreme climate events are expected to become more frequent as a result of climate change and global warming.

Natural disasters such as floods, droughts, storms, tsunamis and unbearable high temperatures are forecast to occur more frequently in the future with their attendant negative impacts on the social, economic, health and environmental systems of our world. Extreme temperature events are of particular importance due to their severe impact on the environment, the economy and the society. It is recognized that developing countries such as Ghana will bear more the brunt of climate change and global warming because they have not developed adequate mitigation plans to arrest the situation.

The increase in greenhouse gases concentrations in the atmosphere is regarded as one of the causes of the rise in global temperature. The greenhouse gases which are normally associated with temperature rise are carbon dioxide, methane and nitrous oxides. The total greenhouse gas emission in Ghana was estimated at 12.2 $MtCO_2e$ and 23.9 $MtCO_2e$ in 2006 (Felix *et al*., 2014). Another cause is burning of fossil fuels for energy. Many people who live in cold countries of the world resort to fossil energy to heat their houses and for cooking. Other natural causes listed include solar activity, volcanic eruptions and changes in orbital characteristics (Fischer *et al*., 2007)

Several studies have shown that drought coupled with extreme hot temperatures and low humidity can increase the risk of wildfires (IPCC, 2012). It has also been found that the agricultural sector is more severely influenced by extreme hot temperature. This is because different crop species are very sensitive to extreme hot or cold temperatures (Hatfield *et al*., 2011).

Another research carried by Prasad *et al*.(2006) and Rodriguez-Puebla *et al*. (2007), revealed that grains and cereals produce very low yields when extremely cold or hot temperatures persist for a long time. Scientist in the agricultural sector have shown that higher night temperatures have given rise to increased respiration hence reducing the net gain in the form of grain yield. They found that current rise in temperature is likely to continue during this century and extreme events associated with the rise are expected to increase in frequency, intensity and persistency (Rasul *et al*., 2008). Furthermore, when extreme hot temperatures persist, water resources are affected as more water is demanded by both man and other living organisms (Colombo *et al*., 1999). They also found that there is increased demand for electricity during periods of extreme hot temperatures.

A lot of research has also been done on the effects of extremely high temperatures on human health, comfort and mortality (Garcia-Herrera *et al*. 2005; Tobias *et al*. 2010). An example is the high human mortality rate that was recorded in Europe during the summer heat waves in 2003 and 2010. In the 2003 heat waves and drought in Europe for instance, more than 30,000 fatalities with France losing more than 15,000 lives were recorded. It also led to the destruction of large forested areas by fires and a monetary damage of about US$14 billionfrom crop losses (Nicholls and Alexander, 2007; WHO, 2003; Koppe *et al*. 2004; Garcia-Herrera *et al*., 2008). Human beings are warm blooded, and so, we have the physiological ability to regulate our body's internal temperature which is kept at 37℃ ± 2℃. If the body's core temperature rises or falls beyond this, then serious illness or even death may occur. Exposure to extreme natural heat may also result in death because it exacerbates preexisting chronic conditions (e.g. cardiovascular, cerebral and respiratory diseases). Patients receiving psychotropic drug treatment for mental disorders and those taking medications that affect the body's heat regulatory system are more susceptible (Jeffrey *et al*. 2014).

In the 21st century, scientists have been faced with the challenging problem of global warming. This menace has affected global air temperature and rainfall (Ayuketang *et al*. 2014). The global average temperature has increased by 0.4℃ in the past 25 years and is forecast to further increase by between 1℃ and 3℃ by the year 2100 (Michelle, A., *et al*. 2010). Ghana, not being an island, therefore, needs some information on trends of temperature, to be able to forecast and take precautionary measures to curb, reduce or withstand it.

Ghana ranks high amongst African countries most exposed to risks from multiple weather related hazards such as floods, droughts and high temperatures (UNDP/NADMO 2009). For instance, flood occurrence has been an annual ritual in northern Ghana especially in the Upper East region for the past decade and it is still causing widespread destruction increasing the risk of vulnerability and general underdevelopment of the region (UNDP/NADMO 2009).

In Ghana, the Upper East Region is considered as one of the poorest regions. It consists of a poor subsistence agricultural region, where the population suffers from low levels of education and poor infrastructure and other forms of state investments. The threat of climate change which leads to very high temperatures and the variability of other climatic factors, risk making the people more vulnerable than before (GSS, 2010). Among the most vulnerable in the region are the small-scale farmers, whose livelihoods are directly dependent on the climate conditions. The region is also characterized by one erratic rainy season from May/June to September/October. It has a long spell of dry season which spans from November to somewhere April. Temperatures during this period can be as low as 15℃ at night but goes to more than 35℃ during the day (GSS, 2010).

Many research works have been undertaken worldwide concerning the effects of extremely high temperatures using extreme value theory. Hasan, *et al*. (2013) modeled extreme temperatures in Malaysia using generalized extreme value distribution and Ayuketang *et al*. (2014) used the same method in Cameroon. Again Kysely, (2002) modeled extreme temperature events using a stochastic and extreme value distribution approach in the Czech Republic. A similar work was studied by Politano, (2008) in China. However, in Ghana no one has modeled extreme temperature prediction in the Upper East Region. This study will employ some time series models together with extreme value theory on observed temperature data to predict extreme temperature behavior in the Upper East Region.

1.2 Statement of the problem

The effects of global rising temperatures have been of great concern to many people worldwide because of its negative consequences. Research conducted by various climate scientists have revealed that increase in global temperatures affect all aspects of our lives. It has impacted negatively on our agricultural production, health, economic and social wellbeing.

Firstly, in agriculture, it has been discovered that productivity of wheat and other crop species fall markedly at high temperatures in India and other parts of the world (Kothawale *et al*., 2002). Again, the impact of extreme temperature on human health has been well documented in many studies. The effects include heat exhaustion and potential death from heatstroke (Smoyer-Tomic *et al*., 2003). Another consequence of high temperatures on human health is the provision of favorable conditions for the spread of some diseases, especially cerebrospinal meningitis (Boko *et al*., 1997). In 2012, eighty one cases of the disease were recorded with several fatalities in the region during the hot and dry season (Dukic *et al*., 2012). A research carried out in the Kassena Nankana district of the Upper East Region showed that months with higher temperatures tend to co-occur with the higher meningitis incidence and months with lower temperatures recorded low meningitis incidence (Samuel *et al*., 2014).

Further, higher temperatures have implications for energy production and demand. Ghana as a developing country is still struggling to meet its energy demand. Our ability to predict extreme temperatures will serve as a guide to government to take adequate measures to meet our energy demands. The prediction of extreme temperatures is critical information which is needed to assess the impact of potential climate change on human beings and the natural environment. Such information is also vital for long-term planning at the regional and national levels for further study and analysis leading to adaptation strategies.

However, modelling and predicting rare events is statistically difficult because it is related to very high or very low values. Standard statistical theory in practice describes observations with reference to the centre of a probability distribution ignoring the tails. If the interest lies in such rare observations then it requires suitable and robust methods to quantify the distribution of the tails. It based on this that this research seeks to use extreme value theory to model and predict extreme temperatures in the Upper East Region.

1.3 General objective

The main objective of this study is to develop a model which can be used to predict extreme temperature return levels for given return periods in the Upper East Region of Ghana.

1.4 Specific objectives

i. To determine the monthly effects of extreme temperature in the region.

ii. To determine any trend that characterizes monthly maximum temperatures.

iii. To determine the return levels of extreme temperature in the region.

1.5 Significance of the study

Extreme temperature events are rare in nature but they have a strong impact on mankind. Therefore, for Ghana's adaptation policies for these extreme events, it requires the availability of precise and up-to-date estimates so that she can prepare for them. This study serves this purpose. There is overwhelming scientific evidence and consensus that climate change is largely human induced (Rosenzzweig, *et al.*, 2007; Sachs, 2008). It will serve as a guide for people to adopt behavioural change to protect the earth rather than endanger it. Finally, apart from serving as a guide for policy makers and stakeholders to mitigate climate change and globe warming, it will serve as reference material to provoke further research on not only extreme temperature but other climate change related variables that have negative consequences on our lives.

1.6 Structure of the thesis

The study is presented in five chapters. Chapter one comprises the background of the study, problem statement, objectives and the significance of the study. Chapter two examined and reviewed relevant literature on the subject matter. The methodology and data description are covered in chapter three. Also chapter four deals with the analysis, discussions and inferences drawn from the analysis of the data. The final chapter is five which covers the summary of major findings, conclusions and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

A review of the background of extreme value theory and its fields of application is given in this chapter. The chapter also discusses the causes of global warming which results in very high temperatures. Furthermore, it looks at the effects of extreme rise in temperatures on man, plants, animals and other areas. The chapter again reviewed empirical research on extreme temperature using various models. Finally, it concluded by examining the various extreme value models.

2.1 Background of extreme value theory and fields of application

Emil Julius Gumbel, a German mathematician is considered to be the founder of extreme value theory. He developed a distribution type known as Gumbel distribution which is used to find the maximum or minimum of various distributions. Historically his original focus was to predict the maximum level of a river but later the theory was found as a tool for modeling the extreme behaviour of other observed physical processes (Coles, 2001).

Fisher and Tippet (1928) first developed the statistical methods of extreme value theory and Gnedenko (1948) formalized extreme value distribution to which block maxima converges (David *et al*., 2011). Jenkinson (1955) developed generalized extreme value distribution by combining three single models namely Gumbel, Frechet and Weibull families together.

Extreme value theory has been used in the applied sciences since the 1940s mostly in areas such as engineering where extreme values can affect the failure of a given system (Castillo, 1988). An example of application in this area was the estimation of very high wind magnitudes needed to design bridges and buildings to withstand the extreme forces associated with these rarely experienced winds (Coles, 2001). A further application of extreme value models in the context of corrosion engineering are described by Scarf and Tawn (1996). In sports, Robinson and Tawn (1995) applied extreme value theory to study extreme athletic records.

Extreme value theory (EVT) also became popular in the field of hydrology for forecasting the probability of flood events (Jarvis, *et al.*, 1936; Gumbel, 1941). Furthermore, EVT has been increasingly used by insurance and financial institutions as a risk management tool to predict the occurrence of catastrophic losses such as credit default and stock market crashes (Embrechts *et al*., 1999). Again, the theory is applied to earthquake prediction, especially for estimating the magnitude of long return period earthquakes (Al-Abbasi and Fahmi, 1985).

Moreover, EVT is utilized in determining exposure to food chemicals (Coles, 2001; Tresson *et al*., 2004). Finally in climate prediction, EVT is used to forecast the probability of extreme weather events such as heavy rainfall and snowfall, heat waves and droughts (Claudette, 1992).

2.2 Causes of global warming and temperature rise

Global warming is considered a threat to the world and so is the subject of discourse at many international and local conferences. It is mainly considered as a problem caused by man through his activities. One of the main causes is the burning of fossil fuels which include coal, oil and gas. These substances release carbon dioxide into the atmosphere. The atmosphere gets polluted with the carbon dioxide which blankets the earth and traps in heat. The trapped heat then causes the earth to warm, leading to extreme temperatures (Jesper, 2005).

In addition other atmospheric greenhouse gases such as chlorofluorocarbons and their substitutes, for example, methane, nitrous oxide cause increase in global warming. Another cause in the increase of global warming is the generation of electricity from thermal plants. The generation of electricity from hydropower has become unreliable because of severe drought. Many countries have resorted to energy from thermal power which depends on fossil fuels. The combustion of these chemicals contributes to global warming (Jesper, 2005).

Apart from these, severe drought is also attributed to increases in temperatures and global warming. Emissions of fumes from factories, vehicles and other household sources increase global temperatures (Rasul *et al*., 2008). Over the past 40 years, human activities such as burning of fossil fuels for energy and transportation, changing land use, deforestation, land clearing, oil drilling, coal mining and agriculture have accelerated the release of CO2 into the atmosphere, way beyond the rate of release from natural processes. This has thereby speeded up the warming of the earth within a short period of time, to the extent that the earth's systems are increasingly unable to adapt. Finally, bush burning for farming and other activities are noted to contribute to global warming.

2.3 Effects of extreme rise in global temperatures

Extreme high temperatures have affected agriculture, health and the economies of many countries. Firstly, according to Rasul *et al*., (2008), crop growth and development is mainly a function of temperature if water is available to the optimum satisfaction. However, high temperatures both at night and day increase the rate of respiration of crops and hence affect their yield. Physiologically, higher temperatures induce higher rate of growth but overall growing periods become shorter. This shortened reproductive stage due to high temperatures limit carbohydrate accumulation resulting in overall yield reduction (Kwabena, A.K, 2013). Freezing or low temperatures, on the other hand, during the growing season damage floral buds of fruit trees. Extreme heat can also wither plants and lead to heat stress in livestock (Smith, 2001).

Also, Assan (2008) stated that an element of climate of vital importance to agricultural and rural livelihoods is temperature. He studied the trend of temperature increase in Ghana from 1961 to 2003 and found that from a low of 27.6 degrees Celsius in 1961, temperatures have increased to 29.5 degrees Celsius in 2003, showing a significant increase of 1.9 degrees Celsius. He asserted that increases in temperature lead to increases in evaporation and evapotranspiration rates and these together reduce soil moisture.

Secondly, extreme high temperatures have been link to increase in some infectious diseases, one of which is the dreaded Cerebrospinal Meningitis (CSM) (Samuel *et al*., 2014). In a study by Dudic *et al*., (2012), they found that a 10 degrees Celsius increase in the monthly average maximum temperature is associated with a 6-fold increase in the mean monthly number of meningitis cases when all other weather variables were held constant. It is found in several studies that many epidemics occur during the dry and hot seasons when temperatures are high but stop or drop significantly with the onset of the rainy season when temperatures are low but resuming again when the dry season starts (de Chabalier *et al*., 2000; Greenwood *et al*., 1995; Salih *et al*,. 1990).Additionally, several studies have identified that very high temperatures lead to high mortality rates in humans (Basu, 2009; Basu and Ostro, 2008; Kovats and Hajat, 2008). The elderly, women and children are particularly vulnerable to extreme temperatures (Hajat *et al*., 2007). People with particular diseases such as cardiovascular, respiratory problems, diabetes, and mental disorder are more sensitive to extreme temperature than healthy people (Basu *et al*., 2005; McMichael *et al*., 2008; Stafoggia *et al*., 2006).

To add, extreme high temperatures also have an indirect effect on the economy of Ghana (Ntiamoah *et al*., 2008). In Ghana a greater proportion of national income comes from agriculture, especially in the production of cocoa. Ghana is one of the world's top producers and exporters of cocoa, and the sector has played a key role in the nation economic development as it contributes about 3.4% to gross domestic product (Ntiamoah *et al*., 2008). About halve of Ghana's cocoa is grown under low shade. Due to the increase in temperature, there will be a serious reduction in the suitability within the current cocoa-growing areas. The increase in temperature will increase evapotranspiration of the cocoa trees which will result in low yields.

In addition, when extreme hot temperatures persist, water resources are affected due to the increase in evaporation and the high demand for water by both mankind and other living creatures (Kankam-Yeboah *et al*., 2010). It also results in the drying up of some rivers especially in the dry season which were hitherto perennial rivers.

Finally, increase in temperatures cause sea levels to rise as the sea water expands and ice melts. These cause floods along the coast, affecting and displacing settlements which are closed to the sea. In Ghana, the town of Keta in the Volta region is an example of towns affected by rise in sea levels (Ghana News Agency, 2015).

2.4 Empirical research on extreme temperature behaviour

Research on extreme events such as floods, droughts, rainfall and temperature has taken centre stage among many statisticians and scientists worldwide due their devastating effects when they do occur. One of the most researched into areas is extreme temperature which causes severe damage when it occurs. Many of the research on extreme events have relied on extreme value theory and statistical models to estimate and predict them, notwithstanding their rare nature.

Firstly, Brown *et al*., (2008) analyzed the observed daily temperature anomalies with regard to the normal climate from 1961 to 1990 and concluded that since 1950, extreme daily maximum and minimum temperatures have warmed over most regions of Europe showing a significant positive trend in extreme temperature anomalies for both upper and lower tails of their distributions. They applied a point process (PP) approach to model extreme temperatures in Mainland Spain, and found out that this approach provided more reliable results with less uncertainty in the estimation than other statistical models in extreme value theory, for example, the Block Maxima approach. They added that climate processes possess seasonal or trend effects due to different climatic patterns or long climate changes and these need to be accounted for. They concluded that more complex models which are non-stationary and use more data reduce uncertainties of the statistical modeling of extreme events.

Also Hasan *et al*., (2013) modelled extreme temperature behaviour in Malaysia using generalized extreme value (GEV) distribution. Extreme temperatures of several weather stations were modeled by fitting the monthly maximum to the GEV distribution. The Mann-Kendall test for trend showed the existence of trend for some stations and absence in other stations. Therefore, the annual maximum temperatures were modelled by applying stationary and non-stationary GEV distribution to the different stations. In the end the non-stationary model was the best because it explained much of the variation in the data.

Furthermore, Kysely, (2002) compared stochastic modeling with extreme value distribution approach to estimate temperature extremes (annual maxima and heat waves) in the Czech Republic. The stochastic approach consisted o modeling of time series of the daily maximum temperature using the first order autoregressive (AR(1)) model whilst the other model fitted the extreme value distribution to the sample of annual temperature peaks. They realized that the AR (1) model was able to reproduce the main characteristics of heat waves. However, they recommended that the estimated probabilities should be treated as upper limits because of the deficiencies in simulating the temperature variability inherent in the AR(1) model. They drew the conclusion that though theoretical extreme value distribution did not yield good results when applied to maximum annual lengths of heat waves and periods of tropical days, is was still the best method for estimating the probability and recurrence time annual one-day temperature extreme.

Again, Hasan *et al*., (2012) modelled extreme temperature using generalized extreme value (GEV) distribution in Malaysia using Penang as a case study. They studied extreme maximum temperatures using 10 years of data. Maximums of five different time periods, namely, weekly, biweekly, monthly, quarterly and half-yearly, were fitted to the GEV distribution. The results showed that only weekly, biweekly and monthly maximums were significant to be fitted to the GEV model. Also, both Augmented Dickey Fuller (ADF) and Kwiatkowski, Phillips, Schmid and Shin (KPSS) stationarity tests detected no stochastic trends for maximum temperatures. However, the Mann-Kendall (MK) test showed that all three selection periods had a decreasing trend. When the Kolmogorov-Smirnov and Anderson-Darling goodness of fit tests were conducted, they revealed that all three selection period maximum values converged to the GEV distribution.

Next, Politano, (2008) also contributed to the study of extreme temperature events in the Mediterranean, a region which extends over three continents, namely, Europe, Africa and Asia. The study focused on seasonal and annual trends to detect any differences in trends between annual and seasonal scales. Politano analyzed trends in daily maximum and minimum temperatures based on the assessment of five temperature indices. Eighty stations distributed across the whole Mediterranean region were selected and used for calculating temperature based climate indices. The results showed statistically significant changes in minimum and maximum temperature extremes between 1950 and 1999. Trends in temperature indices related an increase in both minimum and maximum temperatures.

In addition, Ayuketang *et al*., (2014) modeled extreme temperature in Cameroon by using a GEV distribution. The study used 20 years of annual maximum temperature provided by the Yauonde meteorological station. In order to have enough data for modeling purposes, five different selection periods, namely; monthly, bimonthly, quarterly, half yearly and yearly selection periods were used. When these selection periods were fitted to the GEV distribution, monthly, bimonthly and quarterly selection periods were the best. It also came to light that the best fit for the data was the three parameter GEV distribution model when compared with the Gumbel, Frechet and Weibull models. A normality test conducted using the Anderson-Darling test rejected the null hypothesis that the distribution of the data was normal with unspecified mean and standard deviation in favor of the alternative hypothesis at 5% significant level. The Mann-Kendall trend test showed that all the selection periods had a decreasing as time increases. Also the Kolmogorov-Smirnov and Anderson-Darling goodness of fit tests revealed that all selection periods converged to the GEV distribution with monthly maximum having the best convergence. This conclusion arrived at is similar to that by Husna *et al*., (2012).

In another study in Pakistan, Muhammed *et al*., (2013) modeled the relationship between electricity consumption and mean maximum temperature index. They developed an autoregressive integrated moving average time series forecast model for temperature index. The forecast values of the mean monthly maximum temperature showed an increasing trend. Also, a linear trend model for detecting consumption was developed in the same study as a function of temperature. Electricity consumption revealed a significant increasing trend due to increase in temperature. The forecast model revealed that electricity consumption and mean monthly maximum temperature were increasing with the passage of time.

Nury *et al*., (2012) also contributed to the study of extreme temperature in Bangladesh. They developed an autoregressive integrated moving average model which they used to carry out short-term predictions of monthly maximum and minimum in two towns of north-east Bangladesh. Based on inspection of the partial autocorrelation function (PACF) and the autocorrelation function (ACF) plots the most appropriate orders of ARIMA models were determined and evaluated using the AIC criterion. For prediction of the maximum and minimum temperatures in Syhet station, ARIMA (1, 1, 1) $(1,1,1)_{12}$ and ARIMA (1, 1, 1)$(0,1,1)_{12}$ were obtained respectively. The models for prediction of maximum and minimum temperatures at Moulvibazar station were respectively, ARIMA (1, 1, 0) $(1,1,1)_{12}$ and ARIMA (0, 1, 1) $(1,1,1)_{12}$.

In addition, Gillian *et al*., (2005) modeled the temperature variations in the Antartic Peninsula using multiple regression models with correlated errors admitting ARIMA models with non-Gaussian innovations having GEV distribution. The time series models they fitted indicated that the mean of the minimum temperatures was likely to increase over the next 50 years with temperatures going above zero degrees Celsius during summer months.

Zografia *et al*., (2012) investigated the statistical evidence of global warming by identifying shifts in seasonal mean of daily average temperatures over time and seasonal variance of temperature residuals. They proposed a seasonal mean Lasso-type technique based on a multiplicative structure of Fourier and GARCH terms in volatility. The model described well the stylized facts of temperature, namely, seasonality, inter-temporal correlations and the heteroscedastic behavior of the residuals. The application to European temperature data indicated that the multiplicative model for seasonal variance performed better in terms of the out of the sample forecast than the other models proposed for modeling temperature dynamics. They equated global warming to an increase in extreme temperatures.

Again, Xier, (2009) analyzed monthly temperature series for Stockholm by applying two statistical methods namely, a generalized linear model (GLM) and ARIMA model to fit the data for three different periods. A forecast of the monthly temperatures was compared with the true values. The results showed that seasonal ARIMA model for the series fitted the data better than the generalized linear model.

Also, Inderjeet *et al*., (2008) used the Box-Jenkins time series seasonal ARIMA approach to for predicting extreme temperature and rainfall on a monthly scales base in India. They discovered that the seasonal ARIMA model provided reliable and satisfactory predictions for extreme temperature and rainfall parameters on a monthly scale.

Kysely *et al*., (2010) also contributed to the modeling of extreme temperature in Czech Republic. They presented a methodology for estimating high quantiles of distributions of daily temperature in a non-stationary context based on the Peaks-Over-Threshold (POT) analysis with a time-dependent threshold expressed in terms of regression quantiles. Extreme value models were applied to estimate 20 year return values of maximum daily temperatures over Europe in transient global climate model (GCM) simulations for the 21st century.

The estimates based on the non-stationary extreme value models were compared with results of classical stationary POT models of temperature extremes. A comparison of scenarios of changes in the 20 year return temperatures based on the non-stationary POT models and stationary POT models was done. The results showed that the stationary extreme value POT models in temperature data from GCM scenarios yielded bias patterns. The non-stationary models, on the contrary, were robust. They concluded that the POT methodology for modeling extremes should be preferred over the block maxima approach due to the increase in the amount of data that enters the estimation.

Finally, Bommier, (2014) analyzed extreme temperatures in Uppsala in Sweden during the time period 1840-2012, using extreme value theory based on the block maxima and POT approaches. The block maxima parameter estimation was based on the GEV distribution, whilst the POT method used the generalized Pareto distribution (GPD) for parameter estimation. The temperature series data were analyzed on a monthly basis. Comparing the results of the two estimation methods for the same time series data gave different values which led to different estimation for return periods. For maxima, estimates for of the return periods were larger with the GEV model than with the POT method, whilst for the minima, the GEV model gave lower return period estimates than the POT method. The study drew the conclusion that to take into account the effect of seasonality, it is better to perform a time-varying threshold analysis on the whole series instead of cutting the temperature data into different series based on each month.

2.5 Extreme value theory models

Extreme value theory models are used for asymptotically approximating the tails of the distribution function. The traditional estimation methods based on density estimation are not very favorable in assessing extreme quantiles, since they produce good fit in areas where most of the data fall but in reality, only fewer observations fall (Bommier, 2014). There are many models used in extreme value theory each with their pros and cons.

One approach to model extreme events is called the block maxima method which constitutes the maximum order statistics. This is based on the classical types of extreme value distributions, namely the Gumbel, Frechet and Weibull distributions. Each of these distributions gives different parameter estimates (Coles, 2001). As a result, they are combined into a common distribution called the generalized extreme value (GEV) distribution. A GEV distribution is fitted to the series of extreme observations. This method is, however bedeviled with limitations, in particular, since possible dependence is not taken into account in the model. It is also criticized for its inefficient use of data as it wastes data especially if we have other data on extremes or in cases where there are variations of data within blocks. However, it is considered the best model for accounting for seasonality in the data. Also, it is able to model a sequence of random variables even when their independence is questionable (Coles, 2001).

In order to solve the problem of data scarcity for the model estimation in the block maxima approach, other better models have been developed. One of these models is the *r* largest order statistic model. This is based on the behaviour of the *r* largest order statistics within a block for small values of *r*. In this case, *r* stands for the number of peaks maximums at each block. Unfortunately, like in the block maxima approach, there is the problem of block size which amounts to a trade-off between bias and variance. Also, the number of order statistics used in each block comprises a bias trade-off: small values of r generate few data leading to high variance and large values of *r* are likely to violate the asymptotic support for the model, resulting to bias (Sanchez, 2014). Though the *r* largest order statistic model is a better approach when compared with to the block maxima method, it is unusual to have data of this form. If an entire series of, say, hourly, or daily observations is available, then better use is made of the data by avoiding altogether the procedure of blocking (Coles, 2001).

Another approach for extreme event analysis is called Poisson-General Pareto distribution (Poisson-GPD) model for excesses (Sanchez, 2014; Coles, 2001). This model is a joint distribution; the GDP, for the excesses values and a Poisson distribution for the number of excesses over a level in any given year or period. The Poisson-GPD model for excesses is related to the Peaks-Over-Threshold (POT) model. It originated in hydrology for carrying out statistical modeling of floods (Todorov and Zelenhasic, 1970).

Next is the Point Process (PP) approach for modeling extreme events. According to Sanchez (2014) the PP method provides and interpretation of extreme value that unifies all the other asymptotic models, namely the block maxima, the *r* largest order statistic and the POT models. Coles (2001), however, stated that it is only applicable to extreme data provided they satisfy the independent and identically distributed assumptions. He concluded that because of the connections between the POT and PP approaches, any inferences made using the former model could equally be made using the latter model.

Finally, the peaks-over-threshold (POT) has been developed to overcome the shortcomings of the other models described above. This procedure provides a model for independent exceedances over a high threshold. One of the advantages of the POT method is that it can make use of all the extreme data. Inference consists of fitting the generalized Pareto family to the observed threshold exceedances, followed by model verification and extrapolation. This approach contrasts with the block maxima approach through the characterization of an observation as extreme if it exceeds a high threshold (Coles, 2001). However, this method also has disadvantages. Coles, (2001) stated that the selection of the threshold process is always a trade-off between bias and variance. If the threshold is too low the asymptotic arguments underlying the derivation of the generalized

Pareto distribution model are violated. In contrast, too high a threshold will generate fewer excesses to estimate the shape and scale parameters. This will lead to high estimation variances. However, there are two methods for overcoming the issue of selection of threshold suggested by Coles (2001). One is an exploratory technique carried out prior to model estimation; the other is an assessment of the stability of parameter estimates, based on the fitting of models across a range of different thresholds. In more detail, the first method is based on the mean of the generalized Pareto distribution using a diagnostic plot called the mean residual life plots.

The second procedure for threshold selection is to estimate the model at a range of thresholds. This is also based on a diagnostic plot called stability plot. A lot of other plots have been described for overcoming the threshold selection problems. A major concern for GPD models is that extremes may occur in clusters and such data need to be declustered before they can be modelled by the GPD. Declustering involves filtering the exceedances such that they are approximately independent, generally keeping the highest observation in each cluster (Coles, 2001). According to Coles (2001), this is simple to do but it has major drawbacks. Firstly, it leads to substantial wastage of information in discarding all data except the cluster maxima. Furthermore, the results are sensitive to the arbitrariness with which clusters are determined. Consequently, declustering can introduce a lot of bias in the modelling process.

2.6 Conclusion

This chapter looked at several methods and analysis procedures for extreme value theory events. It exposed the reasons for the adoption of various modeling procedures especially in the area of climate change variables. It also revealed the main models used in extreme value theory and their properties. This serves as a guide to choose the most appropriate model to analyze extreme temperature events. From the relevant literature given, both the peaks-over threshold model and the block maxima model based on the GEV distribution are candidate models for extreme data. Both of these methods are applied to model the maximum temperature data.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter gives a brief description of the climatic conditions of the study area. It also describes some methodological approaches which are used in modelling extreme values. Further, it describes some tests which are used to test for stationarity. In addition the chapter describes maximum likelihood estimation of GEV and GPD models. Moreover, it describes some diagnostic procedures in extreme value theory. The chapter concludes with model selection criteria.

3.1 Study Area

The Upper East region (UER) lies within latitudes $10^{0}30'$ and $11^{0}15'$ North of the equator and longitudes 30 W and 18 E. It has a relatively dry climate and belongs to the Sahel and semi-arid zones. It has a geography that makes it susceptible to desertification. UER has a rainfall pattern which is unimodal and mainly occurs between April-May and September-October follows by a long dry, cold and windy season called harmattan. Temperatures during this period can be as low as 18 degrees centigrade at night, but can go to more than 35 degrees centigrade during the daytime. Humidity is, however, very low making the daytime high temperature very uncomfortable. Rainfall levels have decreased while temperature levels have risen with the tendency to experience more extreme and unreliable climate and weather variability in the future (Issahaku *et. al.*, 2016).

The Upper East Region is located in the north-eastern corner of Ghana. It is bordered to the north by Burkina Faso, the east by the Republic of Togo, the west by Sissala District in Upper West and the south by West Mamprusi District in Northern Region. (Blench, 2006)

3.2 Source of Data

The objectives of any research may not be achieved without the analysis of some form of empirical data or information. In line with this, secondary data comprising records of maximum monthly temperatures were obtained from the meteorological services department in the Navrongo municipality of the Upper East Region of Ghana. The data spanned from 1983 to 2014 with the maximum temperature for each of the twelve months in a year chosen. The data were analyzed using R version 3.2.2 with a package called *in2extreme* developed by the National Centre for Atmospheric Research, Colorado, USA, for the analysis of extreme data.

3.3 The classical extreme value distributions

The basic construction of the generalized extreme value distribution, according to Coles (2001), is as follows: Assume that the random variable X_i measures a daily, weekly, monthly or yearly quantity such as maximum temperature. Let M_n = max $\{X_1, X_2, \dots, X_n\}$ where $X_1, \dots, X_n$ is a sequence of independent random variables and M_n corresponds to the maximum of n observations with a common distribution function F then, the distribution of M_n can be determined, theoretically, for all values of n. This is given by

$$\Pr\{M_n \leq z\} = \Pr\{X_1 \leq z, \dots, X_n \leq z\}$$

$$= \Pr\{X_1 \leq z\} \times \dots \times Pr\{X_n \leq z\}$$

$$= \{F(z)\}^n \qquad (3.1)$$

According to Coles (2001), this formulation is not helpful in practice because very small discrepancies in the estimates of F from observed data which are substituted into (3.1) can lead to substantial discrepancies forF^n. If the population distribution function F were known, the distribution function of M_n could be determined exactly. However, the population distribution function is often unknown. Therefore, the distribution of M_n is approximated by modeling F^n by means of asymptotic theory of M_n. This, however, has a disadvantage that as n$\rightarrow\infty$, the distribution of M_n degenerates to a point mass at the upper end point of F. In order to avert this problem, a linear renormalization of M_n is allowed which is analogous to the central limit theorem. The linear renormalization of M_n is given by the following formula:

$$M_n{}^* = \frac{M_n - b_n}{a_n}, \text{ for sequences } a_n > 0 \; and b_n. \tag{3.2}$$

When suitable a_n and b_n are chosen the distribution of M_n is stabilized. This is formulated in a theorem called the extremal types theorem by Fisher *et al* (1928 and 1943):

Extremal Types Theorem: If there exist sequences of constants $\{a_n > 0\}$ *and* $\{b_n\}$, *as* $n \rightarrow \infty$, *such that*

$Pr\left\{\left(\frac{M_n - b_n}{a_n}\right) \leq z\right\} \rightarrow G(z)$ *where* G *a non-degenerate distribution is function then* G *belongs to one of the following families:*

$$G(z)=exp\left\{-exp\left[-\left(\frac{z-\mu}{\sigma}\right)\right]\right\}, -\infty < z < \infty; \tag{3.3}$$

$$G(z) = \begin{cases} 0, & z \leq \mu; \\ exp\left\{-\left(-\frac{z-\mu}{\sigma}\right)\right\}, & z > \mu; \end{cases} \quad (3.4)$$

$$G(z) = \begin{cases} exp\left\{-\left[-\left(\frac{z-\mu}{\sigma}\right)^{-\varepsilon}\right]\right\}, & z \leq \mu; \\ 1, & z > \mu; \end{cases} \quad (3.5)$$

For parameters $\sigma > 0, \mu \epsilon R$ and $\varepsilon > 0$. Here, ε is the shape parameter, σ is the scale parameter and μ is the location parameter. The three distributions are referred to as the extreme value distributions. Equations (3.3), (3.4) and (3.5) are known respectively as the Gumbel, Frechet and Weibull families of distributions.

The three types of distributions stated above have distinct forms of behaviour corresponding to different forms of tail behaviour for the distribution function F of the X_i. The question, therefore, becomes which of the three distributions should be used when analyzing a set of data. To solve this problem the three models were reformulated into a single model by Jenkinson in 1955 (Coles, 2001). The reformulation resulted in a model called the Generalized Extreme Value (GEV) family of distributions. This is given by the following equation:

$$G(z) = exp\left\{-\left[1 + \varepsilon\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\varepsilon}\right\} \quad (3.6)$$

This is defined on the set $\left\{z: 1 + \varepsilon\left(\frac{z-\mu}{\sigma}\right), \sigma > 0\right\}$ with ε, σ and μ representing respectively the shape, scale and location parameters. The parameters also satisfy $-\infty < \mu < \infty$, $-\infty < \varepsilon < \infty$ and $\sigma > 0$. When $\varepsilon > 0$ the distribution is known as the Frechet distribution and it has a fat tail. The larger the shape parameter is, the more fat-tailed the distribution. Also if $\varepsilon < 0$, the distribution is the Weibull distribution and finally a Gumbel distribution when $\varepsilon = 0$.

According to Coles (2001), the unification of the three families of extreme value distributions into a single family distribution simplifies statistical implementation. The type II and type III classes of extreme value distributions given by equations (3.4) and (3.5) correspond respectively to the cases where $\varepsilon > 0$ and $\varepsilon < 0$ in this parameterization. The type I class indicated by equation (3.3) corresponds to the Gumbel distribution family for $\varepsilon = 0$

Fundamental to the study of extremes is the concept of max-stability. According to Coles (2001), a distribution is max-stable if and only if it is of the same type as an extreme value distribution. The family of extreme value distributions can therefore be characterized by the class of max-stable distributions.

3.4 The Peaks-Over-Threshold model

This approach contrasts with the block maxima model through the characterization of an observation as extreme if it exceeds a given high threshold. It has the advantage over the block maxima model in that data are more efficiently used. In the block maxima model data are wasted if one block happens to contain more extreme events than other blocks (Coles, 2001; Beilant *et al.*, 2004 and Babanazarov, 2006). The events exceeding some chosen high threshold have an approximate Generalized Pareto distribution (GPD) that governs the intensity of the events.

By the Peaks-Over-Threshold (POT) method, if $X_1, X_2,\ldots X_n$ is a sequence of independent and identically distributed random variables with a marginal distribution function F, then extreme events are those of the X_i that exceed some high threshold denoted as u. If an arbitrary term in the X_i sequence is X, then a description of the stochastic behaviour of extreme events is given by the conditional probability

$$\Pr\{X > u + y/x > u\} = \frac{1-F(u+y)}{1-F(u)},\ y > 0 \tag{3.7}$$

Again, just as in the block maxima approach, since the parent distribution of F is unknown, the distribution of threshold excesses as given in equation (3.7) are unknown. In this in practical applications, approximations that are broadly applicable for high values of threshold are sought similar to the distribution of maxima of long sequences in the GEV case described earlier. This is provided by a General Pareto Distribution (GPD). The GPD asymptotic model description is given by the following theorem:

Theorem 3.1: Let $X_1, X_2, \ldots$ *be a sequence of independent random variables with a common distribution function* F*, and let* $M_n = max\{X_1, \ldots, X_n\}$*. Denote an arbitrary term in the* X_i *sequence by* X*, and suppose that* F *satisfies the extremal types theorem so that for large* n*,*

$Pr\{M_n \leq z\} \approx G(z)$*, where G(z) is the families of GEV distributions, then for large* u *the distribution function of* $(x-u)$*, conditional on x>* u *is approximately*

$$H(y) = 1 - \left(1 + {}^{\varepsilon y}/_{\sigma_u}\right)^{-1/\varepsilon} \tag{3.8},$$

defined on $\{y: y > 0\ and (1 + \varepsilon y/\sigma_u) > 0\}$ where $\sigma_u = \sigma + \varepsilon(u + \mu)$, σ_u is the GPD scale parameter which is dependent on the threshold σ and ε are the corresponding scale and shape parameters of the GEV distribution and y is the threshold excesses (Coles 2001).

In modelling extreme events by the POT method, inference consists of fitting the generalized Pareto family distributions to the observed threshold of excesses, followed by verification of the threshold and extrapolation.

3.5 Selection of Threshold in the POT model

The issue of threshold selection is similar to that of selection of block size in the block maxima approach. The choice of the threshold is not straightforward and usually a compromise has to be found. According to Bommier (2006) and Coles (2001), a high threshold value reduces the bias as this satisfies the convergence towards the extreme value theory but however increases the variance for the estimators of the parameters of the GPD, as there will be fewer data from which to estimate the parameters. A low threshold value on the other hand, results in the opposite i.e. a high bias but a low variance of the estimators, but there is more data with which to estimate the parameters. Consequently, various graphical techniques have been proposed for use in selecting an appropriate threshold. These include mean excess plot, parameter stability plot and selection based on empirical quantiles.

3.5.0 Mean excess plot

Davison and Smith (1990) suggest this graphical method for the selection of the threshold.

The method is based on the mean of the GPD given by $E(Y) = \frac{\sigma}{1-\varepsilon}$. Suppose the GPD is valid as a model for the excesses of a threshold u_0 generated by a series $X_1,\dots,X_n$. By the mean formula, then, $E(X - u_0 / x > u_0) = \frac{\sigma_{u0}}{1-\varepsilon}$ provided that $\varepsilon < 1$ and where σ_{u0} is the GPD scale parameter for exceedances over threshold u_0. The threshold stability property of the GPD means that if the GPD is a valid model for excesses over some threshold u_0, then it is valid for excesses over all thresholds $u > u_0$. Thus, for all $u > u_0, E(X - u/X > u) = \frac{\sigma_u}{1-\varepsilon} = \frac{\sigma_{u0}+\varepsilon u}{1-\varepsilon}$

Where $\sigma_u = \sigma + \varepsilon(u - \mu)$. Thus for all $u > u_0$, $E(X - u/X > u)$ is a linear function of u. Also, $E(X - u/X > u)$ is the mean of excesses of the threshold u, and can be estimated by the sample mean of the threshold excesses. This leads to the mean residual life plot defined by the locus of points

$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}\left(x_{(i)} - u\right)\right); u < x_{max}\right\}$,

where $x_{(1)},\ldots, x_{(n_u)}$ consists of the n_u observations that exceed u, and x_{max} is the largest of the X_i. If the GPD assumption is correct, then the plot should be linear with the intercept to be $\frac{\sigma_u}{1-\varepsilon}$ and slope is $\frac{\varepsilon}{1-\varepsilon}$. The mean excess plot of the data can be used to distinguish between light- and heavy-tailed models. The plot of a heavy-tailed distribution shows an upward trend, a medium tail shows a horizontal line, and the plot is downward-sloped for light-tailed data.

3.5.1 Parameter stability plot

Another graphical method which is widely used to determine the threshold u is the parameter stability plot. The idea of this plot is that if the excesses of a high threshold

u_0 follows a GPD with parameters ε and σ_{u0} , then for any threshold u such that u >u_0,

the excesses still follow a GPD with shape parameter $\varepsilon_u = \varepsilon$ and scale parameter $\sigma_u = \sigma_{u0} + \varepsilon(\text{u}-u_0)$. Letting $\sigma^* = \sigma_{u0} - \varepsilon_u\text{u}$, this new parameterization does not depend on u any longer, given that u_0 is a reasonable threshold. The plot defined by the locus of points $\{(u, \sigma^*); u < x_{max}\}$ and $\{(u, \varepsilon_u); u < x_{max}\}$, where x_{max} is the maximum of the observations. Estimates of σ^* and ε_u are constant for all asymptotic approximations. The threshold is chosen at the value where the shape and scale parameters remain constant.

3.5.2 Empirical quantiles

According to Coles (2001) and Sanchez (2014) the simplest way to select a threshold is to choose from the raw data at a specified empirical quantiles in the range of 90% to 97%. This procedure consists in choosing one of the sample points as a threshold. The choice is practically equivalent to estimation of the kth upper order statistic X_{n-k+1} from the ordered sequence

$X_{(1)}, \dots, X_n$. Frequently used is the 90% quantile, but this is inappropriate from a theoretical point of view (Scarrott and MacDonald, 2012).

3.6 Tests for stationarity

Time series data such as temperature data may be stationary or nonstationary. Stationary series are characterized by a statistical equilibrium around a constant mean level as well as a constant dispersion around that mean level (Box and Jenkins, 1976).

Nonstationary series that lack mean stationarity have no mean attractor toward which the levels tend over time. According to Greene (1997), unstable and indefinitely growing variances inherent in nonstationary series not only complicate significance tests but also render forecasting problematic as well.

According to Bommier (2014), the Peaks-Over-Threshold method is only valid for extreme if we can assume stationarity of the data. Temperature data are likely to be nonstationary as they have a tendency to show strong seasonal patterns. Temperatures may rise during the dry season but fall during the rainy season. Therefore, to prepare data that are likely to be nonstationary for any statistical modelling, it is important to subject the data to tests of stationarity to transform them to stationarity before analyzing them.

Various methods exists to transform the data before any analysis if they are nonstationary, such as by taking the natural logarithm, by taking a difference or by taking residuals from a regression. There are objective tests which are conducted to determine the stationarity or otherwise of a time series. These include graphical techniques such as autocorrelation function (ACF) and partial autocorrelation function (PACF) and unit root tests such as Augmented

Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. This study employs the ADF and KPSS tests to test for stationarity.

3.6.0 The ADF test

The Dickey-Fuller (DF) and Augmented Dickey-Fuller (ADF) tests have been used to check the stationarity and presence of unit root of a process. The DF test is considered valid for only AR(1) and when the residuals are not autocorrelated. However, when there is higher order correlation or where there is autocorrelation in the residuals, the ADF test is employed to test for unit root. The test presumes that the errors are independent of one another. In other words they are distributed as white noise and homogenous.

There are three versions of ADF which can be used to test for the presence of unit roots. These are test for a unit root, test for a unit root with drift and test for unit root with drift and a deterministic time trend. These situations have their respective equations given by

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p-1} \varphi_i y_{t-i} + \varepsilon_t \quad (3.9)$$

$$\Delta y_t = \beta_0 + \gamma y_{t-1} + \sum_{i=1}^{p-1} \varphi_i y_{t-i} + \varepsilon_t \quad (3.10)$$

$$\Delta y_t = \beta_0 + \gamma y_{t-1} + \sum_{i=1}^{p-1} \varphi_i y_{t-i} + \beta_1 t + \varepsilon_t \quad (3.11)$$

where y_t is the time series being tested, β_0 is the drift term, t is a linear trend term, β_1 is the coeffiecient of the linear trend term, φ_i are parameters of the model, $p-1$ is the number of lags which are added to ensure the model residuals are white noise. Also $\Delta y_t = y_t - y_{t-1}$.

The null and alternative hypotheses are respectively

H_0:Ɣ = 0 (nonstationary)

H_1: Ɣ <0 (stationary)

The test statistic is $\tau = \frac{Ɣ}{\sqrt{var(Ɣ)}}$.

This value is compared to the corresponding critical value at different significant levels.

3.6.1 The KPSS test

Kwiatkowski *et al.*,(1988) proposed an alternative test in which y_t is assumed to be stationary under the null hypothesis. The test can be computed by firstly regressing the dependent variable y_t on a constant and a time trend variable t. KPSS assesses the null hypothesis that a time series is trend stationary against the alternative that it is a nonstationary unit root process. The test uses the structural model

$$y_t = c_t + \delta t + u_{1t} \tag{3.12}$$

where $c_t = c_{t-1} + u_{2t}$, and δ is the trend coefficient, u_{1t} is a stationary process, and u_{2t} is an independent and identically distributed process with mean zero and varianceσ^2. The null and alternative hypotheses are respectively

$$H_0: \sigma^2 = 0$$

$$H_1: \sigma^2 > 0$$

The null hypothesis implies that the random walk term is constant and acts as the model intercept and the alternative hypothesis introduces the unit root in the random walk. The test statistic is given as

$$\text{KPSS} = \frac{\sum_{t=1}^{T} s_t{}^2}{s^2 T^2} \tag{3.13}$$

$T$is the sample size, s^2 is the estimate of the long-run variance and s_t sum of the errors.

3.7 Model selection criteria

When there are two competing candidate models for a set of data, it important to subject them to test to see which of them better fits the data well. There are many measures that can be used for estimating how well the model fits the data. Two of these models employed in this study are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC is a measure which uses the log-likelihood but adds a penalizing term associated with the number of variables. The fit of a model can be improved by adding more variables. As a result the AIC tries to balance the goodness-of-fit versus the inclusion of variables in the model. The AIC is given as

$$\text{AIC}=2\text{k}-2ln(L) \tag{3.14}$$

where,k is the number of unknown parameters included in the model,lnL is the log-likelihood function of the model.If the model errors are normally and independently distributed and n the number of observations with the residual sum of squares (RSS) defined byRSS= $\sum_{i=1}^{n} \hat{\varepsilon}_i{}^2$,then the AIC is redefined as

$$\text{AIC}= 2k + n[ln(2\pi RSS/n) + 1] \tag{3.15}$$

A variation of the AIC which is used to test model fit is the Consistent AIC which includes an added penalty for models that have a greater number of parameters, k, for a given sample size n. It is defined by

$$\text{CAIC}= -2lnL + [ln(1) + 1] \tag{3.16}$$

Another comparative fit measure which is the Bayesian information criterion (BIC). It is given by the following formula

$$\text{BIC}= 2\ln(L) + kln(n) \qquad (3.17)$$

where,n is the sample size$k$ is the number of free parameters to be estimated

L is the maximized value of the log-likelihood function for the estimated model

Under the assumption that the model errors or disturbances are normally distributed, the BIC becomes

$$\text{BIC}= \quad nln\left(\frac{RSS}{n}\right) + kln(n) \qquad (3.18)$$

The model with the lowest BIC value is preferable.

3.8 Model Diagnostics

The reason for fitting any statistical model to data is to draw inferences about some aspects of the population from which the data were drawn. Since conclusions drawn can be sensitive to the accuracy of the fitted model, it is important to check that the model fits the data well. The reliability of fitting a result of a statistical model is assessed by goodness-of fit methods. In extreme value modelling there are various methods which are used for model diagnostics. These include probability-probability (PP) plots, quantile-quantile (QQ) plots, return level plots and empirical distribution function plots. A brief description of the first three plots is given below.

3.8.0 QQ plot

The QQ plot graphs the sample quantiles against the theoretical quantiles of the distribution F and then a visual check is made to see whether or not the points are closed to a straight line. This plot if it fits the data well should converge to a straight line as the sample size increases. It is a graphical device used to test of goodness of fit of a sample X_1 ,…, X_n to some distribution F in an exploratory way. It measures how close the sample quantiles are to the theoretical quantile. Rather than considering individual quantiles the QQ plot considers the sample quantiles against the theoretical quantiles of a specified target distribution F.

For a quantile plot, a given ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ from a population with estimated distribution function $\hat{\square}$, has a quantile plot consisting of the points $\left\{\left(\hat{F}^{-1}(\frac{i}{n+1}\right), x_{(i)}) : i = 1, \dots, n\right\}$.

3.8.1 Probability-probability plot

P-P plot (probability-probability plot), also known as percent-percent plot, is a graphical technique to assess if a fitting result of probability distribution is a reasonable model by comparing theoretical and empirical probability. Given an ordered sample of independent observations $x_{(1)} \leq x_{(2)} \leq \cdots x_{(n)}$ from a population with estimated distribution function $\hat{F}$, then a probability plot consists of the points $\left\{\left(F(x_{(i)}\right), \frac{i}{n+1}) : i = 1, \dots, n\right\}$. If $\hat{F}$ is a reasonable model for the population distribution function, the points of the probability plot should lie in a straight line.

3.8.2 Return level plot of GEV distribution

Return level plots are considered convenient for both presentation and validation of extreme value models. The return level plot for the GEV distribution is stated in equation (3.19) which is obtained by inverting the GEV distribution function.

$$z_p = \begin{cases} \mu - \frac{\sigma}{\varepsilon}[1 - \{-log(1-p)\}^{-\varepsilon}], \ for \ \neq 0 \\ \mu - \sigma log\{-log(1-p)\}, \ for \ \varepsilon \ = 0 \end{cases} \tag{3.19}$$

If $-\log(1-p) = y_p$ so that the equation becomes after substitution,

$$z_p = \begin{cases} \mu - \frac{\sigma}{\varepsilon}\left[1 - \{y_p\}^{-\varepsilon}\right], \ for \ \neq 0 \\ \mu - \sigma log\{-log(y_p)\}, \ for \ \varepsilon \ = 0, \end{cases} \tag{3.20}$$

a plot of z_p against y_p on a logarithmic scale will produce a straight line when ε= 0 but has no finite limit when ε is less than zero.

3.8.3 Return levels of GPD

It often of interest to evaluate return levels, for example the N-year return level denoted by x_N that is exceeded once in N years (Coles, 2001). Given that the GPD model with parameters ε and σ is most suitable for exceedances, then the probability of exceedances of a variable X over a high threshold u is written as

$$\text{P}\{X > x/X > u\} = \left[1 + \varepsilon\left(\frac{x-u}{\sigma}\right)\right]^{-1/\varepsilon} \tag{3.21}$$

provided $x > u$ and $\varepsilon \neq 0$.

Also letting $\pi_u = p\{X > u\}$ where π_u represents the probability of occurrence of an excess of a high threshold u where the subscript u emphasizes that this value depends on the choice of threshold u then

$$P\{X > u\} = \pi_u \left[1 + \varepsilon\left(\frac{x-u}{\sigma}\right)\right]^{-1/\varepsilon} \quad (3.22)$$

Therefore, the level x_m that is exceeded on average once every m observations will be obtained by solving the following equation

$$\pi_u[1 + \varepsilon(x - u/\sigma)]^{-1/\varepsilon} = \frac{1}{m} \quad (3.23)$$

Rearrangement of equation 3.23 results in

$$x_m = u + \frac{\sigma}{\varepsilon}\left[(m\pi_u)^{\varepsilon} - 1\right] \quad (3.24)$$

Equation 3.23 according to Coles (2001) is only valid when m is large enough to ensure that $x > u$.

In the case where $\varepsilon = 0$, applying the same preceding procedures to the GPD leads to

$$x_m = u + \sigma log(m\pi_u) \quad (3.25)$$

where x_m is called the m-observation return level. In the case where the N-year return level is of interest and given that the number of observations per year is n_y, then

$$m = Nn_y \quad (3.26)$$

Hence, the N-year return level is given by

$$x_N = \begin{cases} u + \frac{\sigma}{\varepsilon}\left[\left(Nn_y\pi_u\right)^{\varepsilon} - 1\right], & \varepsilon \neq 0 \\ u + \sigma log\left(Nn_y\pi_u\right), & \varepsilon = 0 \end{cases} \quad (3.27)$$

An estimate of π_u is also given the

$$\pi_u = \frac{k}{n} \tag{3.28}$$

where k is the number of exceedances and n is the sample size. Coles (2001), stated that the number of exceedances of u follows a Binomial distribution, Bin(n, π_u).

CONCLUSION

This chapter looked two main tests for testing for stationarity, namely the ADF and KPSS tests. Further, it described the block maxim and peaks over threshold models for modelling extreme temperature. It also dealt with various diagnostics for the models and their return levels.

CHAPTER FOUR

ANALYSIS AND DICUSSION OF RESULTS

4.0 Introduction

In this chapter, the analysis of data and discussion of results obtained from the study were presented. The chapter is divided into three main subsections, namely; preliminary analysis, further analysis and discussion of results. The main packages used for the analysis in R were *in2extReme, evd, evir* and *extremes* developed by Gilleland *et al.,* (2014) for the analysis of extreme data.

4.1 Preliminary analysis

In this section, the descriptive statistics of the maximum temperature returns from January 1983 to December 2014 were explained. The maximum and minimum values of the maximum temperature returns for the entire period were 0.1671 and -0.1640 respectively. Also, the average maximum temperature returns for the whole period was 0.0002. Again, the returns showed positive skewness with a value of 0.0400 but platykurtic in nature because the excess kurtosis is negative (-0.6900). Further, the coefficient of variation is 35873.3300. This large value means that there was increased variability in the returns of the maximum temperature. These details are displayed in Table 4.1.

Table 4.1: Descriptive statistics for maximum temperature returns

Variable	Mean	Min	Max	CV	Skewness	Kurtosis
Returns	0.0002	- 0.1640	0.1671	35873.33	0.0400	-0.6900

To further investigate the characteristics of the monthly maximum temperature returns, a descriptive analysis was conducted. The analysis showed that April has the highest of the maximum temperature returns and the smallest occurred in June as shown in Table 4.2. Also, the minimum and maximum temperature returns occurred in June and April respectively. Again, the month of March shows more variability since it has the highest coefficient of variation and January has the least variability. In terms of skewness, January, February, March, May, August and November were negatively skewed whilst the rest of the months were positively skewed. Finally, Table 4.2 showed that May, June, August, September and October were platykurtic and the rest of the months were leptokurtic in nature. In terms of extreme value modelling, a negative or positive kurtosis value for the returns represents a light or heavy tail distribution. From Table 4.2, April, November and March have the heaviest tails while August and June have the lightest tails.

Table 4.2: Monthly descriptive statistics for maximum temperature returns

Month	Mean	Min	Max	CV	Skewness	Kurtosis
January	-0.0111	-0.0993	0.0651	-361.11	-0.42	0.31
February	0.0677	-0.0304	0.1404	-110.20	-0.62	0.16
March	0.0404	-0.1521	0.1412	117.01	-1.88	8.63
April	-0.0141	-0.0669	0.1671	-300.26	2.68	10.28
May	-0.0732	-0.1384	-0.0132	-39.25	-0.05	- 0.08
June	-0.0843	-0.1640	0.0000	-49.04	0.23	-0.34
July	-0.0553	-0.1323	0.0541	-59.10	0.93	3.44
August	-0.0239	-0.0601	0.0096	-84.89	-0.32	-1.00
September	0.0281	-0.0158	0.0706	74.57	0.36	-0.01

October	0.0856	0.0272	0.1406	32.29	32.29	-0.14
November	0.0725	-0.0721	0.1117	47.86	-2.50	9.02
December	-0.0307	-0.1071	0.0604	-110.20	0.50	1.21

A scatter plot of the raw maximum temperatures for each year from 1983 to 2014 was plotted as shown in Figure 4.1. The plot does not give any strong evidence that the pattern of variation in temperature has changed over the observation period but a careful look at the plot suggests that the pattern of variation in temperatures has not remained constant. There is a discernible increase in the data over time. The increase seems slighter in more recent years. From the plot, between 1980 and 1990, there was a decrease in maximum temperatures. Thereafter, temperatures rose higher across the years with the highest occurring between 2000 and 2005.

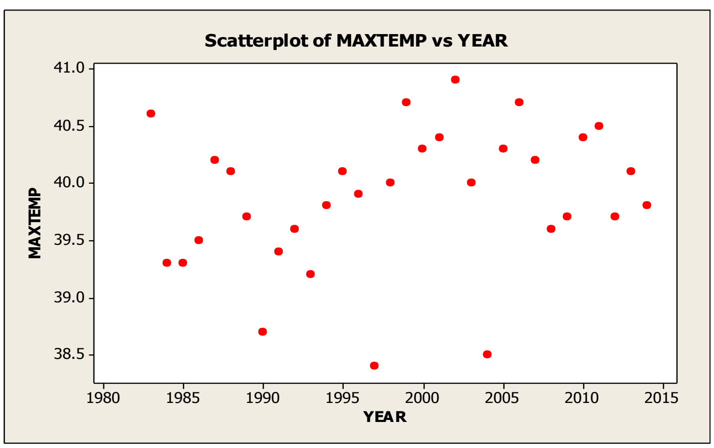


Figure 4.1: scatter plot of maximum annual temperature

To further explore the data, the maximum temperature returns were plotted against time in years. From the plot shown in Figure 4.2, there was no clear evidence of significant trend.

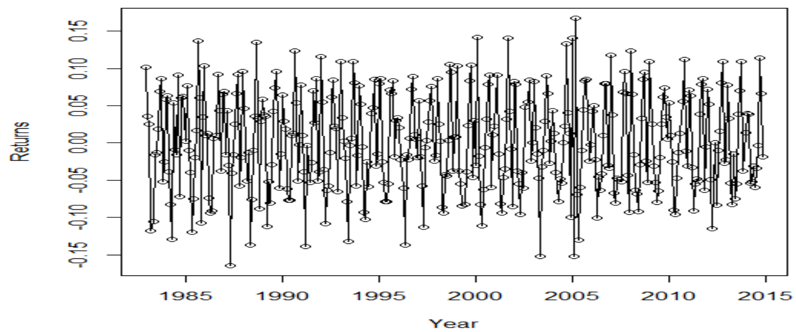


Figure 4.2: Time series plot of returns against year.

An investigation was conducted on the kind of trend that the maximum temperatures exhibit. As a result, linear, log-linear, quadratic and log-quadratic trends models were employed to determine the type of trend. Table 4.3 gives the AIC, BIC and adjusted R-squared values. The results revealed that the log-quadratic trend model was the best model because it has the lowest AIC, BIC and highest adjusted R-squared values.

Table 4.3 Trend analysis of maximum temperature

Model	Adjusted R-squared	AIC	BIC
Linear	-0.0010	-793.7009	-785.7996
Quadratic	0.9985	-557.0181	-545.1662
Log-linear	-0.001942	1936.5670	1944.468
Log-quadratic	0.9992*	-813.2593*	-801.4074*

*: means best model on the selection criteria

The parameters of the log-quadratic trend model were estimated. As shown in Table 4.4, all the parameters were significant at the 5% level of significance. The quadratic trend model revealed that the maximum temperatures did not exhibit linearity. The R-squared was about 99.9% which is a reflection that the trend is responsible for a large part of the variation in the maximum temperatures. Thus a model for the quadratic trend model is given by

$$lnmaxtemp = -27.1363 - 0.0048T + 4.9229T^2$$

where T is time.

Table 4.4: Estimated parameters of the log-quadratic trend model

Variable	Coefficient	Standard error	T-statistic	*P*-value
Constant	-27.1363	0.0889	-305.1314	0.0000
Time	-0.0048	0.0038	-0.0124	0.0001
$(Time)^2$	4.9229	0.0070	702.1311	0.0000
AIC = -813.2593	R-Squared = 0.9992			
BIC = -801.407	Adj. R-Squared = 0.9992			

In order to measure the monthly changes in temperature, the maximum temperature returns was regressed on a linear trend and full set of dummy variables excluding an intercept. The insignificance of the coefficients does not really matter as they depict incremental month effects of the year. From Table 4.5, the months of April through to August and December saw significant negative seasonality in maximum temperature returns whilst the rest of the months exhibited significant positive seasonality.

Table 4.5: Estimates of regression on linear trend and periodic dummies

VARIABLE	**Coefficient**	***Std. Error***	***T-statistic***	***p-value***
January	-0.0111	0.0063	-1.7502	0.0809
February	0.0677	0.0062	10.8989	0.0000*
March	0.0404	0.0062	6.4972	0.0000*
April	-0.0141	0.0062	-2.2695	0.0238*
May	-0.0732	0.0062	-11.7732	0.0000*
June	-0.0843	0.0062	-13.5678	0.0000*
July	-0.0553	0.0062	-8.8917	0.0000*
August	-0.0239	0.0062	-3.843	0.0001*
September	0.0281	0.0062	4.5199	0.0000*
October	0.0856	0.0062	13.7682	0.0000*
November	0.0725	0.0062	11.6669	0.0000*
December	-0.0306	0.0062	-4.9316	0.0000*

*means significant at 5% level of significance

Since the estimated coefficients of the dummy variables in Table 4.5 are incremental month effects of each year, their significance does not matter. Hence using Halvorsen and Palmquist (1980) approach of interpreting differential coefficients in semi-logarithmic equations, the differential coefficients were transformed to show differential effects in terms of percentage change. The effect for each month, for all January, February, through to December for the entire period (from 1983 to 2014) were calculated by using an exponential transformation and then converted into a percentage change. From the results, the months of January, April, May, June, July, August and December decreased the percentage maximum temperature returns by 1.1039, 1.4001, 7.0585, 8.0845, 5.3799, 2.3617 and 3.0137 respectively in maximum temperature returns. However, the months of February, March, September, October and November indicated percentage increments in maximum temperature returns of 7.0044, 4.1227, 2.8499, 8.9370 and 7.5193 respectively. Table 4.6 displays these results.

Table 4.6: Monthly effects of maximum temperature returns

Month	Coefficients	Percent effect
January	-0.0111	-1.1039
February	0.0677	7.0044
March	0.0404	4.1227
April	-0.0141	-1.4001
May	-0.0732	-7.0585
June	-0.0843	-8.0845
July	-0.0553	-5.3799
August	-0.0239	-2.3617
September	0.0281	2.8499
October	0.0856	8.9370
November	0.0725	7.5193
December	-0.0306	-3.0137

NB: Effect of January = $(e^{-0.0111} - 1) \times 100\%$

4.2 Further analysis

4.2.1 Unit root tests

In order to test the stationarity of the maximum temperature returns, KPSS and ADF tests were conducted. The KPSS test results shown in Table 4.7 revealed that the returns series were stationary.

Table 4.7: KPSS test of maximum temperature returns in level form

Test	Test Statistic	Critical value
KPSS	0.0485691	0.462

To further confirm that the maximum temperature returns were stationary, an ADF test with a constant and a constant with quadratic trend was conducted. The test results in Table 4.8 revealed that the maximum temperature returns were stationary at the 5% significance level.

Table 4.8: ADF test for maximum temperature returns in level form

Test	Constant		Constant+ Trend	
	Test Statistic	*P*-value	Test Statistic	*P*-value
ADF	-4.6738	0.0000	-5.27798	0.0000

4.2.2 Fitting the generalized extreme value (GEV) distribution

The stationary maximum temperature return series was modelled using the generalized extreme distribution. Table 4.9 displays the parameter estimates, AIC and BIC values of the fitted model. The negative value of the shape parameter suggests that the Weibull distribution which is in the family of GEV distributions fits the data well, tentatively.

Table 4.9: Parameter estimates for GEV distribution

Parameter	Estimate	Standard error	AIC	BIC
Location(μ)	-0.02257	0.00368	-1003.605	-991.760
Scale (σ)	0.06501	0.00267		
Shape (ε)	-0.29423	0.03480		

Negative log-likelihood value of GEV estimated: -504.8023

The GEV distribution is given by

$$z_p = -0.0226 + \frac{0.0650}{0.2942}[1 - \{-log(1-p)\}^{0.2942}] \ .$$

Based on the GEV distribution model, return periods of 2, 20 and 100 years were also estimated. The return levels corresponding to return periods of 2, 20 and 100 years were 0.0000, 0.1062 and 0.1429 respectively. The return levels clearly show an increasing trend as the years increase. However, return periods corresponding to long periods such as 100 are often considered and those corresponding to shorter periods such as 2 and 20 ignored (Porcenich, 2002).

Some graphical techniques were then employed to ascertain the fitness of the GEV distribution to the data. The plots used were the QQ-plot, density plot, return level plot and a QQ-plot based on randomly generated data from the fitted GEV distribution function as shown in Figure 4.3. Figure 4.3 (a) shows a QQ-plot of the empirical data quantiles plotted against those derived from the GEV distribution function. The plot is reasonably a straight line along the unit diagonal which indicates that the asymptotic assumption for using the GEV distribution function is met. Figure 4.3 (b) shows a QQ-plot for randomly generated data from the fitted GEV distribution function, against the empirical data quantiles along with 95% confidence bands, a 1-1 line and a fitted regression line. All the points lie in a straight line and within the confidence interval. This plot is similar to the QQ-plot in the upper left panel, (a) and, therefore, confirms that the GEV model is an appropriate model for the data.

To further buttress the fit of the GEV distribution, the density plot shown in figure 4.3 (c) shows good agreement between the empirical density shown by the solid black line and that of the fitted GEV distribution function shown by the blue line. Finally, the return level plot is shown on the lower right panel (Figure 4.3 (d)). It is plotted on the log scale so that a heavy tail case is concave, a bounded-tail case is convex and the light-tail case is linear. A visual inspection shows that it is convex which suggests that the shape parameter is negative. This confirms the negative value in the shape parameter as given in Table 4.9. The convexity of the return level plot also shows that the distribution that best fits the data from the GEV distribution family is of the Weibull type distribution which has an upper bound. Furthermore, the return level plot shows the return periods in years for 2, 5, 20, 100, 500 return periods.

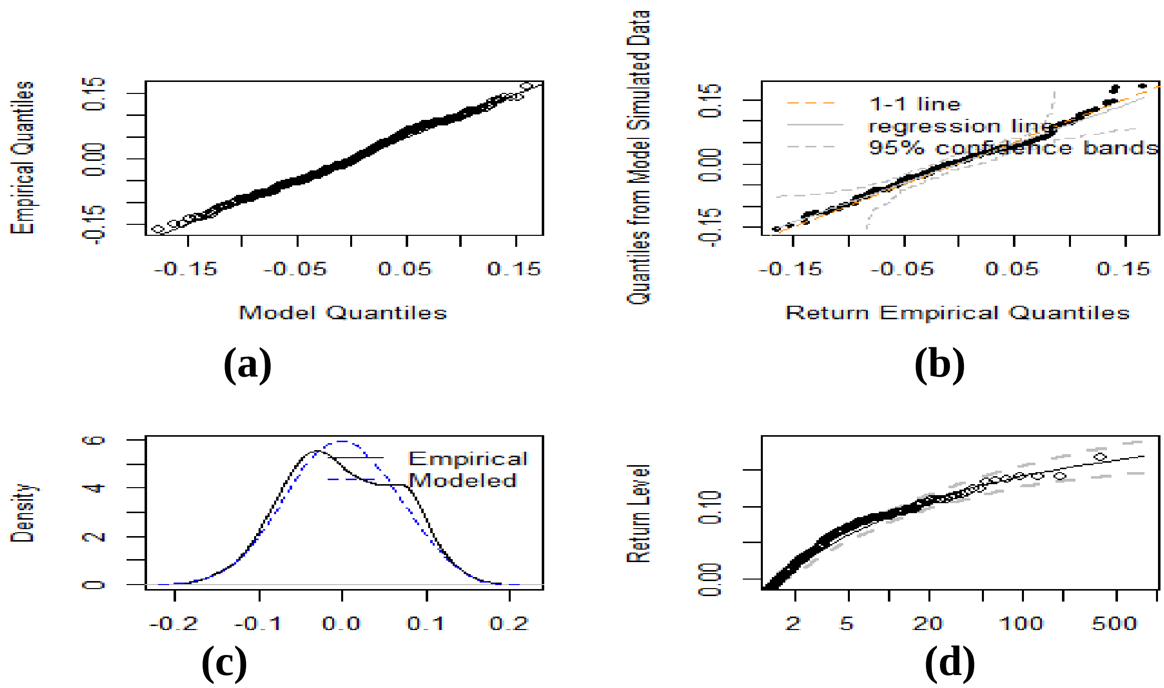
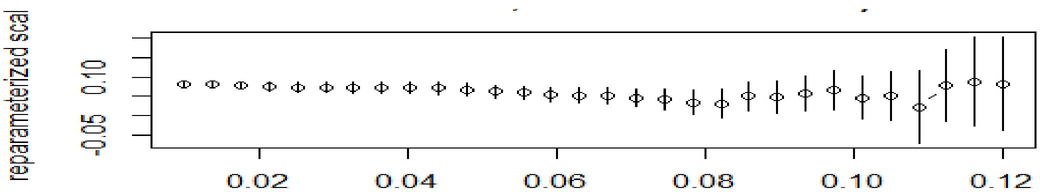


Figure 4.3 Diagnostic plots of the GEV distribution

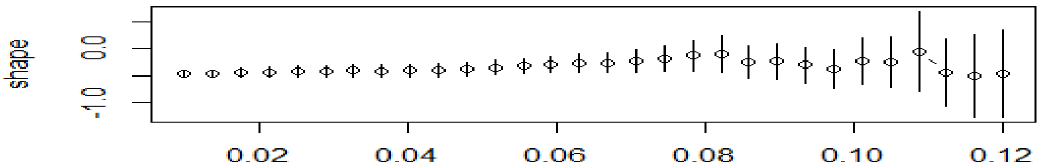
4.2.3 Fitting the Generalized Pareto distribution (GPD)

Next, the GPD was fitted to the maximum temperature returns. The first step in fitting the GPD is the selection of an appropriate threshold level for the tails of the distribution. Two methods for selecting an adequate threshold, namely parameter stability plot and mean residual life plot were used in the selection of the threshold. Firstly, in the parameter stability plot a range of thresholds were arbitrary selected with the nature of the data in mind for a total number of 30 thresholds. Based on variation of the minimum and maximum thresholds using the *in2extreme package* in R, it was revealed that a minimum threshold of 0.01 and a maximum threshold of 0.12 yielded the best stability of the parameters. The choice of minimum threshold and maximum threshold is based on the output generated by the software for arbitrary choices made based on the data and examining the stability plot. The stability plots shown in Figure 4.4 (a) and (b) show that the parameter estimates do not appear to vary considerably for even small values.

Vertical line segments display the 95% normal approximation confidence intervals for each parameter estimate. Therefore, any threshold that lies between 0.01 and 0.12 could be used to fit the GPD. Based on this criterion a threshold of 0.07 was chosen bearing in mind the fact that too high a threshold will result in few excesses that will result in a large variance.



(a)



(b)

Figure 4.4 Parameter stability plots

The second method which was used to choose the appropriate threshold is the mean excess plot. The idea behind it is to find the lowest threshold whereby a straight line could be drawn from that point to higher values and still be within the uncertainty bounds indicated by gray dashed lines in Figure 4.5. The plot confirms that 0.07 or any value slightly greater than this value is a good threshold choice. The downward behaviour of the plot also suggests a light-tail distribution.

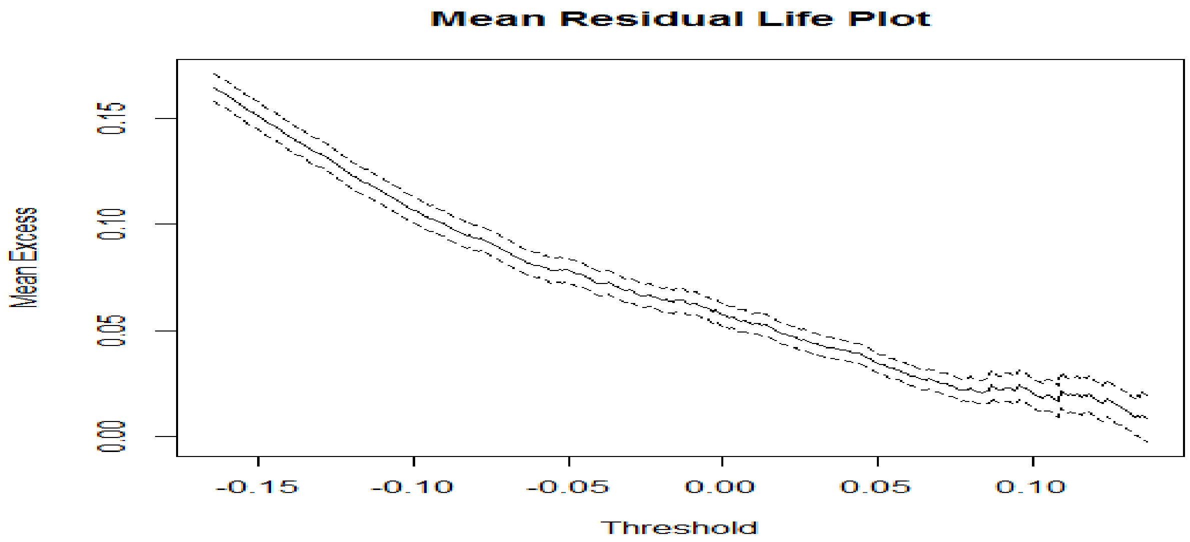


Figure 4.5 Mean residual life plot

Using the threshold value of 0.07 the estimated parameters of the GPD are shown in table 4.10. The shape parameter which is dominant in determining the qualitative behaviour of the GP distribution is negative. The value of the shape parameter of the GPD (-0.23359) is almost the same as the estimated shape parameter value (-0.29423) in the GEV estimate. This shows that the distribution of excesses has an upper bound or upper end point and also is short-tailed. Just as in the GEV case, this is a Weibull distribution in the family of the generalized Pareto distributions.

Table 4.10: Parameter estimates for GP distribution

Parameter	Estimate	Standard error	AIC	BIC
Scale (σ)	0.03114	0.00483	-374.403	-369.906
Shape (ε)	-0.23359	0.10377		

Negative log-likelihood value of GPD:-189.2017

Hence using the estimated parameters of the GPD model and the chosen threshold of 0.0700, the model can be express by the equation $H(y) = 1 - \left(1 + \frac{-0.23359y}{0.3114}\right)^{1/0.23359}$ (3.29),

where y=$X_i - 0.0700$

To further confirm that the threshold selected is good to use in fitting the GPD, diagnostic plots were plotted based on the selected threshold of 0.0700. Figure 4.6 (a) to (d) indicate that the assumptions for fitting the GPD to excesses over threshold were met. The diagnostic plots agreed with those of the GEV distribution function. The QQ-plot figure 4.6 (a) shows that all the points are approximately linearly distributed along the unit diagonal showing a good fit of the GPD for maximum temperature returns. This agrees with QQ-plot generated by randomly selected data from the GPD against the empirical quantiles in figure 4.6 (b). The empirical density plot in (a) also affirms how adequate the GPD is in terms of modelling the data. It was observed that the number of excesses is 70 for the chosen threshold. The return level plot in (d) is also convex as in the GEV distribution case. Apart from a few points at the upper portion which show departure, the rest of the points lie on the line.

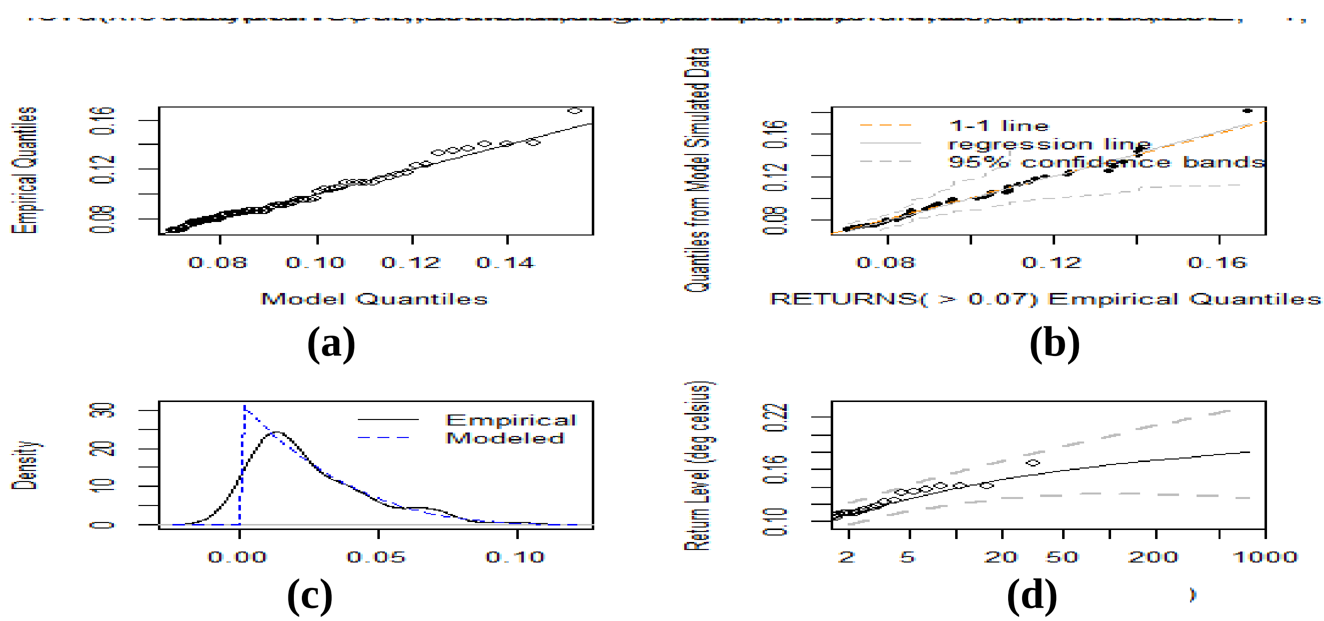


Figure 4.6 Diagnostic plots for GPD

4.2.4 Comparative analysis of the GPD and GEV models

The diagnostic plots suggest that both the GEV and GPD fit the data of maximum temperature returns well. However, their AIC and BIC values which are shown in Tables 4.6 and 4.7 clearly revealed that the GEV distribution is superior in fitting the data because it has the lowest AIC and BIC values. In order to validate this conclusion, further goodness-of-fit tests were conducted using two nonparametric tests, namely, Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) tests. Table 4.11 shows the ranks of the GEV and GP distributions based on the two tests. The results indicate clearly that the GEV distribution is best based on the ranks.

Table 4.11: KS and AD tests for GEV and GPD

Distribution	Kolmogorov-Smirnov		Anderson-Darling	
	statistic	rank	statistic	rank
GEV*	0.0447*	1*	0.981	1*
GPD	0.0572	2	57.732	2

***: means best based on the ranks**

Also the likelihood ratio test was used to compare the fit of the two models to the data where the null model is the GPD with two parameters and the alternative model is the GEV distribution model with three parameters. The test yielded a *p*-value of 0.0000 which is less than the 5% level of significance and hence the GPD model was rejected in favour of the GEV model. The critical value of the chi-square distribution with one degree of freedom was 3.8415 and the test statistic was 631.2. This further affirms the rejection of the null hypothesis in favour of the alternative as shown in Table 4.12

Table 4.12 Likelihood ratio test

Test Statistic	Critical value	Significance level	*P*-value
631.2000	3.8415	0.0500	0.0000

It is therefore concluded that the GEV distribution model is best in forecasting long return periods of extreme temperature.

4.3 Discussion of Results

The descriptive statistics revealed that the data were positively skewed but platykurtic in nature. Clearly, this indicates that the distribution of the maximum temperature returns is distributed far away from the mean. Also the large value in the coefficient of variation showed that there was increased variability in the maximum temperature returns. Further, the results revealed that there was trend in the data. Among the trend models the one that best described the data was the log-quadratic trend model. It explained about 99.8 percent of the variation in maximum temperatures. The log-quadratic trend model implies that the maximum temperatures were not stable over time.

To determine the monthly effects in maximum temperature within a year the maximum temperature returns was regressed on a linear trend with full set of periodic dummy variables. The monthly effects expressed as a percentage revealed that the months January and April through to August and December decreased the temperature returns by 1.1, 1.4, 7.1, 8.1, 5.4, 2.4 and 3.0 respectively. The decrease in temperature in January and December could be attributed to severe harmattan which is associated with these months. The significant decrease in May, June and July could be due to the rainfall associated with these months which usually result in lower temperatures. The increase in March corresponds to the heat period of the year when temperatures rise before the rainy season which begins in late April. Further it is observed that by October the rainy season comes to an end so that temperatures begin to increase slightly. This however does not last long before the harmattan sets in again by late November. The results corroborated a study by Brown *et al.*, (2008) who analyzed daily maximum temperatures in Europe from 1961 to 1990 found out that since 1950 there have been gradual increase in daily temperatures. The results are also consistent with Hasan*et al*., (2012) who conducted a study using maximum temperature in Malaysia. Their study also revealed that the GEV model was more flexible in modelling weather data than the GPD model.

The prediction of the occurrence or reoccurrence of extreme events such as temperature cannot be overemphasized because their coming often results in disastrous outcomes. Extreme high temperatures affect agriculture, health and even the energy sector. Diseases such as cerebrospinal meningitis are favoured by high temperatures. Also crops may wither and more energy may be demanded to cool residential areas and offices. All these drain the coffers of the economy.

Two extreme value theory models were employed to predict long return periods of extreme temperature. The models were the generalized extreme value distribution and the generalized Pareto distribution. After taking the models through diagnostic checks both models proved to be appropriate for predicting long return periods of extreme temperature.

Furthermore, to identify which of the two models has the best prediction ability for long return periods of extreme temperature; a comparative analysis between the two models was carried out using two nonparametric tests based on ranks. These were the Kolmogorov-Smirnov and

Anderson-Darling tests. The tests revealed that the GEV distribution was better than the GPD model. Again, in terms of information criteria the GEV distribution had the lowest values. Therefore a likelihood ratio test was performed to check whether there was significant difference in the performance of the two models in predicting extreme temperature based on their negative log-likelihood values. The results of the test revealed that the GEV distribution model was better than the GPD in terms of modelling extreme temperature return levels for long return periods.

4.4 Conclusion

The chapter discussed the results of the study. It analyzed and presented the details of the major findings of the study.

CHAPTER FIVE

CONLUSION AND RECOMMENDATION

5.0 Introduction

This chapter presents the conclusion and recommendations of the study. It is divided into two subsections, namely, conclusion and recommendations.

5.1 Conclusion

In this study, the monthly maximum temperatures from January, 1983 up to December, 2014 were studied using two extreme value distribution models. Before fitting the models to the data, the maximum temperature data were transformed logarithmically in order to ensure that they are stable.

A plot of the maximum temperature against time did reveal any noticeable trend. However, a trend analysis revealed that the maximum temperature values followed a log-quadratic trend. An exploration of the monthly effects on the maximum temperature returns showed that temperature increases during some months in the dry season (February, March, October and November) but decreases during the rainy season (April, May, June, July, August) and harmattan (December and January).

Between the two models developed, namely, the generalized extreme value and generalized Pareto distributions, only the former was adequate based on evidence from diagnostic plots, model stability checks and model comparison techniques. The return levels revealed that temperatures are rising and can reach unbearable levels in the far future. A trend analysis also showed that the maximum temperature returns followed a loq-quadratic trend model. The percentage effect of each month's temperature showed that there was a decrease in cold seasons and increase in hot seasons but with few anomalies where some months recorded increases even within the cold seasons.

5.2 Recommendations

The following recommendations were made based on the outcome of the study;

i. Farmers should be encouraged on climate change adaptation measures through agro-forestry, soil and water management. Alternative sources of water for crop irrigation should be encouraged such as water harvesting culture during the rainy season for later use.

ii. The Ministries of Land and Natural Resources and Agriculture should design education programmmes on local radio to sensitize people especially famers on best practices that can increase their farm outputs while at the same maintain the natural environment. Crops which are tolerant to adverse climatic conditions should be cultivated.

iii. It is further recommended that the study be replicated in other regions to assess the extent of global warming so that mitigation measures could be adopted to reduce it.

REFERENCES

Ayuketang, N., and Joseph, E., (2014). Modelling Extreme Temperature in Cameroon using Generalized Extreme Value Distribution. University of Buea and AIMS-Cameroon. Unpublished M. Sc. Thesis.

Al-Abbasi, J. N., and Fahmi, K. J., (1985). Estimating Maximum Magnitude Earthquakes in Iraq Using Extreme Value Statistics.*Geophysical Journal International***82** (3): 535-548.

Assan, J. K., (2008). Generational Differences in Internal Migration: Derelict Economies Exploitative Employment and Likelihood Discontent, *International Development planning Review*, **301**(**4**): 377-398.

Babanazarov, B., (2006) Distribution of New temperature Extremes. Texas Technical University of Health Sciences. Unpublished M.Sc. Thesis.

Basu, R., and Ostro, B. D., (2008). A Multi-Country Analysis Identifying the Populations Vulnerable to Mortality Associated with High Ambient Temperature in California. *American Journal of Epidemiology*, **168**, 632-637.

Basu, R., and Samet, J. M., (2002). Relation between Elevated Ambient Temperature and Mortality: a Review of the Epidemiologic Evidence. *Epidemiologic Reviews***,** **24**, 190-202.

Beirlant, J., Goegbeur,Y., Segers, J., and Teugels, J., (2004). Statistics of Extremes. Theory and Applications. John Wiley and Sons.

Blench, R., (2006). Background Conditions in the Upper East Region of Northern Ghana, 2005. Working paper. Rome: IFAD. Electronics ms.

Boko, M., Besancenot, J. P., and Oke, P. C., (1997). Weather Conditions and Cerebrospinal Meningitis in Benin (Gulf of Guinea, West Africa). European *Journal of Epidemiology*, **13**(**7**), 807-815.

Bommier, E., (2014). Peaks-Over-Threshold Modelling of Environmental data. University of Uppsala, Sweden. An Unpublished M.Sc. Thesis.

Bonsal, B. R., Zhang, X., Vincent, L. A., and Hogg, W. D. (2001). Characteristics of Daily and Extreme Temperatures Over Canada. *Journal of Climate*, **14**, 1959-1976.

Box, G.E.P., and Jenkins, G.M., (1976). Time Series Analysis, Forecasting and control. Oakland, CA.: Holden Day.

Brooks, C., (2012). Introductory econometrics for finance. United Kingdom, Cambridge University Press.

Brown, S., Caesar., J and Ferro, C., (2008). Global Changes in Daily Temperature since 1950, *Journal of Geophysical Research: Atmosphere* (1984-2012), 13.

Castillo, E., (1988). Extreme Value theory in Engineering: Academic press.

Claudette, A. M., (1992). Regional Frequency Analysis of Seasonal Rainfall and Snowfall, for the Southern interior of British Columbia. Unpublished M. Sc. Thesis.McGill University.

Coles, S., (2001). An Introduction to Statistical Modelling of Extreme Values.Springer-Verlag.

Coles, S.G., and Tawn, T.A., (1996b). Modelling Extremes of the Areal Process. Journal of the Royal Statistical Society. B58, 329-347.

Colombo, A. F., Etkim, D., and Karney, B. W., (1999). Climate Variability and the Frequency of Extreme Temperature Events for Nine Sites Across Canada: Implications for Power Usage, *J. Clim*., **12**, 2490-2502.

David, F., Valeries, L., Giorgio, T., and Sandro, V., (2011). Numerical Convergence of theBlock Maxima Approach to the Generalized Extreme Value Distribution, *Journal State Phys*., **145**, 1156-1180, Springer, 2011.

Davison,A.C. and Smith, R.L.(1990). Models or Exceedances over High Thresholds. *Journal of the Royal Society*, **B 52**, 393-442

De Chabalier, F., Djingarey, M. H., HASSNE, A., and Chippanx, J. P. C. (2000). Meningitis Seasonal Pattern in Africa and Detection of Epidemics: A Retrospective Study in Niger; 1990-1998. *Trans. Roy. Soc. Trop. Med. Hyg*. 2000, **94**, 664-668.

Diebold, F. X., (1998). Elements of Forecasting. Cincinnati: Southwestern College Publishing

Dominici, F., McDemolt, A., Daniels, M., Zeger, S. L., and Samet, J. M., (2005). Revised Analysis of the National Morbidity, Mortality and Air pollution Study: Mortality among residents of 90 cities. *Journal of Toxicology and Environmental Health*.; part A, **68** (13/14), 1071-1092.

Dukic, V., Hayden, M., Forgor,A.A., and Akweongo, P., (2012). The Role of Weather in Meningitis Outbreaks in Navrongo, Ghana:A Generalized Additive Modelling Approach.*Journal of Agricultural, Biological and Environmental Statistics*, vol.17, **3**, 442-460.

Embrechts, P., Klupelburg, C., and Mikosch, T., (1998).Modelling Extremal Events for Insurance and Finance. Springer Verlag, New York.

Embrechts, P., Resnick, S. I., and Samarodnitsky, G., (1999). Extreme Value Theory as a Risk Management tool.*North American Actuarial Journal***3**(2):30-41. Environmental Protection Agency (EPA) (2000).

Issahaku, A. B., Campion, B., and Edzieyie (2016). Rainfall and temperature changeds and variability in the Upper East Region of Ghana, *Earth and space science,3. Doi.10.1002/2016EA00161*

Felix, A. A., and Amuakwa-Mensah, F., (2014). Climate Change and Variability in Ghana: Stock taking.

Fischer, E. M., Seneviratne, S. L., Luthi, D., and Schar, C., (2007). Contribution of Land Atmosphere Coupling to Recent European Heatwaves.Geophys. Res. Lett., 34, L06, 707.

Fisher, R., and Tippet, 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.*

Frei, C., and Schar, C., (2001). Detection Probability of Trends in Rare Events. Theory and Applications to Heavy Precipitation in the Alpine Region. *Journal Clim.*,**14**, 1568-1584.

Gnedenko, B., (1948). Sur la distributions limite du terme maximum d'une serie. *The annals of Mathematics, 44(3):423-453.*

Galambos, J., Liegh, S., and Simiu, E.,(1994). Extreme value theory and applications. Kluwer, Amsterdam.

Garcia-Herrera, R., Diaz, J., Trigo, R., Lutherbacher, J., and Fisher, E., (2010). A Review of European Summer Heat Waves of 2003, *Crit. Rev. Environ. Sci. Technol*., **40**, 267-306

Ghana statistical service, (GSS2010) Population and Housing Census Summary Report of Final Results; Sakoa Press: Accra, Ghana.

Gilleland, E. and Katz, R.W.(2011). New Software to Analyze How Extremes Change over Time.EOS, Transactions American Geophysical Union, 92, 13-14.

Gillian, L. H., and Tata, S. R., (2005). Statistical Analysis and Time Series Models for Minimum and Maximum Temperatures in the Antartic Peninsula.

Greene, W. H., (1997). Econometric analysis 3rd edition. Englewood Cliffs, NJ: Prentice Hall.

Greenwood, B. M., Bradley, A.K., Wall, R.A., Meningococcal Disease and Season in Sub-S aharan Africa. *Lancet* 1985, **2** 829-830.

Gumbel, E. J.,(1941). The Return Period of Flood Flows. The Annals of Mathematical Statistics **12**(**2**): 163-190.

Hajat, S., Kovats, R. S., and Lachonycz, K. (2007). Heat-Related and Cold-Related Deaths in England and Wales: Who is at Risk? *Occupational and Environmental Medicine,* 64 (2), 93- 100.

Halvorsen, R., and Palmquist, R., (1980). The Interpretation of Dummy Variables in Semi-Logaritmic Equations.

Hartfield, J. L., Boote, K. J., Kimball, B., Ziska, L., Izaurelde, R. C., Ort, D., Thomson, A. M., and Wolfe, D., (2011). Climate Impacts on Agriculture: Implications for Crop Production, Agron. J., **103**, 351-370.

Hasan, H., Salam, N., and Adam, B.M., (2013). Modelling Annual Extreme Temperature Using Generalized Extreme Value Distribution: A Case Study in Malaysia. *International journal Mathematical, Computational, Physical, Electrical and Computer Engineering* vol:7,No:6, 2013.

Hasan, H., Yeong, W.C., (2012). Extreme Value Modelling and Prediction of Extreme Rainfall: A case study of Penang, AIP conf. Proc, V1309, 372-393.

Inderjeet, K., and Sabita, M. S., (2008). Seasonal ARIMA Model for Forecasting of Monthly Rainfall and Temperature.*Journal of Environmental Research and Development, vol*. 3 No 2.

IPCC, (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A special report of working groups I and II of the International Panel on Climate Change.Cambridge University Press, Cambridge, UK.

Jarvis , C. S., Mendehall, C. S and Harold, L. I., (1936). Floods in the United States, magnitude and frequency.*U.S.A Geol*. *Survey water-supply paper* 771-497.

Jeffrey, A. W and Samuel, M., (2014) National Health Statistics Reports, U.S.A.

Jenkinson, J.A.,(1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Events. *Quarterly Journal of the Royal Meteorological Society* ***81***, *158-172.*

Jesper, R., (2005). Extreme Value Modelling: A Preliminary Analysis of Monthly Precipitation at Havana. *School of Technology and Society, Dept. of Applied Mathematics, Malmo Univ*.

Kankam-Yeboah, K., Amisigo, B., and Obuobi, E., (2010). Climate Change Impacts on water Resources in Ghana (CSIR/ WRI).

Koppe, C., Kovats, R., Jendritzky, G., Menne, B.,(2004). Heat-waves: Impacts and Responses. *Techn. Ber.,World Health Organization Regional Office for Europe, Copenhagen, Sweden: Health and Global Environmental Changes Series 2.*

Kothawale, D. R., and Rupa, K. K., (2000). Tropospheric Temperature Variation over India and Links with the Indian Summer Monsoon: 1971-2000, Mansam, 53.

Kotz, S., and Nadarajah, S.(2000). Extreme Value Distributions:Theory and Applications Imperial College Press, London.

Kovats, R.S., and Hajat, S., (2008). Heat Stress and Public Health: *A critical Review. Annual Review of Public Health, 29,* ***41****-55.*

Kwabena, A. K., (2013). Assessing the Impacts of Climate Change on the Availability of Stored Water for Irrigation Purposes in the Semi-arid Areas of Ghana, a Case Study of Tono and Vea Irrigation Projects.

Kysely, J., (2012). Probability Estimates of Extreme Temperature Events: Stochastic Modelling Approach Versus Extreme Value Distributions. *Institute of Atmospheric Physics, Acad.Sci. Czech Republic.*

Kysely, J., Picek, J., and Beranova (2010). Estimating Extremes in Climate Change Simulations

Using Peaks-Over-Threshold Method with a Non-Stationary Threshold.Institute of Atmospheric Physics.*Journal of Global and Planetary Change*, 72 (2010) 55-68.

Leadbetter, M. R., Lindgren, C., and Rootzen (1980). Extremes and Related Properties of Random Sequences and Process, Springer-Verlag, New York.

McMichael, A. J., Wilkinson, P., Kovats, R. S., Pattenden, S., and Armstrong, B., (2008). International Study of Temperature, Heat and Urban Mortality: the ISOTHURM Project. *International Journal of Epidemiology*, **37** (5), 1121-1131.

Michelle, J. F.B., Manabe, S., Tokioka, T., and Meleshko, V., (2010). High Temperatures and Health Outcomes: A Review of the literature. Climate change, The IPCC Scientific Assessment, Cambridge University Press, UK, 131-172.

Muhammed, A., Muhammed, J., and Muhammed, S., (2013). Relationship Between Extreme Temperature and Electricity Demand in Pakistan.*International Journal of Energy and Environmental Engineering*. 4:36

Nejc, B., Brilly, M., and Sraj M., (2014) Comparison between the Peaks over Threshold Method and the Annual Maximum Method for Flood Frequency Analysis, *Hydrological Sciences Journal*, 59:5, 959-977.

Nicholls, N., and Alexander, B., (2007). Detecting, Understanding and Attributing Climate

Change.*Australian Greenhouse Office, Dept. of Environmental and water resources*.

Ntiamoah, A., Afrane, G., (2008). Environmental Impacts of Cocoa Production and Processing in Ghana: Life Cycle Assessment Approach. *Journal. Clean. Prod*. 2008, **16**, 1735-1740.

Nury, A. H., Koch, M., and Alam, M. J.B., (2012). Time Series Analysis and Forecasting of Temperatures in the Sylnet Division of Bangladesh.

Palutikof, J. P., Davies, T. D., and Kelly, P. M., (1998). A Data Bank of Wind Speed Records for the British Isles and Offshore Waters. Cambridge University Press, Cambridge.

Politano, L., (2008). Extreme Temperature Events in the Mediterranean. Institute of Geography

Porcernich, J.M.(2002). Application of Extreme value theory and threshold models to Hydrological events. MSc. Thesis. Univ. of Colorado, Denver.

And Oeschger Centre for Climate Change Research. Univ. of Bern.

Prasad, P., Boote, K. J., Allen Jr., Sheely, J., and Thomas, J., (2006). Species, ecotype and Cultivar Differences in Spikelet Fertility and Harvest Index of Rice in Response to High Temperature stress, Field Crops Res., **95**, 398- 411.

Rasul, G., Dahe, Q., and Chaudhry, Q. Z., (2008). Global Warming and Melting of Glaciers along slopes of HKH ranges, Pakistan. Journal of Meteorology, vol.5, no. 9, pp.63-76.

Robinson, M. E., and Tawn, J. A. (1995). Statistics for Exceptional Athletics Records Applied

Statistics 44, 499-511.

Rodriguez-Puebla, C.S., Ayuso, S., Frias, M., and Garcia-Casado (2007). Effects of Climate Variation on Winter Cereal Production in Spain, Climate Research**, 34**, 223.

Rodriguez-Puebla, C., Encinas, A. H., Gracia- Casado, L. A., and Nieto, S., (2010). Trends in Warm Days and Cold Nights over the Iberian Peninsula: Relationships to large-scale variables, *Clim. Change*, **100**, 667-684.

Rosenzweig, C., Tubiello, F. N. (2007). Adaptation and Mitigation Strategies in Agriculture: an Analysis of Potential Synergies.Mitig.Adapt.Strat.Glob.Chang. 2007, doi: 10. 1007/s 11027-007-9103-8.

Sachs, J. D. (2008). Commonwealth: Economics of a Crowded Planet; The Penguin Press: New York, NY, USA 2008.

Salih, M. A., Danielsson, D., Backman, A., Caugant, D. A., Achtman, M., and Olcen, P., (2000). Characterization of Epidemic and Non-epidemic Neissera Meningitis Serogroup A *Clin.*Strains from Sudan and Sweden.*Journal Microbiol.* 1990, **28**, 1711-1719.

Samuel, N. A. C., Vivian, A, N., (2014). Climate Change and Cerebrospinal Meningitis in the Ghana Meningitis Belt.

Sanchez, M. C., (2014). Statistical Modelling and Analysis of Summer Very Hot Events in

Mainland Spain. An unpublished PhD. Thesis. Dept. of Geog., Centre for Climate Change,

Virgili University.

Scarf, P. A., and Laycock, P. J., (1996).Estimation of extreme in corrosion engineering.*Journal of Applied Statistics***23**, 621-643.

Scarf, P.A., and Tawn, J.A., (1996). An extreme value theory for dependent observations. *Journal of Hydrology*. **101,** 227-250.

Scarrot, C., and McDonald, A., (2012). A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. REVSTAT-Statistical Journal, **10**(**1**) : 33-60, 2012.

Smith, B. (2001). Smith, B., Climate Change Adaptation in Light of Sustainable Agriculture. *Journal of Sustainable Agriculture. Vol.* ***27****(1) 113-123.*

Smoyer-Tomic, K., Kuhn, R. and Hudson, A., (2003). Heat Waves Hazards: An Overview of Heat Wave Impacts in Canada.*Natural hazards*.**28**, 463-485.

Stern, N., (2007). Review on the Economics of Climate Change, Cambridge University Press.

Staffogia, M., Forastiere, F., Agostini, D., Biggeri, A., Bisanti, L., Cadum, E., *et al*., (2006). Vulnerability to Heat-related Mortality: a Multicity Population-based, Case-Crossover analysis. *Epidemiology*, **17**(**3**), 315-323.

Tawn, J. A. (1995). Statistics for Exceptional Athletics Records Applied Statistics **44**, 499-511.

Tobias, A., De Olla, P., Linares, C., Bleda, M., Cayla, J., and Diaz, J., (2010). Short-term Effects of Extreme Hot Summer Temperature on Total Daily Mortality in Barcelona, Spain, *International Journal of Biometry*, **54**, 115-117.

Tresson, J, Crepet, A., Bertail, P., Feinberg, M. H., and Leblanc, J. C., (2004). Probabilistic Exposer Assessment to Food Chemicals Based on Extreme Value Theory. Application to Heavy Metals from Fish and Sea Products. *Food and Chemical Toxicology* **42**(**8**) 1349-1358.

World Health Organization (2003). The Health Impacts of 2003 Summer Heat-waves: Briefing note for the Delegations of the Fifty-third Sessions of WHO Regional Committee for Europe.

World Health Organization, (2004). Action Plan for Malaria Control: Preventive Action and Intensification of the Struggle against Malaria Control, Mimeograph. WHO, Geneva.

Xier Li (2009). Applying GLM Model and ARIMA Model to the Analysis of Monthly Temperature of Stockholm. D-Level Essays in Statistics; Dalarna University, Dept of Econs.

Zelenhasic, E., (1970). Theoretical Probability Distributions for Flood Peaks, *Hydrol. Paper N0. 42. Fort Collins: Colorado State University*.

Zografia, A., and Brenda, L. C., (2012).Statistical Modelling of Temperature Risk. University of Berlin. *SFB 649 Discussion Paper* 2012-029.