MODELLING TEMPERATURE EXTREMES IN THE LIMPOPO PROVINCE OF SOUTH AFRICA USING EXTREME VALUE THEORY

by

ANNA MAMODUPI SEIMELA

DISSERTATION

Submitted in fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

in the

FACULTY OF SCIENCE AND AGRICULTURE
(School of Mathematical and Computer Sciences)

at the

UNIVERSITY OF LIMPOPO

SUPERVISOR: Dr. D Maposa

CO-SUPERVISOR: Prof. M Lesaoana

2021

Declaration

I, **Anna Mamodupi Seimela**, hereby declare that the dissertation submitted to the University of Limpopo, for the Master of Science degree in Statistics has not been submitted by me for a degree at this or any university; that it is my work in design and execution and that all material contained herein has been duly acknowledged.

Signature:.....*AMS*.....Date:...09 September 2020.....

Seimela, A M.

Abstract

Temperature extremes have a crucial impact on agricultural, economic, health and energy sectors due to the occurrence of climate extreme events such as heat waves and cold waves. Limpopo province is among the hottest provinces of South Africa and experiences little rainfall which affect the water availability, food production and biodiversity. In the Limpopo province, temperature extremes are expected to become more frequent as a result of climate change. The aim of this study was to model temperature extremes in the Limpopo province of South Africa using extreme value theory (EVT). The stationarity of the data was tested using augmented Dickey-Fuller (ADF), Phillips-Peron (PP) and Kwiatkowski-Phillips-Schmit-Shin (KPSS). Four candidate parent distributions: normal, log-normal, gamma and Weibull distributions, were fitted to the average monthly maximum and minimum daily temperatures. Prior to the selection of the parent distributions, the data set at each station was subjected to normality test using the Shapiro-Wilk (SW) and Jarque-Bera (JB) tests. The stationarity and normality tests revealed that the maximum and minimum temperature data series at all the stations are neither stationary nor normally distributed. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to select the best fitting distribution at a particular site. The findings revealed that both maximum and minimum temperatures series at all the stations belong to the Weibull domain of attraction. The findings from the Mann-Kendall (M-K) test and time series plots trend analyses showed that there is a monotonic downward and upward long-term trend in minimum and maximum temperature data, respectively.

Two fundamental approaches of EVT, block maxima and peaks-over-threshold (POT), were used in this dissertation. The generalised extreme value (GEV), generalised Pareto (GP) and Poisson point process distributions were fitted to the data set for each station. In order to account for climate change impact, non-stationary models were considered with Seasonal Oscillation Index (SOI) as covariates of the parameters of the GEV distribution. The findings revealed that both the maximum and minimum temperature data can be modelled by the Weibull family of distribution. The EVT return level analysis findings of above 40^0C for maximum temperature suggests impending heat waves and droughts in the Limpopo province. The bivariate conditional extremes approach with a time-varying threshold was used. The findings revealed both significant positive and negative extremal dependence in some pairs of meteorological stations. The findings of this study play an important role in revealing information useful to meteorologists, climatologists, agriculturalists and planners in the energy sector where temperature extremes play an important role.

The scientific contribution of this study was to reduce the risk and impact of temperature extremes on agricultural, energy and health sectors in the Limpopo province. An understanding of temperature extremes will help government and other stakeholders to formulate mitigation strategies that will minimise the negative impact resulting from temperature extremes in the Limpopo province. Among the major contributions of the study was the use of a penalised cubic smoothing spline to perform a nonlinear detrending of the temperature data, before fitting bivariate time-varying threshold excess models based on Laplace margins, to capture the climate change effects in the data. Future studies may consider exploring the use of extreme value copulas, as well as spatio-temporal dependence between temperature extremes using the conditional extremes model of Heffernan and Tawn (2004).

Dedication

This work is wholeheartedly dedicated to my mother, Mantoa and my late father Pegedi. My mother, a strong and gentle soul, who taught me to trust in God and believe in hard work, gave me the inspiration and strength when I thought of giving up. To my son Kgethego and siblings, for tirelessly supporting and encouraging me during my study period.

Acknowledgments

I thank the Almighty God for his guidance throughout my study period. Special thanks go to my supervisors, Dr. Daniel Maposa and Prof. 'Maseka Lesaoana, for their encouragement, tireless, unreserved support and guidance throughout the research process. I also sincerely thank my family and friends for supporting, advising and encouraging me throughout the duration of my master of science degree.

I am very grateful to the South African Weather Service (SAWS) and the National Ocean and Atmospheric Administration (NOAA), for providing data used for this dissertation. I am also greatly indebted to the DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS) of South Africa for providing funds for this study.

Contents

List of Figures

List of Tables

Research Outputs

The following sections give a list of research outputs from this dissertation.

JOURNAL ARTICLES PUBLISHED

1. Maposa D., Seimela A.M., Sigauke C. and J.J. Cochran (2020). Modelling temperature extremes in the Limpopo province: Bivariate time-varying threshold excess approach. *Natural Hazards*. Available at: `https://doi.org/10.1007/s11069-021-04608-w`.

2. Seimela A.M. and Maposa D. (2020). An investigation of parent distributions and long-term trends of average maximum and minimum temperature in the Limpopo province of South Africa. *Journal of Statistics Applications & Probability*, **10**(1), 17-33. Available at: `http://dx.doi.org/10.18576/jsap/100103`.

SCIENTIFIC PAPERS PRESENTED AT CONFERENCES

1. Seimela A.M., Maposa D. and Lesaoana M. (2019). An investigation of parent distributions and long-term trends of average maximum and minimum temperature in the Limpopo province of South Africa. 27-30 November 2019, **Nelson Mandela University, Port Elizabeth, South Africa**.

2. Seimela A.M., Maposa D. and Lesaoana M. (2019). 10^{th} Faculty of Science and Agriculture Postgraduate Research Day Seminar [FSA-PGRD], (2019). Fitting parent distributions to average maximum and minimum temperatures in the Limpopo province. 19-20 September 2019, **The Ranch Hotel, Polokwane, South Africa**.

Chapter 1

Introduction and background

1.1 Introduction

Over the last five decades, South Africa has experienced a considerable increase in mean annual temperatures with hot and cold extremes increasing and decreasing in frequency across the country respectively (DEA, 2013). Temperature is one of the main climatic elements that can indicate climate change as climate change seems to be one of the most important issues in the recent two decades (Worku et al., 2019; Roy and Das, 2013, 2012). Climate change is a measurable reality posing noteworthy social, economic and environmental risks and challenges globally (DEA, 2017). Global warming and its associated increase in temperature extremes pose a substantial challenge on natural systems.

Limpopo province is one of the nine provinces of South Africa. The province is among the lowest ranked in terms of regional gross domestic product (GPD) per capita and it is the most vulnerable province to climate change impacts. How-

ever, the province is one of South Africa's richest agricultural regions (Maponya and Mpandeli, 2012b). Drought is the main problem in the province affecting agricultural sectors due to high temperatures and unreliable rainfall in the province (Mpandeli and Maponya, 2013; Tshiala et al., 2011). The recent high temperatures in the province was experienced in the western bushveld and the lowveld in October 2019 (Phophi et al., 2020).

According to Tarmizi et al. (2019), human activities are the major causes of climate change because of the burning of fossil fuel that produces gases like carbon dioxide, methane and nitrous oxides which lead to global warming. These gases are then released into the atmosphere. The atmosphere gets polluted with the carbon dioxide which covers the earth and absorbs heat, then the absorbed heat causes the earth to warm, leading to extreme temperatures (Mori, 2016). According to Ochanda (2016), climate change can influence nature and threaten humans in different aspects of life, both economically and socially. It is widely believed that the changing temperature due to global warming is permanently changing the earth's climate. That is, increase in temperature is likely to lead to a global increase in drought condition, decrease in water supplies due to evapotranspiration and increase in agricultural demand (Nhamo et al., 2019; Ochanda, 2016).

The temperature extremes and frequent flooding affect agricultural production in the province, leading to scarce food and water resources, which is a big threat to a country like South Africa, where the population is rapidly growing (Andersson and Nilsson, 2018). South Africa is also concerned about public health around extreme hot events rather than extreme cold events and how the impact of these events may change in the future (Wright et al., 2014). In the past four decades (1980-2015), Southern Africa experienced 491 climate disasters (meteorological, hydrological, and climatological) that resulted in 110 978

deaths, left 2.49 million people homeless and affected an estimated 140 million people (Reddy and Vincent, 2017).

1.2 Background

South Africa is one of the most vulnerable countries to climate variability and events due to its exposure to extreme weather events and sensitive economy (Jimmy et al., 2019). Extreme climate and weather events such as flooding, drought, heat waves and cold waves have negative impacts on the society, environment and resource management, particularly in developing countries like South Africa (Gebrechorkos et al., 2019; Wolf et al., 2010). Extreme climate events are expected to become more frequent as a result of climate change (Hales et al., 2003). Climate change can be due to natural and anthropogenic effects (John et al., 2014; Nasri and Modarres, 2009). Climate change has resulted in rising temperature trends with associated changes in temperature extremes across the globe, which has the potential to impact on human health. It is generally anticipated that as the planet heats up, climate variability will increase (Jimmy et al., 2019; Mothupi et al., 2016; Kruger and Sekele, 2013).

Temperature extremes, which are attributed to an increasing concentration of greenhouse gases, are natural phenomena that affect our socio-economic activities (Senyolo et al., 2018). For instance, extremely high temperatures and prolonged heat waves can damage agricultural production, increase energy and water consumption and also badly affect human well-being, human health and even cause loss of human lives (Hasan et al., 2012), while extreme cold temperatures can also lead to high energy consumption particularly during the winter period in South Africa and other parts of Southern Africa.

During the 21^{st} century, the global surface temperature has increased by about 0.85^0C and many areas have experienced significant warming (Toros et al., 2019). Kruger and Shongwe (2004) found considerable increase in temperature between 1960 and 2003 for the three stations Bela Bela, Polokwane and Musina situated in the Limpopo province in north-eastern South Africa.

1.3 Problem statement

Many areas of society throughout the world are susceptible to the effects of temperature extremes (Raghavendra et al., 2019; Nemukula et al., 2018; Keellings and Waylen, 2015). Temperature extremes such as heat waves and cold waves are deadly natural hazards, although they occur less frequently and are more difficult to detect than a hurricane or a cyclone (Chan and Nadarajah, 2015; Henderson and Muller, 1997). Heat waves are reportedly occurring more frequently across much of the globe including South Africa, and under a global warming climate they are expected to increase in frequency, intensity, and duration (Sigauke and Nemukula, 2018; Keellings and Waylen, 2015; Coumou et al., 2013; Grumm, 2011). Climate change is regarded as the most contributing factor to recent increases in global temperatures (Winter, 2016; Beniston et al., 2007).

Temperature extremes have a major impact on agricultural, economic, health and energy sectors (Raggad, 2018; John et al., 2014; Reich et al., 2014). For instance, extreme high temperatures such as heat waves may result in loss of plant and animal species, losses in economic goods, high energy demand for air conditioning, death resulting from heart attacks, heat cramps, fainting, heat strokes and heat exhaustion (Bowling et al., 2018; Sigauke and Nemukula, 2018; Chan and Nadarajah, 2015). Extreme low temperatures such as cold waves may result in freezing and bursting of water pipelines, a rise in the

demand for fuels and electricity, animals not able to graze and dying of starvation, frostbites in humans and animals, and other serious medical ailments (Reich et al., 2014; Chikobvu and Sigauke, 2013; Nury et al., 2013).

In Africa, the impact of a changing climate varies by region (Sigauke and Nemukula, 2018; Wright et al., 2014; Yamba et al., 2011). By the end of the century, Southern Africa is expected to experience an average temperature increase of about two degrees Celsius higher than the predicted average global increase (Wright et al., 2014). Changing weather conditions increase electricity demand due to the fact that in winter heating systems are used, while in summer air conditioning appliances are used (Sigauke and Nemukula, 2018; Chan and Nadarajah, 2015). This creates a big problem, particularly in South Africa where the national electricity supplier, ESKOM, is already battling with meeting the demands of the nation in energy supply (Sigauke and Nemukula, 2018; Chikobvu and Sigauke, 2013). Over recent decades, ESKOM has experienced an increased demand for electricity, placing pressure on its supply side, consequently leading to rolling blackouts (Hohne et al., 2019). According to Yamba et al. (2011), energy demand is expected to change drastically in South Africa as a result of increasing temperatures and changing weather patterns, subsequently affecting heating and cooling demands.

In Limpopo province, temperature extremes have effects on human health and agricultural production. These temperature extremes result in the transmission and spread of respiratory, diarrhoea and malaria diseases to humans (Thompson et al., 2012). Temperature extremes are also a major threat to food security as they have a strong impact on food production, access and distribution (Maponya and Mpandeli, 2012b). Limpopo is a drought prone province which faces challenges of drought from time to time. During drought, there is an increased risk for wildfires and dust storms. Agricultural sector is one of

the industries that have been hardest hit by the drought. As a result of the severe drought, the province experienced reduced grazing and water for livestock and irrigation which negatively impacted the agricultural sector (Maponya and Mpandeli, 2012a).

1.4 Rationale

Temperature extremes may have significant effects on various sectors of South Africa and its economy. Sectors that are most vulnerable to temperature extremes in South Africa are agriculture, health and energy. There is a critical need to understand the variability and persistence of temperature extremes in order to enhance climate impact studies and to act as a baseline for global climate change (Henderson and Muller, 1997).

Chikobvu and Sigauke (2013) investigated the influence of temperature on average daily electricity demand in South Africa using a piecewise linear regression model and the generalised extreme value (GEV) distribution approach for the period 2000-2010. Their empirical results showed that electricity demand in South Africa is highly sensitive to cold temperatures. The results also revealed the Weibull as the appropriate distribution to model daily temperature extremes. Their study focused on the block maxima approach. This study will use both the block maxima and peaks-over-threshold (POT) approaches, where a comparative analysis of the GEV with a generalised Pareto (GP) distributions, Poisson point process and bivariate realisations will be performed.

Nemukula et al. (2018) used bivariate threshold excess to model extreme high temperature in Limpopo province of South Africa. The aim was to model the extremal dependence of temperature at three meteorological stations in the Limpopo province. A penalised cubic smoothing spline was used for non-linear

detrending of the data. The present study will adopt a similar approach, but with more meteorological stations of large average monthly temperature data sets.

In South Africa, few studies have been conducted using extreme value theory (EVT) to model temperature extremes, and the majority of those that used EVT mainly focused on the GEV and GP distributions. Again, most of the research conducted on temperature extremes used daily data. This study will use average monthly maximum and minimum temperature data to model temperature extremes using Poisson point process and bivariate threshold excess, since literature is scarce in South Africa on studies conducted using these approaches. The GEV and GP distributions will also be explored in the absence and presence of some covariates. The present study will also investigate the domain of attraction of the parent distributions of the maximum and minimum temperatures at the selected sites. The purpose of studying temperature extremes in the Limpopo province is to reduce the associated risk and impact of these temperature extremes on agricultural, health, economic and energy sectors. Hence, modelling temperature extremes using EVT will help decision-makers in the agriculture, economic, health, water and energy sectors in the Limpopo province of South Africa to plan ahead in order to counter climate related disasters.

1.5 Aim and objectives

1.5.1 Aim

The aim of this study is to model the extreme maximum and minimum temperatures in the Limpopo province of South Africa using extreme value theory.

1.5.2 Objectives

The objectives of the study are to:

1. Fit the parent distributions to average maximum and minimum temperatures and select the best fitting distribution(s).

2. Investigate the long-term trends of the maximum and minimum temperatures for four selected weather stations in the province.

3. Model the tail behaviour of the block maxima and minima temperature extremes in the province in the presence and absence of some covariates.

4. Model the extremal behaviour of the peaks-over-threshold temperature extremes in the province using the GP distribution and Poisson point process approaches.

5. Model maximum temperature extremes in the province using the bivariate conditional extremes approach with a time-varying threshold.

6. Perform a comparative analysis of the various modelling approaches used in the study.

1.6 Significance of the study

Temperature extremes are expected to increase in South Africa. The study is intended to enhance awareness and understanding of temperature extremes for the community and decision-makers in the Limpopo province to develop tools that will reduce the negative impact of extreme temperature events in the agricultural, economic, health and energy sectors. The major findings of this study are of paramount importance in the sense that developing statistical models for temperature extremes will help in the mitigation strategies and reduce the deleterious impacts of these temperature extremes on humans and

property. This will also go a long way in reducing the amount of aid money required for post-disaster operations.

1.7 Structure of the dissertation

This dissertation is divided into five chapters. Chapter 1 describes the background, problem statement, rationale, aim and objectives and significance of the study. Chapter 2 explores the previous literature of similar and related studies that have been done. Chapter 3 details the research methods that are used in the data analysis. In Chapter 4 the data is analysed and the results are presented in the form of graphs and tables. Chapter 5 gives the concluding remarks and recommendations of the study.

Chapter 2

Literature review

2.1 Introduction

This chapter outlines the related literature review of several studies worldwide. These include the application of extreme value theory (EVT), and analysis of long-term trends of temperature extremes using Mann-Kendall (M-K) test statistics and other time series techniques. EVT provides a rigorous framework for analysis of climate extremes and their return levels. It is widely applicable in a wide range of disciplines including finance, insurance, engineering, hydrology and climatology (Rasaki et al., 2018; Mothupi et al., 2016). M-K test has been widely used in hydrological studies (Salman et al., 2017). Its advantages are that it is distribution-free, robust against outliers, and has a higher power than many other commonly used tests.

2.2 Climate of Southern Africa and South Africa

The climate across Southern Africa varies from arid conditions in the west to humid subtropical conditions in the north and east, while much of the central part of Southern Africa is classified as semi-arid (Daron, 2014; Hulme et al., 2001). The climate of South Africa is highly variable and unpredictable and the region is prone to extreme weather conditions such as drought, floods and heat waves (Makate et al., 2019; Nhemachena and Hassan, 2007). One of the main components of climate change is temperature, which has increased globally by approximately 0.9 degree Celsius over the past 100 years (Prokosch et al., 2019). Climate change with expected long-term changes in rainfall and extreme temperature patterns are expected to have considerable negative effects on agriculture, food and water security, and economic growth in Southern Africa including South Africa (Nhemachena and Hassan, 2007).

2.3 Climate of Limpopo province

Limpopo province is one of the developing provinces in South Africa in terms of regional GDP per capita and is one of the most vulnerable provinces to climate change impact due to its exposure to extreme weather events and sensitive economies (LEO, 2016; Tshiala and Olwoch, 2010). The province has three distinct climatic regions: the lowveld, middleveld and highveld (Tshiala et al., 2011). Climate change mostly affects the rural poor communities where the economy is fully dependent on rain-fed agriculture (Worku et al., 2019).

Agriculture, mining and tourism are the three pillars of the economy of Limpopo province. However, the agricultural sector has been heavily affected by drought for a very long time, with particular negative impacts on rain-fed agriculture (Phophi et al., 2020; Mpandeli and Maponya, 2013). The extreme droughts that affected many areas in Limpopo province were those in the 1970s, 1980s, 1990s,

2000, 2005, 2012 and 2015 (Vogel and Olivier, 2019; Mpandeli and Maponya, 2013). Even though drought has been a problem in the province due to high temperatures, there are other socio-economic factors affecting and constraining livelihoods in Limpopo, such as HIV/AIDS, poverty and a high unemployment rate (Phophi et al., 2020; CSP, 2015; Mpandeli and Maponya, 2013).

2.4 Heat and cold waves

A heat wave is a prolonged period of excessively hot weather, lasting for several days or weeks (Hasan et al., 2012). Heat waves are expected to become more intense, more frequent and last longer than what has been observed in recent years (Kauwe et al., 2019; Hatfield and Prueger, 2015; Teskey et al., 2015). Temperature over subtropical Southern Africa have risen more than twice the global rate over the last five decades (Reddy and Vincent, 2017). It is, therefore, likely that there will be further heat waves across South Africa and the continent of Africa in general (Russo et al., 2016).

Globally, heat waves cause human mortality more than any other natural hazard like hurricanes, lightning, tornadoes, floods and earthquakes (Luber and McGeehin, 2008). Increasing temperatures will cause a shift in crop growing seasons which affects food security and changes in the distribution of disease vectors, putting more people at risk from diseases such as dengue fever and malaria (Singh et al., 2014). In South Africa heat waves occur mostly in summer, and occasionally in spring and early autumn (Phillips, 2020).

Globally, heat waves claim hundreds and thousands of lives each year (Raghavendra et al., 2019). The recent heat wave events such as the 2019 European heat wave resulted in the deaths of at least 15 people. Five died in France, four in Germany, three in the United Kingdom, two in Spain and one in Italy (Irfan,

2019). In South Africa, the recent heat waves that occurred in October 2019 in the Limpopo province were extremely severe with maximum temperatures of over 40^0C in the western bushveld and the lowveld.

Contrary to a heat wave, a cold wave is a weather phenomenon that is distinguished by a cooling of the air (Tomczyk et al., 2019). A cold wave can cause death and injury to livestock and wildlife. Exposure to cold weather mandates greater caloric intake for all animals, including humans, and if a cold wave is accompanied by heavy and persistent snow, grazing animals may be unable to reach needed food intake and die of hypothermia or starvation (Barnett et al., 2012).

2.5 Impact of temperature extremes

2.5.1 Agriculture

The agricultural sector in Africa has been impacted by flooding, droughts, soil erosion, land degradation and deforestation, leading to human migration within Africa and to other continents (Gonzalez-Sanchez et al., 2019). Agriculture is the most important sector for the African continent and on average, 70% of the people in the continent are employed in this sector (Pereira, 2017). Limpopo province is extremely vulnerable to climate variability and change as agriculture production depends on climatic conditions and mainly on the quality of the rainy seasons (Musvoto, 2019; Musetha, 2016; Maponya and Mpandeli, 2012a).

Higher temperature increases atmospheric vapour pressure, which in turn increases evapotranspiration leading to more rapid soil drying and increased drought severity (Musvoto, 2019; Teskey et al., 2015; Nelson et al., 2009). Climate extreme events are projected to increase in the warming climate (Kauwe

et al., 2019; Coumou and Robinson, 2013; Miller et al., 2008) and this can minimise crop production since it is known that environmental factors such as temperature and soil moisture are determinant factors of crop yield (Mbokodo, 2017).

2.5.2 Economy

The South African economy is dependent on primary sectors such as agriculture and mining. Climate extremes have substantial control on the day-to-day economic development of the Limpopo province and the whole of South Africa (DEA, 2019). The likely outcomes of the effects of climate extremes are loss of jobs in agricultural value chains, reduced local food security, increasing food prices, the consolidation of production to fewer producers, a decline in export earnings, and a greater dependency on food produce imported from other provinces of South Africa and from other countries (Piennar and Boonzaaier, 2018; CDKN, 2012).

In the Western Cape province of South Africa, losses in agricultural production over the three-year drought period 2015-2017, had a serious economic impact, which included job losses. The number of jobs lost were approximately 35 000 over the three-year period, with a further knock-on effect in the processing and food sectors (Archer et al., 2019; Piennar and Boonzaaier, 2018).

2.5.3 Human health

Temperature extremes are prominent threats to human health. Heat waves and cold waves cause serious effects on human health. High temperature can trigger a multitude of medical conditions including heat stress, heat exhaustion and heat stroke; with the elderly, the young and people with existing medical conditions being the most vulnerable (Henderson and Muller, 1997). The

number of emergency hospital admissions and ambulance call-outs has been observed to increase during heat waves and cold waves (Zuo et al., 2015). In Chicago in the United States of America in 1995, heat waves claimed approximately 739 deaths over a period of five days (Hintz et al., 2018). In Europe 2018, cold waves occurred, which resulted in 137 deaths (Herring et al., 2020).

2.6 Application of EVT on climate extremes

Diriba et al. (2015) used the GP distribution to model extreme daily temperature at Port Elizabeth, in the Eastern Cape province of South Africa for the period 1949-2013. Since extremes in minimum and maximum temperature series do not follow a normal distribution, the non-parametric methods of Kendall's tau test and Sen's slope estimator were used for the trend analysis. The maximum likelihood estimator (MLE) for parameter estimation and confidence intervals on profile likelihood functions were used. Their results showed a significant positive trend for the extreme annual minimum temperature. The return level estimates showed that by the end of the 21^{st} century the extreme summer temperature would have increased by 5^oC, whereas the change in the winter minimum temperature would be less severe with an increase of 2^oC.

Bhagwandin (2017) applied multivariate EVT to climate data in the Western Cape province for the period 1965-2015. The block maxima, threshold excess and point process approaches were used on the weather data, rainfall, wind speed and temperature maxima. Their results showed that models under the block maxima approach do not perform well in modelling the weather variables at the five stations in both the univariate and multivariate cases as many useful observations were discarded. Also, threshold excess and point process approaches performed better in modelling the weather extremes.

Mothupi et al. (2016) estimated extreme quantiles of the maximum surface air temperature for the Sir Seretse Khama International Airport weather station in Gaborone for the period 1985-2015, using the generalised extreme value (GEV) distribution. Their aim was to predict the future behaviour of the quarterly maximum surface air temperature by estimating their high quantiles using the GEV distribution. The Kwiatkowski-Philips-Schmidt-Shin (KPSS) test of stationarity on the series revealed that the maximum temperature was stationary. The seasonal M-K test for trend showed no presence of monotonic trend. The parameters were estimated using the MLE method. The Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) goodness-of-fit tests showed that the distribution gives a reasonable fit to the quarterly maximum surface air temperature.

Sampson and Kwadwo (2019) modelled the extreme maximum temperature in Ghana using EVT for the period 1900-2013. Their aim was to inform decision-makers to help them plan appropriate risk mitigating measures to reduce the damage caused by drought. The block maxima with GEV distribution, and the GP distribution with all excesses and decluster peaks were used on 113 years of monthly temperature data. Two statistical tests for stationarity, namely: augmented Dickey-Fuller (ADF) and M-K tests were performed. In the GEV modelling, the model selection criteria namely: Akaike's information criterion (AIC), likelihood-ratio test and the diagnostic checking indicate that the model with linear trend in location parameter was appropriate. In fitting the GP distribution, the results from the parameter estimation showed that GP with all excesses better fits the data than the GP with decluster peaks. The results were also confirmed by diagnostic tests which also revealed the same conclusion. This suggests that in 20 years to come, maximum temperature in Ghana would exceed $36.3^{o}C$, which indicates a drought period.

Ngailo et al. (2016) modelled extreme maximum rainfall using EVT in Tanzania. The aim of their study was to determine the best fitting distribution to the extreme daily rainfall for 14 stations in Tanzania for the periods 1961-2014 and 1984-2014. They used stationary and non-stationary processes. Their empirical results indicated significant increases in annual maxima over the period 1961-2014 in Dar es Salaam confirming trends over the region, while the trends for other stations were insignificant. The model fit suggested that the Gumbel distribution provides the most appropriate model for the annual maximum of daily rainfall, while the exponential distribution provided a reasonable model for the daily rainfall data over the threshold value of 99% for all the 14 stations considered in the study in Tanzania.

Lazoglou et al. (2019) analysed extreme precipitation and temperature events in the Mediterranean region for the periods 1901-2000, 1916-2015 and 1951-2010 using EVT. They used both the block maxima and peaks-over-threshold (POT) approaches and as a result both the GEV and GP distributions were used to fit the data. Their results were compared in order to select the most appropriate distribution. Their results revealed that the GP distribution is the appropriate approach for both precipitation and temperature extreme events. Also, the GEV distribution with the Bayesian parameter estimation method was found to be appropriate, especially for the greatest values of extremes.

Andersson and Nilsson (2018) used EVT approach to model risk of extreme rainfall in Bangladesh for the period 1980-2016. The aim of their study was to make a univariate extreme value analysis to calculate the extreme rainfall pattern over the next 10, 50 and 100 years. The block maxima and POT approaches were used. GEV and GP distributions were fitted to the daily rainfall data. Their results showed that the return levels for both GEV and GP distributions indicated an increase in rainfall for most of the stations.

Gao and Zheng (2018) applied non-stationary extreme value analysis to model temperature extremes in China for the period 1956-2013. The annual temperature extremes were fitted to GEV distribution with time-varying parameters. The year index and the two climate indices namely; western Pacific substropical high and Arctic oscillation, were chosen as covariates in non-stationary GEV distribution of annual extremes of daily maximum and minimum temperatures. The best fitting GEV model to the 20-year return levels of annual warm and cold extremes in the periods 1961-1980 and 1991-2010 were computed and compared to assess the changes of temperature extremes. Their results revealed that the changes of 20-year return levels of both maximum and minimum temperatures were jointly determined by trend and distributional changes of annual temperature extremes.

Brown et al. (2008) analysed the observed daily temperature anomalies with regard to the normal climate using POT approach for the period 1961-1990. Their results revealed that extreme daily maximum and minimum temperatures have warmed for most regions of Europe since 1950, showing a significant positive trend in extreme temperature anomalies for both upper and lower tails of their distributions. For the non-stationary POT model, the scale and shape parameters were assumed to be time-varying parameters, generating time-varying T-year return levels. In applying non-stationary POT based model, there is an additional important step, threshold selection, which should be carefully considered in non-stationary extreme value modelling of climate extremes.

Osman et al. (2015) applied hybrid POT and GP distribution to model extreme temperature in Ireland under global warming for the period 1961-2000. Their study used a combination of POT, that is, GP distribution in which the parameter of the distribution was allowed to vary with a dominant feature of climate

at the location. Their results indicated that considerable changes in extreme temperature events, which include hotter summers and mild winters, are projected to occur in Ireland over the course of the present century.

Hasan et al. (2012) modelled extreme temperature behaviour in Penang, Malaysia using GEV distribution for the period 2000-2009. They studied extreme maximum temperature 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. The ADF and KPSS stationarity tests both detected no stochastic trend for maximum temperature. However, the M-K test showed that all three selected periods had decreasing trends. The K-S and A-D goodness-of-fit tests showed that all three selected period maximums converge to the GEV distribution.

Hasan et al. (2013) modelled annual extreme temperature in Malaysia using GEV distribution for the period 1981-2011. Extreme temperature of several stations were modelled by fitting the annual maximum to the GEV distribution. The M-K results showed existence of trend for some stations and absence in other stations. The annual maximum temperature was modelled by applying stationary and non-stationary GEV distribution to the different stations. The parameters were estimated using the MLE and likelihood ratio test was used to determine the best-fitting model. The non-stationary models were recommended to describe extreme temperature series in some stations.

Ayuketang and Joseph (2014) modelled extreme temperature in Cameroon using GEV distribution for the period 1993-2012. Maximums of five different selection periods, namely: monthly, bi-monthly, quarterly, half yearly and yearly were fitted to the GEV distribution. The results showed that the monthly, bi-

monthly and quarterly periods were the best selection periods to be fitted to the GEV distribution. It was also found that the three parameter GEV model best fit the data as compared to Gumbel, Frechet and Weibull family of distributions. A normality test conducted using the A-D test rejected the null hypothesis that the distribution of the data was normal in favour of the alternative hypothesis at 5% significance level. The M-K test showed that all selection periods have decreasing trends, while the K-S and A-D goodness-of-fit tests showed that all three selected periods converge to the GEV distribution, with monthly maximum having the best convergence.

Worku et al. (2019) examined the changes in rainfall, minimum and maximum temperature extremes of Jemma Sub-Basin of the Upper Blue Nile Basin for the period 1981-2014. Ten rainfall and 12 temperature indices were used to study changes in rainfall and temperature extremes. The non-parametric M-K, seasonal M-K, and Sen's slope estimator were used to estimate annual trends. The results showed an increasing trend of annual and summer rainfall in more than 78% of the stations and a decreasing trend of spring rainfall in most of the stations. Annual maximum and minimum temperature and extreme temperature indices showed warming trend in the sub-basin.

2.7 Summary of the chapter

In this chapter relevant literature was reviewed. The general background of temperature extremes and application of EVT have been discussed. It was found that there is limited literature on the use of EVT in South Africa.

Chapter 3

Methodology

3.1 Introduction

This chapter mainly presents statistical tests and methods used to analyse the temperature data in Limpopo province.

3.2 Research methodology

3.2.1 Data source and study area

The study uses time series secondary data. The average daily maximum and minimum temperature data measured in degrees Celsius (^0C) were obtained from South Africa Weather Service (SAWS) database. The data are monthly maxima and minima average daily temperatures. The study area covers the following meteorological stations of the Limpopo province in South Africa: Mara (1949-2018), Messina [also locally known as Musina] (1934-2009), Polokwane (1932-2018) and Thabazimbi (1994-2018). Some other meteorological stations

in the Limpopo province such as Giyani, Lephalale, Marken, Mokopane, Thohoyandou and Tzaneen were excluded from this study due to massive missing values. Figure 3.1 shows the four meteorological stations in the Limpopo province where the study is being conducted.

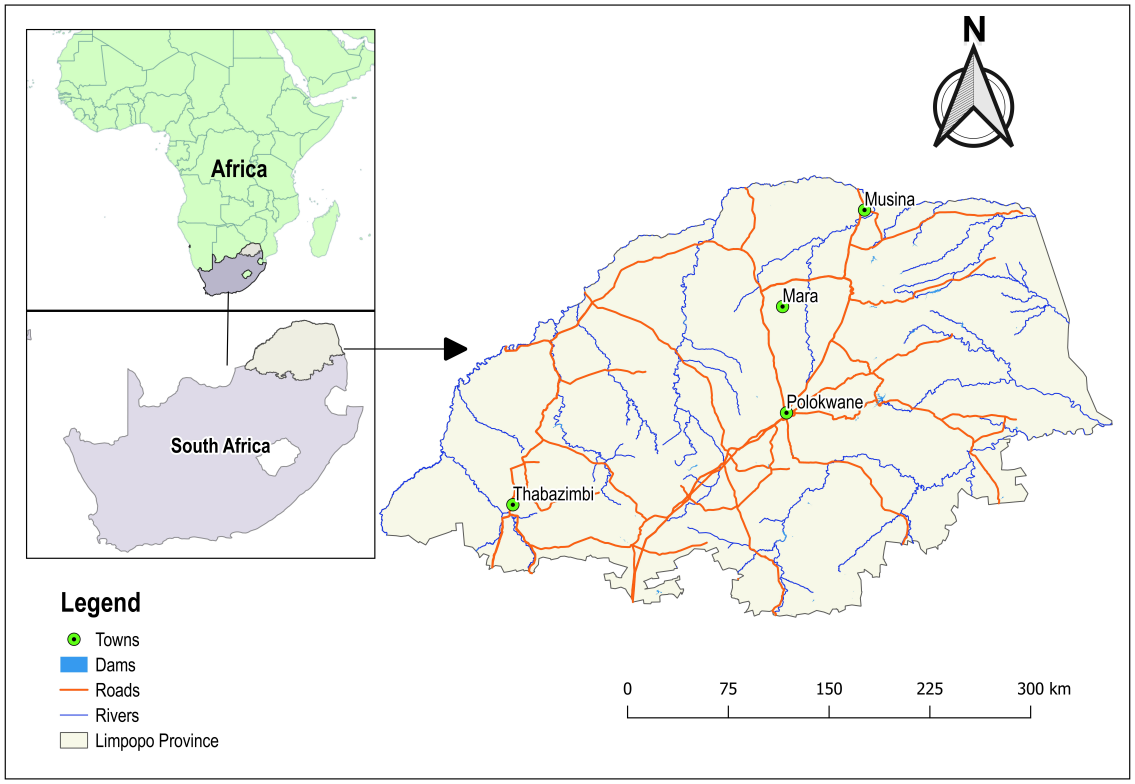

Figure 3.1: Study area map showing the four meteorological stations in the Limpopo province of South Africa. Source: [Author's own contribution]

3.2.2 Tests for stationarity

Augmented Dickey-Fuller test

Augmented Dickey-Fuller (ADF) is a test for a unit root in a time series sample and its fundamental aim is to test the null hypothesis that a unit root is present in a time series sample, that is, $\Phi = 1$, in

$$y_t = \Phi y_{t-1} + v_t, \tag{3.1}$$

where y_t is the maximum and minimum temperature data and v_t is an independently and identically distributed zero-mean error term.
against the one-sided alternative, $\Phi < 1$ (Alhassan and Tela, 2019).

In other words, the hypothesis to be tested is:

H_0: the series has a unit root,

against

H_1: the series is stationary.

If the value of the ADF test statistic is negative, then the null hypothesis is rejected. The more negative the test statistic is, the stronger the rejection of the hypothesis that there is a unit root at some level of significance. Alternatively, if the p-value of the ADF test statistic is small, that is, smaller than the level of significance, then the null hypothesis is rejected (Ochanda, 2016).

Phillips-Peron test

Phillips-Peron (PP) test is a unit root test and it is used to test the null hypothesis that a time series is integral of order 1 (Ochanda, 2016). The null hypothesis is that there is a unit root against the alternative that there is no unit root. If the p-value is greater than the level of significance, then we fail to reject the null hypothesis.

Kwiatkowski-Phillips-Schmid-Shin test

Kwiatkowski-Phillips-Schmit-Shin (KPSS) test is used to test the null hypothesis of trend stationarity against an alternative of a unit root (Ochanda, 2016). If the p-value from the test statistic is greater than the level of significance,

then we fail to reject the null hypothesis and conclude that the data is trend stationary.

3.2.3 Testing for normality

Shapiro-Wilk test

Shapiro-Wilk tests the null hypothesis that a sample $X_1, ..., X_n$ comes from a normally distributed population.

The test statistic is given by

$$W = \frac{(\sum_{i=1}^{n}(a_i)(x_i))^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{3.2}$$

where x_i is the i^{th} order statistics, that is, the i^{th} smallest number in the sample,

$$\bar{x} = \frac{(x_i + ... + x_n)}{n}. \tag{3.3}$$

The constants $a_i s$ in (3.2) are given by

$$(a_1, ..., a_n) = \frac{m^t(V^{-1})}{(m^t(V^{-1}(V^{-1})m)^{1/2}}, \tag{3.4}$$

where

$$m = (m_1, ..., m_n)^T, \tag{3.5}$$

$m_1, ..., m_n$ are the expected values of the order statistics of independent and identically distributed (iid) random variables from the standard normal distribution, and V in (3.4) is the covariance matrix of those order statistics (Chifurira, 2018). Shapiro-Wilk is fairly powerful omnibus test, not good with small samples or discrete data. Good power with symmetrical, short and long tails. Good with asymmetry.

Jarque-Bera test

Jarque-Bera (JB) test is defined as:

$$JB = \frac{n-k+1}{6}(s^2 + \frac{1}{4}(c-3)^2), \tag{3.6}$$

where

n is the number of observations,
s is the sample skewness,
c is the sample kurtosis,
k is the number of regressors (Kim et al., 2020; Gel and Gastwirth, 2008).

Additionally, the sample skewness s and sample kurtosis c are given by

$$s = \frac{\hat{\mu_3}}{\hat{\sigma_3}}, \tag{3.7}$$

$$c = \frac{\hat{\mu_4}}{\hat{\sigma_4}}, \tag{3.8}$$

respectively, where $\hat{\mu_3}$ and $\hat{\mu_4}$ are the estimates of the third and fourth central moments (Kim et al., 2020; Gel and Gastwirth, 2008). JB test is good with symmetric and long-tailed distributions, less powerful with asymmetry, and poor power with bimodal data.

3.2.4 Parent distributions

This subsection presents the theoretical framework of the parent distributions investigated in this study.

Normal distribution

A normal distribution is symmetrical and has a bell-shaped density curve with a single peak (Chikobvu and Chifurira, 2015). The normal density function is given by

$$f(x|\sigma,\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad \text{with } \sigma > 0, \tag{3.9}$$

where μ is the mean, x is the sample mean and σ is the standard deviation, known as the location and scale parameters of the distribution, respectively (Al-Suhili and Khanbilvard, 2014).

Log-normal distribution

The log-normal distribution probability density function (PDF) is given by

$$f(x|\sigma,\mu) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(logx-\mu)^2}{2\sigma^2}\right), \tag{3.10}$$

with μ and σ as mean and standard deviation, known as the location and scale parameters of the distribution, respectively (Al-Suhili and Khanbilvard, 2014).

Weibull distribution

The Weibull distribution PDF is given by

$$f(x|\alpha,\beta) = \left(\frac{\alpha}{\beta^{-\alpha}}\right)\left(x^{\alpha-1}\right) \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right), \tag{3.11}$$

with α and β as shape and scale parameters, respectively.

Gamma distribution

The PDF of the gamma distribution is given by

$$f(x|\alpha,\beta) = \frac{1}{\beta^{\alpha}}\left(\frac{1}{\Gamma(\alpha)}\right)\left(x^{\alpha-1}\right)\exp\left(\frac{x}{\beta}\right), \tag{3.12}$$

with α and β as the shape and scale parameters, respectively.

3.2.5 Maximum likelihood parameter estimation

The maximum likelihood estimator (MLE) is mostly used for fitting probability distributions to data. For an assumed distribution, it produces asymptotically efficient and unbiased estimates (Strupczewski et al., 2001).
Let $X_1, X_2, X_3, ..., X_n$ be a joint density denoted by

$$f_{\theta}(x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n|\theta). \tag{3.13}$$

Given observed values $X_1 = x_1, X_2 = x_2, ..., X_n = x_n$, the likelihood of θ is the function

$$L(\theta|x) = f(x_1, x_2, ..., x_n|\theta), \tag{3.14}$$

considered as a function of θ.

The MLE of θ is that value of θ that maximises $f(\theta)$. If the x_i's are independent and identically distributed (iid), then the likelihood simplifies to

$$likelihood = L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta), \tag{3.15}$$

where $L(\theta|x)$ is the likelihood of the set of parameter θ given the observation X and $\prod_{i=1}^{n} f(x_i|\theta)$ is the probability density function of the probability model (Maposa, 2016).

Maximising the log likelihood, we get

$$\log likelihood = \ln[L(\theta|x)] = \frac{1}{n}\sum_{i=1}^{n}\ln[f(x_i|\theta)]. \tag{3.16}$$

3.2.6 Model selection criterion

The goodness-of-fit of the distributions is assessed using Akaike's information criterion (AIC) and Bayesian information criterion (BIC). AIC and BIC can be used for nested and non-nested models.

Akaike's information criterion

The models are selected according to the values of AIC. The model or distribution with the lowest value of AIC is chosen to be the best (Kim et al., 2020).

The AIC is defined as:

$$AIC = 2K - 2\log(L), \tag{3.17}$$

where
K is the number of parameters in the statistical model,
L is the maximum values of the likelihood function for the estimated models.

Bayesian information criterion

The BIC assesses goodness-of-fit of a distribution or model, but avoids overfitting by penalising additional degree of freedom. The model with the lowest BIC value is chosen as the best (Kim et al., 2020).

The BIC is defined as:

$$BIC = \log(n)k - 2\log(\widehat{L}), \tag{3.18}$$

where

$\widehat{L}$ is the maximised value of likelihood function of the model,

n is the number of data points,

k is the number of free parameters to be estimated.

3.3 Likelihood ratio test for nested models

The likelihood ratio (LR) test compares the log likelihoods of the two models and tests whether this difference is statistically significant. If the difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model. The LR test statistic is given by:

$$Q = 2\log(L_B|L_A) = 2(\log L_B - \log L_A)),$$

which is equal to twice the $\log$ of the ratio of the likelihoods. The value of L_B must be larger than or equal to that of L_A because model A is a special case of model B.

3.3.1 Trend analysis

Mann-Kendall test

The non-parametric Mann-Kendall (M-K) test is commonly used to detect monotonic trends in series of environmental, climate and hydrological data (Ongoma and Chen, 2017; Karmeshu, 2012). The advantage of M-K test is that it is a non-parametric test and does not require the data to be normally distributed. The null hypothesis, H_0, is that there is no trend or serial correlation from a population, while the alternative hypothesis, H_1, states that there is a monotonic increasing or decreasing trend (Ndlovu and Demlie, 2018; Pohlert, 2020).

Let $x_1, x_2, ..., x_n$ be a sequence of random variables, then the M-K statistic can be calculated using

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k), \tag{3.19}$$

where x_j denotes the ordered data values, n is the length of observation and S is the M-K statistic. The sign$(x_j - x_k)$ expression in (3.19) is given by

$$sign(x_j - x_k) \quad = \quad \begin{cases} 1 & \text{if } \ x_j - x_k > 0, \\ 0 & \text{if } \ x_j - x_k = 0, \\ -1 & \text{if } \ x_j - x_k < 0. \end{cases} \tag{3.20}$$

For $n \geq 10$, the statistic S is approximately normally distributed with mean zero $(E(S) = 0)$ and variance as follows:

$$v(s) = \frac{n(n-1)(2n+5) - \sum_{k=1}^{nk} t_k(k)(k-1)(2k+5)}{18}, \tag{3.21}$$

where t_k denotes the number of duplicates to extend k. Equation (3.21) is used in case of tie values in time series where nk is the total number of ties in a dataset. In case of $n \geq 10$, the standardised test statistic for M-K can be written as:

$$Z_s = \tau \quad = \quad \begin{cases} \frac{s-1}{\sqrt{v(s)}}, & \text{if } \ s > 0, \\ 0, & \text{if } \ s = 0, \\ \frac{s+1}{\sqrt{v(s)}}, & \text{if } \ s < 0. \end{cases} \tag{3.22}$$

The test statistic Z_s, also referred to as Kendall's τ, is used to measure the significance of trend. A positive Z_s or τ value indicates an upward trend, whereas a negative value indicates a downward trend (Meng et al., 2018).

3.4 Generalised extreme value distribution model

3.4.1 Block maximum approach

The block maxima (BM) model uses a series of extreme values formed by selecting the highest value in a year or a block of time and then proceeds with fitting a statistical distribution to this extracted series. The BM approach models data by fitting the GEV distribution to a set of block maxima data (Jakata and Chikobvu, 2019).

Let $X_1, X_2, ..., X_n$ be a sequence of independent random variables having a common distribution function F. The model focuses on statistical behaviour of

$$M_n = Max(X_1, ..., X_n), \tag{3.23}$$

where X_i represent values of a process measured on a regular time scale, for example, hourly measurements of sea level or daily mean temperatures so that M_n represents the maximum of the process over n time unit of observation. If n is the number of observations in a year, then M_n corresponds to the annual maximum. Then the distribution of M_n is

$$Pr\{M_n \leq x\} = Pr\{X_1 \leq x, ..., X_n \leq x\} = Pr\{X_1 \leq x\} \times ... \times Pr\{X_n \leq x\} = F(x)^n. \tag{3.24}$$

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x),\ as\ \ n \rightarrow \infty \tag{3.25}$$

where G is a non-degenerate distribution function, then G belongs to one of the following:

$$Type\ I : G(x) = \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\}, \quad -\infty < x < \infty \tag{3.26}$$

$$Type\ II : G(x) \quad = \quad \begin{cases} 0, & x \leq b, \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\}, & x > b; \end{cases} \tag{3.27}$$

$$Type\ III : G(x) \quad = \quad \begin{cases} \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{\alpha}\right]\right\}, & x < b\}, \\ 1, & x \geq b \end{cases} \tag{3.28}$$

for parameters $a > 0$ and b. In the case of families II and III, $\alpha > 0$ (Maposa, 2016).

The GEV distribution consists of parametric distributions, namely: the Gumbel, Fréchet, and Weibull distributions. The GEV cumulative distribution function (CDF) is given by

$$G(\mu, \sigma, \xi; x) \quad = \quad \begin{cases} \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right), & \text{for } 1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0,\ \xi \neq 0, \\ \exp\left(-\exp(-\frac{x-\mu}{\sigma})\right), & x \in \mathbb{R},\ \ \xi = 0, \end{cases} \tag{3.29}$$

where μ, σ and ξ are location, scale and shape parameters, respectively (Dey et al., 2020; Maposa, 2016; Maposa et al., 2014).

The estimates of the extreme quantiles are obtained from the quantile function, X_p, given by

$$X_p = G^{-1}(1-p) \quad = \quad \begin{cases} \mu - \frac{\sigma}{\xi}\left[1-\{-\log\left(1-p\right)\}^{-\xi}\right], & \xi \neq 0, \\ \mu - \sigma\log\left(-\log\ \left(1-p\right)\right), & \xi = 0, \end{cases} \tag{3.30}$$

where p is the exceedance probability. In (3.30) as $p \to 0$ and $\xi < 0$, we get $X_p = \mu - \frac{\sigma}{\xi}$ (Nemukula and Sigauke, 2018; Maposa et al., 2014; Brown et al.,

2008).

3.4.2 Parameter estimation

Maximum likelihood estimation

Under the assumption that $X_1, ..., X_n$ are iid random variables having the GEV distribution, the log-likelihood for the GEV parameters when $\xi \neq 0$ is

$$\begin{aligned} \ell(\mu, \sigma, \xi) &= -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^{m} \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] \\ &\quad - \sum_{i=1}^{m} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)^{-1/\xi}\right], \end{aligned} \tag{3.31}$$

provided that

$$1 + \xi \left(\tfrac{x_i - \mu}{\sigma}\right) > 0, \ for\ i = 1, ..., m.$$

For $\xi = 0$

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^{m} \left(\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^{m} \exp\left\{-\left(\frac{x_i - \mu}{\sigma}\right)\right\}. \tag{3.32}$$

Maximisation of (3.31) and (3.32) with respect to the parameters (μ, σ, ξ) leads to the maximum likelihood estimate with respect to the whole GEV family. Advantage of MLE method over other methods of parameter estimation is its adaptability to changes in model structures. This allows for the incorporation of model parameters when they change as functions of the covariates (Osman et al., 2015; Coles, 2001).

3.4.3 Inference for return level

By substitution of the maximum likelihood estimates of the GEV parameters into (3.30), the maximum likelihood estimate of X_p, for $0 < p < 1$, the $\frac{1}{p}$ return

level, is obtained as (Brown et al., 2008):

$$\hat{X}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \{-log\ (1-p)\}^{-\hat{\xi}}\right], & \xi \neq 0, \\ \hat{\mu} - \hat{\sigma}\ log\ \{-log\ (1-p)\}, & \xi = 0. \end{cases} \tag{3.33}$$

3.5 Model checking

3.5.1 Probability-probability plot

A probability-probability (P-P) plot is a comparison of the empirical and fitted distribution functions (Coles, 2001). With ordered block maximum data $x_{(1)} \leq, ..., \leq x_{(m)}$; the empirical distribution function evaluated at $x_{(i)}$ is given by $\tilde{G}(x_{(i)}) = \frac{i}{m+1}$. When parameter estimates are substituted into (3.29), the corresponding model-based estimates are

$$\hat{G}(x_{(i)}) = \exp\left\{-\left[1 + \hat{\xi}\left(\frac{x_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-\frac{1}{\hat{\xi}}}\right\}. \tag{3.34}$$

The model is working well if $\tilde{G}(x_{(i)}) \approx \hat{G}(x_{(i)})$ for each i, so that the plot points $\left\{\left(\tilde{G}(x_{(i)}), \hat{G}(x_{(i)})\right)\right\}, \quad i = 1, ..., m$, lie close to the unit diagonal. The weakness of the probability plot for extreme value models is that both $\hat{G}(x_{(i)})$ and $\tilde{G}(x_{(i)})$ are bound to approach 1 as $x_{(i)}$ increases, while it is usually the accuracy of the model for large values of x that is of greatest concern (Coles, 2001).

3.5.2 Quantile-quantile plot

The quantile- quantile (Q-Q) plot consists of the points $\left\{\left(\hat{G^{-1}}\left(\frac{i}{m+1}\right), x_{(i)}\right), i = 1, ..., m\right\}$, where

$$\hat{G}^{-1}\left(\frac{i}{m+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \left\{-\log\left(\frac{i}{m+1}\right)\right\}^{-\hat{\xi}}\right]. \tag{3.35}$$

Departures from linearity in the quantile plot indicate failure in the model (Coles, 2001).

3.5.3 Return level plot

The return level plot, comprising a graph of $x_p = \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right]$ against $y_p = -\log(1-p)$ on a logarithmic scale, is particularly convenient for interpreting extreme value models. The tail of the distribution is compressed, so that return level estimates for long return periods are displayed, while the linearity of the plot in the case $\xi = 0$ provides a baseline against which to judge the effect of the estimated shape parameter. A fitted model of the return level plot consists of the locus of points $\{(\log y_p, \hat{x}_p) : 0 < p < 1\}$ (Coles, 2001).

3.5.4 Goodness-of-fit tests

Two goodness-of-fit (GoF) tests, Anderson-Darling (A-D) and Kolmogrov-Smirnov (K-S) were applied at 5% level of significance. The advantage of the A-D test over the K-S test is that it is especially sensitive towards differences at the tail of distribution. There is also evidence that A-D test is better capable of detecting very small differences, even between large sample sizes (Maposa, 2016).

Let $x_1, x_2, ..., x_n$ be a sample of n annual maximum and minimum temperature observed and suppose the CDF of the random variable X is $F(.)$. The A-D and K-S tests are presented in detail in the next two subsections.

Anderson-Darling test

The A-D test is based on a comparison between the fit of an observed CDF to a theoretical (expected) CDF. The test statistic, denoted, A^2, is given by

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln F(x_i) + \ln\left(1 - F_{n-i+1}(x)\right)\right], \tag{3.36}$$

where $F_n(x)$ is the empirical CDF, $F(x)$ is the theoretical CDF, $x_i's$ are the ordered data, n is the sample size (Maposa, 2016).

The hypothesis for GoF under the A-D test procedure is
H_0: The temperature data follow a specified distribution,
H_1: The temperature data do not follow the specified distribution.

The null hypothesis, H_0, for A-D test is rejected at 5% level of significance if the p-value is less than 5%, and conclude that the data does not come from the specified distribution (Maposa, 2016; Sukla et al., 2014).

Kolmogorov-Smirnov test

The K-S test is a non-parametric test applied to test whether the sample under consideration is from a reference probability distribution or to compare whether two samples come from identical distribution (Masereka et al., 2018). The K-S test statistic is calculated from the largest vertical distance, D_{max}, between the empirical (observed) CDF, $F_n(x)$, and the theoretical CDF, $F(x)$. The test statistic of the K-S test is given by

$$D_{max} = \mathbf{Max}_x \left| F^*(x) - F(x) \right|. \tag{3.37}$$

The hypothesis for the GoF under the K-S test procedure is
$H_0 : \ F(x) = F^*(x)$ for all x from $-\infty$ to ∞ (The data follow a specified distribution)
$H_1 : \ F(x) \neq F^*(x)$ for at least one value of x (The data do not follow the specified distribution).

The H_0 is rejected at 5% level of significance if D_{max} calculated is greater than the tabulated value $D_{0.05} = 1.36/\sqrt{n}$ (Maposa, 2016; Sukla et al., 2014).

3.6 The peaks-over-threshold approach

The peaks-over-threshold (POT) technique considers all sample values that exceed a predefined upper threshold u. The POT is based on fitting GP distribution to all values that exceed a certain threshold u (Bommier, 2014; Coles, 2001).

3.6.1 Generalised Pareto distribution

Let $X_1, X_2, ...$ be a sequence of iid random variables, having marginal distribution function F. The conditional probability is given by

$$F_u(y) = Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0, \tag{3.38}$$

where u is the threshold, $y = x - u$ are the excesses. If the parent distribution F was known, the distribution of threshold exceedances in (3.38) would also be known (Diriba and Debusho, 2020; Maposa et al., 2017).

The following theorem gives the connection between GEV and GP distributions.

Theorem

Let $X_1, X_2, ...$ be a sequence of iid random variables with common distribution function F and $M_n = max\{X_1, ..., X_n\}$. Denote an arbitrary term in the X_i sequence by X, and suppose that F satisfies

$$pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \rightarrow G(z) \qquad \text{as} \quad n \rightarrow \infty. \tag{3.39}$$

For large n, $Pr\{M_n \leq z\} \approx G(z)$, where

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{3.40}$$

for some $\mu, \sigma > 0$ and ξ.

Then for large enough u, the distribution function of $(X-u)$, conditional on $X > u$, is approximately,

$$H(y) = 1 - \left(1+\frac{\xi y}{\widetilde{\sigma}}\right)^{-\frac{1}{\xi}}, \tag{3.41}$$

defined on $\{y : y > 0\}$ and $\left(1+\frac{\xi y}{\widetilde{\sigma}}\right) > 0$, where $\widetilde{\sigma} = \sigma + \xi(u+\mu)$. If $\xi < 0$, the distribution of excesses has an upper bound of $u - \frac{\widetilde{\sigma}}{\xi}$; if $\xi > 0$ the distribution has no upper limit. The distribution is also unbounded if $\xi = 0$, which should be interpreted by taking the limit $\xi \to 0$ in (3.41), leading to

$$H(y) = 1 - \exp\left(-\frac{y}{\widetilde{\sigma}}\right), \qquad y > 0, \tag{3.42}$$

which corresponds to an exponential distribution with parameters $\frac{1}{\widetilde{\sigma}}$ (Ragno et al., 2019).

3.7 Parameter estimation

3.7.1 Maximum likelihood estimation

Let $y_1, ..., y_m$ be a random sample of m excesses of a threshold u, the log likelihood function for a GP distribution is given by

$$\ell(\sigma_u, \xi) = -m\log\sigma_u - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m}\log\left(1+\frac{\xi y_i}{\sigma_u}\right), \quad \xi \neq 0, \tag{3.43}$$

provided $\left(1+\frac{\xi y_i}{\sigma_u}\right) > 0$ for $i = 1, ..., m$. Otherwise, $\ell(\sigma, \xi) = -\infty$. The likelihood can also be reduced to

$$\ell(\sigma_u) = -m \log \sigma_u - \frac{1}{\sigma_u} \sum_{i=1}^{m} y_i, \qquad \xi = 0. \tag{3.44}$$

3.7.2 Return levels

Suppose that GP distribution with parameters σ and ξ is a model for exceedances of a threshold u by a variable X. That is, for $x > u$,

$$Pr\{X > x | X > u\} = \left[1 + \xi \left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}. \tag{3.45}$$

Then it follows that

$$Pr\{X > x\} = \zeta_u \left[1 + \xi \left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \tag{3.46}$$

where $\zeta_u = Pr\{X > u\}$. The level x_m that is exceeded, on average, once every m observations can be written as:

$$\zeta_u \left[1 + \xi \left(\frac{x_m - u}{\sigma}\right)\right]^{-\frac{1}{\xi}} = \frac{1}{m}, \tag{3.47}$$

which can be simplified as:

$$x_m \quad = \quad \begin{cases} u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right] & \xi \neq 0, \\ u + \sigma \log(m\zeta_u) & \xi = 0, \end{cases} \tag{3.48}$$

provided m is sufficiently large to ensure that $x_m > u$ (Bommier, 2014; Li et al., 2014). An estimate of ζ_u is also given as:

$$\zeta_u = \frac{k}{n}, \tag{3.49}$$

where k is the number of exceedances and n is the sample size. The number of exceedances of u follows a binomial distribution, $Bin(n, \zeta_u)$ (Coles, 2001).

3.7.3 Threshold selection

Mean excess plot

The mean excess plot is based on the mean of the GP distribution. If Y has a GP distribution with parameters σ and ξ, then

$$E(Y) = \frac{\sigma}{1-\xi}, \tag{3.50}$$

provided $\xi < 1$. Suppose the GP distribution is valid as a model for the excesses of a threshold μ_0 generated by a series $X_1, ..., X_n$, of which an arbitrary term is denoted by X. Using (3.50),

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1-\xi}, \tag{3.51}$$

provided $\xi < 1$ and where σ_{u_0} is the GP distribution scale parameter for exceedances over threshold u_0 (Kajambeu, 2016). If the GP distribution is valid for excesses of the threshold u_0, then it is valid for all thresholds $u > u_0$. Thus, for all $u > u_0$,

$$E(X - u | X > u) = \quad \tfrac{\sigma_u}{1-\xi} = \quad \frac{\sigma_{u_0} + \xi u}{1-\xi}, \tag{3.52}$$

where $\sigma_u = \sigma + \xi(u - \mu)$. Then, 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 the excesses of the threshold u and can be estimated by the sample mean of the threshold excesses (Zhou et al., 2017). This leads to the following procedure, the locus of points

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_i - u) \right) : u < x_{max} \right\}, \tag{3.53}$$

where $x_{(1)}, ..., x_{(n_u)}$ consist of the n_u observations that exceed u, and x_{max} is the largest of the X_i (Andersson and Nilsson, 2018; Coles, 2001).

Parameter stability plot

The concept of the parameter stability plot is that if the exceedances of a high threshold u_0 follow a GP distribution with parameters ξ and σ_{u_0}, then for any threshold u such that $u > u_0$, the exceedances still follow a GP distribution with shape parameter $\xi_u = \xi$ and scale parameter $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$ (Bommier, 2014).
Let

$$\sigma^* = \sigma_u - \xi_u u. \tag{3.54}$$

This new parameterisation in (3.54) does not depend on u any longer, given that u_0 is a reasonably high threshold.

The parameter stability plot is defined by the locus of points

$$\{(u, \sigma^*); u < x_{max}\} \text{ and } \{(u, \xi_u); u < x_{max}\}, \tag{3.55}$$

where x_{max} is the maximum of the observations. Hence, estimates of σ^* and ξ_u are constant for all $u > u_0$ if u_0 is an appropriate threshold for the asymptotic approximation. The threshold should be chosen at the value where the shape and scale parameters remain constant (Bommier, 2014).

3.8 Model checking

3.8.1 Probability-probability plots

Assuming a threshold u, threshold excesses $y_{(1)} \leq ... \leq y_{(k)}$ and an estimated model $\hat{H}$, the P-P plot consists of the pairs

$$\left\{ \left(\frac{1}{m+1} \right), \hat{H}(y_{(i)}); i = 1, ..., k \right\}, \tag{3.56}$$

where $\hat{H}(y) = 1 - \left(1 + \frac{\hat{\xi}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\xi}}}$, provided $\hat{\xi} \neq 0$ (Andersson and Nilsson, 2018).

3.8.2 Quantile-quantile plot

If $\hat{\xi} = 0$, the Q-Q plot is constructed using (3.43) in place (3.42). Assuming $\hat{\xi} \neq 0$, Q-Q plot consists of the pairs

$$\left\{ \left(\hat{H}^{-1} \left(\frac{1}{m+1}, y_{(i)} \right), i = 1, ..., m \right) \right\}, \tag{3.57}$$

where $\hat{H}^{-1}(y) = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[y^{-\hat{\xi}} - 1 \right]$ (Andersson and Nilsson, 2018).

3.8.3 A return level plot

A return level plot consists of the locus points $\{(m, \hat{x}_m)\}$ for large values of m, where $\hat{x}_m$ is the estimated m-observation return level:

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[(m\hat{\zeta_u})^{\hat{\xi}} - 1 \right],$$

provided $\hat{\xi} = 0$ (Maposa, 2016).

3.8.4 Declustering

This technique works by filtering of the dependent observations to obtain a set of threshold excesses that are approximately independent. According to Coles

(2001) declustering works as follows:

- it uses an empirical rule to define cluster of exceedances;
- it identifies the maximum excess within each cluster;
- it assumes cluster maxima to be independent, with condition excess distribution given by the GP distribution;
- it fits the GP distribution to the cluster maxima.

3.8.5 Poisson point process

A Poisson process limit for extremes

The point process is defined on a sequence of point process M_m on $\Re^2$ by

$$M_m = \left\{ \left(\frac{1}{m+1}, \frac{X_i - b_m}{a_m} \right) : i = 1, ..., m \right\}, \tag{3.58}$$

where time axis is through $(0, 1)$; and the second point ensures stability in the occurrence of extremes as $m \to \infty$ such that on $[0, 1] \times [\tau, \infty)$, $T_m \to T$ as $m \to \infty$, where T is heterogeneous Poisson process (Coles, 2001).

Defining a set $A = (t_1, t_2) \times (x, \infty)$ provided $t_1 < t_2$, then when $z \geq u$, the intensity measure is

$$\Lambda(A) = (t_1 - t_2) \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{\frac{1}{\xi}}, \tag{3.59}$$

valid for $1 + \xi(\frac{x-\mu}{\sigma}) > 0$. When u is sufficiently large on $(0, 1) \times [u, \infty]$, M_m in (3.58) is said to follow a Poisson process (Gilleland and Katz, 2016).

3.8.6 Maximum likelihood estimation

The likelihood for the point process takes on the form of the GEV distribution,

$$\Lambda(A) = n_y(t_2 - t_1)\left[1 + \xi(\frac{x-\mu}{\sigma})\right]^{-\frac{1}{\xi}}, \tag{3.60}$$

where n_y is the number of years of observation (Coles, 2001).

The general form of the Poisson process likelihood is given when $[t_1, t_2] = [0, 1]$ is substituted into (3.60),

$$\begin{aligned} L_A(\mu, \sigma, \xi; x_1, ..., x_n) &= \exp\{-\Lambda(A)\} \prod_{i=1}^{N(A)} \lambda(t_i, x_i) \times \\ &\alpha \exp\left\{-n_y\left[1 + \xi(\frac{u-\mu}{\sigma})\right]^{\frac{1}{\xi}}\right\} \prod_{i=1}^{N(A)} \sigma^{-1}\left[1 + \xi(\frac{x_i-\mu}{\sigma})\right]^{-\frac{1}{\xi}-1}. \end{aligned} \tag{3.61}$$

3.8.7 Return levels

The m-observation return level z_m is the level exceeded, on average, once every m years (Coles, 2001). It is given by

$$1 - \tfrac{1}{m} = Pr\{max(X_1, ..., X_n) \leq z_m\} \approx \textstyle\prod_{i=1}^{n} pi,$$

where

$$pi = \begin{cases} 1 - n^{-1}\left[1 + \xi_i(\frac{z_m-\mu_i}{\sigma_i})\right]^{-\frac{1}{\xi}}, & \text{if } [1 + \xi_i(\frac{z_m-\mu_i}{\sigma_i})] > 0, \\ 1, & \text{otherwise}, \end{cases} \tag{3.62}$$

and (μ_i, σ_i, ξ_i) are the parameters of the point process model for observation i. Taking logarithms,

$$\sum_{i=1}^{n} \log pi = \log(1 - \frac{1}{m}), \tag{3.63}$$

which can easily be solved for z_m using standard numerical methods for non-linear equations.

3.9 Extremes of a non-stationary series with time-varying parameters

The time-varying parameters in non-stationary GEV distribution could be modelled as a function of time or other climate indicators (Gao and Zheng, 2018). Suppose $X_1, ..., X_n$ constitute average temperatures that are distributed as in (3.29), then X_t, where t is the annual maximum temperature in the t^{th} year, follows GEV distribution $(\mu(t), \sigma, \xi)$ where

$$\mu(t) = \mu_0 + \mu_1 t, \tag{3.64}$$

for a linear variation in mean with an intercept parameter μ_0 and a slope parameter μ_1 which expresses the annual rate of change in annual average temperature (Nemukula, 2018). A quadratic model in μ is given by

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2, \tag{3.65}$$

or a change point model,

$$\mu(t) \quad = \quad \begin{cases} \mu_1 & \text{for} \quad t \leq t_0, \\ \mu_2 & \text{for} \quad t > t_0. \end{cases} \tag{3.66}$$

Non-stationarity may also be expressed in terms of exponential variation in the scale parameter given by

$$\sigma(t) = \exp(\sigma_0 + \sigma_1(t)), \tag{3.67}$$

where the exponential function is used to ensure the positivity of σ (Jiang and Kang, 2019).

In case the average daily temperature is related to other variables such as time, the covariates are included in the non-stationary models. This suggests the model Z_t, the annual average temperatures in year t: $Z_t \sim \text{GEV}(\mu(t), \sigma, \xi)$, where

$$\mu(t) = \mu_0 + \mu_1 \text{SOI}(t), \tag{3.68}$$

and $\text{SOI}(t)$ denotes the Southern Oscillation Index in year t (Nemukula, 2018).

The maximum log-likelihood of the non-stationary $\text{GEV}(\mu(t), \sigma(t), \xi(t))$ is given by

$$\begin{aligned} \ell(\mu, \sigma, \xi) &= -\sum_{t=1}^{m} \left\{ \log \sigma(t) + \left(1 + \frac{1}{\xi(t)}\right) \log \left[1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right)\right] \right\} \\ &\quad -\sum_{t=1}^{m} \left\{ \left[1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right)\right]^{-\frac{1}{\xi(t)}} \right\}, \end{aligned} \tag{3.69}$$

where $1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right) > 0, \text{for } \ t = 1, 2, ..., m$ (Chen and Zhao, 2020).

Suppose that $X_1, ..., X_n$ are average daily temperatures that follow a GP distribution. Let the season that contains day t be denoted by $s(t)$ and $u_{s(t)}$ be the threshold for particular seasons, then X_t follows $\text{GP}(\sigma, \xi)$ with the conditional season model given by

$$(X_t - u_{s(t)} | X_t > u_{s(t)}) \sim \text{GP}(\sigma_{s(t)}, \xi_{s(t)}), \tag{3.70}$$

where $(\sigma_{s(t)}, \xi_{s(t)})$ are GP distribution parameters in season $s(t)$ (Nemukula, 2018).

3.10 Model choice

3.10.1 Deviance statistics

Maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other. With models $M_i \subset M_o$, $i = 1, 2, ..., m$, where m denotes the number of the models, the deviance statistics is given as

$$D = 2\{\ell_i(M_i) - \ell_o(M_o)\}, \tag{3.71}$$

where $\ell_i(M_i)$ and $\ell_o(M_o)$ are maximised log-likelihood functions of M_i and M_o, respectively (Chen and Zhao, 2020; Kajambeu, 2016). Large values of D indicate that model M_i explains substantially more of the variation in the data than M_o, while small values of D suggest that the increase in model size does not bring worthwhile improvements in the model's capacity to explain the data. Model M_o is rejected by a test at the α-level of significance if $D > c_\alpha$, where c_α is the (1-α) critical value. The distribution of D is given by the χ^2_k distribution with k degrees of freedom, where k is equal to the difference in the dimensionality of M_i and M_o (Coles, 2001).

3.11 Conditional multivariate extreme value modelling

3.11.1 Time-varying threshold excess

A time-varying threshold $\pi(t)$ is a penalised cubic smoothing with a positive shift operator given by

$$\pi(t) = \sum_{i=1}^{n} \left(y_i - f(t_i)\right)^2 + \lambda \int \left(f''(t)\right)^2 dt + \tau, \tag{3.72}$$

where y_i denotes maximum temperature, λ is a smoothing parameter and $\tau \in \Re$ is a shift factor which should be large enough to allow asymptotic conditions to be satisfied when we fit the GP distribution. We extract observations above the time-varying threshold without the shift factor. The positive shift factor τ is then estimated using high quantiles in the 'texmex' R package (Sigauke and Bere, 2017).

3.11.2 Bivariate threshold excess model

A class of approximations to the tail of the distribution function F for the threshold u, with exceedances denoted by a variable X, on the condition that $X > u$ for large enough u, is given by

$$G(x) = 1 - \zeta \left\{ 1 + \frac{\xi}{\sigma_u}(x-u) \right\}^{-\frac{1}{\xi}}, \quad x > u, \tag{3.73}$$

where $\zeta = Pr(X > u)$, $\xi \neq 0$ and $\sigma_u > 0$ for a family defined on $\{(1+\xi(x-u)/\sigma_u) > 0$ and $x - u : x - u > 0\}$. Hence $F(x) \approx G(x)$ on $x > u$ for large enough threshold u with parameters ζ, ξ and σ_u (Nemukula et al., 2018).

Suppose $\{(x_1, y_1), (x_2, y_2), ...\}$ are independent realisations of a random vector (X, Y) with joint distribution function $F(x, y)$. The bivariate threshold excess model approximates the joint distribution $F(x, y)$ on regions of the form $x > u_x, y > u_y$ for large enough u_x and u_y. For suitable threshold u_x and u_y, each of the two marginal distributions of F can be approximated in the form of a univariate GP distribution, with respective parameter sets (μ_x, ξ_x) and (μ_y, ξ_y) (Nemukula et al., 2018; Coles, 2001).

The transformations

$$\tilde{X} = -\left(\log\left\{1 - \zeta_x\left[1 + \frac{\xi_x(X - u_x)}{\sigma_x}\right]^{-\frac{1}{\xi_x}}\right\}\right)^{-1} \tag{3.74}$$

and

$$\tilde{Y} = -\left(\log\left\{1 - \zeta_y\left[1 + \frac{\xi_y(Y - u_y)}{\sigma_y}\right]^{-\frac{1}{\xi_y}}\right\}\right)^{-1} \tag{3.75}$$

induce a variable $(\tilde{X}, \tilde{Y})$ whose distribution function $\tilde{F}$ has margins that are approximately standard Fr*é*chet for $X > u_x$ and $Y > u_y$. The joint distribution can be expressed as

$$F(x, y) \approx G(x, y) = \exp\{-V(\tilde{x}, \tilde{y})\}, \quad x > u_x,\ y > u_y, \tag{3.76}$$

where

$$V(\tilde{x}, \tilde{y}) = 2\int_0^1 max\left(\frac{w}{\tilde{x}}, \frac{1-w}{\tilde{y}}\right) dH(w), \tag{3.77}$$

and H is a distribution function on [0,1] satisfying the mean constraint

$$\int_0^1 w dH(w) = \frac{1}{2}. \tag{3.78}$$

3.11.3 Marginal transformation: Laplace margins

Suppose $\mathbf{X} = (\mathbf{X_1}, ..., \mathbf{X_d})$ is a d-dimension random variable whose marginal distributions are arbitrary, and $\widehat{F_j}$ denotes an estimate of the j^{th}, $j = 1, ..., d$, marginal distribution function with G indicating the distribution function of the standardised marginal distribution to be determined. The transformation of the original vector variable $\mathbf{X}$ is done to obtain $\mathbf{Y} = (\mathbf{Y_1}, ..., \mathbf{Y_d})$, where $\mathbf{Y}$ is a variable which has standardised marginal distributions due to the use of the

probability integral transformation (Nemukula et al., 2018):

$$\mathbf{Y_j} = (\mathbf{G^{-1}}(\widehat{\mathbf{F}}_\mathbf{j}(\mathbf{X_j}))), \mathbf{j} = \mathbf{1}, ..., \mathbf{d}. \tag{3.79}$$

Let G denotes the Laplace margins and assume $\mathbf{Y} = \{Y_1, ..., Y_d\}$ are marginally Laplace distributed. If the $\mathbf{Y}_j$ variable exceeds a sufficiently high threshold u, then the Heffernan and Tawn (2004) regression type model is given as:

$$\mathbf{Y}_{-j} = \alpha_{|j}\mathbf{Y}_j + (\mathbf{Y}_j)^{\beta_{|j}}\,\mathbf{R}_{|j}, \tag{3.80}$$

where $\mathbf{R}_{|}j$ is a residual vector, $\alpha_{|j}$ and $\beta_{|j}$ are $(d-1)$ dimensional parameter vectors satisfying $(\alpha_{|j}, \beta_{|j}) \,\epsilon\, [-1, 1]^{d-1} \times (-\infty, 1)^{d-1}$, with elements $\alpha_{i|j}$ and $\beta_{i|j}$, $i \,\epsilon\, \{1, ..., d\}$, $i \neq j$ (Nemukula et al., 2018). In the present study the bivariate threshold excess is applied only to the maximum temperature data.

Chapter 4

Results and discussion

4.1 Introduction

This chapter presents the analysis and discussion of results using methods that were presented in Chapter 3 for the following meteorological stations: Mara (1949-2018), Messina (1934-2009), Polokwane (1932-2018) and Thabazimbi (1994-2018) (Figure 3.1). The block maxima (BM) approach was used to analyse annual maxima temperature data, where the years were taken as independent and identically distributed (iid) blocks. R software was used to analyse the data.

4.2 Descriptive statistics

Descriptive statistics analysis was performed to obtain variation, centralisation and distribution of maximum and minimum temperature data. The mean is utilised to measure central tendency of the data while standard deviation is

utilised to measure variation of the data. Skewness and Kurtosis are used to obtain the shape of the distribution for monthly temperature data.

4.2.1 Mara temperature record (1949-2018)

Table 4.1: Descriptive statistics of Mara maximum and minimum temperatures.

Variable	Min	Max	Mean	Standard deviation	Range	Skewness	Kurtosis
Max temp	19.10	34.40	27.39	2.90	15.3	-0.24	-0.71
Min temp	1.10	20.00	12.42	4.92	18.9	-0.42	-1.21

Table 4.1 shows the descriptive statistics of both maximum and minimum temperature data for Mara meteorological station. The highest maximum temperature for Mara was 34.40 degrees Celsius recorded in December 2006, while the lowest maximum temperature was 19.10 degrees Celsius recorded in July 2009. The lowest minimum temperature for Mara was 1.10 degrees Celsius recorded in June 2010, while the highest minimum temperature was 20.00 degrees Celsius recorded in December 2015. The minimum temperature for Mara has a standard deviation of 4.92 degrees Celsius, while the maximum temperature has a standard deviation of 2.90 degrees Celsius. For maximum and minimum temperatures, the skewness is -0.24 and -0.42 respectively. These values imply that the distribution of the data is skewed to the left or negatively skewed. For the kurtosis of the maximum and minimum temperatures, we have -0.71 and -1.21 respectively, implying that the distribution of the data is platykurtic, since the computed values are less than 3.

4.2.2 Messina temperature record (1934-2009)

Table 4.2 shows the descriptive statistics of both maximum and minimum temperature data for Messina meteorological station. The highest maximum tem-

Table 4.2: Descriptive statistics of Messina maximum and minimum temperatures.

Variable	Min	Max	Mean	Standard deviation	Range	Skewness	Kurtosis
Max temp	7.50	37.70	29.96	3.35	30.2	-0.52	1.36
Min temp	3.30	24.00	15.39	5.23	20.7	-0.43	-1.16

perature for Messina was 37.70 degrees Celsius recorded in December 2009, while the lowest maximum temperature was 7.50 degrees Celsius recorded in May 1938. The lowest minimum temperature for Messina was 3.30 degrees Celsius recorded in June 1943, while the highest minimum temperature was 24.00 degrees Celsius recorded in January 1983. The minimum temperature for Messina has a standard deviation of 5.23 degrees Celsius, while the maximum temperature has a standard deviation of 3.35 degrees Celsius. For maximum and minimum temperatures, the skewness is -0.53 and -0.43 respectively. These values imply that the distribution of the data is skewed to the left or negatively skewed. For the kurtosis of the maximum and minimum temperatures, we have 1.36 and -1.1.6 respectively, implying that the distribution of the data is platykurtic, since the computed values are less than 3.

4.2.3 Polokwane temperature record (1932-2018)

Table 4.3: Descriptive statistics of Polokwane maximum and minimum temperatures.

Variable	Min	Max	Mean	Standard deviation	Range	Skewness	Kurtosis
Max temp	16.50	30.90	25.02	2.98	14.4	-0.44	-0.73
Min temp	2.90	18.70	11.70	4.63	16.6	-0.37	-1.73

Table 4.3 shows the descriptive statistics of both maximum and minimum temperature data for Polokwane meteorological station. The highest maximum temperature for Polokwane was 30.90 degrees Celsius recorded in Decem-

ber 1972, while the lowest maximum temperature was 16.50 degrees Celsius recorded in June 1968. The lowest minimum temperature for Polokwane was 2.9 degrees Celsius recorded in June 1972, while the highest minimum temperature was 18.7 degrees Celsius recorded in January 2005. The minimum temperature for Polokwane has a standard deviation of 4.63 degrees Celsius, while the maximum temperature has a standard deviation of 2.98 degrees Celsius. For maximum and minimum temperatures, the skewness is -0.44 and -0.37 respectively. These values imply that the distribution of the data is skewed to the left or negatively skewed. For the kurtosis of the maximum and minimum temperatures, we have -0.73 and -1.73 respectively, implying that the distribution of the data is platykurtic, since the computed values are less than 3.

4.2.4 Thabazimbi temperature record (1994-2018)

Table 4.4: Descriptive statistics of Thabazimbi maximum and minimum temperatures.

Variable	Min	Max	Mean	Standard deviation	Range	Skewness	Kurtosis
Max temp	21.30	39.60	29.34	3.853	18.3	-0.256	-0.838
Min temp	0.10	24.00	13.350	6.489	24.1	-0.484	-1.229

Table 4.4 shows the descriptive statistics of both maximum and minimum temperature data for Thabazimbi meteorological station. The highest maximum temperature for Thabazimbi was 39.60 degrees Celsius recorded in December 1996, while the lowest maximum temperature was 21.30 degrees Celsius recorded in July 1996. The lowest minimum temperature for Thabazimbi was 0.1 degrees Celsius recorded in July 1994, while the highest minimum temperature was 24.00 degrees Celsius recorded in December 1997. The minimum temperature for Thabazimbi has a standard deviation of 6.489 degrees Celsius, while the maximum temperature has a standard deviation of 3.853 degrees Celsius. For maximum and minimum temperatures, the skewness is -0.256

and -0.484 respectively. These values imply that the distribution of the data is skewed to the left or negatively skewed. For the kurtosis of the maximum and minimum temperatures, we have -0.838 and -1.229 respectively, implying that the distribution of the data is platykurtic, since the computed values are less than 3.

4.2.5 Summary of descriptive statistics

In the four stations, maximum and minimum temperature series are not stable throughout the year. From May both maximum and minimum temperatures drop significantly till July and start rising again. The lowest temperatures are experienced in June, July and August with July being the coldest month of the year. High temperatures are experienced in the months of December, January, February and March.

4.3 Stationarity test results

The stationarity test results are presented in Table 4.5 and Table 4.6 for the average maximum and minimum temperatures, respectively. The significance level, α, was taken as 0.05. The ADF test results for both maximum and minimum temperatures for all stations are greater than the significance level (p > 0.05), suggesting that we fail to reject the null hypothesis, H_0, and conclude that the temperature series is not stationary or it has a unit root. The KPSS test results for both maximum and minimum temperatures for all stations are less than the significance level (p < 0.05), suggesting that we reject the null hypothesis, H_0, and conclude that the temperature series for all stations are not trend stationary.

The PP test results for maximum temperature for all stations are less than the

Table 4.5: ADF, KPSS and PP test results for maximum temperature.

Station name	Test name	Test statistics	p-value
Mara	ADF	-1.905	0.619
	KPSS	2.516	0.01
	PP	-159.3	0.01
Messina	ADF	-1.883	0.628
	KPSS	2.840	0.01
	PP	-176.64	0.01
Polokwane	ADF	-1.871	0.633
	KPSS	2.705	0.01
	PP	-184.73	0.01
Thabazimbi	ADF	-1.468	0.800
	KPSS	1.165	0.01
	PP	-53.973	0.01

Table 4.6: ADF, KPSS and PP test results for minimum temperature.

Station name	Test name	Test statistics	p-value
Mara	ADF	-0.852	0.957
	KPSS	3.008	0.01
	PP	-13.359	0.365
Messina	ADF	-0.837	0.958
	KPSS	3.144	0.01
	PP	-19.306	0.082
Polokwane	ADF	-1.271	0.887
	KPSS	3.178	0.01
	PP	-40.249	0.01
Thabazimbi	ADF	-0.860	0.956
	KPSS	1.201	0.01
	PP	-6.779	0.730

significance level ($p < 0.05$), suggesting that we reject the null hypothesis, H_0, and conclude that the temperature series for all the stations are stationary, while for minimum temperature we fail to reject the null hypothesis, H_0, since p-value is greater than the significance level ($p > 0.05$) and conclude that the temperature series for all stations are not stationary, except for Polokwane minimum temperature series where the null hypothesis is rejected since $p < 0.05$ and conclude that the temperature series is stationary.

4.4 Normality test results

The normality test results are presented in Table 4.7 and Table 4.8 for the average maximum and minimum temperatures, respectively. The significance level, α, was taken as 0.05. The p-values for both maximum and minimum temperatures for all stations are less than the significance level (p< 0.05), suggesting that the null hypothesis, H_0, is rejected and conclude that the temperature series for all the stations is not normally distributed.

Table 4.7: Shapiro-Wilk (SW) and Jarque-Bera (JB) test results for maximum temperature.

Station name	Test name	Test statistics	p-value
Mara	Shapiro-Wilk	0.981	<0.001
	Jarque-Bera	25.554	<0.001
Messina	Shapiro-Wilk	0.970	<0.001
	Jarque-Bera	110.69	<0.001
Polokwane	Shapiro-Wilk	0.967	< 0.001
	Jarque-Bera	27.43	<0.001
Thabazimbi	Shapiro-Wilk	0.966	<0.001
	Jarque-Bera	11.771	0.003

Table 4.8: Shapiro-Wilk (SW) and Jarque-Bera (JB) test results for minimum temperature.

Station name	Test name	Test statistics	p-value
Mara	Shapiro-Wilk	0.915	<0.001
	Jarque-Bera	75.121	<0.001
Messina	Shapiro-Wilk	0.921	<0.001
	Jarque-Bera	78.103	<0.001
Polokwane	Shapiro-Wilk	0.930	<0.001
	Jarque-Bera	68.317	<0.001
Thabazimbi	Shapiro-Wilk	0.894	<0.001
	Jarque-Bera	30.388	<0.001

4.5 Parent distribution selection and diagnostic statistics test results

The four candidate parent distributions were fitted to the data. The best fitting parent distribution was selected according to the values of Akaike information criterion (AIC) and Bayesian information criterion (BIC). The lowest value of AIC and BIC was chosen to be the best.

Table 4.9: Summary of parent distributions information criteria model selection tests for maximum temperature.

Station name	Distribution	AIC value	BIC value
Mara	Normal	4231.283	4240.75
	Log-normal	4262.193	4271.66
	Gamma	4249.286	4258.753
	Weibull	**4221.385**	**4230.852**
Messina	Normal	4733.907	4743.512
	Log-normal	4858.128	4867.732
	Gamma	4801.965	4811.57
	Weibull	**4709.401**	**4719.006**
Polokwane	Normal	5305.021	5314.9
	Log-normal	5305.021	5314.9
	Gamma	5348.834	5358.712
	Weibull	**5282.204**	**5292.082**
Thabazimbi	Normal	1663.706	1671.114
	Log-normal	1677.577	1684.984
	Gamma	1671.713	1679.121
	Weibull	**1656.678**	**1664.085**

Table 4.10: Summary of parent distributions information criteria model selection tests for minimum temperature.

Station name	Distribution	AIC value	BIC value
Mara	Normal	5064.259	5073.726
	Log-normal	5309.453	5318.92
	Gamma	5183.499	5192.966
	Weibull	**5050.361**	**5059.828**
Messina	Normal	5536.302	5542.907
	Log-normal	5740.935	5750.557
	Gamma	5643.047	5652.652
	Weibull	**5500.478**	**5510.083**
Polokwane	Normal	**6247.431**	**6257.333**
	Log-normal	6647.782	6657.684
	Gamma	6419.108	6429.009
	Weibull	**6253.977**	**6263.878**
Thabazimbi	Normal	**1976.296**	**1983.703**
	Log-normal	2156.169	2163.577
	Gamma	2049.755	2057.162
	Weibull	**1999.005**	**2006.412**

The parent distributions goodness-of-fit test results are presented in Table 4.9 and Table 4.10 for the average maximum and minimum temperatures, respectively. The results revealed that the best fitting parent distribution for maximum and minimum temperatures is in the Weibull domain of attraction at all the stations, except for Thabazimbi and Polokwane, where the best fitting parent distributions for the minimum temperature were found to be in the domain of attraction of the normal distribution for both stations. However, given that the results from the SW and JB normality tests in Table 4.7 and Table 4.8 showed that the maximum and minimum temperature series for all the stations are not normally distributed, this therefore means that the minimum temperature series for Polokwane and Thabazimbi cannot be in the normal domain of attraction.

4.5.1 Diagnostic plots illustrating the fit of the average maximum and minimum temperatures

The diagnostic statistics goodness-of-fit tests, quantile-quantile (Q-Q) plots, probability-probability (P-P) plots, empirical and theoretical density plots and cumulative distribution function (CDF) plots are presented in Figure 4.1 to Figure 4.10. The results of the diagnostic plots for Figures 4.1, 4.3, 4.5 and 4.8 suggest that the maximum temperature series for all the four stations in the Limpopo province can be suitably modelled by a distribution in the Weibull domain of attraction. The minimum temperature series for all the stations, Figures 4.2, 4.4, 4.6 and 4.9 present the results of the diagnostic plots for the Weibull distribution, while Figures 4.7 and 4.10 present the results of the diagnostic plots for the normal distribution. The main thrust for presenting both the Weibull and normal distributions diagnostic plots for some stations were to make a visual comparative analysis of the plots.

The visual comparative analysis was necessary for Polokwane and Thabazimbi where the model selection information criteria, AIC and BIC, results contradict those of the SW and JB normality tests. Based on the AIC and BIC results, the Weibull distribution was the second best distribution for both Polokwane and Thabazimbi with values very close to those of the normal distribution, while there was no doubt about the performance of the Weibull class of distributions for both Mara and Messina minimum temperature series. However, Figures 4.6 and 4.7 for the Polokwane minimum temperature series Weibull and normal diagnostic plots, respectively, and Figures 4.9 and 4.10 for the Thabazimbi minimum temperature series Weibull and normal diagnostic plots, respectively, do not exhibit much difference in their goodness-of-fit results for both stations. Combining these findings with those from normality tests and information criteria model selection we can conclude that minimum temperature series parent distribution for all the stations in this study follow the Weibull domain of

attraction.

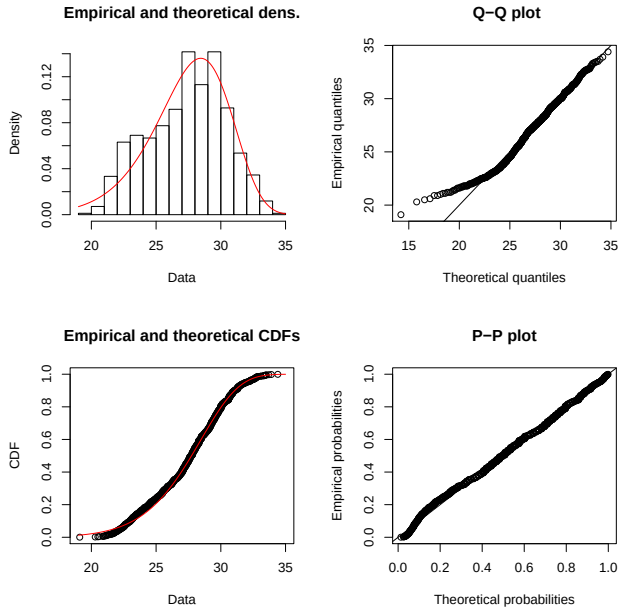

Figure 4.1: Diagnostic plots illustrating the fit of the Weibull family of distributions to average maximum temperature for Mara.

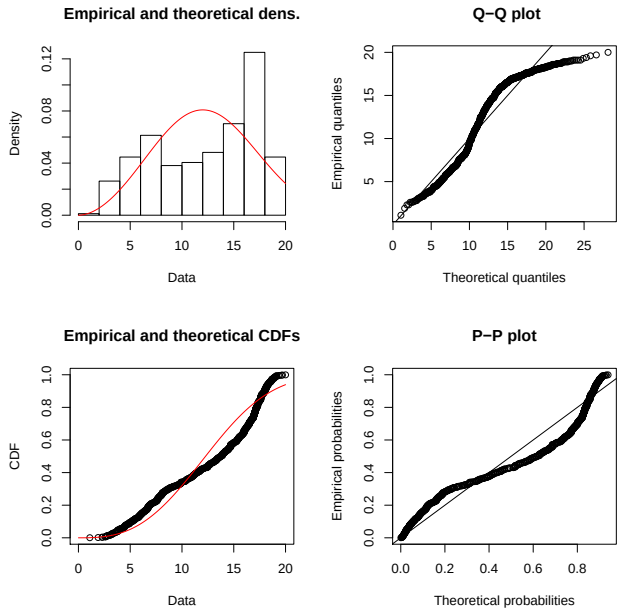

Figure 4.2: Diagnostic plots illustrating the fit of the Weibull family of distributions to average minimum temperature for Mara.

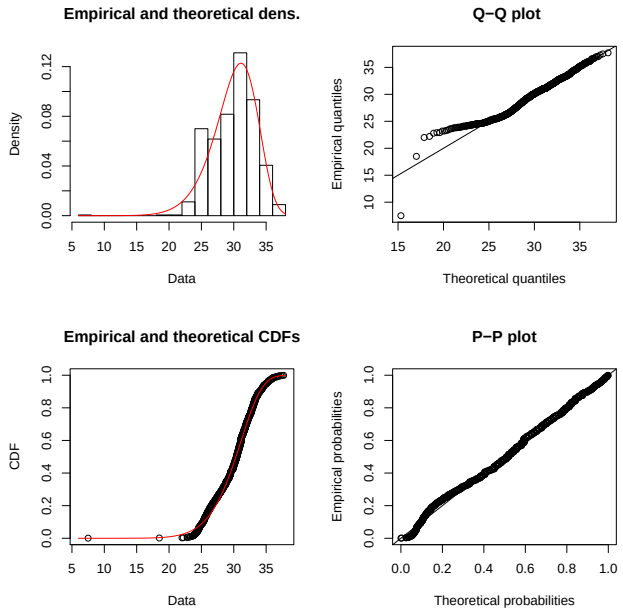

Figure 4.3: Diagnostic plots illustrating the fit of the Weibull family of distributions to average maximum temperature for Messina.

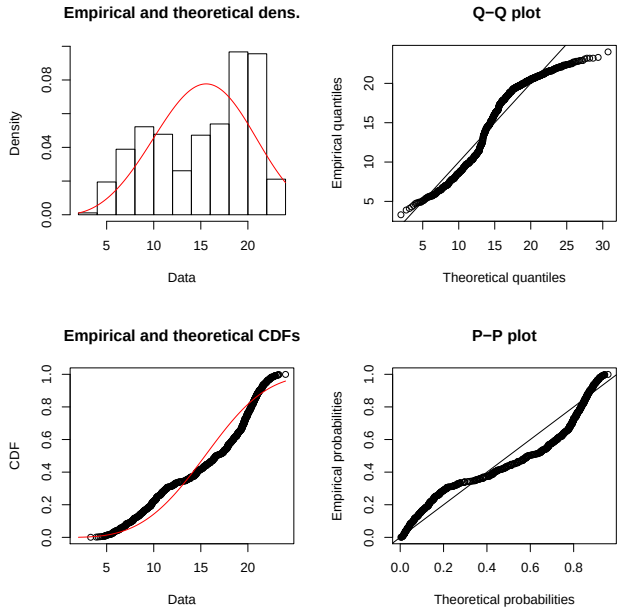

Figure 4.4: Diagnostic plots illustrating the fit of the Weibull family of distributions to average minimum temperature for Messina.

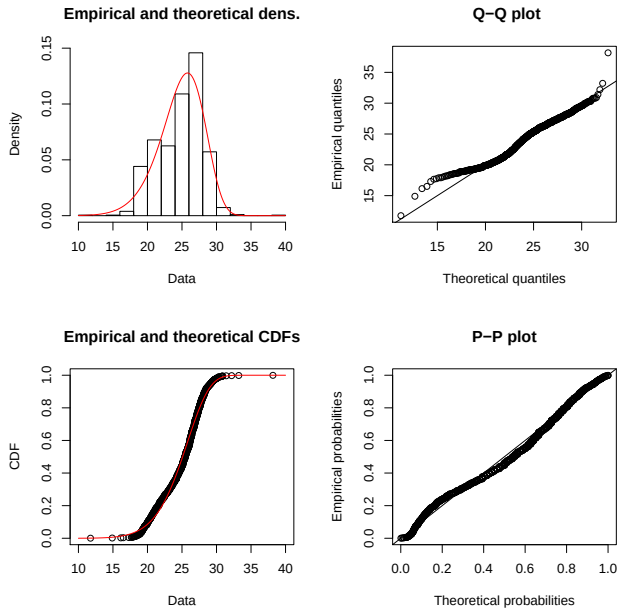

Figure 4.5: Diagnostic plots illustrating the fit of the Weibull family of distributions to average maximum temperature for Polokwane.

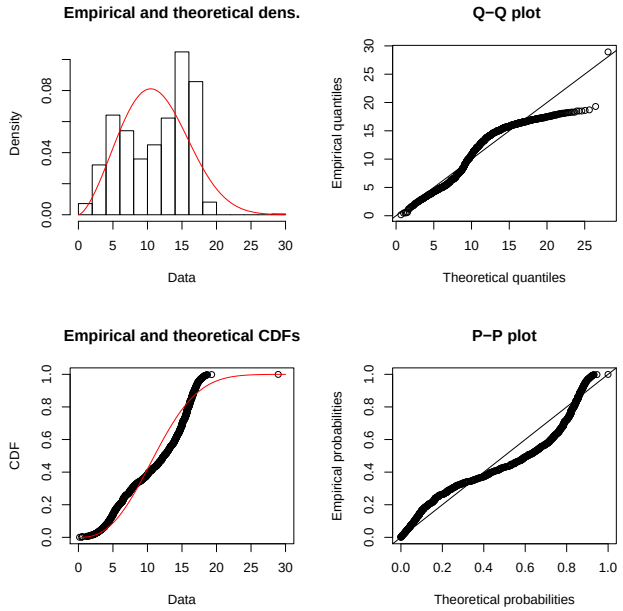

Figure 4.6: Diagnostic plots illustrating the fit of the Weibull family of distributions to average minimum temperature for Polokwane.

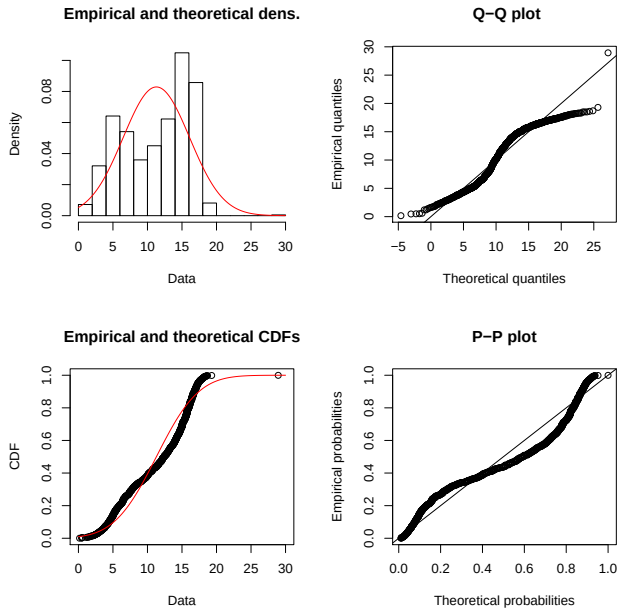

Figure 4.7: Diagnostic plots illustrating the fit of the normal distribution to average minimum temperature for Polokwane.

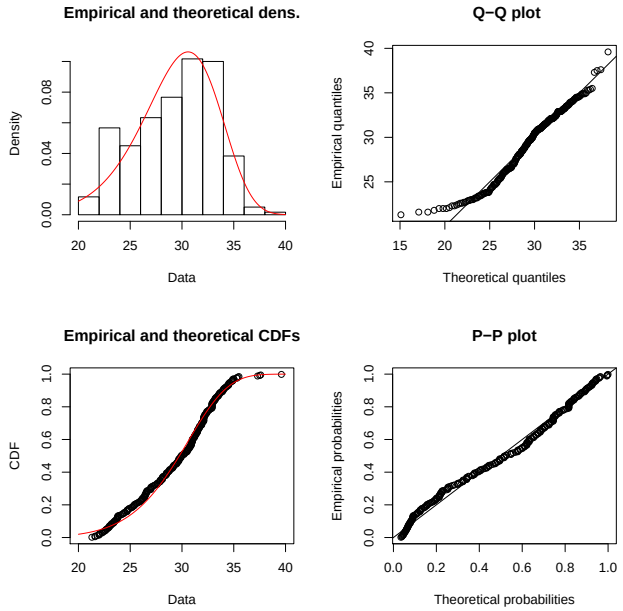

Figure 4.8: Diagnostic plots illustrating the fit of the Weibull family of distributions to average maximum temperature for Thabazimbi.

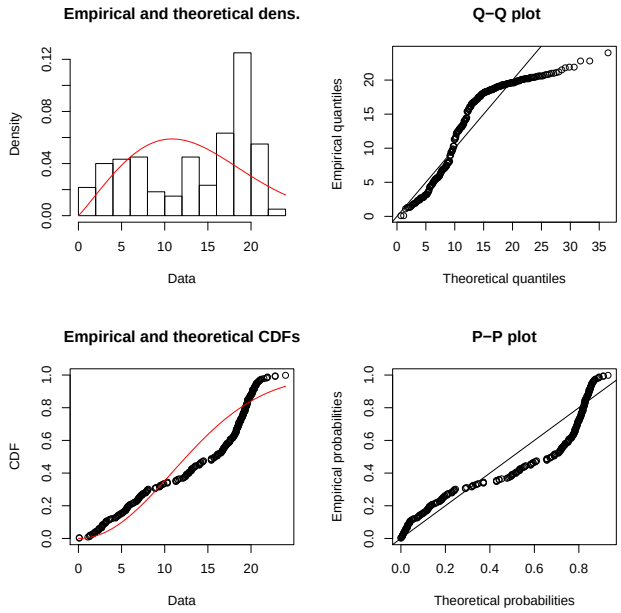

Figure 4.9: Diagnostic plots illustrating the fit of the Weibull family of distributions to average minimum temperature for Thabazimbi.

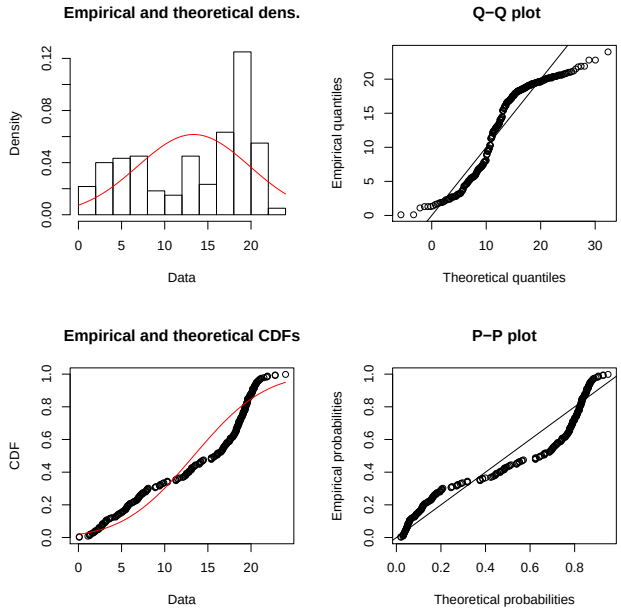

Figure 4.10: Diagnostic plots illustrating the fit of the normal distribution to average minimum temperature for Thabazimbi.

4.6 Long-term trends analysis results

The Mann-Kendall (M-K) test statistic trend analysis results are presented in Table 4.11 for the average maximum and minimum temperatures. The significance level, α, was taken as 0.05. The p-values for minimum temperature for Mara, Messina and Polokwane meteorological stations are less than the significance level ($p < 0.05$), suggesting that there is a significant trend in the minimum temperature data for these stations, while for the maximum temperature the results suggest that there is no significant trend in the data for all the stations ($p > 0.05$). For both maximum and minimum temperatures in Thabazimbi, the results revealed that there is no significant trend in the data since p-values are greater than the significance level ($p > 0.05$).

Table 4.11: Summary of Mann-Kendall trend analysis.

Station name		M-K test statistic(S)	Kendall's τ	Var(S)	P-value
Mara	Max	71.987	0.138	7233.832	0.659
	Min	-155.681	-1.099	7234.027	0.0004
Messina	Max	26.031	0.0160	8896.256	0.886
	Min	-120.334	-0.741	8896.724	0.014
Polokwane	Max	-119.951	-0.153	3645.197	0.589
	Min	-157.297	-0.718	3773.997	0.0105
Thabazimbi	Max	60.096	0.907	1216.053	0.085
	Min	-24.605	-0.371	1216.122	0.481

The M-K test statistic and Kendall's τ values are all negative for the minimum temperature suggesting that there is a monotonic decreasing trend in minimum temperature for all the stations. Therefore, there is a significant monotonic decreasing trend in minimum temperature for the stations Mara, Messina and Polokwane ($\tau < 0$ and $p < 0.05$), while for Thabazimbi the monotonic decreasing trend in minimum temperature is insignificant ($\tau < 0$ and $p > 0.05$). For maximum temperature, there is an insignificant monotonic increasing trend in maximum temperature for Mara, Messina and Thabazimbi

($\tau > 0$ and p> 0.05), while for Polokwane there is an insignificant monotonic decreasing trend in maximum temperature ($\tau < 0$ and p> 0.05).

These findings are also supported by the time series plot results in Figure 4.11 to Figure 4.14. For both maximum and minimum temperature data, the results show that there is no evidence of systematic variation about the mean on the time series plots; and there is a clear upward and downward trend indicating that the series is non-stationary.

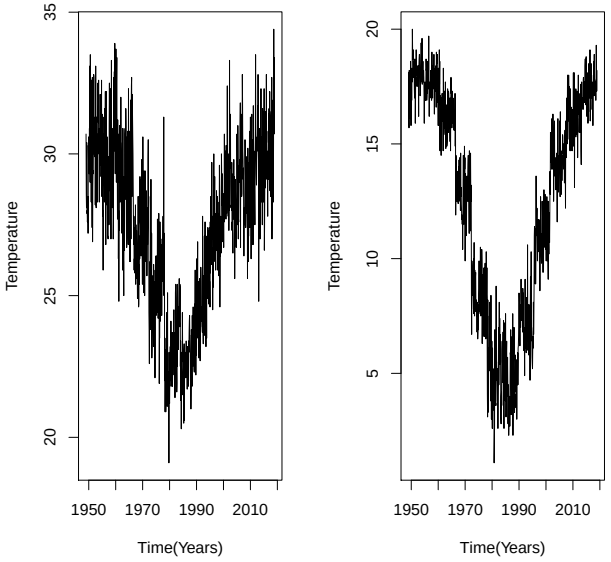

Figure 4.11: Time series plot of Mara maximum (left panel) and minimum (right panel) temperatures from 1949 to 2018.

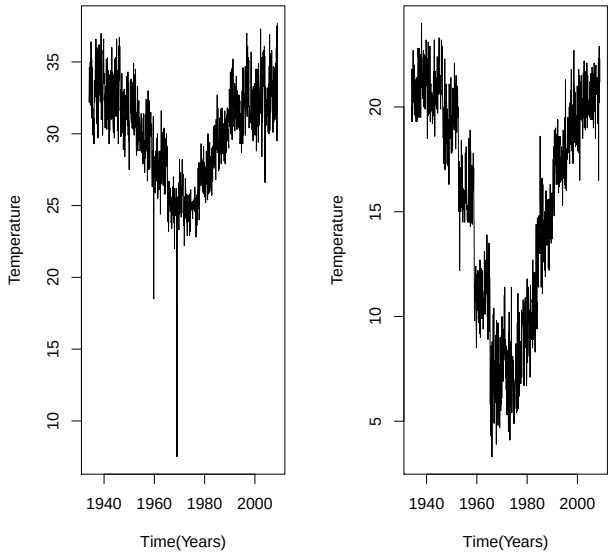

Figure 4.12: Time series plot of Messina maximum (left panel) and minimum (right panel) temperatures from 1934 to 2009.

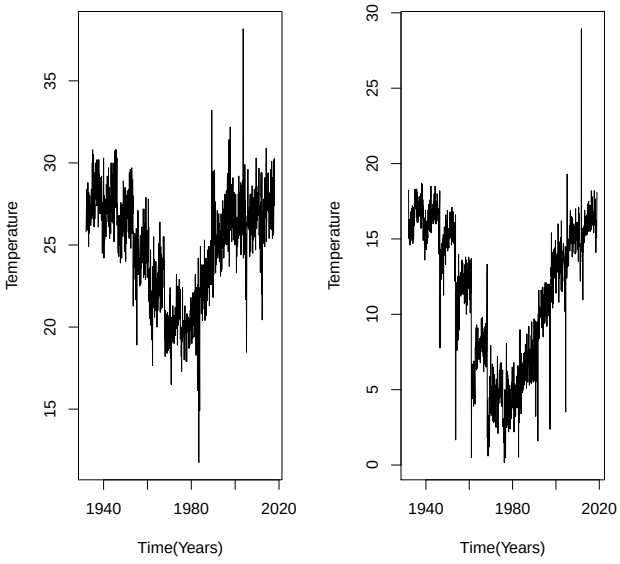

Figure 4.13: Time series plot of Polokwane maximum (left panel) and minimum (right panel) temperatures from 1932 to 2018.

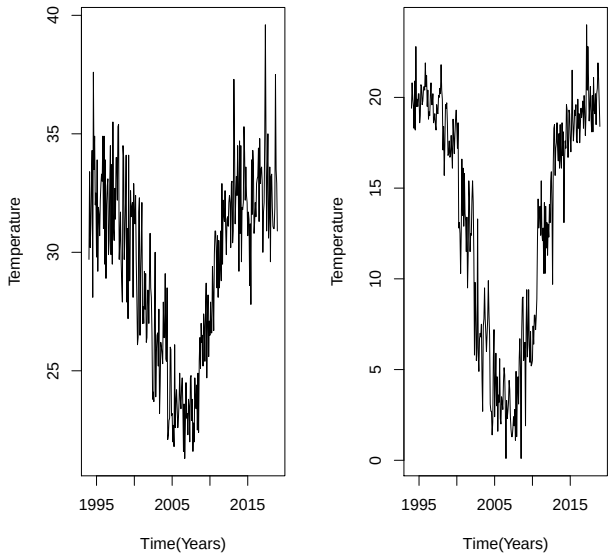

Figure 4.14: Time series plot of Thabazimbi maximum (left panel) and minimum (right panel) temperatures from 1994 to 2018.

4.7 Fitting the GEV distribution

Table 4.12 and Table 4.13 show the maximum likelihood estimates of the stationary generalised extreme value (GEV) distribution (M_0) with standard errors and 95% confidence intervals (CI) for parameter estimates. The estimates and standard errors were combined to give approximate CI. In particular, a 95% CI for ξ was obtained as $\xi \pm 1.96(SE)$. The results revealed that both maximum and minimum temperature data for Mara and Messina meteorological stations can be modelled by the Weibull family of distribution since $\xi < 0$ and CI is significantly different from zero, while both maximum and minimum temperature data for Polokwane and Thabazimbi can be modelled by the Gumbel family of distribution since CI for ξ is not significantly different from zero.

Table 4.12: Maximum likelihood estimates of the GEV distribution parameters with standard errors and 95% CI for maximum temperature.

Station name	Parameter	Estimate	SE	95% CI of parameter estimates
Mara	μ	30.929	0.186	(30.564,31.294)
	σ	1.412	0.134	(1.149,1.675)
	ξ	-0.338	0.078	(-0.491,-0.185)
Messina	μ	34.135	0.199	(33.745,34.525)
	σ	1.569	0.143	(1.289,1.849)
	ξ	-0.369	0.076	(-0.518,-0.220)
Polokwane	μ	28.246	0.131	(27.989,28.503)
	σ	1.086	0.097	(0.902,1.270)
	ξ	-0.082	0.076	(-0.231,0.067)
Thabazimbi	μ	33.643	0.288	(33.079,34.207)
	σ	1.313	0.201	(0.919,1.707)
	ξ	-0.008	0.117	(-0.237,0.221)

Table 4.13: Maximum likelihood estimates of the GEV distribution parameters with standard errors and 95% CI for minimum temperature.

Station name	Parameter	Estimate	SE	95% CI of parameter estimates
Mara	$\hat{\mu}$	18.035	0.080	(17.878,18.192)
	$\hat{\sigma}$	0.612	0.056	(0.502,0.722)
	$\hat{\xi}$	-0.228	0.069	(-0.363,-0.093)
Messina	$\hat{\mu}$	21.166	0.132	(20.907,21.425)
	$\hat{\sigma}$	1.063	0.092	(0.883,1.243)
	$\hat{\xi}$	-0.330	0.055	(-0.438,-0.222)
Polokwane	$\hat{\mu}$	16.651	0.110	(16.435,16.867)
	$\hat{\sigma}$	0.954	0.075	(0.807,1.101)
	$\hat{\xi}$	0.032	0.039	(-0.044,0.108)
Thabazimbi	$\hat{\mu}$	20.266	0.209	(19.856,20.676)
	$\hat{\sigma}$	0.951	0.144	(0.669,1.233)
	$\hat{\xi}$	-0.062	0.118	(-0.293,0.169)

4.7.1 Diagnostic analysis

The diagnostic plots for assessing the accuracy of the GEV model fitted to all meteorological stations are displayed in Figure 4.15 to Figure 4.22. The P-P plot, Q-Q plot, return level plot and density plot for both maximum and minimum temperatures for Mara, Messina and Polokwane meteorological stations, show that the plotted points are linear, suggesting that the GEV distribution fits the temperature data well. For Thabazimbi, maximum and minimum temperatures show that there are deviations from linearity, suggesting that the GEV distribution model does not fit the temperature data well.

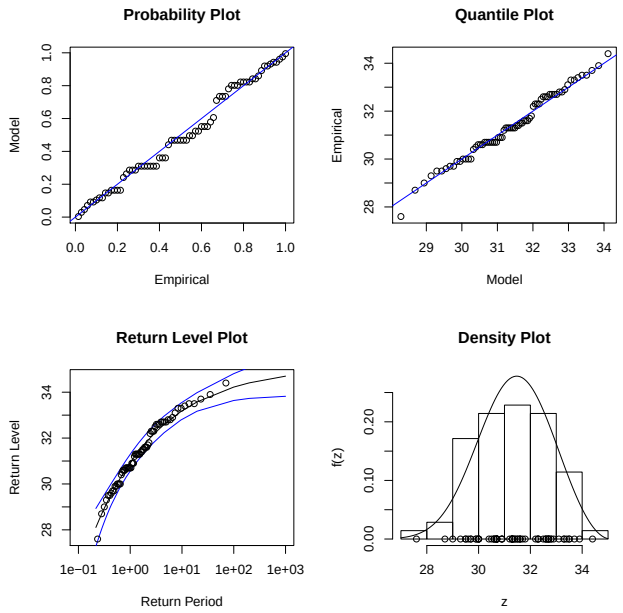

Figure 4.15: Diagnostic plots illustrating the fit of the Mara maximum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

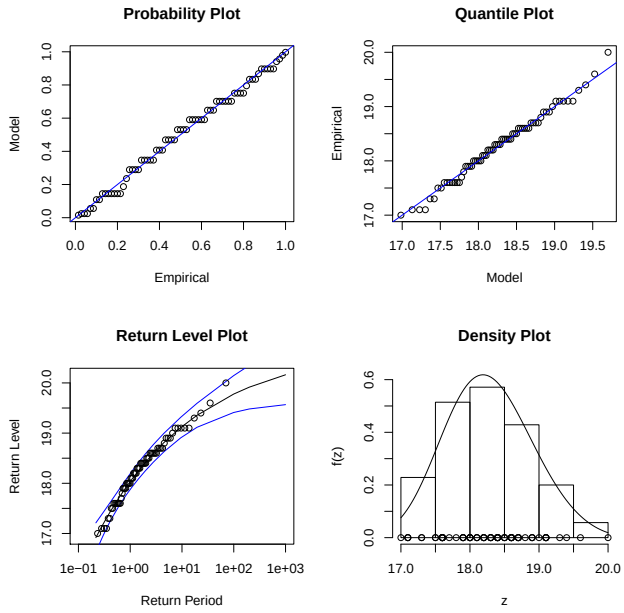

Figure 4.16: Diagnostic plots illustrating the fit of the Mara minimum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

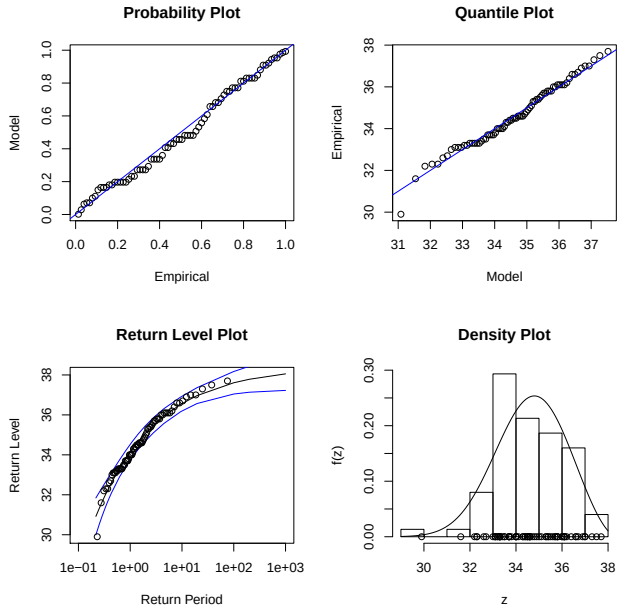

Figure 4.17: Diagnostic plots illustrating the fit of the Messina maximum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

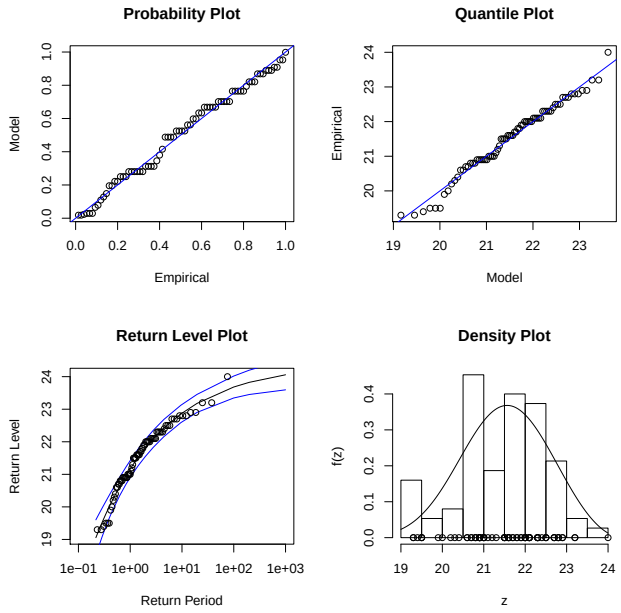

Figure 4.18: Diagnostic plots illustrating the fit of the Messina minimum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

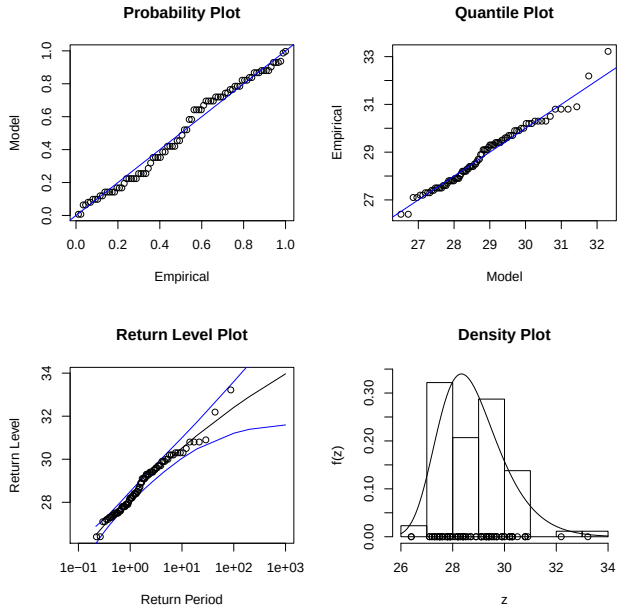

Figure 4.19: Diagnostic plots illustrating the fit of the Polokwane maximum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

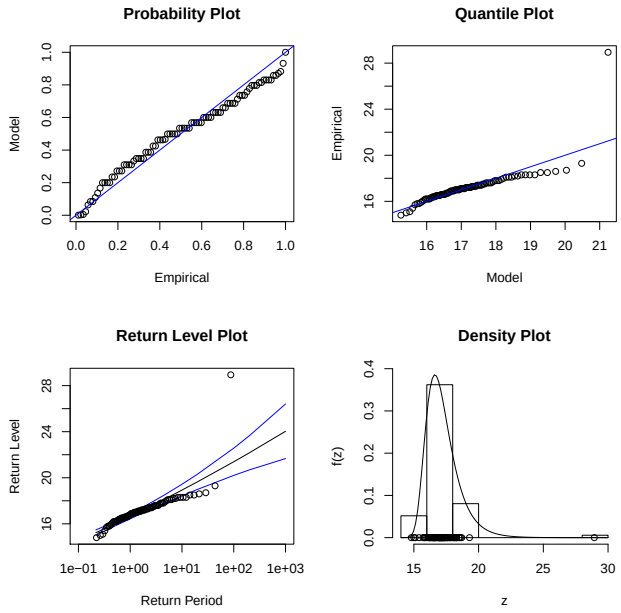

Figure 4.20: Diagnostic plots illustrating the fit of the Polokwane minimum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

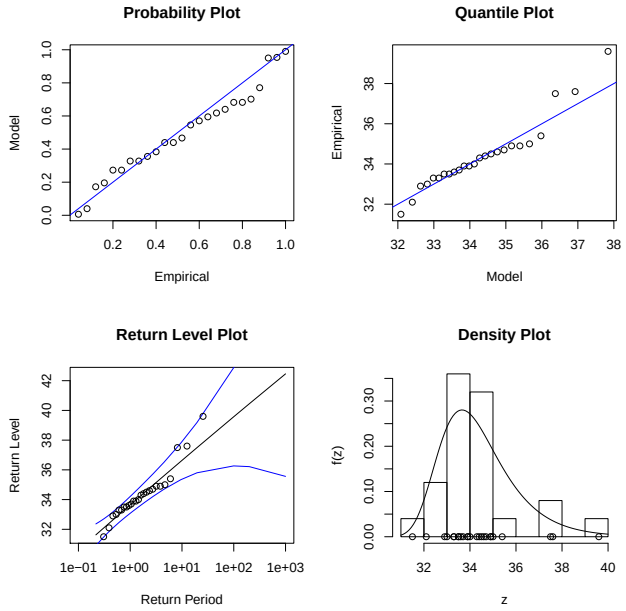

Figure 4.21: Diagnostic plots illustrating the fit of the Thabazimbi maximum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

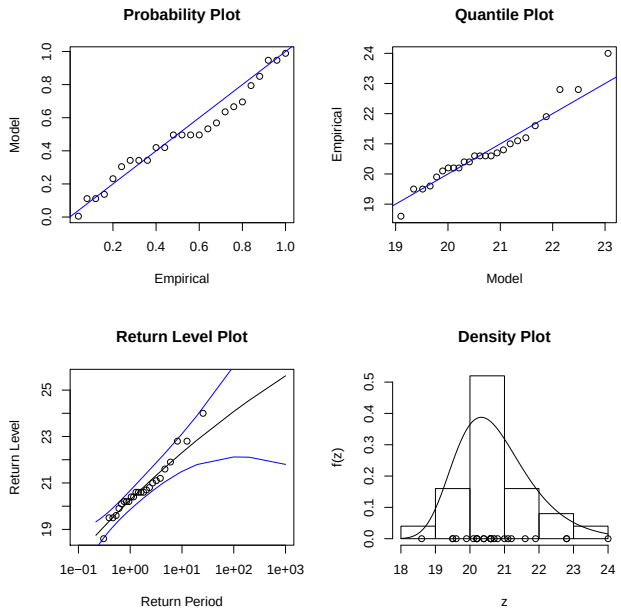

Figure 4.22: Diagnostic plots illustrating the fit of the Thabazimbi minimum temperature to the GEV distribution, (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

4.7.2 Goodness-of-fit test

The goodness-of-fit test results are presented in Table 4.14 and Table 4.15 for average maximum and minimum temperatures, respectively. The significance level, α, was taken as 0.05. The p-values for both maximum and minimum temperatures are greater than the significance level (p> 0.05). This suggests that we fail to reject the null hypothesis, H_0, and conclude that the temperature data follow a specified distribution, except for Polokwane minimum temperature for A-D test results where p-value is less than the significance level (p< 0.05), suggesting that the temperature data do not follow the specified distribution. The Messina minimum temperature data produced not applicable (N/A) results for both tests.

Table 4.14: Goodness-of-fit test for the maximum temperature.

Station name	Test	Test statistic	p-value
Mara	A-D	0.412	0.332
	K-S	0.077	0.795
Messina	A-D	0.400	0.354
	K-S	0.0798	0.748
Polokwane	A-D	0.635	0.095
	K-S	0.090	0.479
Thabazimbi	A-D	0.507	0.182
	K-S	0.138	0.726

Table 4.15: Goodness-of-fit test for the minimum temperature.

Station name	Test	Test statistic	p-value
Mara	A-D	0.292	0.596
	K-S	0.069	0.888
Messina	A-D	N/A	N/A
	K-S	N/A	N/A
Polokwane	A-D	1.597	0.000
	K-S	0.100	0.340
Thabazimbi	A-D	0.438	0.271
	K-S	0.112	0.915

Key: N/A represents Not Applicable

4.7.3 Return level analysis

Table 4.16 presents results for return periods and their corresponding return levels for both average monthly maximum and minimum temperatures based on the GEV model. The results in Table 4.16 revealed that return levels for each station are increasing with increasing return periods. In comparison, Thabazimbi meteorological station has highest return levels for maximum temperature, while Polokwane has highest return levels for minimum temperature data.

Thabazimbi has the highest 10-year return level for maximum temperature of 36.570^0C, while Polokwane has the least maximum temperature 10-year return level of 30.476^0C. This suggests that, in Thabazimbi, a maximum temperature of 36.570^0C is expected to be exceeded at least once in 10 years. Messina has the highest 10-year return level for minimum temperature of 22.854^0C, while Polokwane has the least minimum temperature 10-year return level of 18.878^0C. This suggests that, in Messina, a minimum temperature of 22.854^0C is expected to be exceeded at least once in 10 years.

Thabazimbi has the highest 100-year return level for maximum temperature of 39.569^0C, while Polokwane has the least maximum temperature 100-year return level of 32.403^0C. This suggests that, in Thabazimbi, a maximum temperature of 39.569^0C is expected to be exceeded at least once in 100 years. Thabazimbi has the highest 100-year return level for minimum temperature of 24.074^0C, while Mara has the least minimum temperature 100-year return level of 19.778^0C. This suggests that, in Thabazimbi, a minimum temperature of 24.074^0C is expected to be exceeded at least once in 100 years. The results in Table 4.16 also revealed that some stations like Thabazimbi may experience maximum temperature of over 41^0C at least once in 500 years.

For Messina and Thabazimbi maximum temperatures, the extremely high 10-year and 100-year return levels indicate impending heat waves in the Limpopo province. These findings are in agreement with those of Kruger and Shongwe (2004) and Phophi et al. (2020) who also found increasing temperatures in the Limpopo province.

Table 4.16: GEV model return periods (years) and their corresponding return levels (^oC).

Station name		10 years	20 years	50 years	100 years	500 years
Mara	Max	33.154	33.575	33.988	34.222	34.592
	Min	19.112	19.355	19.616	19.778	20.067
Messina	Max	36.533	36.965	37.379	37.608	37.957
	Min	22.854	23.178	23.498	23.681	23.972
Polokwane	Max	30.476	31.107	31.869	32.403	33.527
	Min	18.878	19.625	20.619	21.383	23.217
Thabazimbi	Max	36.570	37.495	38.684	39.569	41.596
	Min	22.264	22.847	23.563	24.074	25.174

4.8 Fitting non-stationary GEV distribution

The models that are considered in this section are:

$x \backsim$ GEVD$(\mu(t), \sigma(t), \xi(t))$

Model M_0 taken as the reference model such that

$M_0 : \mu(t) = \mu, \sigma(t) = \sigma, \xi(t) = \xi$;

Model M_1 has a linear trend in the location parameter such that

M_1: $\mu(t) = \mu + \mu_1 t, \sigma(t) = \sigma, \xi(t) = \xi$;

Model M_2 has a linear trend in the location and scale parameters such that

$M_2 : \mu(t) = \mu + \mu_1 t, \log \sigma(t) = \sigma_0 + \sigma_1 t, \xi(t) = \xi$;

Model M_3 has a quadratic trend in the location parameter such that

$M_3 : \mu(t) = \mu + \mu_1 t + \mu_2 t^2, \sigma(t) = \sigma, \xi(t) = \xi$;

Model M_4 has a quadratic trend in the location parameter and a linear trend in the scale parameter such that

$M_4 : \mu(t) = \mu + \mu_1 t + \mu_2 t^2, \log \sigma(t) = \sigma_0 + \sigma_1 t, \xi(t) = \xi$;

Model M_5 has a quadratic trend in both the location and scale parameters such that

$M_5 : \mu(t) = \mu + \mu_1 t + \mu_2 t^2$, $\log \sigma(t) = \sigma_0 + \sigma_1 t + \sigma_2 t^2$, $\xi(t) = \xi$;

Model M_6 contain SOI (Southern Oscillation Index) covariate term in the location parameter of the GEV distribution such that

$M_6 : \mu(t) = \mu_0 + \mu_1 SOI(t), \sigma(t) = \sigma, \xi(t) = \xi$;

Model M_7 includes both a trend and SOI term in the location parameter of the GEV distribution such that

$M_7 : \mu(t) = \mu_0 + \mu_1 SOI(t) + \mu_2 t, \sigma(t) = \sigma, \xi(t) = \xi$.

The likelihood ratio test p-values for each model parameter are calculated in R software using *"tb1=abs((parameter estimate)/standard error); tb1"*, *"pt(tb1,nrow,lower.tail=FALSE)"*.

4.8.1 Non-stationary GEV models with a trend

Mara models

Mara maximum temperature data

Consider the pair of models (M_0, M_1) from Table 4.17, where M_0 is taken as the reference model: $\chi^2_{1,0.05} = 3.841$, $D = 2[-115.376 - (-121.302)] = 11.852$. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001 for model M_1, which indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). Since D is greater than the critical value ($D > \chi^2_{1,0.05}$), we conclude that the model M_1 provides significant improvement in fit over the stationary GEV model, that is, model M_1 is worthwhile. The other pair is (M_0, M_2) with $\chi^2_{2,0.05} = 5.991$ and $D = 5.578$. The likelihood ratio test for $\mu_1 = 0$ has p-value $= 0.091$ and $\sigma_1 = 0$ has p-value $= 0.0440$, which indicates that the linear trend in the location parameter is not significant at 5% significance level (p-value > 0.05), whereas the linear trend in the scale parameter is significant (p-value< 0.05) in the model. The D statistic is less than the critical value ($D < \chi^2_{2,0.05}$), which shows that model M_2 does not provide any improvement in fit over the stationary GEV model, that is, model M_2 is not worthwhile.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 11.900 and 11.592 for M_3 and M_4, respectively. For the model M_3, the likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ both have p-value < 0.001. This indicates that the linear and quadratic trends in the location parameter are significant at 5% significance level (p-value < 0.05). For the model M_4, the likelihood ratio test for $\mu_1 = 0$ has p-value$= 0.484$; for $\mu_2 = 0$, it has p-value < 0.001; and for $\sigma_1 = 0$, it has p-value$= 0.169$. The results show that the linear trend in the location and scale parameters is not significant (p-value > 0.05), while the quadratic trend in the location parameter is significant at 5% significance level. Since D statistic for

each model is greater than the corresponding critical value, we conclude that there is significant improvement in fit over the stationary GEV model, that is, model M_3 and M_4 are worthwhile.

The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = 5.124$. The likelihood ratio test for $\mu_1 = 0, \mu_2 = 0, \sigma_1 = 0$ and $\sigma_2 = 0$ has p-value < 0.001, which shows that the quadratic trend in both the location and scale parameters is highly significant (p-value < 0.05). However, the overall model is not worthwhile since the D statistic is less than the critical value, i.e. $D < \chi^2_{4,0.05}$.

The best two competing non-stationary models for Mara maximum temperature data based on their deviance statistics and likelihood ratio test results are respectively: model (4.1) as the main model with a linear and quadratic trend in the location parameter and model (4.2) as the alternative model with a linear trend in the location parameter. The main model, M_3, is given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.323(x_i - (29.962 + 0.032t_i - 0.000t_i^2))}{1.279}\right)^{\frac{1}{0.323}}\right\}, \tag{4.1}$$

where x_i is the average monthly maximum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_1, is

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.327(x_i - (30.045 + 0.026t_i))}{1.282}\right)^{\frac{1}{0.327}}\right\}, \tag{4.2}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Table 4.17: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Mara maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	30.929			1.412			-0.338	121.3019
M_1	30.045	0.026		1.282			-0.327	115.376
M_2	30.841	0.011		1.320	0.001		-0.414	118.513
M_3	29.962	0.032	-0.000	1.279			-0.323	115.353
M_4	30.382	-0.000	0.000	1.450	-0.004		-0.353	115.506
M_5	31.361	-0.050	0.001	2.000	-0.040	0.000	-0.241	118.740

Key: NLLH denotes Negative Log-likelihood

Mara minimum temperature data

The model pair (M_0, M_1) from Table 4.18 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 5.462. The likelihood ratio test for $\mu_1 = 0$ has p-value=0.008 for model M_1, which indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). Since D is greater than the critical value ($D > \chi^2_{1,0.05}$), we conclude that the model M_1 provides significant improvement in fit over the stationary GEV model, that is, model M_1 is worthwhile. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = -4.872$. In model M_2, the likelihood ratio test for $\mu_1 = 0$ has p-value$= 0.473$ and for $\sigma_1 = 0$, it has p-value$= 0.065$. This indicates that the linear trend in both the location and scale parameters are not significant at 5% significance level (p-value > 0.05). The D statistic is less than the critical value ($D < \chi^2_{2,0.05}$), which shows that the model M_2 does not provide any improvement in fit over the stationary GEV model, that is, model M_2 is not worthwhile.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 5.92 and 6.578 for M_3 and M_4, respectively. For model M_3, the likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001. This indicates that the linear trend in the location parameter is significant at 5% significance level. For the model M_4, the likelihood ratio test for $\mu_1 = 0$ has

p-value < 0.001 and for $\sigma_1 = 0$, it has p-value$= 0.159$, which shows that the linear trend in the location parameter is significant, while for the scale parameter it is not significant at 5% significance level. Since D statistic is less than the corresponding critical value for both M_3 and M_4, we conclude that there is no significant improvement in fit over the stationary GEV model.

The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = -4.634$. The likelihood ratio test for $\mu_1 = 0$ has p value=0.473, $\mu_2 = 0$ and $\sigma_2 = 0$ have p-value < 0.001 and $\sigma_1 = 0$ has p-value=0.016. This indicates that the linear trend in the location parameter is not significant at 5% significance level (p-value > 0.05), while the linear trend in the scale parameter, and quadratic trend in both the location and scale parameters are significant at 5% significance level (p-value < 0.05). Since D statistic is less than the critical value, i.e. $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in fit over the stationary GEV model, that is, model M_5 is not worthwhile.

Table 4.18: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Mara minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	18.035			0.611			-0.228	66.719
M_1	17.725	0.009		0.871			-0.177	63.988
M_2	18.097	0.000		0.801	-0.006		0.011	69.155
M_3	17.620	0.018	-0.000	0.570			-0.000	63.759
M_4	17.640	0.016	-0.000	0.640	-0.002		-0.188	63.430
M_5	17.988	-0.000	0.000	0.809	-0.007	0.000	0.100	69.036

Key: NLLH denotes Negative Log-likelihood

The best non-stationary model, M_1, for Mara minimum temperature data has a linear trend in the location parameter given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.177(x_i - (17.725 + 0.009t_i))}{0.571}\right)^{\frac{1}{0.177}}\right\}, \qquad (4.3)$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.1) and (4.2) and (4.3), that is, -0.323 (p-value = 0.001), -0.327 (p-value = 0.001) and -0.177 (p-value = 0.010), are significantly different from zero (p-value < 0.05), implying that the shape parameters for the three models are significantly negative. This suggests that both the maximum and minimum temperature data at Mara meteorological station can be modelled by Weibull family of distribution, which is short-tailed. The diagnostic plots for the non-stationary GEV models in (4.1) and (4.3) are presented in Figure 4.23 and Figure 4.24, respectively. The residual probability plots for both models suggest a good fit to the data.

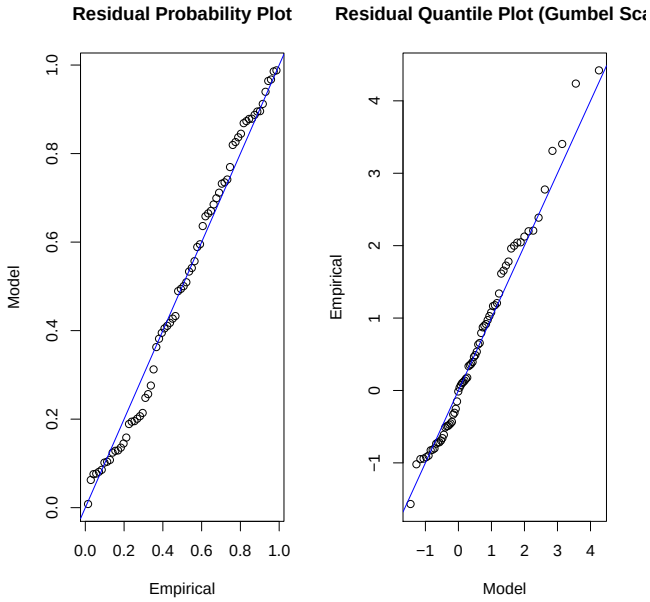

Figure 4.23: Diagnostic plots for the non-stationary GEV best fitting model (with a linear trend in the location parameter) at Mara maximum temperature.

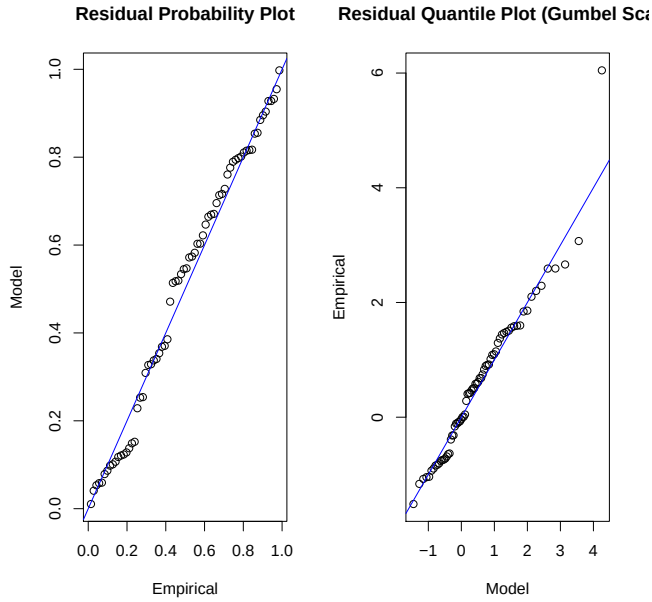

Figure 4.24: Diagnostic plots for the non-stationary GEV best fitting model (with a linear trend in the location parameter) at Mara minimum temperature.

Messina models

Messina maximum temperature data

The model pair (M_0, M_1) from Table 4.19 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 5.106. The likelihood ratio test for $\mu_1 = 0$, has p-value=0.012. This shows that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). The results showed that there is significant improvement in fit over the stationary GEV model for M_1 since D statistic is greater than the critical value, i.e. $D > \chi^2_{1,0.05}$. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 4.924$. The likelihood ratio test for $\mu_1 = 0$ has p-value=0.204, and for $\sigma_1 = 0$ it has p-value=0.053. This indicates that the linear trend in the location and scale parameters are not significant at 5% significance level (p-value > 0.05). Since D is less than the critical value $(D < \chi^2_{2,0.05})$, we

conclude that there is no significant improvement in fit over stationary GEV model, that is, model M_2 is not worthwhile.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 12.938 and 13.444 for M_3 and M_4, respectively. For the model M_3, the likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value < 0.001. This indicates that the linear and quadratic trends in the location parameter are significant at 5% significance level (p-value < 0.05). For the model M_4, the likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$, have p-value < 0.001; and for $\sigma_1 = 0$, it has p-value$= 0.225$. This indicates that the linear and quadratic trends in the location parameter are significant at 5% significance level (p-value < 0.05), while the linear trend in the scale parameter is not significant at 5% significance level (p-value > 0.05). Since the D statistic values are greater than the corresponding critical values for both model M_3 and M_4, we conclude that there is significant improvement in fit over the stationary GEV model.

The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = -14.266$. Since D statistic is less than the critical value, i.e. $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in fit over the stationary GEV model, that is, model M_5 is not worthwhile.

Table 4.19: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Messina maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	34.135			1.569			-0.369	136.473
M_1	33.617	0.015		1.540			-0.403	133.920
M_2	33.756	0.013		1.561	-0.000		-0.424	134.011
M_3	34.529	-0.055	0.001	1.465			-0.408	130.004
M_4	34.434	-0.054	0.001	1.613	-0.004		-0.399	129.751
M_5	33.887	-0.029	0.001	3.043	-0.062	0.000	0.172	143.606

Key: NLLH denotes Negative Log-likelihood

The best two competing non-stationary models for Messina maximum temperature data based on their deviance statistics and likelihood ratio test results are respectively: model (4.4) as the main model with a linear trend and quadratic trend in the location parameter, and model (4.5) as alternative model with a linear trend in the location parameter. The main model, M_3, is given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.408(x_i - (34.529 - 0.055t_i + 0.001t_i^2))}{1.465}\right)^{\frac{1}{0.408}}\right\}, \tag{4.4}$$

where x_i is the average monthly maximum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_1, is

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.403(x_i - (33.617 + 0.015t_i))}{1.540}\right)^{\frac{1}{0.403}}\right\}, \tag{4.5}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Messina minimum temperature data

The model pair (M_0, M_1) from Table 4.20 has $\chi^2_{1,0.05} = 3.841$ and D statistic value 10.074. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001. This indicates

that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05) for M_1. Since D statistic is greater than the critical value, i.e. $D > \chi^2_{1,0.05}$, we conclude that there is significant improvement in fit over the stationary GEV model, that is, model M_1 is worthwhile. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 10.554$. For the model M_2, the likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001 and for $\sigma_1 = 0$, it has p-value $= 0.172$. This indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05), while the linear trend in the scale parameter is not significant at 5% significance level (p-value > 0.05). Since D is greater than the critical value $(D > \chi^2_{2,0.05})$, we conclude that there is significant improvement in fit over the stationary GEV model, that is model M_2 is worthwhile.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 10.430 and 11.028 for M_3 and M_4, respectively. For the model M_3, the likelihood ratio test for $\mu_1 = 0$ has p-value $= 0.275$ and $\mu_2 = 0$, it has p-value < 0.001. This indicates that the linear trend in the location parameter is not significant at 5% significance level (p-value > 0.05), while the linear trend in the scale parameter is significant at 5% significance level (p-value < 0.05).

For the model M_4, the likelihood ratio test for $\mu_1 = 0$ has p-value $= 0.256$; for $\mu_2 = 0$, it has p-value < 0.001; and for $\sigma_1 = 0$, it has p-value $= 0.155$. This indicates that the linear trend in the location and scale parameters is not significant at 5% significance level (p-value > 0.05), while the quadratic trend in the location parameter is significant at 5% significance level (p-value < 0.05). Since D statistic is greater than the critical value, we conclude that there is significant improvement in fit over the stationary GEV model. The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = -6.714$. Since D statistic is less than the

critical value, i.e. $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in the model.

Table 4.20: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Messina minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	21.166			1.063			-0.330	108.048
M_1	20.531	0.017		0.990			-0.329	103.011
M_2	20.458	0.019		1.097	-0.003		-0.314	102.771
M_3	20.703	0.003	0.000	0.979			-0.313	102.833
M_4	20.756	-0.003	0.000	1.108	-0.004		-0.285	102.534
M_5	21.066	-0.032	0.000	1.272	-0.008	0.000	0.094	111.404

Key: NLLH denotes Negative Log-likelihood

The best non-stationary model, M_1, for Messina minimum temperature data has a linear trend in the location parameter given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.329(x_i - (20.531 + 0.017t_i))}{0.990}\right)^{\frac{1}{0.329}}\right\}, \qquad (4.6)$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.4) and (4.5) and (4.6), that is, -0.408 (p-value < 0.001), -0.403 (p-value < 0.001) and -0.329 (p-value < 0.001), are significantly different from zero (p-value < 0.05), implying that the shape parameters for the three models are significantly negative. This suggests that both the maximum and minimum temperature data at Messina meteorological station can be modelled by Weibull family of distribution which is short-tailed. The diagnostic plots for the non-stationary GEV models in (4.4), (4.5) and (4.6) are presented in Figure 4.25, Figure 4.26 and Figure 4.27, respectively. The residual probability plots for all the three models suggest a good fit to the data.

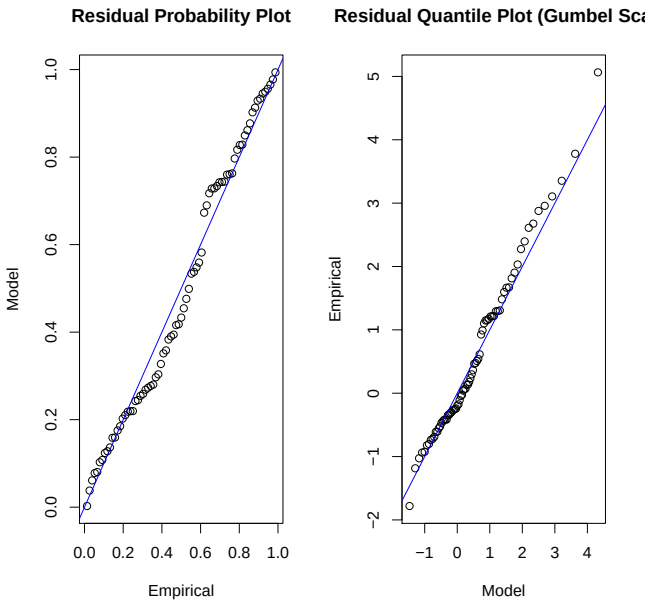

Figure 4.25: Diagnostic plots for the non-stationary GEV best fitting model (with a quadratic trend in the location parameter) at Messina maximum temperature.

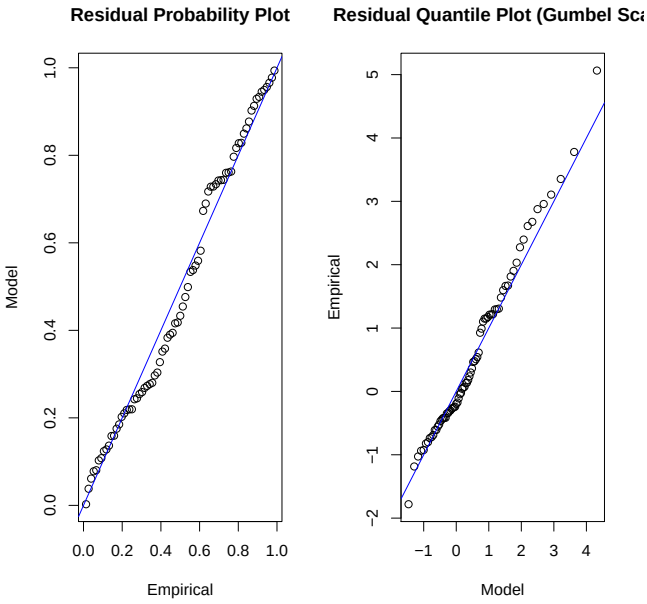

Figure 4.26: Diagnostic plots for the non-stationary GEV best fitting model (with a linear trend in the location parameter) at Messina maximum temperature.

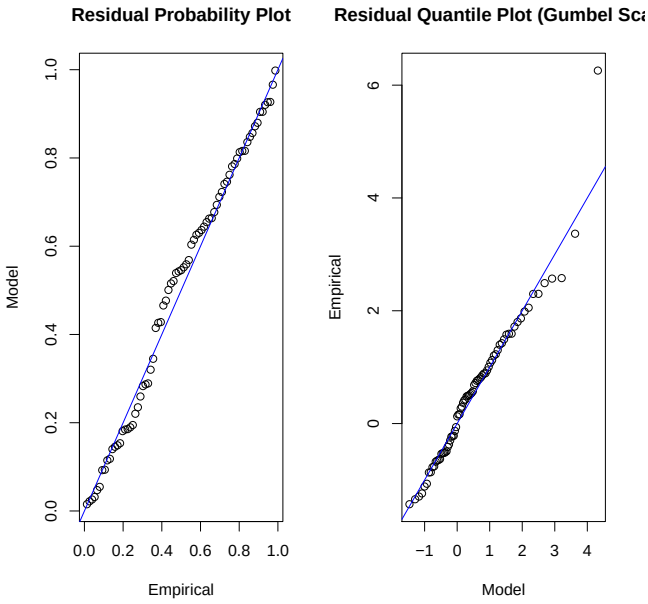

Figure 4.27: Diagnostic plots for the non-stationary GEV best fitting model (with a linear trend in the location parameter) at Messina minimum temperature.

Polokwane models

Polokwane maximum temperature data

The model pair (M_0, M_1) from Table 4.21 has $\chi^2_{1,0.05} = 3.841$ and D statistic value of 4.988. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.011. This indicates that for model M_1, the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). Since D statistic is greater than the critical value ($D < \chi^2_{1,0.05}$), we conclude that there is significant improvement in fit over the stationary GEV model, that is, model M_1 is worthwhile.

The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 3.360$. Since D is less than the critical value ($D < \chi^2_{2,0.05}$), we conclude that there is no significant improvement in adding the linear trend component in the location and scale parameters. The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 5.256 and 5.616 for M_3 and M_4, respectively. The results show that the D statistics for both model M_3 and M_4 are less than the corresponding critical values. Therefore, we conclude that there is no significant improvement in fit over the stationary GEV model, after adding the quadratic trend component in the location parameter and the linear trend component in the scale parameter. Also, since the model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = -0.048$, i.e. $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in the model.

Table 4.21: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Polokwane maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	28.246						-0.082	140.433
M_1	27.773	0.011		1.037			-0.048	137.939
M_2	28.175	0.004		1.232	-0.003		-0.116	138.753
M_3	27.927	0.001	0.000	1.041			-0.057	137.805
M_4	27.977	-0.000	0.000	1.139	-0.002		-0.081	137.6250
M_5	28.222	-0.000	0.000	1.485	-0.019	0.000	0.100	140.457

Key: NLLH denotes Negative Log-likelihood

The best non-stationary model, M_1, for Polokwane maximum temperature data has a linear trend in the location parameter given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.048(x_i - (27.773 + 0.011t_i))}{1.037}\right)^{\frac{1}{0.048}}\right\}, \qquad (4.7)$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Polokwane minimum temperature data

The model pair (M_0, M_1) from Table 4.22 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 25.850. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001 for M_1. This indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). Since D statistic is greater than the critical value $D > \chi^2_{1,0.05}$, we conclude that there is significant improvement in fit over the stationary GEV model, that is, model M_1 is worthwhile. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 26.288$. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001 and $\sigma_1 = 0$ has p-value = 0.224. This indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05), while the linear trend in the scale parameter is not significant

at 5% significance level (p-value > 0.05). Since D is greater than the critical value, i.e. $(D > \chi^2_{2,0.05})$, we conclude that there is significant improvement in fit over the stationary GEV model.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic value of 53.656 and 53.964 for M_3 and M_4, respectively. For model M_3, the likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value < 0.001. This indicates that the linear trend in the location parameter is significant at 5% significance level (p-value < 0.05). For model M_4, the likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value < 0.001, while for $\sigma_1 = 0$, it has p-value$= 0.258$. This indicates that the linear and quadratic trends in the location parameters are significant at 5% significance level (p-value < 0.05), while the linear trend in the scale parameter is not significant at 5% significance level (p-value > 0.05). Since D statistics for both model M_3 and M_4 are greater than the corresponding critical values, we conclude that there is significant improvement in fit over the stationary GEV model.

Table 4.22: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Polokwane minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	16.651			0.954			0.032	131.198
M_1	15.993	0.016		0.802			0.067	118.273
M_2	16.087	0.014		0.881	-0.002		0.055	118.054
M_3	15.023	0.080	-0.001	0.647			0.134	104.370
M_4	15.061	0.079	-0.001	0.695	-0.002		0.129	104.216
M_5	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Key: NLLH denotes Negative Log-likelihood

The best two competing non-stationary models for Polokwane minimum temperature data based on their deviance statistics and likelihood ratio test results are respectively: model (4.8) as the main model with linear and quadratic

trends in the location parameter and model (4.9) as the alternative model with a linear trend in the location parameter. The main model, M_3, is given by:

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{0.134(x_i-(15.023+0.080t_i-0.001t_i^2))}{0.647}\right)^{-\frac{1}{0.134}}\right\}, \tag{4.8}$$

where x_i is the average monthly minimum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_1, is given by:

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{0.067(x_i-(15.993+0.016t_i))}{0.802}\right)^{-\frac{1}{0.067}}\right\}, \tag{4.9}$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameter in model (4.7), that is, -0.048 (p-value = 0.261) is not significantly different from zero (p-value > 0.05); while the shape parameter in model (4.8) and model (4.9), that is, 0.134 (p-value = 0.008) and 0.067 (p-value = 0.049) respectively, are significantly different from zero, implying that the shape parameters for the two models are significantly negative. This suggests that the maximum temperature data at Polokwane station can be modelled by Gumbel family of distribution which is light-tailed, while minimum temperature data at Polokwane station can be modelled by Weibull family of distribution which is short-tailed. The diagnostic plots for the non-stationary GEV models in (4.7) and (4.8) are presented in Figure 4.28 and Figure 4.29, respectively. The residual probability plots for all models suggest a good fit to the data.

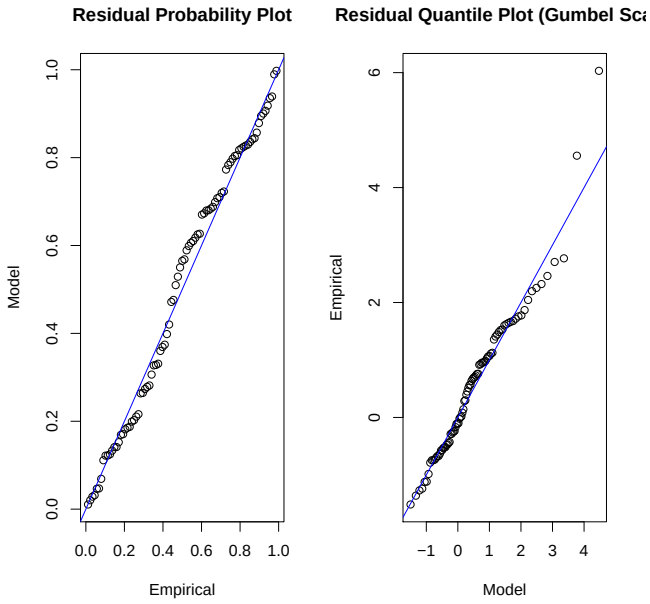

Figure 4.28: Diagnostic plots for the non-stationary GEV best fitting model (with a linear trend in the location parameter) at Polokwane maximum temperature.

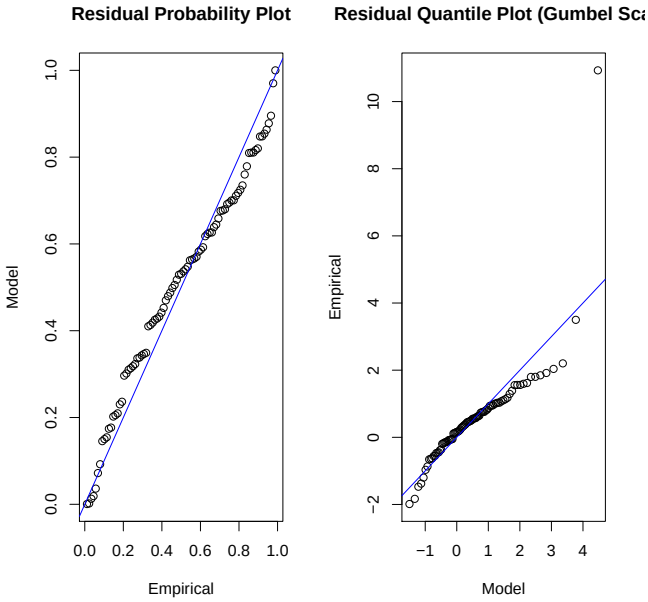

Figure 4.29: Diagnostic plots for the non-stationary GEV best fitting model (with linear and quadratic trends in the location parameter) at Polokwane minimum temperature.

Thabazimbi models

Thabazimbi maximum temperature data

The model pair (M_0, M_1) from Table 4.23 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 0.234. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.319. This indicates that, for model M_1, the linear trend in the location parameter is not significant at 5% significance level (p-value > 0.05). Since D statistic is less than the critical value, i.e. ($D < \chi^2_{1,0.05}$), we conclude that there is no significant improvement in fit over the stationary GEV model, that is, M_1 is not worthwhile. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 0.192$. For the model M_2, the likelihood ratio test for $\mu_1 = 0$ has p-value$= 0.395$; and for $\sigma_1 = 0$, it has p-value $= 0.457$. This indicates that the linear trend in the location and scale parameters is not significant at 5% significance level (p-value > 0.05). Since D

is less than the critical value, i.e. $(D < \chi^2_{2,0.05})$, we conclude that there is no significant improvement in fit over the stationary GEV model, that is, M_2 is not worthwhile.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic values of 0.238 and 0.238 for M_3 and M_4, respectively. For model M_3, the likelihood ratio test for $\mu_1 = 0$ has p-value $= 0.428$; and for $\mu_2 = 0$, it has p-value $= 0.474$. This indicates that the linear and quadratic trends in the location parameter are not significant at 5% significant level (p-value > 0.05). The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.435; for $\mu_2 = 0$, it has p-value = 0.475; and for $\sigma_1 = 0$, it has p-value = 0.499. This indicates that the linear and quadratic trends in the location and scale parameters are not significant at 5% significance level (p-value > 0.05). The results show that there is no significant improvement in fit over the stationary GEV model, that is, M_3 and M_4 are not worthwhile. The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = 0.022$. Since $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in fit over the stationary GEV model.

Table 4.23: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Thabazimbi maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	33.643			1.313			-0.008	46.002
M_1	33.423	0.017		1.311			-0.011	45.885
M_2	33.551	0.010		1.372	-0.004		-0.022	45.906
M_3	33.389	0.027	-0.000	1.311			-0.012	45.883
M_4	33.387	0.027	-0.000	1.309	0.000		-0.011	45.883
M_5	33.761	-0.012	0.001	1.108	0.071	-0.003	0.017	45.991

Key: NLLH denotes Negative Log-likelihood

For Thabazimbi maximum temperature data, the best model is stationary GEV model, M_0, and it is given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.008(x_i - 33.643)}{1.313}\right)^{\frac{1}{0.008}}\right\}, \qquad (4.10)$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Thabazimbi minimum temperature data

The model pair (M_0, M_1) from Table 4.24 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 0.384. Since D statistic is less than the critical value $(D < \chi^2_{1,0.05})$, we conclude that there is no significant improvement in fit over the stationary GEV model. The model pair (M_0, M_2) has $\chi^2_{2,0.05} = 5.991$ and $D = 4.684$. Since D is less than the critical value, i.e. $(D < \chi^2_{2,0.05})$, we conclude that there is no significant improvement in adding the linear trend component in the location and scale parameters.

The quadratic model pairs (M_0, M_3) and (M_0, M_4) have $\chi^2_{2,0.05} = 5.991$ and $\chi^2_{3,0.05} = 7.815$, respectively; with D statistic value of 2.730 and 7.710 for M_3 and M_4, respectively. Since D statistic values are less than the corresponding critical values for both M_3 and M_4, we conclude that there is no significant improvement in adding the quadratic trend component in the location and scale parameters, that is, M_3 and M_4 are not worthwhile. The model pair (M_0, M_5) has $\chi^2_{4,0.05} = 9.488$ and $D = 8.936$. Since $D < \chi^2_{4,0.05}$, we conclude that there is no significant improvement in fit over the stationary GEV model.

Table 4.24: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Thabazimbi minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma_0}$	$\hat{\sigma_1}$	$\hat{\sigma_2}$	$\hat{\xi}$	NLLH
M_0	20.266			0.951			-0.062	37.210
M_1	20.009	0.019		0.935			-0042	37.018
M_2	20.142	0.024		1.648	-0.050		-0.199	34.868
M_3	20.477	-0.116	0.006	0.854			0.027	35.845
M_4	19.993	-0.029	0.003	1.609	-0.052		-0.138	33.355
M_5	21.245	-0.0190	0.008	2.365	-0.173	0.004	-0.298	32.742

Key: NLLH denotes Negative Log-likelihood

For Thabazimbi minimum temperature data, the best model is stationary GEV model, M_0, and it is given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.062(x_i - 20.266)}{0.951}\right)^{\frac{1}{0.062}}\right\}, \qquad (4.11)$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.10) and (4.11), that is, -0.008 (p-value = 0.473), -0.062 (p-value = 0.302) are not significantly different from zero (p-value > 0.05). This suggests that the maximum and minimum temperature data at Thabazimbi meteorological station can be modelled by Gumbel family of distribution which is light-tailed.

4.8.2 Non-stationary GEV models with trend and SOI covariates

Mara model

Figure 4.30 presents the scatter plot for the Mara station between average monthly temperatures and SOI. From the scatter plot (Figure 4.30), there is no clear positive linear relationship between average monthly temperatures and

SOI. These findings are investigated further using results in Table 4.25 and Table 4.26.

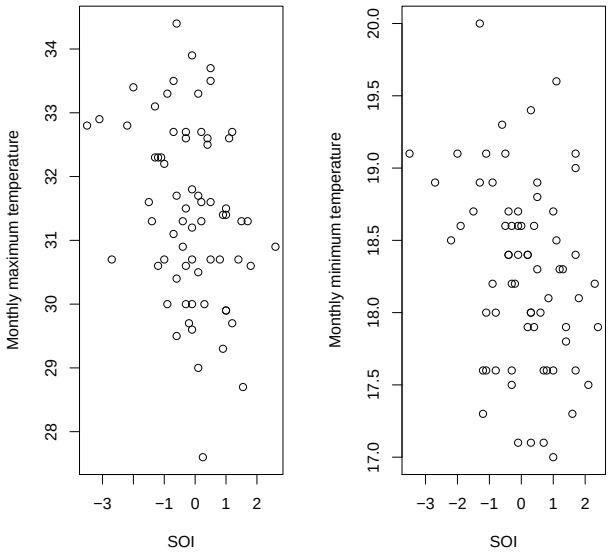

Figure 4.30: Scatter plot of average maximum temperatures and the Southern Oscillation Index (SOI) at Mara maximum (left panel) and minimum (right panel) temperatures from 1951 to 2018.

Mara maximum temperature data

The model pair (M_0, M_6) from Table 4.25 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 13.430. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.007. This indicates that the SOI term in location parameter is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{1,0.05}$). The overall model M_6 is significant and provides an improvement in fit over the stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 24.244$. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.005 and for $\mu_2 = 0$, it has p-value = 0.160. This implies that additional of trend term in model M_7 is not significant, although the SOI term

remains significant. Therefore, model M_7 is worthwhile ($D > \chi^2_{2,0.05}$), but not significant in the additional trend term.

Table 4.25: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Mara maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	30.929			1.412	-0.338	121.302
M_6	30.950	-0.371		1.348	-0.338	114.587
M_7	30.079	-0.351	0.025	1.206	-0.285	109.180

Key: NLLH denotes Negative Log-likelihood

The best non-stationary model, M_6, for Mara maximum temperature data has a SOI term in the location parameter given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.338(x_i - (30.950 - 0.371SOI(t_i)))}{1.348}\right)^{\frac{1}{0.338}}\right\}, \tag{4.12}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Mara minimum temperature data

The model pair (M_0, M_6) from Table 4.26 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 6.658. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.008. This implies that the SOI term in the location parameter is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{1,0.05}$). Thus, overall, model M_6 is significant and provides an improvement in fit over the stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 12.194$. The likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value < 0.001. This implies that the model M_7 is significant and worthwhile ($D > \chi^2_{2,0.05}$). Thus, the model M_7 is significant and provides an

improvement in fit over the stationary GEV model.

Table 4.26: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Mara minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	18.035			0.611	-0.228	66.719
M_6	18.065	-0.155		0.607	-0.249	63.390
M_7	17.756	-0.149	0.009	0.566	-0.205	60.622

Key: NLLH denotes Negative Log-likelihood

The best two competing models for Mara minimum temperature data with a SOI term in the location parameter based on their deviance statistics and likelihood ratio test results are respectively: model (4.13) as the main model with a linear trend and a SOI term in location parameter, and model (4.14) as the alternative model with a SOI term only in the location parameter. The main model, M_7, is given by:

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.205(x_i-(17.756-0.149SOI(t_i)+0.009t_i))}{0.566}\right)^{\frac{1}{0.205}}\right\}, \tag{4.13}$$

where x_i is the average monthly minimum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_6, is

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.249(x_i-(18.065-0.155SOI(t_i)))}{0.607}\right)^{\frac{1}{0.249}}\right\}, \tag{4.14}$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.12), (4.13) and (4.14), that is, -0.338 (p-value < 0.001), -0.205 (p-value=0.001) and -0.249 (p-value=0.005), are signifi-

cantly different from zero (p-value < 0.05), implying that the shape parameters for the three models are significantly negative. This suggests that the maximum and minimum temperature data at Mara meteorological station can be modelled by Weibull family of distribution which is short-tailed.

Messina model

Figure 4.31 presents the scatter plot for the Messina station between average monthly temperatures and SOI. From the scatter plot (Figure 4.31), there is no clear positive linear relationship between average monthly temperatures and SOI. These findings are investigated further using results in Table 4.27 and Table 4.28.

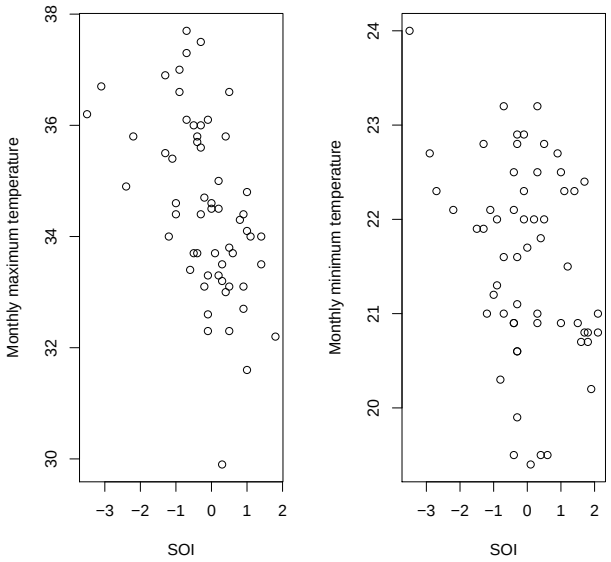

Figure 4.31: Scatter plot of average maximum temperatures and the Southern Oscillation Index (SOI) at Messina maximum (left panel) and minimum (right panel) temperatures from 1951 to 2009.

Messina maximum temperature data

The model pair (M_0, M_6) from Table 4.27 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 73.160. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001. This implies that the SOI term in the location parameter is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{1,0.05}$). The results show that model M_6 is significant and provides an improvement in fit over stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 85.312$. The likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value < 0.001. This implies that the model M_7 is significant and worthwhile ($D > \chi^2_{2,0.05}$).

Table 4.27: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Messina maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	34.135			1.569	-0.369	136.473
M_6	33.934	-0.832		1.398	-0.329	99.893
M_7	33.909	-0.659	0.036	1.123	-0.287	93.817

Key: NLLH denotes Negative Log-likelihood

The best two competing models for Messina maximum temperature data with a SOI term in the location parameter based on their deviance statistics are respectively: model (4.15) as the main model with a linear trend and a SOI term in location parameter, and model (4.16) as the alternative model with a SOI term only in the location parameter. The main model, M_7, is given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.287(x_i - (33.909 - 0.659SOI(t_i) + 0.036t_i))}{1.123}\right)^{\frac{1}{0.287}}\right\}, \tag{4.15}$$

where x_i is the average monthly maximum temperature and t_i is the time in

years. The alternative non-stationary GEV model, M_6, is

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.329(x_i-(33.934-0.832SOI(t_i)))}{1.398}\right)^{\frac{1}{0.329}}\right\}, \tag{4.16}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Messina minimum temperature data

The model pair (M_0, M_6) from Table 4.28 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 58.570. The likelihood ratio test for $\mu_1 = 0$ has p-value < 0.001. This implies that the SOI term in the location parameter is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{1,0.05}$). Thus, the model M_6 is significant and provides an improvement in fit over stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 68.342$. The likelihood ratio tests for $\mu_1 = 0$ and $\mu_2 = 0$ have p-value = 0.002. This implies that the model M_7 is significant and worthwhile ($D > \chi^2_{2,0.05}$). Thus, the model M_7 is significant and provides an improvement in fit over the stationary GEV model.

Table 4.28: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Messina minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	21.166			1.063	-0.330	108.048
M_6	21.315	-0.284		1.052	-0.487	78.763
M_7	20.549	-0.277	0.025	0.886	-0.321	73.877

Key: NLLH denotes Negative Log-likelihood

The best two competing models for Messina minimum temperature data with a SOI term in the location parameter based on their deviance statistics are respectively: model (4.17) as the main model with a linear trend and a SOI term in location parameter, and model (4.18) as the alternative model with a

SOI term only in the location parameter. The main model, M_7, is given by:

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.321(x_i-(20.549-0.277SOI(t_i)+0.025t_i))}{0.886}\right)^{\frac{1}{0.321}}\right\}, \tag{4.17}$$

where x_i is the average monthly minimum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_6, is

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.487(x_i-(21.315-0.284SOI(t_i)))}{1.052}\right)^{\frac{1}{0.487}}\right\}, \tag{4.18}$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.15), (4.16), (4.17) and (4.18), that is, -0.287, -0.329, -0.321 and -0.487, are significantly different from zero (p-value < 0.05), implying that the shape parameters for the four models are significantly negative. This suggests that the maximum and minimum temperature data at Messina meteorological station can be modelled by Weibull family of distribution which is short-tailed.

Polokwane model

Figure 4.32 presents the scatter plot for the Polokwane station between average monthly temperatures and SOI. From the scatter plot (Figure 4.32), there is no clear positive linear relationship between average monthly temperatures and SOI. These findings are investigated further using results in Table 4.29 and Table 4.30.

Polokwane maximum temperature data

The model pair (M_0, M_6) from Table 4.29 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 80.866. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.003.

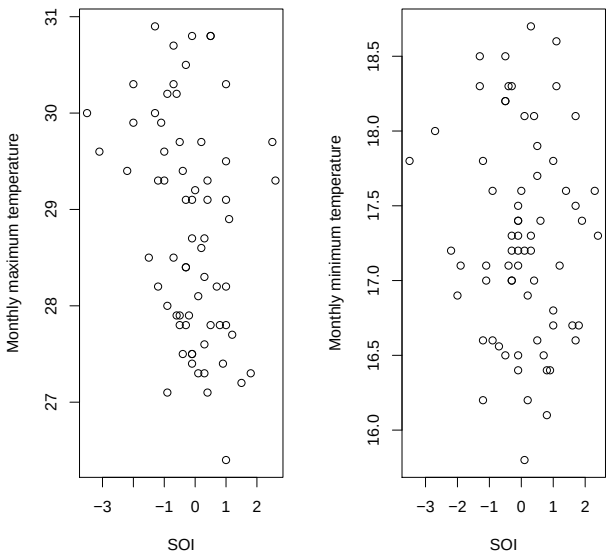

Figure 4.32: Scatter plot of average maximum temperatures and the Southern Oscillation Index (SOI) at Polokwane maximum (left panel) and minimum (right panel) temperatures from 1951 to 2018.

This implies that the SOI term in the location parameter is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{1,0.05}$). Therefore, model M_6 provides an improvement in fit over the stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 86.036$. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.005 and for $\mu_2 = 0$, it has p-value = 0.007. This implies that model M_7 is significant (p-value < 0.05) and worthwhile ($D > \chi^2_{2,0.05}$). Thus, overall model M_7 provides an improvement in fit over the stationary GEV model.

Table 4.29: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Polokwane maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	28.246			1.086	-0.082	140.433
M_6	28.298	-0.376		0.982	-0.163	100
M_7	27.822	-0.309	0.016	0.985	-0.240	97.415

Key: NLLH denotes Negative Log-likelihood

The best two competing models for Polokwane maximum temperature data with a SOI term in the location parameter based on their deviance statistics

are respectively: model (4.19) as the main model with a linear trend and a SOI term in location parameter, and model (4.20) as the alternative model with a SOI term only in the location parameter. The main model, M_7, is given by:

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.240(x_i-(27.822-0.309SOI(t_i)+0.016t_i))}{0.985}\right)^{\frac{1}{0.240}}\right\}, \tag{4.19}$$

where x_i is the average monthly maximum temperature and t_i is the time in years. The alternative non-stationary GEV model, M_6, is

$$G(\mu,\sigma,\xi,x_i,t)=\exp\left\{-\left(1+\frac{-0.163(x_i-(28.298-0.376SOI(t_i)))}{0.982}\right)^{\frac{1}{0.163}}\right\}, \tag{4.20}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Polokwane minimum temperature data

The model pair (M_0, M_6) from Table 4.30 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 123.954. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.247. This implies that the SOI term in the location parameter is not significant (p-value > 0.05). Therefore, model M_6 provides an improvement in fit over the stationary GEV model, although the SOI term is not significant. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 127.034$. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.208 and for $\mu_2 = 0$, it has p-value = 0.034. This implies that the additional trend term in model M_7 is significant, although the SOI term remains not significant. Therefore, model M_7 is worthwhile ($D > \chi^2_{2,0.05}$), but not significant in the SOI trend term.

Table 4.30: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Polokwane minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	16.651			0.954	0.032	131.198
M_6	17.045	-0.050		0.658	-0.264	69.221
M_7	16.794	-0.058	0.008	0.660	-0.303	67.681

Key: NLLH denotes Negative Log-likelihood

The best non-stationary model, M_7, for Polokwane minimum temperature data has a linear trend and a SOI term in the location parameter given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.303(x_i - (16.794 - 0.058SOI(t_i) + 0.008t_i))}{0.660}\right)^{\frac{1}{0.303}}\right\}, \tag{4.21}$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.19) and (4.20), that is, -0.240, -0.163, are not significantly different from zero (p-value > 0.05), implying that the shape parameters for the two models are not significantly negative. The shape parameter in model (4.21), that is, -0.303 is significantly different from zero. This suggests that the maximum temperature data at Polokwane station can be modelled by Gumbel family of distribution which is light-tailed, while minimum temperature data at Polokwane station can be modelled by Weibull family of distribution which is short-tailed.

Thabazimbi model

Figure 4.33 presents a scatter plot for Thabazimbi station between average monthly temperatures and SOI. Results in Figure 4.33 show that there is a probabilistic linear relationship between the average monthly temperatures and SOI with considerable randomness. This relationship is further investigated using Table 4.31 and Table 4.32.

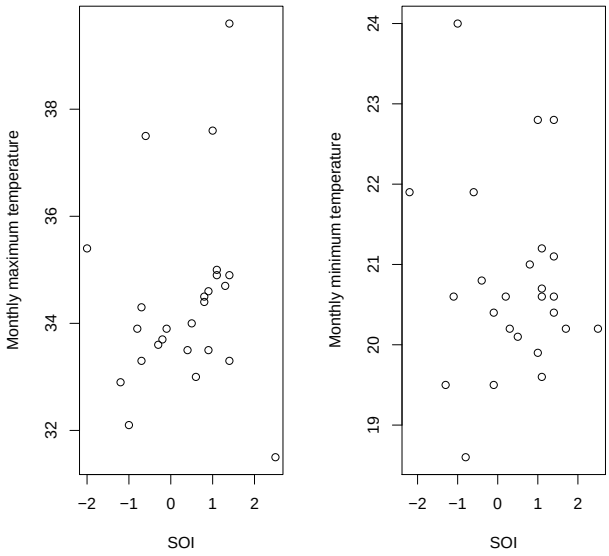

Figure 4.33: Scatter plot of average maximum temperatures and the Southern Oscillation Index (SOI) at Thabazimbi maximum (left panel) and minimum (right panel) temperatures from 1994 to 2018.

Thabazimbi maximum temperature data

The model pair (M_0, M_6) from Table 4.31 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 0.370. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.264. This implies that the SOI term in the location parameter is not significant (p-value > 0.05) and not worthwhile ($D < \chi^2_{1,0.05}$). Therefore, model M_6 is not significant and does not improve the stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 1.982$. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.068 and for $\mu_2 = 0$, it has has p-value = 0.096 . This implies that the SOI and the additional trend term in the location parameter are not significant (p-value> 0.05) and model M_7 is not worthwhile ($D < \chi^2_{2,0.05}$).

Table 4.31: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Thabazimbi maximum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	33.364			1.313	-0.008	46.002
M_6	33.664	-0.130		1.277	-0.023	45.817
M_7	33.892	-0.383	0.060	1.163	0.125	45.011

Key: NLLH denotes Negative Log-likelihood

The best model for Thabazimbi maximum temperature data is the stationary GEV model, M_0, given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.008(x_i - 33.643)}{1.313}\right)^{\frac{1}{0.008}}\right\}, \tag{4.22}$$

where x_i is the average monthly maximum temperature and t_i is the time in years.

Thabazimbi minimum temperature data

The model pair (M_0, M_6) from Table 4.32 has $\chi^2_{1,0.05} = 3.841$ and a D statistic value of 0.506. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.134. This implies that the SOI term in the location parameter is not significant (p-value > 0.05). Since D is less than the critical value, i.e. ($D < \chi^2_{1,0.05}$), we conclude that there is no significant improvement in fit over the stationary GEV model. The model pair (M_0, M_7) has $\chi^2_{2,0.05} = 5.991$ and $D = 1.144$. The likelihood ratio test for $\mu_1 = 0$ has p-value = 0.110 and for $\mu_2 = 0$, it has p-value = 0.217. This implies that the SOI and trend term in the location parameter are not significant (p-value > 0.05). Since D is less than the critical value, i.e. $(D < \chi^2_{2,0.05})$, we conclude that there is no significant improvement in fit over the stationary GEV model, that is, model M_7 is not worthwhile.

Table 4.32: Parameter and maximum likelihood estimates of the non-stationary GEV distribution for Thabazimbi minimum temperature.

Model	$\hat{\mu_0}$	$\hat{\mu_1}$	$\hat{\mu_2}$	$\hat{\sigma}$	$\hat{\xi}$	NLLH
M_0	20.266			0.951	-0.062	37.210
M_6	20.111	0.236		0.875	0.056	36.957
M_7	19.816	0.236	0.023	0.860	0.065	36.638

Key: NLLH denotes Negative Log-likelihood

The best model for Thabazimbi minimum temperature data is the stationary GEV model, M_0, given by:

$$G(\mu, \sigma, \xi, x_i, t) = \exp\left\{-\left(1 + \frac{-0.062(x_i - 20.266)}{0.951}\right)^{\frac{1}{0.062}}\right\}, \tag{4.23}$$

where x_i is the average monthly minimum temperature and t_i is the time in years.

The shape parameters in models (4.22) and (4.23), that is, -0.008 (p-value = 0.473), -0.062 (p-value = 0.302), are not significantly different from zero (p-value > 0.05). This suggests that the maximum and minimum temperature data at Thabazimbi meteorological station can be modelled by Gumbel family of distribution which is light-tailed.

4.9 Peaks-over-threshold (POT) approaches

This section presents results of the POT approaches which include threshold selection, declustering, generalised Pareto (GP) and Poisson point process distributions.

4.9.1 Threshold selection

Thresholds were chosen based on mean residual life and parameter stability plots. Figure 4.34 to Figure 4.37 present the mean residual life plots and Fig-

ure 4.38 to Figure 4.45 present parameter stability plots for average monthly maximum and minimum temperature data. Table 4.33 presents the selected threshold values for each meteorological station obtained from Figure 4.34 to Figure 4.45 for both maximum and minimum temperature data. The thresholds were chosen in order to meet the requirements of the bias-variance threshold trade-off balance such that they were high enough and low enough to have sufficient data to estimate the parameters.

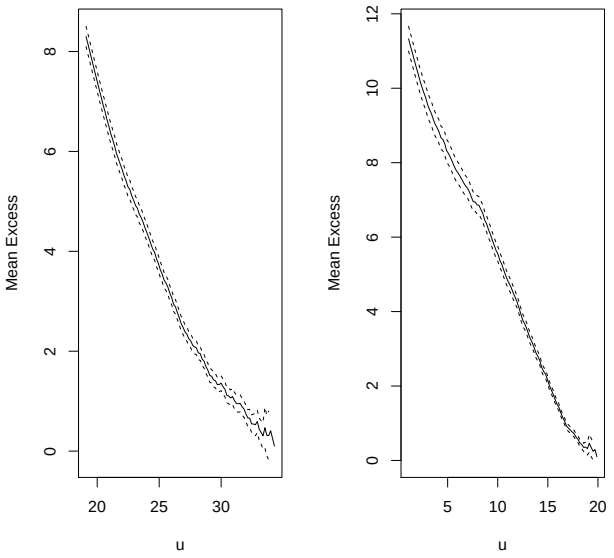

Figure 4.34: Mean residual life plots for Mara maximum (left panel) and minimum (right panel) temperature.

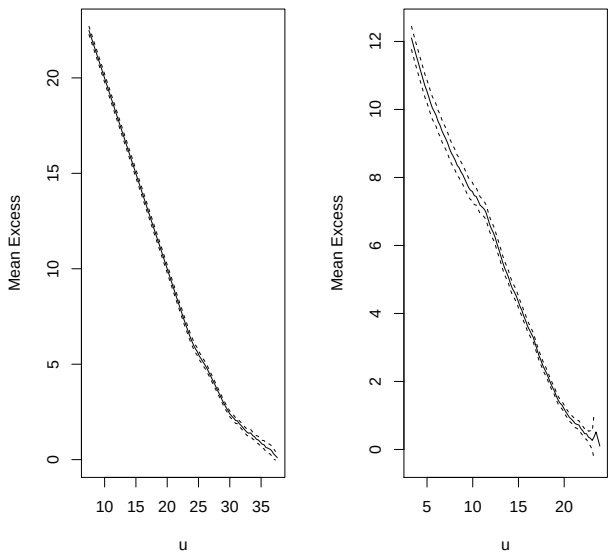

Figure 4.35: Mean residual life plots for Messina maximum (left panel) and minimum (right panel) temperature.

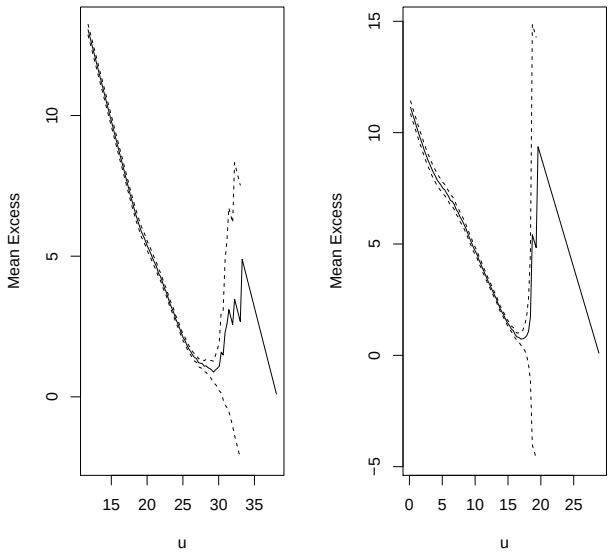

Figure 4.36: Mean residual life plots for Polokwane maximum (left panel) and minimum (right panel) temperature.

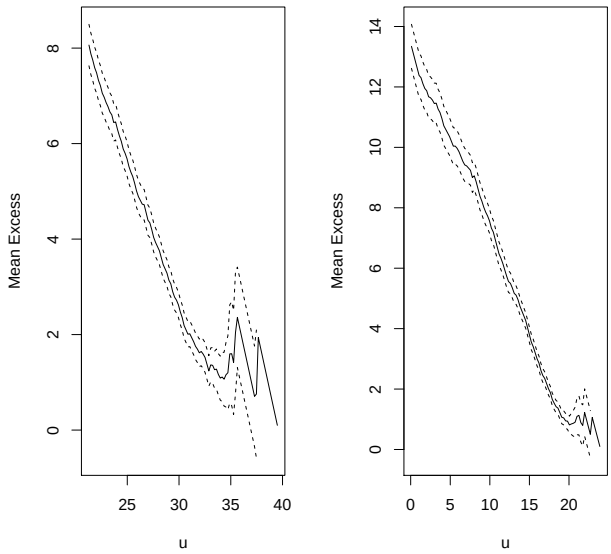

Figure 4.37: Mean residual life plots for Thabazimbi maximum (left panel) and minimum (right panel) temperature.

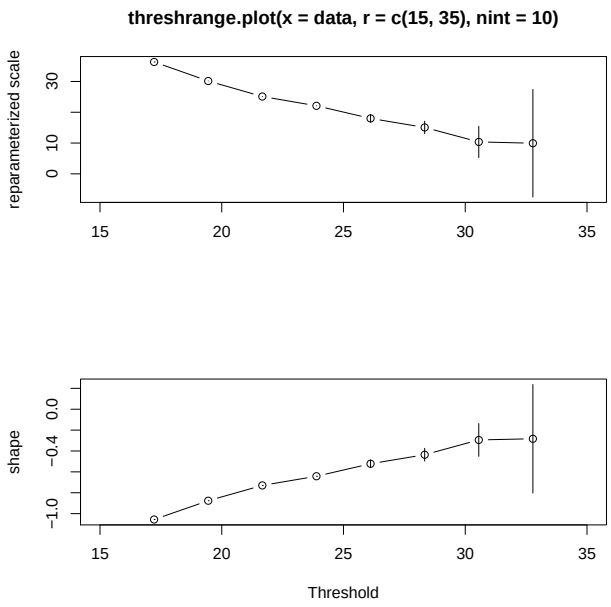

Figure 4.38: Parameter stability plots for Mara maximum temperature.

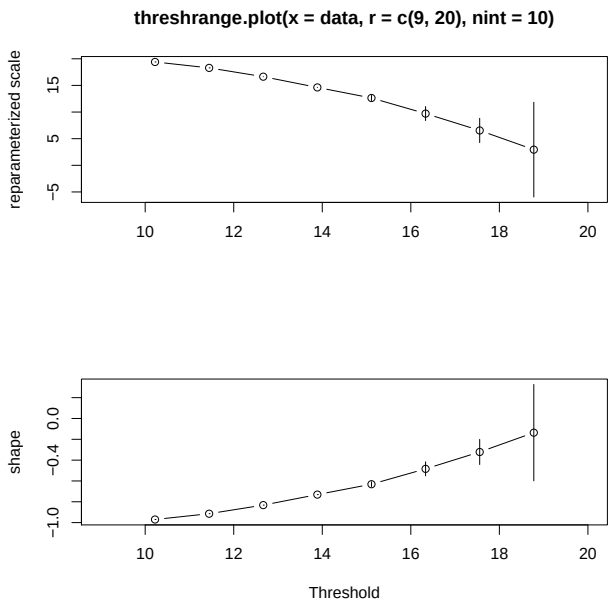

Figure 4.39: Parameter stability plots for Mara minimum temperature.

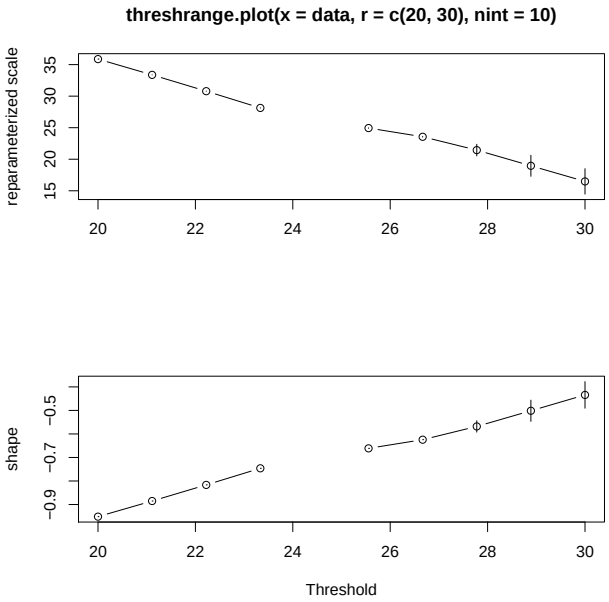

Figure 4.40: Parameter stability plots for Messina maximum temperature.

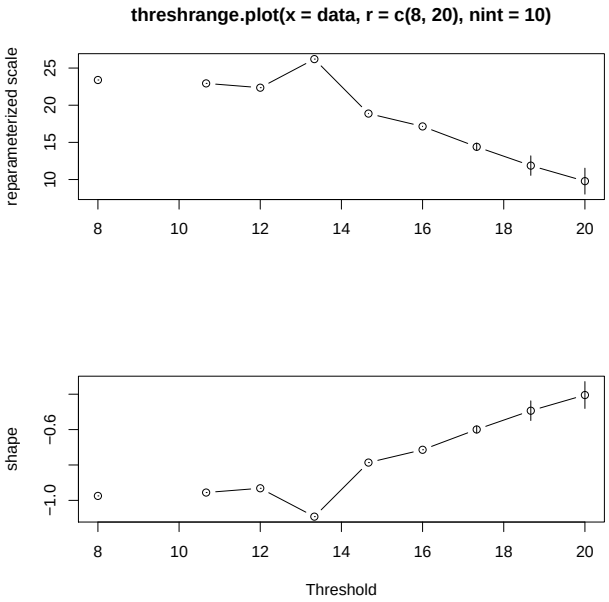

Figure 4.41: Parameter stability plots for Messina minimum temperature.

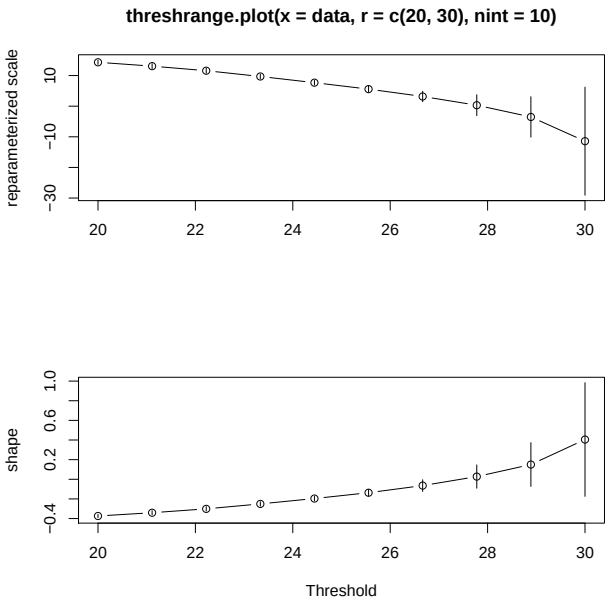

Figure 4.42: Parameter stability plots for Polokwane maximum temperature.

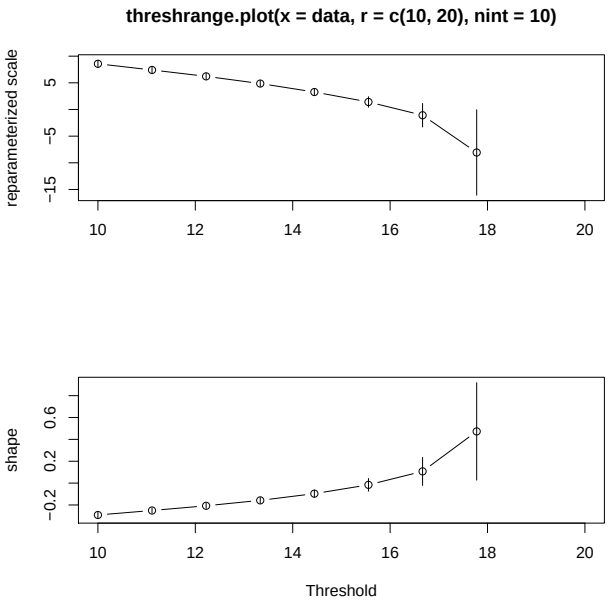

Figure 4.43: Parameter stability plots for Polokwane minimum temperature.

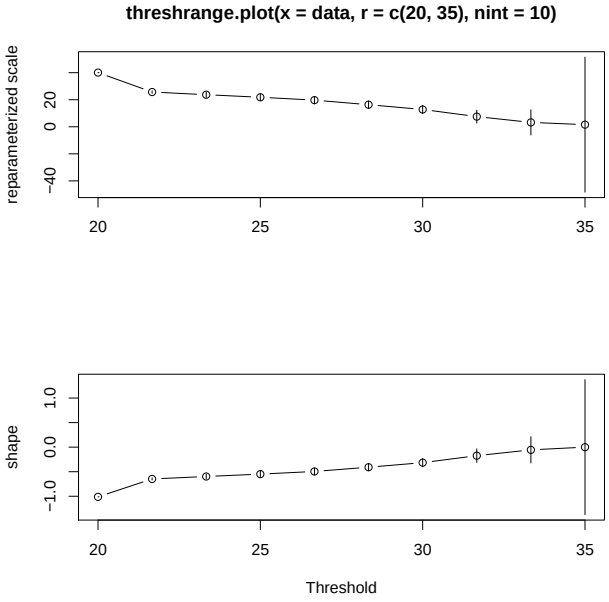

Figure 4.44: Parameter stability plots for Thabazimbi maximum temperature.

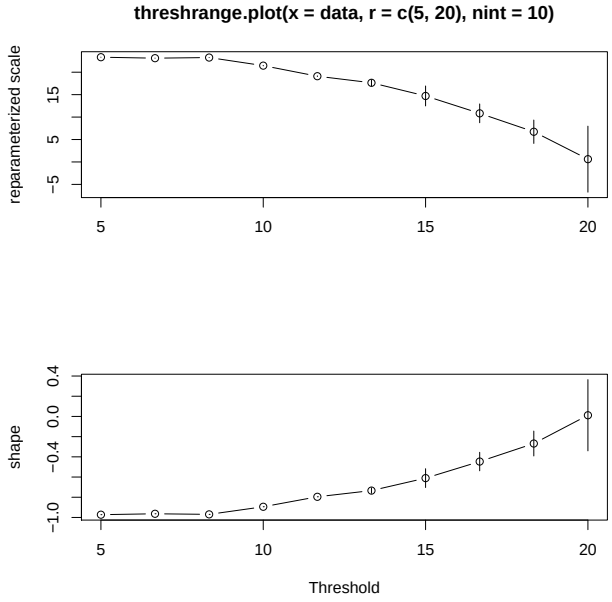

Figure 4.45: Parameter stability plots for Thabazimbi minimum temperature.

Table 4.33: Threshold values for the average maximum and minimum temperatures in all meteorological stations.

Station name	Threshold	
Mara	Max	27
	Min	11
Messina	Max	27
	Min	13.7
Polokwane	Max	22.3
	Min	11
Thabazimbi	Max	30
	Min	8

4.9.2 Declustering

Since the exceedances above the threshold could not be assumed to be independent from each other, declustering of the cluster maxima and minima was performed and the results are presented in Figure 4.46 to Figure 4.53 for all meteorological stations cluster maxima and minima. The declustered exceedances are assumed independent from each other.

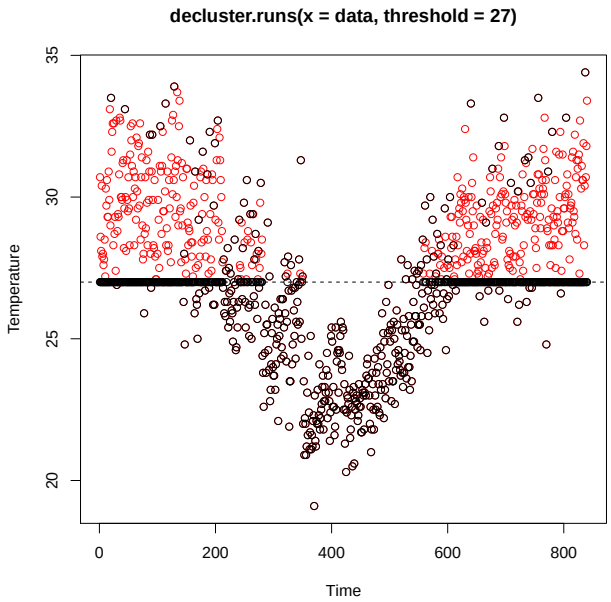

Figure 4.46: Mara declustered maximum temperature showing cluster maxima above 27 ^oC threshold.

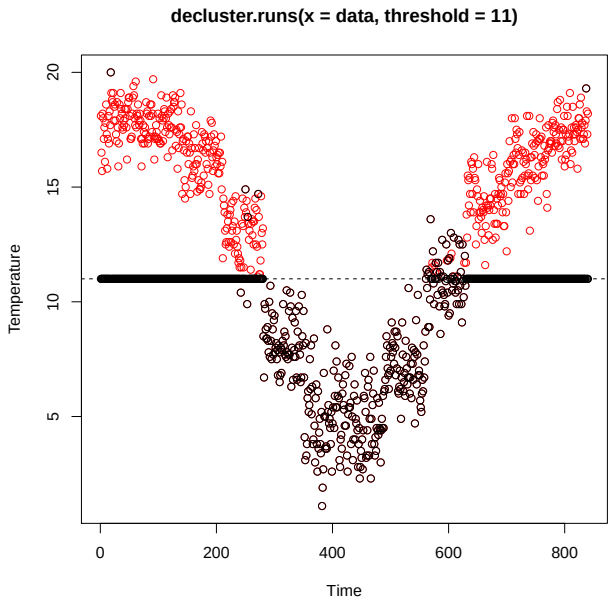

Figure 4.47: Mara declustered minimum temperature showing cluster minima above 11 ^oC threshold.

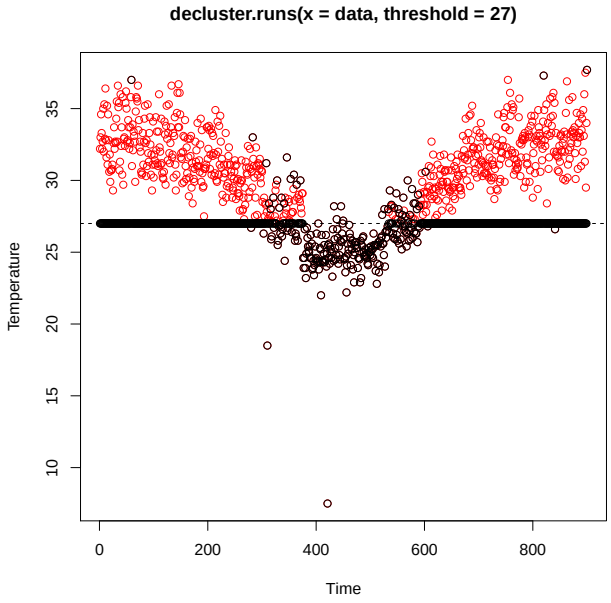

Figure 4.48: Messina declustered maximum temperature showing cluster maxima above 27 ^oC threshold.

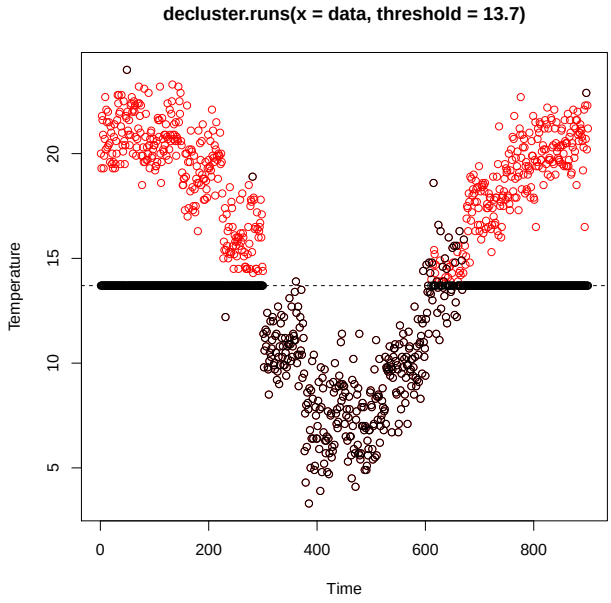

Figure 4.49: Messina declustered minimum temperature showing cluster maxima above 13.7 ^oC threshold.

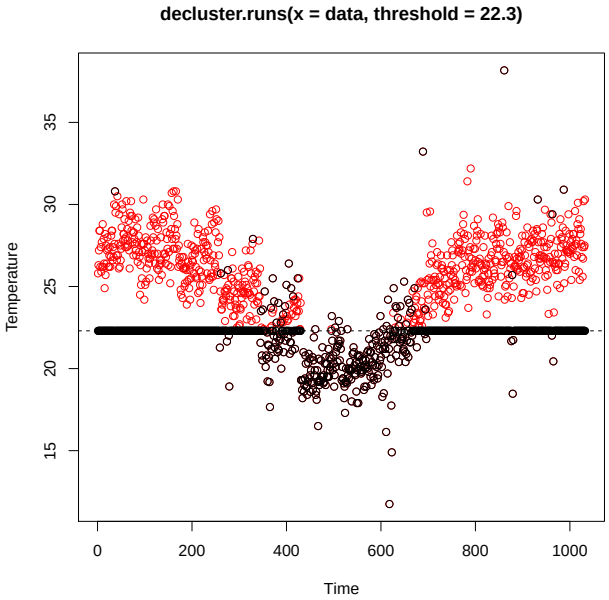

Figure 4.50: Polokwane declustered maximum temperature showing cluster minima above 22.3 ^oC threshold.

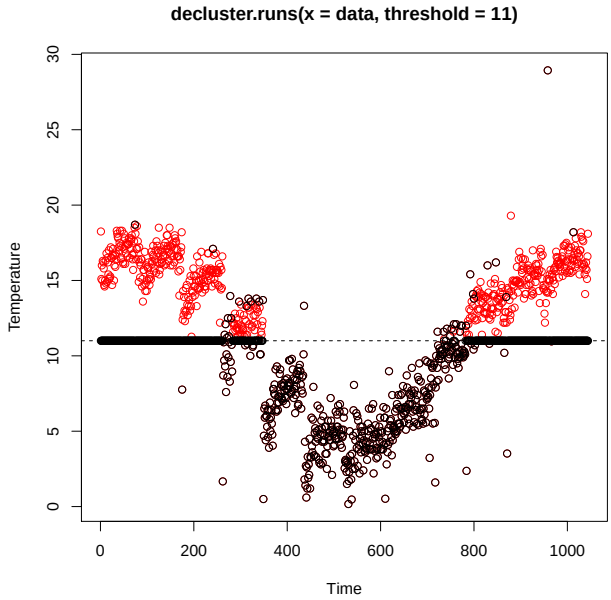

Figure 4.51: Polokwane declustered minimum temperature showing cluster minima above 11 ^oC threshold.

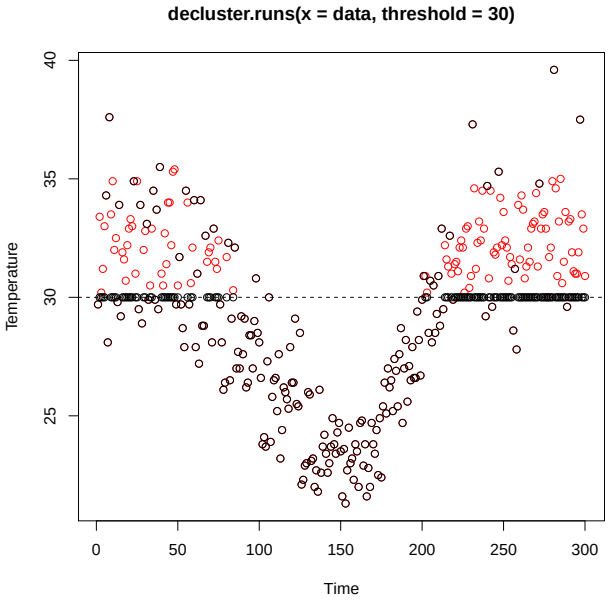

Figure 4.52: Thabazimbi declustered maximum temperature showing cluster maxima above 30 ^oC threshold.

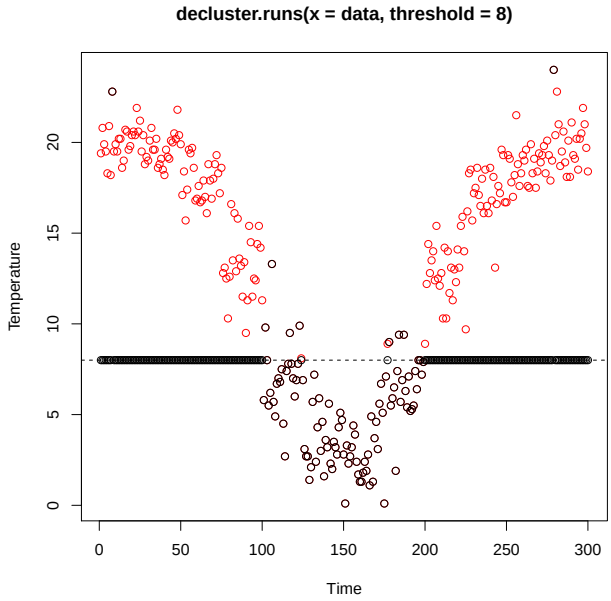

Figure 4.53: Thabazimbi declustered minimum temperature showing cluster minima above 8 ^oC threshold.

4.9.3 GP distribution approach

The generalised Pareto (GP) distribution parameter estimation results are presented in Table 4.34. The results revealed that both maximum and minimum temperature data for Mara, Messina, Polokwane and Thabazimbi can be modelled by the Weibull family of distribution since the values of all shape parameters are negative ($\hat{\xi} < 0$) and confidence intervals (CIs) for shape parameters are all significantly different from zero.

Table 4.34: Maximum likelihood estimates of the GP distribution parameters with standard errors in the parentheses and 95% CI for maximum temperature.

Station name		$\hat{\sigma}$	$\hat{\xi}$	95 % CI for $\hat{\xi}$
Mara	Max	3.540(0.163)	-0.471(0.024)	(-0.518, -0.424)
	Min	8.418(0.000)	-0.935(0.000)	(-0.935, -0.935)
Messina	Max	6.586(0.031)	-0.613(0.001)	(-0.615, -0.611)
	Min	8.581(0.000)	-0.833(0.000)	(-0.833, -0.833)
Polokwane	Max	4.818(0.172)	-0.298(0.012)	(-0.322, -0.274)
	Min	4.747(0.197)	-0.256(0.013)	(-0.281, -0.231)
Thabazimbi	Max	3.315(0.295)	-0.316(0.043)	(-0.400, -0.232)
	Min	15.906(0.000)	-0.994(0.000)	(-0.994, -0.994)

The diagnostic plots for assessing the accuracy of the GP model fitted to all meteorological stations are presented in Figure 4.54 to Figure 4.61. The P-P, Q-Q and return level plots for maximum temperature data of the following stations: Mara and Messina show that the plotted points are linear, suggesting that GP distribution fit temperature data well. The P-P, Q-Q, return level and density plots for Mara and Messina minimum temperature data deviate from linearity, suggesting that the GP distribution does not fit the data well. The P-P, Q-Q and return level plots for both Polokwane maximum and minimum temperature data deviate from linearity, suggesting that the GP distribution does not fit the data well. The P-P, Q-Q, return level and density plots for Thabazimbi maximum temperature data revealed that the plotted points are linear, suggesting that GP distribution fit temperature data well, while minimum tem-

perature plots deviate from linearity, suggesting that the GP distribution does not fit the data well.

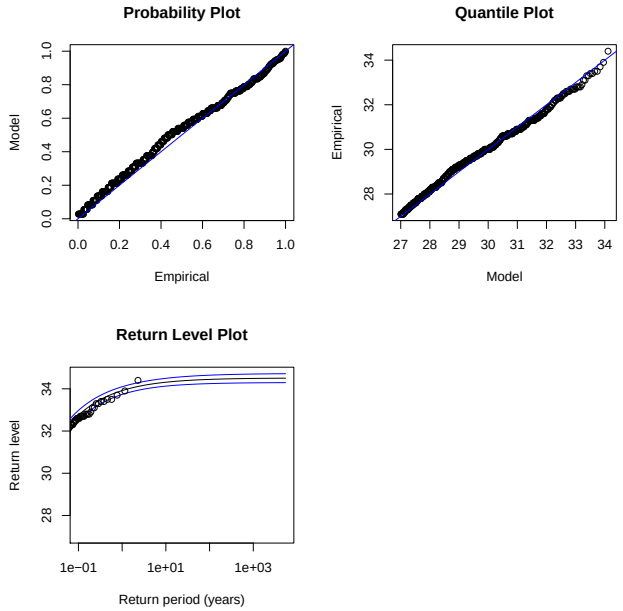

Figure 4.54: Diagnostic plots illustrating the fit of the GP distribution to Mara maximum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel) and (c) Return level (bottom left panel).

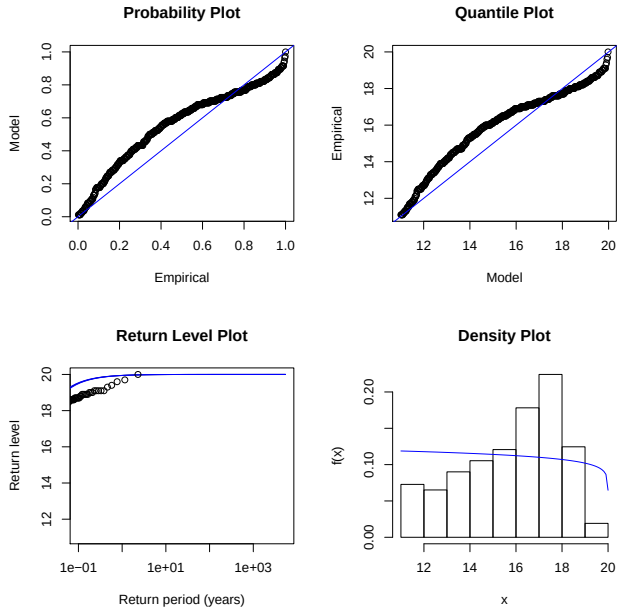

Figure 4.55: Diagnostic plots illustrating the fit of the GP distribution to Mara minimum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

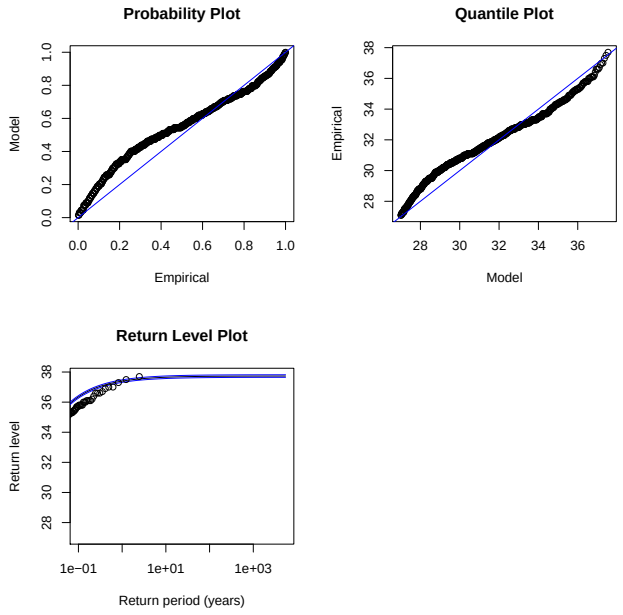

Figure 4.56: Diagnostic plots illustrating the fit of the GP distribution to Messina maximum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel) and (c) Return level (bottom left panel).

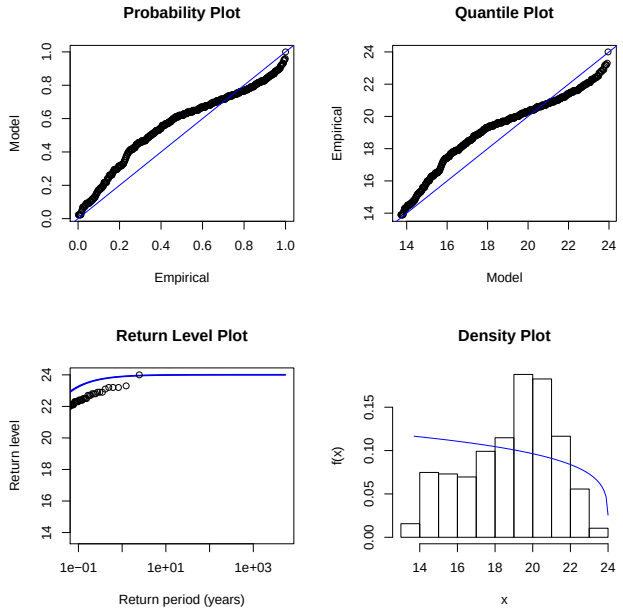

Figure 4.57: Diagnostic plots illustrating the fit of the GP distribution to Messina minimum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

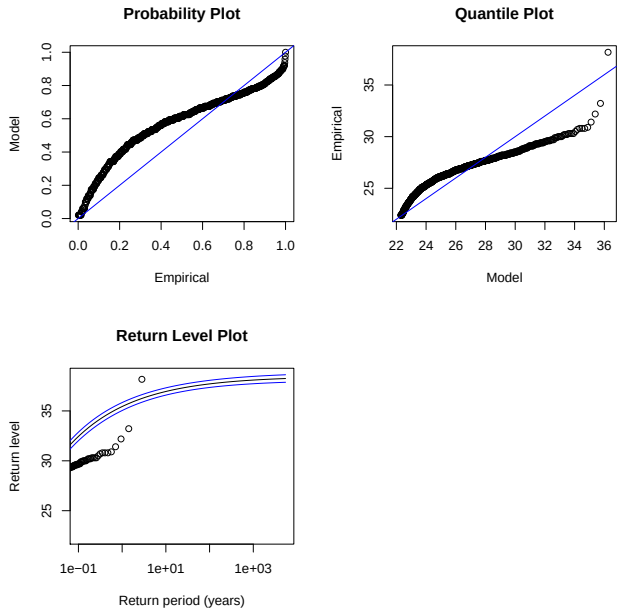

Figure 4.58: Diagnostic plots illustrating the fit of the GP distribution to Polokwane maximum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel) and (c) Return level (bottom left panel).

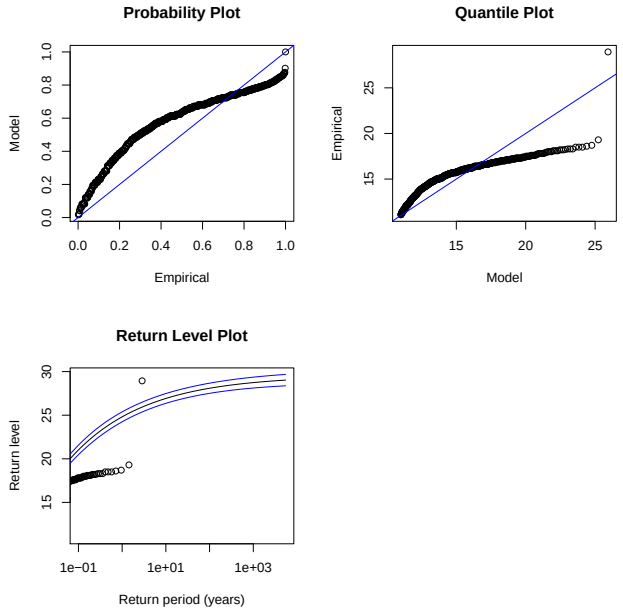

Figure 4.59: Diagnostic plots illustrating the fit of the GP distribution to Polokwane minimum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel) and (c) Return level (bottom left panel).

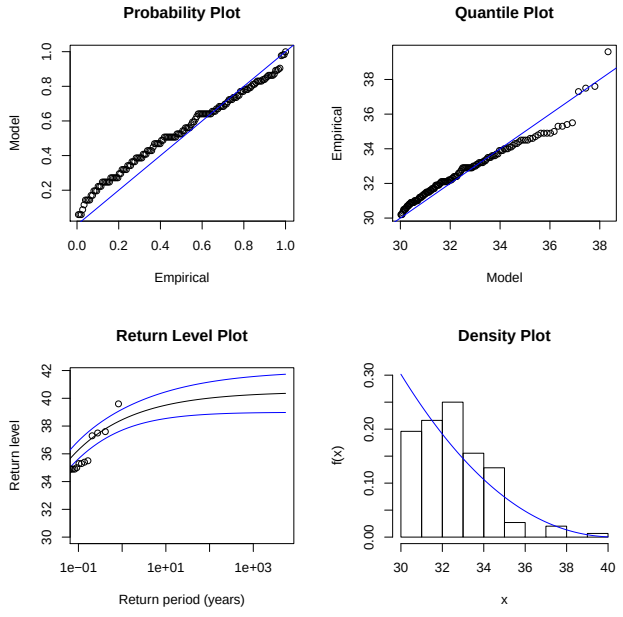

Figure 4.60: Diagnostic plots illustrating the fit of the GP distribution to Thabazimbi maximum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

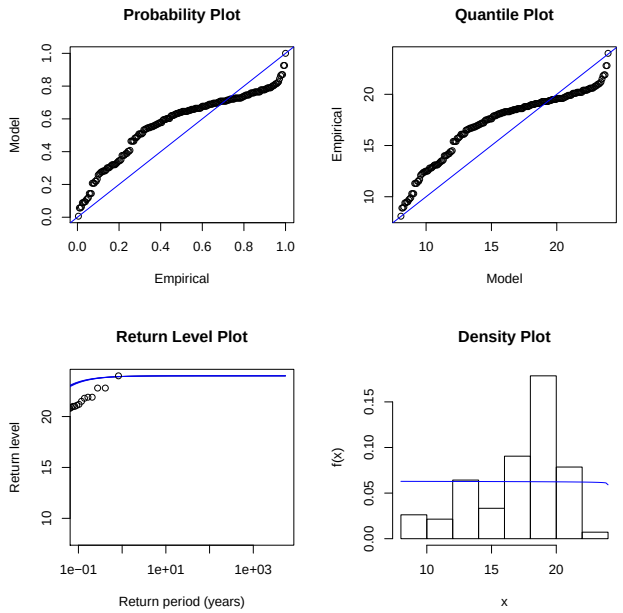

Figure 4.61: Diagnostic plots illustrating the fit of the GP distribution to Thabazimbi minimum temperature: (a) P-P plot (top left panel), (b) Q-Q plot (top right panel), (c) Return level (bottom left panel) and (d) Density plot (bottom right panel).

Return level analysis for GP distribution

Table 4.35 presents results for return periods and their corresponding return levels for both average monthly maximum and minimum temperatures based on the GP distribution. The results in Table 4.35 revealed that return levels for each station are increasing with increasing return periods. In comparison, Thabazimbi meteorological station has highest return levels for maximum temperature, while Polokwane has highest return levels for minimum temperature data.

Thabazimbi has the highest 10-year return level for maximum temperature of 37.124^0C, while Polokwane has the least maximum temperature 10-year return level of 32.939^0C. This suggests that, in Thabazimbi, a maximum temperature of 37.124^0C is expected to be exceeded at least once in 10 years. Thabazimbi has the highest 10-year return level for minimum temperature of 23.553^0C, while Mara has the least minimum temperature 10-year return level of 19.960^0C. This suggests that, in Thabazimbi, a minimum temperature of 23.553^0C is expected to be exceeded at least once in 10 years.

Thabazimbi has the highest 100-year return level for maximum temperature of 38.864^0C, while Mara has the least maximum temperature 100-year return level of 34.049^0C. This suggests that, in Thabazimbi, a maximum temperature of 38.864^0C is expected to be exceeded at least once in 100 years. Polokwane has the highest 100-year return level for minimum temperature of 25.451^0C, while Mara has the least minimum temperature 100-year return level of 19.965^0C. This suggests that, in Polokwane, a minimum temperature of 25.451^0C is expected to be exceeded at least once in 100 years.

The extremely high 10-year and 100-year maximum temperature return levels, particularly for Messina and Thabazimbi, indicate impending heat waves

in the Limpopo province. These findings are in support of those from the GEV distribution.

Table 4.35: GP model return periods (years) and their corresponding return levels (^oc).

Station name		10 years	20 years	50 years	100 years	500 years
Mara	Max	33.135	33.520	33.869	34.049	34.296
	Min	19.690	19.839	19.932	19.965	19.993
Messina	Max	36.554	36.963	37.295	37.447	37.628
	Min	23.490	23.716	23.871	23.930	23.986
Polokwane	Max	32.939	33.972	35.049	35.689	36.752
	Min	22.162	23.363	24.656	25.451	26.835
Thabazimbi	Max	37.124	37.786	38.466	38.864	39.512
	Min	23.553	23.776	23.910	23.955	23.991

4.9.4 Poisson point process approach

The Poisson point process approach parameter estimation results are presented in Table 4.36. The results revealed that both maximum and minimum temperature data for Mara, Messina, Polokwane and Thabazimbi can be modelled by the Weibull family of distribution since the values of all shape parameters are negative ($\hat{\xi} < 0$) and CI are all significantly different from zero.

Table 4.36: Maximum likelihood estimates of the Poisson point process parameters with standard errors in the parentheses and 95% CI for maximum and minimum temperatures.

Station name		$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	95 % CI for $\hat{\xi}$
Mara	Max	33.918(0.093)	0.282(0.026)	-0.471(0.025)	(-0.520, -0.422)
	Min	19.945(N/A)	0.0527(N/A)	-0.935(N/A)	(N/A)
Messina	Max	37.393(0.053)	0.207(0.021)	-0.614(0.024)	(-0.661, -0.567)
	Min	23.895(0.023)	0.092(0.014)	-0.833(0.035)	(-0.902, -0.764)
Polokwane	Max	35.439(0.202)	0.906(0.034)	-0.298(0.012)	(-0.322, -0.274)
	Min	24.814(0.291)	1.212(0.044)	-0.256(0.013)	(-0.281, -0.231)
Thabazimbi	Max	38.457(0.385)	0.642(0.099)	-0.316(0.043)	(-0.400, -0.232)
	Min	23.936(0.006)	0.064(0.006)	-0.994(0.018)	(-1.029, -0.959)

Key: N/A represents Not Applicable.

The diagnostic plots for assessing the accuracy of the Poisson point process model fitted to all meteorological stations are presented in Figure 4.62 to Figure 4.69. The P-P and Q-Q plots for maximum temperature of the following stations: Mara, Messina and Thabazimbi show that the plotted points are linear, suggesting that the distribution fits temperature data well, while minimum temperature data shows that the plotted points deviate from linearity, suggesting that the distribution does not fit the data well. For Polokwane, both maximum and minimum temperature data show that there are deviations from linearity suggesting that the distribution does not fit the data well.

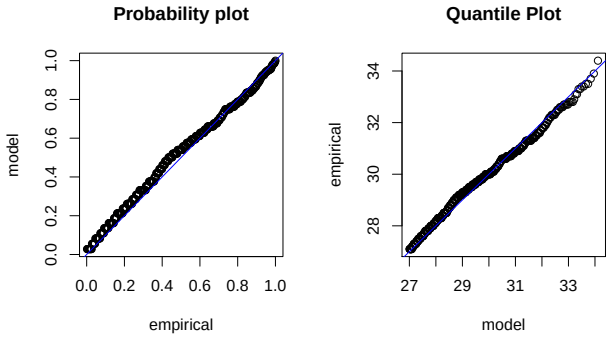

Figure 4.62: Diagnostic plots of stationary Poisson point process model fitted to Mara maximum temperature data.

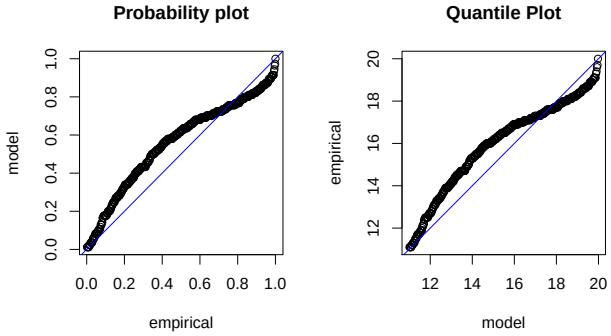

Figure 4.63: Diagnostic plots of stationary Poisson point process model fitted to Mara minimum temperature data.

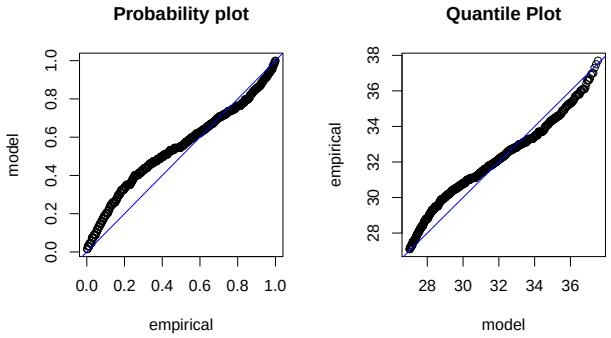

Figure 4.64: Diagnostic plots of stationary Poisson point process model fitted to Messina maximum temperature data.

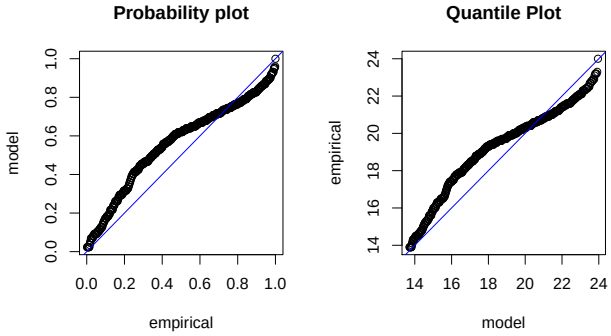

Figure 4.65: Diagnostic plots of stationary Poisson point process model fitted to Messina minimum temperature data.

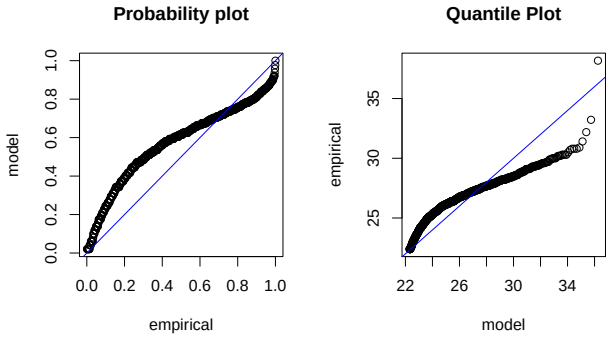

Figure 4.66: Diagnostic plots of stationary Poisson point process model fitted to Polokwane maximum temperature data.

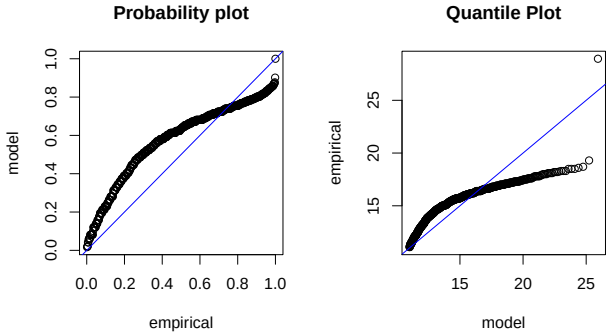

Figure 4.67: Diagnostic plots of stationary Poisson point process model fitted to Polokwane minimum temperature data.

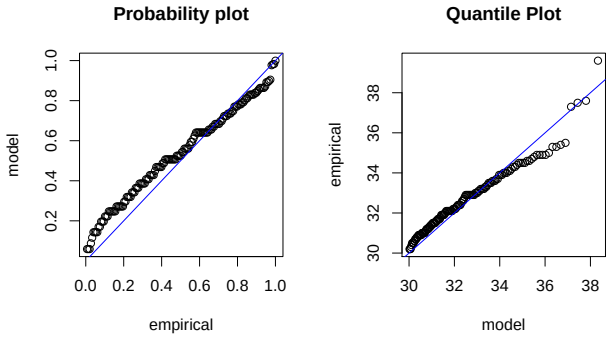

Figure 4.68: Diagnostic plots of stationary Poisson point process model fitted to Thabazimbi maximum temperature data.

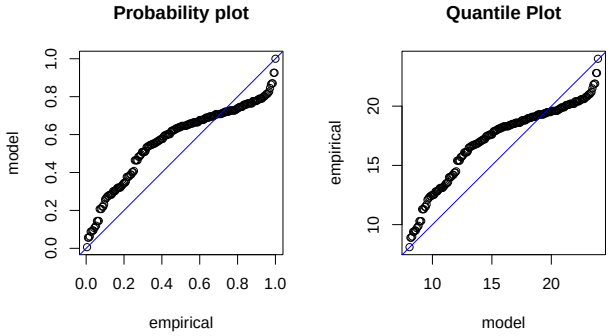

Figure 4.69: Diagnostic plots of stationary Poisson point process model fitted to Thabazimbi minimum temperature data.

Return level analysis for Poisson point process

Table 4.37 presents results for return periods and their corresponding return levels for both average monthly maximum and minimum temperatures based on the Poisson point process model. The results in Table 4.37 revealed that return levels for each station are increasing with increasing return periods. In comparison, Thabazimbi meteorological station has highest return levels for maximum temperature, while Polokwane has highest return levels for minimum temperature data.

Thabazimbi has the highest 10-year return level for maximum temperature of 39.491^0C, while Mara has the least maximum temperature 10-year return level of 34.308^0C. This suggests that, in Thabazimbi, a maximum temperature of 39.491^0C is expected to be exceeded at least once in 10 years. Polokwane has the highest 10-year return level for minimum temperature of 26.887^0C, while Mara has the least minimum temperature 10-year return level of 19.995^0C. This suggests that, in Polokwane, a minimum temperature of 26.887^0C is expected to be exceeded at least once in 10 years.

Thabazimbi has the highest 100-year return level for maximum temperature of 40.014^0C, while Mara has the least maximum temperature 100-year return level of 34.447^0C. This suggests that, in Thabazimbi, a maximum temperature of 40.014^0C is expected to be exceeded at least once in 100 years. Polokwane has the highest 100-year return level for minimum temperature of 28.090^0C, while Mara has the least minimum temperature 100-year return level of 20.001^0C. This suggests that, in Polokwane, a minimum temperature of 28.090^0C is expected to be exceeded at least once in 100 years.

The return level analysis results for Poisson point process further reveal that maximum temperature exceeding 40^0C are expected in the Limpopo province

at least once in the next 100 to 500 years time. These findings are consistent with the return level analysis findings based on both GEV and GP distributions. All these three approaches suggest impending heat waves in the Limpopo province to occur at least once in 10 years to come.

Table 4.37: Poisson point process model return periods (years) and their corresponding return levels (^oC).

Station name		10 years	20 years	50 years	100 years	500 years
Mara	Max	34.308	34.368	34.420	34.447	34.483
	Min	19.995	19.998	19.999	20.001	20.001
Messina	Max	37.646	37.677	37.700	37.711	37.724
	Min	23.988	23.996	24.001	24.003	24.005
Polokwane	Max	36.925	37.225	37.529	37.709	38.004
	Min	26.887	27.335	27.805	28.090	28.584
Thabazimbi	Max	39.491	39.694	39.897	40.014	40.204
	Min	23.994	23.997	23.999	23.910	23.000

4.10 Bivariate threshold excess approach

4.10.1 Introduction

This section of bivariate conditional extremes approach with a time-varying threshold is based on maximum temperature series. The data cover four major meteorological stations of the Limpopo province of South Africa: Mara (1949-2018), Messina [or Musina] (1934-2009), Polokwane (1932-2018) and Thabazimbi (1994-2018) [see Figure 3.1].

As can be noted from the previous paragraph, the longest series spans the period 1932-2018, while the shortest series spans the period 1994-2018. Also one station (Messina) has a series that terminates much earlier (in 2009) than the other three stations, implying that the years 2010-2018 which are included in the other stations are missing for Messina, and hence the period 2010-2018 will be excluded in the bivariate threshold excess approach so that all the stations

terminate in 2009. Therefore, given the differences in the time spans from the four stations and that in bivariate modelling approach the data variables must span the same time period, the section considers only the monthly maximum temperature for the period 1994-2009 for all the stations. This means that all the results in the next subsections and the findings therein are based on the time period 1994-2009.

4.10.2 Time series plot with time-varying threshold excess

Figure 4.70 to Figure 4.73 present the time series plots for Mara, Messina, Polokwane and Thabazimbi average monthly maximum temperature, respectively, with a time-varying threshold.

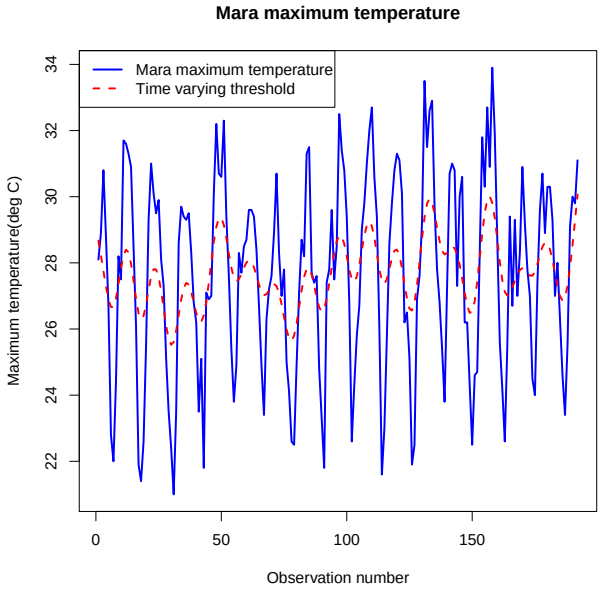

Figure 4.70: Mara maximum temperature with a time-varying threshold.

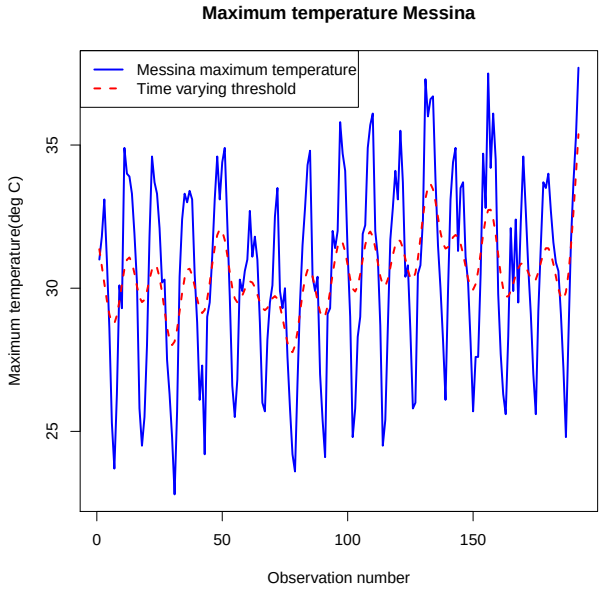

Figure 4.71: Messina maximum temperature with a time-varying threshold.

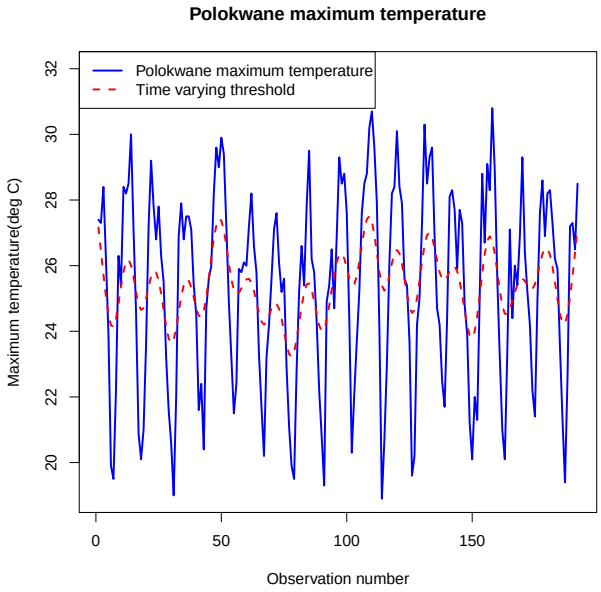

Figure 4.72: Polokwane maximum temperature with a time-varying threshold.

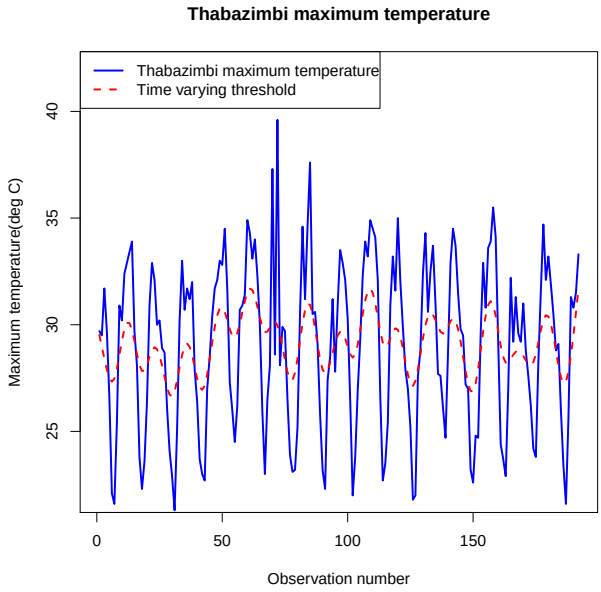

Figure 4.73: Thabazimbi maximum temperature with a time-varying threshold.

4.10.3 Extremal dependence

The pairwise correlation in Figure 4.74 shows the extremal dependence amongst the four stations. The results in Figure 4.74 reveal strong positive correlation between Messina and Mara (0.48), a very weak positive correlation between Polokwane and Mara (0.00115) and a weak positive correlation between Thabazimbi and Mara (0.234), between Polokwane and Messina (0.144) and between Thabazimbi and Polokwane (0.299).

Figure 4.75 and Figure 4.76 present the multivariate conditional Spearman's ρ correlation plots, where ρ denotes the Spearman's coefficient of correlation. The results in Figure 4.75 suggest a weak positive association in Messina and Polokwane, Messina and Thabazimbi, and Polokwane and Thabazimbi, while Figure 4.76 suggests a strong positive association in Mara and Messina, a very weak positive association in Mara and Polokwane, and a weak positive association in Mara and Thabazimbi. The multivariate conditional Spearman's correlation plots in Figure 4.75 and Figure 4.76 support the results of the pairwise correlation in Figure 4.74 by confirming the findings.

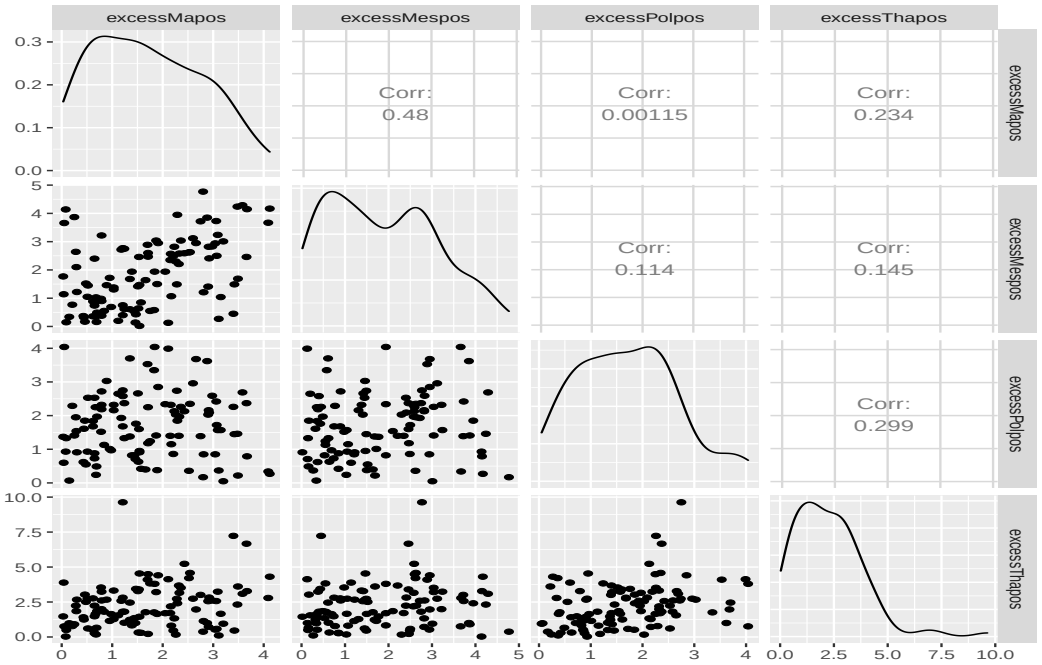

Figure 4.74: Pairwise scatterplot and correlation of the data.

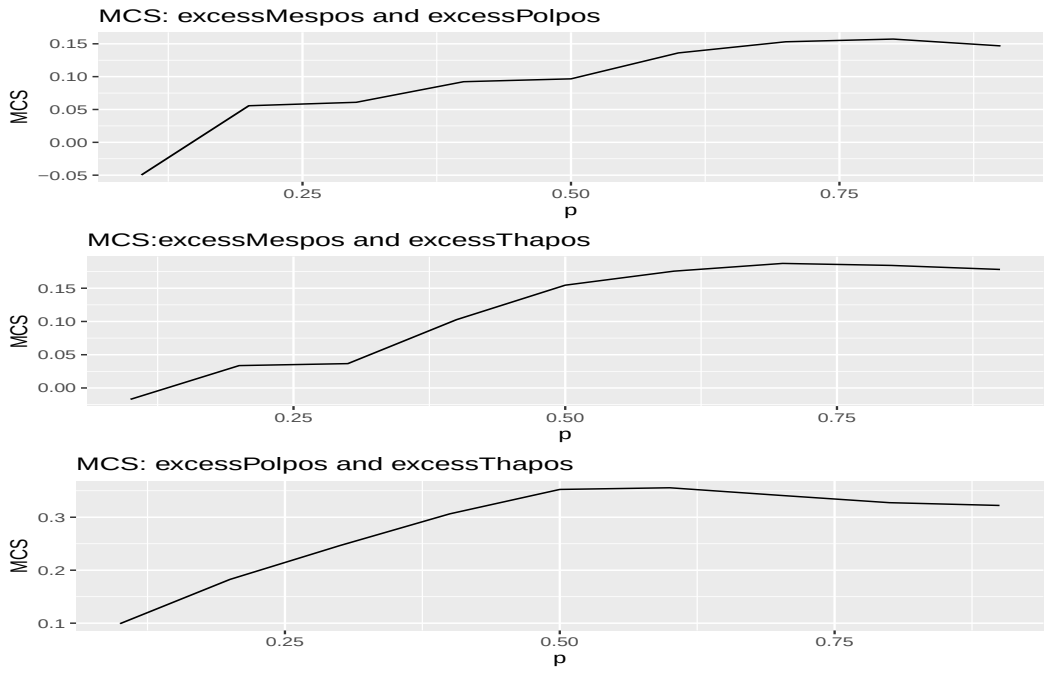

Figure 4.75: Multivariate conditional Spearman's correlation for Messina, Polokwane and Thabazimbi.

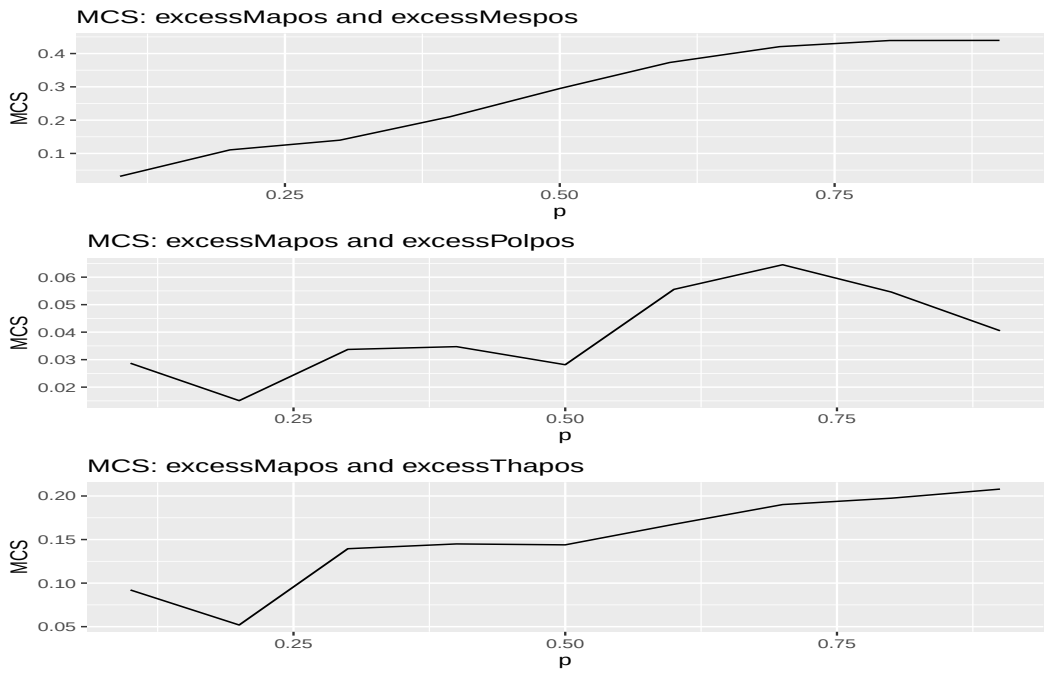

Figure 4.76: Multivariate conditional Spearman's correlation for Mara, Messina, Polokwane and Thabazimbi.

4.10.4 Conditional extremal dependence

Table 4.38 presents the dependence structure conditioning on each station. The estimates of the parameter a in Table 4.38 denote strong negative (or positive) extremal dependence for values close to -1 (or 1), respectively. Conditioning on Thabazimbi station, the estimates of the dependence parameters for Mara, Messina and Polokwane stations are $a = 0.2405$, $a = -0.001963$ and $a = 0.2575$, respectively. This implies that Mara and Polokwane stations have a positive extremal dependence on large values of Thabazimbi, while Messina has a negative extremal dependence on large values of Thabazimbi. Conditioning on Polokwane, the estimates of the dependence parameters for Mara, Messina and Thabazimbi are $a = -0.2902$, $a = -0.3007$ and $a = -0.02911$, respectively. This implies that Mara, Messina and Thabazimbi have a significant negative extremal dependence on large values of Polokwane.

Table 4.38: Estimates of dependence models.

	Dependence parameters	Mara	Messina	Polokwane
Conditioning on	a	0.2405*	-0.001963	0.2575*
Thabazimbi	b	0.2938	-0.128200	0.3147
	Dependence parameters	Mara	Messina	Thabazimbi
Conditioning on	a	-0.2902*	-0.3007*	-0.02911
Polokwane	b	0.0711	0.4061	-0.21600
	Dependence parameters	Mara	Polokwane	Thabazimbi
Conditioning on	a	-0.1475*	-0.59110*	-0.4077*
Messina	b	0.6061	-0.06039	-0.1408
	Dependence parameters	Messina	Polokwane	Thabazimbi
Conditioning on	a	0.4060*	-0.4624*	0.5155*
Mara	b	-0.3634	-0.5239	0.1690

* denotes significant positive or negative values of a at 5% significance level

Conditioning on Messina, the estimates of the dependence parameters for Mara, Polokwane and Thabazimbi are $a = -0.1475$, $a = -0.59110$ and $a = -0.4077$, respectively. This implies that Mara, Polokwane and Thabazimbi have significant negative extremal dependence on large values of Messina, the strongest negative being that of Polokwane on large values of Messina. Conditioning on Mara, the estimates of the dependence parameters for Messina and Thabazimbi are $a = 0.4060$ and $a = 0.5155$, respectively, which shows significant positive extremal dependence, while the estimate of the dependence parameter for Polokwane $(a = -0.4624)$ shows significant negative extremal dependence on large values of Mara. These results reveal that Thabazimbi exhibits the strongest positive extremal dependence on large values of Mara.

Figure 4.77 presents the diagnostic plots for conditioning Thabazimbi on Mara, Messina and Polokwane. The plots from top to bottom show; dependence model residuals across the range of the extreme conditioning variable, the absolute values of the centred and scaled values of the residuals across the range of the extreme conditioning variable and the original untransformed data with con-

tours showing quantiles of the fitted conditional model. The results in Figure 4.77 revealed that the parameter estimates are stable at the 75^{th} percentile since the horizontal lines are smoothest.

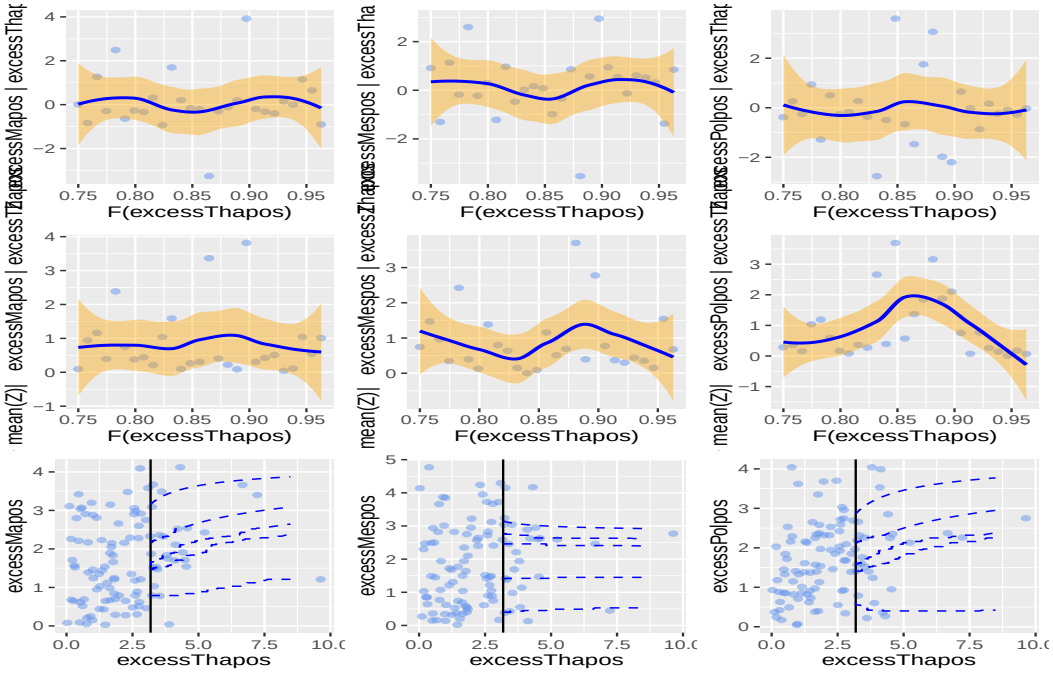

Figure 4.77: Diagnostic plots for conditioning on Thabazimbi.

4.11 Summary of the chapter

This chapter provided statistical modelling of temperature extremes in the four meteorological stations of Limpopo province. The data was first tested for stationarity and normality. The parent distributions of the data were investigated and the long-term trends were tested using M-K test statistic. The GEV, GP, Poisson point process distributions and bivariate threshold excess were fitted to the four meteorological stations of the Limpopo province of South Africa. The data were analysed using R software package. The main findings are as follows:

1. The distribution of the data is skewed to the left and platykurtic as shown by the skewness and kurtosis values.

2. The ADF, KPSS and PP tests were used to test for the stationarity of the maximum and minimum temperature data. The findings revealed that both maximum and minimum temperature data in the four meteorological stations are not stationary.

3. The Shapiro-Wilk and Jarque-Bera tests were used to test the normality of the temperature data. The findings revealed that the temperature data for all the stations are not normally distributed.

4. The four parent distributions namely: normal, log-normal, gamma and Weibull were fitted to the maximum and minimum temperature data. The findings revealed that the parent distribution for the maximum temperature in the Limpopo province lies in the Weibull domain of attraction, while for the minimum temperature the common parent distributions also lie in the Weibull domain of attraction based on the diagnostic tests conducted.

5. The Mann-Kendall (M-K) test and time series plots were used to investigate the long-term trend. The findings revealed both increasing and

decreasing long-term trends of maximum and minimum temperature extremes, respectively.

6. The GEV distribution was fitted to the block maxima and minima. The findings revealed that the maximum and minimum temperature data for Mara and Messina can be modelled by the Weibull family of distribution, while Polokwane and Thabazimbi data can be modelled by the Gumbel family of distribution.

7. The non-stationary GEV distribution with additional SOI covariate term was fitted to the block maxima and minima temperature data. The findings revealed that models with the SOI term, linear and quadratic trends in the location parameter provide significant improvement in fit over the stationary GEV model.

8. The GP distribution was fitted using the peaks-over-threshold approach. The findings revealed that maximum temperature data for the four stations can be modelled by the Weibull family of distribution.

9. The Poisson point process was fitted using the peaks-over-threshold approach. The findings revealed that both maximum and minimum temperature data for the four meteorological stations can be modelled by the Weibull family of distribution.

10. The maximum temperature 10-year return levels for all the various approaches GEV, GP and Poisson point process distributions suggest an impending heat wave and drought in the Limpopo province.

11. The bivariate conditional extremes approach with a time-varying threshold was applied to the maximum temperature data for the four meteorological stations in the study. The findings of the pairwise correlation matrix revealed positive extremal dependence amongst the four stations [Figure 3.1].

12. Conditional extremal dependence for the four meteorological stations revealed significant positive and negative extremal dependence among the stations.

Chapter 5

Conclusion

5.1 Conclusion

The aim of this study was to model the extreme maximum and minimum temperatures in the Limpopo province of South Africa. Temperature extremes accompanied by heat waves and cold waves pose serious challenges in major sectors such as economic, energy, agricultural and health sectors. For example, in the agricultural sector high temperatures lead to droughts and loss of livestock, in the health sector hospitalisation and loss of lives are experienced, while in the energy sector high demand for electricity is experienced due to changing weather conditions that affect the heating and cooling systems. The maximum and minimum temperatures in the Limpopo province are not stationary throughout the year with June, July and August being the coldest months of the year, while December, January, February and March are the hottest months. Recent studies have indicated increasing maximum and minimum temperatures attributed to climate change and global warming.

The present study has investigated the stationarity of the data using ADF, KPSS and PP tests. The findings for ADF and KPSS tests revealed that both maximum and minimum temperature data for the four meteorological stations are not stationary. The Shapiro-Wilk (SW) and Jarque-Bera (JB) tests were used to test for normality of the temperature data. The findings revealed that both maximum and minimum temperature data for the four meteorological stations are not normally distributed. These findings lead to the conclusion that both maximum and minimum temperature data in the four meteorological stations are neither stationary nor normally distributed.

The study also investigated parent distributions for all the stations. The four parent distributions considered in this study are: normal, log-normal, gamma and Weibull. The findings suggest that the underlying distribution of the data belongs to the Weibull domain of attraction (which is a family of short-tailed distributions). The long-term trend of the temperature data was investigated using Mann-Kendall (M-K) test and time series plots. The findings revealed both monotonic increasing and decreasing long-term trends of maximum and minimum temperature extremes, respectively.

The major contribution of the study on parent distributions was on establishing that the underlying distribution of the maximum and minimum temperatures in the Limpopo province belongs to the Weibull domain of attraction, while the major contribution of the M-K test was on detecting the long-term trends of the maximum and minimum temperatures in the Limpopo province which could not be easily detected using time series plots.

The GEV distribution was fitted to both maximum and minimum average monthly temperature data using the block maxima approach. Non-stationary GEV models with linear and quadratic trends in the location and scale parameters

were considered in this study. The findings revealed that the model with the linear and quadratic trends in the location parameter provides significant improvement in fit over the stationary GEV model for Mara, Messina and Polokwane meteorological stations. An alternative modelling approach involved fitting a GEV model with a SOI term covariate in the location parameter of the model was considered. The findings for the non-stationary GEV model with a SOI term covariate revealed that the SOI term and a linear trend in the location parameter provide significant improvement in fit over the stationary GEV model. These climate change findings in this study could be attributed mainly to the effects of global warming and natural modes of interdecadal variability such as the El Niño and La Niña phenomenon (Maposa, 2016).

The GP and Poisson point process distributions were fitted using a peaks-over-threshold approach. The findings revealed that both maximum and minimum temperature data in the four stations can be modelled by the Weibull family of distribution. The return levels for the GEV, GP and Poisson point process distributions revealed that maximum and minimum temperatures in the Limpopo province are expected to increase in the next 10 to 500 years. These findings suggests that temperature extremes in the Limpopo province are expected to occur more frequently, last longer and become more intense.

The major contributions of the three univariate extreme value theory approaches, GEV, GP and Poisson point process distributions were in predicting impending heat waves in the Limpopo province of South Africa, as well as establishing that the distribution of the temperature data in the Limpopo province belongs to the Weibull domain of attraction.

The bivariate conditional extremes approach with a time-varying threshold was also considered in this study. The findings for the pairwise correlation

matrix revealed that Messina, Polokwane and Thabazimbi all have positive extremal dependence on Mara, the strongest being that of Messina and the weakest being that of Polokwane. Other weak pairwise extremal dependence correlations were revealed for the pairs Polokwane and Messina, Thabazimbi and Messina, and Thabazimbi and Polokwane.

Conditioning on Thabazimbi station, the values of the estimated dependence parameters show that the maximum temperatures of Thabazimbi have a positive extremal dependence on Mara and Polokwane, and a negative extremal dependence on Messina. Conditioning on Polokwane station, the values of the estimated dependence parameters show that Mara, Messina and Thabazimbi all have negative extremal dependence on Polokwane, the strongest being that of Messina on Polokwane. Conditioning on Messina station, the values of the estimated dependence parameters show that Mara, Polokwane and Thabazimbi all have negative extremal dependence on large values of Messina, the strongest being that of Polokwane on Messina. Conditioning on Mara station, the values of the estimated dependence parameters show that the maximum temperatures of Mara have positive extremal dependence on Messina and Thabazimbi, and negative extremal dependence on Polokwane. The strongest positive conditional extremal dependence was revealed to be that of Thabazimbi on Mara. This suggests that the occurrence of extreme maximum temperatures in Mara would imply some extreme high temperature dependence impact on Thabazimbi.

The major contribution of this bivariate time-varying threshold excess approach was in using a penalised cubic smoothing spline to perform a nonlinear detrending of the temperature data prior to fitting bivariate threshold excess models based on Laplace margins to positive residuals above the threshold and a positive shift factor as a time-varying threshold to capture the season-

ality or climate change effects in the data. The existing gap in literature was in combining these two approaches in conditional extremes dependence modelling. The overall significance of this bivariate conditional extremes dependence modelling approach lies on quantifying the dependence effects of maximum temperature extremes amongst the various meteorological stations in order to reveal some useful information needed for planning by climatologists, meteorologists, agriculturalists, decision markers and planners in the energy sector.

5.2 Recommendations of the study

The reseacher makes the following recommendations based on the present study:

1. The South Africa Weather Service (SAWS) should make attempts to record and provide researchers with daily temperature data to allow more statistical inferences to be made on the data compared to monthly average temperature data.

2. The SAWS should make efforts to provide researchers with data that does not have many missing values so that the number of meteorological stations in studies of this nature can be increased.

3. The findings in this study are intended to enhance awareness and understanding of temperature extremes for the community and decision-makers in the Limpopo province to improve their mitigation strategies and reduce the deleterious impacts of these temperature extremes on humans and property.

4. The return level analyses suggested impending heat waves. Therefore, the Limpopo province community must brace itself for such heat waves in the near future.

5.3 Limitations of the study

The researcher was aware and mindful of some limitations to the study. For instance, the data that was brought forth before the research was extremely limited hence posing some challenges in terms of the data itself and the analyses that followed. The data had 11 stations, but due to missing values, only four stations were considered. Another challenge was that the data did not span over the same period, making it difficult to model using bivariate conditional extremes time-varying threshold approach.

5.4 Future studies

Future studies in modelling of temperature extremes will consider the use of Bayesian approach using Markov chain Monte Carlo (MCMC) techniques. In extremal dependence modelling of temperature extremes, future studies may consider exploring the use of extreme value copulas while paying special attention to asymptotic independence. With increased availability of meteorological stations, another important consideration in the future will be investigation of spatio-temporal dependence between temperature extremes using the conditional extremes model of Heffernan and Tawn (2004).

References

AL-SUHILI, R. H. AND KHANBILVARD, R. (2014). Frequency analysis of the monthly rainfall data at Sulaimania region, Iraq. *American Journal of Engineering Research*, **3** (5), 212–222.

ALHASSAN, B. G. AND TELA, N. M. (2019). An improvement forecasting using vector autoregression approach on tax revenue in Nigeria. *Journal of Engineering and Applied Sciences*, **11** (2).

ANDERSSON, E. AND NILSSON, E. (2018). *An extreme value approach to modelling risk of extreme rainfall in Bangladesh*. Bachelor Thesis in Mathematical Statistics, Lund University.

ARCHER, E., LANDMAN, W., MALHERBE, J., TADROSS, M., AND PRETORIUS, S. (2019). South Africa's winter rainfall region drought: A region in transition? *Climate Risk Management*, **25**, 100–188.

AYUKETANG, N. AND JOSEPH, E. (2014). *Modelling extreme temperature in Cameroon using generalized extreme value distribution*. University of Buea and AIMS, Cameroon.

BARNETT, A. G., HAJAT, S., GASPARRINI, A., AND ROCKLÖV, J. (2012). Cold and heat waves in the United States. *Environmental Research*, **112**, 218–224.

BENISTON, M., STEPHENSON, D. B., CHRISTENSEN, O. B., FERRO, C. A., FREI, C., HALSNAES, S. G. K., HOLT, T., JYLHÄ, K., KOFFI, B., PALUTIKOF,

J., SCHÖLL, R., SEMMLER, T., AND WOTH, K. (2007). Future extreme events in European climate: An exploration of regional climate model projections. *Climate Change*, **81** (1), 71–95.

BHAGWANDIN, L. (2017). *Multivariate extreme value theory with an application to climate data in the Western Cape province*. MSc dissertation, University of Cape Town.

BOMMIER, E. (2014). *Peaks-over-threshold modelling of environmental data*. Project Report, Uppsala University.

BOWLING, L. C., WIDHALM, M., CHERKAUER, K. A., BECKERMAN, J., BROUDER, S., BUZAN, J., DOERING, O., DUKES, J., EBNER, P., FRANKENBURGER, J., AND GRAMIG, B. (2018). Indiana's agriculture in a changing climate: A report from the Indiana climate change impacts assessment. *Agriculture Report*, Paper 1.

BROWN, S. J., CAESAR, J., AND FERRO, C. A. (2008). Global changes in extreme daily temperature since 1950. *Journal of Geophysical Research: Atmospheres*, **113** (D5).

CDKN (2012). *Managing climate extremes and disasters in Africa: Lesson from the SREX report*. Climate and Development Knowledge Network (CDKN). Last accessed: 2020-25-02.
URL: *https://www.cdkn.org/srex*

CHAN, S. AND NADARAJAH, S. (2015). Extreme value analysis of electricity demand in the UK. *Applied Economics Letters*, **22** (15), 1246–1251.

CHEN, H. AND ZHAO, T. (2020). Modelling power loss during blackouts in China using non-stationary generalised extreme value distribution. *Energy*, **195**, 117044.

CHIFURIRA, R. (2018). *Modelling mean annual rainfall for Zimbabwe*. PhD Thesis, University of the Free State.

CHIKOBVU, D. AND CHIFURIRA, R. (2015). Modelling of extreme minimum rainfall using generalised extreme value distribution for Zimbabwe. *South African Journal of Science*, **111** (9-10), 1–8.

CHIKOBVU, D. AND SIGAUKE, C. (2013). Modelling influence of temperature on daily peak electricity demand in South Africa. *Journal of Energy South Africa*, **24** (4), 63–70.

COLES, S. (2001). *An introduction to statistical modelling of extreme values*. Springer-Verlag, London.

COUMOU, D. AND ROBINSON, A. (2013). Historic and future increase in the global land area affected by monthly heat extremes. *Environmental Research Letters*, **8** (3), 034018.

COUMOU, D., ROBINSON, A., AND RAHMSTORF, S. (2013). Global increase in record breaking monthly-mean temperatures. *Climate Change*, **118** (3-4), 771–782.

CSP (2015). *Adaptation strategies for Limpopo province*. Climate Support Programme (CSP), Limpopo Climate Adptation Strategy Report. Last accessed: 2020-11-27.
URL: *https://www.ledet.gov.za*

DARON, J. D. (2014). Regional climate messages Southern Africa. Scientific report from the CARIAA adaption at scale in semi-arid regions (ASSAR) project. Last accessed: 2020-18-03.
URL: *https://www.assar.uct.ac.za*

DEA (2013). *Climate trends and scenarios*. Department of Envirinmental Affairs (DEA) Report, Pretoria.

DEA (2017). *South Africa's third climate change*. Department of Envirinmental Affairs (DEA) Report, Pretoria.

DEA (2019). *National climate change adaptation strategy*. Department of Envirinmental Affairs (DEA) Report, Pretoria.

DEY, A. K., EDWARDS, A., AND DAS, K. P. (2020). Determinants of high crude oil price: A non-stationary extreme value approach. *Journal of Statistical Theory and Practice*, **14** (1), 4.

DIRIBA, T. A. AND DEBUSHO, L. K. (2020). Modelling dependency effect to extreme value distributions with application to extreme wind speed at Port Elizabeth, South Africa: A frequentist and Bayesian approaches. *Computational Statistics*, 1–31.

DIRIBA, T. A., DEBUSHO, L. K., AND BOTAI, J. (2015). Modeling extreme daily temperature using generalized Pareto distribution at Port Elizabeth. *In Annual Proceedings of the South African Statistical Association (SASA) Conference*, **2015** (1), 41–48.

GAO, M. AND ZHENG, H. (2018). Nonstationary extreme value analysis of temperatures in China. *Stochastic Environmental Research and Risk Assessment*, **32** (5), 1299–1315.

GEBRECHORKOS, S. H., HÜLSMANN, S., AND BERNHOFER, C. (2019). Changes in temperature and precipitation extremes in Ethiopia, Kenya, and Tanzania. *International Journal of Climatology*, **39** (1), 18–30.

GEL, Y. R. AND GASTWIRTH, J. L. (2008). A robust modification of the Jarque-Bera test of normality. *Economic Letter*, **99** (1), 30–32.

GILLELAND, E. AND KATZ, R. W. (2016). ExtRemes 2.0: An extreme value analysis package in R. *Journal of Statistical Software*, **72** (8), 1–39.

GONZALEZ-SANCHEZ, E. J., VEROZ-GONZALEZ, O., CONWAY, G., MORENO-GARCIA, M., KASSAM, A., MKOMWA, S., ORDOÑEZ-FERNANDEZ, R., TRIVIÑO-TARRADAS, P., AND CARBONELL-BOJOLLO, R. (2019). Meta-analysis on carbon sequestration through conservation agriculture in Africa. *Soil and Tillage Research*, **190**, 22–30.

GRUMM, R. H. (2011). The central European and Russian heat event of July-August 2010. *Bulletin of the American Meteorological Society*, **92** (10), 1285–1296.

HALES, S., EDWARDS, S. J., AND KOVATS, R. S. (2003). Impacts on health of climate extremes. *Climate Change and Human Health: Risks and Responses*, 79–96.

HASAN, H. B., RADI, N. A., AND KASSIM, S. (2012). Modeling of extreme temperature using generalized extreme value (GEV) distribution: A case study of Penang. *In Proceedings of the World Congress on Engineering*, **1**, 1–6.

HASAN, H. B., RADI, N. A., AND KASSIM, S. (2013). Modeling annual extreme temperature using generalized extreme value distribution: A case study in Malaysia. *In AIP Conference Proceedings*, **1522** (1), 1195–1203.

HATFIELD, J. L. AND PRUEGER, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, (10), 4–10.

HEFFERNAN, J. E. AND TAWN, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, B*, **66** (3), 497–546.

HENDERSON, K. G. AND MULLER, R. A. (1997). Extreme temperature days in the South Central United States. *Climate Research*, **8** (2), 151–162.

HERRING, S. C., CHRISTIDIS, N., HOELL, A., HOERLING, M. P., AND STOTT, P. A. (2020). Explaining extreme events of 2018 from a climate perspective. *Bulletin of the American Meteorological Society*, **101** (1).

HINTZ, M. J., LUEDERITZ, C., LANG, D. J., AND WEHRDEN, H. V. (2018). Facing the heat: A systematic literature review exploring the transferability of solutions to cope with urban heat waves. *Urban Climate*, **24**, 714–727.

HOHNE, P. A., KUSAKANA, K., AND NUMBI, B. P. (2019). A review of water heating technologies: An application to the South Africa context. *Energy Report*, **5**, 1–19.

HULME, M., DOHERTY, R., NEW, M., AND LISTER, D. (2001). African climate change: 1900-2100. *Climate Research*, **17** (2), 145–168.

IRFAN, U. (2019). *113 degrees in France: Why Europe is so vulnerable to extreme heat*. Wikipedia.
URL: *https://www.vox.com/world/2019/6/26/18744518/heat-wave-2019-europe-france-germany-spain*

JAKATA, O. AND CHIKOBVU, D. (2019). Modelling extreme risk of the South African financial index (j580). *Journal of Economic and Financial Sciences*, **12** (1), 1–7.

JIANG, S. AND KANG, L. (2019). Flood frequency analysis for annual maximum streamflow using a non-stationary GEV model. *In E3S Web of Conferences*, **79**, 03022.

JIMMY, R., GOVENDER, P., BENCHERIF, H., AND MOODLEY, K. M. (2019). Trend-Run model application of surface temperature and its implications for South Africa forestry and reforestation using local weather service data. *Reforesta*, **7**, 50–72.

JOHN, C., GEORGE, U., AND CHUKWUEMEKA, O. S. (2014). Time series analysis and forecasting of monthly maximum temperatures in South Eastern Nigeria. *International Journal of Innovation Research and Development*, **3** (1), 165–171.

KAJAMBEU, R. (2016). *Modelling flood heights of the Limpopo River at Beitbridge Border post using extreme value distribution*. MSc dissertation, University of Venda.

KARMESHU, N. (2012). *Trend detection in annual temperature and precipitation using the Mann-Kendall test: A case study to assess climate change on select states in the Northeastern United States*. MSc dissertation, University of Pennysylvania.

KAUWE, M. G. D., MEDLYN, B. E., PITMAN, A. J., DRAKE, J. E., UKKOLA, A., GRIEBEL, A., PENDALL, E., PROBER, S., AND RODERICK, M. (2019). Examining the evidence for decoupling between photosynthesis and transpiration during heat extremes. *Biogeosciences*, **16** (4), 903–916.

KEELLINGS, D. AND WAYLEN, P. (2015). Investigating teleconnection drivers of bivariate heat waves in Florida using extreme value analysis. *Climate Dynamics*, **44** (11-12), 3383–3391.

KIM, S. Y., SAPOTTA, B., JANG, G., KANG, Y. H., AND KIM, H. G. (2020). Prefeasibility study of photovoltaic power potential based on a skew-normal distribution. *Energies*, **13** (3), 676.

KRUGER, A. C. AND SEKELE, S. S. (2013). Trends in extreme temperature indices in South Africa: 1962-2009. *International Journal of Climatology*, **33** (3), 661–676.

KRUGER, A. C. AND SHONGWE, S. (2004). Temperature trends in South Africa: 1960-2003. *International Journal of Climatology: A Journal of the Royal Meteorology Society*, **24** (15), 1929–1945.

LAZOGLOU, G., ANAGNOSTOPOULOU, C., TOLIKA, K., AND KOLYVA-MACHERA, F. (2019). A review of statistical methods to analyze extreme precipitation and temperature events in the Mediterranean region. *Theoretical and Applied Climatology*, **136** (1-2), 99–117.

LEO (2016). *Economics for the Limpopo province, South Africa*. Limpopo environmental outlook report. Discussion document, First draft. Last accessed: 2020-27-02.
URL: *https://www.ecoafrica.co.za/sites/default/files/project*

LI, Z., LI, C., XU, Z., AND ZHOU, X. (2014). Frequency analysis of precipitation extremes in Heihe River basin based on generalised Pareto distribution. *Stochastic Environment Research and Risk Assessment*, **28** (7), 1709–1721.

LUBER, G. AND MCGEEHIN, M. (2008). Climate change and extreme heat events. *American Journal of Preventive Medicine*, **35** (5), 429–435.

MAKATE, C., MAKATE, M., MANGO, N., AND SIZIBA, S. (2019). Increasing resilience of smallholder farmers to climate change through multiple adoption of proven climate-smart agriculture innovations: Lessons from Southern Africa. *Journal of Environmental Management*, **231**, 858–868.

MAPONYA, P. AND MPANDELI, S. (2012a). Climate change and agriculture production in South Africa: Impacts and adaptation options. *Journal of Agricultural Science*, **4** (10), 48.

MAPONYA, P. AND MPANDELI, S. (2012b). Impact of drought on food scarcity in Limpopo province, South Africa. *African Journal of Agricultural Research*, **7** (37), 5270–5277.

MAPOSA, D. (2016). *Statistics of extreme with applications to extreme flood heights in the lower Limpopo River basin of Mozambique*. PhD Thesis, University of Limpopo.

MAPOSA, D., COCHRAN, J. J., AND LESAOANA, M. (2017). Modelling extreme flood heights in the lower Limpopo River basin of Mozambique using a time-heterogeneous generalised Pareto distribution. *Statistics and Its Interface*, **10** (1), 131–144.

MAPOSA, D., COCHRAN, J. J., LESAOANA, M., AND SIGAUKE, C. (2014). Estimating high quantiles of extreme flood heights in the lower Limpopo River basin of Mozambique using model based Bayesian approach. *Natural Hazard and Earth System Sciences Discussions*, **2** (8), 5401–5425.

MASEREKA, E. M., OCHIENG, G. M., AND SNYMAN, J. (2018). Statistical analysis of annual maximum daily rainfall for Nelspruit and its environs. *Jàmbá: Journal of Disaster Risk Studies*, **10** (1), 1–10.

MBOKODO, I. L. (2017). *Heat waves in South Africa: Observed variability, structure and trends*. MSc dissertation, University of Venda.

MENG, G., JING, L., HONGSHI, H., JIAWEI, X., AND YINGHUA, J. (2018). Detecting global vegetation changes using Mann-Kendal (MK) trend test for 1982-2015 time period. *Chinese Geographical Science*, **28** (6), 907–919.

MILLER, N. L., HAYHOE, K., JIN, J., AND AUFFHAMMER, M. (2008). Climate, extreme heat, and electricity demand in California. *Journal of Applied Meteorology and Climatology*, **47** (6), 1834–1844.

MORI, K. W. (2016). *Modelling extreme temperature behaviour in upper East region, Ghana*. MSc dissertation, University for Development Studies.

MOTHUPI, T., THUPENG, W. M., MASHABE, B., AND MOKOTO, B. (2016). Estimating extreme quantiles of the maximum surface air temperatures for the Sir Seretse Khama International Airport using generalised extreme value distribution. *American Journal of Theoretical and Applied Statistics*, **5** (6), 365–375.

MPANDELI, N. S. AND MAPONYA, P. I. (2013). Coping with climate variability in Limpopo province, South Africa. *Peak Journal of Agricultural Sciences*, **4** (1), 54–64.

MUSETHA, M. A. (2016). *The impact of climate change on agriculture crop production in the Vhembe District Municipality, Limpopo province, South Africa*. MSc dissertation, University of South Africa.

MUSVOTO, C. (2019). *Climate resilience through resources efficiency in smallholder vegetable production in South Africa*. Council for Scientific and Industrial Research, Pretoria, South Africa.

NASRI, M. AND MODARRES, R. (2009). Dry spell trend analysis of Isfahan province, Iran. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **29** (10), 1430–1438.

NDLOVU, M. S. AND DEMLIE, M. (2018). Statistical analysis of groundwater level variability across KwaZulu-Natal province, South Africa. *Environmental Earth Science*, **77** (21), 739.

NELSON, G. C., ROSEGRANT, M. W., KOO, J., ROBERTSON, R., SULSER, T., ZHU, T., RINGLER, T., MSANGI, S., PALAZZO, A., BATKA, M., MAGALHAES, M., VALMONTE-SANTOS, R., EWING, M., AND LEE, D. (2009). *Climate change: Impact on agriculture and cost of adaptation*. International Food Policy Research institute, Washington.

NEMUKULA, M. M. (2018). *Modelling temperature in South Africa using extreme value theory*. MSc dissertation, University of the Witwatersrand, Johannesburg.

NEMUKULA, M. M. AND SIGAUKE, C. (2018). Modelling average maximum daily temperature using r largest order statistics: An application to South African data. *Jàmbá: Journal of Disaster Risk Studies*, **10** (1), 1–11.

NEMUKULA, M. M., SIGAUKE, C., AND MAPOSA, D. (2018). Bivariate threshold excess models with application to extreme high temperatures in Limpopo province of South Africa. *In Annual Proceedings of the South African Statistical Association (SASA) Conference*, **2018** (1), 33–40.

NGAILO, J. T., REUDER, J., RATALEBWA, E., NYIMVUA, S., AND MESQUITA, D. S. M. (2016). Modelling of extreme maximum rainfall using extreme value theory for Tanzania. *International Journal of Scientific and Innovative Mathematical Research*, **4** (3), 34–45.

NHAMO, L., MATCHAYA, G., MABHAUDHI, T., NHLENGETHWA, S., NHEMACHENA, C., AND MPANDELI, S. (2019). Cereal production trends under climate change: Impacts and adaptation strategies in Southern Africa. *Agriculture*, **9** (2), 30.

NHEMACHENA, C. AND HASSAN, R. (2007). *Micro-level analysis of farmers adaption to climate change in Southern Africa*. International Food Policy Research institute, Washington.

NURY, A. H., KOCH, M., AND ALAM, M. J. B. (2013). Time series analysis and forecasting of temperatures in the Sylhet division of Bangladesh. *In:* 4^{th} *International Conference on Environmental Aspect of Bangladesh (ICEAB)*, 24–26.

OCHANDA, O. O. (2016). *Time series analysis and forecasting of monthly air temperature changes in Nairobi, Kenya*. MSc dissertation, University of Nairobi.

ONGOMA, V. AND CHEN, H. (2017). Temporal and spatial variability of temperature and precipitation over East Africa from 1951 to 2010. *Meteorology and Atmospheric Physics*, **129** (2), 131–144.

OSMAN, Y. Z., FEALY, R., AND SWEENEY, J. (2015). Modelling extreme temperatures in Ireland under global warming using a hybrid peak-over-

threshold and generalised Pareto distribution approach. *International Journal of Global Warming*, **7** (1), 21–47.

PEREIRA, L. (2017). *Climate change impacts on agriculture across Africa*. Oxford Research Encyclopedia of Environmental Science. Oxford, UK, Oxford University Press.

PHILLIPS, L. (2020). *Keeping it cool: Dealing with extreme temperatures*. F. Fact-Journals.co.za. Last accessed: 2020-27-02.
URL: *https://www.magzter.com*

PHOPHI, M. M., MAFONGONYA, P., AND LOTTERING, S. (2020). Perceptions of climate change and drivers of insect pest outbreaks in vegetable crops in Limpopo province of South Africa. *Climate*, **8** (2), 27.

PIENNAR, L. AND BOONZAAIER, J. (2018). *Drought policy brief Western Cape agriculture*. Western Cape Department of Agriculture (WCDoA) and the Bureau for Food and Agriculture (BFAP), Elsenburg.

POHLERT, T. (2020). *Non-parametric trend test and change-Point detection*. Creative Commons Licence.
URL: *https://www.creativecommons.org/licenses/by-nd/4.0*

PROKOSCH, J., BERNITZ, Z., BERNITZ, H., ERNI, B., AND ALTWEGG, R. (2019). Are animals shrinking due to climate change? Temperature-mediated selection on body mass in mountain wagtails. *Oecologia*, **189** (3), 841–849.

RAGGAD, B. (2018). Statistical assessment of changes in extreme maximum temperatures over Saudi Arabia, 1985-2014. *Theoretical and Applied Climatology*, **132** (3-4), 1217–1235.

RAGHAVENDRA, A., DAI, A., MILRAD, S. M., AND CLOUTIER-BISBEE, S. R.

(2019). Floridian heatwaves and extreme precipitation: Future climate projection. *Climate Dynamics*, **52** (1-2), 495–508.

RAGNO, E., AGHAKOUCHAK, A., CHENG, L., AND SADEGH, M. (2019). A generalised framework for process-informed non-stationary extreme value analysis. *Advances in Water Resources*, **130**, 270–282.

RASAKI, O. O., EZIKIEL, O., OLUMIDE, A. L., AND ADIGUN, O. A. (2018). On skew generalised extreme value-ARMA model: An application to average monthly temperature (1901-2016) in Nigeria. *International Journal of Statistics and Applied Mathematics*, **3** (5), 20–27.

REDDY, C. L. AND VINCENT, K. (2017). Climate risk and vulnerability, A Handbook for Southern Africa, Second edition, CSIR, Pretoria, South Afica.

REICH, B. J., SHABY, B. A., AND COOLEY, D. (2014). A hierarchical model for serially dependent extremes: A study of heat waves in the Western US. *Journal of Agricultural, Biological, and Environmental Statistics*, **19** (1), 119–135.

ROY, T. D. AND DAS, K. K. (2012). Time series analysis of Dibrugarh air temperature. *Journal of Atmospheric and Earth Environment*, **1** (1), 30–34.

ROY, T. D. AND DAS, K. K. (2013). Temperature trends at four stations of Assam during the period 1981-2010. *International Journal of Scientific and Research Publications*, **3** (6), 1–3.

RUSSO, S., MARCHESE, A. F., SILLMANN, J., AND IMMÃME, G. (2016). When will unusual heat waves become normal in a warming Africa? *Environmental Research Letters*, **11** (5), 054016.

SALMAN, S. A., SHADID, S., ISMAIL, T., CHUNG, E. S., AND AL-ABHADI, A. M. (2017). Long-term trends in daily temperature extremes in Iraq. *Atmospheric Research*, **198**, 97–107.

SAMPSON, T. A. AND KWADWO, N. A. (2019). Statistical modeling of temperature extremes behaviour in Ghana. *Journal of Mathematics and Statistics*, **14**, 275–284.

SENYOLO, M. P., LONG, T. B., BLOK, V., AND OMTA, O. (2018). How the characteristics of innovation impact their adption: An exploration of climate-smart agricultural innovations in South Africa. *Journal of Cleaner Production*, **172**, 3825–3840.

SIGAUKE, C. AND BERE, A. (2017). Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threhold: An application to peak electricity demand. *Energy*, **119**, 152–166.

SIGAUKE, C. AND NEMUKULA, M. M. (2018). Modelling extreme peak electricity demand during a heatwave period: A case study. *Energy Systems*, 1–23.

SINGH, C. S., ANIMA, K., KUMAR, B., GAUTAM, E., AND KUMARI, E. A. (2014). Environmental challenge due to climate change in Bihar, developing state of India. *Journal of Natural Sciences Research*, **4** (13), 21–28.

STRUPCZEWSKI, W. G., SINGH, V. P., AND FELUCH, W. (2001). Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *Journal of Hydrology*, **248** (1-4), 123–142.

SUKLA, M. K., MANGARAJ, A. K., AND SAHOO, L. N. (2014). An investigation on the stochastic modeling of daily rainfall amount in the Mahandi Delta region, India. *Research Journal of Mathematical and Statistics Sciences*, **2** (9), 1–8.

TARMIZI, A. H., HISHAMUDDIN, E., AND RAZAK, R. A. A. (2019). Impartial assessment of oil degradation through partitioning of polar compounds in vegetable oils under simulated frying practice of fast resturants. *Food Control*, **96**, 445–455.

TESKEY, R., WERTIN, T., BAUWERAERTS, I., AMEYE, M., MCGUIRE, M. A., AND STEPPE, K. (2015). Responses of tree species to heat waves and extreme heat events. *Plant, Cell and Environment*, **38** (9), 1699–1712.

THOMPSON, A. A., MATAMALE, L., AND KHARIDZA, S. D. (2012). Impact of climate change on children's health in Limpopo province, South Africa. *International Journal of Environmental Research and Public Health*, **9** (3), 831–854.

TOMCZYK, A. M., BEDNORZ, E., PÓLROLNICZAK, M., AND KOLENDOWICZ, L. (2019). Strong heat and cold waves in Poland in relation with the large-scale atmospheric circulation. *Theorectical and Applied Climatology*, **137** (3-4), 1909–1923.

TOROS, H., MOKARI, M., AND ABBASNIA, M. (2019). Regional variability of temperature extremes in the maritime climate of Turkey: A case study to develop agricultural adaptation strategies under climate change. *Modeling Earth System and Environment*, **5** (3), 857–865.

TSHIALA, M. F. AND OLWOCH, J. M. (2010). Impact of climate variability on tomato production in Limpopo province, South Africa. *Journal of Agricultural Research*, **5** (21), 2945–2951.

TSHIALA, M. F., OLWOCH, J. M., AND ENGELBRECHT, F. A. (2011). Analysis of temperature trends over Limpopo province, South Africa. *Journal of Geography and Geology*, **3** (1), 13.

VOGEL, C. AND OLIVIER, D. (2019). Re-imagining the potential of effective drought responses in South Africa. *Regional Environmental Change*, **19** (6), 1561–1570.

WINTER, H. (2016). *Extreme value modelling of heatwaves*. PhD Thesis, Lancaster University.

WOLF, J., ADGER, W. N., AND LORENZONI, I. (2010). Heat waves and cold spells: An analysis of policy response and perception of vulnerable population in the UK. *Environment and Planning A*, **42** (11), 2721–2734.

WORKU, G., TEFERI, E., BANTIDER, A., AND DILE, Y. T. (2019). Observed changes in extremes of daily rainfall and temperature in Jemma Sub-Basin, Upper Blue Nile Basin. *Theoretical and Applied Climatology*, **135** (3-4), 839–854.

WRIGHT, C. Y., GARLAND, R. M., NORVAL, M., AND VOGEL, C. (2014). Human health impacts in a changing South African climate. *South African Medical Journal*, **104** (8), 579–582.

YAMBA, F. D., WALIMWIPI, H., JAIN, S., ZHOU, P., CUAMBA, B., AND MZEZEWA, C. (2011). Climate change/variability implications on hydroelectricity generation in the Zambezi River Basin. *Mitigation and Adaptation Strategies for Global Change*, **16** (6), 617–628.

ZHOU, G. D., YI, T. H., CHEN, B., AND ZHANG, H. (2017). A generalised Pareto distribution based extreme value model of thermal gradients in a long-span bridge combining parameter updating. *Advances in Structural Engineering*, **20** (2), 202–213.

ZUO, J., PULLEN, S., PALMER, J., BENNETTS, H., CHILESHE, N., AND MA, T. (2015). Impacts of heat waves and corresponding measures: A review. *Journal of Cleaner Production*, **92**, 1–12.

Appendix

SOME SELECTED R CODES

Tempdata $<-$ read.delim("C:/Users/Lerato/Desktop/DATA SET/Tempdata.txt")
View(Tempdata)
data=c(TempdataJAN, Tempdata$FEB,Tempdata$MAR$, Tempdata$APR,TempdataMAY,
Tempdata$JUN,Tempdata$JUL$, Tempdata$AUG,TempdataSEP,
Tempdata$OCT, Tempdata$NOV$, Tempdata$DEC)
View(data)
summary(data)

win.graph()
par(mfrow = c(1,2))

R codes for fitting parent distributions

Fitting normal distribution
normMLE$<-$fitdist(data,"norm",method="mle")
normMLE
summary(normMLE)
plot(normMLE)

Fitting log-normal distribution

```
lnormMLE< −fitdist(data,“lnorm”,method=“mle”)
lnormMLE
summary(lnormMLE)
plot(lnormMLE)

# Fitting gamma distribution
gamMLE< −fitdist(data,“gamma”,method=“mle”)
gamMLE
summary(gamMLE)
plot(gamMLE)

# Fitting Weibull distribution
weiMLE< −fitdist(data,“weibull”,method=“mle”)
weiMLE
summary(weiMLE)
plot(weiMLE)
```

R codes for fitting EVT

```
# Installing packages for Extreme Value Analysis

library(VGAM)
library(extRemes)
library(ismev)
library(“evd”)
library(fitdistrplus)
library(nortest)
library(aTSA)
```

```
library(actuar)
library(EnvStats)

freq=gev.fit(AM)
gev.diag(freq)

# Goodness-of-fit test
library(EnvStats)
gofTest(AM,distribution="gev",test="ks")
gofTest(AM,distribution="gev",test="ad")

# Non-stationary GEV distribution
# Linear model for location parameter
ti=matrix(ncol=1,nrow=75)
ti[,1]=seq(1,75,1)
ti=gev.fit(AM,ydat=ti,mul=1,sigl = NULL)
gev.diag(ti)

# Linear model for location and scale parameters
ti=matrix(ncol=1,nrow=25)
ti[,1]=seq(1,25,1)
ti=gev.fit(AM,ydat=ti,mul=1,sigl = 1)
gev.diag(ti)

# A quadratic model in location parameter
ti2=matrix(ncol=2,nrow=75)
ti2[,1]=seq(1,75,1)
ti2[,2]=(ti2[,1])**2
gev.fit(AM,ydat=ti2,mul=c(1,2))
```

gev.diag(ti)

A quadratic model in location and linear scale parameter

ti2=matrix(ncol=2,nrow=25)

ti2[,1]=seq(1,25,1)

ti2[,2]=(ti2[,1])$* * 2$

gev.fit(AM,ydat=ti2,mul=c(1,2),sigl = 1)

A quadratic model in location and scale parameter

ti2=matrix(ncol=2,nrow=75) ti2[,1]=seq(1,75,1)

ti2[,2]=(ti2[,1])$* * 2$

gev.fit(AM,ydat=ti2,mul=c(1,2),sigl = c(1,2))

Calculating t-ratios and p-values (an example)

tb1=abs((0.004051329)/ 2.000028e-06); tb1

pt(tb1,25,lower.tail=FALSE)

Using POT approach to make inference # Plotting mean residual Life plot

mrp=mrl.plot(data)

Stability plots to validate u

win.graph()

par(mfrow=c(1,1))

threshrange.plot(data, r = c(20, 30), nint=10)

Declustering data

ei$< -$decluster(data,threshold =8)

ei

xcl=ts(ei,start=1994,frequency=12)

plot(ei,col="red",xlab="Time",ylab="Temperature",)

```
# Fitting point process approach
y < − pp.fit(data,8)
pp.diag(y)

# Fitting generalised Pareto distribution
y < − gpd.fit(data,8)
gpd.diag(y)
```

R codes for fitting bivariate

```
tempmaximum < − na.omit(tempmaximum) # removing missing values
attach(tempmaximum)
head(tempmaximum)

plot(Mara, type="l", ylab="Maximum temperature(deg C)",col="blue",lwd=2,
main="Mara maximum temperature",xlab="Observation number")
z < − smooth.spline(time(Mara), Mara)
z
lines(smooth.spline(time(Mara), Mara, spar= 0.5),col="red",lwd=2,lty=2,) #spar=
0.2002359
legend("topleft",col=c("black","red"), lty=1:2,lwd=2,
legend=c("Mara max temp", "Time varying threshold"))

Temp < − data.frame(cbind(excessMapos,excessMespos))
Temp < − na.omit(Temp)

#palette(c("black","purple","cyan","orange"))
summary(Temp,digits=2)
```

```
GGally::ggpairs(Temp)

# Examining pairwise extremal dependence using the
# Multivariate conditional Spearman's correlation coefficient
mcsMara-Mes <- MCS(Temp[, c("excessMapos", "excessMespos")])
mcsMara-Pol <- MCS(Temp[, c("excessMapos", "excessPolpos")])
mcsMara-Tha <- MCS(Temp[, c("excessMara", "excessThapos")])
mcsMes-Pol <- MCS(Temp[, c("excessMespos", "excessPolpos")])
mcsMes-Tha <- MCS(Temp[, c("excessMespos", "excessThapos")])
mcsPol-Tha <- MCS(Temp[, c("excessPolpos", "excessThapos")])

g1 <- ggplot(mcsMara-Mes, main="MCS: excessMapos and excessMespos")
gridExtra::grid.arrange(g1,ncol=1)
g2 <- ggplot(mcsMara-Pol, main="MCS: excessMapos and excessPolpos")
g3 <- ggplot(mcsMara-Tha, main="MCS: excessMapos and excessThapos")
gridExtra::grid.arrange(g1,g2,g3,ncol=1)

g4 <- ggplot(mcsMes-Pol, main="MCS: excessMespos and excessPolpos")
g5 <- ggplot(mcsMes-Tha, main="MCS: excessMespos and excessThapos")
g6 <- ggplot(mcsPol-Tha, main="MCS: excessPolpos and excessThapos")
gridExtra::grid.arrange(g4,g5,g6,ncol=1)

# Dependence model diagnostics
ggplot(mex.Temp)
mrf <- mexRangeFit(marg, "Mara", trace=11)
ggplot(mrf)
# Fitted model parameters
mex.Temp
```