Modelling Temperature in South Africa Using Extreme Value Theory

MASTERS DISSERTATION

Author:
Mr. Murendeni M Nemukula

Supervisor:
Dr. Caston Sigauke

Dissertation submitted for Masters of Science degree in Mathematical Statistics

in the

Faculty of Science,
School of Statistics and Actuarial Science,
University of the Witwatersrand
Johannesburg

January 23, 2018

Abstract

This dissertation focuses on demonstrating the use of extreme value theory in modelling temperature in South Africa. The purpose of modelling temperature is to investigate the frequency of occurrences of extremely low and extremely high temperatures and how they influence the demand of electricity over time. The data comprise a time series of average hourly temperatures that are collected by the South African Weather Service over the period $2000-2010$ and supplied by Eskom. The generalized extreme value distribution (GEVD) for r largest order statistics is fitted to the average maximum daily temperature (non-winter season) using the maximum likelihood estimation method and used to estimate extreme high temperatures which result in high demand of electricity due to use of cooling systems. The estimation of the shape parameter reveals evidence that the Weibull family of distributions is an appropriate fit to the data. A frequency analysis of extreme temperatures is carried out and the results show that most of the extreme temperatures are experienced during the months January, February, November and December of each year. The generalized Pareto distribution (GPD) is firstly used for modelling the average minimum daily temperatures for the period January 2000 to August 2010. A penalized regression cubic smoothing spline is used as a time varying threshold. We then extract excesses above the cubic regression smoothing spline and fit a non-parametric mixture model to get a sufficiently high threshold. The data exhibit evidence of short-range dependence and high seasonality which lead to the declustering of the excesses above the threshold and fit the GPD to cluster maxima. The estimate of the shape parameter shows that the Weibull family of distributions is appropriate in modelling the upper tail of the distribution. The stationary GPD and the piecewise linear regression models are used in modelling the influence of temperature above the reference point of $22°C$ on the demand of electricity. The stationary and non-stationary point process models are fitted and used in determining the frequency of occurrence of extremely high temperatures. The orthogonal and the reparameterization approaches of determining the frequency and intensity of extremes have been used to establish that, extremely hot days occur in frequencies of 21 and 16 days per annum, respectively. For the fact that temperature is established as a major driver of electricity demand, this dissertation is relevant to the system operators, planners and decision makers in Eskom and most of the utility and engineering companies. Our results are further useful to Eskom since it is during the non-winter period that they plan for maintenance of their power plants. Modelling temperature is important for the South African economy since electricity sector is considered as one of the most weather sensitive sectors of the economy. Over and above, the modelling approaches that are presented in this dissertation are relevant for modelling heat waves which impose several impacts on energy, economy and health of our citizens.

Keywords: Extreme value theory, generalized extreme value distribution, generalized Pareto distribution, maxima, order statistics, Poisson point process, temperature.

Declaration

I declare that the dissertation which is hereby submitted for the qualification of Master of Science in Mathematical Statistics at the University of the Witwatersrand, is my own independent original work and has not been handed in before for a qualification at/in another University/Faculty. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references. I further cede copyright of the dissertation to the University of the Witwatersrand.

Signature:.....................................

Date: January 23, 2018

Dedication

This work is heartily dedicated to my parents (Mr. Khathutshelo Adolfas Nemukula and Mrs. Mavhungu Eunice Nemukula) for all the efforts, support and the difficulties that they have been through for the sake of our survival and success.

Acknowledgements

Special thanks go to my supervisor, Dr Caston Sigauke whose knowledge and experience in statistics and research far much exceed mine. I thank him a lot for believing in me, introducing me to many research techniques, more especially extreme value theory, suggesting this project, advices, patience and hence a fruitful supervision. I am indebted for consistently sharing your knowledge, skills, ideas, guidance and criticisms that made me an individual who can author publishable manuscripts.

My gratitude goes to the University of Limpopo for their funds and resources. To Prof M Lesaoana, thank you for all the support that you have shown towards my academic progress. To my line manager, Prof A Tessera, thank you for the departmental support. The financial assistance of the National Research Foundation (NRF) of South Africa towards this dissertation is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily attributed to the NRF. This project would not have been a success without the data that is provided by Eskom which is hereby highly acknowledged.

The courageous words of a friend, Matsato Speedy Ramodipa led to the whole idea of furthering studies with Wits. I am grateful to Speedy for a fruitful friendship. This dissertation would not have been possible without the paramount assistance of Mr. Timotheus Brian Darikwa who gives support in all ways most of times. I am appreciative to Brian for sharing ideas, encouragements and motivation which made this endeavor a masterpiece. I would like to send my sincere appreciations to Prof Y.G Kifle for being there for me when I knew absolutely nothing about typesetting in TeX and programming in R language. To Mr M Matjuda, Dr EB Inyangala, Dr D Maposa, Mr A Boateng, Mr N Maswanganyi and all members of the Department of Statistics and Operations Research at UL, thank you colleagues for your patience and assistance when I used to stuck and knock at your doors times and again. To Mr R Minkah from Stellenbosch University, Ms S Khuluse-Makhanya from CSIR and Mr M Mokhele from Wits University, thank you guys for your contributions with EVT and computer packages.

Visiting Wits University would not have been easy without the kindness of my uncle, Mr Namadzavho B Kone. I am appreciative for your advices, encouragements and hospitality. Thank you once again for discovering at my earliest ages that, disadvantageous conditions such as albinism and short sight do not in anyhow, limit the understanding of Mathematical Sciences. To my parents and siblings, thank you very much for all the support, prayers, encouragements and love that you have shown throughout my Masters studies.

I cannot cease without thanking the almighty God for the gift of life, mercy and wisdom.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

ADED	Average Daily Electricity Demand
ADT	Average Daily Temperature
AHT	Average Hourly Temperature
AIC	Akaike Information Criterion
CI	Confidence interval
Eskom	Electricity Supply Commission
ETT	Extremal Types Theorem
EVA	Extreme Value Analysis
EVD	Extreme Value Distributions
EVI	Extreme Value Index
EVT	Extreme Value Theory
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GCM	Global Climate Models
GCV	Generalized Cross Validation
GEVD	Generalized Extreme Value Distribution
GLR	Generalized Likelihood Ratio
GPD	Generalized Pareto Distribution
i.i.d.	independent and identically distributed
L-MOM	L-Moments Method
MARS	Multivariate Adoptive Regression Splines
MDT	Maximum Daily Temperature
MLE	Maximum Likelihood Estimation
MME	Method of Moments Estimator
MW	Mega watts
P-P plot	Probability-Probability plot
POT	Peaks-Over-Threshold

Q-Q plot	Quantile-Quantile plot
SAWS	South African Weather Service

List of Notation and Special Symbols

$F_X(x)$	distribution function
$f_X(x)$	density function
$\ell(x_i, \theta)$	maximum likelihood function
f_t	cubic spline
M_n	block maxima
$\tilde{M}_n$	block minima
M_n^*	re-scaled block maxima
$\frac{1}{p}$	return period
log	natural logarithm
$G_\xi(x)$	GEVD function for block maxima
$\tilde{G}_\xi(x)$	GEVD function for block minima
$W_\xi(x)$	GPD function for threshold exceedances
$\tilde{\mu}$	location parameter of the GEVD for block minima
$\tilde{\sigma}$	scale parameter of the GEVD for block minima
x_p	quantile function
$x(T)$	$T-$year return level
μ	location parameter
μ_0	intercept coefficient of the location parameter
μ_1	slope coefficient of the location parameter
μ_2	quadratic/parabolic coefficient of the location parameter
σ	scale parameter
σ_0	intercept coefficient of the scale parameter
σ_1	slope coefficient of the scale parameter
σ_2	quadratic/parabolic coefficient of the scale parameter
α	shape parameter
$H(.\|.)$	non-parametric bulk model

$H(.)$	parametric bulk model
ξ	EVI for the generalized Pareto distribution (GPD)
ξ_0	intercept coefficient of EVI for GPD
τ	sufficiently high threshold
ϕ_τ	probability of exceeding the threshold
β	bulk parameter
θ	extremal index
λ	smoothing parameter
λ	intensity of a point process
$\pi(\theta)$	prior distribution
$\pi(\theta \mid x)$	posterior distribution
$\pi(x \mid \theta)$	likelihood function
θ	target parameter
$^\circ$C	degrees celsius

Research Outputs

A list of research outputs from this dissertation is given below.

Peer Reviewed Journal Publications

1. Nemukula, M.M. and Sigauke, C. (2018). A point process characterisation of extreme temperatures: An application to South African data. Environmental Modelling and Assessment. Under review.

2. Nemukula, M.M. and Sigauke, C. (2017). Modelling average maximum daily temperature using *r* largest order statistics: An application to South African data. JÀMBÁ: Journal of Disaster Risk Studies. Accepted for publication on 12 January 2018.

3. Nemukula, M.M. and Sigauke, C. (2015). Modelling average minimum daily temperature using extreme value theory with a time-varying threshold. South African Statistical Journal: Peer-reviewed Proceedings of the 57th Annual Conference of the South African Statistical Association, Congress 1, Vol. 57, pp. 57-64.

National Conference Presentations

1. Nemukula, M.M. and Sigauke, C. (2017). A point process characterisation of extreme temperatures: An application to South African data. 59th Annual Conference of the South African Statistical Association, 27 − 30 November 2017, University of the Free State, ILanga Estate, Bloemfontein, South Africa.

2. Nemukula, M.M. and Sigauke, C. (2016). Modelling average maximum daily temperature using *r* largest order statistics: An application to South African data. 58th Annual Conference of the South African Statistical Association, 28 November − 1 December 2016, University of Cape Town, The River Club, Cape Town, South Africa.

3. Nemukula, M.M. and Sigauke, C. (2015). Modelling average minimum daily temperature using extreme value theory with a time-varying threshold. 57th Annual Conference of the South African Statistical Association, 29 November $-$ 2 December 2015, University of Pretoria, South Africa.

International Conference Presentations

1. Nemukula, M.M. and Sigauke, C. (2017). Modelling average maximum daily temperature using *r* largest order statistics: An application to South African data. 61st International Statistical Institute - World Statistics Congress, 16 $-$ 21 July 2017, Marrakech, Morocco.

2. Nemukula, M.M. and Sigauke, C. (2017). Modelling average minimum daily temperature using extreme value theory with a time-varying threshold. 10th Extreme Value Analysis Conference, 26 $-$ 30 June 2017, Delft University of Technology (TU Delft), Netherlands.

3. Nemukula, M.M. and Sigauke, C. (2017). Modelling average minimum daily temperature using extreme value theory with a time-varying threshold. 4th African International Conference on Statistics, 20 $-$ 23 March 2017, The Ranch Resort, Polokwane, South Africa.

Chapter 1

Introduction

1.1 Introduction

Electricity usage in South Africa is increasing continuously with the increase in population size, technology, electrical appliances and economic growth (Hyndman and Fan, 2010; Muñoz, Sánchez-Úbeda, Cruz and Marín, 2010). The Electricity Supply Commission (Eskom), power utility company of South Africa is expected to increase supply of energy load to meet demand (Ferguson, Wilkinson and Hill, 2000; Eberhard, 2007; Inglesi, 2010, among others). Planners and decision makers in power utility companies like Eskom face uncertainties in demand of electricity. This is because of several factors that include weather conditions like temperature which is a major driver of electricity demand (Hyndman and Fan, 2010; Muñoz et al., 2010). It has apparently been established that, modelling temperature significantly contributes towards the level of reliability of outcomes in energy demand forecasting (Hyndman and Fan, 2010).

Most of typical statistical procedures are based on normal distributions and time series techniques. All of these assume symmetric distributions, which are not suitable for modelling tail behavior of fat (heavily) tailed and asymmetric distributions (Gencay and Selcuk, 2004; Wentzel and Maré, 2007). This problem is overcome by using Extreme Value Theory (EVT) distributions. However, the common limitation that is associated with using EVT models involves shortage of extreme data for statistical modelling (Coles, 2001). The problem that is addressed in this dissertation stems from the need to assess the extent to which EVT models can be used towards modelling temperature.

South Africa faced severe crisis of electricity supply from periodic blackouts and higher electricity prices for both industrial and domestic electricity users ever since 2008 (Inglesi, 2010; Strengers, 2012). This dissertation focuses on modelling temperature using EVT models with an objective of quantifying the effects of extreme low and extreme high temperature on electricity demand in South Africa.

1.2 Background

The frequently applied statistical approaches in estimating recurrence probabilities of exceptional occurrences are the methods that deal with exceptionally thin or heavy tailed functions that represent the concerned variables, leading to EVT that comprises a rich family of techniques that are important in several fields, including finance, hydrology and meteorology (Gencay and Selcuk, 2004; Wentzel and Maré, 2007). Castillo and Hadi (1997) and Gumbel (2012) consider the main idea of EVT as a study of the distribution of a largest or smallest phenomena, including daily average temperature in order to take informed decisions during improbable situations.

When the data consist of only maximum and minimum observations, the block maxima or block minima approach becomes relevant, and the appropriate modelling approach is a well known Generalized Extreme Value Distribution (GEVD) for modelling block maxima and block minima. If the complete time series of data is accessible, the Peaks-Over-Threshold (POT) approach becomes relevant, suggesting a Generalized Pareto Distribution (GPD) for modelling threshold exceedances. Time series observations are known to be dependent and as a result, do not satisfy the condition that, for the application of EVT models, we need independent and identically distributed (i.i.d.) observations. Meteorological data such as hourly average temperatures are known to be naturally grouped or clustered together, thereby exhibiting properties of short-range dependence and strong seasonality, which create a limitation that is addressed by implementing the declustering of extreme values (Smith, 1989).

The EVT approach that combines both stationary and non-stationary GPD and GEVD provides more detailed analysis of extremes (Coles, 2001; Beichelt, 2006) and as a result, a point process characterization of extremes is considered in this dissertation. Though stationarity is a truthful assumption for most of physical processes (Leadbetter, 1983), modelling extremes such as daily temperature that can be affected by seasonal effects that may vary through time invokes the need for non-stationary EVT models such as non-stationary GEVD and GPD.

In this dissertation, Extreme Value Analysis (EVA) framework is based on the following classification; with regards to modelling temperature:

1. For extreme temperatures that result from a maxima or minima of hourly average temperatures within a block of a given length, the GEVD for r largest ordered observations is used.

2. When considering the rest of hourly average temperatures, stationary and non-stationary GPD are used.

3. In case hourly temperatures violate the i.i.d. condition, declustering for dependent series is used.

4. A point process approach is used since it provides more detailed analysis compared to GEVD and GPD. This includes prediction of the frequency of occurrence of the coldest and hottest days.

This dissertation explores and interrogates these classifications in modelling temperature effects on electricity demand over time.

1.3 Purpose of the dissertation

1.3.1 Aim

Temperature is a major driver of electricity demand (Hyndman and Fan, 2010; Muñoz et al., 2010). This dissertation is aimed at presenting the extreme value modelling of temperature for winter and non-winter seasons to investigate their effects on electricity demand in the Republic of South Africa.

1.3.2 Objectives

The key objectives of this dissertation are to

1. assess frequencies of the occurrence of minimum and maximum temperatures in order to determine their influence on electricity demand,

2. fit stationary GEVD on blocks of $r = 1$ and of several values of r largest order maximum and minimum temperatures,

3. detect extreme temperatures over a sufficiently high threshold using stationary GPD, and determine whether or not there is a trend in threshold excesses,
4. investigate the effect of declustering and modelling temporal dependency on the estimated values of several return levels in exceedance process,
5. Model extreme temperature using point process models of extremes.

1.4 Scope of the dissertation

Frequency analyses of the occurrence of minimum and maximum hourly and daily temperatures are assessed using GEVD for r largest order statistics. Analyses based on stationary and non-stationary threshold models are done. Models for extremes of dependent sequences are discussed and fitted. A point process modelling approach is discussed and then applied in modelling extreme temperatures. The method of Maximum Likelihood Estimation (MLE) and the Bayesian framework are used for estimating the target parameters. Several graphical diagnostics and the deviance statistic are used in checking the fit of the models and choosing sufficiently high thresholds was carried out using the POT approach.

Most of the statistical analyses are performed using R, a software package that is useful as a language of programming and statistical analysis (R Core Team, 2013). The following extreme value packages in R are used: ismev by Heffernan and Stephenson (2014), evmix by Hu and Scarrott (2013), texmex by Southworth and Heffernan (2013b), fExtremes by Wuertz (2013), OpenBUGS by Andrew Thomas and Sturtz (2006), evdbayes by Stephenson and Ribatet (2014) and laeken by Alfons and Templ (2013). The motivation and advantages of using R as an appropriate programming language are discussed in Conway (2000). A thorough discussion about historical background and recent developments in the use of R packages and S-Plus code towards modelling extreme values is given in Stephenson and Gilleland (2005) and Gilleland et al. (2013).

Meteorological data that are modelled in this dissertation consist of a national time series of hourly average temperatures that are collected by the South African Weather Service (SAWS) over the period January 2000−August 2010 and supplied by Eskom. Detailed description of data is given in Chapter 3 and some parts of Chapter 4. This dissertation deals only with the average hourly temperature at a national level and as a result, modelling temperature together with its impact on electricity demand are not covered at provincial and lower categories. The limitations in this regard will be discussed in Section 5.5 of Chapter 5.

1.5 Significance of the dissertation

The electricity sector is established in several research articles as one of supreme weather-sensitive sectors in the economy of any country and therefore, precise modelling of electricity demand in the electricity sector is vital (Sigauke, 2014). Amongst weather variables that are significant in predicting demand of electricity load, temperature is established in Hyndman and Fan (2010) as the major one. The crucial role that temperature plays on the demand of electricity load is based on the fact that, heating systems are used in winter to keep warm, whereas air conditioning appliances are desired in summer to keep cool (Muñoz et al., 2010).

The significance of this dissertation is based on presenting the extent to which EVT models can be used towards the modelling of extreme temperature in South Africa. Amongst others, the significance of this dissertation is recognized in terms of investigating the effects of coldest days as well as effects of hottest days on the demand of electricity load in South Africa over time. The results of this dissertation are important to Eskom for scheduling and dispatching of electrical energy. The fact that the demand of electricity is highly sensitive to high temperature is mentioned by several authors in literature. Occurrence of extreme heat impose several impacts which affect among other, health, transportation, energy, agriculture and economy (Meehl and Tebaldi, 2004; Lyon, 2009; Steffen, Hughes and Perkins, 2014). The modelling approaches in this dissertation are important for modelling heat waves.

1.6 Structure of the dissertation

The rest of the dissertation is structured as follows:

Chapter 2 focuses on reviewing theoretical and mathematical backgrounds of EVT, together with its applicability to various suitable areas. Books, published articles and technical reports that deal with the meteorological studies such as modelling temperature, extreme rainfall and floods frequency analyses are summarized.

The detailed procedures of EVT that are used in this dissertation are discussed in Chapter 3. This includes description of data and discussion of the procedures of data analyses. Chapter 4 presents the analyses of data. In Section 4.1, the GEVD for r largest $(r \geq 1)$ order statistics is fitted in order to assess the behavior of the minima and maxima temperature series. The use of the GPD is presented in Section 4.2, where the GPD is used for modelling the cluster maxima of the declustered temperature series and the effects of the hottest days on the average daily electricity demand are assessed. Lastly, the usage of point process analysis is demonstrated under Section 4.3. The frequency of occurrence of extreme high temperatures is calculated and interpreted in Section 4.3.3.

Chapter 5 finalizes the dissertation by summarizing and presenting concluding remarks that are based on the results of Chapter 4. The key findings, contributions and limitations of the dissertation are discussed. Chapter 5 further gives some recommendations and then suggests fields for further studies.

Chapter 2

Literature Survey

2.1 Introduction

The focus of Chapter 2 is on literature review. Stationary and non-stationary GEVD for homogeneous sequences and time-varying parameters respectively, are discussed for *r* largest order statistics. The duality principle is discussed for the asymptotic model for block minima. Stationary and non-stationary GPD are discussed as alternative approaches for modelling extremes. Threshold selection techniques that are discussed are the Pareto quantile, mean excess and threshold stability plots, including extremal mixture models. Declustering is discussed as an essential technique for dealing with stationary but dependent series. The Bayesian inference framework and the MLE technique are discussed for estimating the target parameters. Graphical diagnostic tools are discussed as important criteria for assessing goodness of fit of the models to the data. The discussion about the choice of the best out of several candidate models is based on the deviance statistic. The point process concepts including stationarity, intensity and the Poisson point process with its connection to EVT are discussed together with the point process inference.

2.2 Generalized extreme value distribution for block maxima and block minima

2.2.1 Classical extreme value theory

The GEVD is established in literature as a most suitable approximation for modelling maxima or minima of a long sequence of finite variables. The extreme value distributions that are constituted by the GEVD are justified by Davison and Smith (1990) as the stable limiting distributions of EVT. Suppose that the i.i.d. finite sequence $X_1,\ldots,X_n$ constitutes a simple random sample that is chosen from the variable X that follows a marginal distribution function F. By letting $M_n = \max\{X_1,\ldots,X_n\}$ for the suitable normalizing constants $\{a_n > 0\}$ and $\{b_n\}$, the distribution function of re-scaled block maxima M_n^* is given by:

$$Pr\left\{\frac{(M_n - b_n)}{a_n} \leq x\right\} = F^n(a_n x + b_n) \rightarrow G(x) \quad \text{as} \quad n \rightarrow \infty \tag{2.1}$$

for all $x \in \mathrm{I\!R}$ and a non-degenerate distribution function G. According to Gnedenko (1943), for an appropriate selection of $\{a_n > 0\}$ and $\{b_n\}$, G(x) converges towards one of three Extreme Value Distribution (EVD) families, namely Gumbel, Fréchet and Weibull, depending on the shape parameter α as shown in equation (2.2):

$$G(x) = \begin{cases} \exp\left\{-\exp\left[-\left(\frac{x-b}{a}\right)\right]\right\} & \text{if } -\infty < x < \infty, \quad a > 0 \\ \exp\left\{-\left(\frac{x-b}{a}\right)^{-\alpha}\right\} & \text{if } x > b, \quad a > 0 \\ \exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{\alpha}\right]\right\} & \text{if } x < b, \quad a > 0. \end{cases} \tag{2.2}$$

The classical modelling of extremes was previously based on equation (2.2), whereby the parameters could be estimated and then choose the appropriate model for the data (Millington, Das and Simonovic, 2011). As a result of several challenges including uncertainty associated with the choice of one out of three EVD models, and also because of the inferential weaknesses (Coles, 2001), Fisher and Tippett (1928) obtained limiting forms for the distribution functions of suitably normalised maxima, which is an early contribution to classical EVT. As discussed in Jenkinson (1955), the three EVD families are unified into the GEVD given by:

$$G_\xi(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{2.3}$$

valid for $\left\{x : \mu - \frac{\sigma}{\xi} < x < \infty\right\}$, where $-\infty < \mu < \infty$ is a location parameter, $\sigma > 0$ is a scale parameter and $-\infty < \xi < \infty$ is a shape parameter.

Ordinary GEVD in equation (2.3) converges to one of three EVD families, depending on the rate of decay of the tail that is indexed by ξ, which denotes the Extreme Value Index (EVI) of the GEVD. When $\xi = 0$, $G_\xi(x)$ reduces to a type I or a short-tailed unbounded

Gumbel family of distributions which is defined as a limit of equation (2.3) as $\xi \to 0$ (Coles, 2001; Beirlant, Goegebeur, Segers and Teugels, 2004; De Haan and Ferreira, 2007; Reiss, Thomas and Reiss, 2007, among others)

$$\lim_{\xi \to 0} \left[\exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \right] = \exp \left[- \exp \left\{ - \left(\frac{x-\mu}{\sigma} \right) \right\} \right], -\infty < x < \infty. \tag{2.4}$$

If $\xi > 0$, then $G_\xi(x)$ belongs to a type II family which is a heavy-tailed Fréchet class of distributions that is bounded below by $\mu - \frac{\sigma}{\xi}$ (Coles, 2001; Beirlant et al., 2004; De Haan and Ferreira, 2007; Reiss et al., 2007). When $\xi < 0$, $G_\xi(x)$ is thin-tailed and we get a type III family which is the Weibull class of distributions with an upper bound given by $\mu - \frac{\sigma}{\xi}$ (Coles, 2001; Beirlant et al., 2004; De Haan and Ferreira, 2007; Reiss et al., 2007). The survival distribution of the GEVD in equation (2.3) is

$$Pr(X > x) = 1 - G_\xi(x) = 1 - \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \tag{2.5}$$

defined for $\left\{ x : 1 + \xi \left(\frac{x-\mu}{\sigma} \right) > 0 \right\}$ and $\xi \neq 0$.

By letting $p = Pr(X > x)$ and rearranging equation (2.5), we get the following quantile function:

$$x_p = \mu + \frac{\sigma}{\xi} \left\{ \left[-\ln(1-p)^{-\xi} \right] - 1 \right\}, \quad \xi \neq 0, \tag{2.6}$$

where $G(x_p) = 1 - p$ is a return level related to return period $1/p$ such that x_p is a level that is anticipated to be exceeded once on average in $1/p$ years. As $p \to 0$ and $\xi < 0$, $x_p = \mu - \frac{\sigma}{\xi}$. The quantile function that is specified in equation (2.6) is applied in estimating extreme quantiles and calculating probability of exceedance levels together with $m-$year return levels, where $m = 1/p$.

2.2.2 Generalized extreme value distribution for r largest order statistics

Block maxima model in equation (2.3) is extended to give the GEVD (for $r \geq 1$) within annual blocks, for fixed values of r. We initially define $M_n^j =$ jth largest of $\{X_1, X_2, \ldots, X_n\}$ followed by recognizing limit behavior of the variable, for fixed j, as $n \to \infty$ (Weissman, 1978; Smith, 1986; Coles, 2001; An and Pandey, 2007, among others). If equation (2.1) is fulfilled, then, for fixed j, it follows that

$$Pr\left\{\frac{(M_n^{(j)}-b_n)}{a_n}\leq x\right\}\to G_j(x) \quad \text{as} \quad n\to\infty,$$

defined for $\left\{x:1+\xi\left(\frac{x-\mu}{\sigma}\right)>0\right\}$, where

$$G_j(x)=\exp\{-\tau(x)\}\sum_{s=0}^{j-1}\frac{\tau(x)^s}{s!}, \tag{2.7}$$

with

$$\tau(x)=\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}.$$

The GEVD for r largest order statistics

$$M_n^{(r)}=\left(\frac{M_n^{(1)}-b_n}{a_n},\frac{M_n^{(2)}-b_n}{a_n},\ldots,\frac{M_n^{(r-1)}-b_n}{a_n},\frac{M_n^{(r)}-b_n}{a_n}\right),$$

is the joint probability density function as given in David and Nagaraja (1981), Coles (2001), Soares and Scotto (2004) and Reiss (2012):

$$\begin{aligned} f\left(x^{(1)},\ldots,x^{(r)}\right) &= \exp\left\{-\left[1+\xi\left(\frac{x^{(r)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ &\times \prod_{j=1}^{r}\frac{1}{\sigma}\left[1+\xi\left(\frac{x^{(j)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1,} \end{aligned} \tag{2.8}$$

valid for $-\infty<\mu<\infty,\sigma>0$ and $-\infty<\xi<\infty$; $x^{(r)}\leq x^{(r-1)}\leq\cdots\leq x^{(1)}$; and $x^{(j)}$: $1+\frac{\xi\left(x^{(j)}-\mu\right)}{\sigma}>0$ for $j=1,2,\ldots,r$.

A likelihood function for the r largest order statistics model when $\xi \neq 0$ is given by

$$\begin{aligned} L(\mu,\sigma,\xi) &= \prod_{i=1}^{m}\left(\exp\left\{-\left[1+\xi\left(\frac{x_i^{(r_i)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right. \\ &\quad \times \left.\prod_{j=1}^{r_i}\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i^{(j)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\right) \qquad (2.9) \end{aligned}$$

defined for $1+\xi\left(\frac{x_i^{(j)}-\mu}{\sigma}\right) > 0$, $j = 1,\ldots,r_i$, $i = 1,\ldots,m$.

2.2.3 Asymptotic model for block minima

In Section 2.2, the focus is on GEVD that is suitable for modelling block maxima of any particular sequence. In case there is a desire to consider modelling block minima, the duality principle is used to transform GEVD for block maxima to that of block minima (Coles, 2001; Gilleland and Katz, 2011; Chikobvu and Sigauke, 2013, among others). Suppose that $\tilde{M}_n = \min\{X_1,\ldots,X_n\}$ is the minima of i.i.d. sequence $X_1,\ldots,X_n$. Letting $Y_1,\ldots,Y_n = -X_1,\ldots,-X_n$, where the $-$ sign implies minimum average daily temperature $X_1,\ldots,X_n$ corresponding to large values of $Y_1,\ldots,Y_n$ such that $\tilde{M}_n = -M_n$, the GEVD for block minima is as follows:

$$\begin{aligned} Pr\left\{\tilde{M}_n \leq x\right\} &= Pr\left\{-M_n \leq x\right\} \\ &= Pr\left\{M_n \geq -x\right\} \\ &= 1 - Pr\left\{M_n \leq -x\right\} \\ &= 1-\exp\left\{-\left[1+\xi\left(\frac{-x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ \Rightarrow \tilde{G}_{\xi}(x) &= 1-\exp\left\{-\left[1-\xi\left(\frac{x-\tilde{\mu}}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \qquad (2.10) \end{aligned}$$

valid for $\left\{x : 1-\frac{\xi(x-\tilde{\mu})}{\sigma} > 0\right\}$, where $\tilde{\mu} = -\mu$ such that $-\infty < \mu < \infty$, $\tilde{\sigma} > 0$ and $-\infty < \xi < \infty$ (Coles, 2001).

2.3 Generalized Pareto distribution for threshold exceedances

The GPD is originally pioneered by Balkema and de Haan (1974), then formally introduced by Pickands III (1975) as an appropriate asymptotic model for modelling stochastic behavior of residuals above the threshold. In this dissertation, average daily temperatures are considered as extreme values provided they exceed a suitably high threshold τ that is chosen using several threshold selection tools. Smith (1989) and Coles (2001) consider the POT as a better alternative analysis of extremes compared to the block maxima or block minima approach due to the capability of the POT approach to use as much as possible of available information.

Suppose that $X_1, X_2, \dots, X_n$ denote a stationary process that is distributed as mentioned in Section 2.2 provided equation (2.3) is fulfilled. The interest is on stochastic behavior of threshold excesses that is described by the conditional distribution that is according to Balkema and de Haan (1974).

$$F_\tau(y) = P(X \leq \tau + y | X > \tau) \quad \Rightarrow F_\tau(y) = \frac{F(\tau + y) - F(\tau)}{1 - F(\tau)},$$

where $0 \leq y < \tau_F - \tau$ and $\tau_F = \inf\{x : F(x) = 1\} \leq \infty$ is the upper tail of $F(x)$. If equation (2.3) holds for sufficiently high threshold τ, the CDF of $(X - \tau)$, conditioned on $X > \tau$ is a GPD that is approximated by:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{2.11}$$

(Balkema and de Haan, 1974) provided $\{y | y > 0, (1 + \xi y / \tilde{\sigma}) > 0\}$, and $\tilde{\sigma} = \sigma + \xi(\tau - \mu)$. Pickands III (1975) formally introduced the unified stationary GPD and denoted it by W_ξ given as follows:

$$W_\xi(x, \sigma_\tau) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\tau)}{\sigma_\tau}\right)^{-\frac{1}{\xi}} & \text{if } \xi > 0, x > \tau \\ 1 - \exp\left(-\frac{x-\tau}{\sigma_\tau}\right) & \text{if } \xi = 0, x > \tau \\ 1 - \left(1 + \frac{\xi(x-\tau)}{\sigma_\tau}\right)^{-\frac{1}{\xi}} & \text{if } \xi < 0, \tau < x < \tau - \frac{\sigma_\tau}{\xi}, \end{cases} \tag{2.12}$$

where $-\infty < \xi < \infty$ is the shape parameter that expresses tail behavior of the distribution and $\sigma_\tau > 0$ is the scale parameter which characterizes spread of the distribution such that, $\sigma_\tau > 0$ is reparameterized as $\theta = \log(\sigma_\tau)$, and τ is the threshold that is exceeded by excesses (Beirlant et al., 2004).

The GPD as shown in equation (2.12) results in one of three Pareto distribution families, depending on rate of decay of the tail that is indexed by ξ. When $\xi = 0$, $W_{\xi}(x)$ reduces to a type I or a short-tailed exponential family of distributions with parameter $1/\tilde{\sigma}_{\tau}$ (Coles, 2001; Beirlant et al., 2004; De Haan and Ferreira, 2007; Reiss et al., 2007). The resulting exponential class is unbounded and defined as a limit of equation (2.11) as $\xi \to 0$

$$\lim_{\xi \to 0}\left[1-\left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}\right] = 1-\exp\left(1-\frac{y}{\tilde{\sigma}}\right), \quad y > 0. \tag{2.13}$$

When $\xi > 0$, we get a heavy-tailed type II which is a Pareto class of distributions (Coles, 2001; Beirlant et al., 2004; De Haan and Ferreira, 2007; Reiss et al., 2007). When $\xi < 0$, $W_{\xi}(x)$ approaches a Type III class of GPD that is thin tailed and bounded above by $\tau - \frac{\sigma_{\tau}}{\xi}$ (Coles, 2001; Beirlant et al., 2004; De Haan and Ferreira, 2007; Reiss et al., 2007). An unconditional distribution of variable X is desired when one needs to calculate return levels of the GPD. Suppose that $\phi_{\tau} = Pr(X > \tau)$ and considering the conditional distribution that is such that

$$Pr\{X > x|X > \tau\} = \left(1+\xi\frac{(x-\tau)}{\sigma}\right)^{-\frac{1}{\xi}},$$

we obtain

$$Pr\{X > x\} = \phi_{\tau}\left(1+\xi\frac{(x-\tau)}{\sigma}\right)^{-\frac{1}{\xi}}.$$

The quantile function x_m is obtained as a solution to the following equation:

$$\frac{1}{m} = \phi_{\tau}\left(1+\xi\frac{(x_m-\tau)}{\sigma}\right)^{-\frac{1}{\xi}}$$

which is given by

$$x_m = \tau + \frac{\sigma_{\tau}}{\xi}\left[(m\phi_{\tau})^{\xi}-1\right], \tag{2.14}$$

valid for the values of m such that $x_m > \tau$ and $\xi \neq 0$. This quantile function is used in this dissertation for calculating probabilities of threshold exceedance and the $m-$year return levels. When $\xi = 0$, equation (2.14) is given by $x_m = \tau + \sigma_{\tau}\log(m\phi_{\tau})$, provided m is sufficiently large (Pickands III, 1975; Coles, 2001; Beirlant et al., 2004; Mallor, 2009).

2.4 Threshold selection

The analysis of extremes based on the POT approach is valid provided the threshold above which observations are extreme values is neither too high nor too low. When the threshold is too high, there are few positive excesses above the threshold and hence a large variance. Looking on the other side, a low value of the threshold clues to the destruction of the asymptotic feature of the GPD, implying bias (Pickands III, 1994; Castillo and Hadi, 1997; Coles, 2001, among others). To this effect, the main requirement of the threshold is to be sufficiently high for the purpose of maintaining a balance between bias and variance (Coles, 2001; Smith, 2003; Sugahara et al., 2009). Among several threshold selection tools that are proposed in literature, this section discusses few that are frequently used by most researchers.

Suppose that $x_1,\ldots,x_n$ denote i.i.d. average hourly temperature values and let τ be a sufficiently high threshold. The values $x_{(1)} \leq \cdots \leq x_{(k)}$ are k positive ordered residuals above τ if $\{x_i : x_i > \tau\}$ and the observations $y_j = x_{(j)} - \tau$, for $j = 1,\ldots,k$ are threshold excesses (Pickands III, 1994; Coles, 2001; Beirlant et al., 2004).

2.4.1 Pareto quantile plot

Beirlant, Vynckier and Teugels (1996) view the Pareto quantile plot as the basis for determining a sufficiently high threshold over which the observations are extremes. Suppose $X_1,\ldots,X_n$ is a sequence of continuous random variables. A graphical tool with coordinates $(-\log(1-p_i), \log x_i), \forall i = 1,\ldots,n$ that is plotted for several values of $p_i \in (0,1)$, where $p_i = \frac{i}{n+1}$, is referred to as a Pareto quantile plot. The high threshold is realized at logarithmic observations (vertical axis) on linear pattern of the plot (Beirlant et al., 2004).

2.4.2 Mean excess plot

A function of mean excesses with a finite expectation $E(X) < \infty$ is discussed in Beirlant et al. (2004) as

$$e(t) = E(X - t \mid X > t),$$

where estimate for mean excess function is given by:

$$\hat{e}_n(t) = \frac{\sum_{i=1}^{n} x_i 1_{(t,\infty)}(x_i)}{\sum_{i=1}^{n} 1_{(t,\infty)}(x_i)} - t. \tag{2.15}$$

The indicator function $1_{(t,\infty)}(x_i)$ is defined such that

$$1_{(t,\infty)}(x_i) = \begin{cases} 1 & \text{if } x_i > t, \\ 0 & \text{otherwise }. \end{cases}$$

The mean excess plot is a locus of the points $t = x_{n-k,n}, k = 1, \ldots, n-1$ that are plotted alongside the average residuals $e_{k,n}$, where average residuals are estimated as (Beirlant et al., 2004):

$$e_{k,n} := \hat{e}_n(x_{n-k,n}) = \frac{1}{k} \sum_{j=1}^{k} x_n - j + 1, n - x_{n-k,n}. \tag{2.16}$$

A sufficiently high threshold is chosen as the graph begins to display linearity (Beirlant et al., 2004).

2.4.3 Threshold stability plot

"The GPD models exhibit threshold stability properties" (Hu and Scarrott, 2013, Page 3). Due to this, it has been established to be recommendable, to plot $\hat{\sigma}$ and $\hat{\xi}$ against various threshold values τ_i, $i = 1, 2, \ldots, n$, resulting in a threshold stability plot. This plot is considered in this dissertation as one of the threshold selection techniques because of its several advantages. The threshold stability plot provides the basis for assessing a range of several thresholds for the invariance in extremal index estimator. It further provides the criteria for choosing the least threshold over which the estimates of extremal index are constant (Heffernan and Southworth, 2013).

2.4.4 The use of extreme value mixture models

Threshold selection tools that have been discussed so far were established in literature prior to the existence of extreme value mixture models as threshold selection criteria. The idea of extremal mixture model is presently considered a recent tool that constitutes an important basis for choosing a sufficiently high time-varying threshold (Scarrott and MacDonald, 2012). The main idea of using extremal mixture models as emphasized in Wu, Huang, Long and Peng (2007), is to avoid consequences of fixed threshold approach such as ignorance of uncertainties and inability to compare GPD parameters of an entire model at different thresholds (Hu and Scarrott, 2013; Bommier, 2014). Before fitting the extremal mixture models, the data are initially detrended using a penalized regression cubic smoothing spline given in equation (2.17),

$$\sum_{t=1}^{n}(x_t - f_t)^2 + \lambda \int \left(f_t''\right)^2 \mathrm{d}t \tag{2.17}$$

where x_t is the temperature, f_t is a cubic spline and λ is a smoothing parameter that is chosen based on the Generalized Cross Validation (GCV) approach (Diggle, 1985; Friedman, 1991; Shumway and Stoffer, 2011; Wang, 2011, among others). Both the parametric and non-parametric versions of extreme value mixture models are considered in this dissertation. In the parametric version, extreme value mixture models that are continuous for the threshold are considered for the bulk component as a threshold selection criterion. According to Hu and Scarrott (2013), a simple parametric extreme value mixture modelling is initially established by Behrens, Lopes and Gamerman (2004) as a result of fitting the parametric bulk model below and the GPD above the sufficiently high threshold. This implementation is found to be valid for gamma, Gaussian and Weibull classes of distributions. In this dissertation, the truncated Weibull distribution is fitted to the bulk model and GPD fitted to upper tail of distribution. The cumulative distribution function (cdf) of a bulk based tail fraction model is given by Hu and Scarrott (2013) as:

$$F(x|\tau,\beta,\sigma_\tau,\xi,\phi_\tau) = \begin{cases} H(x|\beta), & x \leq \tau, \\ H(\tau|\beta) + (1 - H(\tau|\beta))(x|\tau,\sigma_\tau,\xi), & x > \tau, \end{cases} \tag{2.18}$$

where $H(.)$ is the cdf of the bulk model with the bulk parameter β. The tail fraction $\phi_\tau = 1 - H(\tau|\beta)$ is expressed as a survival distribution of equation (2.18) assuming its continuity beyond the threshold. According to MacDonald, Scarrott, Lee, Darlow, Reale and Russell (2011), the usual consequences associated to bulk modelling is grounded on misspecification at the end point of the distribution, which is generally accounted for by jointly considering the mixture model which represents the bulk based and the parameterized based tail fraction models. The cdf of the parameterized based tail fraction is as follows:

$$F(x|\tau,\beta,\sigma_\tau,\xi,\phi_\tau) = \begin{cases} (1-\phi_\tau) \times \frac{H(x|\beta)}{H(\tau|\beta)}, & x \leq \tau, \\ (1-\phi_\tau) + \phi_\tau \times G(x|\tau,\sigma_\tau,\xi), & x > \tau, \end{cases} \tag{2.19}$$

where ϕ_τ is the proportion of observations that exceed the threshold such that $0 < \phi_\tau < 1$, $H(.|\beta)$ is a (truncated) Weibull distribution and $G(.|\tau,\sigma_\tau,\xi)$ is the GPD.

The non-parametric version of extremal mixture models was pioneered by Tancredi, Anderson and OHagan (2006) and formally developed by MacDonald, Scarrott and Lee (2013). The procedure is to fit a fixed threshold to positive residuals (excesses) above the time-varying threshold (Hu and Scarrott, 2013), and then fit a kernel density to the bulk model, followed by fitting a GPD to the tails of probability distributions of minima and maxima (Wang, 2011; Hu and Scarrott, 2013). The cdf of extremal mixture model as discussed in MacDonald et al. (2013) is as follows:

$$F(x|\beta,\tau,\sigma_\tau,\xi,\phi_\tau) = \begin{cases} H(x|\beta), & \text{if } x \leq \tau, \\ H(\tau|\beta) + \phi_\tau G(x|\tau,\sigma_\tau,\xi), & \text{if } x > \tau, \end{cases} \tag{2.20}$$

where the bulk model is denoted by $H(.|.)$ and the bulk parameter is given by β. A kernel density is fitted to the bulk model and a GPD to the right end-point. Parameters of the mixture models are determined in this dissertation using MLE technique.

2.5 Extremes of dependent series

According to Coles, Heffernan and Tawn (1999), Coles (2001) and Castillo (2012), stationary series $X_1,\ldots,X_n$ meets the requirements of the $D(u_n)$ condition (Leadbetter, 1983) if, for all $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$\begin{aligned} &|Pr\{X_{i1} \leq u_n,\ldots,X_{ip} \leq u_n, X_{j1} \leq u_n,\ldots,X_{jq} \leq u_n\} \\ -&Pr\{X_{i1} \leq u_n,\ldots,X_{ip} \leq u_n\}Pr\{X_{j1} \leq u_n,\ldots,X_{jq} \leq u_n\}| \leq \alpha(n,l), \end{aligned} \tag{2.21}$$

where $\alpha(n,l_n) \to 0$ for the sequence l_n in a manner that $l_n/n \to 0$ as $n \to \infty$.

The measure for the properties of dependence, given as $u_n = a_n x + b_n$ ensures that the series does not have influence on the result of the limit, implying that if the extreme series exhibits features of long-range dependence, then the distribution laws of limits in series that are independent are similar to those of stationary process such that the parameters of the limit distribution violate the i.i.d. condition (Coles, 2001; Castillo, 2012). For the purpose of examining the impacts of dependence on the series of limit distribution parameters, suppose that $X_1,\ldots$ denote a process that is stationary and distributed with F and suppose further that, $X_1^*,\ldots$ denote variables that constitute an independent sequence that is distributed with F. The comparison of $M_n = \max\{X_1,\ldots,X_n\}$ and $M_n^* = \max\{X_1^*,\ldots,X_n^*\}$ is as follows:

$$Pr\left\{\frac{(M_n^* - b_n)}{a_n} \leq x\right\} \rightarrow G_1(x) \tag{2.22}$$

as $n \rightarrow \infty$ for the sequential constants $\{a_n > 0\}$ and $\{b_n\}$, where G_1 is a marginal distribution which does not degenerate, provided

$$Pr\left\{\frac{(M_n - b_n)}{a_n} \leq x\right\} \rightarrow G_2(x), \tag{2.23}$$

where

$$G_2(x) = G_1^{\theta}(x) \tag{2.24}$$

for an extremal index θ such that $0 < \theta \leq 1$ (Coles, 2001).

The distribution of $G_1^{\theta}(x)$ is a GEVD with parameters μ^*, σ^* and ξ given as follows:

$$\begin{aligned} G_1^{\theta}(x) &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}^{\theta} \\ &= \exp\left\{-\theta\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu^*}{\sigma^*}\right)\right]^{-\frac{1}{\xi}}\right\} \end{aligned} \tag{2.25}$$

where

$$\mu^* = \mu - \frac{\sigma}{\xi}\left(1 - \frac{\xi}{\theta}\right) \quad \text{and} \quad \sigma^* = \sigma\theta^{\xi}.$$

An alternative interpretation of θ is the clustering at extreme level (Coles, 2001). This is emphasized in Leadbetter (1983) as:

$$\theta = \frac{1}{\text{(limiting mean cluster size)}}, \tag{2.26}$$

where the limit extends over clustering exceedances of sufficiently high thresholds such that the extremal index $\theta = 1$ for independent series.

2.5.1 Modelling stationary series

"For a block maxima approach, the $D(u_n)$ condition is fulfilled if the data exhibit strong evidence of weak dependence at extreme levels such that, the block maxima is distributed with the same family of distributions as for independent series" (Coles, 2001, p. 98). The declustering procedure that is used in this dissertation is discussed in Section 3.4 of Chapter 3. However, the declustering approach is characterized by limitations which include among others, the condition that "the outcomes could easily be affected by the manner in which the clusters are determined as well as the possibility of wasting data while discarding data from the cluster block maxima" (Coles, 2001, p. 99).

In addition to that, Ferro and Segers (2003) explain that the manner in which the declustering procedure is chosen significantly affects the estimated values of the cluster parameters, creating a problem that is overcome by switching to a declustering procedure that is not manually performed in a way that, the choice of the parameters is not subjective. The analysis of extremes of stationary processes is carried out in this dissertation using the automatic interval estimator method of Ferro and Segers (2003) that is developed by Southworth and Heffernan (2013a) in texmex package.

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

Suppose $X_1,\ldots,X_n$ constitute average temperatures that are distributed as in equation (2.3), then X_t, where t is the annual maximum temperature in the t^{th} year follows GEVD$(\mu(t),\sigma,\xi)$ where

$$\mu(t) = \mu_0 + \mu_1 t, \tag{2.27}$$

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 (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007). A quadratic non-stationarity in μ of the GEVD is given by the function (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007)

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2. \tag{2.28}$$

To model exponential variation in the scale parameter σ, the function

$$\sigma(t) = \exp(\sigma_0 + \sigma_1 t) \tag{2.29}$$

is used, where the exponential transformation maintains non-negativity of the scale (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007).

Suppose that $X_1, \dots, X_n$ are average daily temperatures that follow a GPD, then let the season that contains day t be denoted by $s(t)$ and $\tau_{s(t)}$ be the threshold for particular seasons, then X_t follows GPD(σ, ξ) with the conditional seasonal model given by:

$$(X_t - \tau_{s(t)} \mid X_t > \tau_{s(t)}) \sim \text{GPD}(\sigma_{s(t)}, \xi_{s(t)}), \tag{2.30}$$

where $(\sigma_{s(t)}, \xi_{s(t)})$ are the GPD parameters in season s(t) (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007). In case the average daily temperature is related to other variables such as time, the covariates are included in the non-stationary models. Suppose that $X_t \sim \text{GEVD}(\mu(t), \sigma(t), \xi(t))$. The following is a maximized log $-$likelihood function of a non-stationary GEVD per equation (2.3):

$$\begin{aligned} \ell(\mu, \sigma, \xi) &= -\sum_{t=1}^{m} \left\{ \log \sigma(t) + \left(1 + \frac{1}{\xi(t)}\right) \log \left[1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right)\right] \right. \\ &+ \left. \left[1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right)\right]^{-\frac{1}{\xi(t)}} \right\}, \end{aligned} \tag{2.31}$$

where

$$1 + \xi(t) \left(\frac{x_t - \mu(t)}{\sigma(t)}\right) > 0, \text{ for } \quad t = 1, 2, \dots, m.$$

For the purpose of modelling the impacts of seasonality in location parameter with T as temperature, the following function is used:

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 \sin\left(\frac{2\pi t}{T}\right) + \mu_3 \cos\left(\frac{2\pi t}{T}\right) \tag{2.32}$$

(Katz, Parlange and Naveau, 2002; Khaliq, Ouarda, Ondo, Gachon and Bobée, 2006). Seasonal effects in the scale parameter σ are modelled using the following function:

$$\log \sigma(t) = \sigma_0 + \sigma_1 t + \sigma_2 \sin\left(\frac{2\pi t}{T}\right) + \sigma_3 \cos\left(\frac{2\pi t}{T}\right) \tag{2.33}$$

(Katz et al., 2002; Khaliq et al., 2006).

2.7 A point process characterization of extremes

The use of the GPD as discussed in Section 2.3 is essential for modelling residuals above the threshold. According to Coles (2001), the point process approach is considered as an alternative extreme value analysis criterion that is more advanced to an extent of jointly representing both the block and threshold approaches. However, it is further indicated in Coles (2001) that the same parametric inference that is concluded based on either the GEVD or the GPD is exactly achieved through using point processes.

Point process refers to stochastic model of points that are randomized in some space, where points may represent times of events or locations of objects that are characterized by a stochastic system (Pickands III, 1971; Karr, 1991; Resnick, 2013, among others). As explained in Smith (1989) and Coles (2001), a point process approach provides the basis for unifying and extending EVT modelling based on both block and threshold methods in the view of high-level exceedances.

To introduce the point process concepts according to Beichelt (2006), let $\mathscr{A} = [0, \infty)$ denote a set of events occurring in particular periods. In this dissertation, the sequence of interarrival times $\{t_1, t_2, \ldots\}$ is used as a representation of a point process as follows:

$$\{y_1, y_2, \ldots\} \quad \text{for} \quad y_i = t_i - t_{i-1} | i = 1, 2, \ldots ; t_0 = 0$$

with the total count of occurrences within the interval $(0, t]$, $t > 0$ given by:

$$n(t) = \max\{n, t_n \leq t\}$$

such that the counting process for $\{t_1, t_2, \ldots\}$ assuming that at most one event occurs at a time, is given by:

$$\{n(t), t \geq 0\}.$$

Considering the event times as random variables, the sequence $\{T_1, T_2, \ldots\}$ constitute a random point process with

$$T_1 < T_2 < \ldots \quad \text{and} \quad Pr\left(\lim_{i \to \infty} T_i = +\infty\right) = 1. \tag{2.34}$$

Let

$$N(t) = \max\{n, T_n \leq t\}$$

denote the arbitrary count of the occurring events within interval $(0,t]$. The stochastic process given by $\{N(t), t \geq 0\}$ within $Z = \{0,1,\dots\}$ is a counting process that belongs to $\{T_1, T_2, \dots\}$, provided the following properties are fulfilled:

1. $N(0) = 0$,
2. $N(s) \leq N(t) \quad \text{for} \quad s \leq t$,
3. For all possible s,t such that $0 \leq s < t$, then $N(s,t) = N(t) - N(s)$ matches the number of occurrences within $(s,t]$.

2.7.1 Stationarity

The major requirement of a strict state of a stationary process is the invariance to the absolute shifts in time. Thus, the shared distribution of the processes amongst the events are as well expected to be invariant (Khuluse, 2010). If strong stationarity exists in the sequential interarrival times $\{Y_1, Y_2, \dots\}$, then a point process $\{T_1, T_2, \dots\}$ is referred to as stationary, meaning that if for any sequential numbers $i_1, i_2, \dots, i_k$ with $1 \leq i_1 < i_2 < \cdots < i_k$, $k = 1,2\dots$ and for any $\tau = 0,1,2,\dots$, the shared densities of the following random vectors agree:

$$\{Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}\} \quad \text{and} \quad \{Y_{i_1+\tau}, Y_{i_2+\tau}, \dots, Y_{i_k+\tau}\}$$

(Finkenstadt and Rootzén, 2003; Beichelt, 2006). However, full stationarity is in most cases impractical since in such situations, point processes end up seeking for certain properties with which stationarity is assumed. According to Beichelt (2006), for a stationary counting process $\{N(t), t \geq 0\}, 0 \leq s < t$ that is valid for $\tau = 0,1,2,\dots$, the cdf of $N(s,t)$ is only dependent on $\tau = t - s > 0$ such that:

$$p_k(\tau) = Pr(N(s, s+\tau) = k); \quad k = 0,1,\dots; \quad s \geq 0, \quad \tau > 0 \tag{2.35}$$

and hence for a point process that fulfills stationarity condition,

$$m(\tau) = m(s, s+\tau) = m(s+\tau) - m(s) \quad \forall s \geq 0, \tau \geq 0. \tag{2.36}$$

Khuluse (2010) discusses that a counting process is called weak stationary provided there is invariance in the mean and variance to absolute shifts in time. To this effect, the counting process possesses homogeneous increments provided the sequential interval fulfills strict stationarity.

2.7.2 The intensity of a point process

For stationary point processes, Beichelt (2006) defines the intensity as an average count of the events within $[0,1]$, given as follows:

$$\begin{aligned}\Lambda(t) &= E(N(t)) \\ &= \sum_{k=0}^{\infty} kp_k(t) \quad t \geq 0 \end{aligned} \tag{2.37}$$

such that λ is the mean of the occurrences in the interval of unit length given by:

$$\lambda = m(s, s+1), \quad s \geq 0. \tag{2.38}$$

Generalizing equation (2.38), the average count of occurrences within $(s,t]$ of the distance $\tau = t - s$ is given by:

$$m(s,t) = \lambda(t-s) = \lambda\tau. \tag{2.39}$$

When the time axis is re-scaled such that the occurrence of events is restricted to the interval $(0,1]$, the average count of the occurrences within $[0,t]$ gives function of trend given by:

$$\Lambda(t) = \int_0^t \lambda(x)\mathrm{d}x, \quad t \geq 0$$

which is the intensity measure for non-stationary processes. The intensity function is $\lambda(.)$ such that the mean for the stationary process is

$$\Lambda(t) = \int_0^t \lambda \mathrm{d}x = \lambda t.$$

Khuluse (2010) emphasizes the importance of a measure of intensity during statistical modelling using point processes.

2.7.3 Poisson point process

Snyder and Miller (2012) refer to a Poisson process as a most simple process associated with counting random numbers of points. $\{N(t), t \geq 0\}$ fulfills requirements of homogeneous Poisson process with the intensity $\lambda > 0$, if these conditions hold:

1. $N(0) = 0$,
2. $\{N(t), t \geq 0\}$ is a stochastic process whose increments are independent,
3. $N(s,t) = N(t) - N(s), 0 \leq s < t$, are Poisson distributed with $\lambda(t-s)$:

$$Pr(N(s,t) = i) = \frac{(\lambda(t-s))^i}{i!} e^{-\lambda(t-s)}; \quad i = 0, 1, \ldots \tag{2.40}$$

(Beichelt, 2006). Like any other parametric model, Poisson process possesses parameters that are estimable in order to understand the features of a population based on the sample observations. The inferential method that will be used is the MLE in which the parametric vector θ of a non-homogeneous Poisson point process is estimated.

Let $\mathscr{A}$ be a set that encloses the sample points with parametric family $\lambda(.;\theta)$ for an intensity function. Let $I_i = [t_i, t_i + \tau_i] \quad \forall i = 1, 2, \ldots, n$ denote the minor intervals surrounding the observations. To be able to transform the intensity function into that of a Poisson process (Snyder and Miller, 2012), we define

$$\mathscr{I} = \mathscr{A} - \cup_{i=1}^{n} I_i$$

where the points contained in an interval $\mathscr{I}$ appear in the interval $\mathscr{A}$ in a manner that several events are discarded within the interval. The likelihood of a count or more within I_i is given by:

$$Pr\left[N\left(I_i\right) = 1\right] = \Lambda(I_i;\theta) \exp\left\{-\Lambda\left(I_i;\theta\right)\right\}$$

where

$$\Lambda\left(I_i;\theta\right) = \int_{t_i}^{t_i+\tau_i} \lambda\left(u\right) \mathrm{d}u \approx \lambda\left(t_i\right) \tau_i,$$

such that

$$Pr\left[N\left(I_i\right) = 1\right] = \lambda\left(t_i\right) \tau_i \exp\left\{-\lambda\left(t_i\right) \tau_i\right\}$$

for $\tau_i \rightarrow 0$. When there is no count of occurrences within interval, then

$$Pr[N(\mathscr{I}) = 0] = \exp\{-\Lambda(\mathscr{I})\} \approx \exp\{-\Lambda(\mathscr{A})\}$$

due to the minor values τ_i (Beichelt, 2006). We then have

$$\begin{aligned} L(\theta;t_1,t_2,\ldots,t_n) &= Pr[N(\mathscr{I})=0,N(I_1)=1,N(I_2)=1,\ldots,N(I_n)=1] \\ &= Pr[N(\mathscr{I})=0]\prod_{i=1}^{n} Pr[N(I_i)=1] \\ &= \exp\{-\Lambda(\mathscr{A};\theta)\}\prod_{i=1}^{n}\lambda(t_i;\theta)\,\tau_i. \end{aligned} \tag{2.41}$$

The resulting likelihood function after dividing through by τ_i for the density is then given by

$$L(\theta;t_1,t_2,\ldots,t_n) = \exp\{-\Lambda(\mathscr{A};\theta)\}\prod_{i=1}^{n}\lambda(t_i;\theta), \tag{2.42}$$

where

$$\Lambda(\mathscr{A};\theta) = \int_{\mathscr{A}} \lambda(u;\theta)\mathrm{d}u$$

(Coles, 2001; Khuluse, 2010, among others). Coles (2001) shows the log likelihood function as:

$$\ell(\lambda(\theta,t)) = \Lambda(\mathscr{A};\theta) + \sum_{i=1}^{n}\log\lambda(t_i;\theta). \tag{2.43}$$

2.7.4 Connection between a Poisson point process and extreme value theory

Most of the asymptotic results in EVT models are obtained through convergence of limiting distributions. The point process approach offers the sophisticated manner in which the outcomes of EVT limit are expressed to an extent that the same analyses and parametric estimations that are conducted with blocks and threshold approaches can exactly be obtained (Coles, 2001; Khuluse, 2010). Even though the point process approach serves as an alternative of classical and POT procedures, point processes possess more attractive features that are neither in the GEVD nor GPD (Coles, 2001), namely; representation of the block maxima and POT approaches simultaneously as well as the capability of estimating not only the frequency, but also the intensity of the occurrence of extreme values.

2.7.5 A Poisson process limit for extremes

A requirement for fitting point process model is that the block maxima M_n of a series $X_1,X_2,\ldots,X_n$ follows GEVD for the normalizing constants $\{a_n>0\}$ and $\{b_n\}$, leading to the point process on $\mathbb{R}^2$ with the coordinates given by

$$T_n = \left\{ \left(\frac{i}{(n+1)}; Y_{n,i} = \frac{(X_i - b_n)}{a_n} \right) : i = 1, \ldots, n \right\}, \tag{2.44}$$

where time axis is through $(0,1)$; and the second point ensures stability in the occurrence of extremes as $n \to \infty$ such that on $[0,1] \times [\tau, \infty), T_n \to T$ as $n \to \infty$, where T is heterogeneous Poisson process (Karr, 1991; Beichelt, 2006; Resnick, 2013, among others). The points in equation (2.44) reveal evidence of non-homogeneous Poisson process that possesses a measure of intensity that is given by Beichelt (2006) as follows:

$$\Lambda\{(t_1, t_2) \times (x, \infty)\} = (t_2 - t_1) \left[1 - \xi \frac{(x - \mu)}{\sigma} \right]^{\frac{1}{\xi}}, \tag{2.45}$$

whenever $0 \le t_1 \le t_2 \le 1$ and $1 - \xi \frac{(x-\mu)}{\sigma} > 0$. The GPD model is applied as the limiting conditional probability (that $Y_{n,i} > \tau + y$ given $Y_{n,i} > \tau$) that is expressed as:

$$\begin{aligned} P(Y_{n,i} > \tau + y | y_{n,i} > \tau) &= \frac{\Lambda\{(0,1) \times (\tau + y, \infty)\}}{\Lambda\{(0,1) \times (\tau, \infty)\}} \\ &= \left[1 - \frac{\xi y}{\sigma - \xi\tau + \xi\mu} \right]^{\frac{1}{\xi}}, \end{aligned} \tag{2.46}$$

which is a GPD with the scale parameter $\sigma - \xi\tau + \xi\mu$. The parameters of this model are estimable using the MLE where the numerical techniques such as Newton-Raphson and quasi-Newton iteration are used to determine the estimates.

$$\begin{aligned} L\left(\mu_{ij}, \sigma_j, \xi; y_{ijm}\right) &= \prod_{i,j} \left[\exp\left\{ -p_{ij} \left(1 + \frac{\xi_j \mu_{ij}}{\sigma_j} \right)^{\frac{1}{\xi_j}} \right\} \right. \\ &\quad \times \left. \prod_{m=1}^{N_{ij}} \left\{ \left(1 - \xi_j \left(y_{ijm} - \mu_{ij}\right)/\sigma_j\right)^{\frac{1}{\xi_j} - 1} / \sigma_j \right\} \right]. \end{aligned} \tag{2.47}$$

Considering a space of the form $A = [0,1] \times [\tau, \infty)$ for a higher threshold τ, all the values of T_n possess a p chance of occurring within A, where

$$p = P\left\{ \frac{(X_i - b_n)}{a_n} > \tau \right\} \approx \frac{1}{n} \left[1 + \xi \left(\frac{\tau - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}}. \tag{2.48}$$

As the binomial mass approaches to the limiting Poisson distribution, then $T_n(A)$ as $n \to \infty$ follows $Poi(\Lambda(A))$ such that for all spaces that satisfy $A = [t_1, t_2] \times [\tau, \infty)$, with $[t_1, t_2] \subset [0, 1]$, the limiting distribution of $T_n(A)$ is also $Poi(\Lambda(A))$, where

$$\Lambda(A) = (t_2 - t_1)\left[1 + \xi\left(\frac{\tau - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \tag{2.49}$$

which arises as a homogeneous consequence of the process in the direction of time (Karr, 1991; Beichelt, 2006). The ordinary GEVD model, the GEVD model for r largest ordered observations and the POT models are all constituted in the Poisson process, which therefore becomes a reason for the Poisson process to be an advanced alternative characterization for all the EVT models (Coles, 2001). Considering the point process in equation (2.44), the re-scaled maxima is similar to the occurrence equivalent of $T_n(A_z) = 0$, where $A_z = (0, 1) \times [z, \infty)$ (Beichelt, 2006). For the threshold models with

$$\Lambda\left(A_z\right) = \Lambda_1\left([t_1, t_2]\right) \times \Lambda_2\left([z, \infty)\right)$$

where

$$\Lambda_1\left([t_1, t_2]\right) = \left(t_2 - t_1\right) \quad \text{and} \quad \Lambda_2\left([z, \infty)\right) = \left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}$$

such that

$$\begin{aligned} P\left\{\frac{(X_i - b_n)}{a_n} > z \mid \frac{(X_i - b_n)}{a_n} > \tau\right\} &= \frac{\Lambda_2[z, \infty)}{\Lambda_2[\tau, \infty)} \\ &= \left[1 + \xi\left(\frac{z - \tau}{\tilde{\sigma}}\right)\right]^{-\frac{1}{\xi}} \end{aligned} \tag{2.50}$$

with $\tilde{\sigma} = \sigma + \xi(\tau - \mu)$.

2.7.6 Inference for a Poisson point process

The likelihood function for determining the parameters μ, σ and ξ is as follows:

$$\begin{aligned} L_A(\mu,\sigma,\xi;x_1,\ldots,x_n) &= \exp\{-\Lambda(A)\}\prod_{i=1}^{N(A)}\lambda(t_i,x_i) \\ &\propto \exp\left\{-n_y\left[1+\xi\left(\frac{\tau-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ &\times \prod_{i=1}^{N(A)}\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}, \end{aligned} \tag{2.51}$$

corresponding to the negative log likelihood function given by:

$$\begin{aligned} -\ell(x_1,x_2,\ldots,x_n) &= n_y\left[1+\xi\left(\frac{\tau-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} + n_\tau \log\sigma \\ &+ \sum_{i=1}^{n_\tau}\left(\frac{1}{\xi}+1\right)\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]. \end{aligned} \tag{2.52}$$

Likelihood function for determine GPD parameters of a Poisson process is as follows:

$$L(\zeta,\tilde{\sigma},\xi;x_1,\ldots,x_n) = (1-\zeta)^{n-n_\tau}\prod_{i=1}^{n_\tau}\zeta\tilde{\sigma}^{-1}\left[1+\xi\left(\frac{x_i-\tau}{\tilde{\sigma}}\right)\right]^{-\frac{1}{\xi}-1}, \tag{2.53}$$

where n_τ is the total count of the residuals above τ (Coles, 2001; Beichelt, 2006; Khuluse, 2010).

2.8 Parametric inference

2.8.1 Method of maximum likelihood estimation

Likelihood methods of finding parameters of EVT models are more reliable compared to others because of several advantages including the adaptability to model change (Miller, Freund and Johnson, 1965; Mukhopadhyay, 2000; Coles, 2001; Ross, 2014; Devore, 2015, among others). The MLE method is applied in parametric inference in this dissertation. Let $X_1,\ldots,X_n$ represent i.i.d. average daily temperatures with densities $f(x_i)$. The likelihood function is as follows:

$$L(x_i,\theta) = \prod_{i=1}^{n} f(x_i,\theta). \tag{2.54}$$

The log likelihood function is given by:

$$\ell(x_i,\theta)=\frac{d}{d\theta}\log\left[\prod_{i=1}^{n}f(x_i,\theta)\right]=\sum_{i=1}^{n}\log f(x_i,\theta)$$

determines the values of θ that maximize $\ell(x_i,\theta)$ (Devore, 2015).

2.8.1.1 Parametric inference: Maximum likelihood estimation for GEVD

Let $X_1,\ldots,X_n$ be the block maxima that are independent random variables whose distribution is GEVD. The log likelihood function for equation (2.8) when $\xi \neq 0$ is as follows:

$$\begin{aligned}\ell(x_i,\mu,\sigma,\xi) &= -n\log\sigma-\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]\\ &- \sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}},\end{aligned}\tag{2.55}$$

where $1+\xi\left(\frac{x_i-\mu}{\sigma}\right)>0$, $i=1,\ldots,n$ (Coles, 2001). If this condition is not fulfilled, then $L(x_1,x_2,\ldots,x_n;\mu,\sigma,\xi)=0$ and $\ell(x_1,x_2,\ldots,x_n;\mu,\sigma,\xi)=-\infty$. In a special situation for $\xi=0$, the log likelihood for equation (2.4) is given as follows:

$$\ell(x_i,\mu,\sigma)=-n\log\sigma-\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)-\sum_{i=1}^{n}\exp\left\{-\left(\frac{x_i-\mu}{\sigma}\right)\right\}\tag{2.56}$$

(Coles, 2001). For the case of $X_1,\ldots,X_n$ following GEVD for r largest order statistics, the likelihood function for equation (2.8) is given as follows:

$$\begin{aligned}L(x_i,\mu,\sigma,\xi) &= \prod_{i=1}^{n}\left(\exp\left\{-\left[1+\xi\left(\frac{x_i^{(r_i)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}\right.\\ &\times \left.\prod_{j=1}^{r_i}\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i^{(j)}-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\right),\end{aligned}\tag{2.57}$$

where $1+\xi\left(\frac{x^{(j)}-\mu}{\sigma}\right)>0$, $j=1,\dots,r_i$, $i=1,\dots,n$; zero otherwise (Coles, 2001). For the case of $\xi=0$, the likelihood function is as follows:

$$\begin{aligned} L(x_i,\mu,\sigma,\xi) &= \prod_{i=1}^{n}\left(\exp\left\{-\exp\left[-\left(\frac{x^{(r_i)}-\mu}{\sigma}\right)\right]\right\}\right.\\ &\quad\times\left.\prod_{j=1}^{r_i}\frac{1}{\sigma}\exp\left[-\left(\frac{x_i^{(j)}-\mu}{\sigma}\right)\right]\right) \end{aligned} \qquad (2.58)$$

(Coles, 2001). Equations (2.55)-(2.58) are maximized by parameter estimates obtained using the numerical techniques for the purpose of determining the maximum likelihood estimates. The MLE method is the most commonly used parametric inference tool in most of the statistical models including EVT. However, MLE is characterized with some limitations when it comes to estimating shape parameter of the GEVD (Sigauke, Chikobvu and Verster, 2012).

The point estimates of shape parameter are computed with regard to certain properties when using the MLE method and there are some situations under which the shape is not estimable with the MLE. Those are the situations under which the Bayesian framework gains preference over the MLE since there are no certain limiting properties that bind the use of the Bayes estimation approach (Beirlant et al., 2004; Sigauke et al., 2012). The following are the properties for the estimation of ξ using MLE (Smith, 1985):

1. When $\xi<-0.5$, the distribution possesses upper tail that is very short bounded and the MLEs are very rarely obtainable.

2. When $\xi>-0.5$, the MLEs fulfill the standard requirements.

3. When $-1<\xi<-0.5$, the MLEs are obtainable in a usual manner but without fulfilling the standard requirements.

4. When $\xi<-1$, the MLEs are normally unobtainable.

2.8.1.2 Parametric inference: Maximum likelihood estimation for GPD

If $y_1,\dots,y_k$ constitute a random sample of k excesses over an upper threshold τ, the log likelihood function for GPD is as follows:

$$\ell(y_i,\sigma_\tau,\xi) = -k\log\sigma_\tau-\left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi y_i}{\sigma_\tau}\right), \qquad (2.59)$$

where $\left(1+\frac{\xi y_i}{\sigma_\tau}\right) > 0$, $\forall i = 1,2,\ldots,k$ and $-\infty$ for $\left(1+\frac{\xi y_i}{\sigma_\tau}\right) < 0$ (Coles, 2001). The log likelihood function $\ell(y, \sigma_\tau, \xi)$ exists as an exponential distribution for $\xi = 0$ which is given by:

$$\ell(y_i, \sigma_\tau) = -k \log \sigma_\tau - \frac{1}{\sigma_\tau} \sum_{i=1}^{k} y_i. \tag{2.60}$$

2.8.2 Bayesian estimation approach

The use of Bayesian parametric estimation approach in EVT is dealt with in this section. This estimation approach is used in this dissertation for estimating the target parameters of EVT models. The advantage of Bayes over MLE approach is its independence of any limiting properties. To define the Bayesian framework according to Beirlant et al. (2004);

let $x = x_1, \ldots, x_n$ be the values that are contained in X that is distributed with $f(x \mid \theta)$. If $\pi(\theta)$ is the density of the prior distribution with parameter $\theta = (\theta_1, \ldots, \theta_p)$ and $\pi(x \mid \theta)$ the likelihood function, we have the posterior distribution given by:

$$\pi(\theta \mid x) = \frac{\pi(x \mid \theta)\pi(\theta)}{\int_\theta \pi(x \mid \theta)\pi(\theta) d\theta} \propto \pi(x \mid \theta)\pi(\theta), \tag{2.61}$$

where the integral extends through $\theta \in \Omega$. Amongst the rest, the attractive features of the Bayesian framework include the following (Behrens et al., 2004; Beirlant et al., 2004, among others):

1. Basis of converting initial idea about the target parameter θ into the posterior distribution $\pi(\theta \mid x)$,

2. Estimates of the target parameter θ are determined by the mean or mode of $\pi(\theta \mid x)$ and the accuracy of estimation is maintained by the posterior distribution itself,

3. The ease in predictions and more reliable interval estimation at suitably high precision of estimation.

Coles and Powell (1996) argue that the GEVD model given in equation (2.3) does not admit any conjugate prior distribution and to that limitation, two approaches based on obtaining conjugate families for sub-classes of models have been proposed.

2.8.3 Model checking and selection

In statistical modelling, it is a usual behavior to assess validity of the assumptions connected to a proposed model. This practice is formally known as a goodness of fit test (Tsujitani, Ohta and Kase, 1980). When statistical modelling is done in several stages, there are usually various candidate models from which the best suitable model must be chosen (Songchitruksa and Tarko, 2006). Initially, we adopt the use of the deviance statistic that is considered by El Adlouni, Ouarda, Zhang, Roy and Bobée (2007) as the classical basis for choosing the most appropriate EVT model under specific situations. This is used for examining fit of the models considering maximum likelihood estimation. The deviance statistic is defined by

$$\begin{aligned} D_{(i,j)} &= 2\ln\left(\frac{\lambda(r_i)}{\lambda(r_j)}\right) \\ &= 2\left\{\ln\lambda(r_i) - \ln\lambda(r_j)\right\} \sim \chi_1^2, \quad \text{for} \quad i,j = 1,2,\ldots,6 \quad (i \neq j), \end{aligned} \tag{2.62}$$

where $\lambda(r_i)$ and $\lambda(r_j)$ show maximum likelihood functions of r_i and r_j respectively (Tsujitani et al., 1980; Songchitruksa and Tarko, 2006; El Adlouni et al., 2007). However, the degrees of freedom in χ_1^2 is a special case that holds for models such as the stationary r largest order statistics possessing same number of parameters. It is not always the case for the models such as non-stationary cases with different number of parameters. A test for appropriateness of candidate model considering r_i compared to r_j at significance level α is to reject r_j whenever $D_{(i,j)} > C_\alpha$, provided C_α is the $(1-\alpha)$ quantile of the χ_1^2 distribution (Smith, 1989; Coles, 2001; Soares and Scotto, 2004). However, the deviance statistic is useful when the models are nested and if not, the use is made of the information criteria techniques such as the AIC or BIC (Burnham and Anderson, 2004; Vrieze, 2012).

2.8.3.1 Graphical model checking

The assessment of EVT models, more especially GEVD and GPD is commonly done using observed data due to difficulty that is associated to extrapolation (Smith, 1987; Coles, 2001; Beirlant et al., 2004, among others). Instead of extrapolation approach, use is made of the graphical diagnostic plots with observed data so that model fit can be examined. In this dissertation, EVT models are assessed through the use of graphical diagnostic plots. The diagnostic plots are constructed as follows:

1. Probability-Probability (P-P) plot

   The comparison of the fitted and the empirical distribution functions is referred to as a P-P plot. Considering the GEVD with the i^{th} ordered block maxima, the evaluated empirical distribution function of $x_{(i)}$ is $\tilde{G}(x_{(i)}) = i/(n+1)$, and the fitted distribution function at the same point is given by

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

   A good model is achieved if $\tilde{G}(x_{(i)}) = \hat{G}(x_{(i)})$. We expect the plot of the points $\big(\tilde{G}(x_{(i)}), \hat{G}(x_{(i)})\big)$, $i = 1, \dots, n$ to approach 45° line.

   Looking at the GPD, suppose that $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(k)}$ are order excesses above sufficiently high threshold τ and let $\hat{H}(y)$ be an estimated GPD. Then the P-P plot of the GPD is given by the point $\left(i/(k+1), \hat{H}(y_{(i)})\right), i = 1, \dots, k$ where $\hat{H}(y)$ is the estimated model of equation (2.12). The disadvantage of the P-P plots for both GEVD and GPD is that the fitted and the empirical functions are bounded to approach 1 and the plot ends up being less informative. To this limitation, preference is gained by the Quantile-Quantile plot (Coles, 2001; Beirlant et al., 2004; Mallor, 2009).

2. Quantile-Quantile (Q-Q) plot

   In Q-Q plot of GEVD, the represented points are $\left(\hat{G}^{-1}(i/(n+1)), x_{(i)}\right)$, $i = 1, \dots, n$, where

   $$\hat{G}^{-1}(i/(n+1)) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left(1 - \left(-\log\left(\frac{i}{n+1}\right)^{-\hat{\xi}}\right)\right), \quad i = 1, \dots, n.$$

   Looking at the GPD, the Q-Q plot represents the points $\left(\hat{H}^{-1}\left(i/(k+1)\right), y_{(i)}\right), i = 1, \dots, k$,

   $$\text{for} \quad \hat{\xi} \neq 0, \quad \hat{H}^{-1}(p) = \frac{\hat{\sigma}}{\hat{\xi}}\left((1-p)^{-\hat{\xi}} - 1\right) \quad \text{and for} \quad \hat{\xi} = 0, \quad \hat{H}(p) = -\hat{\sigma}\ln\left((1-p)\right).$$

   The Q-Q plots are expected to display linear function and the departure from linearity reveals evidence of failure in goodness of fit (Coles, 2001; Beirlant et al., 2004;

Mallor, 2009).

3. Return level plot

   Return level plot is used as one of the graphical diagnostic tools, and is plotted based on the quantile function. Looking at the GEVD, letting $y_p = -\ln(1-p)$, $\hat{x}_p$ is plotted against $\log y_p$ on a logarithmic scale to represent points $(\log y_p, \hat{x}_p)$, $0 < p < 1$ which results into a linear plot in case $\xi = 0$, convex plot with asymptotic limit if $\xi < 0$ as $p \to 0$ and the shape is concave and unbounded for $\xi > 0$. Considering the GPD, return level plot is given by a locus of the points $(m, \hat{x}_m)$ for large values of m, where $\hat{x}_m$ is the estimated $m-$observation return level in equation (2.14). The empirical estimates of the return level function are also added if the return level plot is used as one of the model diagnostic tools. For the suitable models, there must be agreement between the empirical estimates and the model based curve (Coles, 2001; Beirlant et al., 2004; Mallor, 2009).

2.8.4 Assessment for uncertainty of point estimates

The estimation of target parameters in several models including the GEVD and GPD in EVT is characterized by uncertainty and to this effect, there is a need to assess estimates of parameters for uncertainty (Wu et al., 2007). Several methods including bootstrap resampling are proposed in literature as relevant for such an assessment. The bootstrap resampling approach is relevant in EVT as an automatic computer-based method for assessing uncertainty in the estimation of parametric and non-parametric models (Efron and Tibshirani, 1994; Coles and Simiu, 2003; Li et al., 2010, among others).

In the bootstrap approach, the bootstrap densities are fitted to each of the target parameters, resulting in the density plots that are used to assess the fit. The better the fit of the bootstrap densities, the lesser the level of uncertainty. The fitted densities further give more information with which uncertainty is assessed. These include the bootstrap means, bootstrap medians, biases and standard errors of the parameter estimates. We expect these statistics to be smaller as compared to the theoretical values in order to maintain certainty in the parametric estimation. In this dissertation, the bootstrap resampling approach is used for assessing uncertainty in the models that are based on the POT approach.

2.9 Applications of Extreme Value Theory

Section 2.9 focuses on the applications of EVT methodologies. Several studies in which EVT is used are summarized. These include the application of EVT in meteorology, hydrology and other possible areas of application.

2.10 Application of stationary and non-stationary generalized extreme value distribution

Classical EVT emphasizes the detection and analysis of extremes based on the block maxima approach. This further generalizes to the asymptotic behavior of the annual block maxima of r largest order statistics (Coles, 2001). However, most of the research articles that are reviewed in this dissertation use the ordinary GEVD (for $r = 1$) and the MLE for statistical inference on the target parameters. In this dissertation we consider both the MLE and the Bayesian inference for estimating parameters of the models.

Hasan and Kassim (2012) consider the use of GEVD in modelling extreme temperature at Penang, Malaysia. Their study focuses on the occurrences of climatological disorders, including floods and weather conditions such as maximum temperature that may negatively influence living conditions of the Malaysian population at Penang. The GEVD as a classical approach of EVT is used to assess the frequencies and describing the stochastic behavior of maximum temperature as an extreme event. Their study is aimed at reaching informative results that may possibly contribute towards preparing the Malaysian community for extreme temperatures. Average daily temperatures that are collected in degrees Fahrenheit over a period of 32 years from 22 meteorological stations at Penang are used. Hasan and Kassim (2012) use the MLE and the L-Moments Method (LMOM) for parametric inference. Confidence intervals on profile likelihood functions are also used. Models are diagnosed for fit using generalized likelihood ratio (GLR) test as well as relevant graphical diagnostic tools. Results of Hasan and Kassim (2012) establish the need for a non-stationary GEVD model on 11 meteorological stations of which, 8 exhibit the existence of trend. For both models, the estimates of shape parameters in all stations confirm short-tailed GEVD with negative values that are closer to zero, confirming approximation to linearity. Graphical diagnostics also confirm a convincing appropriateness of the models.

The models that are discussed in Hasan and Kassim (2012) are the following:

1. σ, μ and ξ are treated as intercepts, (time-homogeneous GEVD).

2. $\mu(t) = \mu_0 + \mu_1 t$, with ξ and σ as intercepts.

3. $\sigma(t) = \exp(\sigma_0 + \sigma_1 t)$, with μ and ξ as intercepts, μ, σ and ξ are location, scale and shape parameters of the GEVD.

Depending on availability of extremes, threshold based models could provide better results than the block maxima models that are used in Hasan and Kassim (2012). Ferreira, de Haan et al. (2015) give more exploration on the comparison of POT and block approaches.

Smith (1989) and Coles (2001) refer to a statistical model as a stationary model if it does not cater for any variations in the parameters with respect to time or any other covariates. Such a model requires a process $X_1, X_2, \ldots$ to be invariant of any shifts in time (Beichelt, 2006). Observation of trends and non-stationarity injects useful information in the analyses of extremes. Coles (2001) explains that, non-stationarity within the GEVD model can only be taken care of in the location and scale parameters due to the difficulty in modelling the variations in the shape parameter.

Nadarajah (2005) studies the extreme rainfall in West Central Florida to investigate and identify types of non-stationary patterns that arise in forms of trends. The block maxima of daily rainfall are taken annually over a period of 102 years from 14 meteorological stations. The modelling is done using the GEVD in which μ is reparameterized firstly as a linear function and secondly as quadratic function of time for the purpose of modelling non-stationarity. Inference on the target parameters is done using the MLE method and the standard deviations of estimators are determined by inverting the Fisher's matrix. Nadarajah (2005) discusses and fit the following models:

1. σ, μ and ξ are treated as intercepts, (time-homogeneous GEVD).

2. σ and μ are treated as intercepts with $\xi = 0$, (Gumbel family).

3. $\mu(t) = \mu_0 + \mu_1(\text{Year} - t_0 + 1)$, with σ and ξ as intercepts, which is a time-varying model with 4 parameters.

4. $\mu(t) = \mu_0 + \mu_1(\text{Year} - t_0 + 1)$, with σ as an intercept and $\xi = 0$, which is a time-varying model with 3 parameters.

5. $\mu(t) = \mu_0 + \mu_1(\text{Year} - t_0 + 1) + c(\text{Year} - t_0 + 1)^2$, with ξ and σ as intercepts, which is a model with 5 parameters.

6. $\mu(t) = \mu_0 + \mu_1(\text{Year} - t_0 + 1) + c(\text{Year} - t_0 + 1)^2$, with σ as intercept and $\xi = 0$, which is a model with 4 parameters where t_0 in all cases denotes the year on which rainfall records started.

About 57% of meteorological stations in West Central Florida show existence of non-stationarity in forms of several types of trends including downward linear trends, concave and convex-types of quadratic trends. Standard likelihood ratio test statistic shows that the GEVD fits the data better than model 2 (Gumbel family). Nadarajah (2005) further singles out that, of all the 6 fitted models, model 5 provides the most relevant results, which confirm that the quadratic trends significantly contribute towards non-stationarity through time. Even though Coles (2001) states that the threshold exceedance models and the point processes provide better analyses of extremes, the reparameterization of a location parameter as the linear and quadratic functions of time provides more detailed ideas in the analyses.

This dissertation focuses attention on the use of the GEVD that possesses a location parameter μ that is reparameterized as linear and quadratic functions of time. We have also considered the exponential transformation of scale parameter σ for the purpose of modelling non-stationarity and maintaining non-negativity of the scale parameter.

2.10.1 Application of generalized extreme value distribution for r largest order statistics

Ferreira, de Haan et al. (2015) explain that the block approach is a more efficient method under usual practical conditions, but has not been studied thoroughly in comparison to the threshold approach. Davison and Smith (1990) argue that MLE methodologies using r largest order statistics model are more efficient in comparison to the POT approach. These ideas are perhaps in odds with Smith (1986) and Coles (2001) who emphasize that the use of block maxima approach in the presence of an entire time series is wasteful of information. The use of the ordinary GEVD (for $r = 1$) is further criticized by Coles (2001) and other several authors in literature, in favour of the r largest order statistics model. It is argued that major limitation in EVA is the scarcity of extremes and consequently, characterizations of extreme value modelling that are more concrete than a block maxima approach are essential (Weissman, 1978; Smith, 1989; Coles, 2001). Smith (1986), Coles (2001), An and Pandey (2007) and Balakrishnan and Cohen (2014) explain that the GEVD for block maxima gains less preference in comparison to other EVT techniques due to its limitations in providing the detailed analysis of extremes.

In this dissertation, GEVD for $r \geq 1$ largest order statistics is fitted to annual maxima for a sufficiently small number of ordered observations, r. As discussed in Section 2.2, this asymptotic model arises as a generalization of GEVD for block maxima for the purpose of assessing stochastic behavior of several values of order statistics (Coles, 2001; Soares and Scotto, 2004). As a requirement for this model, the data must be in matrix form with rows matching to blocks and columns to order statistics (Heffernan and Stephenson, 2014). The precautionary measure that has to be followed when using GEVD for r largest order statistics is grounded on the tradeoff between bias and variance (Smith, 1986; Arnold, Balakrishnan and Nagaraja, 1992; Balakrishnan and Cohen, 2014). In the application of this technique in this dissertation, the total count of the annual order statistics r is not too small to avoid the likelihood estimators that possess high variance and at the same time, not too large since the asymptotic support may be violated and hence biased (Smith, 1986; Arnold, Balakrishnan and Nagaraja, 1992; Balakrishnan and Cohen, 2014).

Soares and Scotto (2004) focus on a comparative study of extremal characteristics of time series of the substantial wave heights that are analyzed using extreme value theory. The aim is to fit a GEVD (for $r \geq 1$) towards modelling the incidences of extreme sea levels and the extreme value extrapolation. The observations of interest for this study constitute the sea wave data that are collected in Northern North Sea over a period of 23 years $(1976 - 1999)$. The encountered limitation as emphasized in Smith (1986), Wang (1995) and Coles (2001) is that of the missing values for the period $1994 - 1997$ and the problem is addressed by a replacement with the hindcast values. The main property that is tested for the justification of the asymptotic fit of the r largest GEVD is the annual count of the independent observations that are sufficiently large in comparison to the order statistics. Soares and Scotto (2004) have taken into consideration, the tradeoff between bias and variance in order to achieve the asymptotic arguments of the GEVD (for $r \geq 1$). The model is fitted to 10 values of the order statistics that are selected using the deviance function that is given in equation (2.62). The MLE is used in estimating the target parameters whereby, the paramount attention concerns the first 6 order statistics due to reasonable doubt of the fit for $r \geq 7$.

The r largest order statistics model under consideration is given in equation (2.8). Furthermore, the $m-$year return levels are predicted using the quantile function that is discussed in equation (2.6). For evaluating the adequacy of the fitted models, use is made of the GEVD diagnostic plots which show satisfactory fit of the model within the first 6 cases. The results of the study establish the use of 5 order statistics as an appropriate choice. The block maxima approach that is used in Soares and Scotto (2004) is found to be more competitive in comparison to other techniques due to several facts including the decreased values of the standard errors for all the parametric estimates. The modelling further reveals relevance of Weibull family of distributions towards modelling wave heights in Northern North Sea (Soares and Scotto, 2004).

2.11 Application of stationary and non-stationary generalized Pareto distribution for threshold exceedances

"The use of a block maxima approach tends to be wasteful since only the annual maxima or annual minima are considered in place of an entire time series of observations" (Coles, 2001, page 74). To mitigate against this limitation, analysis based on POT approach leads to the GPD for threshold exceedances which possesses potential to utilize as much as possible of the available information (Coles, 2001; Ferreira et al., 2015).

Li, Cai and Campbell (2005) consider the POT approach in modelling daily rainfall observations that are recorded over the period 1930 – 2001 from 5 weather stations that are geographically scattered around South Western Australia. Although stations are found to be in possession of sufficient data, a minor challenge of missing data is encountered and addressed by substitution with the point patched values. The aim is to model rainfall time series above sufficiently high threshold, at the selected meteorological stations. The mean excess plot is used for choosing the adequately high threshold. The conditional distribution of excesses is given in Balkema and de Haan (1974). The change point procedure is applied in an extreme rainfall distribution and the conclusion is a substantial increase in winter rainfall since the mid-twentieth century. The change in winter extreme rainfall is quantified by estimating the tails of extreme rainfall distribution, which reveals existence of spatial variation (Li et al., 2005).

Sugahara, Da Rocha and Silveira (2009) focus on the use of EVT towards modelling rainfall data that are collected over the period 1933 – 2005. The modelling is done through the

POT approach in which several non-stationary GPD models are fitted. The choice of the non-stationary modelling framework is motivated by three reasons, one being that, "over the past several years, the city has been growing in area and in population, mainly after 1930s, following ever-increasing industrialization and consequently, the city has become one of the most populous and largest metropolitan areas of the world" (Sugahara et al., 2009, page 2). It is further highlighted based on the given reasons that, the stationarity assumption is violated due to inappropriate accommodation of the linear or non-linear time-trend components. The time-dependent variable thresholds are determined through the strategy of testing several sufficiently high thresholds so that the most suitable values can be chosen for fitting the GPD models. The following GPD models are fitted:

1. σ is treated as intercept, (time-homogeneous GPD).
2. $\sigma_d = \sigma_0 + \sigma_1 \sin(2\pi t_d/365) + \sigma_2 \cos(2\pi t_d/365)$, (annual cycle model on σ).
3. $\sigma_d = \sigma'_o + \sigma'_1 t_i$, (linear trend model on σ).
4. $\sigma_d = \sigma''_0 + \sigma_1 \sin(2\pi t_d/365) + \sigma_2 \cos(2\pi t_d/365) + \sigma'_1 t_i$, (annual cycle and long term linear trend model).

The target parameters are estimated using the MLE method because of its capability in accommodating non-stationarity features or covariates which in this case are the annual cycles and the long term linear trend, where t_d denotes the specific day of the year and $q_p(t_d)$ is a threshold that depends on the annual day. The level of uncertainty on the parameter estimates is assessed using the bootstrap resampling approach. The quality of the models is assessed using the P-P (and the Q-Q) plots which confirm appropriate fit in all the four cases. The AIC approach was then used for choosing the best of the candidate models. The sensitive outcomes of the study are based on the detected trend because Sao Paulo has long been associated to floods, possibly resulting from urban growth and lack of laws against population density. A field for further research is recommended with the consideration of more stations in order to deduce more interesting conclusions (Sugahara et al., 2009).

Kyselỳ, Picek and Beranová (2010) model the occurrences of non-stationary extreme quantiles in the distribution of daily temperature. A POT approach is used with a threshold that varies with respect to time. The existence of trends within the observations led to the violation of stationarity assumption, leading to a non-stationary GPD with time-varying parameters that are used in Kyselỳ et al. (2010). The study uses data set that was collected over the period of 139 years. The $T-$year return levels $x(T)$ are estimated using a quantile function in equation (2.14).

Bommier (2014) uses the POT approach with a time-varying scale towards modelling environmental data that are collected in 172 years in Sweden. The collected observations represent a long series of daily average temperatures that reveal non-stationarity in terms of a seasonal component, leading to a variation that is addressed by splitting the data into several monthly series. Initially, modelling is done on each monthly series with the restriction of i.i.d. condition as a requirement of the POT approach. For determining sufficiently high threshold in this case, use is made of several techniques such as the dispersion index, rule of thumb and the multiple-threshold models. However, the interpretation of the threshold values is given for all the techniques that are used, except the residual life plot due to the difficulties in its interpretation as emphasized in Smith (1989) and Coles (2001). The following models are fitted to the positive residuals above the threshold using the MLE based on the following assumptions:

1. σ and ξ are stationary (time-homogeneous GPD).
2. $\sigma(t) = \exp(\sigma_0 + \sigma_1 t)$, where only the scale depends on time.

It is then concluded that, the use of a recent threshold selection technique (multiple-threshold model) that is based on the statistical test is among the others, the most reliable due to several defects that are associated with other classical threshold selection criteria that are used (Bommier, 2014).

2.12 Application of declustering dependent series

This section looks at meteorological studies that involve the data that violate the i.i.d. condition. “Environmental data tend to depart from the independent and identically distributed condition in terms of heavy seasonality and exhibition of short-range dependence” (Smith, 1989, page 5). Weather observations such as daily average temperatures are naturally grouped to an extent that, an extremely cold day is probably followed by other days that are also cold, and such clusters of observations are likely to negatively influence the results of researches (Smith, 1989; Katz and Brown, 1992; Bonsal, Zhang, Vincent and Hogg, 2001; Klein Tank and Können, 2003, among others).

The way to address this problem as emphasized in Ferro and Segers (2003), is to decluster the data, possibly using R extreme value packages such as the texmex declustering package of Southworth and Heffernan (2013b). The automatic declustering algorithm of Ferro and Segers (2003) is applied in Southworth and Heffernan (2013b) towards modelling the daily rainfall observations that are collected from the South-West location of England over the period 1914 – 1962. The threshold stability plot is used for parametric estimation of GPD that is fitted to threshold excesses. Threshold stability plots reveal evidence that, the estimates of the parameters are stable at the chosen declustering threshold, but with instability of extremal index. The data reveal evidence of serial dependence and to this effect, the automatic declustering algorithm in texmex package is applied prior to fitting the GPD to cluster maxima. For assessing validity of the GPD fit to cluster maxima, use is made of the Q-Q plots and it is established that rainfall tends to increase in consecutive days.

The comparative analysis is then carried out in order to compare the GPD that is fitted to original threshold excesses to the GPD that is used towards modelling the cluster maxima, and the remarkable findings are that, the GPD that is used towards modelling cluster maxima is characterized with slightly higher standard errors although the parameter estimates of the two models are not significantly different. It is further concluded that the i.i.d. restriction is violated for all the threshold values in both models, but the validity of independence in cluster maxima is achieved. The return levels are calculated by inverting both GPD models, which results into quantile functions (Southworth and Heffernan, 2013b).

It is argued in Bommier (2014) that, the impact of seasonality leads to short-range dependence and to that effect, use is made of declustering strategies as efforts to get rid of the limitations that may arise as consequences of heavy seasonality. In this special case, the i.i.d. restriction is violated so that the data can first be declustered in order to address the dependence features. A run period of 5 days is used with the threshold of 90% quantile in declustering, and it is found that the parameter estimates of the declustered and non-declustered data suggest the Weibull family of distributions as a valid model for the temperatures in Uppsala, Sweden (Bommier, 2014).

2.13 Application of a point process characterization of extremes

In this section we summarize studies in which extreme observations are modelled using point processes. The point process approach that is used in this dissertation is originally introduced by Pickands III (1971), and then studied and applied by several authors and researchers including Smith (1989), Coles (2001) and Beichelt (2006), who emphasize the importance of the point processes in EVT modelling whereby among the rest, the following two main features of Poisson point process are singled out:

1. Poisson point process provides analyses of extremes through merging the block and POT approaches,

2. as compared to the POT approach, point process is highly connected with the variations in the excesses above the threshold.

Smith (1989) uses EVT models in modelling environmental data that are captured in Houston, Texas. The non-stationary GEVD, non-stationary GPD, together with the point processes are used. The aim is to investigate an existence of trend in the data which consist of hourly measurements of ozone over a 13-years period (April, 1973 to December, 1986). The problem to be solved involves estimating frequency with which specified high levels are exceeded as well as the desire to know whether or not there is any evidence of frequency changing over the period of the study. The EVT modelling approaches are applied in the estimation of target parameters using MLE method that is given in equation (2.47). The data violate i.i.d. condition through short-range dependence and seasonal effects, leading to the declustering of a high level exceedance and the $D(u_n)$ condition. The extremal index θ is estimated using equation (2.26), where $0 \leq \theta \leq 1$ measures the amount of clustering in the process and $1/\theta$ is the limiting average size of the cluster (Smith, 1989). The results reveal that the most extreme emissions have been reduced by about 3 parts per 100 million over the period of the study, recommending further analyses of data collected from different sites.

Point process approach is capable of providing more detailed analyses of extreme observations due to its advanced features that are not available in either the GEVD or GPD. To make use of such advantages, this dissertation uses the Poisson point process as an advanced technique of EVT. Khuluse (2010) uses point process approach towards modelling heavy rainfall in South Africa over time and space. The data that are used consist of daily rainfall that is recorded by the SAWS in *mm* from 15 weather stations over a period of 50 years. The usual extreme value challenge that is encountered in Khuluse (2010) is that of missing data of which only 7 of 15 stations have complete data series. However, incompleteness of data is dealt with in a manner that does not result in misleading outcomes.

The paramount attention is limited to the rainfall records over the winter season for several reasons including lower incompleteness rate of 3% during June, July and August.

Scatter diagram is used to study detailed features of data, and the absence of trend component is established. For the distribution of the data in each of the stations, use is made of box plots with stationarity assumption, and it is noted that the distributions are characterized by fat heavy tails. The sufficiently high threshold is determined using several techniques that are supported by an additional sensitivity study as means to guard against uncertainty. Poisson point process is assumed with the consideration of annual winter block maxima, and the validity of Poisson assumption is tested using dispersion index plot. The estimate of shape parameter is found to be negative, suggesting suitability of Weibull family of distributions towards modelling heavy rainfall in Eastern Cape province of South Africa (Khuluse, 2010).

Chapter 3

Methods

3.1 Introduction

In Chapter 2, the detailed theoretical and mathematical backgrounds of EVT are discussed together with the applicability to various suitable areas. Chapter 3 discusses the methodology that is directly followed or applied throughout this dissertation. This involves several quantitative research approaches that pertain to the use of EVT. The specification of meteorological data and its source are described together with the research instruments in Section 3.2. The procedures of three main phases of analysis that are done in Chapter 4 are discussed. Most of statistical analyses are performed using R, a software package that is useful as a language of programming and statistical modelling (R Core Team, 2013). Packages of extreme value analysis that are used for fitting models in this dissertation are discussed in Section 1.4 of Chapter 1.

3.2 Data

There are several predictor variables that are used in modelling electricity demand in the energy sector. Amongst all variables, average daily temperature is the major driver of electricity demand worldwide. This dissertation considers the use of average daily temperature as meteorological data in modelling electricity demand in South Africa over time. The data comprises national time series of average hourly temperatures that are collected by the SAWS over the period $2000 - 2010$ and provided by Eskom. The data is dated from 1 January 2000 to 30 August 2010 and is partitioned into two seasonal versions so that the purpose of this dissertation can be achieved. Maximum temperatures are specified by defining non-winter season as the period from September to April of each year. The rest of the observations from January to August of each year are considered as the data for winter season. This versions of data are then used for the purpose of modelling the occurrences of both the coldest and hottest days.

3.3 Applying the block maxima approach

In this dissertation, use of block maxima approach in modelling average maximum daily temperatures is deemed vital. This approach requires the use of certain maximum observations instead of an entire time series. Raw data which are hourly temperatures are used for calculating average maximum daily temperatures during non-winter season over the period $2000-2010$. This results in 93 480 observations. The r largest order observations in each year are extracted from the average maximum daily temperature, where $r = 10$ is chosen. The data are sorted in matrix form with rows matching to blocks and columns matching to order statistics. We then end up with 11 blocks and 10 order statistics, implying 110 observations. This is done for the purpose of meeting requirements of fitting the GEVD for r largest order statistics. The use in this regard is made of ismev package that is authored by Heffernan and Stephenson (2014).

The MLE method is used for fitting models and determining target parameters. Amongst 10 chosen order statistics, attention is limited to $r \leq 6$ because of reasonable doubt on validity of the model for $r \geq 7$. The models that are used are nested, so likelihood ratio test using deviance statistic is used for selecting the most appropriate candidate model. The fit of the data to candidate models is diagnosed using graphical tools. The upper tail of the distribution is studied by modelling extreme returns using quantile function. Profile log-likelihood plots are used for inferential purpose on the parameters of the best model. The demonstration of the block approach in this dissertation is discussed in Section 4.1 of Chapter 4.

3.4 Applying the Peaks-Over-Threshold (POT) approach

According to Coles (2001), alternative of block approach in analyzing extremes is the use of POT which attains the advantage of utilizing as much as possible of the available information. This approach is used in two phases of analysis in this dissertation. Looking at the second subsection of Section 4.2 in Chapter 4, the POT approach is used in modelling average minimum daily temperature. The average daily temperatures over the winter season have been negated for the purpose of using the duality principle in modelling the occurrences of coldest days.

The second subsection of Section 4.2 demonstrates the application of POT approach in modelling influence of average maximum daily temperature above 22 °C on the demand of electricity over time. Modelling framework includes the piecewise linear regression model that is fitted for the purpose of explaining the impact of maximum temperature or hottest days on electricity demand. The target parameters of the GPD models are estimated in this dissertation using both the MLE technique and Bayesian approach. The corresponding uncertainties of the parameter estimates are assessed using the bootstrap resampling approach.

3.4.1 Threshold selection

The POT approach desires the use of a sufficiently high threshold in fitting the GPD models. This section deals with the manner in which a suitably high time-varying threshold is selected in this dissertation. The main requirement of the threshold is that it should be sufficiently high so that it does not violate the asymptotic property of the GPD (Coles, 2001; Sugahara et al., 2009). There are several techniques in literature that are useful for choosing the threshold values as discussed in Section 2.3 of Chapter 2. In this dissertation, the preference in choosing the thresholds is given to the extremal mixture models due to their capability of assessing and quantifying the uncertainties associated to chosen threshold value.

In Section 4.2 of Chapter 4, a sufficiently high threshold is chosen and used for fitting univariate GPD to average minimum daily temperature. Initially we fit a penalized regression cubic smoothing spline as a time-varying threshold which is given in equation (2.17) of Chapter 2, where x_t is the average minimum daily temperature, f_t is a cubic spline and λ is a smoothing parameter that is chosen based on the Generalized Cross Validation (GCV) approach (Wang, 2011). We then extract the residuals (excesses) above this threshold and fit a fixed threshold above the positive residuals (excesses above the time-varying threshold) using non-parametric extremal mixture models that are given in equation (2.20) of Chapter 2. The procedure that is followed in fitting the extremal mixture models is to fit a kernel density to the bulk model and a GPD fitted to upper end-point of the model. The whole process of determining the sufficiently high threshold using the extreme value mixture models is accomplished using the evmix package that is authored by Hu and Scarrott (2013).

3.4.2 Declustering excesses above the threshold

The time series that is considered for the POT approach in this dissertation has shown evidence of short-range dependence and heavy seasonality. These are the limitations that negatively influence the results of the fitted models. These limitations are avoided in this dissertation by declustering the excesses above the threshold so that the GPD is fitted to cluster maxima instead of original time series.

The declustering process requires the calculation of the extremal index that is calculated in this dissertation using equation (2.26) of Chapter 2. The automatic declustering algorithm which is the interval estimator method of Ferro and Segers (2003) is then used for declustering excesses above the sufficiently high threshold. This is given as:

$$\eta_\tau = \frac{2\left[\sum_{i=1}^{N-1}(T_i - 1)\right]^2}{(N-1)\sum_{i=1}^{N-1}(T_i-1)(T_i-2)}, \tag{3.1}$$

where τ is a sufficiently high threshold and T_i denotes interexceedance times. The extremal index, θ_τ measures the amount of clustering and $0 \le \theta_\tau \le 1$, where $\frac{1}{\theta_\tau}$ is the limiting mean cluster size (Smith, 1989). The following steps have been taken during declustering:

1. Empirical rule is used in terms of defining the exceedance clusters;
2. extreme residuals are noted out of all the clusters;
3. independence in the cluster maxima is assumed with the conditional GPD of the residuals over the sufficiently high threshold;
4. the GPD is then fitted to maxima within the clusters.

This declustering algorithm is applied in Section 4.2 of Chapter 4 where excesses above the threshold are declustered for the purpose of fitting the stationary GPD to cluster maxima. In Section 4.3 of Chapter 4, the declustering approach is again applied before fitting the stationary and non-stationary point process models to cluster maxima.

3.5 Applying the time-homogeneous and non-homogeneous point process approach

Point process models are used in two modelling frameworks of this dissertation. Both the time-homogeneous and non-homogeneous models are fitted to cluster maxima in two cases.

In the first subsection of Section 4.3 in Chapter 4, point process models are fitted towards modelling average hourly temperatures, whereas the second subsection demonstrates the application of the same approach towards modelling maximum daily temperature. Non-homogeneity in modelling location and scale parameters in this dissertation is done using linear, exponential and quadratic transformations. Threshold selection and declustering approaches that are used in this section are discussed in Section 3.4. Stationary point process models in both cases are further used for calculating frequencies and intensities of the occurrence of extremely hot days. This involves applications of the orthogonal and reparameterization techniques that are demonstrated in the last subsection of Section 4.3 of Chapter 4.

Chapter 4

Analyses

Chapter 4 presents the analyses and interpretation of results using methods that were discussed in Chapter 3. The analysis is done in three broad approaches of EVT. Section 4.1 demonstrates the use of annual block maxima approach in which the GEVD for r largest order statistics is fitted. In Section 4.2, the use of POT approach is demonstrated in two phases of analysis. In the first case, the stationary GPD is fitted to average minimum daily temperature with the threshold that is determined using the extremal mixture models, followed by declustering the data and fitting the GPD to cluster maxima. In the second case, threshold approach is further used towards modelling average maximum daily temperatures above 22 °C. The GPD is fitted to cluster maxima after declustering the data. The piecewise linear regression model is used for modelling and explaining the influence of average maximum daily temperature above the upper reference point 22 °C on electricity demand. Section 4.3 demonstrates the use of point process approach towards modelling average hourly temperature as well as modelling maximum daily temperature in South Africa. The orthogonal and reparameterized methods are used for calculating the frequency and intensity of the occurrence of extremely hot days.

4.1 Modelling maximum daily temperature using r largest order statistics

4.1.1 Introduction

Ever since the twenty first century, daily activities of individuals and industries in South Africa are presently relying on the daily consumption of electricity load (Inglesi, 2010). The influence of electricity consumption towards economic growth is evidenced in Ferguson, Wilkinson and Hill (2000), where the economic growth of over 100 countries is established to be highly correlated to the usage of electricity. This is initially based on the fact that electricity is used in both processes of producing and consuming goods and services and consequently, thorough studies in the energy sector and precise modelling of electricity demand are essential for growth of economies of all the countries in general (Hyndman and Fan, 2010; Payne, 2010). For that and many other reasons, the demand for electricity has been studied by several researchers for over three decades as efforts to guard against the consequences of underestimation and overestimation which may lead to a huge cost (Hyndman and Fan, 2010; Payne, 2010).

Hahn, Meyer-Nieberg and Pickl (2009) classify the statistical methodologies that are frequently used in the energy sector into regression analysis, time series, state space and Kalman filtering. However, the limitation that is commonly encountered amongst such techniques is a symmetrical distribution which often leads to undependable estimates since some observations are far from the end-points of the distributions (Soares and Scotto, 2004; Byström, 2005). This creates a problem which can be addressed by the use of EVT. The crucial role that daily temperature plays on the demand of electricity is based on the fact that, the heating systems are used in winter to keep warm whereas the air conditioning appliances are desired in summer to keep cool (Muñoz, Sánchez-Úbeda, Cruz and Marín, 2010).

In this section, use of block maxima approach of EVT towards modelling average maximum daily temperature in South Africa is discussed. The EVT is discussed in Gencay and Selcuk (2004) as a field of mathematical statistics that comprises a rich family of non-symmetric techniques that are suitable for modelling the recurrence behavior of the fat (heavy) tailed distributions. This section is aimed at fitting the GEVD to the annual block maxima in an effort to assess the asymptotic behavior of r largest order statistics within blocks of equal lengths. The meteorological data that are used comprise daily maximum temperatures that are collected by SAWS and supplied by Eskom for the period 2000 to 2010. The objective of this section is to determine the frequency of occurrence of hottest days so that the effects of maximum temperature on the demand of electricity can be quantified.

4.1.2 Models

The focus is on modelling that is based on annual maxima approach of classical EVT. We discuss and demonstrate application of GEVD for r largest order statistics given in equation (2.8) of Chapter 2. The application is on modelling average maximum daily temperature. The models that are used are nested with the target parameters that are determined with the use of MLE method. Likelihood ratio and Wald tests are usually used in selecting the most suitable model. The likelihood ratio test using the deviance statistic is used in this section for selecting the best model. Goodness of fit is done using graphical tools. Upper tail of the distribution is then studied by modelling extreme returns using quantile function.

4.1.3 Data

The data that are used in this section are average maximum daily temperatures that are collected by the SAWS over the period January 2000 to August 2010 and supplied by Eskom. Only the observations for the non-winter seasons have been selected for the purpose of modelling occurrence of hottest days. The details on the manipulation of data to meet requirements of this analysis are given in Section 3.2 of Chapter 3. Time series plot of the data for the non-winter period is given in Figure 4.1.

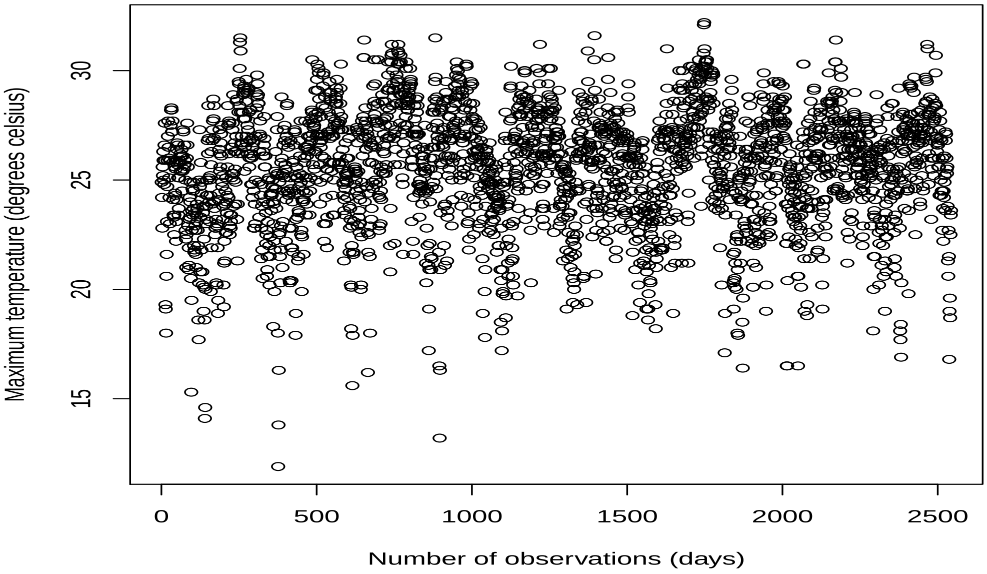


Figure 4.1: Plot of average maximum daily temperature (°C). Only data for the period September to April $(2000-2010)$ are included.

4.1.4 Empirical results and discussion

4.1.4.1 Modelling using the r largest order statistics

In this section we replace the notation of the order statistics r by k for the purpose of being consistent with the diagnostic plots. The pattern in the time series plot given in Figure 4.1 reveals evidence of a seasonal component within the considered observations. The results of the k largest asymptotic order statistics model that is given in equation (2.57) of Chapter 2 and fitted to 10 largest annual temperatures over a period of 11 years are summarized in Table 4.1. Several values of the maximized negative $\log -$likelihood estimates (λ_i), for $i = 1, \ldots, 10$ have led to MLEs of target parameters as illustrated in Table 4.1 with standard errors in parentheses. However, the attention is limited to $k \leq 6$ order statistics because of reasonable doubt on validity of the model for $k \geq 7$.

Despite the diagnostic plots, the doubt is also justified by a significant decrease in the values of λ_i for $i \geq 6$. The observable trend on the values of λ_i is a decrease at $k \leq 2$ that is followed by an increase at $3 \leq k \leq 5$. As anticipated, the notable fact is a decrease in the standard errors of $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ as the values of k increase, which according to Coles (2001) corresponds to the increase in precision of the model. However, the decrease in standard errors seems to be colossal at $k \leq 4$ and then stabilizes at $k \geq 5$. The estimates of a location parameter seem to be stable at $k \geq 2$ whereas the estimates of the scale and shape parameters stabilize at $k \geq 5$.

The sufficient evidence concerning validity of Weibull family of distributions is revealed by the estimates of the shape parameter for all values of k. The graphical diagnostic tools (P-P and Q-Q plots) for assessing accuracy in the fit of annual maximum temperature to the k largest asymptotic GEVD are given for each value of k in Figures 4.2, 4.3 and 4.4. The reasonably good fit of the GEVD to average maximum daily temperatures is achieved if the pattern of the dots are approximately linear. Looking at Figures 4.2 and 4.3, the Q-Q plots for $k = 2, 3$ and 4 display the dots that are near linear. There is doubt on the fit of the model at $k \geq 5$. The model fitted for $k = 4$ with $\lambda_4 = -8.5826$ seems to be the one that possesses a reasonably good fit. The graphical diagnostics for $k = 4$ are given in Figure 4.5. To justify appropriateness of Weibull class of distributions as a proper model for average maximum daily temperatures in South Africa, confidence intervals for ξ_i are estimated. For example, the confidence interval for ξ considering $k = 1$ is given by:

$$\hat{\xi} \pm z_{\alpha/_2} \times (\text{standard error}) \quad \Rightarrow \quad -0.6848 \pm 1.96 \times 0.2376 = (-1.1505; -0.2191).$$

This leads to conclude that at a 95% level of confidence, the value of ξ is expected to be enclosed within interval $(-1.1505; -0.2191)$. Confidence intervals for ξ are then estimated for all values of k, and it is established that all upper limits are negative values and hence enclose the point estimate $\hat{\xi}$. This justifies that the Weibull family of distributions is relevant for modelling maximum daily temperature in South Africa. The results of confidence interval are included in column 6 of Table 4.1.

Table 4.1: Maximized log-likelihoods λ_i, parameter estimates and standard errors (in parentheses) of k largest order statistics model fitted to the temperatures in South Africa with different values of k.

k	λ_i	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$	95% CI for ξ
1	-12.5255	30.8813 (0.3141)	0.9551 (0.2643)	-0.6848 (0.2376)	(-1.1505; -0.2191)
2	-15.9670	31.0993 (0.2243)	0.7540 (0.1133)	-0.6206 (0.1784)	(-0.9703; -0.2709)
3	-11.3280	31.1311 (0.1660)	0.6220 (0.0688)	-0.4803 (0.1173)	(-0.7102; -0.2504)
4	-8.5826	31.1190 (0.1576)	0.6107 (0.0616)	-0.4515 (0.1062)	(-0.6597; -0.2433)
5	-0.8328	31.1405 (0.1427)	0.5743 (0.0551)	-0.4226 (0.0887)	(-0.5965; -0.2487)
6	-11.1508	31.1421 (0.1358)	0.5571 (0.0535)	-0.3963 (0.0824)	(-0.5578; -0.2348)

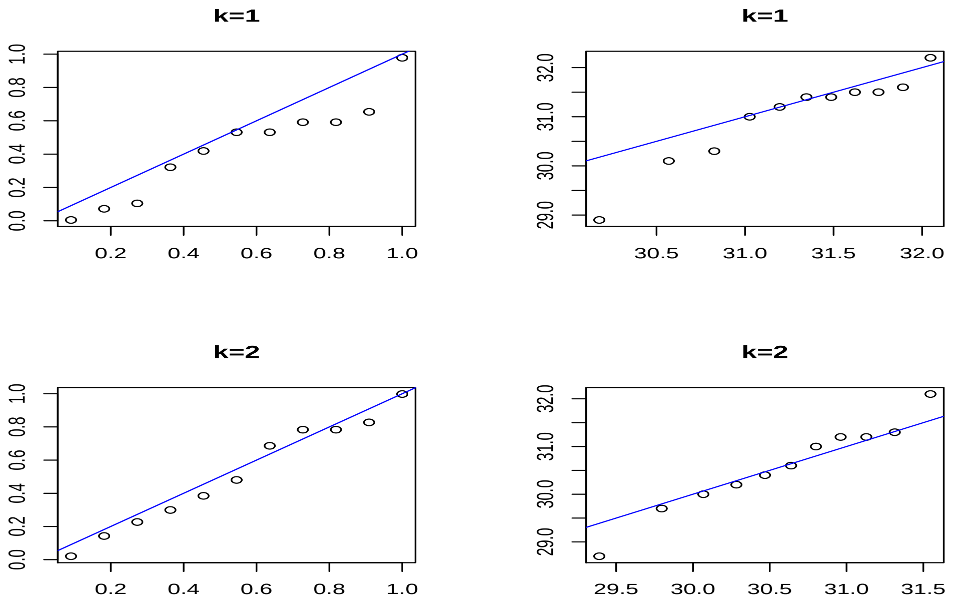


Figure 4.2: Diagnostic plots illustrating the fit of the data (annual average maximum temperature) to the GEVD for r largest order statistics model with $k = 1$ and $k = 2$: From left to right, P-P plot for $k = 1$ (top left panel), Q-Q plot for $k = 1$ (top right panel), P-P plot for $k = 2$ (bottom left panel) and Q-Q plot for $k = 2$ (bottom right panel).

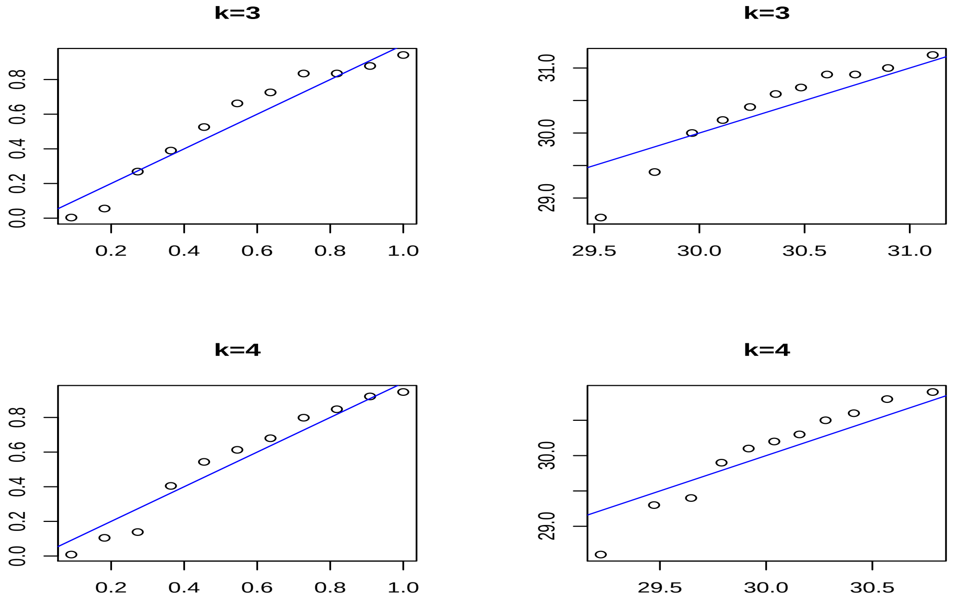


Figure 4.3: Diagnostic plots illustrating the fit of the data (annual average maximum temperature) to the GEVD for r largest order statistics model with $k = 3$ and $k = 4$. From left to right, P-P plot for $k = 3$ (top left panel), Q-Q plot for $k = 3$ (top right panel), P-P plot for $k = 4$ (bottom left panel) and Q-Q plot for $k = 4$ (bottom right panel).

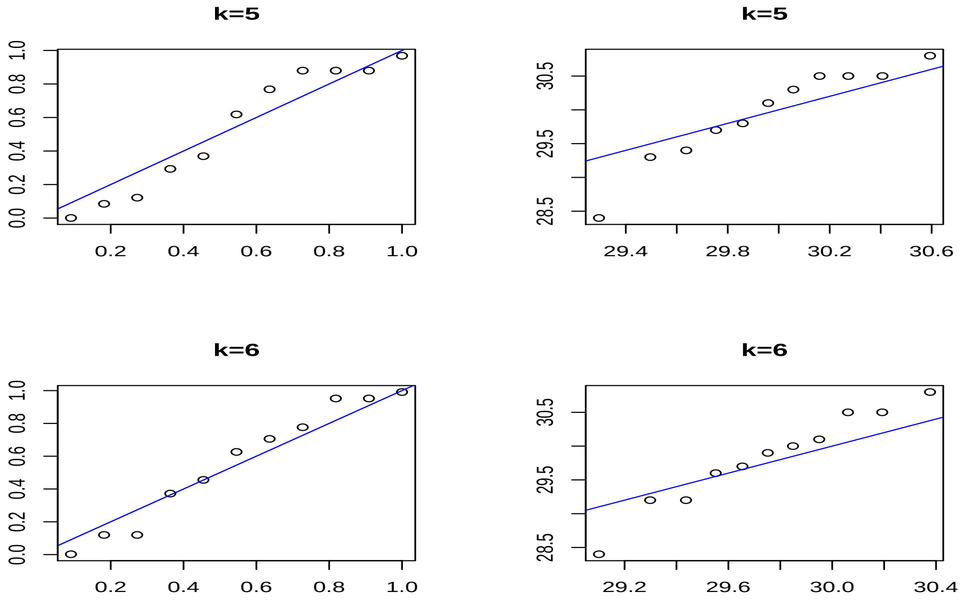


Figure 4.4: Diagnostic plots illustrating the fit of the data (annual average maximum temperature) to the GEVD for r largest order statistics model with $k = 5$ and $k = 6$. From left to right: P-P plot for $k = 5$ (top left panel), Q-Q plot for $k = 5$ (top right panel), P-P plot for $k = 6$ (bottom left panel) and Q-Q plot for $k = 6$ (bottom right panel).

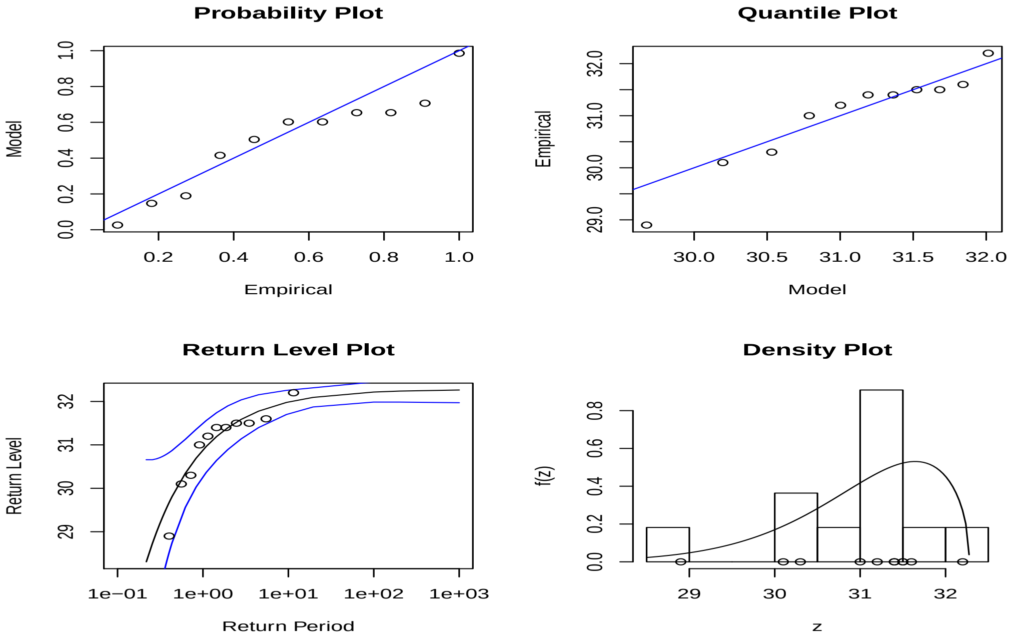


Figure 4.5: Diagnostic plots illustrating the fit of the data (annual average maximum temperature) to the GEVD for r largest order statistics model with $k = 4$. From left to right: P-P plot for $k = 4$ (top left panel), Q-Q plot for $k = 4$ (top right panel), return level plot for $k = 4$ (bottom left panel) and density plot for $k = 4$ (bottom right panel).

4.1.4.2 Return level estimation and model selection

The estimated values in Table 4.1 are used for diagnosing the fit and hence selecting best model. This is done using deviance statistic that is given in equation (2.62) of Chapter 2. Table 4.2 summarizes the results.

Table 4.2: Deviance statistics and $p-$values.

$D_{(1,2)}$	$D_{(2,3)}$	$D_{(3,4)}$	$D_{(4,5)}$	$D_{(5,6)}$
-6.88	9.23	5.49	15.50	-20.64
0.009999	0.009999	0.009999	0.009999	0.009999

Table 4.3: Return values for 5, 10, 25, 50 and 100 years.

Return levels $5-$years	$10-$years	$25-$years	$50-$years	$100-$years
31.78	31.98	32.15	32.24	32.30

The point estimate $\hat{x}_p$ for the 5, 10, 25, 50 and 100 years return levels corresponding to return period $1/p$ are calculated using equation (2.6) of Chapter 2 and the outcomes are summarized in Table 4.3. For example, looking at $25-$years return level, the level $x_{0.04} = 32.15$ is anticipated to be surpassed on average once every 25 years. For accuracy in the estimation of return level and parameters in the model for $k = 4$, profile log$-$likelihoods for μ, σ and ξ are given in Figures 4.6, 4.7 and 4.8, respectively. Table 4.2 shows comparison of models using deviance statistic given in equation (2.62) in Chapter 2. For example, the deviance statistic for comparing $\lambda(k_1)$ and $\lambda(k_2)$ is calculated as $D_{(1,2)} = 2(-15.9670 - (-12.5255)) = -6.88$.

Similar to Soares and Scotto (2004), the deviance statistics in this study are significant when comparing the log$-$likelihoods among $D_{(2,3)}$, $D_{(3,4)}$ and $D_{(4,5)}$. The comparison is invalid for $D_{(1,2)}$ and $D_{(5,6)}$. The test for selecting the best model is conducted at 1% level of significance for which $\chi_1^2 = 6.64$. For $D_{(2,3)}$ and $D_{(4,5)}$, $\lambda(k_3)$ and $\lambda(k_5)$ are rejected since $D_{(2,3)} > 6.64$ and $D_{(4,5)} > 6.64$. It is therefore clear that $D_{(3,4)} = 5.49 < 6.64$ reveals failure to reject $\lambda(k_4)$ which implies the validity of $k = 4$ as an order statistic that possesses a reasonably good fit as suggested by the graphical diagnostic tools. The Q-Q and the P-P plots given in Figure 4.2 show that the Weibull family is the appropriate distribution for the maximum temperatures and also that $k = 4$ possesses the most reasonably good fit of the model out of the 6 order statistics. Equation (2.6) of Chapter 2 is then used in estimating the future return levels (extreme quantiles) for the different return periods as illustrated in Table 4.4. For example, the 90th quantile is calculated as follows:

$$x_{0.1} = 31.1190 + \frac{0.6107}{0.4515}\left\{1 - \left[-\ln(0.90)\right]^{0.4515}\right\} \quad \Rightarrow \quad x_{0.1} = 31.98 \approx 32.0.$$

The frequency of values that are above the predicted tail quantile $x_{0.1} = 32.0$ are then enumerated and concluded to be 2. For the observed frequency of exceedances, we obtain $0.1 \times 44 = 4.4 \approx 4$ where 44 is the number of observations available in 11 years.

Table 4.4: Tail and quantiles estimation for the GEVD for annual maxima with $k = 4$.

Quantiles	Temperatures (x_p)	Observed no. of exceedances	GEVD no. of exceedances
90th	32.0°C	4	2
95th	32.1°C	2	1
97.5th	32.2°C	1	0
99th	32.3°C	0	0

Table 4.4 presents a summary of estimated tail quantiles at different tail probabilities. The tail quantiles (temperatures) are given in column 2. The observed frequency of temperatures that are greater than the predicted end-point quantiles are shown in column 3 while column 4 shows the corresponding number estimated using the GEVD.

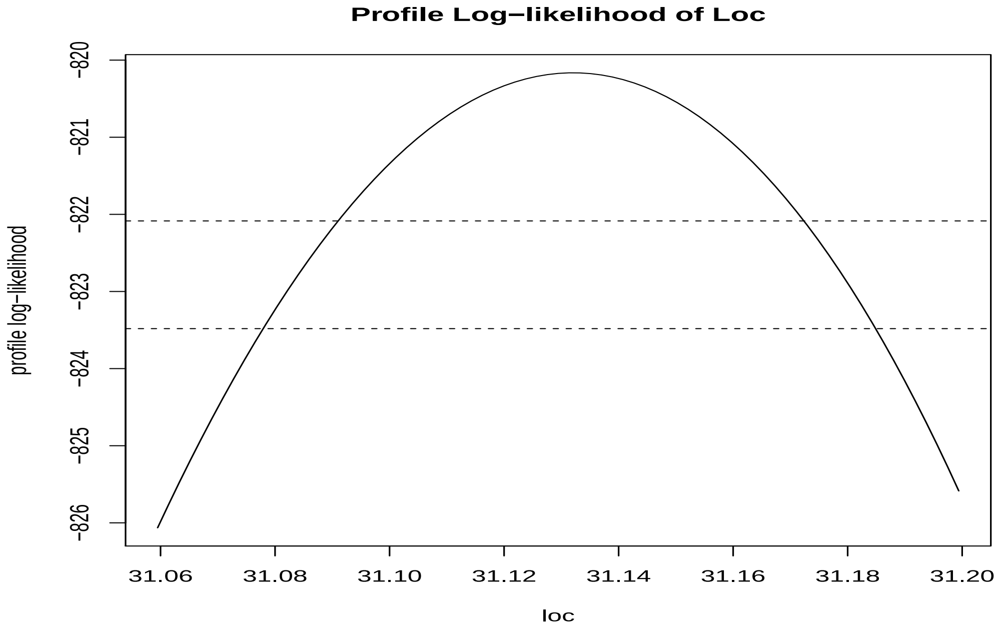


Figure 4.6: Profile log −likelihood for the location parameter μ using estimates of the model with $k = 4$.

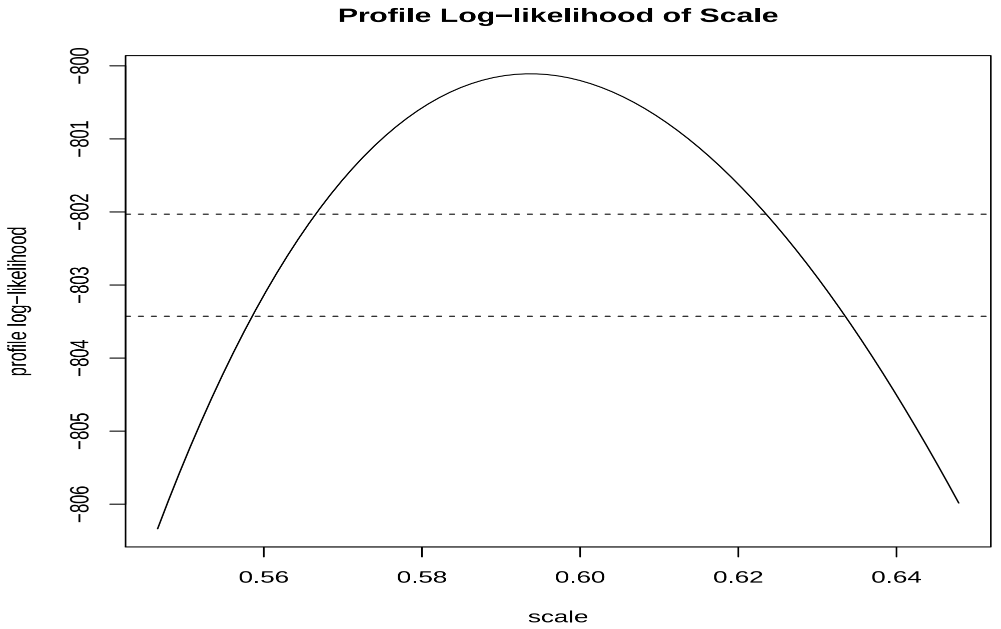


Figure 4.7: Profile log −likelihood for the scale parameter σ using estimates of the model with $k = 4$.

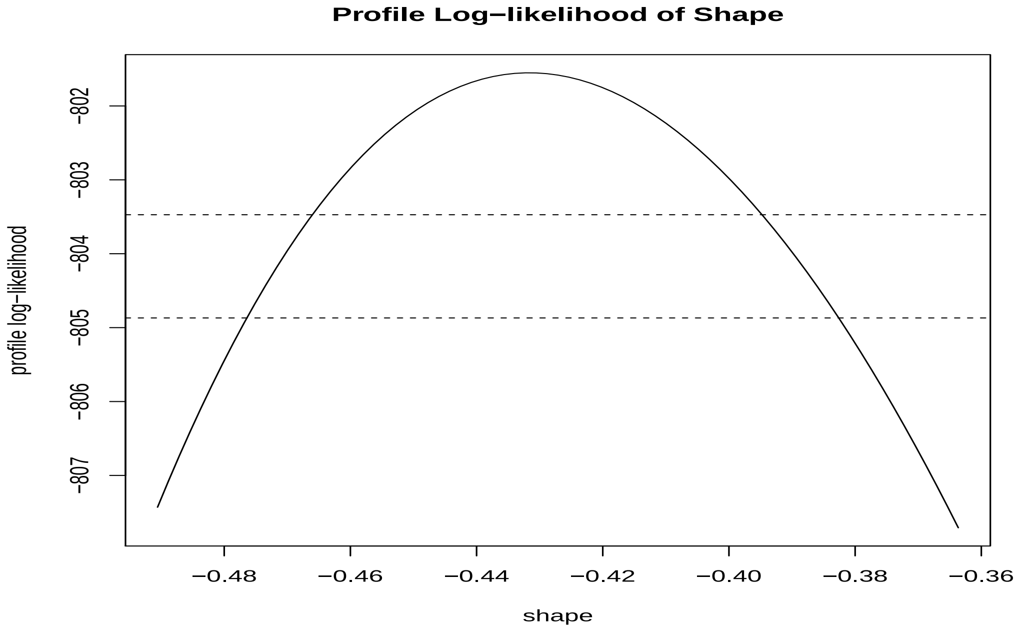


Figure 4.8: Profile log −likelihood for the shape parameter ξ using estimates of the model with $k = 4$.

4.1.5 Conclusion

Modelling frequency of occurrence of hottest days is crucial in energy sector for the load forecasters to assess the impacts of maximum temperature on the demand of electricity load. The maximum daily temperature in South Africa is modelled in this dissertation for the non-winter season (September to April of each of the 11 years) using the GEVD for k largest order statistics. The modelling involves 6 values of k amongst which $k = 4$ is established to be the one that shows a reasonable fit in comparison to the rest. This is established through the use of the diagnostic tools which are the P-P and the Q-Q plots that are given in Figures 4.3 and 4.5 respectively. The choice of $k = 4$ as the best is done using the deviance statistic. The asymptotic behavior of k reveals the Weibull family as an appropriate distribution that can be used for modelling maximum daily temperatures in South Africa. The justification for this validity is done by estimating the confidence interval for the shape parameters that is found within negative intervals for all values of k. The right end-point of the k largest model is assessed using the quantile function. Statistical inference is done by analysing several return levels corresponding to the return periods as well as plotting the profile log −likelihoods for the parameters of the best model.

4.2 Modelling using generalized Pareto distribution

4.2.1 Modelling average minimum daily temperature using extreme value theory with a time-varying threshold

The present section discusses and apply GPD with time-varying threshold in modelling average minimum daily temperature. Winter data with duality aspect is used in order to model frequency of occurrence of extremely cold days.

4.2.1.1 Introduction

Challenges that are continually encountered by electricity sector in South Africa are concerns to both the industrial and domestic electricity users. Inglesi (2010) and Strengers (2012) support the fact that, uncertainties in the energy sector in South Africa result in periodic blackouts and higher electricity prices, which are challenges for South African economy since 2008. In the presence of such, accurate prediction of electricity demand and detailed statistical modelling in the energy sector may help planners and decision makers in planning thoroughly (taking cautious decisions) in the presence of uncertainties. Accurate statistical modelling in the energy sector taking temperature into account guards against economic risks since electricity sector is viewed as one of supreme weather-sensitive sectors of South Africa and any other country in general (Chikobvu and Sigauke, 2012). Amaral, Souza and Stevenson (2008) and Hyndman and Fan (2010) consider daily temperature as one of most relevant weather variables to consider in forecasting electricity load and it is therefore established as a major driver of electricity demand.

This section presents the use of POT technique of EVT towards modelling temperature beyond a suitably high time-varying threshold. The problem to be addressed in this section is the desire to assess the extent to which POT approach can be used in modelling temperature. This section is primarily aimed at fitting a stationary GPD on average minimum daily temperature that is dated from January 2000 to August 2010. Extreme temperatures naturally tend to occur in groups or clusters, leading to a problem that is addressed by declustering of exceedances at a high threshold. Declustering is implemented in this study as a result of dependence and seasonality. Uncertainty in parameter estimates of GPD is assessed with bootstrap resampling approach.

4.2.1.2 Generalized Pareto distribution

Smith (1989), Coles (2001) and Sugahara et al. (2009) consider POT as a better alternative to block maxima or block minima approach due to its capability to use as much as possible of the available information. Suppose that $X_1, X_2, \ldots, X_n$ denote a stationary process of i.i.d. temperatures with a univariate marginal distribution function F. The unified stationary GPD model is given in equation (2.12) of Chapter 2 and in this dissertation we reparameterized scale parameter as follows: $\theta = \log(\sigma_\tau)$, and τ is the threshold that is exceeded by the excesses.

4.2.1.3 Empirical results

Figure 4.9 is a time series plot of negated average minimum daily temperature with a time-varying threshold which is a penalized cubic smoothing spline. We selected smoothing parameter λ based on GCV criterion. The estimated value is $\hat{\lambda} = 0.1719$. An initial threshold is set at zero after fitting time-varying threshold and only positive observations (excesses) above zero are considered. We then determine a sufficiently high threshold by fitting a non-parametric extremal mixture model and exceedances are declustered using Ferro and Segers (2003) intervals estimator method. Figure 4.10 shows threshold estimation using a non-parametric extremal mixture model where a kernel density is fitted to the bulk model and a GPD fitted to the upper end-point of the model. The estimated threshold is $\hat{\tau} = 1.26$.

After declustering we get 176 cluster maxima and the extremal index is estimated as 0.622. A comparative analysis is done by using a direct POT approach. A fixed threshold is determined by using nonparametric extremal mixture models. The exceedances are declustered and a GPD is then fitted to the cluster maxima. Using this direct application of POT, the estimated extremal index is found to be 0.163. This means that exceedances tend to occur in groups of $1/0.163$ which is approximately 6. With this modelling approach, exceedances occur in groups of 1/0.622, which is approximately 1.6. This shows that the cluster maxima are approximately independent. Table 4.5 shows maximum likelihood estimates of GPD parameters. Scale and shape parameters are found to be $\hat{\theta} = 0.3798(0.1121)$ and $\hat{\xi} = -0.1464(0.0837)$ with standard errors in parentheses respectively. Negative value of the EVI reveals evidence that W_ξ belongs to Weibull family of GPD and hence the appropriateness of the Weibull family of distributions in modelling average minimum daily temperatures. The diagnostic plots in Figure 4.11 show an appropriate fit of the GPD. The estimated rate of excess is determined as the frequency of cluster maxima divided by the frequency of the values above the time-varying threshold which is $\frac{176}{1168} \approx 0.15068$. Now, since $\theta = \log(\sigma_\tau)$ this implies that the scale parameter is estimated as $\hat{\sigma}_\tau = e^{\hat{\theta}} = e^{0.3798} \approx 1.46199$. The GPD diagnostic tools are used to assess the goodness of the GPD fit. Both probability and quantile plots show a good fit.

Table 4.5: Parameter estimates of GPD fitted to cluster maxima of the average minimum daily temperature.

	Value	Standard error
$\hat{\theta}$	0.3798	0.1121
$\hat{\xi}$	-0.1464	0.0837
Log. lik	-217.0758	
AIC	438.1515	

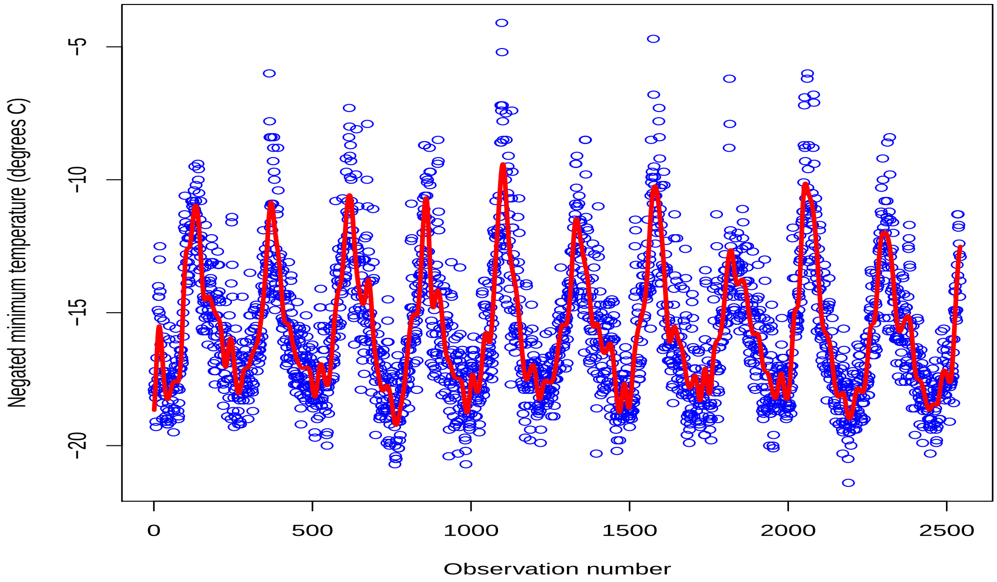


Figure 4.9: A time series plot of the negated average minimum daily temperatures with a time-varying threshold. Blue dots shows the negated observations and the red line shows the smoothing parameter.

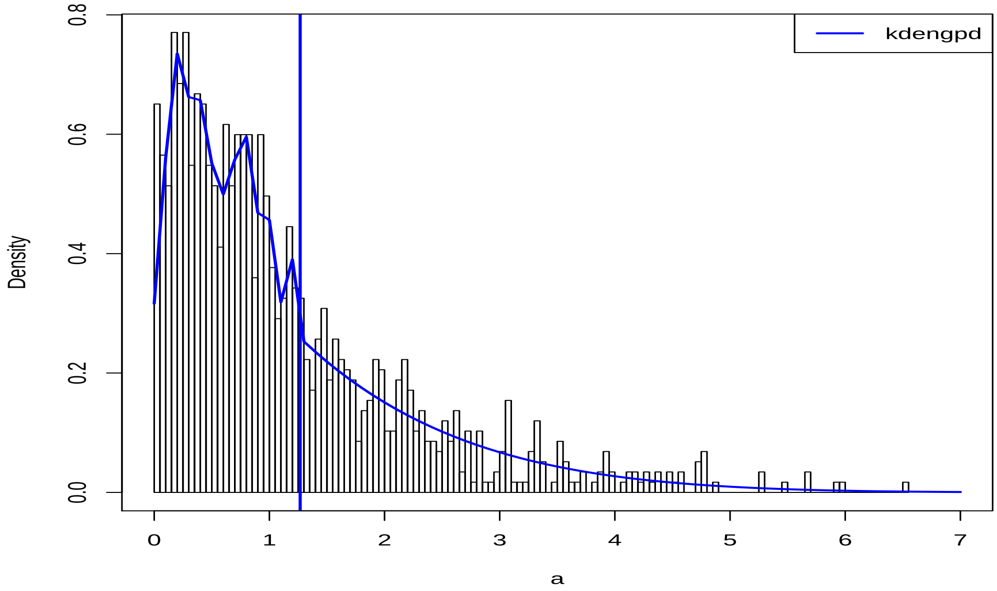


Figure 4.10: Threshold estimation using a non-parametric extremal mixture model where a kernel density is fitted to the bulk model and a GPD fitted to the tail of the distribution ($\hat{\tau} = 1.26$).

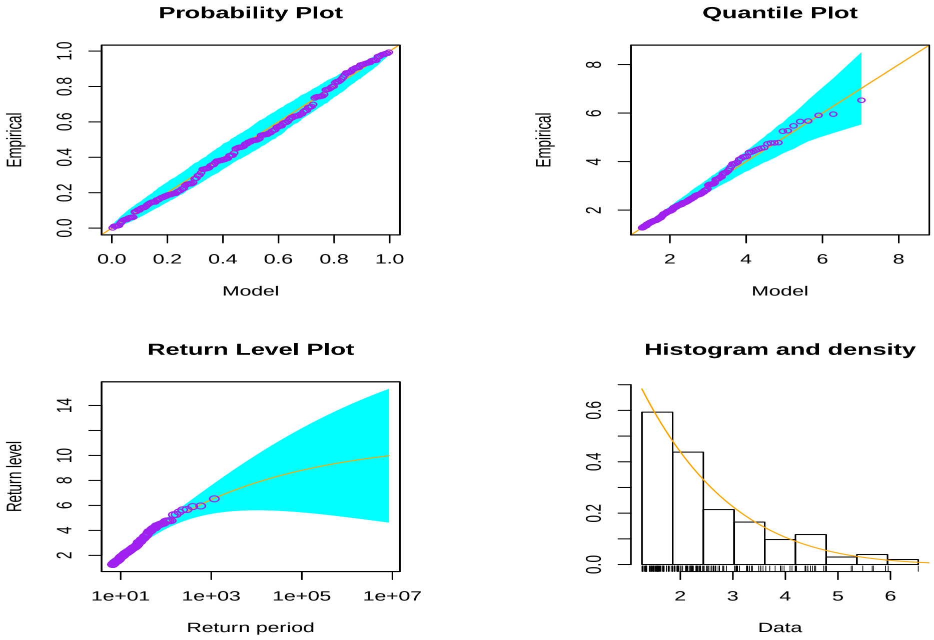


Figure 4.11: Diagnostic plots illustrating the fit of the cluster maxima of the average minimum daily temperature to the GPD. From left to right: P-P plot (top left panel), Q-Q plot (top right panel), return level plot (bottom left panel) and density plot (bottom right panel).

4.2.1.4 Estimating return levels

Extreme quantiles are estimated using the m-observations return level that is given in equation (2.14) of Chapter 2. The return period is in days since the data is average minimum daily temperature and x_m represents the minimum value of x, we expect to see in m observations, with ϕ_τ as the probability of exceeding the threshold τ as discussed in Southworth and Heffernan (2013a). A summary of the estimated return levels up to 10 years is given in Table 4.6.

Table 4.6: Estimating return levels of the average minimum daily temperature.

Number of observations (m)	Years	$\hat{x}_m$ ($\hat{\tau} = 1.2675$)
365	1	5.70
731	2	6.24
1096	3	6.53
1461	4	6.72
1826	5	6.87
2192	6	6.98
2557	7	7.08
2922	8	7.16
3287	9	7.23
3653	10	7.29

4.2.1.5 Parameter uncertainty using parametric bootstrap

The bootstrap resampling approach is relevant in EVT and other statistical models as an automatic computer-based method for assessing uncertainty in the estimation of parametric and non-parametric models (Efron and Tibshirani, 1994; Kyselỳ, 2008; Li, Shao, Xu and Cai, 2010). Though it is theoretically familiar that estimation of shape parameter is more uncertain compared to the scale parameter (Sugahara et al., 2009), this approach is used in this dissertation to assess the uncertainty in both shape and scale parameters. Table 4.7 shows that the biases for $\hat{\theta}$ and $\hat{\xi}$ are given by 0.0135 and -0.0175 respectively, with the standard deviations as 0.1037 and 0.0739 respectively.

Table 4.7: GPD parameter uncertainty using parametric bootstrap resampling approach.

	$\hat{\theta}$	$\hat{\xi}$
Original	0.3798	-0.1464
Bootstrap mean	0.3933	-0.1638
Bias	0.0135	-0.0175
SD	0.1037	0.0739
Bootstrap median	0.3920	-0.1596

4.2.1.6 Conclusion

We discussed a modelling approach which uses a penalized cubic smoothing spline as a time-varying parameter after which we extract excesses. Non-parametric extremal mixture models were used to determine a threshold which is suitably high. Positive residuals above the sufficiently high threshold were declustered using intervals estimator method of Ferro and Segers (2003) and a GPD was fitted to cluster maxima. All the diagnostic plots show a good fit of the GPD. Uncertainty is assessed in parameter estimates using bootstrap resampling approach. Bootstrapping output shows that both estimates have small biases and standard deviations that are less or closer to zero.

4.2.2 Modelling the influence of average maximum daily temperature above 22 degrees celsius on electricity demand

4.2.2.1 Introduction

Over several decades, the use has been made of various statistical approaches including regression analysis and classical time series towards studying the influence of daily average temperature on electricity demand in the energy sector (Muñoz et al., 2010; Mirasgedis, Sarafidis, Georgopoulou, Lalas, Moschovits, Karagiannis and Papakonstantinou, 2006; Hekkenberg, Benders, Moll and Uiterkamp, 2009). The use of EVT models gains preference over other symmetrical distributional approaches due to the capability of EVT in modelling tail behavior of fat (heavy) distributions (Gencay and Selcuk, 2004; García-Cueto and Santillán-Soto, 2012). Chikobvu and Sigauke (2013) model the impact of temperature on average daily electricity demand in South Africa using a piecewise linear regression together with the GEVD and then indicated that weather variables such as temperature and solar radiation are used as predictors in the energy forecasting models.

This subsection discusses modelling of average daily temperature in South Africa using stationary GPD for the period 2000 – 2010. The impact of temperature on the consumption of electricity is explained in this section using piecewise linear regression model that is discussed in Byström (2005) and Chikobvu and Sigauke (2013).

4.2.2.2 Models

Temperature modelling in this subsection is done in two stages. Firstly, a piecewise linear regression model is fitted so that the influence of maximum temperature or hottest days on electricity demand can be explained. The piecewise linear regression model is given by

$$\mathrm{ADED} = \beta_0 + \beta_1 \max(0, t_h - \mathrm{ADT}) + \beta_2 \max(0, \mathrm{ADT} - t_c) + \varepsilon_t, \tag{4.1}$$

where ADED is the (Average Daily Electricity Demand), ADT is the (Average Daily Temperature), t_c and t_h are temperatures that distinguish between the winter, weather neutral and the summer sensitive periods (Vieth, 1989; Byström, 2005; Moral-Carcedo and Vicens-Otero, 2005; Chikobvu and Sigauke, 2013, among others). The target parameters to be estimated are β_0, β_1 and β_2 where ε_t is the random error term that is distributed with $\varepsilon_t \sim N(0, \sigma_t^2)$. The reference winter and summer temperatures t_c and t_h as discussed in Chikobvu and Sigauke (2013), are estimated using the Multivariate Adoptive Regression Splines (MARS) algorithm that is developed by Friedman (1991) and found to be 18 °C and 22 °C respectively.

In the second modelling stage, stationary GPD that is given in equation (2.12) of Chapter 2 is fitted to average daily temperature above 22 °C after determining the threshold using the penalized regression cubic smoothing splines and the non-parametric extremal mixture models that are discussed in Wang (2011) and Hu and Scarrott (2013) as given in equations (2.17) and (2.20) in Chapter 2, respectively. The data are then declustered using automatic declustering algorithm of Ferro and Segers (2003) in texmex R package as discussed in Southworth and Heffernan (2013b). The stationary GPD is then fitted to cluster maxima and the estimated parameters are used to estimate high quantiles and then model the effect/influence of temperature above 22 °C and demonstrate how they affect marginal increases in ADED. For the inference on extreme quantiles, use is made of the m-observation return levels that are discussed in Southworth and Heffernan (2013a) as given in equation (2.14) in Chapter 2.

4.2.2.3 Data

We use the average daily temperatures that are captured in South Africa for the period $2000 - 2010$. The non-winter observations above a reference temperature of 22 °C are considered for the purpose of analyzing impacts of hottest days on ADED as indicated in Figure 4.12.

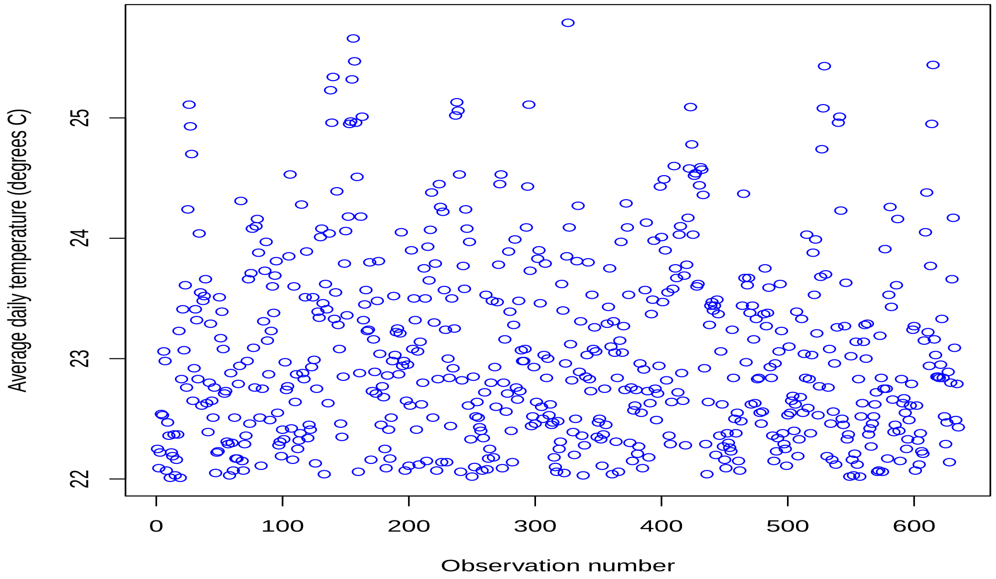


Figure 4.12: Average Daily Temperature (ADT) above the reference temperature of 22 °C.

4.2.2.4 Empirical results

Considering piecewise linear regression output, the model that is given in equation (4.1) identifies winter sensitive, weather neutral and the summer sensitive periods. The parameter estimates together with their standard errors in parentheses are $\hat{\beta}_0 = 23932(150.32)$, $\hat{\beta}_1 = 263(20.74)$ and $\hat{\beta}_3 = 138(38.53)$. The parameter estimates are then substituted in equation (4.1) which results in equation (4.2). Residual analysis was done and the error terms were found to be approximately normally distributed and fluctuated randomly around mean zero. Equation (4.2) is then used for describing the impacts of maximum temperature above 22 °C on the ADED. We adopt the piecewise linear regression model that is developed by Chikobvu and Sigauke (2013) as given in equation (4.2).

$$\text{AD}\hat{\text{E}}\text{D} = 23932 + 263\max(0, 22 - \text{ADT}) + 138\max(0, \text{ADT} - 18). \tag{4.2}$$

For the average daily temperatures that exceed a maximum reference point of 22 °C, equation (4.2) reduces to equation (4.3)

$$\text{AD}\hat{\text{E}}\text{D} = 23932 + 138\max(0, \text{ADT} - 18). \tag{4.3}$$

Equation (4.3) explains that, if temperature rises by 1 °C, (for example, from 22 °C to 23 °C), electricity demand is expected to increase marginally by 138MW. From the estimated electricity demand model in equation (4.3), we further determine the impact of ADT on electricity demand. If ADT rises by 1 °C, (say from 25 °C to 26 °C), the rate of increase on ADED is given by:

$$\begin{aligned}\%\text{increase in } \mathrm{E}(\widehat{\mathrm{ADED}}) &= \frac{\mathrm{E}(\widehat{\mathrm{ADED}}|\mathrm{ADT}=26\,^{\circ}\mathrm{C})-\mathrm{E}(\widehat{\mathrm{ADED}}|\mathrm{ADT}=25\,^{\circ}\mathrm{C})}{\mathrm{E}(\widehat{\mathrm{ADED}}|\mathrm{ADT}=25\,^{\circ}\mathrm{C})}\times 100\% \\ &= \frac{25036-24898}{24898}\times 100\% \\ &= 0.55\%.\end{aligned}$$

This implies that a 1 °C increase on average daily temperature results into about 0.55% rise on average daily electricity demand.

Focusing on the use of GPD towards modelling average daily temperature, a sufficiently high threshold is determined using a non-parametric extremal mixture model that is given in equation (2.20) of Chapter 2, where a kernel density is fitted to bulk model and a GPD fitted to the tail of distribution. A sufficiently high threshold is found to be $\hat{\tau} = 24.1$. Graphical threshold diagnostic tools that are given in Figure 4.14 are the Q-Q plots of normalised interexceedance times against standard exponential quantiles. The extremal index $(\theta = 0.5675)$ is calculated prior to the use of automatic declustering algorithm and there are 34 resulting cluster maxima. The GPD is finally fitted to cluster maxima (positive excesses above the threshold) using MLE method. Table 4.8 shows maximum likelihood estimates of GPD parameters fitted to cluster maxima. The scale and shape parameters are found to be $\hat{\theta} = -0.2662(0.2686)$ and $\hat{\xi} = -0.3497(0.2153)$ with standard errors in parentheses, respectively. Now, since $\theta = \log(\sigma_\tau)$ this implies that the scale parameter is estimated as $\hat{\sigma}_\tau = e^{\hat{\theta}} = e^{-0.26624649} \approx 0.76625$.

Table 4.8: Parameter estimates of GPD fitted to cluster maxima of the ADT above 22 °C.

	Value	Standard error
$\hat{\theta}$	-0.2662	0.2686
$\hat{\xi}$	-0.3479	0.2153
Log. lik	-13.05907	
AIC	30.11813	

Table 4.9: GPD parameter uncertainty using parametric bootstrap for model fitted to the ADT above 22 °C.

	$\hat{\theta}$	$\hat{\xi}$
Original	-0.2662	-0.3497
Bootstrap mean	-0.1817	-0.4744
Bias	0.0845	-0.1247
SD	0.2598	0.2134
Bootstrap median	-0.1778	-0.4590

Sugahara et al. (2009) explain that estimation in parametric models is associated with uncertainties which need to be assessed so that valid inferential conclusions can be drawn. One of the techniques in this regard is the bootstrap resampling approach that is used in this section as a tool for assessing uncertainty in the estimated values of the GPD fitted to cluster maxima. Table 4.9 shows that the biases for $\hat{\theta}$ and $\hat{\xi}$ are given by 0.0845 and -0.1247, with the standard deviations as 0.2598 and 0.2134, respectively. The negative value of EVI reveals evidence that W_ξ belongs to the Weibull family of GPD which is bounded from above. It is important to justify that W_ξ belongs to a Weibull class of distributions before concluding its appropriateness. This is achieved by estimating the confidence interval for ξ which is expected to be enclosed within the interval $(-\infty;0)$. A 95% confidence interval for ξ is estimated as follows:

$$\hat{\xi} \pm z_{\alpha/_2} \times (\text{standard error}) \quad \Rightarrow \quad -0.3497 \pm 1.96 \times 0.2153 = (-0.7717; 0.0723).$$

This leads to conclude that at a 95% level of confidence, the shape parameter ξ lies within the interval $(-0.7717; 0.0723)$. This confidence interval contains zero, meaning appropriateness of the Gumbel family of distributions, or a combination of Gumbel and Weibull families of distributions towards modelling average daily temperature in South Africa. Diagnostic plots in Figure 4.13 supports an appropriate fit of the GPD to cluster maxima. The Bayesian estimation is used for estimating target parameters with the posterior distributions that are given in Figure 4.15. The upper bound of W_{ξ} is calculated and found to be

$$\text{Upper bound} = \hat{\tau} - \frac{\hat{\sigma}_{\tau}}{\hat{\xi}} \Rightarrow 24.1 - \frac{0.76625}{-0.3497045} \approx 26.29.$$

This implies that for any temperature increase above 26.3 °C, there will not be any increase in average daily electricity demand. Return levels for several return periods are determined using equation (2.14) of Chapter 2. The results are summarized in Table 4.10. For example, a 95^{th} quantile is associated to 20−year return level and is found to be $x_{0.05} = 24.6\,^{\circ}\mathrm{C}$. Of the 636 observations in the data, the number of observations that exceed estimated tail quantile $x_{0.05} = 24.6\,^{\circ}\mathrm{C}$ are 411. The observed number of exceedances associated to 95^{th} quantile are determined by using $0.05 \times 427 = 21.35 \approx 21$, where 427 is the number of observations above the maximum reference temperature $t_h = 22\,^{\circ}\mathrm{C}$. The corresponding marginal increase for a rise in temperature from 22 °C to $x_{0.05} = 24.6\,^{\circ}\mathrm{C}$ is finally given by $(24.6 - 22) \times 138 = 358MW$, where 138 is the marginal increase in electricity demand for a 1 °C increase above 22 °C.

Table 4.10: In-sample evaluation of estimated tail quantiles at different probabilities (number of exceedances).

Quantiles	Tail probabilities	Temperature (x_p)	Marginal increase in demand (MW)
90^{th}	0.1	24.2 °C	–
95^{th}	0.05	24.6 °C	55.2
97.5^{th}	0.025	24.9 °C	41.4
99^{th}	0.01	25.3 °C	55.2
99.5^{th}	0.005	25.5 °C	27.6
99.9^{th}	0.001	26 °C	69
99.99^{th}	0.0001	26.4 °C	55.2
99.999^{th}	0.00001	26.6 °C	27.6
99.9999^{th}	0.000001	26.8 °C	27.6
99.99999^{th}	0.0000001	26.9 °C	13.8
99.999999^{th}	0.00000001	27 °C	13.8
99.9999999^{th}	0.000000001	27 °C	–

A summary of the impact of average daily temperature on daily electricity demand is given in Table 4.10 which shows the estimated tail quantiles in column 1 and the corresponding tail probabilities in column 2. The m−year return levels and the estimated tail quantiles (temperature) are shown in columns 3 and 4 respectively. The resultant calculations of the marginal increases in electricity demand for the tail quantiles in column 3 are shown in column 4. It is also noted that the tail quantile converges to 27 °C, which is significantly larger than the upper bound of W_{ξ}.

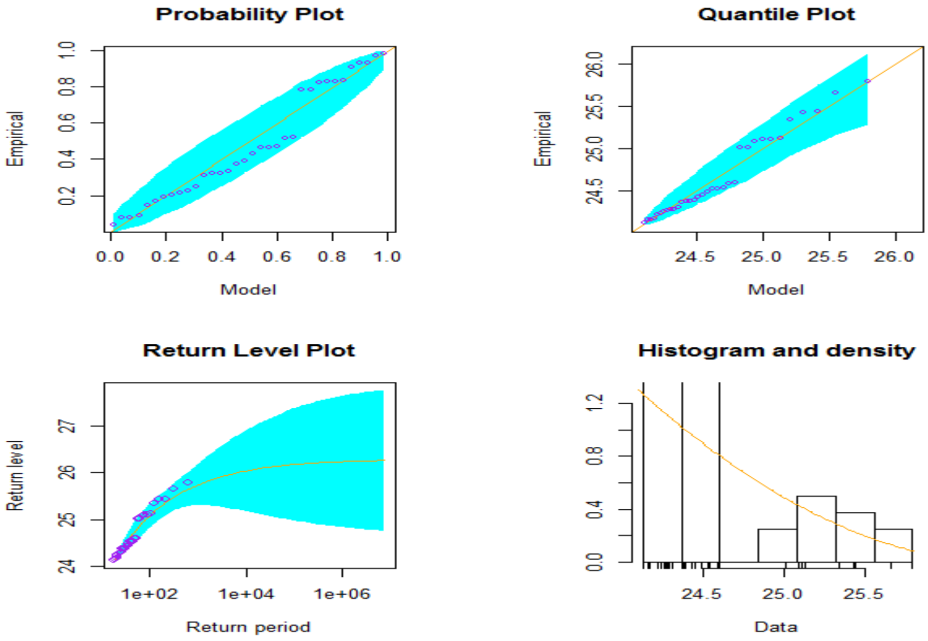


Figure 4.13: Diagnostic plots illustrating the fit of cluster maxima of ADT above 22 °C to the GPD. From left to right: P-P plot (top left panel), Q-Q plot (top right panel), return level plot (bottom left panel) and density plot (bottom right panel).

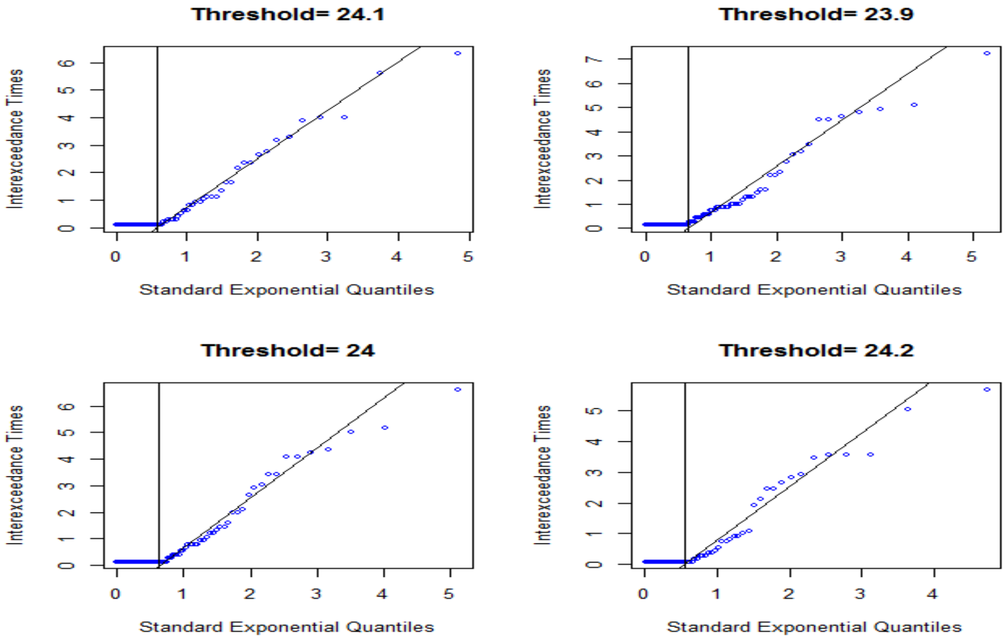


Figure 4.14: Q-Q plots of normalised interexceedance times against standard exponential quantiles. Vertical line shows the $(1-\hat{\theta})$ quantile; sloping line has gradient $1/\hat{\theta}$. Data are ADT above 22 °C that are simulated from a max-autoregressive process with extremal index $\theta = 0.5675$.

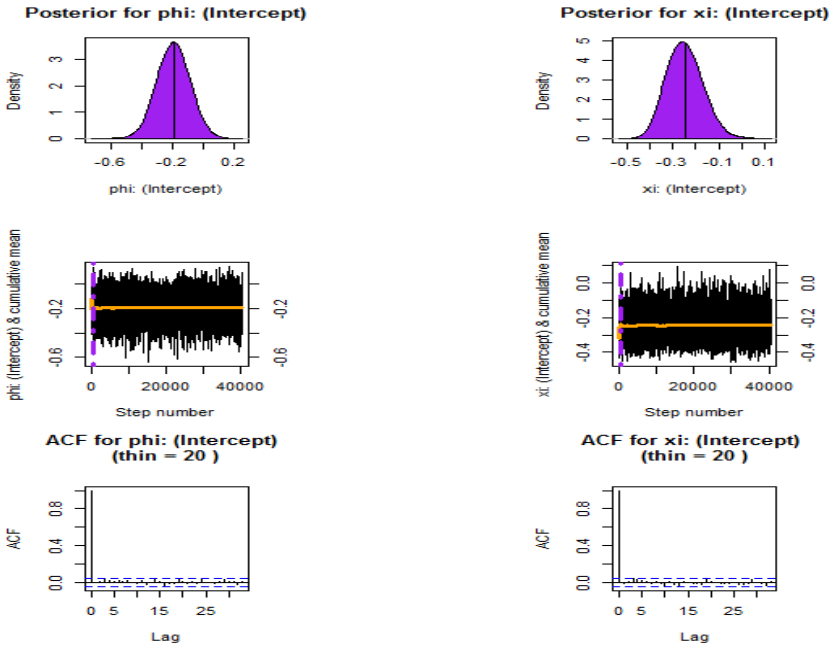


Figure 4.15: Posterior distributions of the GPD parameters fitted to the cluster maxima of the ADT above 22 °C.

4.3 Modelling temperature using point process characterization of extremes

4.3.1 Point process characterization of average hourly temperature

4.3.1.1 Introduction

In Section 4.2, the analyses are based only on GPD modelling of average daily temperature. Point process approach gains preference in comparison to both the block and POT methods due to the fact that, instead of considering separately the time of occurrence of threshold excesses and residuals above the corresponding sufficiently high threshold, point process approach combines the two and treat them as one process based on the bivariate plot (Cox, 1965; Deheuvels, 1983; Ogata and Katsura, 1986; Smith, 2003; Daley and Vere-Jones, 2007; Northrop and Jonathan, 2011; Embrechts, Klüppelberg and Mikosch, 2013, among others). The present section deals with point process modelling approach in which both the frequency and intensity of extreme temperature are analyzed in this dissertation.

4.3.1.2 Models

In this subsection we discuss point process models that are used for modelling temperature in this dissertation. Time-homogeneous and non-homogeneous point process models are discussed and fitted to cluster maxima of average hourly temperature and maximum daily temperature using MLE method. Homogeneous point process model is given in equation (4.4).

$$\begin{aligned} \text{Model} \quad M_0 : \mu(t) &= \mu_0 \\ \sigma(t) &= \sigma_0 \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.4}$$

where all parameters are treated as constants (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007). The non-homogeneous point process models that we discuss involve linear, exponential and quadratic transformations which are as follows:

$$\begin{aligned} \text{Model} \quad M_1 : \mu(t) &= \mu_0 + \mu_1 t \\ \sigma(t) &= \sigma_0 \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.5}$$

where the model is linear in location parameter only (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007),

$$\begin{aligned} \text{Model} \quad M_2 : \mu(t) &= \mu_0 + \mu_1 t \\ \sigma(t) &= exp(\sigma_0 + \sigma_1 t) \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.6}$$

where the model is linear in location parameter and exponential in scale parameter (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007),

$$\begin{aligned} \text{Model} \quad M_3 : \mu(t) &= \mu_0 \\ \sigma(t) &= exp(\sigma_0 + \sigma_1 t) \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.7}$$

where the model is exponential in scale parameter only (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007),

$$\begin{aligned} \text{Model} \quad M_4 : \mu(t) &= \mu_0 + \mu_1 t + \mu_2 t^2 \\ \sigma(t) &= \sigma_0 \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.8}$$

where the model is quadratic in location parameter only (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007) and

$$\begin{aligned} \text{Model} \quad M_5 : \mu(t) &= \mu_0 \\ \sigma(t) &= \sigma_0 + \sigma_1 t + \sigma_2 t^2 \\ \xi(t) &= \xi_0, \end{aligned} \tag{4.9}$$

where the model is quadratic in scale parameter only (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007). The use of exponential transformation of the scale parameter is to have parameters as non-negative values. Non-stationary modelling of shape parameter is not as easy as for location and scale parameters (Kotz and Nadarajah, 2000; Coles, 2001; De Haan and Ferreira, 2007). This dissertation is limited to point process modelling approach that models non-stationarity in location and scale parameters only.

4.3.1.3 Data

One of the basic requirements for point process modelling approach as emphasized in Deheuvels (1983) and Hu (2013), is the high frequency data. In this section, the data comprises nominally 93 480 observations of Average Hourly Temperature (AHT) that were collected by SAWS and provided by Eskom for the period 2000 – 2010. The features of row data are summarized in the time series plot that is given in Figure 4.16 which shows among others, existence of time series components such as seasonality and trend.

Smith (1989) and Goubanova and Li (2007) argue that seasonality component is associated to short-range dependence since a cold day is probably followed by consecutive cold days and that the same holds also for other seasons. Dependence feature leads to a weakness whereby the data are clustered or grouped together, thereby violating i.i.d. condition and hence gives misleading results. This limitation is avoided by declustering the data at high-level exceedance. In this dissertation, temperature observations are declustered using Ferro and Segers (2003) interval estimator method. Positive residuals are extracted and plotted in Figure 4.17 top left panel and does not reveal any evidence of a trend component.

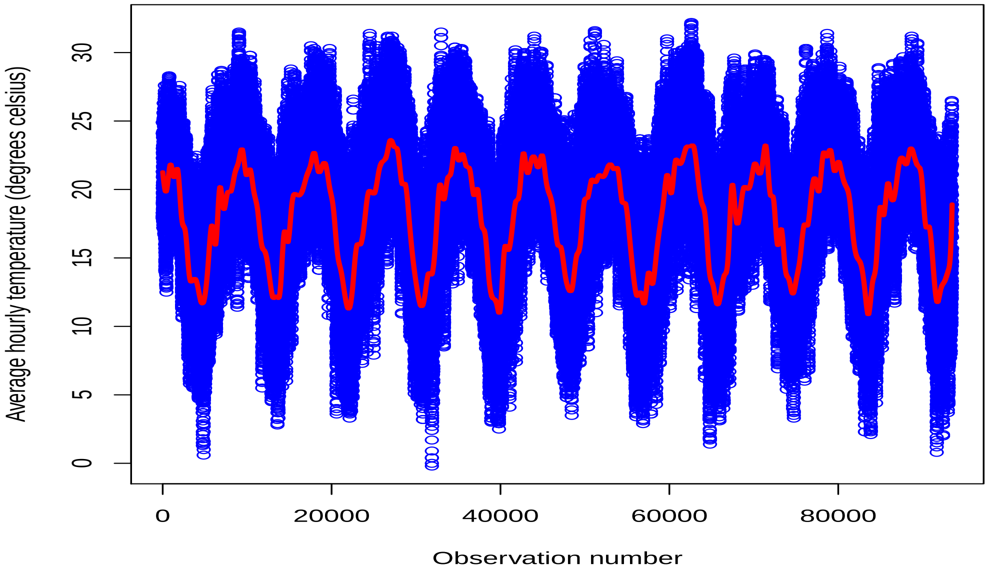


Figure 4.16: A time series plot of the average hourly temperatures with a time-varying threshold. The blue dots show the observations and the red line shows the smoothing parameter.

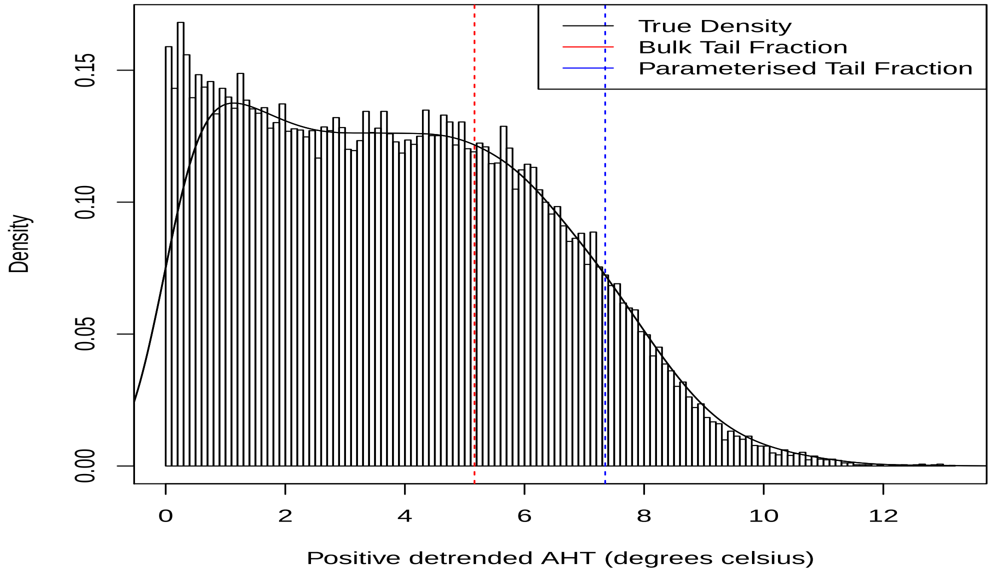


Figure 4.17: Threshold estimation using a parametric extremal mixture model with (truncated) Weibull distribution fitted to the bulk model and a GPD fitted to the upper tail of the distribution. The tail fraction that is based on the bulk model is shown in red whereas the parameterised tail fraction is indicated in blue. The corresponding thresholds are respectively indicated by the vertical dashed lines.

4.3.1.4 Threshold selection

Point process modelling approach desires the use of sufficiently high threshold that is obtainable based on one of the threshold selection tools as discussed in Chapter 2. Data are initially detrended using a penalized regression cubic smoothing spline given in equation (2.17) of Chapter 2, where in this case, y_i denotes average hourly temperature and λ is a smoothing parameter. Initial threshold is set at zero so that extreme value mixture model can be fitted to positive residuals. This is done for the purpose of choosing a sufficiently high fixed threshold.

Parametric version of extremal mixture models is used in this section to determine the sufficiently high threshold whereby the truncated Weibull distribution is fitted to bulk model and GPD fitted to upper tail of the distribution. The cumulative distribution functions for bulk based tail fraction model and parameterized based tail fraction model are given in equations (2.18) and (2.19) in Chapter 2, respectively.

Figure 4.17 shows a plot of threshold estimation using parametric extremal mixture models in which the histogram indicates (in red), tail fraction that is based on bulk model together with a corresponding threshold of $\hat{\tau} = 5.164$. The parameterized based tail fraction is indicated by a broken line in blue, with the threshold of $\hat{\tau} = 7.349$. Estimates of the target parameters determined using MLE method are given in Table 4.11 which shows that the shape parameters for both models reveal appropriateness of Weibull class of distributions towards modelling AHT in South Africa. Estimates of scales, shapes and thresholds possess small values of standard errors. Diagnostic plots for both the bulk based and parameterized tail fraction models confirm good fit of extremal mixture models as shown in Figures 4.19 and 4.20.

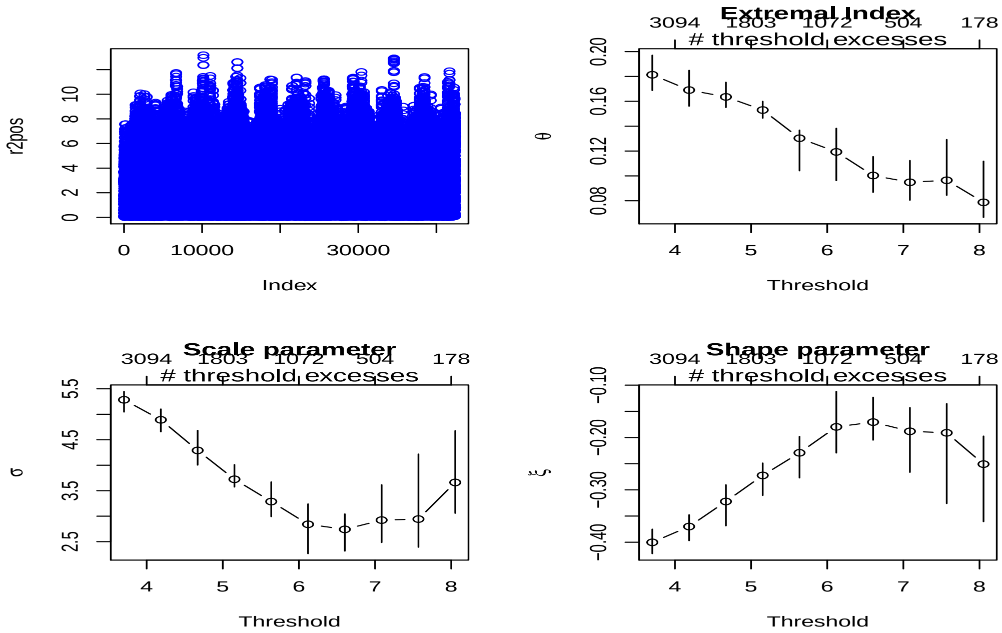


Figure 4.18: From left to right: Positive residuals (top left panel), threshold stability plot for the extremal index (top right panel), threshold stability plot for the scale parameter (bottom left panel) and threshold stability plot for the shape parameter (bottom right panel).

Detailed assessment of thresholds is based on Figure 4.21 which shows several threshold diagnostic plots. Looking at top left panel, threshold stability plot for scale parameter shows estimate of scale parameter plotted against several thresholds. Threshold stability plot at top right panel shows estimate of shape parameter plotted against several thresholds. The main idea of threshold stability plots as emphasized in Heffernan and Southworth (2013), is to assess a range of several thresholds for invariance of extremal index, which then support results of the parametric extremal mixture model in this dissertation. Figure 4.18 shows threshold stability plots of extremal index, scale and shape parameters at the top right and the bottom panels, respectively. At the bottom panel of Figure 4.21 is the mean residual life plot that is helpful in determining threshold of GPD model, unlike the threshold stability plot which assesses thresholds for individual parameter. Unfortunately, the mean residual life plot does not gain paramount preference in determining threshold values due to difficulty in its interpretation (Coles, 2001). Figure 4.22 shows Q-Q plots for normalized inter-exceedance times against standard exponential quantiles. This is useful for assessing several thresholds for estimated value of extremal index.

Table 4.11: Maximum likelihood estimates of the parametric extremal mixture model.

	Bulk based tail fraction model	Parameterized based tail fraction model
log lik.	-94031.36	-93537.13
$\hat{\tau}$	5.1635 (0.0007)	7.3492 (0.0018)
$\hat{\xi}$	-0.2480 (0.0035)	-0.1378 (0.0132)
$\hat{\sigma}_\tau$	2.0285 (0.0183)	1.1723 (0.0241)
ϕ_τ	0.3311	0.0928

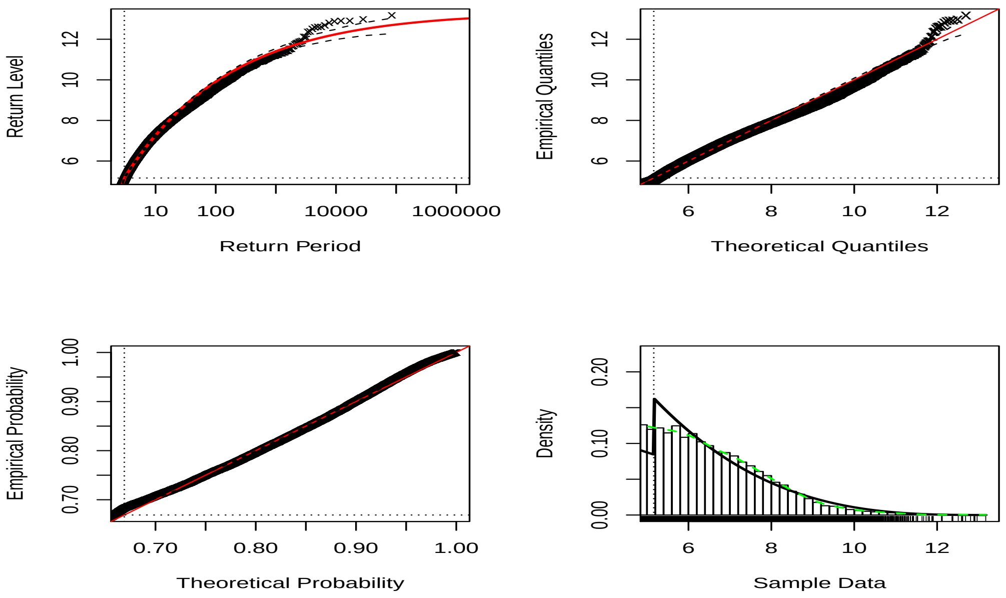


Figure 4.19: Diagnostic plots for the fit of the (truncated) Weibull distribution to the bulk model. From left to right: Return level plot (top left panel), Q-Q plot (top right panel), P-P plot (bottom left panel) and density plot (bottom right panel).

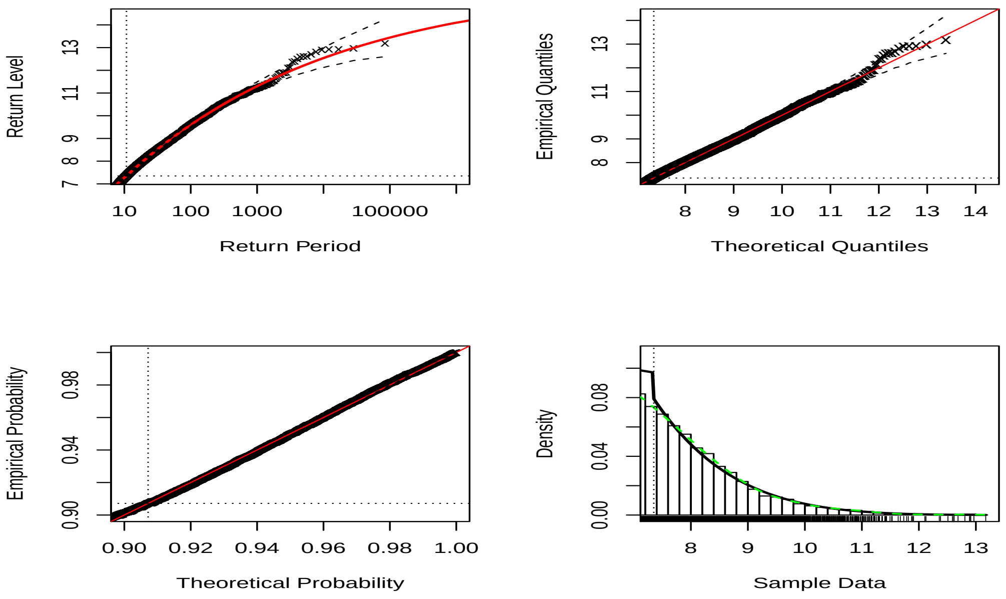


Figure 4.20: Diagnostic plots for the parameterized tail fraction. From left to right: Return level plot (top left panel), Q-Q plot (top right panel), P-P plot (bottom left panel) and density plot (bottom right panel).

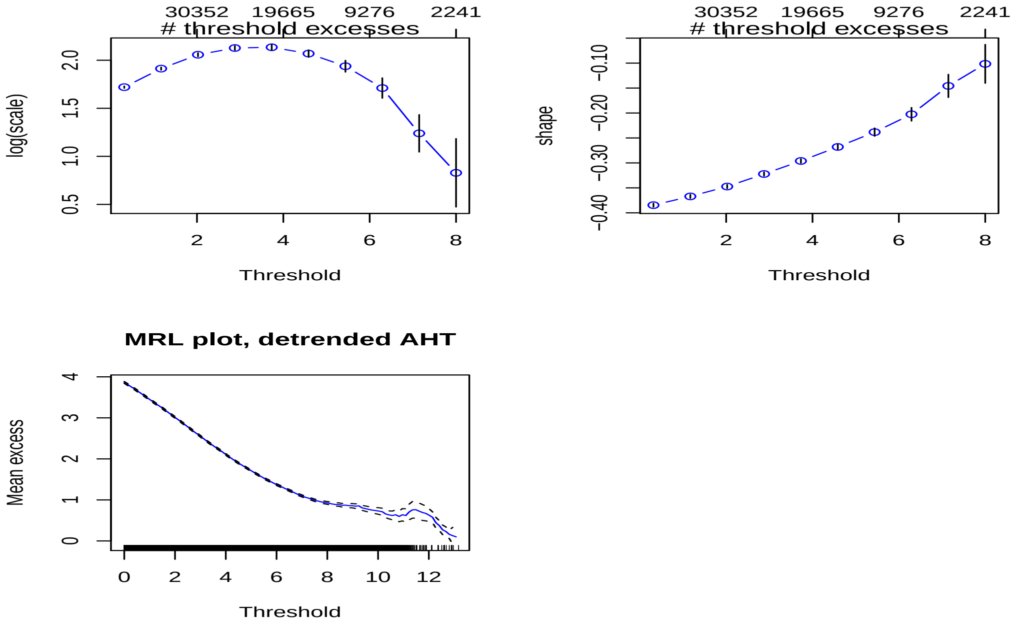


Figure 4.21: Threshold diagnostic plots. From left to right: Scale parameter threshold diagnostic plot (top left panel), Shape parameter threshold diagnostic plot (top right panel) and Mean residual life plot (bottom panel). Data is the average hourly temperature.

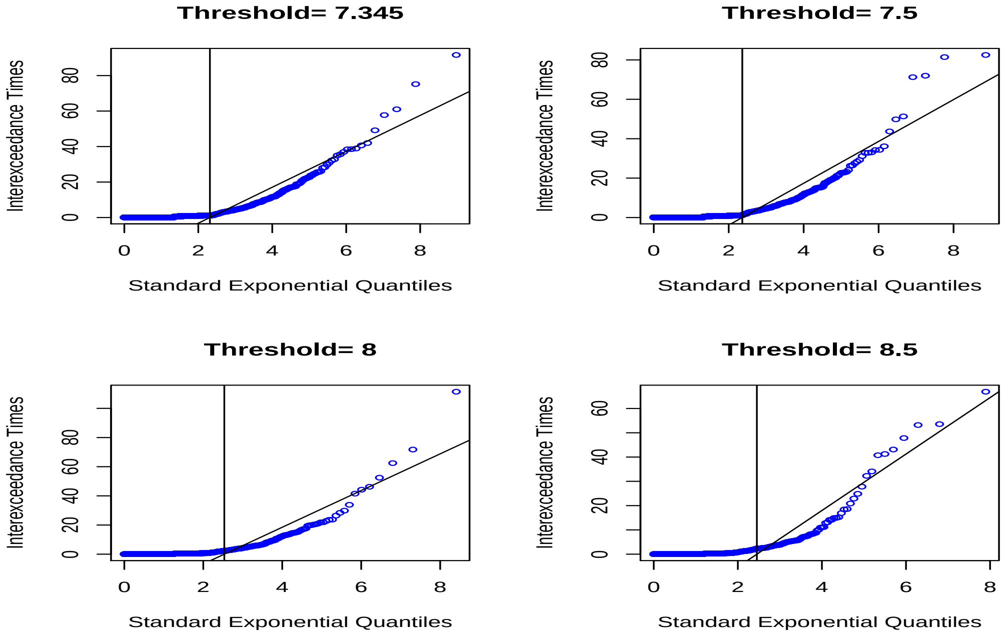


Figure 4.22: Q-Q plots for normalised interexceedance times against standard exponential quantiles. Vertical line shows the $(1-\hat{\theta})$ quantile; sloping line has gradient $1/\hat{\theta}$. Data are average hourly temperatures in degrees celsius that are simulated from a max-autoregressive process with extremal index $\theta = 0.09878$.

4.3.1.5 Empirical results

A sufficiently high threshold $\hat{\tau} = 7.349$ is used for declustering and fitting point process models to the detrended cluster maxima (positive residuals) that are plotted in Figure 4.17 top left panel. Extremal index is estimated prior to declustering and found to be $\theta = 0.09878$, meaning that the average cluster size is $1/0.09878 = 10.1235 \approx 11$. This implies that exceedances occur in groups of approximately 11. For original series of length 42 409, there are 3 957 exceedances above the threshold. There are 378 approximately independent clusters, implying some feature of serial dependence in the data since threshold excesses are not independent. After declustering, the GPD is fitted to cluster maxima as shown in Figure 4.23. Estimates of target parameters of the time-homogeneous and non-homogeneous point process models that are fitted towards modelling AHT are given in Table 4.12 with respective standard errors in parenthesis.

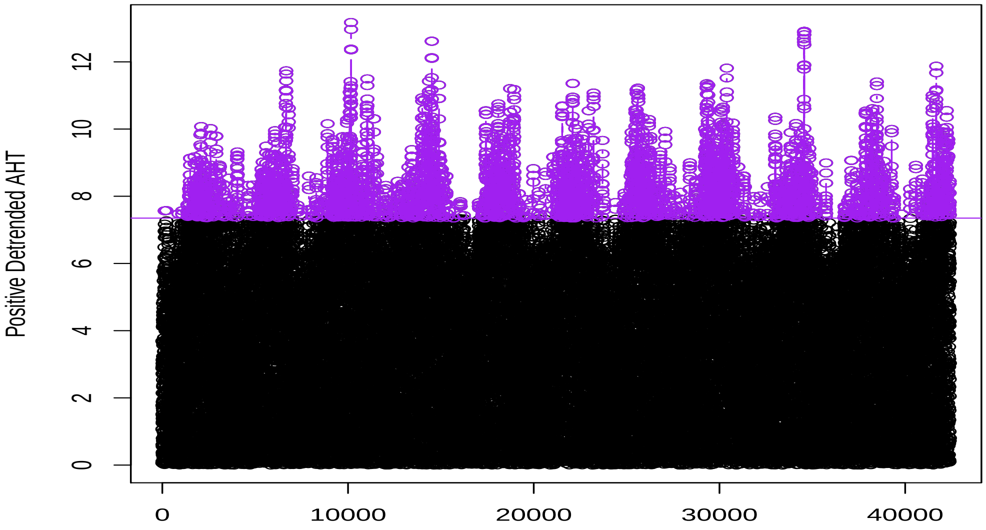


Figure 4.23: GPD fitted to detrended positive residuals of AHT.

Table 4.12: Parameter estimates for time-homogeneous and non-homogeneous point process models fitted to cluster maxima of AHT with standard errors in parenthesis.

Parameter estimate $\hat{\theta}$	Model	Model	Model	Model	Model	Model
	M_0	M_1	M_2	M_3	M_4	M_5
$\hat{\mu}_0$: (Intercept)	12.2531(0.1850)	3.3214(0.2358)	3.3174(0.8986)	3.3305(0.6821)	9.0259(0.2792)	3.3307(0.6830)
$\hat{\mu}_1$: (Slope)		0.0476(0.0001)	0.0193(0.0277)	1.75444(0.7220)	0.2037(0.0049)	1.7466(0.7230)
$\hat{\mu}_2$: (Quadratic)					-0.0085	-0.0004(0.004)
$\hat{\sigma}_0$: (Intercept)	0.4789(0.0531)	1.7641(0.0600)	1.7453(0.8376)	1.7544(0.7220)	1.2548(0.1884)	1.7466(0.7230)
$\hat{\sigma}_1$: (Slope)			0.0093(0.0049)	0.01(0.0004)		0.0125(0.0049)
$\hat{\sigma}_2$: (Quadratic)						-0.000(0.0049)
$\hat{\xi}_0$: (Intercept)	-0.1726(0.0233)	-0.6342(0.0266)	-0.6304(0.3670)	-0.6348(0.3157)	-1.0539(0.1046)	-0.6342(0.3158)
Log. lik	-3915.371	-3915.511	-3916.023	-3915.417	-3670.294	-3915.457

Estimates of time-homogeneous point process model are given in column 2 with standard errors in parentheses. The shape parameter is estimated by $\hat{\xi} = -0.1726(0.0233)$, which reveals appropriateness of Weibull class of distributions towards modelling AHT in South Africa. Standard errors for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ are generally small. Non-homogeneous point process models that are given from equations (4.5) to (4.9) are also fitted using MLE method and the results are included in Table 4.12. Estimation of shape parameters for all non-homogeneous point process models in this section reveals that Weibull class of distributions is an appropriate fit to the data.

Model M_1 is non-stationary with linear transformation of the location parameter. The fit of M_1 is supported by diagnostic plots that are given in Figure 4.24. Model M_2 is non-stationary in both the location and scale parameters. The standard errors are small and the fit of model M_2 is supported by diagnostic plots which show a good fit as given in Figure 4.25. Model M_3 with an exponential transformation of scale parameter also shows a good fit with the diagnostic plots that are given in Figure 4.26. Results of the quadratic models are given in columns 6 and 7 of Table 4.12. The fit of model M_4 seems to be poor as compared to the rest of non-stationary models. This is visible in the diagnostic plots that are given in Figure 4.27. Quadratic transformation of location parameter in model M_4 is therefore not recommendable. However, quadratic transformation is recommended for scale parameter based on reasonably good fit of model M_5. Diagnostic plots for model M_5 are given in Figure 4.28 which does not show any significant doubt about the fit. All linear transformations and quadratic transformation of scale parameter seem to be appropriate non-stationary point process models for modelling AHT in South Africa.

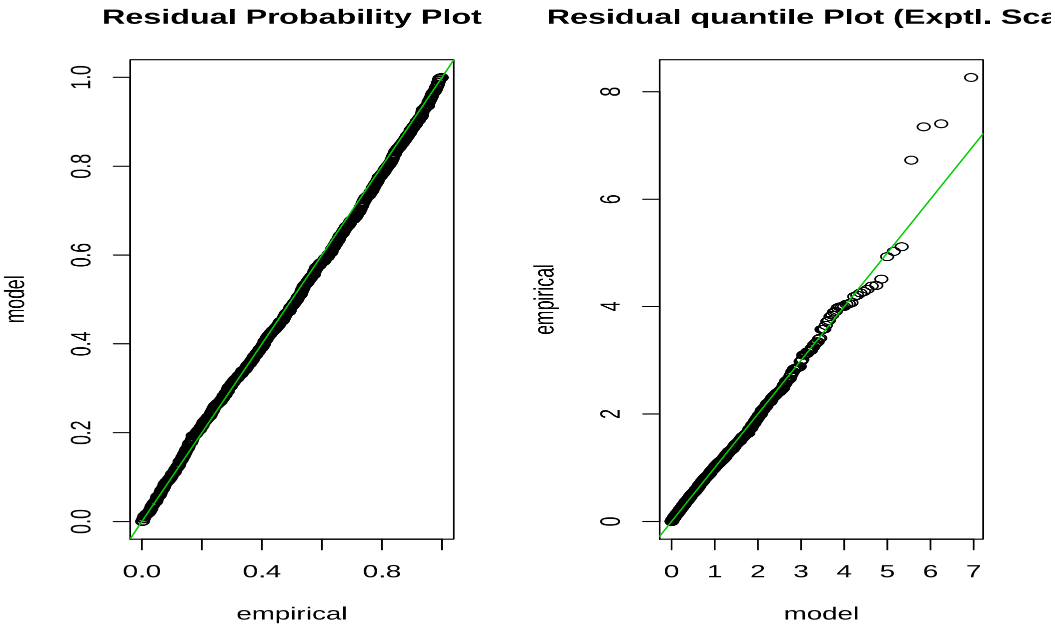


Figure 4.24: Diagnostic plots of the non-stationary point process model M_1 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_1 is linear in location parameter only.

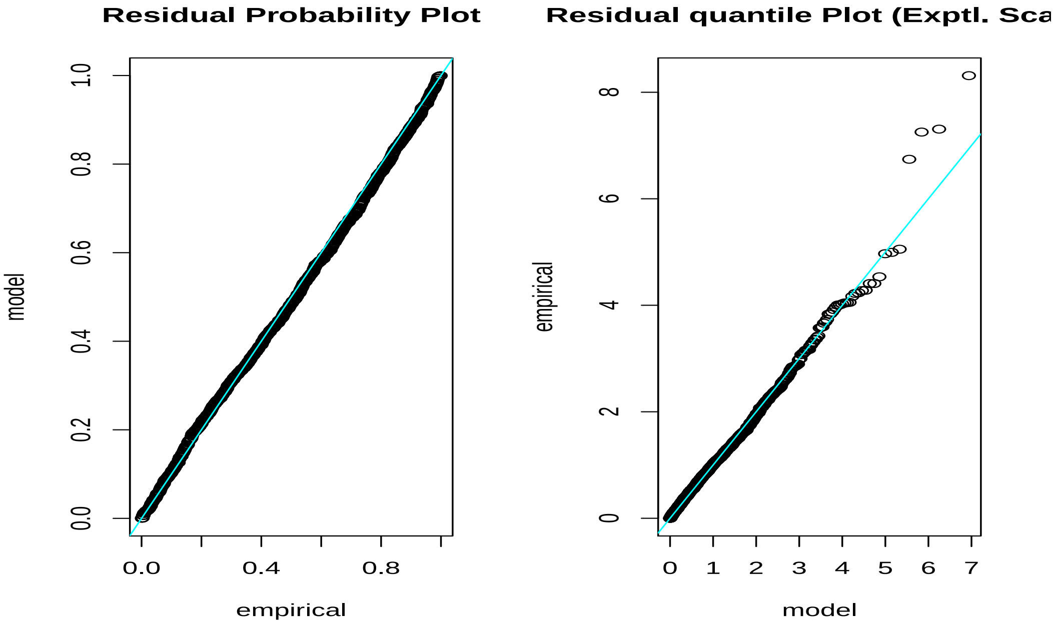


Figure 4.25: Diagnostic plots of the non-stationary point process model M_2 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_2 is linear in location parameter and exponential in scale parameter.

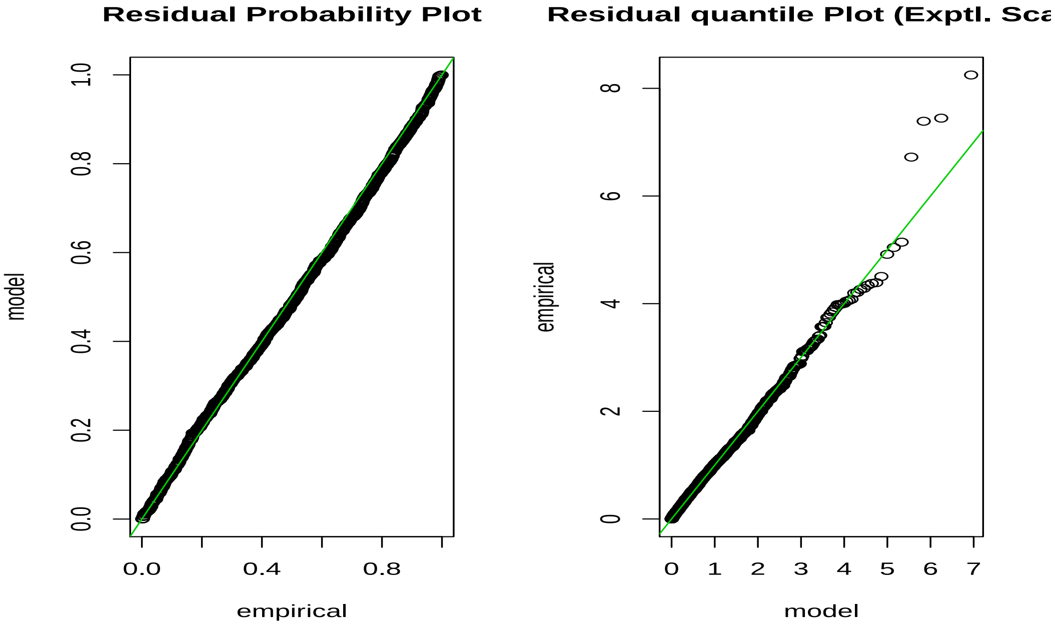


Figure 4.26: Diagnostic plots of the non-stationary point process model M_3 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_3 is exponential in scale parameter only.

.

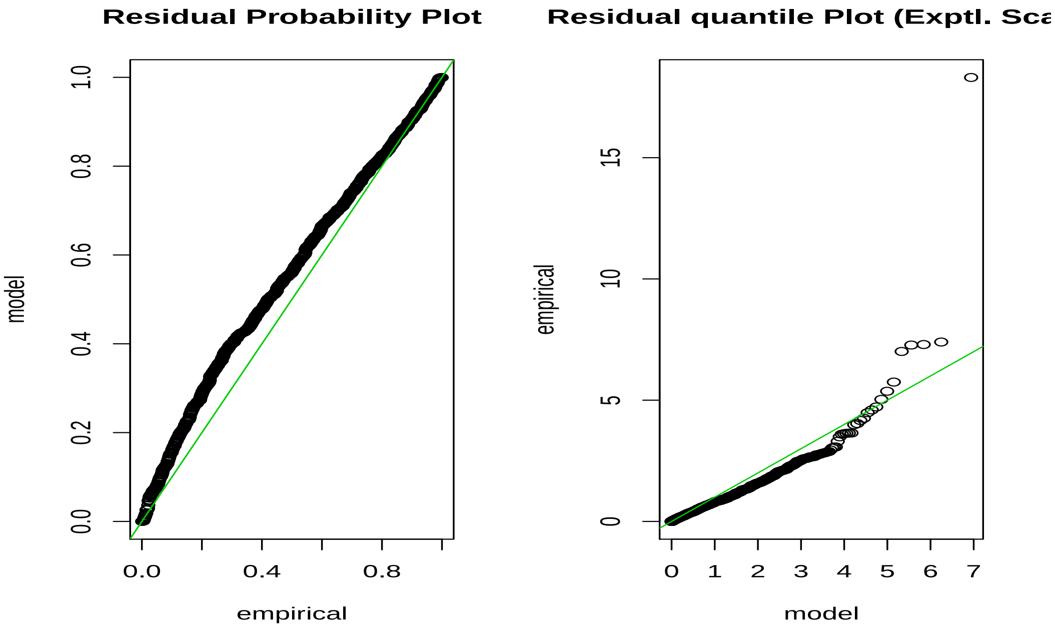


Figure 4.27: Diagnostic plots of the non-stationary point process model M_4 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_4 is quadratic in location parameter only.

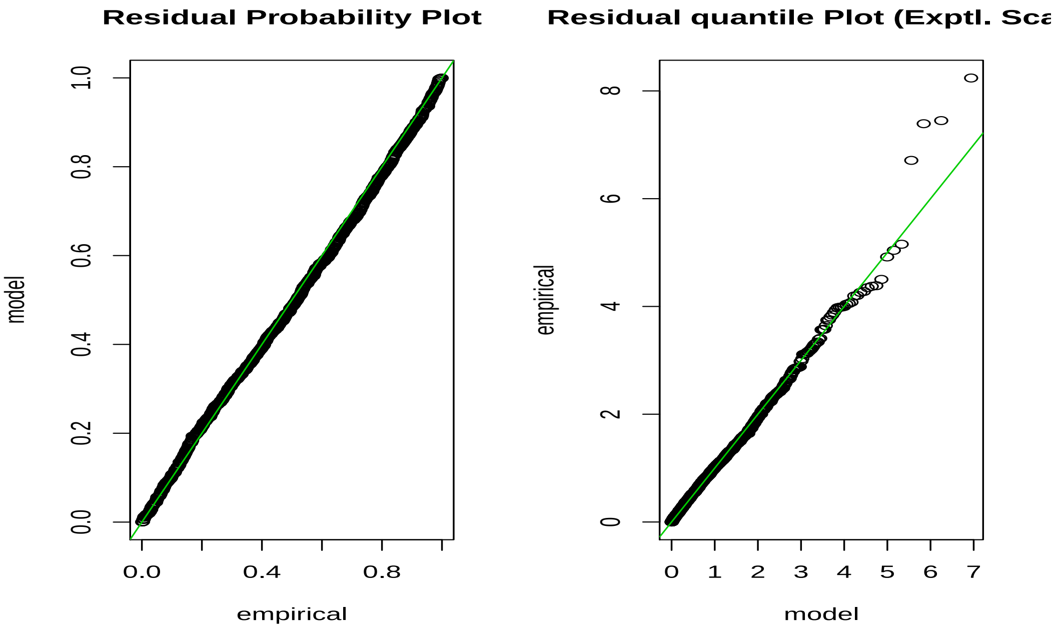


Figure 4.28: Diagnostic plots of the non-stationary point process model M_5 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_5 is quadratic in scale parameter only.

4.3.2 Point process characterization of daily maximum temperature

4.3.2.1 Data

In this section we fit time-homogeneous and non-homogeneous point process models to Daily Maximum Temperature (DMT) in the same fashion that was done towards modelling average hourly temperature. The data that are used in this section constitute a time series of daily maximum temperatures that are collected by SAWS over the period 2000 – 2010 as shown in Figure 4.29. The data are then detrended using penalized regression cubic smoothing splines that is given in equation (2.17) of Chapter 2 as an effort to remove the trend component. The detrended version of data (excesses) is given in Figure 4.30 at the top left panel.

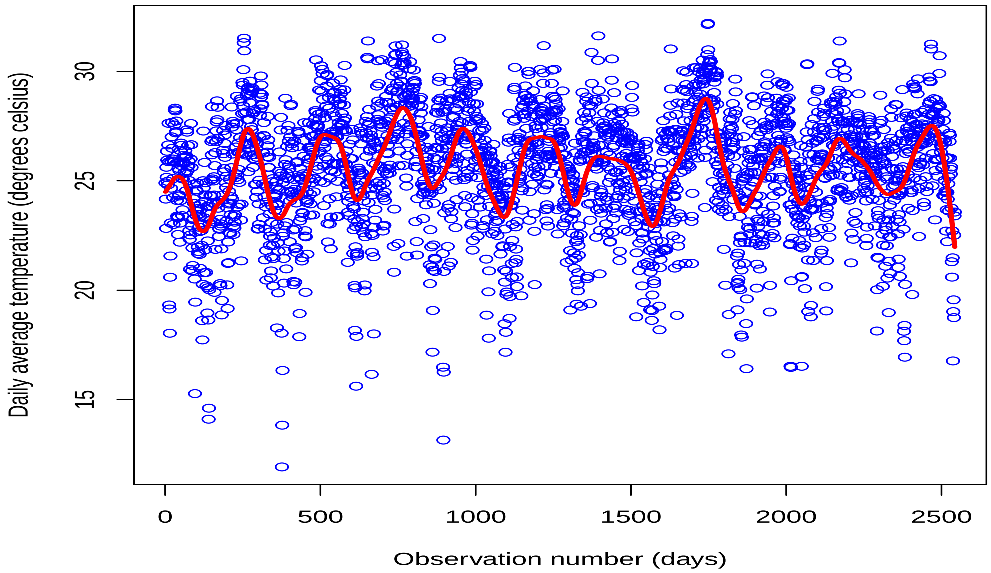


Figure 4.29: Time series plot of the maximum daily temperature. The blue dots indicate the observations and the red line indicates the smoothing parameter.

4.3.2.2 Threshold selection and declustering

Sufficiently high threshold is determined in this section using non-parametric extremal mixture model where a kernel density is fitted to bulk model and GPD fitted to the tail of the distribution in the similar way to Section 4.2 as given in equation (2.20) of Chapter 2. Threshold is found to be $\hat{\tau} = 2.9134$. The extremal index is estimated prior to the declustering process and found to be $\theta = 0.5908436$, which is large enough to give a low rate of exceedance. The rate of exceedance is $1/0.5908436 = 1.69 \approx 2$. We apply interval estimator method of Ferro and Segers (2003) to decluster positive residuals. The original series contains 1 407 observations and there are 81 identified cluster maxima. Excesses above the threshold are found to be 144. The GPD is fitted to cluster maxima shown in Figure 4.33. Threshold diagnostic plots are given in Figure 4.30. Threshold stability plot for extremal index is given at the top right panel and the bottom panel shows threshold stability plot for scale parameter on the left and threshold stability plot for shape parameter on the right. Further assessment of parameter estimates at several thresholds is given in Figure (4.31) which shows at the top left panel, threshold stability plot of logarithmic scale, threshold stability plot of shape parameter at the top right panel and mean residual life plot at the bottom panel. Figure 4.22 shows Q-Q plots for normalized inter-exceedance times against standard exponential quantiles. This is useful for assessing several thresholds for estimated value of extremal index.

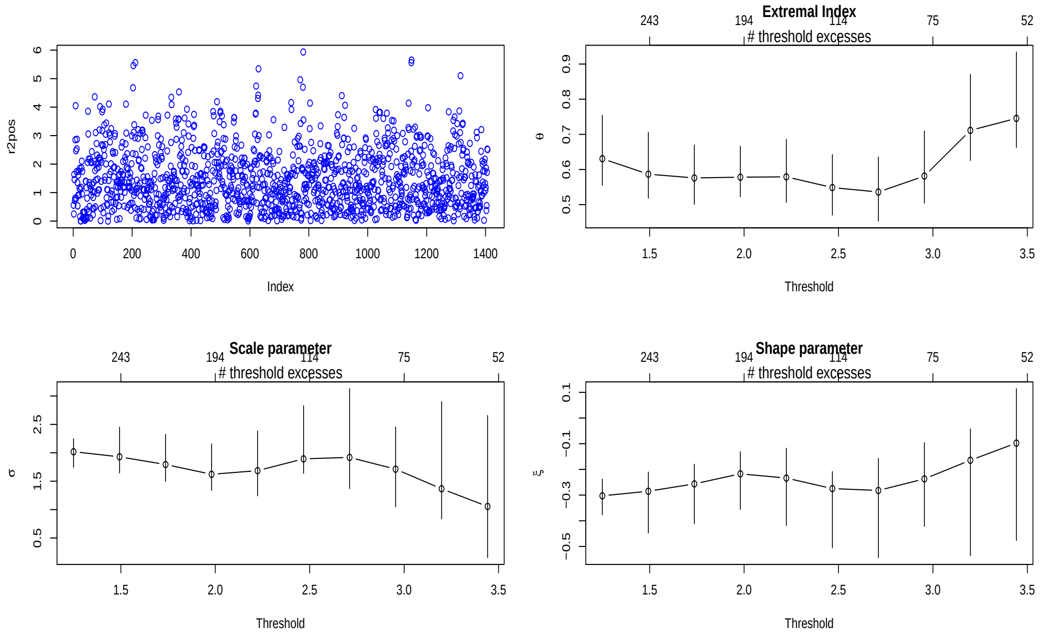


Figure 4.30: From left to right: Positive residuals (excesses) above the threshold (top left panel), threshold stability plot for the extremal index (top right panel), threshold stability plot for the scale parameter (bottom left panel), and threshold stability plot for the shape parameter (bottom right panel).

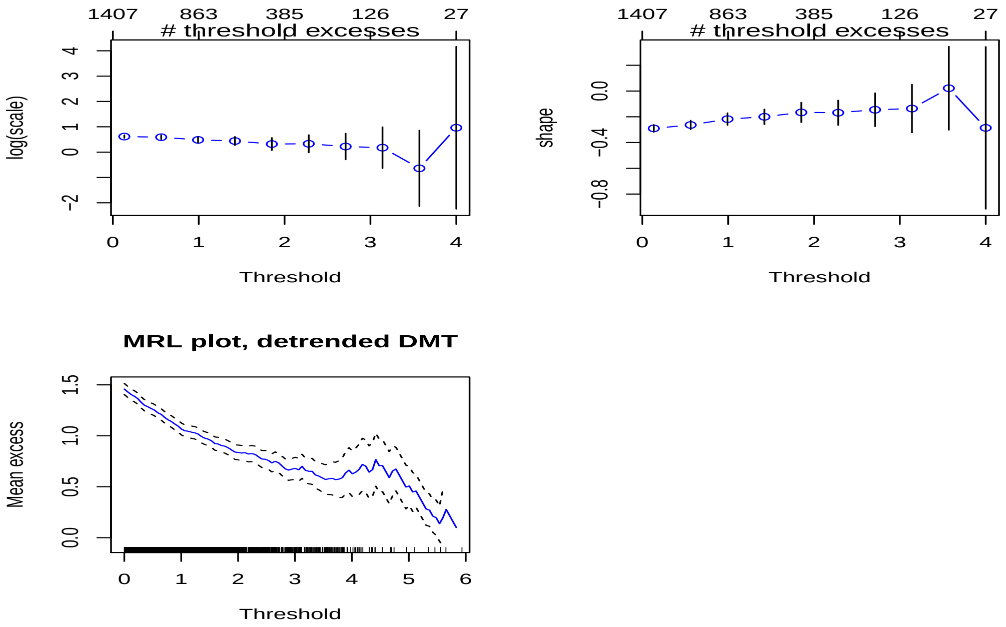


Figure 4.31: Threshold diagnostic plots for the point process models fitted to DMT. From left to right: Scale parameter threshold stability plot (top left panel), shape parameter threshold stability plot (top right panel) and mean residual life plot (bottom panel).

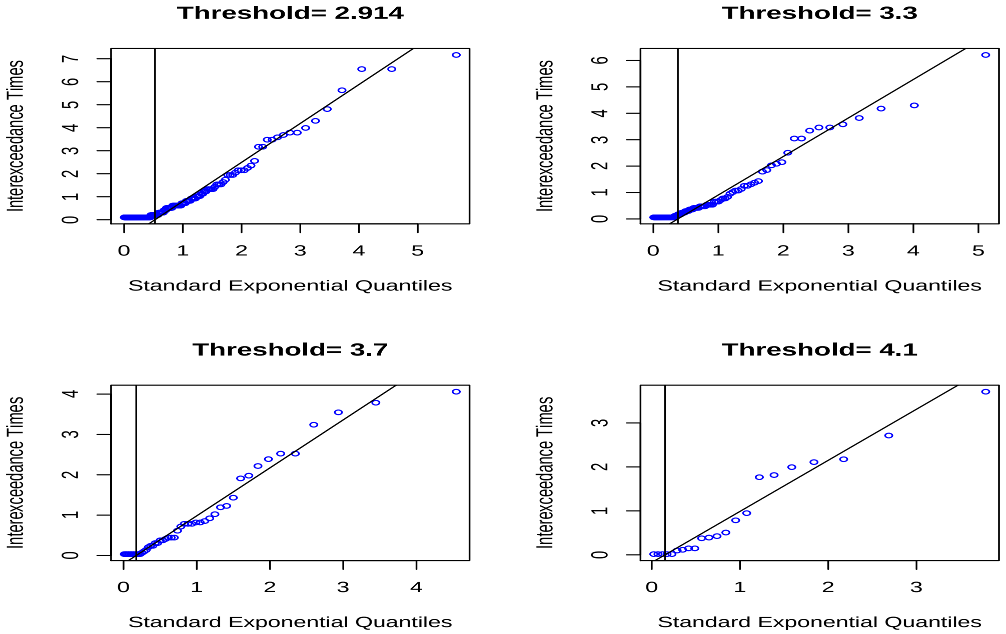


Figure 4.32: Q-Q plots of normalised interexceedance times against standard exponential quantiles. Vertical line shows the $(1-\hat{\theta})$ quantile; sloping line has gradient $1/\hat{\theta}$. Data are maximum daily temperatures in degrees celsius that are simulated from a max-autoregressive process with extremal index $\hat{\theta} = 0.5908436$.

4.3.2.3 Empirical results

The results of modelling DMT using MLE method are discussed in this section. Table 4.13 presents estimation of parameters with standard errors in parentheses, for time-homogeneous and non-homogeneous point process models. Looking at models M_0 and M_1, estimates of parameters look almost the same, implying existence of a zero trend component. If a test of hypotheses of the existence of trend is conducted, the null hypothesis will not be rejected, and the conclusion is an insignificance of a trend component. This implies that linear transformation of location parameter does not contribute much of the difference in the homogeneous model.

Amongst all the parameter estimates for stationary point process model, estimate of shape parameter has smallest standard error. Weibull class of distributions is established in this model as an appropriate fit to DMT based on the estimate of shape parameter which is $\hat{\xi} = -0.1857(0.0948)$, with standard error in parenthesis. Linear transformation is considered towards modelling non-stationarity in location parameter of the model that is shown in column 3 of Table 4.13. Diagnostic plots given in Figure 4.34 for this model do not show a significant difference from stationary point process model. The model that is given in column 4 of Table 4.13 is a linear transformation for modelling non-stationarity in the location parameter and the exponential transformation for modelling scale parameter. This is the only non-homogeneous point process model that does not establish appropriateness of Weibull class of distribution since estimate of shape parameter is $\hat{\xi} = 0.3857$. This model is associated to Pareto family of distributions. The fit of this model is poor as shown in the diagnostic plots that are given in Figure 4.35. The model in which only scale parameter is exponentially transformed is given in column 5 of Table 4.13.

Weibull class of distributions is established as the most appropriate fit to the data based on the estimate of shape parameter. The fit of this model looks better than that of Model M_2 as revealed by the diagnostic plots that are given in Figure 4.36. Point process models with quadratic transformations are models M_4 and M_5 whose results are summarized in columns 6 and 7 of Table 4.13. Both models show that Weibull class of distributions is the most appropriate fit to the data. However, the fit of model M_4 is poor based on the diagnostic plots that are given in Figure 4.37. Quadratic transformation of location parameter is not recommendable for modelling MDT in South Africa. Model M_5 in which scale parameter is quadratically transformed shows a good fit in comparison to model M_4 based on the diagnostic plots that are given in Figure 4.38.

Table 4.13: Parameter estimates for time-homogeneous and non-homogeneous point process models fitted to cluster maxima of MDT with standard errors in parenthesis.

Parameter estimate $\hat{\theta}$	Model	Model	Model	Model	Model	Model
	M_0	M_1	M_2	M_3	M_4	M_5
$\hat{\mu}_0$: (Intercept)	6.2013(0.4741)	6.2085(0.5022)	19.7897(0.8986)	6.2010	7.4027(0.9818)	6.2021
$\hat{\mu}_1$: (Slope)		-0.0001(0.0035)	-0.0469(0.0124)	1.75444(0.7220)	-0.1069(0.0050)	1.7466(0.7230)
$\hat{\mu}_2$: (Quadratic)					0.0014(0.0050)	-0.0004(0.004)
$\hat{\sigma}_0$: (Intercept)	0.3067(0.1396)	0.3069(0.1393)	7.2548(0.8376)	1.1293	0.1122(0.3418)	1.1240(0.7230)
$\hat{\sigma}_1$: (Slope)			-0.0202(0.0013)	-0.0177(0.0004)		0.0208(0.0049)
$\hat{\sigma}_2$: (Quadratic)						-3.2197(0.0050)
$\hat{\xi}_0$: (Intercept)	-0.1857(0.0948)	-0.1855(0.0945)	0.3857	-0.6829	-0.4791(0.7211)	-0.6818
Log. lik	-383.8523	-383.8531	-375.791	-383.8525	-351.6563	-383.857

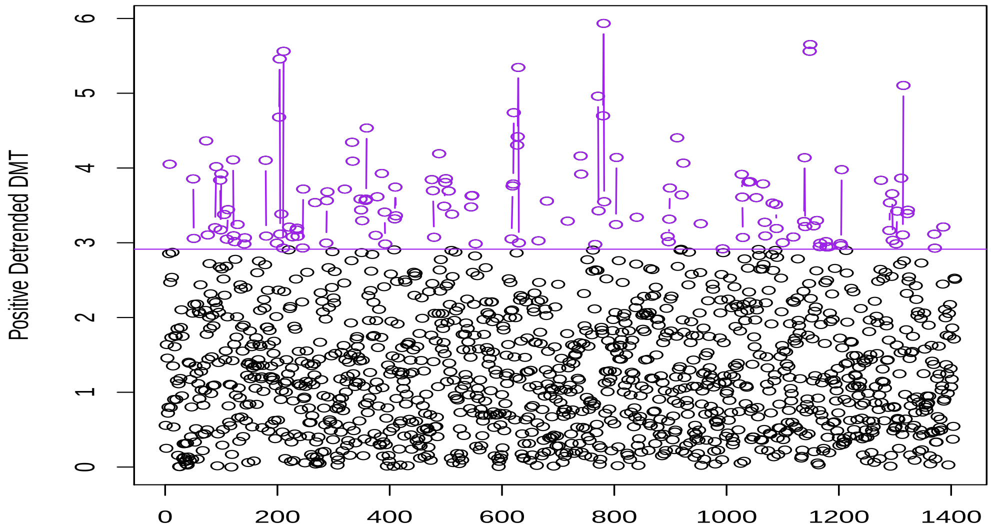


Figure 4.33: GPD fitted to cluster maxima (excesses) of the maximum daily temperature. The extreme observations above the threshold are indicated by the purple dots.

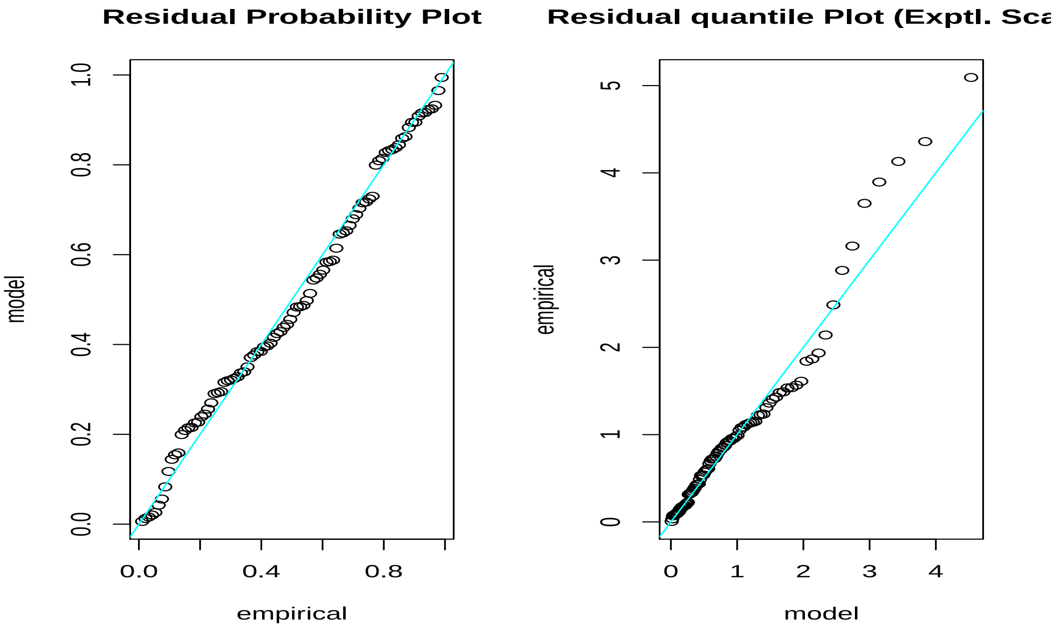


Figure 4.34: Diagnostic plots of the non-stationary point process model M_1 fitted to cluster maxima of AHT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_1 is linear in location parameter only.

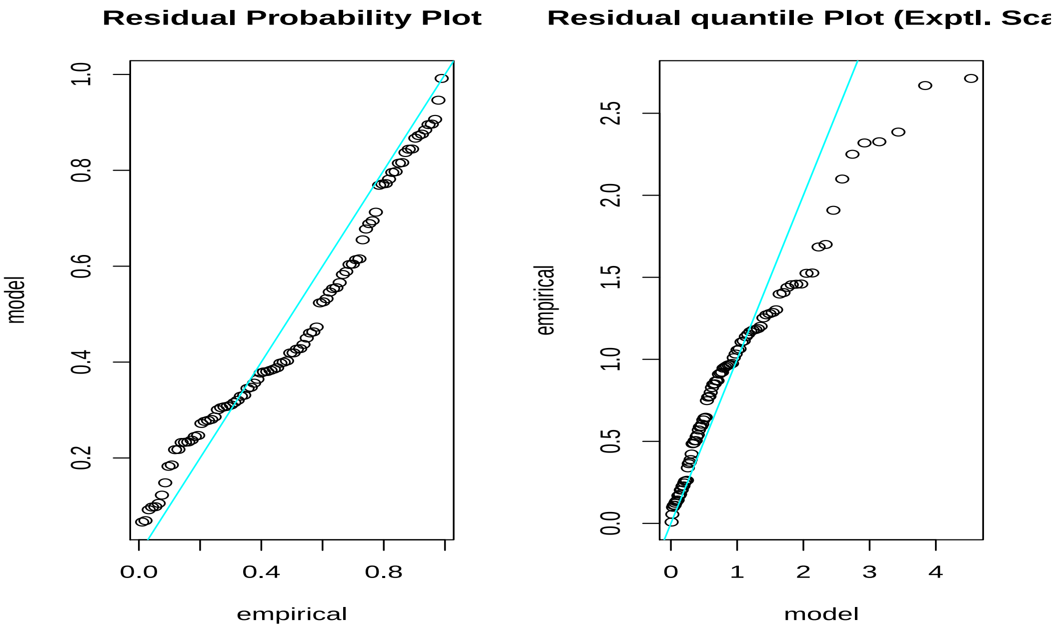


Figure 4.35: Diagnostic plots of the non-stationary point process model M_2 fitted to cluster maxima of DMT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_2 is linear in location parameter and exponential in scale parameter.

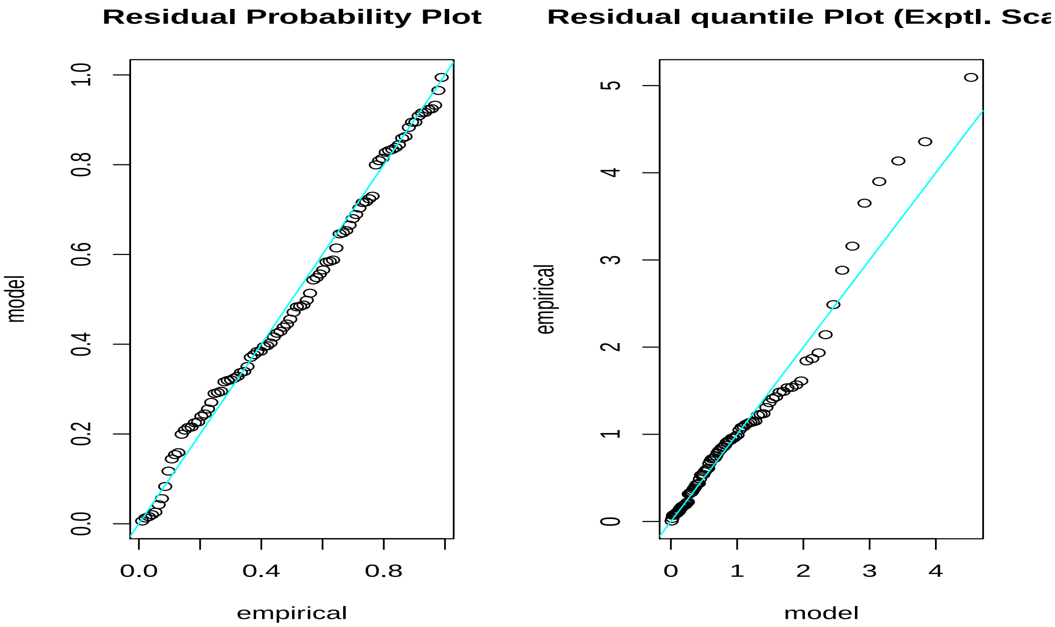
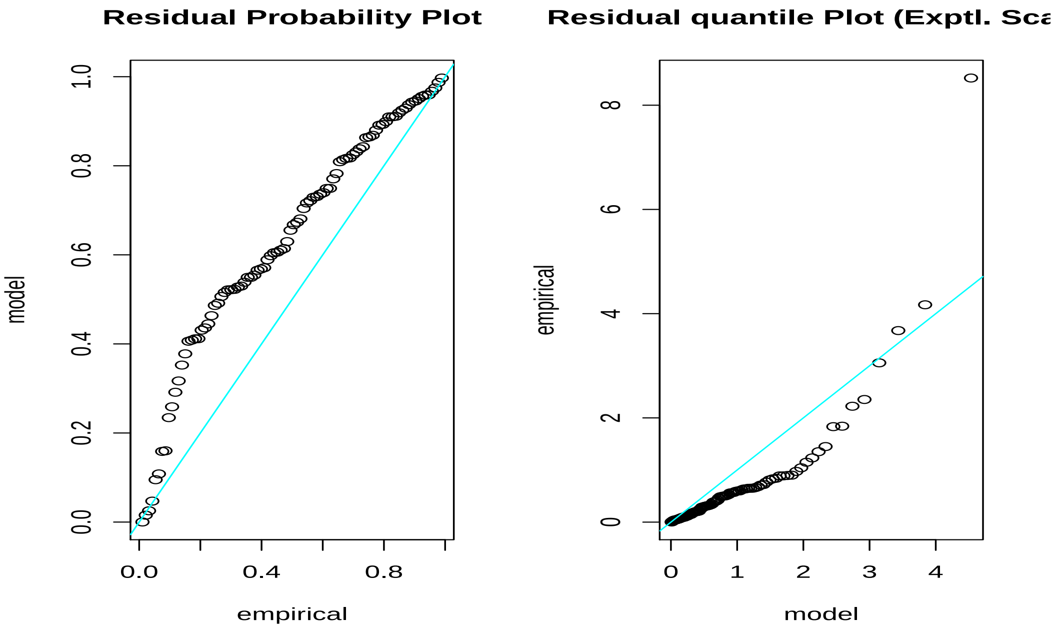


Figure 4.36: Diagnostic plots of the non-stationary point process model M_3 fitted to cluster maxima of DMT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_3 is exponential in scale parameter only.



.

Figure 4.37: Diagnostic plots of the non-stationary point process model M_4 fitted to cluster maxima of DMT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_4 is quadratic in location parameter only.

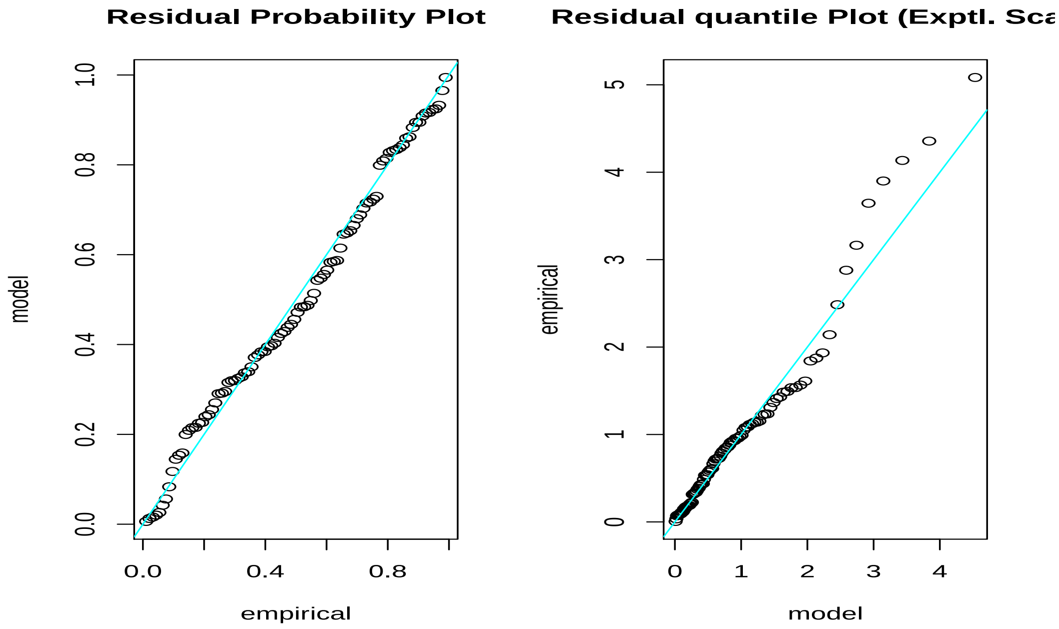


Figure 4.38: Diagnostic plots of the non-stationary point process model M_5 fitted to cluster maxima of DMT. From left to right: Residual probability plot (left panel), residual quantile plot with exponential scale (right panel). The model M_5 is quadratic in scale parameter only.

4.3.3 Frequency of occurrence of extremely high temperatures

The point process approach is an extension of block maxima and POT approaches based on the fact that, block maxima and POT approaches deal only with the frequency of occurrence of extreme values, whereas the point process is further capable of addressing the intensity rate of the occurrence of extremes (Cox, 1965; Deheuvels, 1983; Ogata and Katsura, 1986; Smith, 2003; Daley and Vere-Jones, 2007; Northrop and Jonathan, 2011; Embrechts, Klüppelberg and Mikosch, 2013, among others).

The frequency of occurrence of extremely high and extremely low temperatures constitutes the core of the analyses in this dissertation. This forms part of the important results that are useful to the planners and decision makers in the energy sector. In this section we calculate and interpret the intensity of occurrence of extremely high temperatures based on both the orthogonal and the reparameterization approaches. The orthogonal approach is described as the one that separately estimates the frequency and the intensity rate of extremes and hence difficult to interpret under non-stationary context (Gilleland and Graybill, 2009).

Due to this limitation, we only consider time-homogeneous point process models for calculating intensity rate and frequency using the orthogonal approach in this dissertation.

The intensity of extreme temperatures is calculated using orthogonal approach in this dissertation for the raw data and also with declustering. This is essential for the annual comparison of threshold exceedance rates between clustered and declustered series. Considering the analysis of Maximum Daily Temperature (MDT), the probability of occurrence of extremely high temperatures without declustering is calculated as follows:

$$\begin{aligned}\phi_1 &= \frac{X_i > \tau, i = 1, \ldots, n}{X_i, i = 1, \ldots, n} \\ &= \frac{144}{1407} \\ &= 0.102345.\end{aligned}$$

(Cox, 1965; Smith, 2003). The intensity rate is then calculated as follows:

$$\begin{aligned}\hat{\lambda}_1 &= \phi_1 \times 365 \quad \text{days} \\ &= 0.102345 \times 365 \\ &= 37.3561 \\ &\approx 37.\end{aligned}$$

(Cox, 1965; Smith, 2003). Similarly, we have the following for the declustered series:

$$\begin{aligned}\phi_2 &= \frac{X_i > \tau, i = 1, \ldots, n}{X_i, i = 1, \ldots, n} \\ &= \frac{81}{1407} \\ &= 0.057570.\end{aligned}$$

The intensity is given by:

$$\begin{aligned}\hat{\lambda}_2 &= \phi_2 \times 365 \quad \text{days} \\ &= 0.057570 \times 365 \\ &= 21.0128 \\ &\approx 21.\end{aligned}$$

These calculations imply that the frequency of occurrence of hottest days is about 37 days per year when the observations are not declustered and 21 days per year when there is declustering. The intensity is higher without declustering.

The alternative technique that simultaneously estimates both the frequency and intensity is the reparameterization approach. Gilleland and Graybill (2009) consider this as a method that is difficult to calculate but easy to interpret in the presence of covariates. We adopt the reparameterization that is discussed in Smith (2003) and then calculate the intensity based on point process models that are fitted towards modelling DMT. Smith (2003) discusses the relationship between Poisson−GPD and the GEVD for annual maxima and further highlights the following reparameterization:

$$\hat{\sigma}^* = \sigma + \xi(\tau - \mu) \quad \text{and} \quad \hat{\lambda} = \left[1 + \xi\left(\frac{\tau - \mu}{\hat{\sigma}^*}\right)\right]^{-\frac{1}{\xi}}. \tag{4.10}$$

Applying this reparameterization to the DMT, we have that

$$\begin{aligned}\hat{\sigma}^* &= \sigma + \xi(\tau - \mu)\\ &= 0.3067 - 0.1857(2.9134 - 6.2013)\\ &= 0.91726.\end{aligned}$$

It then follows that:

$$\begin{aligned}\hat{\lambda} &= \left[1 + \xi\left(\frac{\tau - \mu}{\hat{\sigma}^*}\right)\right]^{-\frac{1}{\xi}}\\ &= \left[1 - 0.1857\left(\frac{2.9134 - 6.2013}{0.91726}\right)\right]^{-\frac{1}{-0.1857}}\\ &= (1.6656379)^{5.385}\\ &= 15.6\\ &\approx 16.\end{aligned}$$

The reparameterization approach shows that extremely hot days occur at a frequency of 16 per annum. This results do not significantly differ from calculating the intensity using orthogonal method with declustering. However, the reparameterization is advantageous in terms of the simultaneous estimation of the intensity and the frequency.

We have applied the same approaches on the Average Hourly Temperature (AHT) data.

$$\begin{aligned}\phi_1 &= \frac{X_i > \tau, i = 1, \ldots, n}{X_i, i = 1, \ldots, n} \\ &= \frac{3957}{42409} \\ &= 0.093306.\end{aligned}$$

It then follows that

$$\begin{aligned}\hat{\lambda}_1 &= \phi_1 \times 365 \\ &= 0.0933 \times 365 \\ &= 34.0566 \\ &\approx 34.\end{aligned}$$

After declustering, we have that:

$$\begin{aligned}\phi_2 &= \frac{X_i > \tau, i = 1, \ldots, n}{X_i, i = 1, \ldots, n} \\ &= \frac{378}{42409} \\ &= 0.008913.\end{aligned}$$

The intensity is given by:

$$\begin{aligned}\hat{\lambda}_2 &= \phi_2 \times 365 \quad \text{days} \\ &= 0.008913 \times 365 \\ &= 3.2533 \\ &\approx 3.\end{aligned}$$

Using the reparameterization approach, we have the following:

$$\begin{aligned}\tau &= \hat{\sigma}^* \\ &= \sigma + \xi(\tau - \mu) \\ &= 0.4789 - 0.1726(7.3492 - 12.2531) \\ &= 1.32531.\end{aligned}$$

It then follows that:

$$\begin{aligned}\hat{\lambda} &= \left[1 + \xi\left(\frac{\tau - \mu}{\hat{\sigma}^*}\right)\right]^{-\frac{1}{\xi}} \\ &= \left[1 - 0.1726\left(\frac{7.3492 - 12.2531}{1.32531}\right)\right]^{-\frac{1}{-0.1726}} \\ &= 17.4857 \\ &\approx 18.\end{aligned}$$

The orthogonal approach shows that the intensity is approximately 34 without declustering, and approximately 3 with declustering. This implies that the frequency of occurrence of extremely high AHT is about 34 days in a year when the data are clustered and 3 days per year when the data are declustered. The higher frequency of extremes in the clustered data maybe due to short-range dependence and seasonality component in the data. The reparameterization approach shows even lower frequency of about 18 days per year. This section has presented the analyses of frequency and intensity of the occurrence of extremely high average hourly temperatures and extremely high maximum daily temperatures. These analyses provide useful information to the power utility companies since temperature is a driving factor of electricity demand.

Chapter 5

Conclusion

5.1 Introduction

This dissertation has focussed on presenting the use of EVT towards modelling temperature in South Africa over the period 01 January 2000 to 30 August 2010. Temperature was modelled for the purpose of quantifying the effects of the frequency of occurrence of extremely low and extremely high temperatures on the demand of electricity in South Africa over time. The data constitute a time series of average hourly temperatures that were collected by the South African Weather Service (SAWS) and supplied by Eskom. The data have been divided in to two seasonal versions in order to achieve the aim and objectives of the dissertation. The period from September to April of each year was defined as a non-winter season and the rest of the remaining data were considered for winter season.

The block maxima approach of classical extreme value theory was considered whereby the GEVD for r largest order statistics was fitted to average maximum daily temperature in Section 4.1 of Chapter 4. Under the POT approach, stationary GPD was fitted in two different subsections of Section 4.2 in Chapter 4. The GPD was firstly fitted towards modelling average minimum daily temperature and secondly towards modelling the influence of maximum daily temperature above 22 °C on the demand of electricity. The point process models were also fitted in two subsections of Section 4.3 in Chapter 4. These were first fitted towards modelling average hourly temperatures and then secondly towards modelling maximum daily temperatures. The frequencies and intensities of the occurrence of extreme high temperatures are calculated and discussed in Section 4.3.3 of Chapter 4.

5.2 Summary

In this section we give a brief summary of analyses that were discussed in Chapter 4. In Section 4.1 we discussed an application of block maxima approach of EVT towards modelling average maximum daily temperature data from South Africa over the years 2000 to 2010. The data for non-winter season (September to April of each year) have been considered in order to model the frequency of occurrence of the hottest days. The block approach of EVT was used whereby a stationary GEVD for *r* largest order statistics was fitted to estimate extreme high temperatures which result in high demand of electricity due to use of cooling systems. The MLE method was used to estimate target parameters. The estimation of shape parameter revealed evidence that Weibull class of distributions is a good fit to the data. Extreme quantiles for specified return periods were then estimated.

In Section 4.2 we presented an application of GPD in modelling average minimum daily temperature in South Africa for the winter period January 2000 to August 2010. The winter data are used in order to model the frequency of occurrence of coldest days. The observations over the winter period have actually been negated. A penalized regression cubic smoothing spline was used as a time varying threshold as well as to cater for seasonality. We then extracted excesses (residuals) above the cubic spline and fitted a non-parametric extremal mixture model to get a sufficiently high threshold. The parameters were estimated using both the MLE and the Bayesian methods. The estimate of shape parameter showed that the Weibull family of distributions is appropriate in modelling the upper tail of the distribution of average minimum daily temperature in South Africa. The bootstrap resampling method was used as an assessment tool for uncertainty in the parameter estimation. This resulted in more accurate estimates of return levels.

In the second part of Section 4.2 we used the GPD and the piecewise linear regression models in modelling the influence of maximum daily temperature above a high reference temperature of 22 °C on the demand of electricity. The electricity demand model in equation (4.3) of Chapter 4 was used to determine the impact of ADT on electricity demand and it was established that if ADT increases by 1 °C, then the rate of increase on ADED is 0.55%. It was also established that if temperature increases by 1 °C, (for example, from 22 °C to 23 °C), then the electricity demand is expected to increase marginally by 138MW. The GPD was fitted with the threshold that was determined using non-parametric extremal mixture model and the target parameters were estimated using both the MLE and the Bayesian methods. The data showed elements of dependence and were therefore declustered using interval estimator method of Ferro and Segers (2003). The stationary GPD was finally fitted to cluster maxima and the Weibull class of distributions was established as the best fit to the data. The upper bound of the distribution was calculated and found to be 26.3 °C and it was emphasized that there will not be an increase in ADED for an increase in temperature above the upper bound. The m-observations return level was used for calculating the quantiles (temperature) together with the corresponding increases on ADED.

In Section 4.3 of Chapter 4, the point process models were used in modelling average hourly temperature. The data were detrended using the cubic regression smoothing splines and the models were fitted with the threshold that was determined using the parametric extreme value mixture models, where a truncated Weibull distribution was fitted to the bulk model and the GPD fitted to the tail of the distribution. The advantage of the parametric over the non-parametric version of mixture models is that the parametric does not take time to run the results in evmix R package. Another advantage is based on the use of parameterized tail fraction model which is important for the accountability of the misspecification at the tail of the distribution (MacDonald et al., 2011). Due to the dependence features of the data, the extremal index was calculated and found to be large enough in order to result in a low rate of exceedance. The data were then declustered in the similar manner to Section 4.2 and the MLE method was used to fit the stationary GPD to cluster maxima. We then fitted the time-homogenous and non-homogeneous point process models. Virtually all the models revealed evidence that Weibull class of distributions is appropriate fit to the data.

Amongst the transformation techniques for modelling non-stationarity in the parameters, the use of quadratic transformation of location parameter is not recommended based on the poor fit of the model as shown by the diagnostic plots. The entire procedure of point process modelling was repeated in modelling maximum daily temperature with a threshold that was determined using non-parametric extremal mixture models. The data still revealed evidence of dependence and were hence declustered. The quadratic transformation of the location parameter still shows a poor fit and is therefore not recommended. The stationary model M_0 and model M_1 are almost the same in terms of the parameters that are estimated in Table 4.13, meaning the insignificance of trend component. This implies that the linear transformation of location parameter does not contribute much of the difference from the stationary model.

5.3 Concluding remarks

The concluding remarks based on the analyses in Chapter 4 are listed in this section as follows:

- The frequencies of occurrence of minimum and maximum temperatures are assessed in this dissertation and their effects on electricity demand are determined.

- The use of GEVD for r largest order statistics has been found to be an appropriate block maxima approach for modelling average maximum daily temperature in South Africa.

- Looking at the use of POT approach, the data exhibited evidence of short-range dependence and high seasonality which led to the declustering of excesses above a sufficiently high threshold and fitting the GPD to cluster maxima. The fit of GPD without declustering would be misleading since average daily temperatures are known to be dependent (naturally grouped) because a hot day is likely to be followed by another hot day.

- The POT approach has shown that the use of penalized regression cubic smoothing spline as a time-varying threshold to time series data which exhibit strong seasonality provides a good fit of GPD to cluster maxima.

- The extremal mixture model as a recent criteria for determining sufficiently high or sufficiently low threshold was applied and found to be a convenient basis for determining sufficiently high time-varying thresholds.

- The time-homogeneous and non-stationary point process models were successfully fitted in modelling average hourly temperature and maximum daily temperatures in South Africa.

- The orthogonal and reparameterization approaches of calculating the intensity of extremes have successfully shown the annual frequencies of occurrence of extremely high average hourly temperatures and extremely high average maximum daily temperatures.

5.4 Key findings and contributions

It has been emphasized in several research articles and this dissertation that, among several predictor variables that partake in predicting the demand of electrical energy, temperature is a major driver of electricity demand. Temperature was modelled in this dissertation with the intention of quantifying the effects of extremely low and extremely high temperatures on the demand of electricity. In this section we briefly summarize the key findings and then discuss the contributions of this dissertation.

- The Weibull class of distributions is found to be the best fit to the data in all the modelling frameworks of this dissertation. This implies that the distributions of extremely low and extremely high temperatures in South Africa are thin-tailed.

- Modelling the frequency of occurrence of hottest days in the energy sector is crucial for the load forecasters to determine the effects of maximum daily temperature on the demand of electricity load. A frequency analysis of extreme temperatures was carried out and the results show that most of the extreme temperatures are experienced in January, February, November and December of each year. This point is relevant to the planners in the energy sector because the occurrence of extreme high temperatures implies an increase in electricity demand.

- Modelling maximum daily temperature has revealed that the occurrence of extreme high temperatures result in high demand of electricity due to use of cooling systems. When average hourly temperature increases, people continue to switch on the cooling systems until a point at which virtually all the cooling systems are on, resulting in the increase in electricity demand. At this point (26.3 °C), there would be no further increase in electricity demand.

- Modelling extreme high temperature in this dissertation is found to be useful to decision makers in Eskom, South Africa's power utility company as it is during the non-winter period that they plan for maintenance of their power plants.

- The use of the GPD together with the piecewise linear regression model was found to be a convenient approach for modelling the influence of maximum temperature above 22 °C on the average daily electricity demand.

- The impact of Average Daily Temperature (ADT) on electricity demand was determined and it is established that, if ADT increases by 1 °C, the rate of increase on

Average Daily Electricity Demand (ADED) is 0.55%.

- The orthogonal and the reparameterization approaches were used for determining the frequencies and intensities of occurrence of extreme temperatures. The orthogonal approach was applied to both the clustered and declustered data. The frequency of occurrence of extremes is higher (37 days per annum) for the clustered data and lower (21 days per annum) with declustering. The higher frequency maybe the consequences of short-range dependence and heavy seasonality.

  However, the results of the reparameterization approach are easier to interpret since this approach estimates both the frequency and the intensity simultaneously. It was found that the frequency of occurrence of extreme maximum daily temperature is 16 days per year. This does not differ much from 18 days per year that is calculated based on the average hourly temperature data. The frequency analyses of the occurrence of extreme high and extreme low temperatures provide useful information to the system operators, energy forecasters, planners and decision makers in power utility companies like Eskom because the higher frequency of the occurrence of extreme high temperature implies an increase in the demand of electricity.

5.5 Limitations of the dissertation

There are several predictors of electricity demand that are well known amongst which temperature is the major driver. This dissertation is limited to conclude about the demand of electricity based on temperature only. The data that we used are the average hourly temperatures from the country in general, which makes it a challenge to draw conclusions about the provinces, districts or municipalities. The data that we used are not directly suitable to the use of multivariate EVT which could result in more detailed conclusions.

5.6 Areas for future research

Areas for future research include a detailed simulation study with the analyses based on extremal mixture models, inclusion of covariates in the GPD parameters and the use of Bayesian inference. In the future research, intensity of the point process will be calculated based on the non-stationary models in order to cater for the impact of covariates. We would also consider modelling temperature in South Africa using multivariate extreme value models which will result in more detailed conclusions. This will involve the use of higher techniques beyond the univariate EVT. Dependence could be studied in more details using the Cupula's approach. It could be a good idea also to look at other weather variables that are also vital in predicting the demand of electricity in South Africa. The recent developments in the energy sector involves the use of renewable electrical energy in form of solar systems. It is also our interest to research along these new developments because the non-renewable energy has been studied extensively by several researchers during the past decades.

References

ALFONS, A. AND TEMPL, M. (2013). Estimation of social exclusion indicators from complex surveys: The R package laeken. *Journal of Statistical Software*, **54** (15), pp.1–25.
URL: *http://www.jstatsoft.org/v54/i15/*

AMARAL, L. F., SOUZA, R. C., AND STEVENSON, M. (2008). A Smooth Transition Periodic Autoregressive (STPAR) model for short-term load forecasting. *International Journal of Forecasting*, **24** (4), pp.603–615.

AN, Y. AND PANDEY, M. (2007). The *r* largest order statistics model for extreme wind speed estimation. *Journal of Wind Engineering and Industrial Aerodynamics*, **95** (3), pp.165–182.

ANDREW THOMAS, U. L., BOB O'HARA AND STURTZ, S. (2006). Making BUGS Open. *R News*, **6** (1), pp.12–17.
URL: *http://cran.r-project.org/doc/Rnews/*

ARNOLD, B. C., BALAKRISHNAN, N., AND NAGARAJA, H. N. (1992). *A First Course in Order Statistics*, volume 54. Siam, 305 pages.

BALAKRISHNAN, N. AND COHEN, A. C. (2014). *Order Statistics and Inference: Estimation Methods*. Elsevier, 398 pages.

BALKEMA, A. A. AND DE HAAN, L. (1974). Residual life time at great age. *The Annals of Probability*, pp.792–804.

BEHRENS, C. N., LOPES, H. F., AND GAMERMAN, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, **4** (3), pp.227–244.

BEICHELT, F. (2006). *Stochastic Processes in Science, Engineering and Finance*. Chapman and Hall/CRC, 440 pages.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., AND TEUGELS, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley and Sons, Ltd, 514 pages.

BEIRLANT, J., VYNCKIER, P., AND TEUGELS, J. L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association*, **91** (436), pp.1659–1667.

BOMMIER, E. (2014). Peaks-Over-Threshold modelling of environmental data. *Department of Mathematics, Uppsala University, Sweden: Technical report* 33*,* 3 *September*.

BONSAL, B., ZHANG, X., VINCENT, L., AND HOGG, W. (2001). Characteristics of daily and extreme temperatures over Canada. *Journal of Climate*, **14** (9), pp.1959–1976.

BURNHAM, K. P. AND ANDERSON, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods and Research*, **33** (2), pp.261–304.

BYSTRÖM, H. N. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance*, **14** (1), pp.41–55.

CASTILLO, E. (2012). *Extreme Value Theory in Engineering*. Elsevier, 389 pages.

CASTILLO, E. AND HADI, A. S. (1997). Fitting the Generalized Pareto Distribution (GPD) to data. *Journal of the American Statistical Association*, **92** (440), pp.1609–1620.

CHIKOBVU, D. AND SIGAUKE, C. (2012). A frequentist and Bayesian regression analysis to daily peak electricity load forecasting in South Africa. *African Journal of Business Management*, **6** (40), pp.10524–10533.

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

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics: Springer-Verlag London Ltd, 224 pages.

COLES, S., HEFFERNAN, J., AND TAWN, J. (1999). Dependence measures for extreme value analyses. *Extremes*, **2** (4), pp.339–365.

COLES, S. AND SIMIU, E. (2003). Estimating uncertainty in the extreme value analysis of data generated by a hurricane simulation model. *Journal of Engineering Mechanics*, **129** (11), pp.1288–1294.

COLES, S. G. AND POWELL, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review/Revue Internationale de Statistique*, pp.119–136.

CONWAY, D. (2000). *Object Oriented Perl: A Comprehensive guide to concepts and programming techniques*. Manning Publications Co., Manning Publications, Connecticut, USA, 490 pages.

COX, D. R. (1965). On the estimation of the intensity function of a stationary point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.332–337.

DALEY, D. J. AND VERE-JONES, D. (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Springer Science and Business Media, 536 pages.

DAVID, H. A. AND NAGARAJA, H. N. (1981). *Order Statistics*. Wiley Online Library, 488 pages.

DAVISON, A. C. AND SMITH, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp.393–442.

DE HAAN, L. AND FERREIRA, A. (2007). *Extreme Value Theory: An Introduction*. Springer Science and Business Media, 435 pages.

DEHEUVELS, P. (1983). Point processes and multivariate extreme values. *Journal of Multivariate Analysis*, **13** (2), pp.257–272.

DEVORE, J. L. (2015). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 792 pages.

DIGGLE, P. (1985). A kernel method for smoothing point process data. *Applied Statistics*, pp.138–147.

EBERHARD, A. (2007). The political economy of power sector reform in South Africa. *Center for Environmental Science and Policy, Stanford Institute for International Studies, Stanford University, USA: Technical report 6, April.*

EFRON, B. AND TIBSHIRANI, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC, 456 pages.

EL ADLOUNI, S., OUARDA, T., ZHANG, X., ROY, R., AND BOBÉE, B. (2007). Generalized maximum likelihood estimators for the non-stationary generalized extreme value model. *Water Resources Research*, **43** (3).

EMBRECHTS, P., KLÜPPELBERG, C., AND MIKOSCH, T. (2013). *Modelling Extremal Events: For Insurance and Finance*, volume 33. Springer Science and Business Media, 645 pages.

FERGUSON, R., WILKINSON, W., AND HILL, R. (2000). Electricity use and economic development. *Energy Policy*, **28** (13), pp.923–934.

FERREIRA, A., DE HAAN, L., ET AL. (2015). On the block maxima method in extreme value theory: PWM estimators. *The Annals of Statistics*, **43** (1), pp.276–298.

FERRO, C. A. AND SEGERS, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65** (2), pp.545–556.

FINKENSTADT, B. AND ROOTZÉN, H. (2003). *Extreme Values in Finance, Telecommunications, and the Environment*. CRC Press, 398 pages.

FISHER, R. A. AND TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *In Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24. Cambridge University Press, pp. 180–190.

FRIEDMAN, J. H. (1991). Multivariate Adaptive Regression Splines (MARS). *The Annals of Statistics*, pp.1–67.

GARCÍA-CUETO, O. R. AND SANTILLÁN-SOTO, N. (2012). *Modelling Extreme Climate Events: Two Case Studies in México*. Intech Publishing, Croatia, 244 pages.

GENCAY, R. AND SELCUK, F. (2004). Extreme Value Theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting*, **20** (2), pp.287–303.

GILLELAND, E. AND GRAYBILL, V. (2009). An introduction to the analysis of extreme values using R and extRemes. *National Center for Atmospheric Research, Boulder,*

Colorado, U.S.A, 6th International Conference on Extreme Value Analysis, Technical report, 22 June 2009.
URL: *http://www.ral.ucar.edu/staff/ericg*

GILLELAND, E. AND KATZ, R. W. (2011). New software to analyze how extremes change over time. *Eos*, **92** (2), pp.13–14.

GILLELAND, E., RIBATET, M., AND STEPHENSON, A. G. (2013). A software review for extreme value analysis. *Extremes*, **16** (1), pp.103–119.

GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pp.423–453.

GOUBANOVA, K. AND LI, L. (2007). Extremes in temperature and precipitation around the Mediterranean basin in an ensemble of future climate scenario simulations. *Global and Planetary Change*, **57** (1), pp.27–42.

GUMBEL, E. J. (2012). *Statistics of Extremes*. Dover Publications, Inc. Mineola, New York, 400 pages.

HAHN, H., MEYER-NIEBERG, S., AND PICKL, S. (2009). Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, **199** (3), pp.902–907.

HASAN, A. R. N., H AND KASSIM, S. (2012). Modelling of extreme temperature using Generalized Extreme Value (GEV) Distribution: A case study of Penang, Malaysia. *In World Congress on Engineering*, volume 1. pp. 181–186.

HEFFERNAN, J. E. AND SOUTHWORTH, H. (2013). Extreme value modelling of dependent series using R. *Journal of Statistical Software*, **54** (16), pp.1–25. R package version 2.1.
URL: *http://www.jstatsoft.org/v54/i15/*

HEFFERNAN, J. E. AND STEPHENSON, A. G. (2014). *ismev: An Introduction to Statistical Modelling of Extreme Values*. R package version 1.40.
URL: *http://CRAN.R-project.org/package=ismev*

HEKKENBERG, M., BENDERS, R., MOLL, H., AND UITERKAMP, A. S. (2009). Indications for a changing electricity demand pattern: The temperature dependence of electricity demand in the Netherlands. *Energy Policy*, **37** (4), pp.1542–1551.

HU, Y. (2013). *Extreme Value Mixture Modelling with Simulation Study and Applications in Finance and Insurance*. Master's thesis, University of Canterbury.

HU, Y. AND SCARROTT, C. (2013). evmix: An R package for extreme value mixture modelling, threshold estimation and boundary corrected kernel density estimation. *Journal of Statistical Software*, **56** (18), pp.1–30.
URL: *http://www.math.canterbury.ac.nz/c.scarrott/evmix*

HYNDMAN, R. J. AND FAN, S. (2010). Density forecasting for long-term peak electricity demand. *Power Systems, IEEE Transactions*, **25** (2), pp.1142–1153.

INGLESI, R. (2010). Aggregate electricity demand in South Africa: Conditional forecasts to 2030. *Applied Energy*, **87** (1), pp.197–204.

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

KARR, A. (1991). *Point Processes and their Statistical Inference, 2nd edition*, volume 7. Marcel Dekker, Inc. New York, 512 pages.

KATZ, R. W. AND BROWN, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21** (3), pp.289–302.

KATZ, R. W., PARLANGE, M. B., AND NAVEAU, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, **25** (8), pp.1287–1304.

KHALIQ, M., OUARDA, T., ONDO, J.-C., GACHON, P., AND BOBÉE, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, **329** (3), pp.534–552.

KHULUSE, S. A. (2010). *Modelling Heavy Rainfall Over Time and Space*. Master's thesis, University of the Witwatersrand.

KLEIN TANK, A. AND KÖNNEN, G. (2003). Trends in indices of daily temperature and precipitation extremes in Europe. *Journal of Climate*, **16** (22), pp.3665–3680.

KOTZ, S. AND NADARAJAH, S. (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific, 185 pages.

Kyselỳ, J. (2008). A cautionary note on the use of non-parametric bootstrap for estimating uncertainties in extreme value models. *Journal of Applied Meteorology and Climatology*, **47** (12), pp.3236–3251.

Kyselỳ, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the Peaks-Over-Threshold method with a non-stationary threshold. *Global and Planetary Change*, **72** (1), pp.55–68.

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Probability Theory and Related Fields*, **65** (2), pp.291–306.

Li, Y., Cai, W., and Campbell, E. (2005). Statistical modelling of extreme rainfall in Southwest Western Australia. *Journal of Climate*, **18** (6), pp.852–863.

Li, Z., Shao, Q., Xu, Z., and Cai, X. (2010). Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia Watershed in Northwest China. *Journal of Hydrology*, **385** (1), pp.76–83.

Lyon, B. (2009). Southern Africa summer drought and heat waves: Observations and coupled model behavior. *Journal of Climate*, **22** (22), pp.6033–6046.

MacDonald, A., Scarrott, C., and Lee, D. (2013). Boundary correction, consistency and robustness of kernel densities using extreme value theory. *School of Mathematics and Statistics, University of Canterbury: Technical paper available in http://www.math.canterbury.ac.nz/ c.scarrott*.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis*, **55** (6), pp.2137–2157.

Mallor, F. (2009). An introduction to statistical modelling of extreme values. *Faculty of Economics and Management, Hogeschool Universiteit Brussel (HUB): Technical research paper 36, November*.

Meehl, G. A. and Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305** (5686), pp.994–997.

Miller, I., Freund, J. E., and Johnson, R. A. (1965). *Probability and Statistics for Engineers*, volume 11. Prentice-Hall Englewood Cliffs, NJ, 632 pages.

MILLINGTON, N., DAS, S., AND SIMONOVIC, S. P. (2011). The comparison of the Generalized Extreme Value (GEV), log-Pearson type 3 and Gumbel distributions in the Upper Thames River Watershed under global climate models. *Department of Civil and Environmental Engineering, The University of Western Ontario: Water Resources technical research report* 077*, September*.

MIRASGEDIS, S., SARAFIDIS, Y., GEORGOPOULOU, E., LALAS, D., MOSCHOVITS, M., KARAGIANNIS, F., AND PAPAKONSTANTINOU, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, **31** (2), pp.208–227.

MORAL-CARCEDO, J. AND VICENS-OTERO, J. (2005). Modelling the non-linear response of Spanish electricity demand to temperature variations. *Energy Economics*, **27** (3), pp.477–494.

MUKHOPADHYAY, N. (2000). *Probability and Statistical Inference*. CRC Press, 665 pages.

MUÑOZ, A., SÁNCHEZ-ÚBEDA, E. F., CRUZ, A., AND MARÍN, J. (2010). Short-term forecasting in power systems: A guided tour. *In Handbook of Power Systems II*. Springer, pp. 129–160.

NADARAJAH, S. (2005). Extremes of daily rainfall in West Central Florida. *Climatic Change*, **69** (2-3), pp.325–342.

NORTHROP, P. J. AND JONATHAN, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, **22** (7), 799–809.

OGATA, Y. AND KATSURA, K. (1986). Point process models with linearly parameterized intensity for application to earthquake data. *Journal of Applied Probability*, pp.291–310.

PAYNE, J. E. (2010). A survey of the electricity consumption growth literature. *Applied Energy*, **87** (3), pp.723–731.

PICKANDS III, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, pp.745–756.

PICKANDS III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pp.119–131.

PICKANDS III, J. (1994). Bayes quantile estimation and threshold selection for the generalized Pareto family. *In Extreme Value Theory and Applications*. Springer, pp. 123–138.

R CORE TEAM (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

REISS, R.-D. (2012). *Approximate Distributions of Order Statistics: With Applications to Non-parametric Statistics*. Springer Science and Business Media, 355 pages.

REISS, R.-D., THOMAS, M., AND REISS, R. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and other fields, 3rd edition*, volume 2. Springer, 528 pages.

RESNICK, S. I. (2013). *Extreme Values, Regular Variation and Point Processes*. Springer, 320 pages.

ROSS, S. M. (2014). *Introduction to Probability and Statistics for Engineers and Scientists, 5th edition*. Academic Press, 686 pages.

SCARROTT, C. AND MACDONALD, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, **10** (1), pp.33–60.

SHUMWAY, R. H. AND STOFFER, D. S. (2011). *Time Series Analysis and its Applications: With R Examples, 3rd edition*. Springer Science and Business Media, 596 pages.

SIGAUKE, C. (2014). *Modelling Electricity Demand in South Africa*. Ph.D. thesis, University of the Free State.

SIGAUKE, C., CHIKOBVU, D., AND VERSTER, A. (2012). Modelling daily increases in peak electricity demand using the Generalized Pareto Distribution. *In South African Statistical Journal Proceedings: Proceedings of the 54th Annual Conference of the South African Statistical Association: Congress 1*. Sabinet Online, pp. 58–66.

SMITH, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72** (1), pp.67–90.

SMITH, R. L. (1986). Extreme value theory based on the *r* largest annual events. *Journal of Hydrology*, **86** (1-2), pp.27–43.

SMITH, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, pp.1174–1207.

SMITH, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, pp.367–377.

SMITH, R. L. (2003). Statistics of extremes with applications in environment, insurance and finance. *Extreme Values in Finance, Telecommunications and the Environment*, pp.1–78.

SNYDER, D. L. AND MILLER, M. I. (2012). *Random Point Processes in Time and Space*. Springer-Verlag. New York, Inc, 481 pages.

SOARES, C. G. AND SCOTTO, M. (2004). Application of the *r* largest order statistics for long term predictions of significant wave height. *Coastal Engineering*, **51** (5), pp.387–394.

SONGCHITRUKSA, P. AND TARKO, A. P. (2006). The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*, **38** (4), pp.811–822.

SOUTHWORTH, H. AND HEFFERNAN, J. E. (2013a). *texmex: Statistical modelling of extreme values*. R package version 1.60.
URL: *http://CRAN.R-project.org/package=texmex*

SOUTHWORTH, H. AND HEFFERNAN, J. E. (2013b). Univariate extreme value modelling using R. *R News*, **4** (2), pp.21–52.
URL: *http://cran.r-project.org/doc/Rnews/*

STEFFEN, W., HUGHES, L., AND PERKINS, S. (2014). Heat waves: Hotter, longer, more often. *Climate Council of Australia Limited. Second Major technical report of the Climate Council, 2014*.

STEPHENSON, A. AND GILLELAND, E. (2005). Software for the analysis of extreme events: The current state and future directions. *Extremes*, **8** (3), pp.87–109.

STEPHENSON, A. AND RIBATET, M. (2014). *evdbayes: Bayesian Analysis in Extreme Value Theory*. R package version 1.1.
URL: *http://CRAN.R-project.org/package=evdbayes*

STRENGERS, Y. (2012). Peak electricity demand and social practice theories: Reframing the role of change agents in the energy sector. *Energy Policy*, **44**, pp.226–234.

SUGAHARA, S., DA ROCHA, R. P., AND SILVEIRA, R. (2009). Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology*, **29** (9), pp.1339–1349.

TANCREDI, A., ANDERSON, C., AND OHAGAN, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, **9** (2), pp.87–106.

TSUJITANI, M., OHTA, H., AND KASE, S. (1980). Goodness of fit test for extreme value distributions. *IEEE Transactions on Reliability*, **2**, pp.151–153.

VIETH, E. (1989). Fitting piecewise linear regression functions to biological responses. *Journal of Applied Physiology*, **67** (1), 390–396.

VRIEZE, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods*, **17** (2), pp.228.

WANG, J. Z. (1995). Selection of the k largest order statistics for the domain of attraction of the Gumbel distribution. *Journal of the American Statistical Association*, **90** (431), pp.1055–1061.

WANG, Y. (2011). *Smoothing Splines: Methods and Applications*. Chapman and Hall/CRC, 394 pages.

WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the *k* largest observations. *Journal of the American Statistical Association*, **73** (364), pp.812–815.

WENTZEL, D. AND MARÉ, E. (2007). Extreme value theory: An application to the South African equity market. *Investment Analysts Journal*, **36** (66), pp.73–77.

WU, Z., HUANG, N. E., LONG, S. R., AND PENG, C.-K. (2007). On the trend, detrending, and variability of non-linear and non-stationary time series. *Proceedings of the National Academy of Sciences*, **104** (38), pp.14889–14894.

WUERTZ, D. (2013). *fExtremes: Rmetrics - Extreme Financial Market Data*. R package version 3010.81.
URL: *http://CRAN.R-project.org/package=fExtremes*

Appendices

Some selected R code

R code for fitting stationary GEVD for $r = 1$ using ismev package

```
install.package(“ismev”)
library(ismev)
attach(SummerMax)
head(SummerMax)
tail(SummerMax)
x=gev.fit(SummerMax) ##Fitting the GEVD to Average Daily Maximum Temperature using Ismev
win.graph()
gev.diag(x)
```

R code for fitting GEVD for r largest order statistics using ismev package

```
Library(ismev)
attach(SummerMax)
head(SummerMax)
tail(SummerMax)

win.graph()
x=ts(mx,start=2000,freq=242)
plot(x,xlab=“Year”,ylab=“Maximum Temperature in Degrees Celsius”,col=“blue”)
plot(mx,xlab=“Number of observations (days)”,ylab=“Maximum temperature (degrees cel-
sius)”)

gev.fit(rlargestdata[,2])
win.graph()
gev.diag(gev.fit(rlargestdata[,2]))

rlarg.fit(rlargestdata[,2:3])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:3]))

rlarg.fit(rlargestdata[,2:4])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:4]))

rlarg.fit(rlargestdata[,2:5])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:5]))

rlarg.fit(rlargestdata[,2:6])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:6]))

rlarg.fit(rlargestdata[,2:7])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:7]))

rlarg.fit(rlargestdata[,2:8])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:8]))

rlarg.fit(rlargestdata[,2:9])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:9]))
rlarg.fit(rlargestdata[,2:10])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:10]))

rlarg.fit(rlargestdata[,2:11])
win.graph()
rlarg.diag(rlarg.fit(rlargestdata[,2:11]))
```

R code for fitting stationary GPD using ismev package

```
library(ismev)
attach(cmax)
head(cmax)
tail(cmax)

y=gpd.fit(cmax1,7.345) ##Fitting the GPD to cluster maxima using ismev
win.graph()
gpd.diag(y)
```

R code for fitting cubic regression smoothing splines

```
attach(Smintemp)
head(Smintemp)
tail(Smintemp)

win.graph()
plot(nmin,xlab=“Observation number”, ylab=“Negated minimum temperature (degrees C)”,
col=“blue”)
lines(smooth.spline(time(nmin),nmin, spar=0.1),col=“red”,lwd=3)

plot(nmin, type=“p”, ylab=“Negated minimum temperature (degrees C)”, col=“blue”,xlab=
“Observation number”)
lines(smooth.spline(time(nmin), nmin, spar=0.1719),col=“red”, lwd=3)
smooth.spline(time(nmin), nmin) ## GCV
r2=residuals((smooth.spline(time(nmin), nmin, spar=0.1719)))
plot(r2,col=“blue”,ylab=“Residuals observations”, xlab=“Observation number”)
```

```
r2pos ≤ r2[r2 > 0]
plot(r2pos, ylab=“Residuals above time varying threshold (positive residuals)”, col=“blue”,
xlab=“Observation number”)
tail(r2pos)
```

R code for fitting non-parametric extremal mixture model using evmix package

```
##Nonparametric extreme value mixture models

##Example fit kernel density
attach(cmax)
install.package(“evmix”)
library(evmix)
win.graph()
a=avgT
fit = fkdengpd(a, phiu = FALSE, std.err = FALSE)
hist(a,breaks=100, freq = FALSE, main=“”,xlim = c(22,30))
aa = seq(22,30, 1)
lines(aa, dkdengpd(aa, a, fit$lambda, fit$u, fit$sigmau, fit$xi, fit$phiu), col=“blue”, lwd =
2)
abline(v = fit$u, col=“blue”, lwd = 2)
legend(“topright”, “kdengpd”, col = “blue”,lty = 1, lwd = 2)
box()
fit
```

R code for fitting parametric extremal mixture model using evmix package

```
# The extreme value mixture model with a (truncated) Weibull distribution for the bulk and
GPD upper tail, with bulk model based tail fraction is fitted by default

attach(cmax)
library(evmix)
win.graph()
```

```
fit.bulk = fweibullgpd(a)
with(fit.bulk, lines(aa, dweibullgpd(aa, nmean, nsd, u, sigmau, xi), col = “red”))
abline(v = fit.bulk$u, col = “red”, lty = 2)
fit.bulk
#Parameterised tail fraction requires the option phiu=FALSE to be set:
win.graph()
fit.par = fweibullgpd(a, phiu = FALSE)
with(fit.par, lines(aa, dweibullgpd(aa, nmean, nsd, u, sigmau, xi, phiu),
main=” Histogram of positive detrended AHT”, col = ”blue”))
abline(v = fit.par$u, col = “blue”, lty = 2)
legend(“topright”, c(“True Density”, “Bulk Tail Fraction”,
“Parameterised Tail Fraction”), col=c(“black”, “red”, “blue”), lty = 1)
lines(density(a,adjust=2))
fit.par
## Diagnostic plots for assessing model fit

win.graph()
evmix.diag(fit.bulk)
win.graph()
evmix.diag(fit.par)
```

R code for declustering using texmex package

```
install.package(“texmex”)
library(texmex)
palette(c(“black”, “purple”, “cyan”, “orange”))
set.seed(20120118)

win.graph()
ei = extremalIndex(r2pos, threshold = 1.2675)
ei
dc = declust(ei)
par(mfrow=c(1,1))
plot(dc,col=“blue”, xlab=“Observation number”, ylab=“Positive residuals”)
dc
```

R code for plotting threshold diagnostics using texmex package

```
#Threshold stability plot
#The threshold stability plot examines a range of thresholds for invariance
of extremal index to change in threshold

library(texmex
par(mfrow = c(2, 2))
win.graph()
plot(r2pos, col=“blue”)
extremalIndexRangeFit(r2pos, nboot = 20)
par(mfrow = c(2, 2))
residp = gpdRangeFit(r2pos, umax = 2)
win.graph()
plot(residp,col=“blue”)
mrlresidp = mrl(r2pos)
win.graph()
plot(mrlresidp,col=“blue”, main = “Mean residual plot”)
min(a)
max(a)
win.graph()
plot(dc,col=“blue”, xlab=“Observation number”, ylab=“Positive residuals”)
dc
dc = declust(r2pos, threshold = 1.2675)
par(mfrow = c(2, 2))
ei = extremalIndex(r2pos, threshold = 1.2675)
plot(ei,col=“blue”)
ei = extremalIndex(r2pos, threshold = 0.8)
plot(ei,col=“blue”) ei = extremalIndex(r2pos, threshold = 0.85)
plot(ei,col=“blue”)
ei = extremalIndex(r2pos, threshold = 0.9)
plot(ei,col=“blue”)
```

R code for fitting generalised Pareto distribution to cluster maxima

```
attach(cmax)
install.package("evd")
library(evd)
resid.gpd = evm(dc)
resid.gpd
par(mfrow = c(2, 2))
win.graph()
plot(resid.gpd)

dc$nCluster
length(r2pos)
dc$nCluster/length(r2pos)

evm(r2pos, th = 1.2675)
par(mfrow = c(2,2))
plot(evm(r2pos, th = 1.2675))
declust(r2pos, th = 1.2675, r = 1)
declust(r2pos, th = 1.2675)
m1=evm(dc)
m1
par(mfrow = c(2,2))
plot(m1)
```

R code for fitting parametric bootstrap

```
attach(cmax)
#GPD parameter uncertainty: PARAMETRIC BOOTSRAP

boot = evmBoot(evm(dc), trace = 1001)
summary(boot)
par(mfrow = c(1,2))
plot(boot)
```

R code for fitting Bayesian posterior distributions

```
#BAYESIAN

z = evm(avgT, data = tempabove22, qu = 0.7,method = “simulate”)
z1 = update(z, method = “simulate”, trace = 40001, penalty = “gaussian”)
## 40001 steps taken
## Acceptance rate: 0.356
z1
par(mfrow = c(3, 2))
plot(z1)
z1 = thinAndBurn(z1, burn = 500, thin = 20)
summary(z1)
predict(z1, type = “lp”)
pred = predict(z1, M =1000)# M = 10 predicted return level:
summary(pred)
```

R code for identifying clusters of exceedences using evd package

```
attach(cmax)
install.package(“evd”)
library(evd)
c2 = clusters(r2pos,7.345, cmax = TRUE)
c2
write.table(c2,“ /MREXclustermaxima1.txt”,sep=“”)
```

R code for fitting time-homogeneous point process models using ismev package

```
attach(MrexCmax)
head(MrexCmax)
tail(MrexCmax)
library(ismev)
z=pp.fit(cmax1,7.345) ##Fitting stationary point process to cluster maxima using ismev
```

```
win.graph()
pp.diag(z)
length(cmax1)
tail(cmax1)
```

R code for fitting non-homogeneous point process models using ismev package

```
attach(cmax)
#MODEL 1 with t as covariate
t¡-c(1:206)
#computing/creating the variable $t_i$ #
ti=matrix(ncol=1,nrow=1033)
ti[,1]=seq(1,1033,1)
library(ismev)
ppfit=pp.fit(cmax1,7.345, ydat=ti,mul=1,sigl=NULL) # linear in location only#
win.graph()
pp.diag(ppfit)
ppfit=pp.fit(cmax1,7.345, ydat=ti,mul=1,sigl=1) # linear in both location and scale#
win.graph()
pp.diag(ppfit)
ppfit=pp.fit(cmax1, 7.345, ydat=ti,sigl=1) # linear in scale only#
win.graph()
pp.diag(ppfit)
#summary(ppfit)
#We can also create a quadratic model eg, $u(t) = uo + u1(t) + u2(t^2)$ as follows:
ti2=matrix(ncol=2,nrow=1033)
ti2[,1]=seq(1,1033,1)
ti2[,2]=(ti2[,1])**2
ppfit=pp.fit(cmax1,7.345) # stationary original model, from data in column mpeak#
win.graph()
pp.diag(ppfit)
ppfit=pp.fit(cmax1,7.345, ydat=ti2,mul=c(1,2)) #nonstationary quadratic model in location
parameter
win.graph()
```

```
pp.diag(ppfit)
ppfit=pp.fit(cmax1,7.345, ydat=ti2,sigl=c(1,2)) #nonstationary quadratic model in scale only
win.graph()
pp.diag(ppfit)
```