Multivariate Extreme Value Theory with an Application to Climate Data in the Western Cape Province.

Lipika Bhagwandin

A dissertation presented for the degree of
Master of Science

Department of Statistical Sciences

University of Cape Town

12 March 2017

Supervisor: Dr Şebnem Er

Abstract

An understanding of past and current weather conditions can aid in identifying trends and changes that have occurred in weather patterns. This is particularly important as certain weather conditions can have both a positive and a negative impact on various activities in any region. Together with an ever-changing climate it has become markedly noticeable that there is an upward trend in extreme weather conditions.

The aim of this study is to evaluate the efficacy of univariate and multivariate extreme value theory models on climate data in the Western Cape province of South Africa. Data collected since 1965 from five weather stations viz. Cape Town International Airport, George Airport, Langebaanweg, Plettenberg Bay and Vredendal was modelled and analysed. In the multivariate analysis, multiple variables are modelled at a single location.

Block maxima, threshold excess and point process approaches are used on the weather data, specifically on rainfall, wind speed and temperature maxima. For the block maxima approach, the data is grouped in n-length blocks and the maxima of each block form the dataset to be modelled. The threshold excess and point process approaches use a suitably chosen threshold whereby observations above the threshold are considered as extreme and therefore form the dataset used in the models. Under the threshold excess approach, only observations that exceed the threshold in all components are able to be modelled, whereas exceedances in one and all components simultaneously can be handled by the point process approach.

While the probability of experiencing high levels of rainfall, wind speed and temperature individually and jointly are low, a few conclusions were drawn based on the comparison of the performance of the models. It was found that models under the block maxima approach did not perform well in modelling the weather variables at the five stations in both the univariate and multivariate case as many useful observations are discarded. The threshold excess and point process approaches performed better in modelling the weather extremes. Similar results are achieved between these two approaches in the univariate analysis and there is no outright distinction that favours one approach over the other. In terms of the multivariate case, which is restricted to two variables, the point process approach was able to provide estimates with increased accuracy as in many cases there are more extremes in one component individually than in both components. Specifically, the negative logistic and negative bilogistic models suitably capture the dependence structure between maximum wind speed versus maximum rainfall and maximum wind speed versus maximum temperature at the five weather stations.

The results from the point process models showed very weak dependence between wind

speed and rainfall maxima as well as between wind speed and temperature maxima which may warrant the inclusion of additional variables into the analysis and even a spatial component which is not included in this study.

Keywords: Multivariate extreme value theory; Generalised extreme value distribution; Weather variables; Dependence; Maxima; Fréchet margins; Component-wise; Threshold excess; Point Process.

Contents

List of Figures

List of Tables

List of Abbreviations

EVT	Extreme Value Theory
MEVT	Multivariate Extreme Value Theory
GEV	Generalised Extreme Value
GPD	Generalised Pareto Distribution
MRL Plot	Mean Residual Life Plot
MLE	Maximum Likelihood Estimation
CI	Confidence Interval
N-LL	Negative Log-Likelihood
AIC	Akaike Information Criterion
CT Airport	Cape Town International Airport
SAWS	South African Weather Services

Acknowledgements

The experience from start to finish of my research endeavours can be summed up as a learning one. Not only was I able to learn about a new area of statistics but writing this dissertation has equipped me with skills to deal with various challenges that arise along the way - both in academia and in the practical applications. Of course, this experience and opportunity to pursue a Master's degree could not have been possible without my support system and the assistance of a few people.

First and foremost, I would like to thank my supervisor, Dr Şebnem Er who has been supervising me throughout my postgraduate studies. Thank you for your flexibility of time and your willingness to help me understand difficult concepts. Your enthusiasm towards life and research is contagious and our talks about general day-to-day happenings along with your happy demeanour made this research experience a pleasant one.

To my parents, expressing my appreciation towards you both can never be fully portrayed in words. Thank you for your patience and encouragement throughout the duration of my studies. To my mother, Anusha, thank you for your support and ever assuring words that would always make me feel as though I could achieve anything. To my father, Niresh, who has tirelessly edited my numerous draft versions of this dissertation - thank you for your advice, willingness to read my many versions and constant reassurance. Thank you to my brother, Akshar, who has always been ready and willing to support me especially when it came to technical issues. To my boyfriend, Jarryd, who has seen the good, the bad and the ugly sides of me during my research. Thank you for listening to my rants and frustrations while still continuing to cheer me on and showing understanding during the different stages of my research.

I would also like to express my gratitude to the City of Cape Town for funding my postgraduate studies as well as the South African Weather Services for providing the data used in this study.

Chapter 1

Introduction

1.1 Background

Traditional and popular uses of statistics focuses on the averages of distributions and considers the extreme observations as outliers therefore excluding these observations from the analysis. However, there is much information about the data that can be extracted from these extreme observations that are often disregarded. Extreme observations follow certain distributions which allows for inference about the tails of the distributions. This is where the theory on extreme values comes to light.

Extreme Value Theory (EVT) is a method for the mathematical modelling of extreme events. A distinguishing factor of EVT is that the focus of the analysis is on the tails of a distribution. It allows one to make inference about extreme events occurring over a time period into the future. Accounting for multiple variables working together as predictors of extreme events provides a more accurate representation of how extreme events are ultimately caused from a statistical perspective in what is termed as Multivariate Extreme Value Theory (MEVT).

MEVT can be used to jointly model multiple variables in different scenarios. For instance, MEVT models can be used to model multiple variables at a single location. It can be used to model one variable at different time periods and it can be used to model a single variable at multiple locations (Kotz and Nadarajah, 2000).

1.2 Background to the Study

The climate in the Western Cape province can be classified as dry and hot in summer and wet and cold in winter with wind all year round, formally known as a Mediterranean climate. The weather variables of interest in such a climate would be rainfall, temperature and wind speed maxima. In respect of experiencing extreme heat or cold climatic conditions - these are occurrences that happen on rare occasions. However, over the recent years, there have been an increase in these rare weather conditions. Most noticeably, there have been prolonged periods of droughts over the past few decades occurring when there are lower than average levels of rainfall together with higher temperatures (Araujo *et al.*, 2016). To summarise this, the Southern African Development Community

(SADC) noted that "in the coming decades the SADC is expected to experience higher land and ocean surface temperatures than in the past, which will affect rainfall, winds and the timing and intensity of weather events" (Southern African Development Community, 2012).

Extreme weather events in the Western Cape province include storm surges, flooding and fires. These extreme weather events, in some instances, are a culmination of extreme levels of certain weather variables working together. For example, extended periods of drought together with high levels of wind speed aid in the occurrence in fires (Ziervogel *et al.*, 2014). Thus, increased temperature readings with higher rainfall levels and wind speed makes MEVT an ideal candidate for modelling and analysing the occurrences of the extremes weather events.

1.3 Problem Description

While there have been several studies on extremes of climate data, the applicability and efficacy of EVT and MEVT approaches has not been gauged in a South African context and specifically the Western Cape province. Several different approaches are proposed in MEVT but how well do the approaches perform compared to each other and what causes the differences in the performance of the models under the various approaches.

In particular, which MEVT approaches provide reliable results and how well are the MEVT models in assessing the occurrence of extreme weather events in the province. What are the similarities and differences of the approaches and how suitable are the approaches under a Mediterranean climate.

1.4 Aim of Study

The aim of the study is to assess the efficacy of MEVT as a modelling tool on historic climate data. The climate data is from weather stations across the Western Cape province in South Africa. Two different approaches of univariate EVT and for MEVT, three different approaches, were used to model and analyse extreme weather conditions experienced in the province, a region in the South West of South Africa. Specifically, to apply MEVT models to pairwise combinations of rainfall, temperature and wind speed maxima from five weather stations across the Western Cape province.

Component-wise maxima, threshold excess and point process models are explored and applied to climate data. The efficacy of these models are evaluated in this study in univariate and multivariate analyses. The univariate analysis provides a foundation for the multivariate case in terms of theory and application.

1.5 Scope and Limitations

This study is restricted to the region of South Africa and is specific to the climate experienced in the Western Cape province.

The data is from five weather stations across the Western Cape and the analysis is restricted to a specific period from 1965 to 2015. Three weather variables are used in the models of the different approaches viz. rainfall, temperature and wind speed maxima to look at the efficacy of the approaches in modelling extreme weather events.

While the multivariate methodology extends to two or more variables, the application of the methods is restricted to a bivariate case. Furthermore, a spatial component is not included in the analysis as this would add in another dimension to the study.

1.6 Computation

The software program, R is used to analyse the data. It is freely available for download on https://cran.r-project.org/bin/windows/base/. It is widely used in statistical applications and has packages that perform various statistical routines.

The package *evd* is used extensively in the data analyses as it has both univariate and bivariate capabilities. There are a few extensions to the multivariate case in some of the functions within this package.

1.7 Layout of the Dissertation

Chapter 2 contains a literature review of work published in this area. It solidifies the applicability and viability of MEVT especially in the context of this study and also contains the details of EVT and MEVT methodologies.

Chapter 3 is about the specifics of the data which covers how the data was cleaned, missing values handled, period of the data per station and other finer details. Findings and results of the fitted models are also presented in this chapter.

Finally, Chapter 4 presents the conclusions of the models fitted under the three approaches as well as recommendations and future work.

Chapter 2

Literature Review & Methodology

This chapter describes the origins of EVT and MEVT. It presents a foundation of existing EVT methodologies and explains the differences between the various modelling approaches. Specifically, the details of the block maxima, threshold excess and point process approaches are explained. An overview of applications and studies that have already been conducted within the EVT realm are also explored. There is a differentiation between the univariate EVT and MEVT, each with its own methodologies as MEVT builds onto the univariate approaches.

The methodology is described at a basic level to provide an understanding of the three approaches and thus does not include rigorous proofs.

2.1 History of Extreme Value Theory

EVT concerns the mathematical modelling of extreme events with the focus of the analysis being on the tails of the distribution. The data that generates the tails of the distribution can be considered as extreme observations. The modelling of extreme observations can be helpful in predicting extreme weather events that might occur during a certain period in the future. Therefore, from a statistical perspective, modelling multiple variables together as predictors of extreme events provides a more accurate representation of the occurrence of extreme events. There is also an increased accuracy in prediction with multiple variables as opposed to modelling a single variable as a predictor of extreme events. Modelling extremes of multiple variables together is possible through the use of MEVT methods.

In the first part of this section, explanations of the methodologies relate to univariate cases and is followed by explanations of the multivariate case. The MEVT approaches are an extension to the univariate EVT approaches which takes additional variables into account.

While the earliest recorded works on the asymptotic distributions for largest values can be in found in Fréchet (1927), further developments about the generalised extreme

value limiting distributions were first realised by Ronald Fisher and Leonard Tippet in 1928. In particular, Fisher and Tippett discovered that three types of extreme limit distributions could be used on the sequence of maxima of a random sample. While Richard von Mises (1936) provided sufficient conditions to limiting distributions realised by Fisher and Tippett, it was Gnedenko (1943) who presented a complete characterisation to the non-degenerate limit laws to a sequence of maxima. In 1943, Gnedenko built onto the work of Fisher and Tippett by providing the necessary and sufficient conditions for convergence of extreme order statistics. These results have been captured in what is widely known as the Fisher-Tippett-Gnedenko Theorem (Beirlant *et al.*, 2004).

EVT applications to do with meteorology, hydrology and engineering date back to the 1950's with a few of the earliest studies found in Gumbel (1958). Further contributions to the theory can be seen in the work of de Haan during the 1970s who developed complete characterisations for the domains of attraction for limit laws of the maxima. Contributions to the field include the development of the limiting distributions for largest order statistics and stationary random sequences found in Leadbetter *et al.* (1983). The work of the aforementioned authors have contributed to the theory behind the generalised extreme value (GEV) family of distributions.

MEVT characterisations of asymptotic dependence and independence of variables is documented in Sibuya (1960). Joint tail characterisations of component-wise maxima are based on the work by Resnick (1987). The multivariate extension of the point process is due to the findings of de Haan and Resnick (1977) and de Haan (1985).

The Generalised Extreme Value Family of Distributions

The focus is on the statistical behaviour of

$$M_n = \max\{X_1, X_2, ..., X_n\}$$

$X_1, X_2, ..., X_n$ is a sequence of independent and identically distributed (i.i.d) random variables having a common distribution function F. M_n is standardised for a sequence of constants $a_n > 0$ and b_n to increase the stability of M_n as n increases.

$$\Pr\{M_n \leq x\} = F(x)^n \tag{2.1}$$

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

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

where G is a non-degenerate distribution function and G belongs to one of the following types of distributions according to the Fisher-Tippett-Gnedenko Theorem (Finkenstadt and Rootzen, 2003):

i. The Gumbel distribution

$$G(x) = \exp(-\exp(-x)) \quad -\infty < x < \infty$$

ii. The Fréchet distribution

$$G(x) = \begin{cases} 0 & x < 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}$$

iii. The Weibull distribution

$$G(x) = \begin{cases} \exp(-|x|^{\alpha}) & x < 0 \\ 1 & x > 0 \end{cases}$$

where $a > 0,\ b$ and $\alpha > 0$.

Thus, the GEV distribution is formed from the above-mentioned distributions which is as follows

$$G(x) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{2.2}$$

for $\quad -\infty < \mu < \infty, \quad -\infty < \xi < \infty, \quad \sigma > 0$

and is defined on the set $\{x : 1 + \frac{\xi(x-\mu)}{\sigma} > 0\}$ (Finkenstadt and Rootzen, 2003). G is a member of the GEV family of distributions where σ is the scale parameter, μ is the location parameter, ξ is the shape parameter.

The corresponding GEV distributions can be obtained by varying the values of ξ,
$\xi = 0$ corresponds to the Gumbel distribution
$\xi > 0$ corresponds to the Fréchet distribution
$\xi < 0$ corresponds to the Weibull distribution.

The work on the GEV distributions forms a foundation for the classical EVT models that are widely used in numerous applications and studies.

In addition to the characterisation of extremes under the GEV family, methods for modelling maxima that exceed a threshold has relevance. The works of Balkema and de Haan (1974) and Pickands (1975) showed that the limiting distribution for scaled excesses (over a threshold) can be approximated by a generalised Pareto distribution (GPD). Developments of the asymptotic joint distributions of extreme order statistics can be attributed to Weissman (1978) and Smith (1987). Poisson point process characterisations to excesses above a threshold were given by Smith (1989) and Davison and Smith (1990).

The Generalised Pareto Famiy of Distributions

Considering the random variables of X to have a common distribution function F where the behaviour of $Y = (X - u) > 0$ is described by (Finkenstadt and Rootzen, 2003) conditionally as

$$F(y) = Pr\{Y \leq y | Y > 0\} = \frac{F(u+y) - F(u)}{1 - F(u)} \tag{2.3}$$

For a large value of u, $F(y) \approx H(y)$ and the GPD is defined by

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} \tag{2.4}$$

The upper limits of ξ for different values are as follows (Coles, 2001)

i. $\xi = 0$:

$$H(y) = 1 - \exp\left(\frac{-y}{\sigma}\right),\ y > 0 \text{ as } \xi \to 0 \tag{2.5}$$

which corresponds to the exponential distribution with mean σ.

ii. $\xi > 0$: does not exist.

iii. $\xi < 0$: $-\frac{\sigma}{\xi}$ (finite).

The GPD has links to both the threshold excess and point process approaches, details of which are provided later on in this chapter.

2.2 Extreme Value Theory

Three classical approaches viz. the block maxima, threshold excess and point process approaches are expanded on in this section. An important aspect of EVT is the return levels. While the equations of the return levels are given, it is helpful to provide a conceptual example. For instance, in the analysis of rainfall maxima for a certain region and a particular station, the five year return level was calculated to be 32 mm. This means that on average, the level of 32 mm is expected to be exceeded once over the five years. Moreover, the calculations for the parameter estimates are performed using MLE and in order to use this method an independence assumption must be satisfied. This is discussed in the next section. MRL plots and declustering that are applicable to the threshold excess and point process approaches are expanded on under the threshold excess section. Lastly, diagnostic measures for model fit are provided and are suitable for all of the approaches.

2.2.1 Stationarity

Rainfall data does not exhibit large variation and thus many simplifying assumptions can be made which makes statistical analysis that much easier. The small variation points to the conclusion of the series being a stationary one. Through further analysis, it is found that rainfall has a more complicated structure and instead has a near stationary distribution (Coles, 2001). According to Khuluse (2010), stationarity can be assumed for the rainfall in the Western Cape. It is worth testing the data to ensure the series is stationary (all the random variables have the same probability distribution as each other). To counter the problem of stationarity, the data can be separated into the four

different seasons, namely Summer, Autumn, Winter and Spring. Alternatively, the data can also be divided even further into months, but care needs to be taken to prevent over-fitting of models (Finkenstadt and Rootzen, 2003). While a popular method for testing stationarity is through Mann-Kendall tests (Hasan *et al.*, 2012; Tawn, 1988), the various tests of stationarity are examining the same concept even if the hypotheses differ. Therefore, the Phillips-Perron test was used to determine stationarity in this study.

2.2.2 Block Maxima

The data used in the block maxima approach consists of a series of independent observations that are blocked into sequences of length n. The blocks are made to correspond to time periods (usually of one year duration) (Coles, 2001). In other words, the block maxima can be viewed as the annual maxima where n is the number of observations.

The estimation of parameters can be calculated using probability weighted moments (PWM), method of moments (MOM) (Hosking and Wallis, 1987), elemental percentile method (EPM) (Castillo *et al.*, 2005; Beirlant *et al.*, 2004), Bayesian methods (Beirlant *et al.*, 2004; Katz *et al.*, 2002) maximum likelihood estimation (MLE) (Beirlant *et al.*, 2004; Finkenstadt and Rootzen, 2003; Katz *et al.*, 2002; Coles, 2001). Maximum likelihood is used for all the approaches as it is a flexible method for estimating parameters (Coles, 2001).

The return levels provide a measure that helps with the prediction of the next extreme event. Details of parameter estimation using the MLE and return levels are given below.

Maximum Likelihood Estimation

Estimates for the parameters are found using MLE method. The details of this estimation method are explained in this section.

The log-likelihood of independent variables $X_1, X_2, ..., X_m$ which come from a GEV distribution (Finkenstadt and Rootzen, 2003), is provided as in equation 2.6, when

$\xi \neq 0$

$$l(\mu, \sigma, \xi) = -m\log\sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi} \quad (2.6)$$

subject to $1+\xi(\frac{x_i-\mu}{\sigma}) > 0$.

For $\xi = 0$

$$l(\mu, \sigma) = -m\log\sigma - \sum_{i=1}^{m}\left(\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right] \quad (2.7)$$

Maximising the above equations with respect to the parameters μ, σ and ξ, results in the maximum likelihood estimates for the whole GEV family of distributions.

Return Levels

The return level z_q can be described as the level that is expected to be exceeded on average once every $\frac{1}{q}$ years where $\frac{1}{q}$ is referred to as the return period. The return level is exceeded by the annual maximum with a probability q (Coles, 2001). The return level when $\xi \neq 0$ is calculated as follows:

$$z_q = \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-q)\}^{-\xi}] \tag{2.8}$$

and when $\xi = 0$

$$z_q = \mu - \sigma\log\{-\log(1-q)\} \tag{2.9}$$

which is valid for $0 < q < 1$.

Estimates for the return level are calculated when the maximum likelihood estimates from section 2.2.2 are substituted into equations (2.8) and (2.9). When the return level estimates ($\hat{z_q}$) are plotted, a return level plot is formed (Coles, 2001). Return level plots are a useful tool for model validation as well as presentation as it is simple to understand and interpret.

2.2.3 Threshold Excess

The threshold excess approach can be modelled using the generalised Pareto family of distributions and the interest under this approach is also in the shape parameter ξ.

The underlying principle of this method is to find a suitably high threshold u whereby any observation X_i (from an i.i.d sequence X) that exceeds it, is considered as a maximum observation. The threshold cannot be too low or too high. A threshold that is too low means that the asymptotic properties of the model are not satisfied while a too high threshold leaves too few observations to model (Coles, 2001) and will result in large variances. Tawn and Heffernan (2004) note that an appropriate threshold is one that does not produce significant shifts in estimates (excluding the increased variability) if the threshold was changed to a higher one. There are numerous ways to find a suitable threshold which include parameter stability plots, mean residual life (MRL) plots and Hill estimator. The parameter stability plots model the data using different values of threshold and assesses the stability of the parameter estimates. MRL plots will be used to determine the threshold values for this study. A detailed explanation of the Hill estimator can be found in Beirlant *et al.* (2004, pp.101-107). Similarly, see Scarrott and MacDonald (2012) for a comprehensive review of EVT threshold methods as well as mixture models.

An explanation of MRL plots, parameter estimation through MLE and details of the return levels which is similar to that of block maxima follows. There is an additional topic which is called declustering that relates to dependence with the excess observations (those observations above the chosen threshold).

MRL Plots

A way to visualise exceedances above a threshold is through the use of a mean residual life plot. A MRL plot according to Davison and Smith (1990) can be described as plotting the threshold, u, versus the mean observed excess over u. Formally put,

$$E(Y - u|Y > u) = \frac{\sigma - \xi u}{1 + \xi} \tag{2.10}$$

provided $\xi > -1, u > 0$ and $\sigma - \xi u > 0$.

The MRL plot has a slope of $\frac{-\xi}{1+\xi}$ with an intercept of $\frac{\sigma}{1+\xi}$ (Davison and Smith, 1990). The plot should follow a straight line with the previously mentioned slope and intercept. The threshold value is chosen at a point where there is a deviation in linearity in the plot which leaves room for subjectivity.

Maximum Likelihood Estimation

Suppose that the j excess values for a specific threshold u are $y_1, y_2, ..., y_j$ (Coles, 2001), then

When $\xi \neq 0$ the log-likelihood is:

$$l(\sigma, \xi) = -j\log\sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{j} \log\left(1 + \frac{\xi y_i}{\sigma}\right) \tag{2.11}$$

for $(1 + \frac{\xi y_i}{\sigma}) > 0, i = 1, 2, ..., j$; otherwise $l(\sigma, \xi) = -\infty$.

When $\xi = 0$, the log-likelihood is

$$l(\sigma) = -j\log\sigma - \frac{1}{\sigma} \sum_{i=1}^{j} y_i. \tag{2.12}$$

Return Levels

The p-observation return level z_p is the level exceeded on average once every p years Finkenstadt and Rootzen (2003) and is as follows for $\xi \neq 0$,

$$z_p = u + \frac{\sigma}{\xi}[(p\zeta_u)^{\xi} - 1] \tag{2.13}$$

and when $\xi = 0$

$$z_p = u + \sigma(p\zeta_u) \tag{2.14}$$

which is valid for $0 < p < 1$ and where $\hat{\zeta} = \frac{j}{n}$.

In order to estimate the return levels, the estimates of ξ, σ and ζ_u are needed. The estimates for ξ and σ are found using the maximum likelihood method by equations (2.11) and (2.12). ζ is the proportion of observations that are greater than u and is useful in estimating when extreme events will occur in the future.

Declustering

Within a stationary sequence, there is a tendency for excess observations to cluster together. This is especially applicable to environmental datasets. For instance, rainfall tends to exhibit a pattern where there is rainfall for a few consecutive days until it stops. These consecutive days of rainfall cause the extreme observations to be dependent and needs to be taken into consideration before being modelled. This is where declustering fits in - it segregates the independent excesses from the dependent excesses.

Declustering uses a threshold value to sort the data into clusters with minimum gaps, r, between the clusters. From these clusters, the maximum excesses within the different clusters are found and are assumed to be independent, thus the excesses are fitted to the GPD (Coles, 2001). In order to carry out the declustering analysis, the following terms need to also be defined:

The p-observation return level

$$z_p = u + \frac{\sigma}{\xi}[(p\zeta_u\theta)^{\xi} - 1]$$

with

$$\hat{\zeta}_u = \frac{n_u}{n} \tag{2.15}$$

and

$$\hat{\theta} = \frac{n_c}{n_u} \tag{2.16}$$

where n_u is the number of exceedances above the threshold u and n_c is the number of clusters obtained above the threshold (Coles, 2001).

2.2.4 Point Process

The point process approach also requires a suitable threshold value when modelling the data as the excesses follow a non-homogeneous Poisson process. However, this approach handles exceedances of times and values simultaneously instead of separately as in the threshold excess approach (Finkenstadt and Rootzen, 2003). The Poisson process handles randomly scattered excesses well and naturally contains points that are independent of each other (Coles, 2001). This approach is particularly applicable in the multivariate case. The block maxima and threshold excess approaches are a derivation from this approach.

Taking a random sample $Z_1, Z_2, ..., Z_n$ with a distribution function F (Smith, 1989), a point process can be defined as

$$N_n = \left(\frac{i}{(n+1)}, Y_{n,i}\right) \tag{2.17}$$

for all i and where $Y_{n,i} = \frac{(Z_i - b_n)}{a_n}$. N_n is defined on $\mathbb{R}^2$

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

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi\frac{z-\mu}{\sigma}\right)^{1/\xi} \tag{2.18}$$

valid for $1 + \frac{\xi(z-\mu)}{\sigma} > 0$.

When u is sufficiently large on (0,1) x $[u, \infty]$, N_n is said to follow a Poisson process.

Poisson processes can be classified as homogeneous and non-homogeneous where a homogeneous Poisson process has parameters μ, σ, ξ that are constant. Non-homogeneous processes allow for these parameters to vary (Finkenstadt and Rootzen, 2003).

Maximum Likelihood Estimation

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

$$\Lambda(A) = n_z(t2 - t1)\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right] \tag{2.19}$$

where n_z represents the number of years of the data.

The general form of the Poisson process is obtained when $[t_1, t_2] = [0, 1]$ is substituted into equation (2.19) (Coles, 2001),

$$L_A(\mu, \sigma, \xi; z_1, z_2, ..., z_n) = \exp\{-\Lambda(A)\} \prod_{i=1}^{N(A)} \lambda(t_i, x_i)$$

$$\propto \exp\left\{ - n_y\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \prod_{i=1}^{N(A)} \frac{1}{\sigma}\left[1 + \xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-1/\xi-1} \tag{2.20}$$

For a detailed description and derivation of the point process, see Coles (2001, pp.124 - 141).

Return Levels

The v-observation return level z_v is the level exceeded on average once every v years (Coles, 2001). It can be written by first defining p_i, when $1 + \xi_i\left(\frac{z_v-\mu_i}{\sigma_i}\right) > 0$

$$p_i = 1 - \frac{1}{n}\left[1 + \xi_i\left(\frac{z_v-\mu_i}{\sigma_i}\right)\right]^{\frac{1}{\xi_i}} \tag{2.21}$$

and when $\left(\frac{z_v-\mu_i}{\sigma_i}\right) < 0$ then $p_i = 1$.

It follows that z_v satisfies,

$$1 - \frac{1}{v} = \Pr\{\max(Z_1, Z_2, ..., Z_n \leq z_v)\} \approx \prod_{i=1}^{n} p_i \tag{2.22}$$

thus using a log form of equation (2.22) gives

$$\sum_{i=1}^{n} \log p_i = \log\left(1 - \frac{1}{v}\right). \tag{2.23}$$

2.2.5 Diagnostics

Validating the model and checking the model fit can be done through graphical methods such as probability-probability and quantile-quantile plots. Anderson-Darling and Kolmogorov-Smirnov (Davison and Smith, 1990; Chikobvu and Chifurira, 2015) statistics can be used as well.

2.3 Studies on EVT

To illustrate the use of the generalised distributions, several studies are described in this section. EVT spans many fields and is especially popular in the insurance and finance industry, hydrology and environmental studies. EVT is also applicable in engineering applications.

Insurance claims for an oil company were modelled using the threshold excess and point process approaches in a study by Smith (2004). The data spans over 15 years with 393 observations being used in the analysis. Varying thresholds are used in both approaches with the parameters in the point process approach held constant. The models from the threshold excess and point process approaches yield similar results. However, the point process model may be favoured because the parameters are unlikely to change with varying thresholds. Furthermore, inference on insurance data can be made on examination of the characteristics of the data and on the shape parameters. Insurance data is associated with long-tailed and skewed distributions which can be seen by looking at the shape parameter estimates which are close to 1 under both approaches (Smith, 2004). Furthermore, vehicle claim sizes from a reinsurance company are modelled and compared using various approaches in a study by Beirlant *et al.* (2001). The data consists of 371 observations and spans from 1988 - 2001. The thresholds are found through the Hill estimator instead of MRL plots but caution has to be exercised with this estimator. It is sometimes referred to as the horror plot as it does not reveal much at times (Mikosch *et al.*, 2006).

Value at Risk (VaR) is modelled using the block maxima and threshold excess approaches in a study by Singh *et al.* (2013). The period of the data ranges from 1973 - 2010 and consists of daily returns from the USA's Standard & Poors 500 and Australians Securities Exchange All Ordinaries indices. In addition to fitting the above two classical approaches, a version of the (dynamic) threshold excess approach which incorporates ARCH/GARCH methods is also used to include volatility of the returns. The threshold

excess fared better in modelling VaR with the dynamic approach providing results that adequately account for changes in market conditions. Similarly, daily returns for 3 companies - Citibank, General Electric and Pfizer for the period 1982 - 2001 are modelled to analyse VaR. A traditional point process and one with GARCH models to incorporate dependence are applied with a constant threshold to the returns. The study found that the inclusion of a GARCH process produces models with a better fit compared to when the GARCH is excluded from the analysis. The reason for this is because the volatility of the stocks are now captured in the models (Smith, 2004).

Continuing on the topic of VaR, Altun and Tatlidil (2015) model financial returns using GARCH-EVT models as at times such a series exhibit heavy tails which violates the normality assumption. However, McNeil and Frey (2000) noted that there are instances when the conditional distribution of returns can be light-tailed. A consequence of not taking the heavy tails into consideration when modelling is that the predications of VaR are often underestimated. The data used in the study returns from the Standard & Poors 500, Istanbul Stock Exchange - 100 and the Nikkei - 225 indices to compare the performance of GARCH-EVT model against four other GARCH models (GARCH-normal, GARCH-student-t, GARCH-GED and GARCH-SGED). Although the GARCH-EVT model can handle non-normal data, the model is outperformed by the other models tested.

In another instance, Davison and Smith (1990) modelled two different types of datasets in a study. The first was modelling of 3524 observations of high exposures of nuclear data. The threshold excess approach is used where varying thresholds are used, as covariates are included in the models. The data was split into a training set and a cross validation set. Results of the modelling showed that improved fit at higher threshold values as the shape parameter tends to display homogeneity among the models. The second study was on the exceedances from the River Nidd, a dataset that exhibited seasonality and dependence. The data spans the years 1934 - 1969 and varying thresholds are used with MRL plots to choose the suitable thresholds. Furthermore, graphical methods as well as formal tests were used to assess model fit. There diagnostic methods include quantile-quantile plots (graphical) and formal statistics namely Anderson-Darling and Kolmogorov-Smirnov tests. The GPD did not provide a good fit of the data with the chosen thresholds as the largest two observations deviated from linearity to a greater extent than the other excesses (based on looking at the quantile-quantile plot). Bayesian methods are also employed to construct confidence intervals for the return levels owing to a small sample size.

Reiss and Thomas (2007) analysed river data over a period of 31 years. The data is collected from the Moselle River which flows through Luxembourg, Germany and France. The threshold excess approach was used in this study and there is some notable dependence within the clusters of excesses. A constant threshold is used and the excesses are declustered with r=7 (expectation that flood levels can reach above a threshold for 7 consecutive days). This approach was found to be suitable for modelling the flood levels at the Moselle River.

Hourly reading of ozone levels were analysed as in an attempt to look at ways of controlling air pollution in Houston, Texas in the USA from 1973 - 1986. Non-homogeneous Poisson process models were fitted with a constant threshold as well as with varying thresholds as part of a sensitivity analysis. The idea here was to model the trend within the data which is a form of covariate (Reiss and Thomas, 2007).

A South African EVT study compared the performance of a Generalised Pareo-type model with a GPD model in modelling the extreme daily peak electricity demand. The daily peak demand data for years 2000 to 2009 was used while restricting the period to only the winter months as this is when electricity demand is at its peak. The study found that while the Generalised Pareto-type model had a simpler structure, the GPD model provided an improved fit for this dataset (Chikobvu *et al.*, 2012).

Moreover, a popular area in EVT is on climate modelling. For instance, in a study by Nadarajah and Choi (2007), the block maxima approach was used in the analysis of annual maxima of daily rainfall in South Korea. The Gumbel distribution was found to be suitable in modelling the rainfall maxima. The data covered the period from 1961 to 2001 for five locations in South Korea to provide a fairly accurate representation of the country. While these estimates are a sufficient measure for predicting the probability of future flood damage for this particular study, the structure and climate of South Korea may differ from the Western Cape. This could lead to different distributions being used to describe the behaviour of the tails.

Rainfall maxima in Germany are fitted to models under the non-homogeneous point process (with a constant threshold) and threshold excess approaches (with covariates to account for non-stationarity). Friederichs (2010) compared the performance of these two approaches with constant and varying shape parameters. The data used for this study consisted of daily rainfall in mm recorded at 2000 stations from the German National Meteorological service for the period from 1958 to 2001. The models where the shape parameter remained constant provided results with less uncertainty and less complexity compared to models where the shape parameter varied.

In a further study, the daily temperatures in Malaysia were analysed by using models from the block maxima approach. The dataset spans from 1981 to 2012 and two types of models are applied. In the first model - referred to as the classical EVT model - all parameters are kept constant. In the second model, covariates are included. The rationale behind the difference in models is to handle any non-stationarity within the data (Hasan *et al.*, 2013). Similarly, Finkenstadt and Rootzen (2003) suggest removing the seasonal trend before modelling the data as a method for handling non-stationarity. Another method proposed by (Finkenstadt and Rootzen, 2003; Coles, 2001; Smith, 1989) is to use separate season models and with varying thresholds in case of the threshold excess and point process approaches. Overall, only once an exploratory analysis of the data is performed, can the appropriate model(s) be selected.

In another study examining daily maxima of temperature, both the GPD and point process distributions suitably modelled the data. Three sites are used in the California

Central Valley during summer for the years from 1951 to 2005 (Katz and Grotjahn, 2011). There was no spatial analysis performed in the study, thus the distances between the stations are irrelevant. However, it was concluded that the temperature maxima were similar across the Valley (Katz and Grotjahn, 2011), perhaps because the sites exhibited similar characteristics to each other. Against this background, a spatial analysis of the data is not performed in this study. Accordingly, the distances between the stations is not factored into the analysis. Instead, the weather stations are analysed individually.

While EVT does not strictly apply to maxima, only maxima is modelled and analysed in this study. The minima are also included as extremes and although there are far less studies on minima than there are on maxima, it can be useful in the modelling of extreme events. For instance, the effect of warmer temperatures due to urbanisation is looked at by analysing a series of daily minimum temperatures in Pheonix, Arizona in the USA. The block minima approach is fitted to the data which ranges from 1948 to 1990 and is restricted to the summer months. In this example, the trend component is fitted using the block maxima method. It is concluded that the best fitting model is achieved when trend is modelled in the location parameter μ. Reiss and Thomas (2007) who fitted the models, also noted that similar results would be achieved under the point process approach. Just as trend is a form of covariate, cycles within data can also be considered as a covariate (Reiss and Thomas, 2007).

Coles and Walshaw (1994), modelled wind speeds together with direction as it has an impact on buildings. Hourly maximum wind gust speeds in the United Kingdom are used. A period of six years (1975-1980) was analysed by modelling the data to a GEV distribution. Furthermore, modelling extreme wind speeds and the sea level extremes are of importance for environment control and disaster management. Wind speeds can help clear up the air by clearing out the pollution, whereas the high sea levels can cause floods to occur (Coles, 2001). In this case the flood would be the extreme event and the return levels help in estimating the next occurrence.

A disadvantage of the block maxima approach is that it requires the data to be split into blocks of equal length when applying it to the data. The problem that arises is that because only maxima values are being used, the blocks could be too small or too large. These problems could cause bias in estimation and extrapolation or a large estimation variance respectively. Coles (2001) mentions that the block maxima approach could also lead to the distribution assumptions being violated such as the data all having a common distribution. The aforementioned challenge with the block maxima was encountered in a previous Honours research project which focused on a univariate analysis of rainfall over a 20 year period (1980 - 2014) in the Western Cape. It was found that the block maxima approach does not fare well in modelling extremes as it also does not effectively make use of the available data. Therefore, it may be concluded the use of threshold excess and point process approaches are more effective and accurate ways of modelling extremes.

One of the challenges with the threshold excess and point process approaches is that the manner in which the threshold values are chosen comes with a degree of subjectivity. So the choice of a suitable threshold for a dataset may vary from investigator to

investigator. However, from the classical EVT approaches the drawbacks from the block maxima approach pose more of an issue than from the point process and threshold. An advantage of the point process approach is that it characterises the exceedances under a Poisson rate with the parameters being independent of the threshold (Ledford and Tawn, 1996; Scarrott and MacDonald, 2012). For this reason, the point process models can incorporate non-stationarity and random effects.

Moreover, a univariate analysis often may not accurately reflect the reality of conditions being modelled, therefore employing techniques of a multivariate nature are better suited. In other words, using multiple variables together in the analysis can provide results that are more accurate.

2.4 Multivariate Extreme Value Theory

MEVT involves two or more dependent random variables where the interest is in modelling the joint distribution of these variables as noted by R.L Smith (Galambos *et al.*, 1993, p.232). The shift away from the univariate analysis to a multidimensional space comes with an increased complexity that needs to be taken into account. An important determination is deciphering the term *multivariate* in an EVT context. The term *multivariate* can refer to a single variable at different locations; a single variable at different times or multiple (different) variables at the same location. The focus of this study is on multiple variables at a single location.

The next question is how to capture any dependence among the variables and how to incorporate it into the analysis. Building onto the aforementioned subject, Beirlant *et al.* (2004) and Tawn (1988) reinforce the fact that a finite set of parametric distributions does not exist to characterise the dependence structure within MEVT.

In terms of modelling multivariate extremes, Beirlant *et al.* (2004) and Coles (2001) suggest splitting the analysis into two parts which look at the marginal distributions and dependence structure separately. The first part concerning the marginal distributions can be handled using the univariate techniques that have been explained in the previous section. The dependence structure is modelled using a range of models described further in this section. Alternatively, the marginal and dependence structure can be estimated simultaneously which is an efficient method but with the added downside of a more complicated likelihood (Stephenson *et al.*, 2005). In the case where the marginals and dependence structure is estimated separately, the marginal distribution will be known. Whereas on the other hand, unlike the case where the marginals and dependence structure are estimated simultaneously, the marginal distribution is unknown. The approaches to modelling the data extend from the univariate case, namely component-wise maxima, point process and threshold excess approaches. As seen in the univariate section, the component-wise approach consists of the data being partitioned in different blocks of equal length whereby the maxima/minima for each block becomes the series of data. As with the univariate cases, the methodology explained can be applied for minima as well, although there is mention of maxima only. Coles *et al.* (1999) mentions that a component-wise approach is applicable when the data consists

of annual maxima only. There is also a choice between using parametric methods or non-parametric methods for the modelling of the data.

For example, in a study by Zhang and Singh (2007) copulas are used to model hourly rainfall in Louisiana in the United States of America specifically from Amite River basin over the period 1960-2001. Copulas can be used to model extremes without requiring the assumptions that the parametric methods need. However, Coles *et al.* (1999) noted non-parametric methods do not always fulfil the asymptotic characterisations required in the extreme value models as a reason to prefer parametric methods. The parametric methods are also less complicated than the non-parametric methods and were used in this study. Similarly, Smith *et al.* (1990) noted that a primary use of non-parametric methods is to indicate which appropriate parametric dependence models to use in the analysis.

On inspection of parameter estimation, as in the univariate case, there are a variety of techniques for parameter estimation in these models such as method of moments, probability of weighted moments, Bayesian methods and maximum likelihood estimation. In an earlier study by Gumbel and Goldstein (1964), the method of moments was used for parameter estimation of two different sets of data. MEVT was applied to the first dataset which consisted of the oldest ages at death among males and females from Sweden during 1905 to 1958. The second dataset has to do with flood data at the Ocmulgee River in Georgia. This data of annual discharges was from two stations - one upstream and the other downstream from 1910 to 1948. The most common approach, however, is to use the maximum likelihood method (Khuluse, 2010; Nadarajah and Choi, 2007; Coles, 2001; Kotz and Nadarajah, 2000 ; Tawn, 1988; Prescott and Walden, 1980). This method is popular and the most applicable as it produces estimators that are invariant to the marginal distribution (Tawn, 1988). The point process approach is able to handle situations where there are exceedances of a threshold in one variable and both variable, hence the use of a Poisson likelihood. The threshold excess model is only able to handle situation when there is exceedance in both variables only. Therefore, a censored likelihood can be used Beirlant *et al.* (2004). This means there are more observations that can be modelled under the point process approach.

Consequently, there are significant disadvantages to the component-wise maxima approach as Coles *et al.* (1999) mentions that the component-wise approach is appropriate when only the annual maxima are available. Thus, when there is a full dataset available, the threshold excess and point process approaches should be favoured. Firstly, the data is reduced to only one set of observations per block which means discarding valuable data from the analysis. Secondly, there is no way of telling whether the set of maxima occurs simultaneously or not (Beirlant *et al.*, 2004). The uncertainty of when the maxima occurs presents some difficulty in accurate depiction of extreme events.

The applicability of MEVT as opposed to univariate EVT provides deeper insight and improved realities of the environmental processes. Thus, the main focus of this study is on the multivariate aspect of EVT. The multivariate theory section starts with the details of the marginal transformations and is followed by explanations of the component-wise,

threshold excess and point process approaches. There is a dedicated category for dependence as it is the essence of a multivariate EVT analysis. This section concludes with studies on MEVT.

Details of the marginal transformation to standard Fréchet margins are explained in the next section as it forms part of the first step of fitting MEVT models to a dataset.

2.4.1 Marginal Transformation

To achieve unit Fréchet margins $X_{i,j}$ for random vectors $(\tilde{X}_{i,1}, \tilde{X}_{i,2}, ..., \tilde{X}_{i,d})$, the following transformation defined by Coles and Tawn (1994) is used,

when $\tilde{x}_j > u_j$

$$X_{i,j} = -\left(\log\left[1 - \{1 - F_j(u_j)\}\left\{1 - \xi_j\left(\frac{\tilde{X}_{i,j} - u_j}{\sigma_j}\right)\right\}^{\frac{1}{\xi_j}}\right]\right)^{-1} \tag{2.24}$$

and when $\tilde{x}_j \leq u_j$

$$X_{i,j} = \frac{-1}{\log F_j(\tilde{X}_{i,j})} \tag{2.25}$$

for $i = 1, 2, ..., n$ and $j = 1, 2, ..., d$.

F_j is estimated by the empirical transformation for $\tilde{X}_j$ which is defined as $F_j = \frac{R(\tilde{X}_j)}{(n+1)}$ where $R(\tilde{X}_j)$ is the rank of $\tilde{X}_j$ (Coles and Tawn, 1991). The estimates for the remaining parameters are calculated through MLE. The standardised Fréchet margins are used in the component-wise, threshold excess and point process sections that follow.

An important note about the transformations is that the marginals can take on the form of a variety of distributions, it does not have to be from the Fréchet distribution.

2.4.2 Component-Wise

Similar to the univariate approach of block maxima, the observations are split in blocks where the maxima taken from each block form the dataset to be modelled. In a multivariate case, the maxima of the number of variables are taken per year to create the component-wise dataset.

Defining $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ as an independent sequence of random vectors which have standard Fréchet margins with a distribution function $F(x, y)$, the limiting joint distribution is

$$M_n = \left(\frac{\max\{X_i\}}{n}, \frac{\max\{Y_i\}}{n}\right) \quad \text{for } i=1,\ 2,..,n$$

(M_n is the renormalised vector).

$$Pr\{M_{x,n} \leq x, M_{x,n} \leq y\} \to G(x, y) \tag{2.26}$$

$$G(x,y) = \exp\{-V(x,y)\} \tag{2.27}$$

where

$$V\{x,y\} = 2\int_0^1 \max\left(\frac{w}{x}, \frac{1-w}{y}\right) dH(w) \tag{2.28}$$

for $x, y > 0$.

G is a non-degenerate distribution function and H is a distribution function on [0,1] which satisfies the following mean constraint

$$\int_0^1 w dH(w) = \frac{1}{2} \tag{2.29}$$

H in equation (2.28) can be differentiable or non-differentiable. More detail can be found in Coles (2001).

Maximum Likelihood Estimation

Coles (2001) considers a sequence of independent maxima with standard Fréchet margins (empirical transformation in section 2.4.1 $(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)$ that have been sorted into the desired blocks, where the probability density function is defined as

$$g(x,y) = \{V_x(x,y)V_y(x,y) - V_{xy}(x,y)\} \exp\{-V(x,y)\} \tag{2.30}$$

provided $x > 0, y > 0$ and where V_x V_y are partial derivatives with V_{xy} as the mixed derivative of V.

The likelihood is written as

$$L(\theta; (x_1, y_1), (x_2, y_2), ..., (x_m, y_m)) = \prod_{i=1}^{m} g(x_i, y_i) \tag{2.31}$$

where the series are considered as independent sequences that follow the GEV distribution.

2.4.3 Threshold Excess

A multivariate threshold excess approach requires a suitable threshold for each of the variables. The threshold can be chosen using the univariate techniques mentioned in section 2.2.3. Another technique for finding the threshold in a bivariate case is to use a bivariate threshold choice plot, details of which can be found in Beirlant *et al.* (2004).

Defining $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ as an independent sequence of random observations with a distribution function $F(x, y)$ with standard Fréchet margins as defined in section 2.4.1. The marginals are valid for when $X > u_x$ and $Y > u_y$. Coles (2001) states that under the homogeneity property of V from equation (2.27), for large n, it follows that

$$F(x,y) = \left\{F^n(x,y)\right\}^{\frac{1}{n}} \approx \left[\exp\left\{-V\left(\frac{x}{n}, \frac{y}{n}\right)\right\}\right]^{\frac{1}{n}} = \exp\{-V(x,y)\} \tag{2.32}$$

thus

$$F(x,y) \approx G(x,y) = \exp\{-V(x,y)\} \tag{2.33}$$

for $x > u_x$ and $y > u_y$ as per equation (2.27).

Maximum Likelihood Estimation

The likelihood for observations $x_1, x_2, ..., x_d$ that the components $j_1, j_2, ..., j_n$ exceed the thresholds according to Ledford and Tawn (1996) can be shown as

$$\left.\frac{\partial^n F(y_1, y_2, ..., y_d)}{\partial y_{j_1}, y_{j_2}, ..., y_{j_n}}\right|_{y_j=\{\max(u_j,x_j),\ j=1,2,...,d\}} \tag{2.34}$$

In the bivariate case, the marginal observations which do not exceed the relevant thresholds, are censored. Thus the contribution from these observations to the censored likelihood is an indication that it falls below the respective thresholds (Ledford and Tawn, 1996). The thresholds for each of the variables enable the outcome space to be broken up into 4 regions which is shown in Figure 2.1 (reproduced using the *wavesurge* dataset in the *ismev* package in the program R). Since the focus is on joint exceedances, region $R_{1,1}$ is of interest.

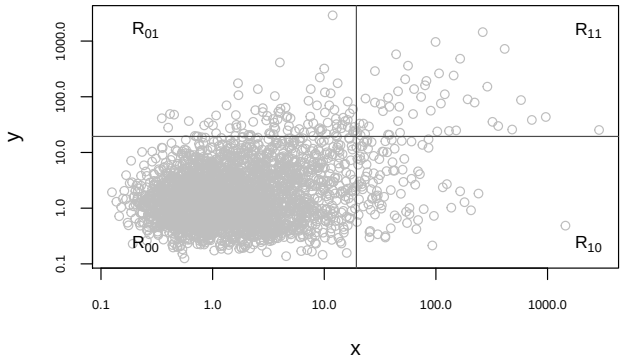


Figure 2.1: Regions contributing to the likelihood.

The likelihood with the unknown parameter θ is

$$L(\theta; (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)) = \prod_{k=1}^{n} \psi(\theta; (x_k, y_k)) \tag{2.35}$$

where equation (2.34) would be applied to the specific region.

2.4.4 Point Process

The point process approach uses a threshold as described in section 2.2.3. The threshold values for this approach and the threshold excess approach intersect the x and y axes at

the same points. This is to allow for comparison of models across approaches. One of the differences between the point process approach is that there is change to pseudo-polar coordinates which enables a curved threshold boundary to be formed. The threshold excess approach has a linear threshold boundary. Figures 3.23 - 3.30 in Chapter 3 show the difference in threshold boundaries for the two approaches.

Defining $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ as an independent sequence of random observations (standard Fréchet margins as seen in section 2.4.1) with a distribution function $F(x, y)$, the marginals have a distribution that follows on from equation (2.26).

The point process sequence N_n, as defined by Coles *et al.* (1999), is of the form

$$N_n = \left\{ \left(\frac{x_1}{n}, \frac{y_1}{n} \right), \left(\frac{x_2}{n}, \frac{y_2}{n} \right), ..., \left(\frac{x_n}{n}, \frac{y_n}{n} \right) \right\}$$

for $i = 1, 2, ..., n$ as $n \to \infty$ and $N_n \to$ a Poisson process, N on $\mathbb{R}^2_+$.

Transformation to pseudo-polar coordinates allows for a simpler representation of the limiting process. The measure of distance from the origin r and the measure of angle w on a [0,1] scale (Coles, 2001) is defined as

$$r = x + y \quad \text{and} \quad w = \frac{x}{r} \tag{2.36}$$

the intensity function of N is represented as

$$v(dr \text{ x } dw) = \frac{dr}{r^2} \text{ x } 2dH(w). \tag{2.37}$$

The H in the intensity measure is related to equation (2.27).

The use of the Poisson process is assumed to be an approximation to N_n in a suitable region. For instance, for a suitable choice of r_0, in the region $A = \left\{ (x, y) : \frac{x}{n} + \frac{y}{n} > r_0 \right\}$, the following result is achieved

$$\Lambda(A) = 2 \int_A \frac{dr}{r^2} dH(w) = 2 \int_{r=r_0}^{\infty} \frac{dr}{r^2} \int_{w=0}^{1} dH(w) = \frac{2}{r_0} \tag{2.38}$$

For the details limiting distribution of the component-wise derivation from the point process, refer to Coles (2001, p.157).

Maximum Likelihood Estimation

A non-homogeneous Poisson process of N_n which is in a region A, the likelihood over A is

$$L_A \left(\theta; \left\{ \frac{\boldsymbol{X_j}}{n} \right\} \right) = \exp\{-\mu(A)\} \prod_{j=1}^{n_A} \mu(dr_j \text{ x } d\boldsymbol{w_j}) \tag{2.39}$$

provided H is differentiable and where r_j and $\boldsymbol{w_j}$ are shown in equation (2.36).

The region A is bounded from $\mathbf{0}$ by a distance that is determined by the rate of convergence and θ defines the measure parameters (Coles and Tawn, 1991). The curved threshold is shown in Figure 2.2 (reproduced using the *wavesurge* dataset in the *ismev* package in the program R)

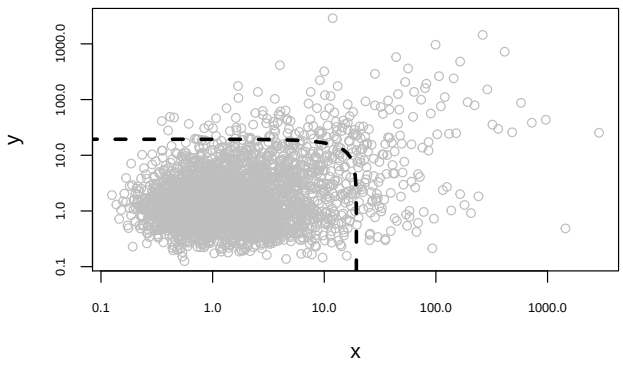


Figure 2.2: Regions contributing to the Possion likelihood.

The observations above the curved threshold in Figure 2.2 contribute to the Poisson likelihood.

2.5 Dependence Structure and Diagnostics

2.5.1 Dependence

There are numerous parametric models available to capture the dependence structure with the MEVT models however only a few of these models are used in this study. MEVT models cannot account for dependence in any other way besides classifying the structure as asymptotically dependent or perfectly independent (Coles *et al.*, 1999). Symmetric and asymmetric models are included and according to Coles and Tawn (1991) cover the different possible dependence structures encountered in datasets. The models presented in this section are chosen based on what has been widely used in studies (Beirlant *et al.*, 2004; Coles, 2001; Bortot *et al.*, 2000; Coles and Tawn, 1994; Joe *et al.*, 1992; Coles and Tawn, 1991; Joe, 1987). Morton and Bowers (1996) also noted that accuracy of the estimates are not affected critically and that dependence structure is described equally well using any model. For further dependence models, refer to the aforementioned cited studies as explaining all existing models would require a few chapters on its own.

For the family of models given below, α and β are a measure of the dependence strength.

i. Logistic model introduced by Gumbel (1960):

$$G(x,y) = \exp\{-(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}\} \quad x > 0, y > 0 \tag{2.40}$$

for $\alpha \in (0,1)$,

with a density function (when H is differentiable),

$$h_1(w) = \frac{1}{2}(\alpha^{-1}-1)\{w(1-w)\}^{-1-\frac{1}{\alpha}}\{w^{\frac{-1}{\alpha}} + (1-w)^{\frac{-1}{\alpha}}\}^{\alpha-2}. \tag{2.41}$$

ii. Negative logistic model introduced by Galambos (1975):

$$G(x,y) = \exp\left\{\frac{1}{x} + \frac{1}{y} - (x^{\alpha}+y^{\alpha})^{\frac{1}{\alpha}}\right\} \quad x>0, y>0 \tag{2.42}$$

for $\alpha > 0$,

with a density function

$$h_2(w) = (1+\alpha)\{w(1-w)^{-\alpha-2}\}\{w^{-\alpha}+(1-w)^{-\alpha}\}^{-\frac{1}{\alpha}-2} \tag{2.43}$$

iii. Bilogistic (asymmetric) model introduced by Joe *et al.* (1992):

$$G(x,y) = \exp\{-xq^{1-\alpha} - y(1-q)^{1-\beta}\} \tag{2.44}$$

with a density function

$$h_3(w) = \frac{1}{2}(1-\alpha)(1-w)^{-1}w^{-2}(1-q)q^{1-\alpha}\{\alpha(1-q)+\beta q\}^{-1} \tag{2.45}$$

for $0 \le w < 1$, $0 \le \alpha < 1$, $0 \le \beta < 1$ where $q = q(w;\alpha,\beta)$ is a solution to equation

$$(1-\alpha)(1-w)(1-q)^{\beta} - (1-\beta)wq^{\alpha} = 0. \tag{2.46}$$

The term $(\alpha - \beta)$ provides a measure of asymmetry in the dependence because the logistic model is formed when $\alpha = \beta$.

iv. Negative bilogistic introduced by Coles and Tawn (1994):

$$G(x,y) = \exp\{-x - y + xq^{1+\alpha} + y(1-q)^{1+\beta}\} \tag{2.47}$$

for $\alpha > 0$ and $\beta > 0$,

with a density function

$$h_4(w) = -\frac{(1-\alpha)(1-q)q^{1-\alpha}}{(1-w)w^2\{(1-q)\alpha + q\beta\}} \tag{2.48}$$

with $q = q(w;\alpha,\beta)$ as per equation (2.46).

v. Dirichlet (asymmetric) model introduced by Coles and Tawn (1991)

$$G(x,y) = \exp\{-x[1-\Gamma(q;\alpha+1,\beta)] - y[1-\Gamma(q;\alpha,\beta+1)]\} \tag{2.49}$$

for $\alpha > 0$, $\beta > 0$ where $q = \frac{\alpha y}{\alpha y + \beta x}$ and $\Gamma(q; \alpha + 1, \beta)$ is evaluated at q.

The corresponding density function is

$$h_5(w) = \frac{\alpha\beta\Gamma(\alpha + \beta + 1)(\alpha w)^{\alpha-1}\{\beta(1-w)\}^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)\{\alpha w + \beta(1-w)\}^{\alpha+\beta+1}}. \tag{2.50}$$

The strength of dependence for each model (according the parameter) is summarised in Table 2.1.

Table 2.1: Strength of Dependence

Model	Dependence	Independence
Logistic	$\alpha \to 0$	$\alpha \to 1$
Negative logistic	$\alpha \to 0$	$\alpha \to \infty$
Bilogistic	$\alpha = \beta \to 0$	$\alpha = \beta \to 1$
Negative bilogistic	$\alpha = \beta \to 0$	$\alpha = \beta \to \infty$
Dirichlet	$\alpha = \beta \to \infty$	$\alpha = \beta \to 0$

2.5.2 Diagnostics

When the decision relates to model selection of dependence structure, then for nested models, standard likelihood ratio tests can be employed. For non-nested models, a goodness of fit statistic such as the Akaike Information Criterion (AIC) can be used (Vuong, 1989; Coles and Tawn, 1991).

AIC is based on the log-likelihood function and adjusts for the number of parameters that are estimated Dobson and Barnett (2008) which is as follows

$$AIC = -2\log(L) + 2p \tag{2.51}$$

where p represents the number of estimated parameters and L denotes the likelihood.

2.6 Studies on MEVT

There are many studies on MEVT that span numerous different fields and is widely used in environmental applications. A majority of the publications are limited to bivariate studies with a few extending the study to a trivariate study. The difference between the bivariate and trivariate estimates is seen by narrower profile likelihood intervals for the shape and scale parameters. For a comprehensive list of dependence models see Coles and Tawn (1991).

There are very few studies on trivariate studies as the increase in complexity in computation of MLE and dependence models is to a greater extent compared to the bivariate models. In an oceanographic application, the hourly surge levels at three coastal sites (Lowestoft, Immingham and Sheerness) in Britain for the periods 1970 - 1976 and 1980 - 1988 are modelled using the point process approach. Bivariate and trivariate models are applied to the data. The choice of the dependence models are 'informally' determined

through a plot of r and w using equation (2.36), thus the logistic, negative logistic and, symmetric Dirichlet and Dirichlet models are used to capture the dependence structure. In the trivariate case, the time series logistic model results in adequately fitting model as the "structure of this model closely resembles the dynamic process of surge propagation" (Coles and Tawn, 1991).

In addition, Tawn (1990) used a logistic model to capture the dependence structure among the maximum sea levels at three sites on the coast of England. Bivariate extreme value distributions were fitted to pairwise combinations of data from the sites, namely Southend, Kings Lynn and Sheerness. The data was for a forty year period and extensions to trivariate models was also performed in this study. It was noted that the extremes did not always occur from the same particular extreme weather event e.g. storm, because of the location of the sites. This means that there were instances where extremes recorded during a weather event could be traced at two of the stations that were closer to each other. Hence, a spatial component plays a role in the dependence structure as sites that are closer to each other would exhibit stronger dependence compared to sites that are further away from each other.

A further study of surge levels at a different port i.e. Newlyn in Cornwall was conducted by Coles and Tawn (1994). Bivariate models are fitted to wave height (recorded every three hours) and surge levels (hour records) from 1971 - 1977 to assess the structure design of sea walls. MRL plots are utilised for marginal threshold selection with six different dependence models being used which are the logistic, negative logistic, bilogistic, negative bilogistic (proposed in the study), symmetric Dirichlet and Dirichlet models. Model selection is executed by looking at the negative log-likelihood values for nested models and AIC for non-nested models. The negative bilogistic model provided the best fit for this dataset. Even though the surge levels are analysed in two different studies, the difference in the manner in which the variables were modelled plays a role. The aforementioned wave and surge dataset was used in the study by Dixon and Tawn (1995) who wanted to improve on the classical threshold excess models by proposing a semi-parametric approach where more conservative thresholds lead to improved estimates. MRL plots are also used in this study to determine a suitable threshold.

For instance, Coles and Tawn (1991) model the same variable at different sites, while Coles and Tawn (1994) model two different variables at one site. Both studies are multivariate in character and such disparities should be taken into account when analysing results. In a similar instance, Tawn (1988) analyse the annual maxima of sea levels at the Lowestoft and Sheerness ports in Britain for the assessment of flood detection. The data is standardised to Gumbel margins as there are claims in the study that there is more stability in this transformation compared to exponential margins. Mixed, asymmetric mixed, logistic and asymmetric logistic models are applied to characterise the dependence structure. In this case the selected model is the asymmetric logistic model.

Studies on multivariate datasets using the component-wise approach are few and infrequent. The reason for the lack of studies can be attributed to the inefficient use of the data and disaccord of grouping of the maxima during any block. Stephenson *et al.*

(2005) offers an improvement to the traditional component-wise approach by incorporating the dates of the occurrence of the maxima into the models. The dates for annual maxima of sea levels at three sites - Dover, Newlyn and Harwich - in England for a period of 1912 - 1992 are incorporated into the analysis. The sites are fairly far from each other between 90km and 500km apart and no spatial analysis is performed. The data is standardised to have Gumbel margins. Based on pairwise plots of the data at each site, the logistic model is chosen to represent the dependence structure with a linear trend model for the location parameters to account for the trend in the sea level maxima. The inclusion of dates into the component-wise models allows for inference on whether the occurrence of a particular extreme event was captured by recorded maxima at the sites at the same time. This of course presents an advantage over the somewhat wasteful traditional component-wise models. In addition to the bivariate model fitted, a logistic trivariate model as defined by Coles and Tawn (1990) is also fitted to the data which allows for increased inference on the dependence between the sites. Overall the inclusion of dates in the component-wise models is a simple step that provides increased inferential capabilities.

In a similar instance, the maxima wind speed and wave heights were analysed from 1990 - 1994. The study looked at the two variables to assess the mooring forces at the Shell UK Exploration and Production North Cormorant Platform in the northern North sea under extreme conditions. Bivariate extreme value distributions were used, specifically the point process approach using only the logistic model to capture the dependence structure between the variables (Morton and Bowers, 1996). The interest of the study was to determine the impact of the dependence between the variables as extreme levels of wind speeds and wave heights contribute to mooring forces. Thus, the same model was applied to two other sites. In particular, Shell UK Exploration and Auk Cormorant Platform and on a databuoy at West Shetland in the United Kingdom for the years 1990 - 1994 and 1984 - 1988, respectively. The outcome showed the dependence between the variables remained the at a constant level at all the sites even though there were different time periods used in one case.

Furthermore, in a technical report by Joe (1989), two different air pollutant datasets were examined. The first dataset consisted of extreme ozone concentrations from 1983 to 1987. The data was collected from five monitoring sites in the San Francisco Bay Area. To counter the seasonality problem, certain months were excluded from the analysis. The second dataset consisted of nitrogen dioxide, sulphur dioxide and extreme ozone concentrations collected from four stations in the Great Vancouver district and the data ranged from 1984 - 1987. The multivariate analysis of the air pollutants was conducted by modelling the marginals and dependence structure separately through the use of copulas to fit multivariate non-normal data.

Moreover, a new conditional method proposed by Tawn and Heffernan (2004), the effects of air pollution on health are investigated by modelling different pollutants together. A conditional method is proposed in this study to model extreme levels of air pollutants as this method can be readily applied to problems where there are more than three variables as opposed to the existing methods. This entails fitting a MEVT model where

the pollutants are conditional on a particular pollutant exceeding a certain level. The data that is used consists of the maxima of hourly means for oxides, nitrogen dioxide, nitrogen oxide, sulphur dioxide and particulate matter in Leeds in the United Kingdom. To deal with seasonality in the data, it is split into summer and winter seasons with separate models being fit to each period. Whether the conditional method outweighs the classical methods remains to be determined.

In another study by Zheng *et al.* (2014) which compared the performance of the point process, threshold excess and conditional methods for risk analysis - the results showed an underestimation of the return levels. An advantage of the conditional method over the classical threshold excess approach is that conditional models are able to handle situations where there is exceedance is at least one margin.

In a similar area, sulphate and nitrate concentrations were measured at single wet deposition monitoring station in the USA. The data consists of 504 pairwise sulphate and nitrate concentrations. In this study, Joe *et al.* (1992) modelled the data to compare parametric point process models using log-likelihoods and non-parametric methods of polar coordinates. Two parametric dependence models were used, namely the logistic and bilogistic models. In this case both the parametric and non-parametric methods showed that the sulphate and nitrate data are best modelled with a bilogistic model capturing the dependence structure.

Certain multivariate datasets exhibit near independence between the variables. Ledford and Tawn (1996) provide a foundation for characterising cases like this through techniques explained in section 2.5.2. The dataset of wave and surge levels from Coles and Tawn (1994) contrasted with rainfall and wind data at Eskdalemuir Observatory in Scotland which has first been analysed by Anderson and Nadarajah (1993). Logistic bivariate models are fitted to both sets of data, one with varying thresholds. The results show a stronger bias for the rainfall and wind data suggesting that asymptotically dependent models were not appropriate for that dataset. On the other hand, the wave and surge data showed a good fit with the asymptotic dependence models.

In contrast, bivariate extreme value models were used to monitor the vital signs of patients by using a heart rate and breathing dataset. This data was collected at the University of Pittsburgh Medical Center using 332 high-risk adult patients. The focus on modelling the extremes of the vital signs is that when abnormally high levels are reached, it would serve as an indication of the occurrence of an imminent adverse event. Both a bivariate guassian extreme value and classical threshold models were used in this study to determine which performs better in detecting the possibility of an adverse event occurring (Hugueny *et al.*, 2009). In this study, the bivariate guassian model outperformed the classical threshold model.

From our research of the literature, there have been no studies on the MEVT with an application to climate data in the Western Cape province.

Chapter 3

Application of MEVT to Climate Data in the Western Cape Province

3.1 Data

The weather datasets analysed in this study are daily readings of maximum rainfall, daily maximum temperature and daily maximum wind speed for stations across the Western Cape from 1965 to 2015. These stations include Plettenberg Bay, Vredendaal, Langebaanweg Cape Town International Airport and George Airport. The weather data was provided by the South African Weather Services (SAWS). Table 3.1 shows the characteristics of the stations. The temperature was recorded in degrees Celsius (^{0}C), rainfall in millimetres (mm) and wind speed in metres per second (m/s) where 1 m/s is equivalent to 1.944 knots or 3.6 km/h. Wind speed was recorded at either of the times 08:00, 14:00 or 20:00.

The data is provided in an Excel spreadsheet with the following symbols and explanations summarised in Table 3.1.

Table 3.1: Summary of data symbols with explanations

Symbol	Explanation
—	Data is not yet available
(empty)	No rainfall for that day
0.0	Calm conditions for wind speed
***	Missing data
=	Values are unreliable due to missing daily values
"A" or "B"	No rainfall recorded for that day but if there was rainfall, it is incorporated into the total amount for that period.
C	Rainfall reading for that day was accumulated over a number of days.

Rainfall is only recorded if the amount of rain was greater than 0.1 mm. The blank cells

for rainfall are recoded to a zero which means that no rainfall fell that day. The cells with "A", "B" and "C" are also recoded and left as a blank cell indicating that the data for that day is missing and therefore unreliable.

In order to satisfy the stationary criteria mentioned in section 2.2.1 of Chapter 2, the data is broken up into the four seasons namely, Summer (December - February), Autumn (March - May), Winter (June - August) and Spring (September - November).

3.1.1 Missing Values

If data for a day was missing then the average of the day before and after was taken as the value for that missing day. Moreover, if data for two or more consecutive days missing then the moving averages were used for the missing days. If however, data over a month or thirty one consecutive days were missing then the data is excluded from the analysis. Three of the stations only have data that starts later than 1965, therefore not much can be done about those missing years worth of data.

The data is rearranged so that the daily dates are put into the first column with the observations of all the subsequent stations in the columns next to it. The monthly and annual maxima for each the different stations are also found. A summary of the characteristics of the five stations is shown in Table 3.2 and a map of the Western Cape province with the locations of the stations is shown in Appendix A.

Table 3.2: Characteristics of the five weather stations

	Longitude (^{0}C)	Latitude (^{0}C)	Altitude (m)	Period
CT Airport	18.597	-33.969	44	01/01/1965 - 18/05/2015
George Airport	22.381	-34.004	193	01/01/1966 - 18/05/2015
Langebaanweg	18.157	-32.972	31	01/01/1980 - 12/08/2014
Plettenberg Bay	23.325	-34.089	138	01/01/1988 - 18/05/2015
Vredendal	18.496	-31.673	42	01/01/1965 - 18/05/2015

3.2 Findings

The results of modelling the data are shown below. It begins with an exploratory analysis of the data per variable per station and is then categorised further into the seasons per variable per station for the univariate analyses. The Phillips-Perron test for stationarity is also mentioned. The results for the models fitted using the threshold excess and point process approaches are shown in the univariate section. Specifically, the MRL plots, threshold choices and extremal indices for each variable during the different seasons at each station is presented. Since the threshold choice remains the same for the threshold excess and point process approaches, the aforementioned is only shown once for rainfall, temperature and wind speed maxima. Subsequently, the threshold excess parameter estimates, return level estimates and plots as well as quantile-quantile plots are shown for each variable. Following on from there, the point process parameter estimates, return level estimates and plots as well as quantile-quantile plots are presented. The univariate section ends off with a comparison of the approaches.

Furthermore, the multivariate section begins with the results for component-wise maxima approach for maximum wind speed and maximum rainfall as well as for maximum wind speed and maximum temperature using the log dependence model. Following that are the bivariate exploratory plots that have been transformed to standard Fréchet margins and shown on log axes. The plots include the thresholds of the threshold excess and point process approaches, followed by the dependence estimates of the pairwise combinations using the log model. Subsequently, the dependence estimates for the point process approach are shown using the log, negative logistic, bilogistic, negative bilogistic and dirichlet dependence models. Finally, the section ends of with a comparison of the three approaches.

3.2.1 Exploratory Analysis

The exploratory plots of the data are useful for a visual inspection of how the data is distributed. It allows for missing values, maxima and minima among other things to be seen. The exploratory analysis is split according to each variable and then split further into each season per variable. These plots also assist in informally assessing the threshold choice used in the threshold excess approach.

The recorded daily rainfall across the different stations are shown in Figure 3.1.

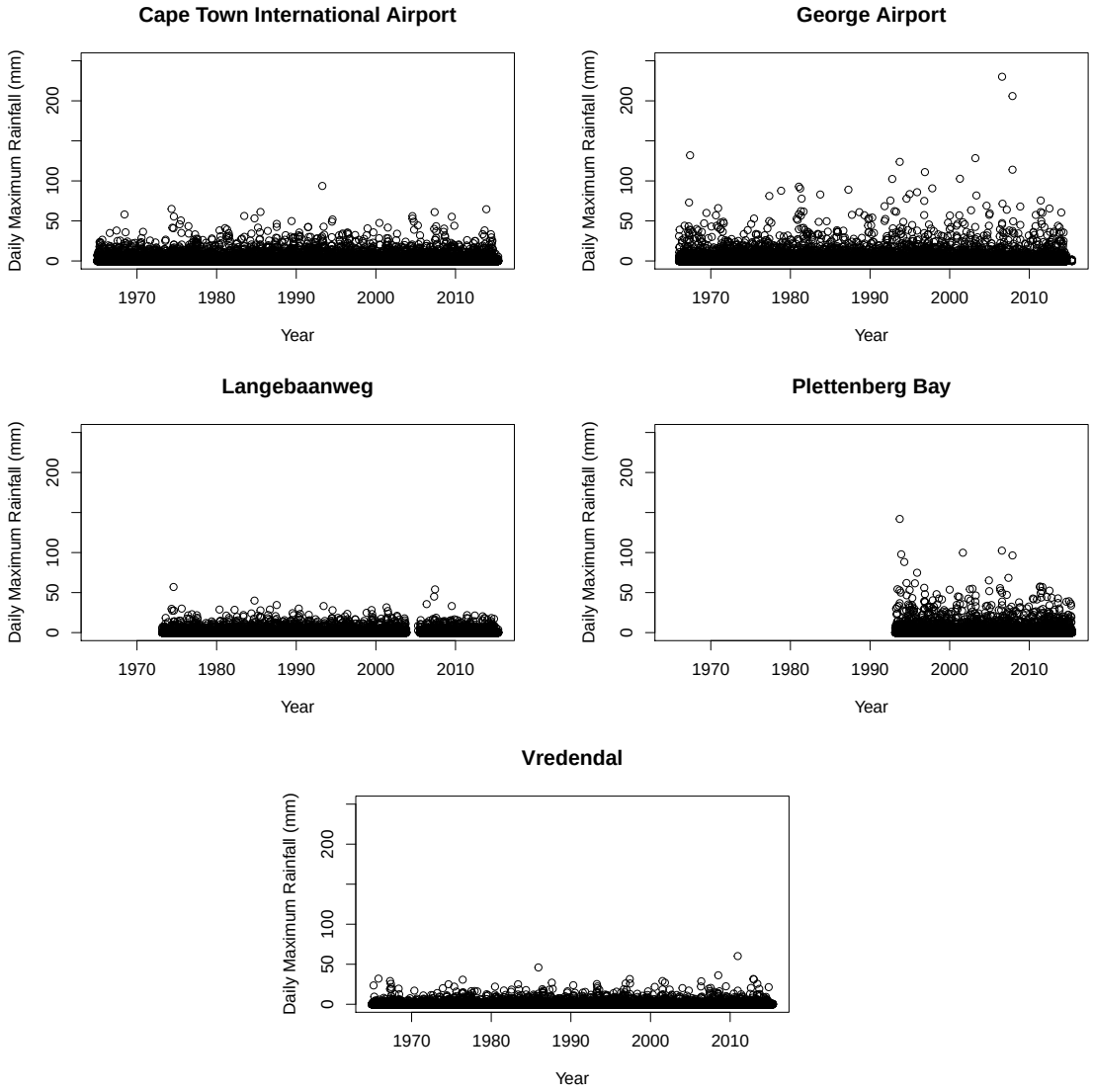


Figure 3.1: Recorded daily rainfall across the different stations (1 January 1965 - 18 May 2015).

From the above exploratory analysis of rainfall, it is evident that over the 50 year period there have been higher rainfall figures recorded at George Airport compared to the other stations. Similarly, for Plettenberg Bay, there is also higher rainfall experienced when comparing the last 20 years of data i.e. 1995 - 2015. These two stations are located in the south-east of the Western Cape province compared to the other three stations which are along the west coast of the province. Overall the rainfall plots seem to suggest that the Western Cape has certain areas that can be classified as drier than others (West Coast) as well as areas that are more lush with greater amounts of greenery compared to the other stations, which are situated in the east of the province. It also indicates that because Vredendal is in the latitude band of the Northern Cape province, that it may experience more of the climatic conditions characteristic of that province. Likewise, since Plettenberg Bay and George Airport are in close proximity to the Eastern Cape, those stations may capture the climate that is more characteristic of that province than it is with the Western Cape.

The recorded daily maximum temperatures across the different stations are shown in Figure 3.2.

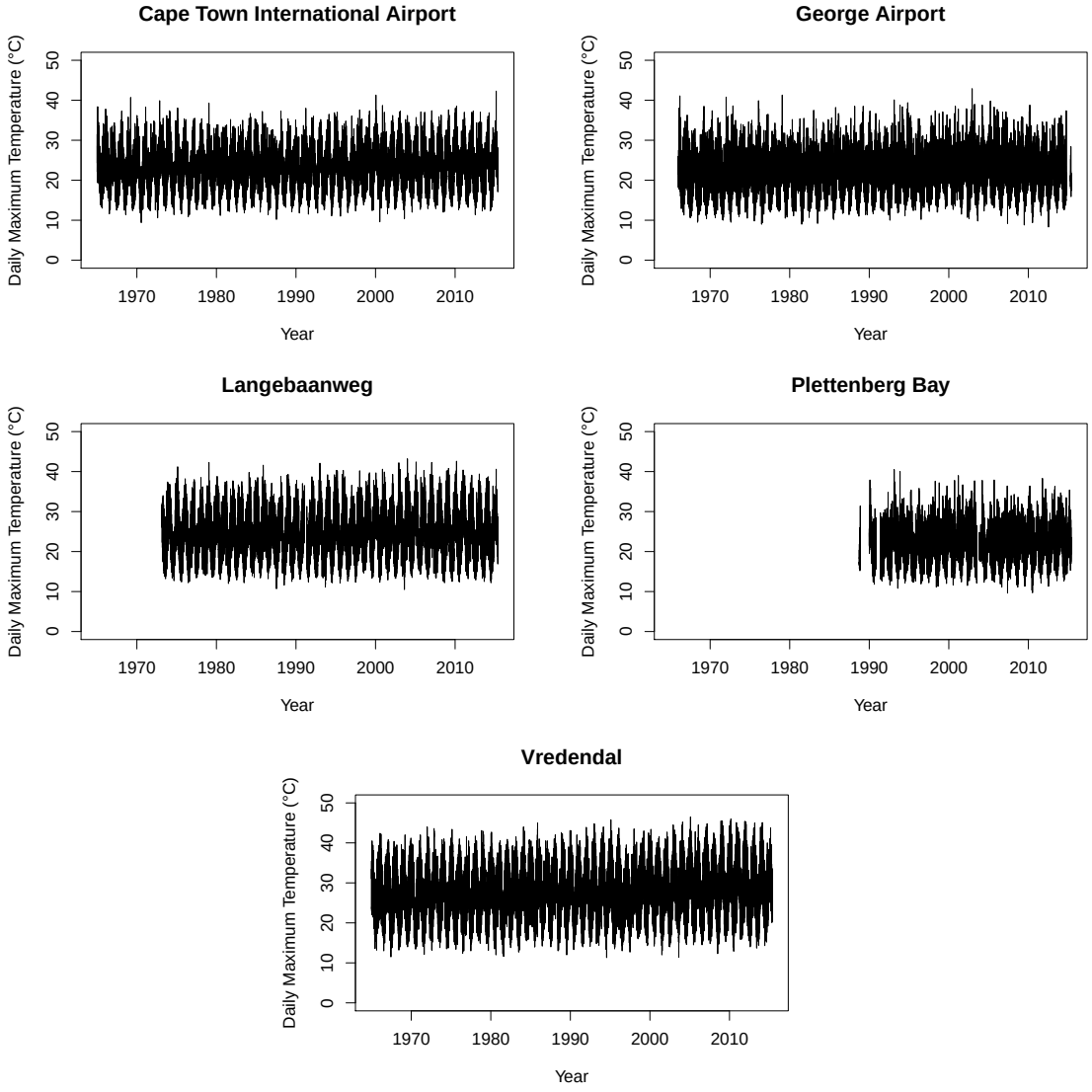


Figure 3.2: Recorded daily maximum temperatures across the different stations (1 January 1965 - 18 May 2015).

In terms of maximum temperature as can be seen in Figure 3.2, Vredendal experienced higher temperatures overall compared to the other four stations. Langebaanweg also recorded higher temperatures on average compared to Cape Town International Airport, George Airport and Plettenberg Bay.

The recorded daily wind speed across only four different stations are shown in Figure 3.3. Langebaanweg is excluded from the wind analysis as there is no wind speed data available for analysis in this study.

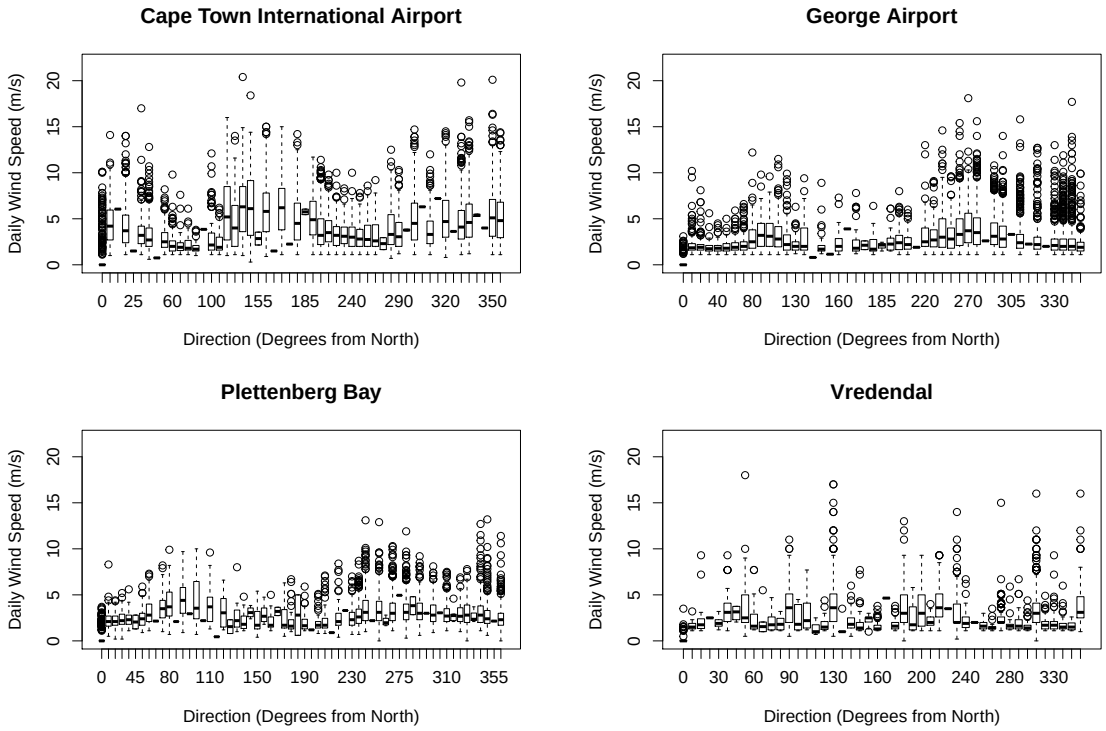


Figure 3.3: Recorded wind speeds versus wind direction across the different stations (1 January 1965 - 18 May 2015).

Figure 3.3 shows the wind speed levels reached above 20 m/s at only Cape Town International Airport in a east-south and north-west direction. Higher wind speeds are recorded for George Airport and Vredendal in comparison to Plettenberg Bay. While the direction is shown on the exploratory plots, this variable is not part of the modelling analysis.

Exploratory Analysis of the data broken up into the four seasons

Maximum Rainfall

Figure 3.4 and Figures A.2 - A.5 (Appendix A) are plots of the daily rainfall for Cape Town International Airport and the four other stations divided into seasons, respectively.

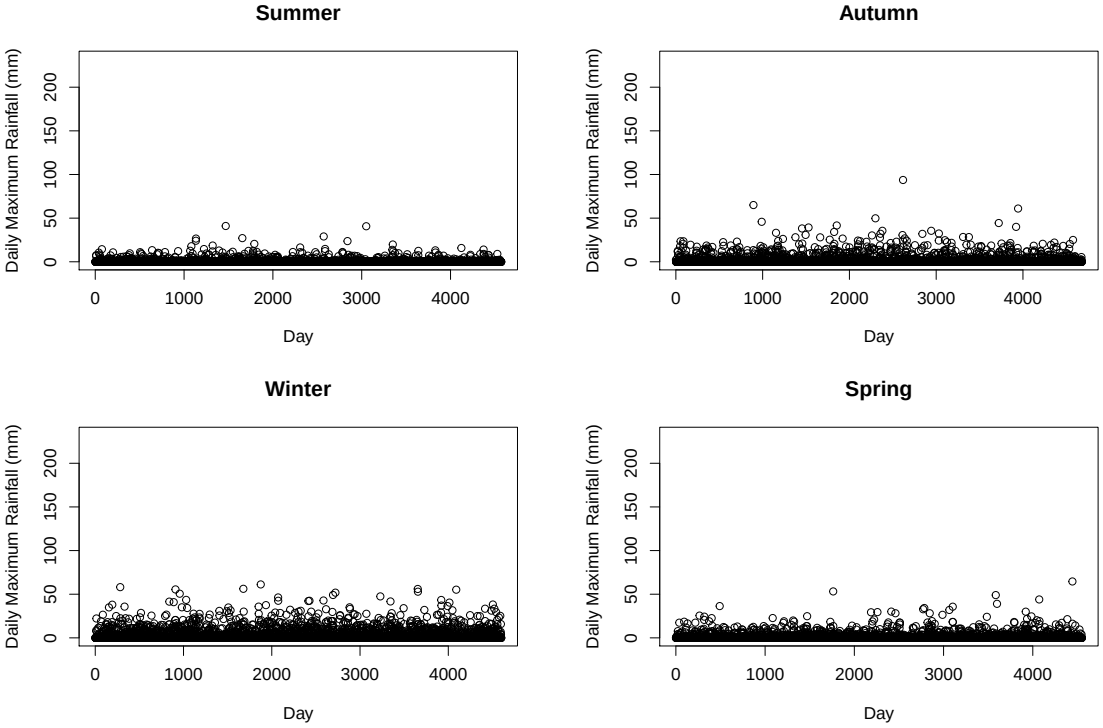


Figure 3.4: Daily rainfall broken up into seasons for Cape Town International Airport

From the exploratory plots shown Figure 3.4, there are higher average rainfall levels recorded during winter and spring at Cape Town International Airport. There are several days with zero rainfall recorded with the average rainfall being below 50 mm for all the seasons. Figure A.2 shows that high levels of rainfall are recorded during the autumn and spring seasons at George Airport with a reading above 200 mm during winter. Summer data at Langebaanweg shows lower levels of rainfall compared to any of the other seasons (Figure A.3). Rainfall levels in all the seasons at Plettenberg Bay shown in Figure A.4 are relatively high with a few observations almost reaching 100 mm and others that are above 100 mm. For Vredendal (Figure A.5) with the exception of summer, the other three seasons display rainfall levels that are below 50 mm. Summer noticeably has two observations that are higher than the rest of the observations with one above 50 mm.

Maximum Temperature

Figure 3.5 and Figures A.6 - A.9 in Appendix A are plots of the daily maximum temperature for Cape Town International Airport and the other four stations broken up into seasons, respectively.

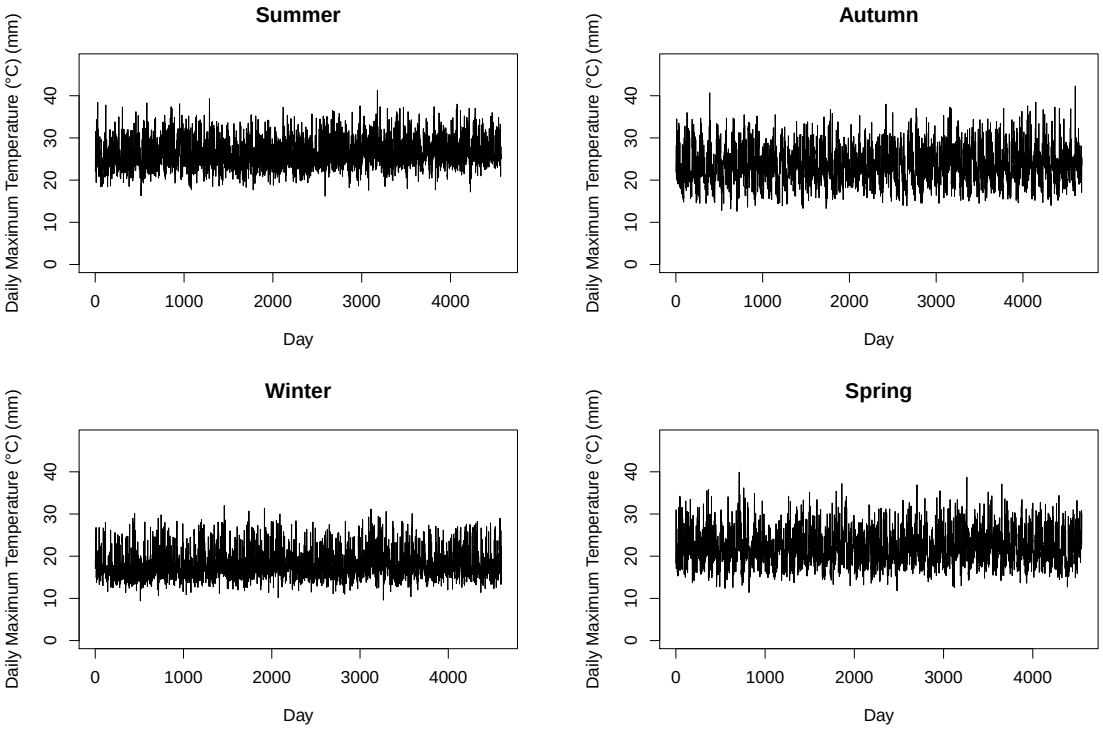


Figure 3.5: Daily maximum temperature broken up into seasons for Cape Town International Airport

The maximum temperatures for the winter season at Cape Town International Airport shown in Figure 3.5 is contained in a lower interval compared to the other three seasons. George Airport is the only weather station that recorded temperatures below 10 (^{0}C). Figure A.6 shows that the recorded maximum temperatures for the winter season also fall within a lower interval compared to the other seasons. For the maximum temperatures recorded at Langebaanweg shown in Figure A.7, the gap in the autumn plot represents some missing data at this station during this season. Winter has a narrower and lower interval for the recorded temperatures compared to the other three seasons. The missing data across the seasons for Plettenberg Bay is clearly seen in the plots in Figure A.8. The interval for summer does not spread as wide as the intervals of autumn, winter and spring. Looking at Figure A.9, summer at Vredendal has a noticeably higher interval than other three seasons. The recorded temperatures for winter is contained in a lower interval compared to the other three stations.

Maximum Wind Speed

Figure 3.6 (shown below) and Figures A.10 - A.12 in Appendix A are plots of the daily maximum wind speed for Cape Town International Airport and other four stations broken up into season, respectively.

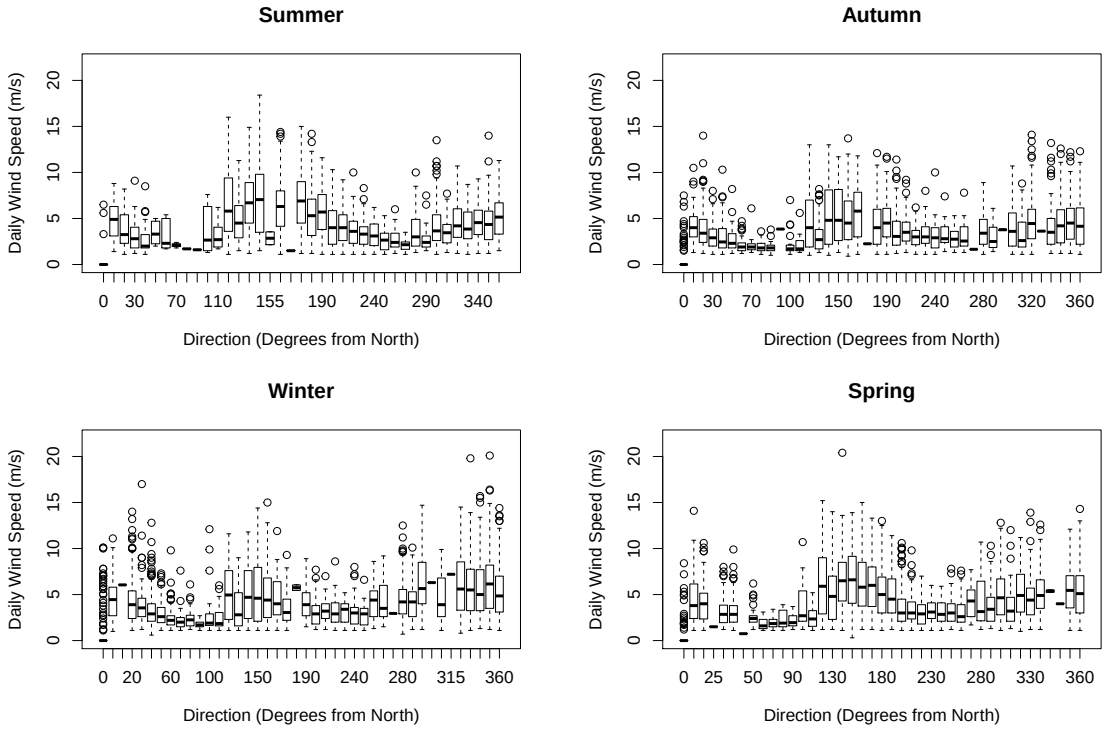


Figure 3.6: Daily wind speed broken up into seasons for Cape Town International Airport

Wind speeds recorded at Cape Town International Airport show that in a few instances speeds reached approximately 20 m/s during the winter and spring seasons. The observations recorded during autumn are consistently below 15 m/s. Wind speed levels during summer at George Airport (Figure A.10) shows the recorded observations to be lower than 15 m/s with a few noticeably high observations recorded during autumn and winter at this station. Wind speed levels were below 15 m/s for all seasons at Plettenberg Bay (Figure A.11). The wind speed plots for Vredendal (Figure A.12) shows that with the exception of the spring season, a few of the recorded observations reach above 15 m/s.

Table 3.3 shows the maximum temperature, maximum rainfall and maximum wind speed recorded at each station.

Table 3.3: Maxima of the weather variables

Station	Maximum Temperature (^{0}C)	Maximum Rainfall (mm)	Maximum Wind Speed (m/s)
CT Airport	41.3	93.7	20.4
George Airport	40.1	230.1	18.1
Langebaanweg	43.2	20.4	-
Plettenberg Bay*	40.5	142.0	13.2
Vredendal	46.5	60.2	18.0

**over 20 years*

Table 3.3 shows that the maximum temperature of 46.5 ^{0}C was recorded at Vredendal followed by Langebaanweg with a high of 43.2 ^{0}C. George Airport, Plettenberg Bay and Cape Town International Airport all recorded maximum temperatures around the 40 ^{0}C mark.

The rainfall recorded at the stations show that George Airport and Plettenberg Bay experienced rainfall of over 100 mm, while there were lower amounts of rainfall recorded at the stations in the north i.e. Langebaanweg and Vredendal. The 'centrally' located station viz Cape Town International Airport, experienced in-between levels of rainfall of all the other stations.

Cape Town International Airport experiences higher wind speeds as measured over the 50 year period with a maximum speed of 20.4 m/s. Following this is George Airport and Vredendal with maximum wind speeds of 18.1 m/s and 18 m/s, respectively.

3.2.2 Stationarity Testing

Stationarity in the data is assessed using the Phillips-Perron test as the data is used as a time-series. The Phillips-Perron tests for unit roots with more relaxed error distribution assumptions than the Dickey-Fuller test (Enders, 2004). Specifically, the hypotheses for the test are as follows

H_0: The series contains a unit root
H_1: The series is stationary

Tables A.1, A.2 and A.3 in Appendix A show that the p-values are very small ($<<$ 0.01) which indicates strong evidence against the null hypothesis. Thus, the series can be accepted as stationary as the hypothesis of the series having a unit root is rejected.

3.3 Univariate Analysis: Threshold Excess

A univariate analysis of the data is useful as the threshold value for each variable can be determined and subsequently used in the multivariate analysis. Another reason for performing a univariate analysis first is that it assists in model validity (Coles and Tawn, 1991). Dependence within each variable dataset can also be checked and declustered if

needed.

The threshold excess section is structured according to maximum rainfall, maximum temperature and maximum wind speed. Each variable follows the same procedure which includes threshold (u) choice through MRL plots, declustering, parameter estimates, return level estimates and plots followed by quantile-quantile plots.

3.3.1 Maximum Rainfall

This subsection contains detailed explanations of the techniques used as it is the first time that these terms are being used in this study. The same techniques are used for the subsequent variables. The following sections show and explain MRL plots, declustering of the data at the chosen thresholds, parameter estimates, return level estimates with plots and quantile-quantile plots for rainfall.

Mean Residual Life Plots

A method for finding a threshold value for this approach is to use mean residual life plots. To choose a suitable u value, it is of interest to see where the plots deviate from a linear pattern (Coles, 2001). The dashed line indicates the 95% confidence interval.

Figure 3.7 shows the mean residual life plots for each season for Cape Town International Airport. The MRL plots for the other four stations are shown in Appendix A (Figures A.13 - A.16) as the same process is followed to find the threshold value.

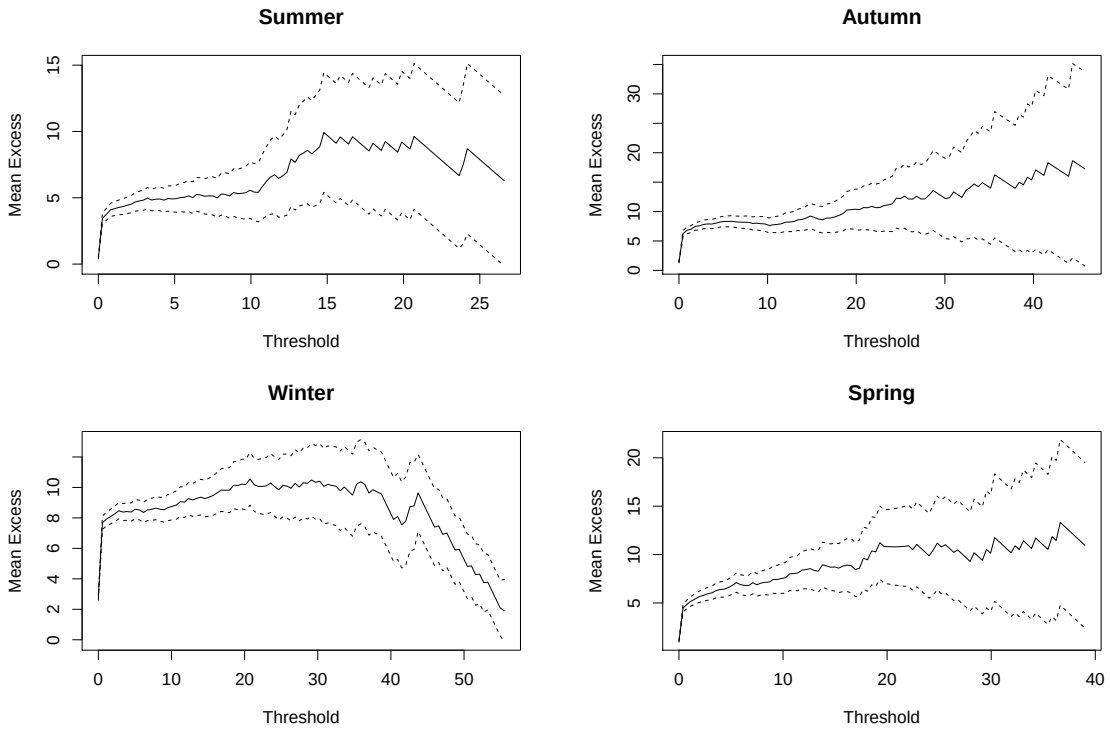


Figure 3.7: Mean residual life plots for rainfall at Cape Town International Airport

In order to look for a suitable u for each variable, the exploratory plots in Figure 3.7 for Cape Town International Airport and Figures A.13 - A.16 for the four other stations are examined. From the MRL plots and exploratory plots, the thresholds can be found so that inferences can be made about the data. Since there is some subjectivity in choosing the threshold just by looking at plots, it is important to keep in mind that the threshold should not be a too high or a too low value (Beirlant *et al.*, 2004).

Starting with Cape Town International Airport, the summer plot is linear from $u = 0$ mm up to $u \approx 10$ mm. Only 45 observations are greater than 10 mm, thus $u = 5.20$ mm will be used as the confidence intervals are still fairly narrow at these values. For the remaining three seasons, an appropriate threshold for autumn is chosen at 10.60 mm, 22.32 mm for winter and 9.65 mm for spring.

Following the same process for the George Airport, an appropriate threshold for summer is 17.43 mm with threshold values of 19.69 mm, 12.78 mm and 17.40 mm for autumn, winter and spring respectively. Summer, autumn, winter and spring for Langebaanweg have thresholds of 3.60 mm, 11.27 mm, 15.10 mm and 7.78 mm respectively. Plettenberg Bay has a threshold for summer of 11.80 mm and for autumn a value of 20.20 mm and 22.29 mm for winter with 23.20 mm for spring. Lastly, Vredendal has relatively low threshold values across all seasons with 1.60 mm for summer, 3 mm for autumn, 9.30 mm for Winter and 2.54 mm for spring. These threshold values and corresponding quantiles are summarised in Table 3.4 below. The respective quantiles are useful when fitting the multivariate models.

Table 3.4: Threshold values for rainfall

	Season	Threshold	Quantile
CT Airport	Summer	5.20	97.2
	Autumn	10.60	95.4
	Winter	22.32	97.7
	Spring	9.65	97.0
George Airport	Summer	17.43	97.4
	Autumn	19.69	97.7
	Winter	12.78	96.7
	Spring	17.40	97.1
Langebaanweg	Summer	3.60	98.2
	Autumn	11.27	98.5
	Winter	15.10	98.3
	Spring	7.78	98.4
Plettenberg Bay	Summer	11.80	96.5
	Autumn	20.20	97.5
	Winter	22.29	97.5
	Spring	23.20	98.0
Vredendal	Summer	1.60	97.8
	Autumn	3.00	96.2
	Winter	2.70	91.2
	Spring	2.54	96.8

The aforementioned thresholds are used when fitting these models using the threshold excess and point process approaches.

Declustering

It is expected that if heavy rainfall is experienced, it does not occur just on one day but rather over a period of days. This makes a few of the observations dependent and the temporal dependence has to be examined. A way to deal with the dependence is to decluster the data at a suitable u with different values of r where r is the run length (Coles, 2001). For example, the values of r could be 1 and 3 as heavy rainfall experienced on a given day could affect rainfall into the future, up to 3 days ahead. A longer period than 3 days may not be a viable length of time as it is too long a period to make inferences based on rainfall experienced at present.

To determine which stations require declustering, a sensitivity analysis is performed. Specifically, the extremal index ($\hat{\theta}$) for each station is calculated for the selected thresholds and varying values of r. When $\hat{\theta}$=1, it means that there is no dependence in the excesses so therefore choosing values $\hat{\theta}$ that fall far below 0.8 (Khuluse, 2010) are required to be declustered. The reason for this is that extremal index estimates of 0.8 and greater show weak dependence among the excesses, hence they do not need to be declustered. The results are summarised in Table 3.5 below for rainfall with the estimates for temperature and wind speed shown in Tables A.4 and A.5.

Table 3.5: Extremal index estimates for maximum rainfall

	Season	Threshold	r=1	r=2	r=3
CT Airport	Summer	5.20	0.995	1.000	1.000
	Autumn	10.60	1.000	1.000	1.000
	Winter	22.32	1.000	1.000	1.000
	Spring	9.65	0.990	1.000	1.000
George Airport	Summer	17.43	0.954	1.000	1.000
	Autumn	19.69	0.880	0.890	0.912
	Winter	12.78	1.000	1.000	1.000
	Spring	17.40	0.969	1.000	1.000
Langebaanweg	Summer	3.60	1.000	1.000	1.000
	Autumn	11.27	1.000	1.000	1.000
	Winter	15.10	1.000	1.000	1.000
	Spring	7.78	1.000	1.000	1.000
Plettenberg Bay	Summer	11.80	0.954	0.996	1.000
	Autumn	20.20	1.000	1.000	1.000
	Winter	22.29	1.000	1.000	1.000
	Spring	23.20	1.000	1.000	1.000
Vredendal	Summer	1.60	1.000	1.000	1.000
	Autumn	3.00	1.000	1.000	1.000
	Winter	2.70	1.000	1.000	1.000
	Spring	2.54	0.941	0.991	1.000

The results from Tables 3.5 for rainfall as well as in Tables A.4 and A.5 in Appendix A for temperature and wind speed, respectively suggest that declustering any of the stations is not necessary. The extremal indices for all seasons across the stations are close to 1 (greater than 0.8) and thus satisfy the assumption of independence.

The next step is to fit the data to the Generalised Pareto Distribution (GPD) for a suitable value of u and r. Since there is trade-off between independent observations and exceedances that are not too high or too low, the number of clusters n_c and number of exceedances n_u are taken into consideration. Table 3.6 shows the estimates of the fitted models according to the different thresholds with the standard errors of the estimates shown in parentheses.

Parameter Estimation

The estimates for σ and ξ of rainfall at each station are summarised in Table 3.6 together with the respective 95% confidence intervals - shown in brackets - and the proportion of exceedance ζ_u. Standard errors are shown in parentheses.

Table 3.6: Threshold excess parameter estimates for rainfall (mm).

		u	ζ_u	σ	ξ
CT Airport	Summer	5.20	128/4571 = 0.028	4.21 (0.55)	0.15 (0.10)
	95% CI			[3.13; 5.28]	[-0.03; 0.35]
	Autumn	10.60	214/4679 = 0.046	6.25 (0.65)	0.20 (0.08)
	95% CI			[4.96; 7.53]	[0.04; 0.36]
	Winter	22.32	106/4600 = 0.023	11.16 (1.72)	-0.10 (0.12)
	95% CI			[7.79; 14.52]	[-0.34; 0.13]
	Spring	9.65	137/4550 = 0.030	6.03 (0.83)	0.20 (0.11)
	95% CI			[4.40; 7.66]	[-0.01; 0.42]
George Airport	Summer	17.43	115/4391 = 0.026	12.47 (1.79)	0.05 (0.11)
	95% CI			[8.96; 15.97]	[-0.16; 0.27]
	Autumn	19.69	105/4562 = 0.023	11.35 (1.95)	0.35 (0.15)
	95% CI			[7.52; 15.18]	[0.06; 0.63]
	Winter	12.78	149/4508 = 0.033	12.25 (1.69)	0.25 (0.11)
	95% CI			[8.94; 15.57]	[0.03; 0.47]
	Spring	17.40	126/4369 = 0.029	15.69 (2.10)	0.21 (0.10)
	95% CI			[11.57; 19.81]	[0.01; 0.41]
Langebaan-weg	Summer	3.60	64/3609 = 0.018	4.45 (0.93)	-0.04 (0.17)
	95% CI			[2.62; 6.28]	[-0.37; 0.29]
	Autumn	11.27	57/3757 = 0.015	6.27 (1.25)	0.04 (0.15)
	95% CI			[3.82; 8.73]	[-0.26; 0.33]
	Winter	15.10	64/3735 = 0.017	4.70 (0.93)	0.24 (0.15)
	95% CI			[2.89; 6.52]	[-0.07; 0.54]
	Spring	7.78	60/3701 = 0.016	5.22 (0.91)	0.03 (0.12)
	95% CI			[3.43; 7.00]	[-0.21; 0.26]
Plettenberg Bay	Summer	11.80	69/1985 = 0.035	8.07 (1.63)	0.30 (0.17)
	95% CI			[4.87; 11.26]	[-0.02; 0.63]
	Autumn	20.20	50/2100 = 0.024	12.42 (2.68)	0.06 (0.17)
	95% CI			[7.17; 17.67]	[-0.27; 0.38]
	Winter	22.29	51/2024 = 0.025	14.19 (3.08)	0.09 (0.16)
	95% CI			[8.16; 20.22]	[-0.23; 0.42]
	Spring	23.20	39/1996 = 0.020	13.76 (3.10)	0.18 (0.16)
	95% CI			[7.68; 19.85]	[-0.14; 0.50]
Vredendal	Summer	1.60	98/4571 = 0.018	2.10 (0.35)	0.56 (0.15)
	95% CI			[1.40; 2.79]	[0.27; 0.85]
	Autumn	3.00	174/4679 = 0.037	5.25 (0.64)	0.07 (0.10)
	95% CI			[3.99; 6.51]	[-0.12; 0.26]
	Winter	9.30	96/4600 = 0.021	4.19 (0.32)	0.04 (0.12)
	95% CI			[3.56; 4.81]	[-0.20; 0.29]
	Spring	2.54	146/4550 = 0.032	3.26 (0.42)	0.19 (0.10)
	95% CI			[2.43; 4.09]	[-0.01; 0.39]

At all the stations there are positive estimates for the shape parameter ξ across all the seasons except for the winter season at Cape Town International Airport and the summer season at Langebaanweg station. The positive estimates corresponds to the Pareto distribution which is characterised by the polynomial tail behaviour. Hence the rainfall extremes with these positive shape estimates can be described as following a Pareto distribution. The exceptions with the negative shape parameter estimate can better be described by the Weibull distribution instead as it is characterised as being a bounded distribution.

Furthermore, the proportions of exceedances are also not too high or too low given the size of the data at each station. It can be seen $\hat{\zeta_u}$ values for all the stations are not high at all. This means that the probability of exceedance of a specific level of rainfall is not high.

Return Levels

The estimates of the various return levels are summarised in Table 3.7 where the confidence intervals are shown in brackets.

Table 3.7: Threshold excess return level estimates for rainfall (mm)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	5.20	21.42	28.05	33.75	40.09
	95% CI		[17.39; 25.45]	[20.76; 35.45]	[22.76; 44.73]	[24.14; 56.03]
	Autumn	10.60	42.27	54.87	66.04	78.85
	95% CI		[34.72; 49.83]	[41.05; 68.68]	[45.23; 86.84]	[48.60; 109.10]
	Winter	22.32	49.58	56.83	61.87	66.55
	95% CI		[43.87; 55.29]	[48.17; 65.49]	[50.04; 73.69]	[50.92; 82.18]
	Spring	9.65	35.48	46.77	56.79	68.32
	95% CI		[28.49; 42.47]	[33.28; 60.26]	[35.75; 77.84]	[36.84; 99.81]
George Airport	Summer	17.43	57.27	70.97	81.79	93.01
	95% CI		[48.07; 66.48]	[55.69; 86.81]	[58.69; 104.89]	[60.63; 125.40]
	Autumn	19.69	74.41	107.21	139.98	181.69
	95% CI		[55.25; 93.57]	[64.19; 150.23]	[65.53; 214.43]	[58.91; 304.48]
	Winter	12.78	72.47	100.49	126.40	157.22
	95% CI		[55.07; 89.87]	[65.58; 135.41]	[70.60; 182.21]	[71.77; 242.68]
	Spring	17.40	84.07	113.84	140.43	171.13
	95% CI		[65.94; 102.19]	[79.60; 148.07]	[87.47; 193.39]	[92.23; 250.03]
Langebaan-weg	Summer	3.60	14.38	17.95	20.56	23.09
	95% CI		[11.46; 17.29]	[13.16; 22.74]	[13.59; 27.52]	[13.36; 32.81]
	Autumn	11.27	27.04	33.342	38.39	43.49
	95% CI		[22.40; 31.68]	[25.71; 41.14]	[27.02; 49.76]	[27.25; 59.73]
	Winter	15.10	31.35	40.07	48.05	57.43
	95% CI		[25.71; 36.99]	[28.67; 51.48]	[29.39; 66.70]	[28.23; 86.64]
	Spring	7.78	21.08	26.22	30.19	34.23
	95% CI		[17.32; 24.84]	[20.35; 32.10]	[21.94; 38.44]	[22.88; 45.58]
Plettenberg Bay	Summer	11.80	56.15	78.89	100.82	127.89
	95% CI		[35.57; 76.73]	[36.48; 121.31]	[31.57; 170.08]	[19.36; 236.42]
	Autumn	20.20	59.77	72.46	82.06	93.07
	95% CI		[46.96; 72.57]	[51.84; 93.07]	[51.96; 112.16]	[50.48; 132.85]
	Winter	22.29	69.75	87.54	102.04	117.50
	95% CI		[52.59; 86.91]	[56.97; 118.11]	[56.48; 147.60]	[52.30; 182.71]
	Spring	23.20	70.20	92.40	111.82	133.84
	95% CI		[50.37; 90.02]	[56.07; 128.73]	[56.13; 167.52]	[51.28; 216.40]
Vredendal	Summer	1.60	15.32	27.03	40.87	61.27
	95% CI		[9.39; 21.26]	[11.77; 42.29]	[11.49; 70.25]	[7.05; 115.48]
	Autumn	3.00	22.56	28.86	33.91	39.22
	95% CI		[18.31; 26.80]	[21.41; 36.31]	[23.08; 44.74]	[24.10; 54.34]
	Winter	9.30	23.79	29.58	34.29	39.30
	95% CI		[20.01; 27.58]	[23.42; 35.74]	[25.76; 42.82]	[27.85; 50.75]
	Spring	2.54	16.57	22.46	27.64	33.55
	95% CI		[12.92; 20.21]	[15.60; 29.32]	[17.09; 38.19]	[17.97; 49.12]

At Vredendal in autumn for instance, the 2-year return level estimate shows that the daily maximum rainfall is expected to exceed approximately 23 mm on average once every 2 years. Also, the daily rainfall is expected to exceed (on average) approximately 29 mm once every 5 years, 34 mm once every 10 years and 39 mm once every 50 years.

From Table 3.7, it can be seen that there is a noticeable difference in the return level estimates for Plettenberg Bay and George Airport. The estimates are much higher compared to the other stations. This is expected since the exploratory plots in Figure 3.1 showed that the highest levels of rain were recorded at these two stations. The confidence intervals for these two stations are also wider than the other three stations which

can be attributed to the high levels of rain recorded. The overall trend in estimates as the return level increases is that the confidence intervals get wider which suggests the uncertainty of prediction for long periods of time.

To be able to see the prediction capabilities of the fitted models, the return level plots are shown in Figure 3.8 for Cape Town International Airport and Figures A.24 - A.27 (Appendix A) for the other four stations.

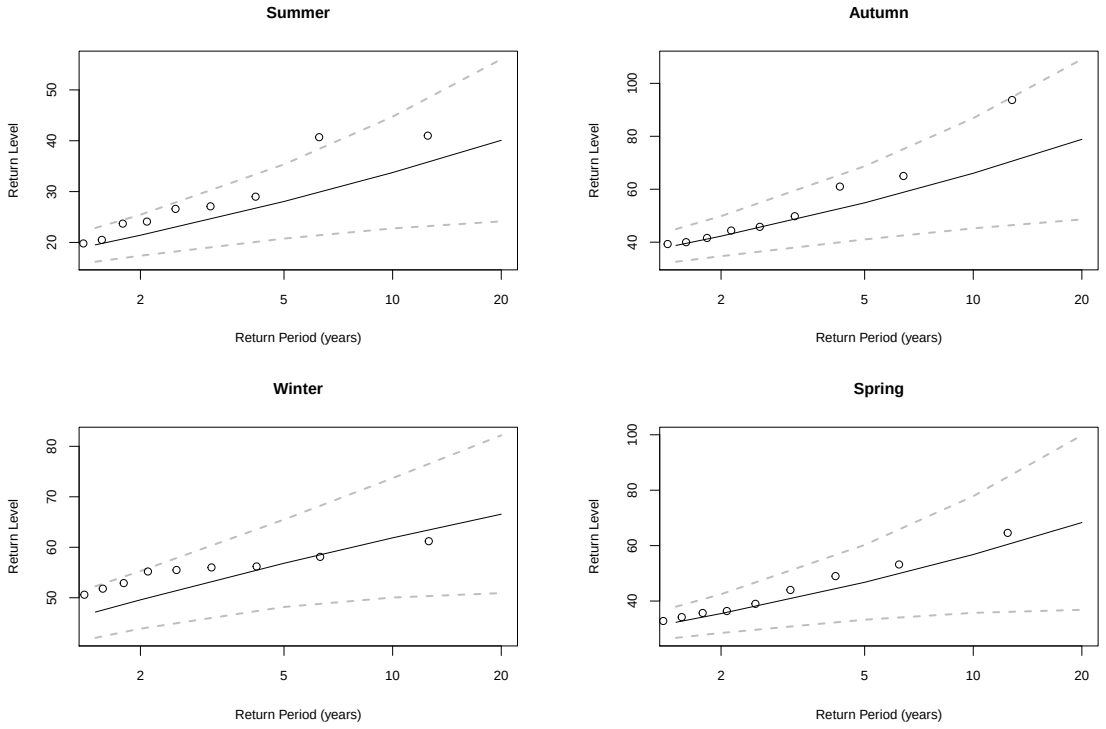


Figure 3.8: Return level plots for rainfall at Cape Town International Airport

The plots shown in Figures 3.8 and A.25 show that the shapes of the plots for winter at Cape Town international Airport and summer at Langebaanwag season are convex and different from the other three respective seasons. This convex shape mirrors the negative estimates of the shape parameter whereas the positive shape estimates for remaining seasons of the each station are concave. George Airport, Plettenberg Bay and Vredendal have return level plots that are concave as these extremes follow a Pareto distribution.

In terms of predictive ability, estimates for the 20-year return period are subject to large variability and inaccuracy as the confidence intervals are wide at the 20-year and above return period. Moreover, the estimates of the return levels across all stations should go up to at most 5 years ahead to have accurate estimates. The 5-year return period is sufficient with the exception of Plettenberg Bay across all seasons. The reason for this is that at a return period of 5 years, the confidence interval is wide (much wider than the other stations at this return period). This suggests that increased accuracy in estimates can be obtained by having a shorter return period of about 3 years for this station.

Overall, the predictive power of the fitted model is not powerful for long return periods which brings in the uncertainty of attempting to predict future events.

Diagnostic Plots

The quantile-quantile plots show a plot of the ranked observations x_k against $F^{-1}\left(\frac{k-\frac{1}{4}}{n+\frac{1}{2}}\right)$ where the inverse distribution function is defined as $F^{-1}(u)$. The interest is to see if the plots produce a straight line through the origin for a good model fit.

The quantile-quantile plots for Cape Town International Airport are shown in Figure 3.9 and in Figures A.28 - A.31 for the other four stations, respectively.

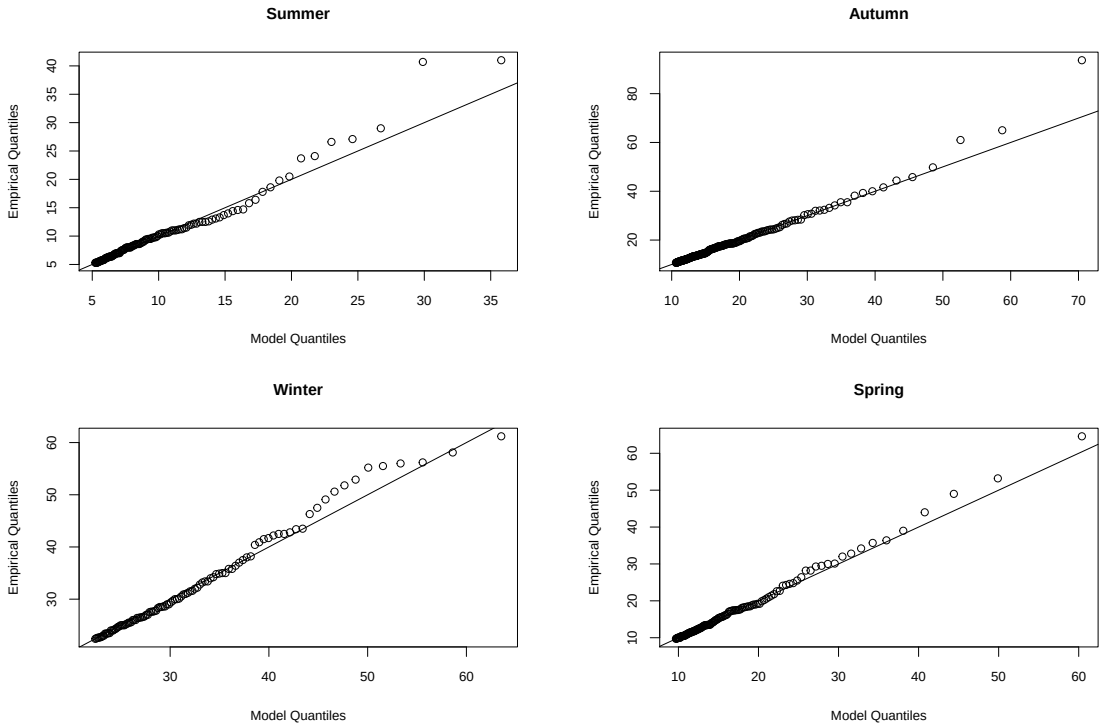


Figure 3.9: Quantile-quantile plots for rainfall at Cape Town International Airport

From the quantile-quantile plots in Figure 3.9 and Figures A.28 - A.31, there are slight deviations from linearity for George Airport and Plettenberg Bay with a slighter larger deviation for the Summer seasons at Vredendal and Langebaanweg and does not impact on the validity of the fitted models.

3.3.2 Maximum Temperature

This subsection contains explanations of the techniques being used as it is the first time that these terms are used in this study. The same techniques are used for the subsequent variables. The following sections show and explain MRL plots, declustering of the data at the chosen thresholds, parameter estimates, return level estimates with plots and quantile-quantile plots for temperature.

Mean Residual Life Plots

The MRL plots for maximum temperatures at Cape Town International Airport are shown in Figure 3.10. Figures A.17 - A.20 (Appendix A) show the MRL plots of temperature for the other four stations.

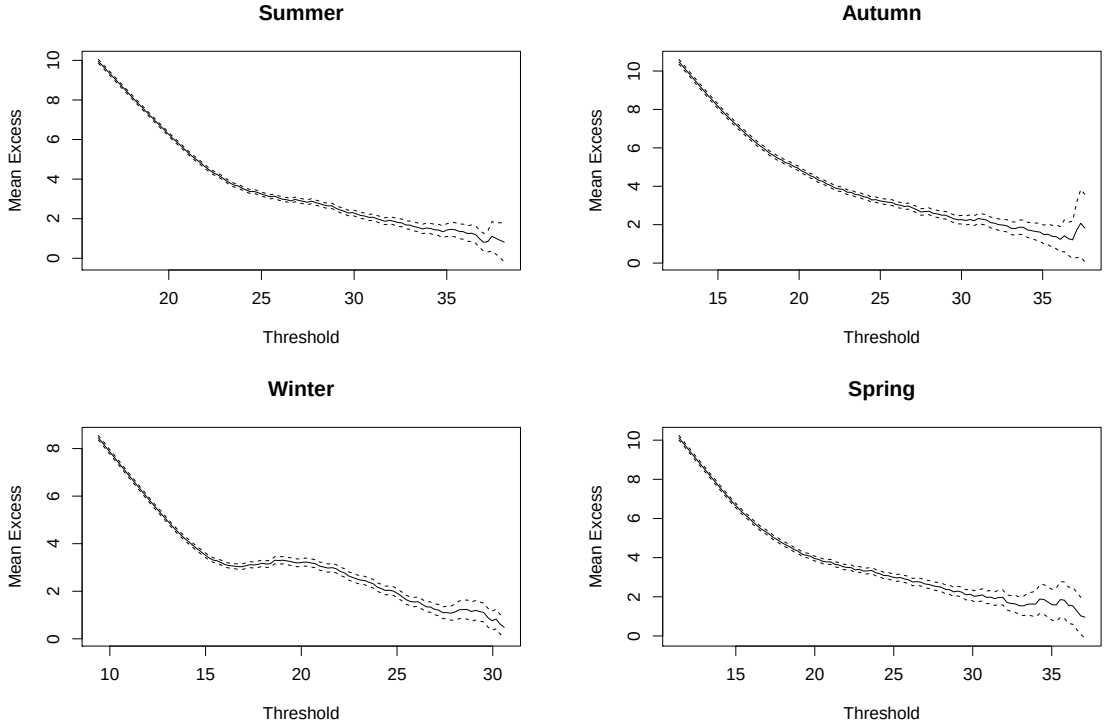


Figure 3.10: Mean residual life plots for temperature at Cape Town International Airport

Table 3.8 summarises the chosen thresholds for temperature with the corresponding quantiles.

Table 3.8: Threshold values for temperature

	Season	Threshold	Quantile (%)
CT Airport	Summer	32.00	93.6
	Autumn	30.00	93.0
	Winter	25.00	95.2
	Spring	26.90	89.4
George Airport	Summer	32.00	97.3
	Autumn	33.00	97.7
	Winter	27.00	94.3
	Spring	30.99	97.3
Langebaanweg	Summer	34.00	90.8
	Autumn	30.00	84.0
	Winter	25.00	91.7
	Spring	29.00	88.6
Plettenberg Bay	Summer	28.00	94.7
	Autumn	30.00	95.0
	Winter	25.00	90.0
	Spring	26.68	95.6
Vredendal	Summer	39.10	95.4
	Autumn	37.00	94.1
	Winter	30.90	96.4
	Spring	34.00	91.9

It can be seen that the MRL plots in Figures 3.10 and A.17 - A.20 for temperature are different to the plots seen in the rainfall analysis. There are narrower confidence intervals and are smoother than the rainfall MRL plots.

Extremal Indices

The extremal indices calculated in Table A.4 (Appendix A) show that the temperature data need not be declustered at the chosen thresholds. Table 3.9 shows the parameter estimates for maximum temperature.

Parameter Estimation

The estimates for σ and ξ of temperature at each station are summarised in Table 3.9 together with the respective 95% confidence intervals - shown in brackets - and the proportion of exceedance ζ_u. Standard errors are shown in parentheses.

Table 3.9: Threshold excess parameter estimates for temperature (^{0}C)

		u	ζ_u	σ	ξ
CT Airport	Summer	32.00	283/4571 = 0.062	2.28 (0.16)	-0.19 (0.04)
	95% CI			[1.96; 2.59]	[-0.27; -0.11]
	Autumn	30.00	322/4679 = 0.069	2.57 (0.17)	-0.13 (0.05)
	95% CI			[2.20; 2.94]	[-0.22; -0.03]
	Winter	25.00	214/4600 = 0.047	2.53 (0.20)	-0.33 (0.05)
	95% CI			[2.14; 2.92]	[-0.41; -0.24]
	Spring	26.90	478/4550 = 0.011	3.27 (0.18)	-0.20 (0.03)
	95% CI			[2.92; 3.62]	[-0.27; -0.14]
George Airport	Summer	32.00	116/4391 = 0.026	2.25 (0.30)	-0.03 (0.10)
	95% CI			[1.66; 2.84]	[-0.22; 0.16]
	Autumn	33.00	104/4540 = 0.023	2.60 (0.33)	-0.24 (0.09)
	95% CI			[1.95; 3.26]	[-0.41; -0.08]
	Winter	27.00	255/4508 = 0.057	1.92 (0.14)	-0.23 (0.04)
	95% CI			[1.64; 2.21]	[-0.32; -0.15]
	Spring	30.99	120/4411 = 0.0.027	3.29 (0.46)	-0.29 (0.11)
	95% CI			[2.38; 4.20]	[-0.51; -0.08]
Langebaan-weg	Summer	34.00	343/3762 = 0.091	3.13 (0.21)	-0.30 (0.04)
	95% CI			[2.73; 3.54]	[-0.38; -0.21]
	Autumn	30.00	609/3882 = 0.157	3.79 (0.17)	-0.28 (0.02)
	95% CI			[3.45; 4.13]	[-0.32; -0.23]
	Winter	25.00	319/3864 = 0.083	2.21 (0.15)	-0.25 (0.04)
	95% CI			[1.92; 2.50]	[-0.33; -0.17]
	Spring	29.00	472/4411 = 0.123	3.83 (0.22)	-0.25 (0.04)
	95% CI			[3.41; 4.26]	[-0.32; -0.18]
Plettenberg Bay	Summer	28.00	116/2227 = 0.052	2.73 (0.37)	-0.08 (0.10)
	95% CI			[2.01; 3.45]	[-0.27; 0.11]
	Autumn	30.00	113/2227 = 0.049	3.02 (0.41)	-0.23 (0.10)
	95% CI			[2.22; 3.82]	[-0.43, -0.04]
	Winter	25.00	222/2246 = 0.099	3.29 (0.24)	-0.35 (0.03)
	95% CI			[2.83; 3.75]	[-0.41; -0.28]
	Spring	26.68	100/2253 = 0.044	3.74 (0.53)	-0.14 (0.10)
	95% CI			[2.69; 4.78]	[-0.34; 0.06]
Vredendal	Summer	39.10	197/4571 = 0.043	2.89 (0.24)	-0.35 (0.05)
	95% CI			[2.42; 3.36]	[-0.45; -0.24]
	Autumn	37.00	267/4679 = 0.057	2.57 (0.19)	-0.24 (0.04)
	95% CI			[2.21; 2.94]	[-0.32; -0.15]
	Winter	30.90	163/4600 = 0.035	1.83 (0.19)	-0.25 (0.07)
	95% CI			[1.45; 2.20]	[-0.40; -0.11]
	Spring	34.00	485/4549 = 0.107	3.28 (0.19)	-0.23 (0.04)
	95% CI			[2.92; 3.65]	[-0.30; -0.16]

All stations for all seasons have estimates for the shape parameter that are negative which corresponds to the bounded Weibull distribution. Also the proportions of exceedances are slightly higher than that of rain which may translate to higher temperatures.

Return Levels

The estimates of the various return levels are shown in Table 3.10 with the confidence intervals shown in brackets.

Table 3.10: Threshold excess return level estimates for temperature (^{0}C)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	32.00	38.14	39.05	39.64	40.16
	95% CI		[37.35; 38.93]	[38.07; 40.03]	[38.50; 40.79]	[38.83; 41.49]
	Autumn	30.00	37.93	39.28	40.21	41.05
	95% CI		[36.92; 38.93]	[37.93; 40.64]	[38.53; 41.88]	[39.02; 43.09]
	Winter	25.00	30.311	30.95	31.32	31.61
	95% CI		[28.19; 31.44]	[29.46; 32.44]	[29.48; 33.16]	[29.32; 33.91]
	Spring	26.90	36.33	37.45	38.16	38.78
	95% CI		[35.36; 37.29]	[36.25; 38.64]	[36.77; 39.55]	[37.17; 40.38]
George Airport	Summer	32.00	38.38	40.25	41.63	42.99
	95% CI		[37.06; 39.70]	[38.20; 42.30]	[38.81; 44.46]	[39.21; 46.76]
	Autumn	33.00	38.30	39.37	40.04	40.60
	95% CI		[37.07; 39.53]	[37.81; 40.93]	[38.15; 41.93]	[38.33; 42.87]
	Winter	27.00	31.90	32.50	32.85	33.14
	95% CI		[30.99; 32.81]	[31.33; 33.66]	[31.43; 34.27]	[31.41; 34.86]
	Spring	30.99	37.54	38.63	39.29	39.82
	95% CI		[35.91; 39.16]	[36.45; 40.82]	[36.55; 42.02]	[36.45; 43.19]
Langebaan-weg	Summer	34.00	41.54	42.26	42.70	43.05
	95% CI		[40.31; 42.76]	[40.68; 43.85]	[40.77; 44.63]	[40.71; 45.39]
	Autumn	30.00	39.97	40.79	41.28	41.69
	95% CI		[38.82; 41.13]	[39.33; 42.25]	[39.53; 43.03]	[39.60; 43.79]
	Winter	25.00	30.67	31.32	31.72	32.06
	95% CI		[29.88; 31.46]	[30.33; 32.31]	[30.55; 32.89]	[30.69; 33.43]
	Spring	29.00	39.39	40.42	41.06	41.60
	95% CI		[38.14; 40.64]	[38.85; 42.00]	[39.20; 42.92]	[39.42; 43.77]
Plettenberg Bay	Summer	28.00	36.61	38.42	39.69	40.90
	95% CI		[34.69; 38.54]	[35.59; 41.26]	[35.97; 43.42]	[36.18; 45.63]
	Autumn	30.00	37.34	38.42	39.10	39.67
	95% CI		[35.64; 39.03]	[36.17; 40.67]	[36.35; 41.85]	[36.38; 42.96]
	Winter	25.00	32.35	32.94	33.28	33.54
	95% CI		[30.49; 34.21]	[30.42; 35.46]	[30.10; 36.46]	[29.52; 37.57]
	Spring	26.68	37.00	38.98	40.33	41.55
	95% CI		[34.68; 39.31]	[35.73; 42.24]	[36.18; 44.47]	[36.41; 46.69]
Vredendal	Summer	39.10	45.01	45.66	46.04	46.32
	95% CI		[43.65; 46.38]	[43.81; 47.52]	[43.69; 48.38]	[43.37; 49.29]
	Autumn	37.00	43.38	44.26	44.81	45.28
	95% CI		[42.47; 44.29]	[43.13; 45.39]	[43.48; 46.14]	[43.72; 46.84]
	Winter	30.90	34.94	35.59	35.99	36.32
	95% CI		[34.16; 35.72]	[34.59; 36.60]	[34.78; 37.21]	[34.87; 37.79]
	Spring	34.00	42.99	43.97	44.59	45.11
	95% CI		[41.97; 44.01]	[42.68; 45.26]	[43.07; 46.11]	[43.34; 46.89]

For all seasons, Langebaanweg and Vredendal have higher return levels for the different return periods compared to the rest of the stations. Langebaanweg and Vredendal are situated closer to each other (approximately 197 km) than any other station in this study and experience drier conditions as seen in the exploratory plots.

Figures 3.11 and A.32 - A.35 show the return level plots for temperature.

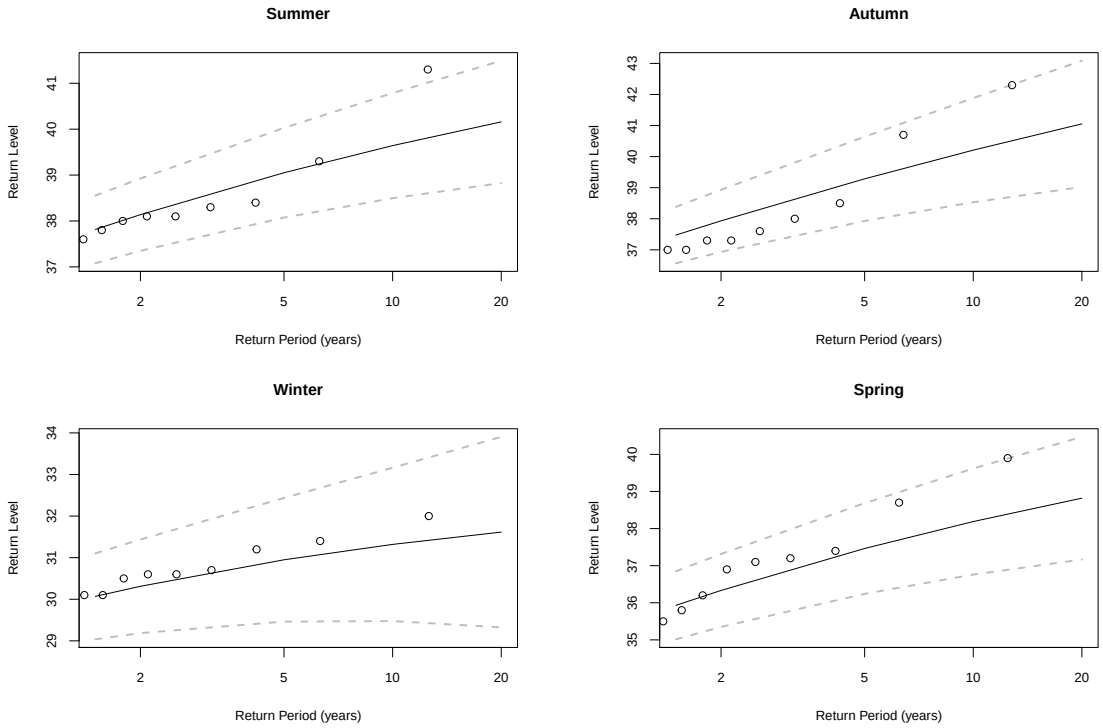


Figure 3.11: Return level plots for temperature at Cape Town International Airport

The negative estimates for the shape parameters in Table 3.10 is reflected by the convex shape of the return level plots (Figure 3.11 for Cape Town International Airport and Figures A.32 - A.35 for the other four stations). The general trend of the confidence intervals for all the stations is that is starts off fairly wide and continues to get even wider which indicates the uncertainty in prediction power as the return period increases.

Diagnostic Plots

Figure 3.12 and Figures A.36 - A.39 show the quantile-quantile plots for Cape Town International Airport and the other stations, respectively.

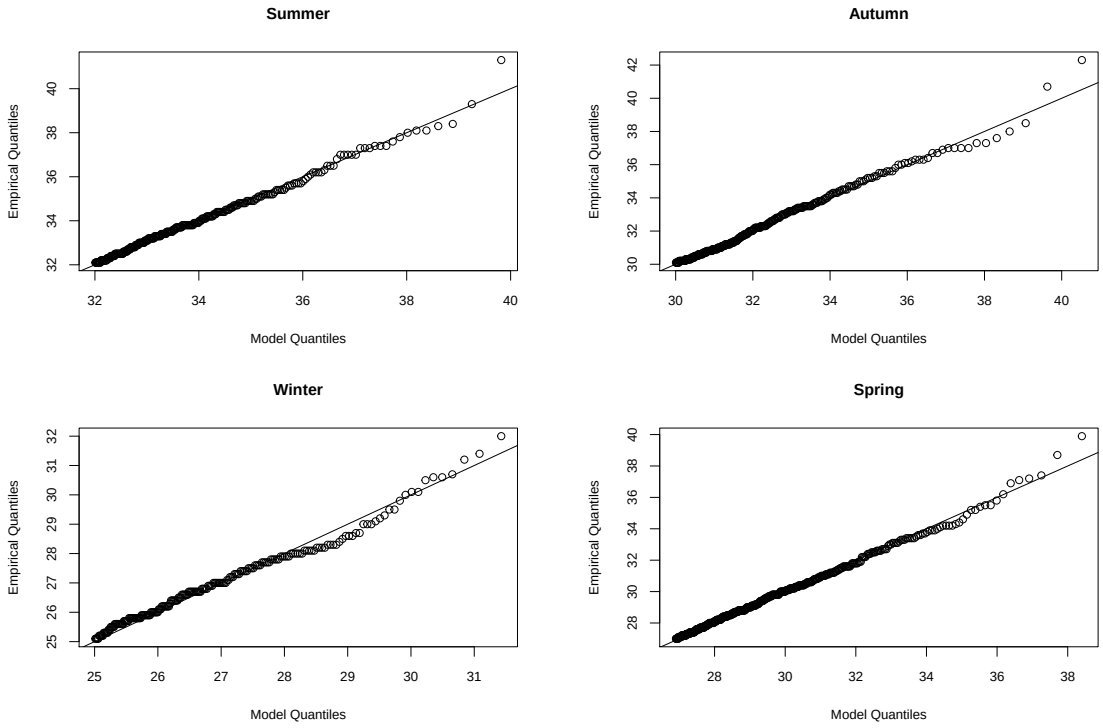


Figure 3.12: Quantile-quantile plots for temperature at Cape Town International Airport

From the above plots, the models fit the data well as the plots are close to linearity.

3.3.3 Maximum Wind Speed

The following sections show and explain MRL plots, declustering of the data at the chosen thresholds, parameter estimates, return level estimates with plots and quantile-quantile plots for wind speed. The wind speed data for the Langebaanweg station is not available and is excluded in the analysis below.

Figure 3.13 shows the MRL plot for Cape Town International Airport. Figures A.21 - A.23 show the wind speed mean residual life plots for other three stations.

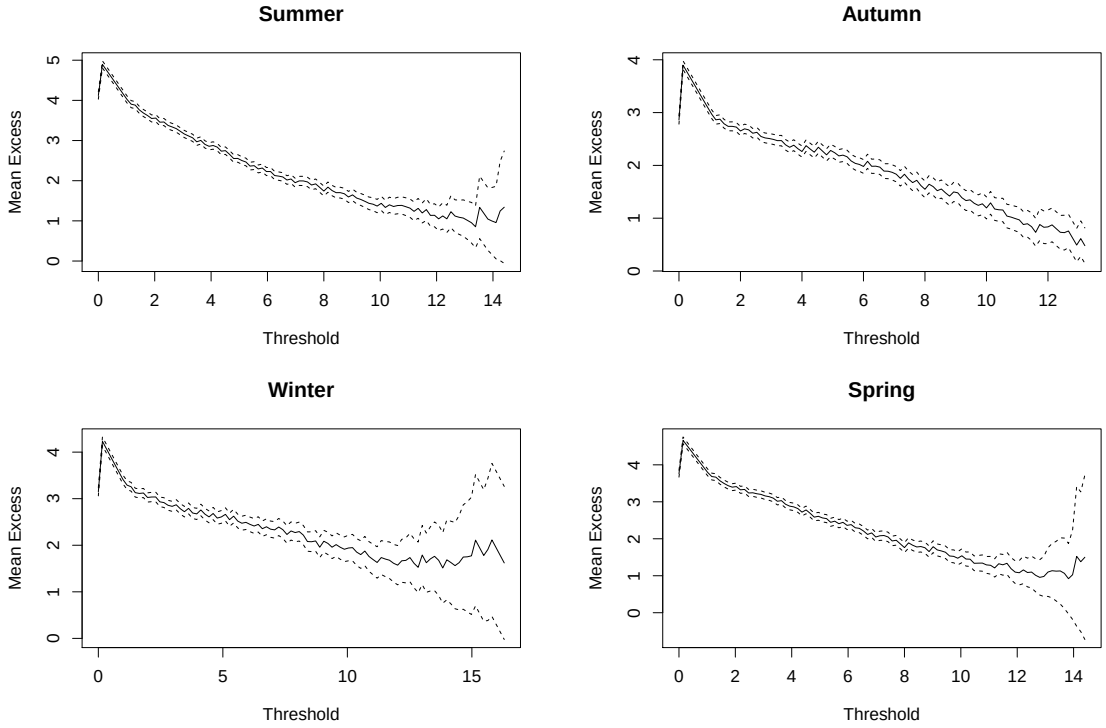


Figure 3.13: Mean residual life plots for wind speed at Cape Town International Airport

These MRL plots for wind speed are not as smooth as the temperature MRL plots, in particular the wind speed data for Vredendal station deviates from linearity in several instances. Table 3.11 shows the chosen thresholds with the corresponding quantiles.

Table 3.11: Threshold values for wind speed

	Season	Threshold	Quantile (%)
CT Airport	Summer	10.20	95.8
	Autumn	9.00	96.2
	Winter	10.00	96.5
	Spring	9.30	93.9
George Airport	Summer	5.20	93.0
	Autumn	6.00	95.6
	Winter	7.00	94.7
	Spring	6.00	93.4
Plettenberg Bay	Summer	5.00	93.0
	Autumn	6.00	96.8
	Winter	6.00	93.1
	Spring	6.00	94.6
Vredendal	Summer	5.10	96.4
	Autumn	2.70	94.2
	Winter	1.90	87.2
	Spring	3.00	90.0

Extremal Indices

The extremal indices calculated in Table A.5 Appendix A show that the wind speed data need not be declustered at the chosen thresholds. Table 3.12 shows the parameter estimates for maximum wind speed.

Parameter Estimation

The estimates for σ and ξ of wind speed at each station are summarised in Table 3.12 together with the respective 95% confidence intervals - shown in brackets - and the proportion of exceedance ζ_u. Standard errors are shown in parentheses.

Table 3.12: Threshold excess parameter estimates for wind speed (m/s)

		u	ζ_u	σ	ξ
CT Airport	Summer	10.20	183/4571 = 0.040	1.52 (0.14)	-0.09 (0.06)
	95% CI			[1.24; 1.81]	[-0.21; 0.02]
	Autumn	9.00	147/4679 = 0.031	2.10 (0.21)	-0.36 (0.07)
	95% CI			[1.68; 2.51]	[-0.49; -0.23]
	Winter	10.00	153/4600 = 0.033	2.20 (0.22)	-0.11 (0.06)
	95% CI			[1.76; 2.64]	[-0.23; 0.01]
	Spring	11.00	107/4550 = 0.024	1.38 (0.16)	-0.03 (0.07)
	95% CI			[1.06; 1.70]	[-0.17, 0.10]
George Airport	Summer	5.20	224/3339 = 0.067	1.53 (0.13)	-0.11 (0.05)
	95% CI			[1.27; 1.78]	[-0.22; -0.01]
	Autumn	6.00	144/3436 = 0.042	2.06 (0.22)	-0.05 (0.07)
	95% CI			[1.62; 2.49]	[-0.19; 0.08]
	Winter	7.00	182/3496 = 0.041	2.27 (0.22)	-0.10 (0.06)
	95% CI			[1.84; 2.71]	[-0.23; 0.02]
	Spring	6.00	213/3410 = 0.062	1.57 (0.16)	0.03 (0.08)
	95% CI			[1.25; 1.90]	[-0.13; 0.18]
Plettenberg Bay	Summer	5.00	147/2168 = 0.068	1.39 (0.14)	-0.22 (0.06)
	95% CI			[1.12; 1.66]	[-0.34; -0.10]
	Autumn	6.00	69/2253 = 0.031	1.60 (0.24)	-0.09 (0.10)
	95% CI			[1.08; 2.03]	[-0.29; 0.10]
	Winter	6.00	111/2300 = 0.048	1.51 (0.21)	-0.02 (0.10)
	95% CI			[1.10; 1.92]	[-0.22; 0.17]
	Spring	6.00	123/2315 = 0.053	1.32 (0.16)	-0.06 (0.08)
	95% CI			[1.00; 1.64]	[-0.22; 0.11]
Vredendal	Summer	5.10	113/4567 = 0.025	2.69 (0.29)	-0.15 (0.06)
	95% CI			[2.12; 3.27]	[-0.27; -0.04]
	Autumn	2.70	270/4672 = 0.058	1.99 (0.15)	-0.05 (0.05)
	95% CI			[1.69; 2.30]	[-0.14; 0.05]
	Winter	1.90	583/4584 = 0.127	1.28 (0.10)	0.30 (0.07)
	95% CI			[1.08; 1.48]	[0.17; 0.44]
	Spring	3.00	437/4545 = 0.096	2.61 (0.15)	-0.19 (0.03)
	95% CI			[2.32; 2.90]	[-0.25; -0.12]

The probability of exceedances is much higher than what is seen in the rainfall and temperature analysis. The shape parameter estimates calculated are negative which means that the Weibull distribution can be used to describe the wind speed extremes found at these stations.

Return Levels

The estimates of the various return levels are summarised in Table 3.13 where the confidence intervals are shown in brackets.

Table 3.13: Threshold excess return level estimates for wind speed (m/s)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	10.20	14.61	15.58	16.27	16.91
	95% CI		[13.92; 15.30]	[14.64; 16.53]	[15.08; 17.45]	[15.44; 18.38]
	Autumn	9.00	12.94	13.47	13.77	14.01
	95% CI		[11.85; 14.04]	[11.87; 15.68]	[11.87; 15.68]	[11.58; 16.43]
	Winter	10.00	15.91	17.25	18.18	19.04
	95% CI		[14.92; 16.90]	[15.94; 18.57]	[16.55; 19.82]	[17.03; 21.05]
	Spring	11.00	14.75	15.88	16.71	17.52
	95% CI		[13.99; 15.51]	[14.81; 16.94]	[15.33; 18.08]	[15.77; 19.27]
George Airport	Summer	5.20	10.03	10.90	11.50	12.05
	95% CI		[9.31; 10.75]	[9.93; 11.87]	[10.30; 12.70]	[10.59; 13.52]
	Autumn	6.00	12.42	13.95	15.06	16.13
	95% CI		[11.25; 13.60]	[12.28; 15.63]	[12.89; 17.22]	[13.38; 18.87]
	Winter	7.00	13.879	15.25	16.20	17.08
	95% CI		[12.77; 15.01]	[13.70; 16.80]	[14.24; 18.15]	[14.67; 19.49]
	Spring	6.00	12.35	13.97	15.23	16.51
	95% CI		[11.05; 13.65]	[11.88; 16.06]	[12.36; 18.10]	[12.71; 20.31]
Plettenberg Bay	Summer	5.00	8.63	9.12	9.42	9.69
	95% CI		[7.94; 9.32]	[8.25; 9.98]	[8.41; 10.44]	[8.50, 10.88]
	Autumn	6.00	10.19	11.21	11.93	12.60
	95% CI		[9.13; 11.24]	[9.75; 12.67]	[10.07; 13.78]	[10.27; 14.93]
	Winter	6.00	11.15	12.40	13.33	14.24
	95% CI		[9.92; 12.38]	[10.50; 14.30]	[10.77; 15.89]	[10.91; 17.58]
	Spring	6.00	10.37	11.34	12.03	12.70
	95% CI		[9.45; 11.30]	[9.98; 12.69]	[10.26; 13.80]	[10.46; 14.94]
Vredendal	Summer	5.10	11.40	12.87	13.86	14.75
	95% CI		[10.18; 12.62]	[11.37; 14.38]	[12.10; 15.62]	[12.68; 16.81]
	Autumn	2.70	9.55	11.05	12.15	13.21
	95% CI		[8.60; 10.50]	[9.70; 12.41]	[10.42; 13.88]	[11.04; 15.38]
	Winter	1.90	14.26	19.54	24.63	30.89
	95% CI		[10.84; 17.69]	[13.15; 25.94]	[14.82; 34.43]	[16.30; 45.47]
	Spring	3.00	10.66	11.66	12.31	12.88
	95% CI		[9.87; 11.46]	[10.68; 12.64]	[11.16; 13.45]	[11.55, 14.20]

Higher return levels are seen at Cape Town International Airport compared to the other four stations with Vredendal and George Airport having similar return levels and lower return levels seen at Plettenberg Bay.

Figure 3.14 and Figures A.40 - A.42 show the return level plots for wind speed at Cape Town International Airport and the other four stations, respectively.

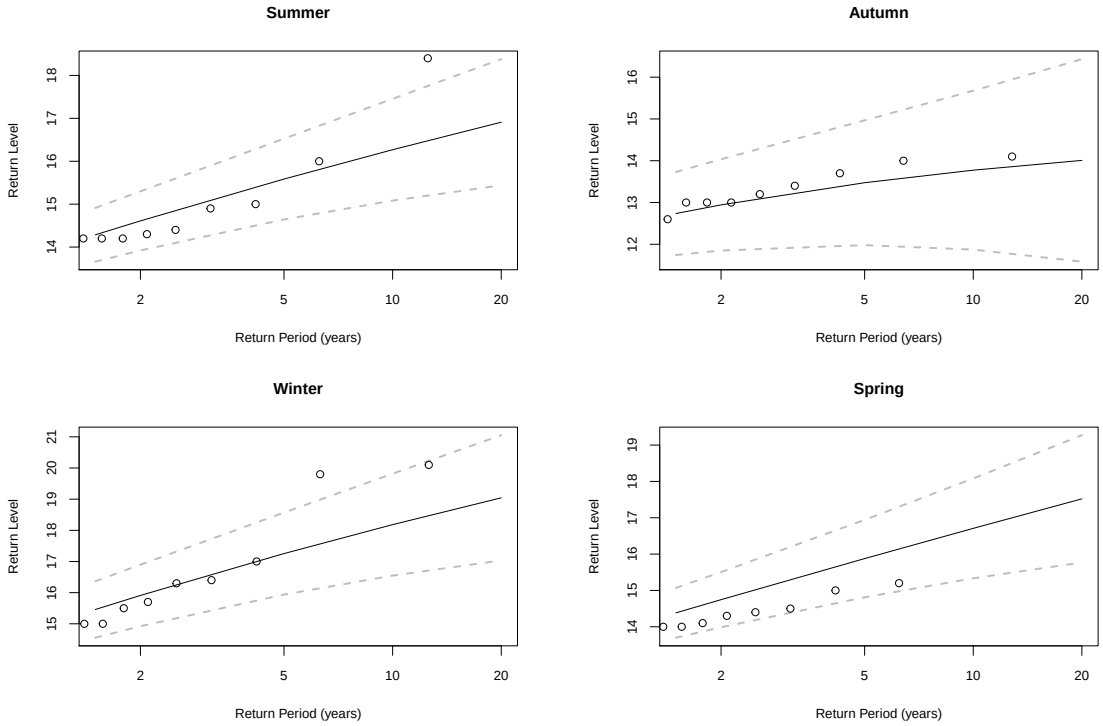


Figure 3.14: Return level plots for wind speed at Cape Town International Airport

Figure 3.14 and Figures A.40 - A.42 show convex shaped return level plots with wide confidence intervals that get wider as the return period increases. The negative estimates for the shape parameters in Table 3.10 is seen by the convex shape of the return level plots (Figure 3.14 for Cape Town International Airport and Figures A.40 - A.42 for the other four stations).

Diagnostic Plots

The quantile-quantile plots of the stations are show in Figures 3.15, A.43 - A.45.

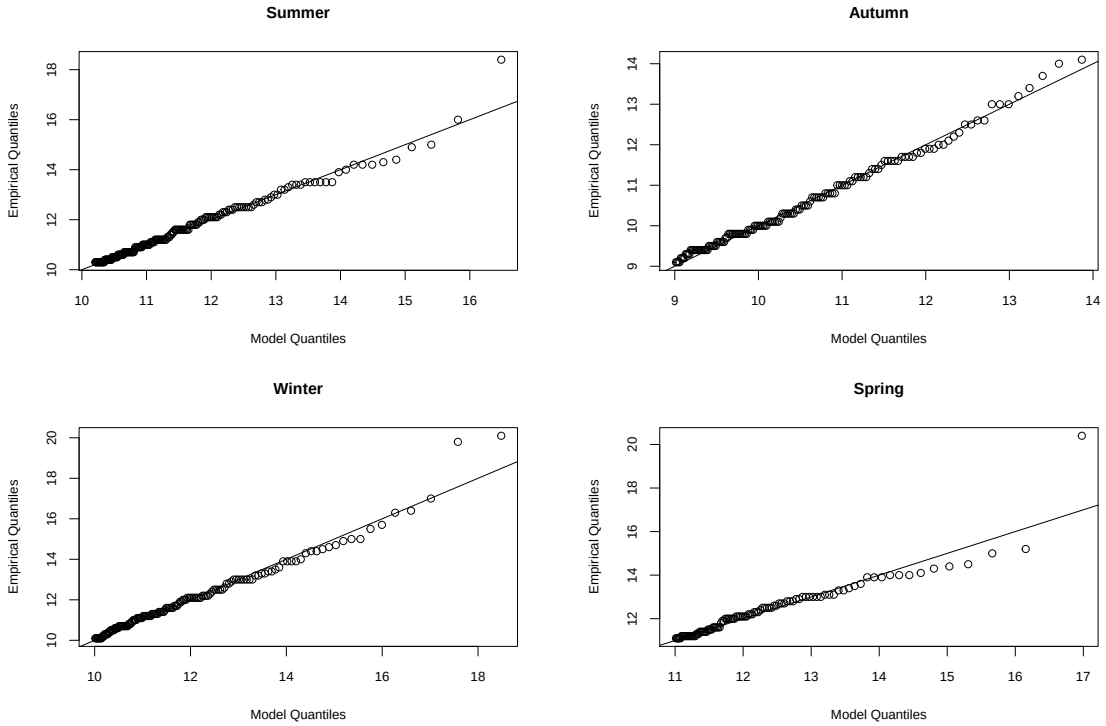


Figure 3.15: Quantile-quantile plots for wind speed at Cape Town International Airport

Figures 3.15 and A.43 - A.44 show a fairly straight line through the origin for Cape Town International Airport, George Airport and Plettenberg Bay, respectively. The quantile-quantile plot for Vredendal shown in Figure A.45 shows a line that deviates from linearity which indicates a poorly fitting model.

3.4 Univariate Analysis: Point Process

The point process approach also requires a threshold whereby observations which fall above the threshold are considered as extreme. Therefore, the same threshold values used in the threshold excess models are used for the point process models to enable comparison of the two approaches. Hence, the thresholds have already been declustered and all that remains is parameter estimation and assessing model fit.

3.4.1 Maximum Rainfall

Parameter Estimation

The estimates for μ, σ and ξ of rainfall at each station are summarised in Table 3.14 together with the respective 95% confidence intervals.

Table 3.14: Point process parameter estimates for maximum rainfall (mm)

		u	μ	σ	ξ
CT Airport	Summer	5.20	16.99 (1.38)	6.05 (1.06)	0.15 (0.10)
	95% CI		[14.29; 19.69]	[3.97; 8.12]	[-0.03; 0.35]
	Autumn	10.60	34.16 (2.45)	10.92 (1.96)	0.20 (0.08)
	95% CI		[29.36; 38.96]	[7.07; 14.77]	[0.04; 0.36]
	Winter	22.32	43.61 (2.04)	8.92 (1.38)	-0.10 (0.12)
	95% CI		[39.62; 47.62]	[6.21; 11.63]	[-0.34; 0.13]
	Spring	9.65	28.22 (2.25)	9.76 (1.93)	0.20 (0.11)
	95% CI		[23.82; 32.63]	[5.97; 13.55]	[-0.01; 0.42]
George Airport	Summer	17.43	47.34 (3.27)	14.06 (2.43)	0.05 (0.11)
	95% CI		[40.92; 53.77]	[9.31; 18.82]	[-0.16; 0.27]
	Autumn	19.69	55.61 (5.56)	23.80 (5.60)	0.35 (0.15)
	95% CI		[44.72; 66.51]	[12.82; 34.79]	[0.06; 0.63]
	Winter	12.78	55.34 (5.34)	22.88 (4.89)	0.25 (0.11)
	95% CI		[44.72; 65.63]	[13.30; 32.47]	[0.03; 0.47]
	Spring	17.40	64.99 (5.95)	25.54 (4.86)	0.21 (0.10)
	95% CI		[53.31; 76.67]	[16.01; 35.06]	[0.01; 0.41]
Langebaan-weg	Summer	3.60	11.58 (1.09)	4.10 (0.76)	-0.04 (0.17)
	95% CI		[9.44, 13.72]	[2.61; 5.59]	[-0.37; 0.29]
	Autumn	11.27	22.36 (1.77)	6.68 (1.18)	0.04 (0.15)
	95% CI		[18.89; 25.82]	[4.37; 8.99]	[-0.26; 0.33]
	Winter	15.10	25.89 (1.89)	7.25 (1.60)	0.24 (0.15)
	95% CI		[22.18; 29.60]	[4.10; 10.39]	[-0.07; 0.54]
	Spring	7.78	17.27 (1.44)	5.45 (0.89)	0.03 (0.12)
	95% CI		[14.44; 20.09]	[3.71; 7.20]	[-0.21, 0.26]
Plettenberg Bay	Summer	11.80	42.67 (6.11)	17.42 (5.74)	0.30 (0.17)
	95% CI		[30.69; 54.64]	[6.16; 28.67]	[-0.02; 0.63]
	Autumn	20.20	49.63 (4.75)	14.07 (3.63)	0.06 (0.17)
	95% CI		[40.32; 58.94]	[6.96; 21.17]	[-0.27; 0.38]
	Winter	22.29	57.27 (6.01)	17.44 (4.60)	0.09 (0.17)
	95% CI		[45.50; 69.04]	[8.42; 26.44]	[-0.23; 0.42]
	Spring	23.20	55.73 (6.96)	19.68 (5.28)	0.18 (0.16)
	95% CI		[42.08; 69.38]	[9.33; 30.03]	[-0.14; 0.50]
Vredendal	Summer	2.00	9.71 (1.55)	6.63 (1.77)	0.56 (0.15)
	95% CI		[6.66; 12.75]	[3.17; 10.10]	[0.27; 0.85]
	Autumn	3.00	18.06 (1.43)	6.33 (1.34)	0.07 (0.10)
	95% CI		[15.25; 20.86]	[21.04; 35.13]	[-0.12; 0.26]
	Winter	9.30	19.69 (1.24)	5.35 (0.91)	0.04 (0.12)
	95% CI		[17.26; 22.11]	[3.57; 7.12]	[-0.20; 0.29]
	Spring	2.54	12.74 (1.19)	5.17 (0.95)	0.19 (0.10)
	95% CI		[10.41; 15.06]	[3.24; 7.11]	[-0.01; 0.39]

Table 3.14 shows that negative estimates for the shape parameter at three stations. In particular, winter at Cape Town International Airport, summer at Langebaanweg and autumn at Plettenberg Bay. The negative shape parameter estimate is indicative of the maxima following a Weibull distribution. The remaining stations show positive non-zero estimates across the seasons which indicates a Fréchet distribution.

Return Levels

The estimates of the various return levels are shown in Table 3.15 with the confidence intervals shown in brackets.

Table 3.15: Point process return level estimates for maximum rainfall (mm)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	5.20	19.27	27.21	33.30	39.84
	95% CI		[15.88; 22.66]	[20.28; 34.14]	[22.53; 44.06]	[24.02; 55.66]
	Autumn	10.60	38.31	53.24	65.14	78.35
	95% CI		[32.13; 44.49]	[40.19; 66.29]	[44.81; 85.47]	[48.33; 108.33]
	Winter	22.32	46.82	56.00	61.49	66.38
	95% CI		[42.29; 51.36]	[48.23; 63.76]	[50.35; 72.65]	[51.28; 81.49]
	Spring	9.65	31.94	45.31	56.00	67.87
	95% CI		[26.23; 37.64]	[32.60; 58.02]	[35.45; 76.54]	[36.68; 99.07]
George Airport	Summer	17.43	52.55	69.31	80.98	92.62
	95% CI		[44.75; 60.35]	[54.26; 84.35]	[58.34; 103.61]	[60.46; 124.78]
	Autumn	19.69	64.92	102.44	136.75	176.20
	95% CI		[50.02; 79.81]	[62.83; 142.04]	[65.26; 208.23]	[59.26; 299.13]
	Winter	12.78	63.96	96.85	124.37	156.11
	95% CI		[50.05; 77.87]	[64.08; 129.62]	[69.93; 178.82]	[71.40; 240.83]
	Spring	17.40	74.71	109.91	138.18	169.77
	95% CI		[59.74; 89.69]	[77.59; 142.22]	[86.52; 189.85]	[91.75; 247.79]
Langebaan-weg	Summer	3.60	13.07	17.53	20.36	22.99
	95% CI		[10.63; 15.51]	[13.12; 21.96]	[13.65; 27.08]	[13.44; 32.55]
	Autumn	11.27	24.82	32.65	38.01	43.30
	95% CI		[20.76; 28.88]	[25.32; 39.98]	[26.90; 49.14]	[27.21; 59.39]
	Winter	15.10	28.66	38.93	47.40	57.06
	95% CI		[23.92; 33.41]	[28.13; 49.73]	[29.15; 65.65]	[28.10; 86.03]
	Spring	7.78	19.27	25.60	29.89	34.08
	95% CI		[15.97; 22.58]	[19.98; 31.22]	[21.80; 37.98]	[22.82; 45.35]
Plettenberg Bay	Summer	11.80	49.42	75.73	98.85	126.54
	95% CI		[33.24; 65.59]	[36.29; 115.18]	[31.83; 165.87]	[19.78; 233.30]
	Autumn	20.20	55.33	70.42	80.20	89.42
	95% CI		[44.85; 65.82]	[52.24; 88.61]	[54.07; 106.32]	[53.64; 125.19]
	Winter	22.29	63.77	85.33	100.90	116.88
	95% CI		[49.28; 78.26]	[56.36; 114.30]	[56.35; 145.44]	[52.30; 181.46]
	Spring	23.20	63.19	89.67	110.46	133.25
	95% CI		[46.12; 80.26]	[54.73; 124.61]	[55.35; 165.56]	[50.60; 215.89]
Vredendal	Summer	2.00	12.40	25.30	39.63	60.37
	95% CI		[7.95; 16.85]	[11.27; 39.33]	[11.25; 68.01]	[6.89; 113.84]
	Autumn	3.00	20.41	28.08	33.52	39.02
	95% CI		[16.89; 23.92]	[21.04; 35.13]	[22.93; 44.11]	[24.04; 54.00]
	Winter	9.30	21.66	29.98	32.34	36.65
	95% CI		[18.76; 24.56]	[22.40; 33.56]	[23.81; 40.86]	[24.38; 48.93]
	Spring	2.54	14.70	21.70	27.23	33.31
	95% CI		[11.71; 17.69]	[15.22; 28.18]	[16.92; 37.53]	[17.88; 48.75]

George Airport and Plettenberg Bay have higher return levels across all the seasons for the respective years compared to the other stations. These two stations have had higher recorded rainfall across the seasons compared to the other stations. As seen with the return level estimates under the threshold excess approach in Table 3.7, the confidence intervals become increasingly wider as the return period increases. The wider intervals reflect the uncertainty associated with the prediction levels far into the future.

Figure 3.16 and Figures A.46 - A.49 show the return level plots for rainfall at Cape Town International Airport and the other three stations, respectively.

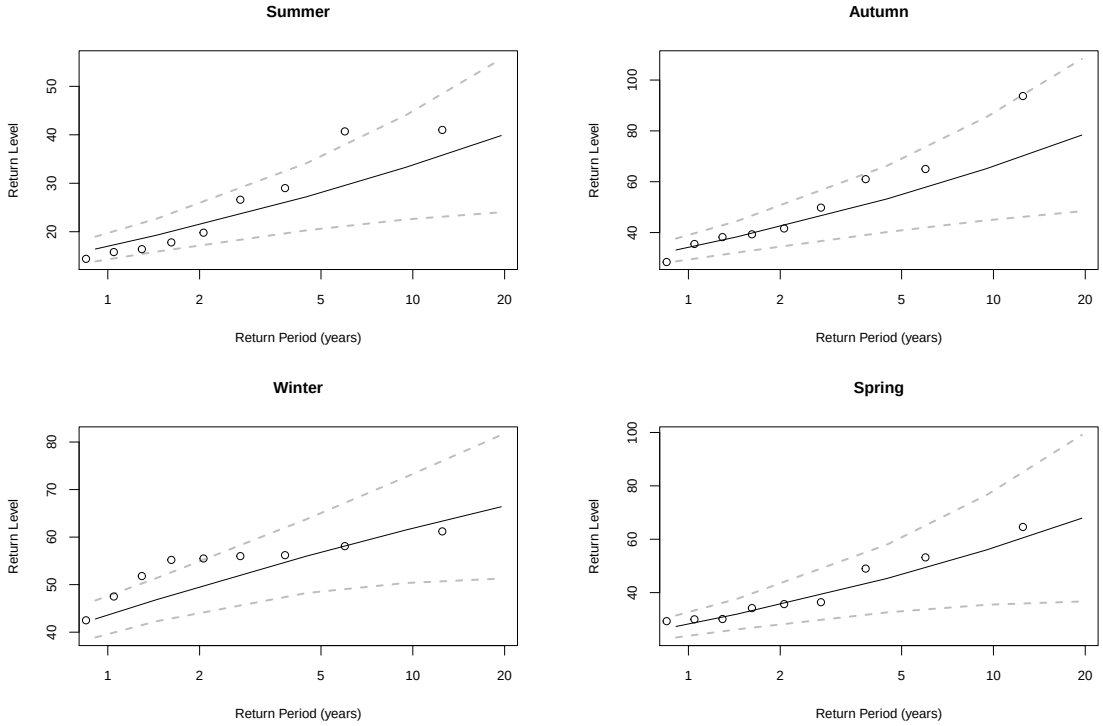


Figure 3.16: Return level plots for rainfall at Cape Town International Airport

The aforementioned plots show concave shaped return level plots with the exception of winter at Cape Town International Airport and summer at Langebaanweg which have negative shape parameter estimates. There is a general trend of confidence intervals that get wider as the return period increases.

Diagnostic Plots

The quantile-quantile plots of the stations are show in Figures 3.17 for Cape Town International Airport and A.50 - A.53 for the other stations.

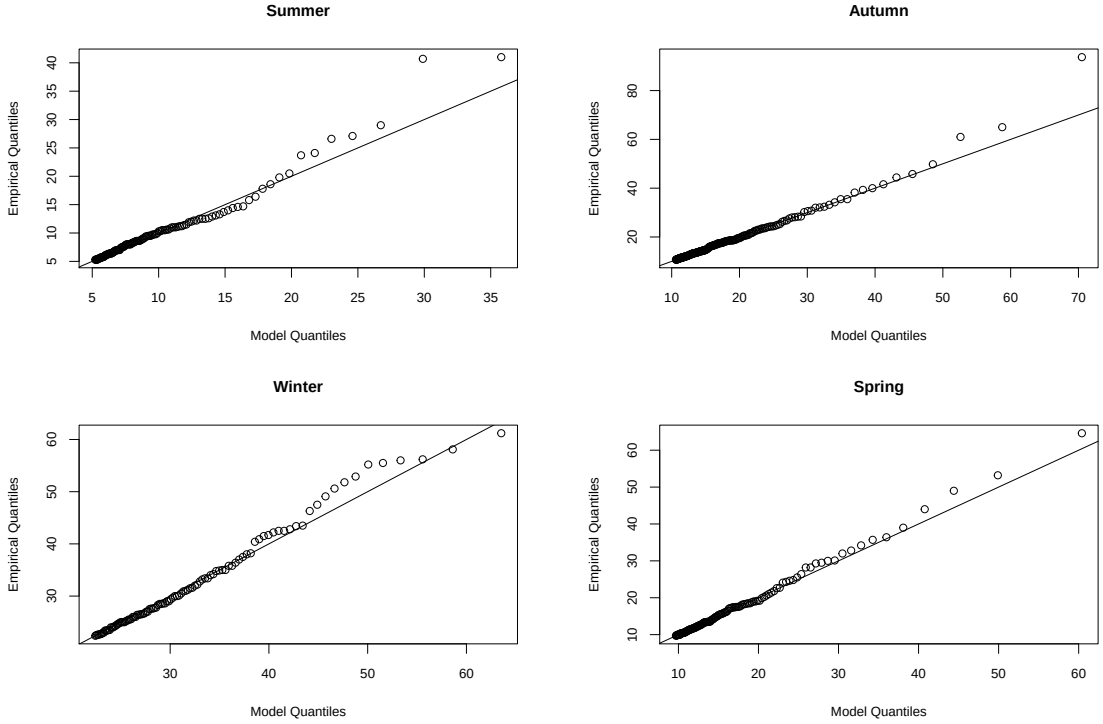


Figure 3.17: Quantile-quantile plots for rainfall at Cape Town International Airport

The fit of the models to the data for Cape Town International Airport show deviations from linearity at the higher quantiles as seen in Figure 3.17. Figures A.50 - A.52 show the similar patterns of model fit. The plots in Figure A.53 shows that there is deviations from a straight line for all quantile levels and indicates not a good fit.

3.4.2 Maximum Temperature

Parameter Estimation

The estimates for μ, σ and ξ of temperature at each station are summarised in Table 3.16 together with the respective 95% confidence intervals.

Table 3.16: Point process parameter estimates for temperature (^{0}C)

		u	μ	σ	ξ
CT Airport	Summer	32.00	37.33 (0.26)	1.24 (0.12)	-0.19 (0.04)
	95% CI		[36.83; 37.83]	[1.03; 1.45]	[-0.27; -0.11]
	Autumn	30.00	36.80 (0.36)	1.71 (0.19)	-0.13 (0.05)
	95% CI		[36.09; 37.51]	[1.33; 2.08]	[-0.22; -0.03]
	Winter	25.00	29.69 (0.21)	1.01 (0.08)	-0.33 (0.05)
	95% CI		[29.27; 30.10]	[0.86; 1.16]	[-0.41; -0.24]
	Spring	26.90	35.33 (0.31)	1.56 (0.13)	-0.20 (0.03)
	95% CI		[34.71; 35.94]	[1.30; 1.71]	[-0.27; -0.14]
George Airport	Summer	32.00	36.93 (0.49)	2.11 (0.31)	-0.03 (0.10)
	95% CI		[35.98; 37.89]	[1.49; 2.72]	[-0.22; 0.16]
	Autumn	33.00	37.31 (0.36)	1.55 (0.16)	-0.24 (0.09)
	95% CI		[36.61; 38.01]	[1.22; 1.87]	[-0.41; -0.08]
	Winter	27.00	31.17 (0.20)	0.95 (0.08)	-0.23 (0.04)
	95% CI		[30.79; 31.56]	[0.78; 1.11]	[-0.32; -0.15]
	Spring	30.99	36.49 (0.38)	1.68 (0.24)	-0.29 (0.11)
	95% CI		[35.74; 37.24]	[1.20; 2.15]	[-0.51; -0.08]
Langebaan-weg	Summer	34.00	40.84 (0.25)	1.11 (0.11)	-0.30 (0.04)
	95% CI		[40.35; 41.33]	[0.90; 1.32]	[-0.38; -0.21]
	Autumn	30.00	39.20 (0.24)	1.23 (0.08)	-0.28 (0.02)
	95% CI		[38.73; 39.67]	[1.08; 1.38]	[-0.32; -0.23]
	Winter	25.00	30.07 (0.21)	0.94 (0.08)	-0.25 (0.04)
	95% CI		[29.66; 30.47]	[0.78; 1.11]	[-0.33; -0.17]
	Spring	29.00	38.44 (0.33)	1.49 (0.14)	-0.25 (0.04)
	95% CI		[37.80; 39.09]	[1.25; 1.77]	[-0.32; -0.18]
Plettenberg Bay	Summer	28.00	35.16 (0.69)	2.15 (0.44)	-0.08 (0.10)
	95% CI		[33.81; 36.51]	[1.30; 3.01]	[-0.27; 0.11]
	Autumn	30.00	36.35 (0.48)	1.54 (0.29)	-0.23 (0.10)
	95% CI		[35.41; 37.28]	[0.97; 2.11]	[-0.43; -0.04]
	Winter	25.00	31.76 (0.25)	0.95 (0.07)	-0.35 (0.03)
	95% CI		[31.28; 32.25]	[0.82; 1.09]	[-0.41; -0.28]
	Spring	26.68	35.32 (0.80)	2.54 (0.49)	-0.14 (0.10)
	95% CI		[33.75; 36.88]	[1.58; 3.50]	[-0.34; 0.06]
Vredendal	Summer	39.10	42.58 (0.26)	1.26 (0.10)	-0.35 (0.05)
	95% CI		[43.93; 44.80]	[0.86; 1.25]	[-0.45; -0.24]
	Autumn	37.00	42.58 (0.26)	1.26 (0.10)	-0.24 (0.04)
	95% CI		[42.07; 43.08]	[1.05; 1.46]	[-0.32; -0.15]
	Winter	30.90	34.34 (0.21)	0.95 (0.11)	-0.25 (0.07)
	95% CI		[33.93; 34.74]	[0.74; 1.16]	[-0.40; -0.11]
	Spring	34.00	42.09 (0.29)	1.40 (0.13)	-0.23 (0.04)
	95% CI		[41.52; 42.66]	[1.14; 1.66]	[-0.30; -0.16]

Table 3.16 shows that the estimates for the shape parameter estimates are negative for all the stations across the seasons. Thus, the maxima at the 5 stations follow a Weibull distribution. This negative shape parameters agree to what was found with the threshold excess approach in Table 3.10. In particular, the shape parameter estimates and corresponding confidence intervals are the same.

Return Levels

The estimates of the various return levels are shown in Table 3.17 with the confidence intervals shown in brackets.

Table 3.17: Point process return level estimates for temperature (^{0}C)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	32.00	37.77	38.95	39.60	40.14
	95% CI		[37.23; 38.32]	[38.23; 39.67]	[38.73; 40.47]	[39.11; 41.17]
	Autumn	30.00	37.41	39.13	40.14	41.02
	95% CI		[36.60; 38.22]	[37.91; 40.35]	[38.58; 41.71]	[39.09; 42.96]
	Winter	25.00	30.03	30.88	31.29	31.60
	95% CI		[29.61; 30.46]	[30.40; 31.36]	[30.75; 31.84]	[30.99; 32.22]
	Spring	26.90	35.88	37.34	38.14	38.80
	95% CI		[35.21; 36.55]	[36.44; 38.24]	[37.06; 39.21]	[37.54; 40.05]
George Airport	Summer	32.00	37.70	40.02	41.52	42.93
	95% CI		[36.58; 38.81]	[38.10; 41.94]	[38.79; 41.94]	[39.22; 46.63]
	Autumn	33.00	37.85	39.26	39.99	40.58
	95% CI		[37.12; 38.59]	[38.30; 40.21]	[38.80; 41.18]	[39.12; 42.05]
	Winter	27.00	31.74	32.49	32.86	33.13
	95% CI		[31.42; 32.07]	[32.15; 32.84]	[32.49; 33.23]	[32.73; 33.53]
	Spring	30.99	37.07	38.52	39.24	39.80
	95% CI		[36.28; 37.86]	[37.36; 39.67]	[37.71; 40.77]	[37.87; 41.73]
Langebaan-weg	Summer	34.00	41.23	42.19	42.67	43.04
	95% CI		[40.70; 41.75]	[41.51; 42.86]	[41.88; 43.46]	[42.13; 43.95]
	Autumn	30.00	39.63	40.70	41.25	41.68
	95% CI		[39.13; 40.12]	[40.12; 41.29]	[40.59; 41.91]	[40.95; 42.41]
	Winter	25.00	30.40	31.25	31.69	32.05
	95% CI		[29.96; 30.83]	[30.69; 31.81]	[31.03; 32.35]	[31.29; 32.81]
	Spring	29.00	38.97	40.31	41.01	41.48
	95% CI		[38.26; 39.68]	[39.36; 41.26]	[39.89; 42.14]	[40.28; 42.87]
Plettenberg Bay	Summer	28.00	35.94	38.21	39.60	40.86
	95% CI		[34.35; 37.53]	[35.55; 40.86]	[36.01; 43.19]	[36.23; 45.48]
	Autumn	30.00	36.89	38.30	39.05	39.66
	95% CI		[35.84; 37.94]	[36.71; 39.89]	[37.01; 41.09]	[37.15; 42.16]
	Winter	25.00	32.09	32.88	33.26	33.54
	95% CI		[31.60; 34.58]	[32.38; 33.39]	[32.73; 33.79]	[32.97; 34.10]
	Spring	26.68	36.22	38.76	40.23	41.51
	95% CI		[34.43; 38.02]	[35.88; 41.64]	[36.40; 44.06]	[36.64; 46.37]
Vredendal	Summer	39.10	44.73	45.60	46.01	46.32
	95% CI		[44.28; 45.18]	[45.05; 46.15]	[45.33; 46.66]	[45.57; 47.07]
	Autumn	37.00	43.02	44.17	44.77	45.26
	95% CI		[42.48; 43.56]	[43.48; 44.85]	[43.96; 45.58]	[44.32; 46.21]
	Winter	30.90	42.59	43.87	44.55	45.10
	95% CI		[41.96; 43.21]	[43.01; 44.73]	[43.51; 45.58]	[43.89; 46.30]
	Spring	34.00	41.96	43.21	43.01	45.58
	95% CI		[41.96; 43.21]	[43.01; 44.73]	[43.51; 45.58]	[43.89; 46.30]

Table 3.17 shows that there is not one specific station that exhibits much greater return levels that the other stations compared to what is seen in the rainfall case. Confidence intervals get wider as the return period increases but to great extent e.g. the 20-year return level confidence interval width at Vredendal for winter is less than 2.

Figure 3.18 and Figures A.54 - A.57 show the return level plots for temperature for Cape Town International Airport and the other four stations, respectively.

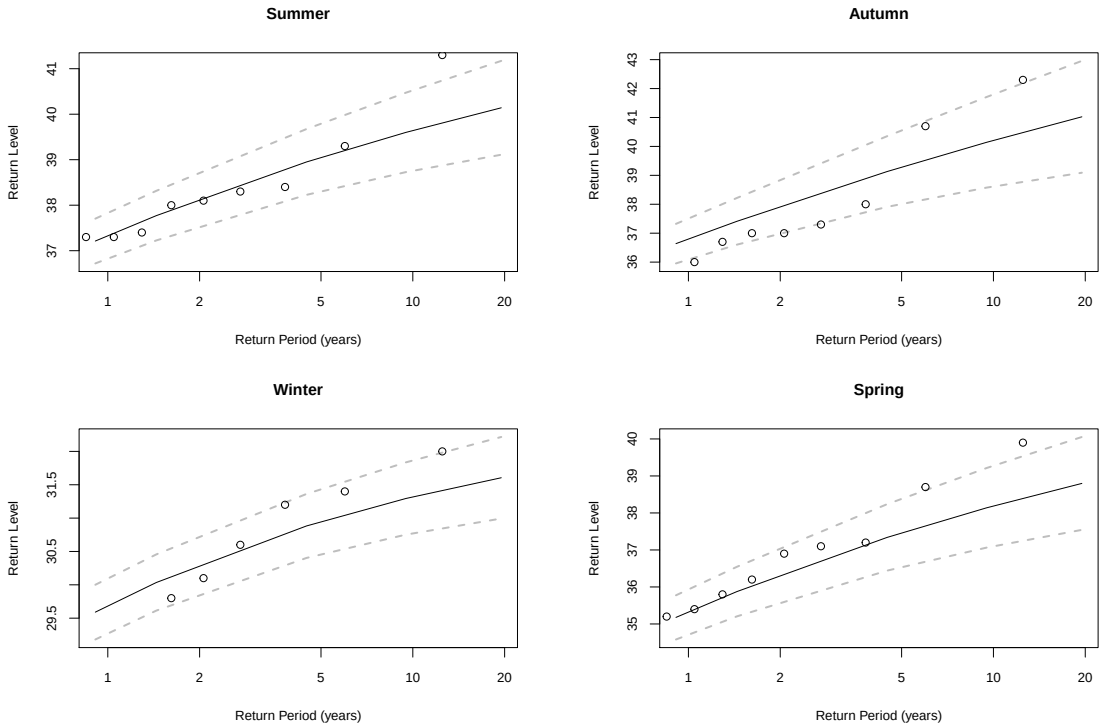


Figure 3.18: Return level plots for temperature at Cape Town International Airport

The above plots show convex shaped return levels which mirror the negative estimates of the shape parameter in Table 3.16.

Diagnostic Plots

The quantile-quantile plots of the stations are show in Figure 3.19 and Figures A.58 - A.61.

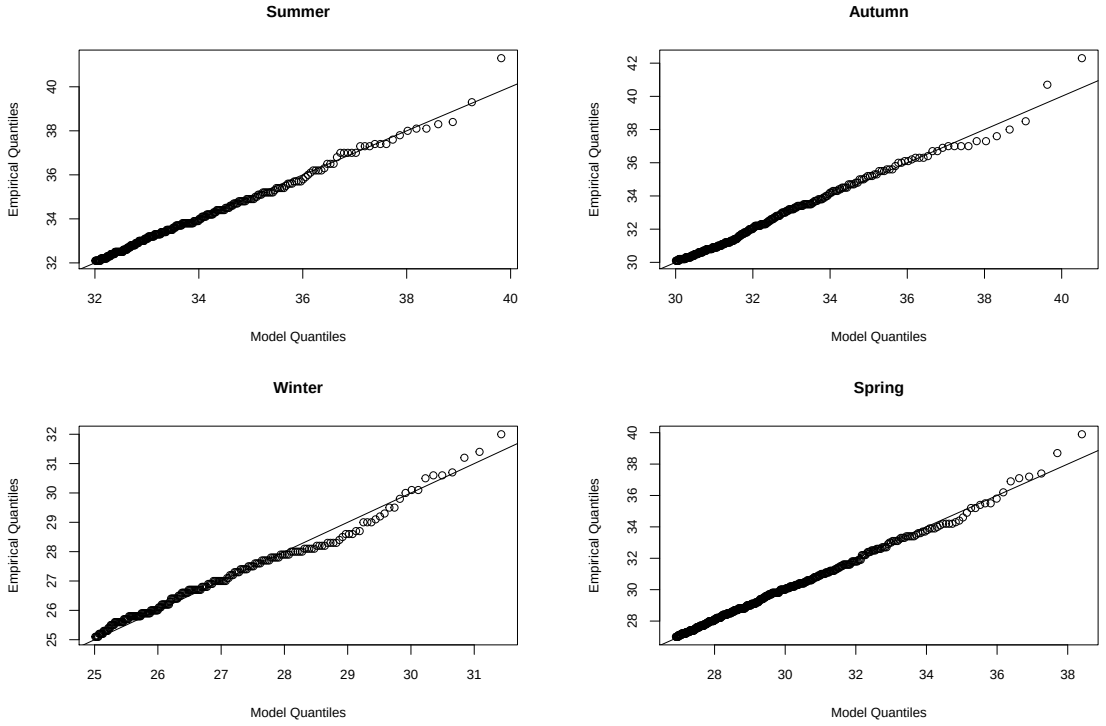


Figure 3.19: Quantile-quantile plots for temperature at Cape Town International Airport

The quantile-quantile plots for Cape Town International Airport in Figures 3.19 and the other 4 stations in A.58 - A.61 show a general good fit as there not much deviation from linearity.

3.4.3 Maximum Wind Speed

Parameter Estimation

The estimates for μ, σ and ξ for wind speed at each station are summarised in Table 3.18 together with the respective 95% confidence intervals.

Table 3.18: Point process return parameter estimates for wind speed (m/s)

		u	μ	σ	ξ
CT Airport	Summer	10.20	13.81 (0.26)	1.19 (0.13)	-0.09 (0.06)
	95% CI		[13.31; 14.32]	[0.92; 1.45]	[-0.21, 0.02]
	Autumn	9.00	12.41 (0.19)	0.87 (0.08)	-0.36 (0.07)
	95% CI		[12.04; 12.78]	[0.72; 1.03]	[-0.49; -0.23]
	Winter	10.00	14.80 (0.37)	1.67 (0.18)	-0.11 (0.06)
	95% CI		[14.08; 15.52]	[1.31; 3.02]	[-0.23; 0.01]
	Spring	11.00	13.87 (0.29)	1.28 (0.15)	-0.03 (0.07)
	95% CI		[13.29; 14.44]	[0.98; 1.58]	[-0.17; 0.10]
George Airport	Summer	5.20	9.32 (0.27)	1.07 (0.14)	-0.11 (0.05)
	95% CI		[8.80; 9.84]	[0.81; 1.34]	[-0.22; -0.01]
	Autumn	6.00	11.22 (0.45)	1.77 (0.24)	-0.05 (0.07)
	95% CI		[10.34; 12.09]	[1.30; 2.25]	[-0.19; 0.08]
	Winter	7.00	12.77 (0.41)	1.67 (0.22)	-0.10 (0.06)
	95% CI		[11.96; 13.58]	[1.24; 2.11]	[-0.23; 0.02]
	Spring	6.00	11.14 (0.45)	1.72 (0.31)	0.03 (0.08)
	95% CI		[10.26; 12.03]	[1.10; 2.33]	[-0.13; 0.18]
Plettenberg Bay	Summer	5.00	8.19 (0.20)	0.68 (0.09)	-0.22 (0.06)
	95% CI		[7.79; 8.60]	[0.51; 0.86]	[-0.34; -0.10]
	Autumn	6.00	9.36 (0.39)	1.24 (0.21)	-0.09 (0.10)
	95% CI		[8.59; 10.13]	[0.83; 1.65]	[-0.29; 0.10]
	Winter	6.00	10.18 (0.44)	1.41 (0.29)	-0.02 (0.10)
	95% CI		[9.32; 11.05]	[0.84; 1.97]	[-0.22; 0.17]
	Spring	6.00	9.61 (0.34)	1.12 (0.20)	-0.06 (0.08)
	95% CI		[0.94; 10.28]	[0.72; 1.52]	[-0.22; 0.11]
Vredendal	Summer	5.10	10.13 (0.44)	1.92 (0.18)	-0.15 (0.06)
	95% CI		[9.28; 10.99]	[1.57; 2.27]	[-0.27; -0.04]
	Autumn	2.70	8.37 (0.36)	1.73 (0.19)	-0.05 (0.05)
	95% CI		[7.66; 9.08]	[1.35; 2.11]	[-0.14; 0.05]
	Winter	1.90	11.13 (1.04)	4.06 (0.85)	0.30 (0.07)
	95% CI		[9.11; 13.16]	[2.38; 5.73]	[0.17; 0.44]
	Spring	3.00	9.79 (0.27)	1.34 (0.11)	-0.19 (0.03)
	95% CI		[9.27; 10.31]	[1.14; 1.55]	[-0.25; -0.12]

The point process estimates for wind speed in Table 3.18 show agreement with the threshold excess estimates of the shape parameter. The same shape parameter estimates and confidence intervals are found from both approaches. The shape parameter estimates are all negative with the exception of winter at Vredendal where $\hat{\xi} > 0$ and is better described to follow a Fréchet family of distributions instead of a Weibull distribution as in the other stations.

Return Levels

The estimates of the various return levels for maximum wind speed are shown in Table 3.19 with the confidence intervals shown in brackets.

Table 3.19: Point process return level estimates for maximum wind speed (m/s)

		u	2-year	5-year	10-year	20-year
CT Airport	Summer	10.20	14.28	15.50	16.22	16.86
	95% CI		[13.71; 14.86]	[14.68; 16.32]	[15.18; 17.26]	[15.56; 18.16]
	Autumn	9.00	12.71	13.42	13.75	14.00
	95% CI		[12.33; 13.08]	[12.99; 13.85]	[13.26; 14.25]	[13.43; 14.57]
	Winter	10.00	15.40	17.10	18.12	19.01
	95% CI		[14.59; 16.20]	[15.93; 18.27]	[16.60; 19.63]	[17.11; 20.91]
	Spring	11.00	14.33	15.74	16.65	17.49
	95% CI		[13.67; 14.99]	[14.74; 16.75]	[15.31; 17.98]	[15.77; 19.21]
George Airport	Summer	5.20	9.69	11.02	11.89	12.71
	95% CI		[8.96; 10.43]	[9.72; 12.33]	[10.04; 13.74]	[10.23; 15.19]
	Autumn	6.00	11.86	13.77	14.98	16.09
	95% CI		[10.86; 12.87]	[12.20; 15.35]	[12.88; 17.08]	[13.39; 18.78]
	Winter	7.00	13.34	15.10	16.16	17.12
	95% CI		[12.40; 14.28]	[13.64; 16.55]	[14.28; 18.04]	[14.77; 19.46]
	Spring	6.00	11.87	13.42	14.34	15.14
	95% CI		[11.87; 12.73]	[12.15; 14.69]	[12.68; 16.00]	[13.04; 17.24]
Plettenberg Bay	Summer	5.00	8.43	9.06	9.40	9.68
	95% CI		[8.00; 8.87]	[8.49; 9.64]	[8.70; 10.10]	[8.86; 10.50]
	Autumn	6.00	9.80	11.09	11.87	12.57
	95% CI		[8.93; 10.68]	[9.77; 12.42]	[10.12; 13.63]	[10.33; 14.82]
	Winter	6.00	10.70	12.25	13.26	14.21
	95% CI		[9.66; 11.73]	[10.45; 14.05]	[10.76; 15.76]	[10.91; 17.51]
	Spring	6.00	10.02	11.22	11.98	12.68
	95% CI		[9.23; 10.80]	[9.94; 12.50]	[10.26; 13.70]	[10.47; 14.88]
Vredendal	Summer	5.10	10.81	12.71	13.79	14.72
	95% CI		[9.89; 11.75]	[11.49; 13.92]	[12.32; 15.26]	[12.96; 16.47]
	Autumn	2.70	9.00	10.88	12.07	13.17
	95% CI		[8.18; 9.82]	[9.59; 12.16]	[10.38; 13.75]	[11.03; 15.30]
	Winter	1.90	12.71	18.83	24.20	30.63
	95% CI		[10.01; 15.40]	[12.89; 24.77]	[14.74; 33.67]	[16.32; 33.67]
	Spring	2.00	10.27	11.55	12.26	12.86
	95% CI		[9.70; 10.84]	[10.79; 12.31]	[11.36; 13.17]	[11.80; 13.92]

Cape Town International Airport has slightly higher return level estimates over the various return periods compared to the other 3 stations as seen in Table 3.13. Again the same pattern emerges i.e. increasing confidence intervals is seen for wind speed at these stations as the return period increases.

Figure 3.20 and Figures A.62 - A.64 show the return level plots for wind speed at Cape Town International Airport and the other three stations, respectively.

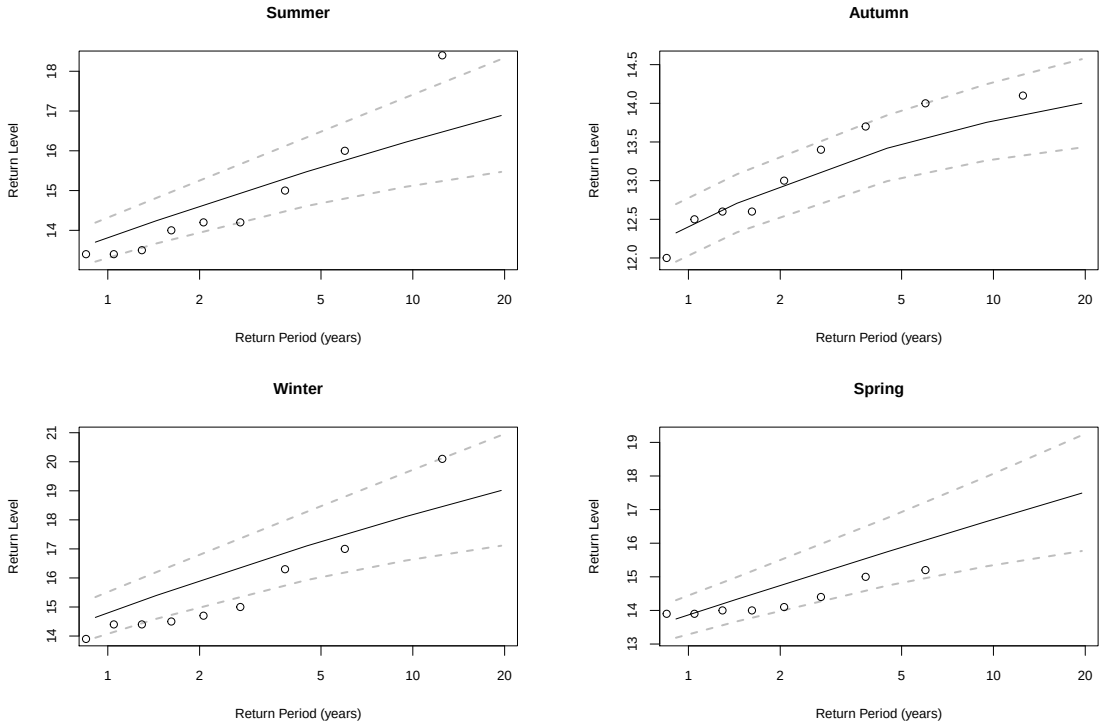


Figure 3.20: Return level plots for wind speed at Cape Town International Airport

Figure 3.20 and Figures A.62 - A.64 show convex shaped return level plots with confidence intervals that get wider as the return period increases. Estimates that are outside the confidence interval indicates that the prediction of return levels for a future period is poor.

Diagnostic Plots

The quantile-quantile plots of the stations are show in Figure 3.21 and Figures A.65 - A.67.

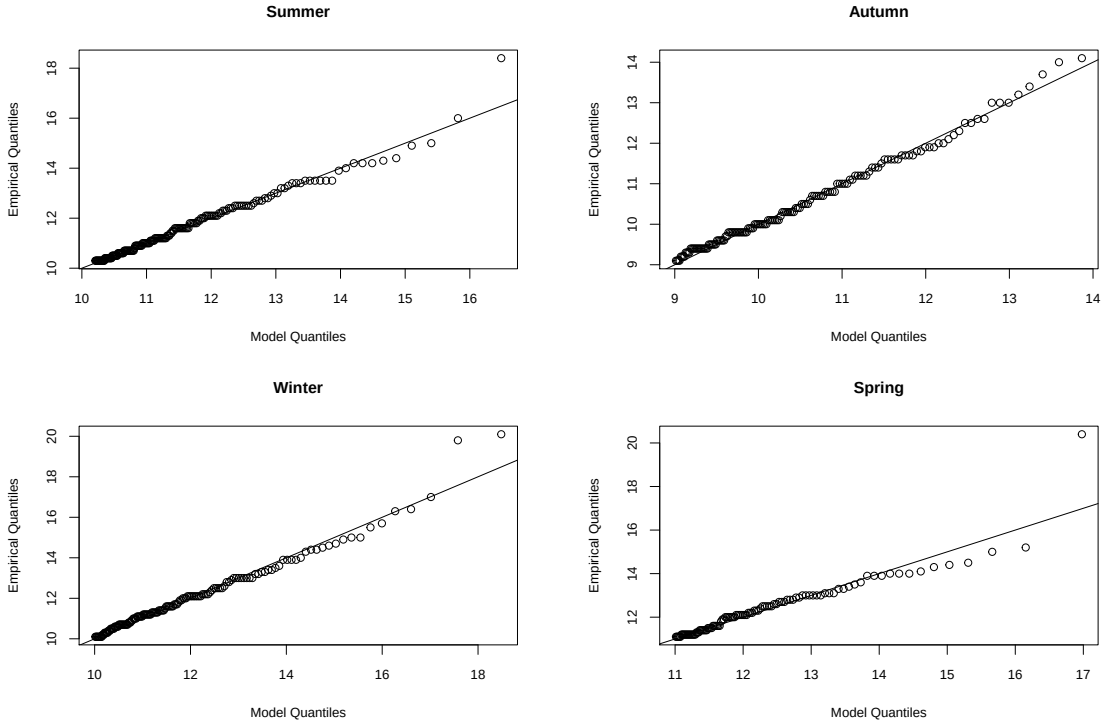


Figure 3.21: Quantile-quantile plots for wind speed at Cape Town International Airport

The quantile-quantile plots of Cape Town International Airport in Figure 3.21, George Airport and Plettenberg Bay in Figures A.65 - A.66, respectively, show a fairly good fit of the models to the wind speed data. Figure A.67 shows there is deviations from linearity for wind speed at Vredendal and the models are not a good fit for the models.

3.5 Multivariate Analysis

In this section, bivariate models are fitted to the climate data. Specifically, two different combinations of the variables are modelled - rainfall maxima with wind speed maxima and wind speed maxima with temperature maxima. Maximum rainfall with maximum temperature is not modelled because in the Western Cape province the combination of those two variables does not form part of the weather patterns.

3.5.1 Component-Wise Maxima

Parameter estimates are shown in Table 3.20 as well as negative log-likelihood (N-LL) which aids in model selection. Standard errors are shown in parentheses.

Table 3.20: Parameter estimates of the component-wise approach using the log model.

Station	Season	Wind Speed and Rainfall		Temperature and Wind Speed	
		α	N-LL	α	N-LL
CT Airport	Summer	0.9652 (0.067)	209.20	0.9993 (2 x 10^{-6})	209.94
	Autumn	0.8027 (0.101)	207.82	0.9996 (2 x 10^{-6})	209.81
	Winter	0.9692 (0.101)	205.26	0.9992 (2 x 10^{-6})	205.94
	Spring	0.9996 (2 x 10^{-6})	205.44	0.9742 (0.071)	205.18
George Airport	Summer	0.9739 (0.079)	149.89	0.9998 (2 x 10^{-6})	150.17
	Autumn	0.9993 (2 x 10^{-6})	208.89	0.9391 (0.117)	149.65
	Winter	0.9991 (2 x 10^{-6})	205.46	0.9992 (2 x 10^{-6})	150.28
	Spring	0.8862 (0.110)	149.50	0.9999 (2 x 10^{-6})	150.08
Plettenberg Bay	Summer	0.9992 (2 x 10^{-6})	170.77	0.9991 (2 x 10^{-6})	90.28
	Autumn	0.9992 (2 x 10^{-6})	78.24	0.9991 (2 x 10^{-6})	90.70
	Winter	0.9997 (2 x 10^{-6})	86.73	0.8991 (0.125)	94.58
	Spring	0.8129 (0.158)	68.86	0.8601 (0.136)	90.00
Vredendal	Summer	0.9997 (2 x 10^{-6})	201.56	0.9998 (2 x 10^{-6})	201.44
	Autumn	0.9882 (0.097)	196.25	0.9994 (2 x 10^{-6})	197.09
	Winter	0.9999 (2 x 10^{-6})	184.20	0.9992 (2 x 10^{-6})	184.19
	Spring	0.9993 (2 x 10^{-6})	197.04	0.9999 (2 x 10^{-6})	193.03

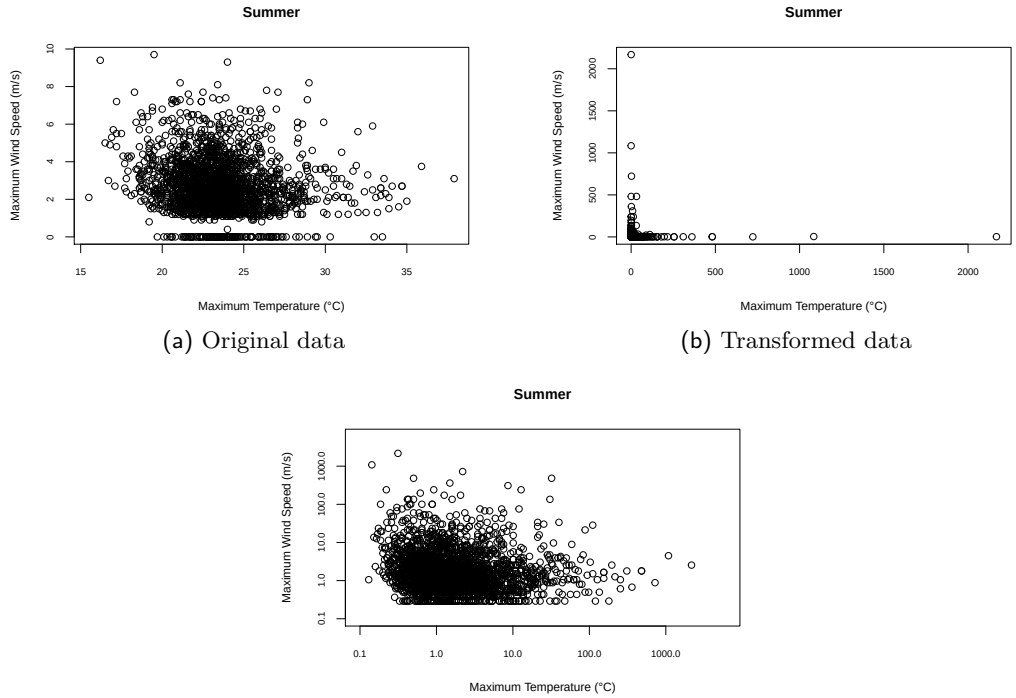
Table 3.20 shows that wind speed and rainfall exhibit weak dependence as the estimates are close to 1 ($\hat{\alpha} \geq 0.9652$) at the weather stations across all the seasons with a few exceptions. These exceptions include autumn at Cape Town International Airport, spring at George Airport and spring at Plettenberg Bay which have dependence estimates that are < 0.9. While the dependence estimates at the 3 aforementioned instances tend further away from compared to the other estimates, it is still a form of weak dependence. The standard errors associated with $\hat{\alpha} \geq 0.9991$ (almost 1) are extremely small which brings in to question the accuracy of these estimates. The N-LL vary according for the different seasons for any station. This may be attributed to fewer observations being modelled due to missing observations. For example, Plettenberg Bay had consecutive months of missing data and has noticeably lower N-LL values than the other stations.

The estimates for temperature and wind speed for winter and spring at Plettenberg Bay of 0.8991 and 0.8601 indicate less independence compared to the other stations which have estimates of $\hat{\alpha} \geq 0.9391$. Overall, there is still weak dependence exhibited by these stations under the component-wise approach. The standard errors again are extremely small (2 x 10^{-6}) .

Moreover, it should be noted that component-wise maxima do not provide information about the time structure of the data (Rakonczai, 2009). In other words, the maximum observations per block may not have occurred on the same day or even the same month for instance. The set-up of this approach therefore makes it difficult to provide accurate results for forecasting the occurrence of extreme events. Analysis of this approach in a univariate case is also not a reliable way to model extremes and so the results in Table 3.20 are not unexpected. Not knowing if the maxima have been observed at the same time makes the analysis of extreme events a difficult one, thus modelling the data under this approach is not pursued any further.

Bivariate Plots

Figure 3.22 shows the transformation (as mentioned in section 2.4.1) of maximum wind speed and maximum temperature from the raw data to standard Fréchet margins to the transformed data on a log scale.



(a) Original data

(b) Transformed data

(c) Transformed data on log axes

Figure 3.22: Wind Speed versus Maximum Temperature during Summer at Plettenberg Bay.

Figure 3.22 is included here to illustrate the transition of the data when transforming to standard Fréchet margins. The same process applies to all the stations for all the seasons with the two different pairwise combinations although the figures showing all the transformations are not included. In the case of the component-wise approach, the maxima from each block is transformed.

Maximum Wind Speed and Maximum Rainfall

For the component-wise approach, there are only 51 pairwise extreme observations with which to model. The interest in the threshold excess approach is in the top right quadrant (Figures 3.23 - 3.26) where x and y are extreme simultaneously. In terms of the point process approach, the change from Cartesian coordinates to angular and radial (pseudo-polar) coordinates is required as the structure of intensity function of the poisson process is easily stated in this form. This means that the the curved threshold allows for a greater number of observations to be incorporated into the analyses.

Figures 3.23 - 3.26 show the transformed wind speed and rainfall maxima on log scale with point process (- - -) and threshold excess thresholds (—) . To be able to compare

the results of the different approaches, the point process threshold is chosen to intersect the x and y axes at the same points that the threshold excess thresholds intersect the axes. The gaps between the data that are on some of the figures above are due to the zero observations.

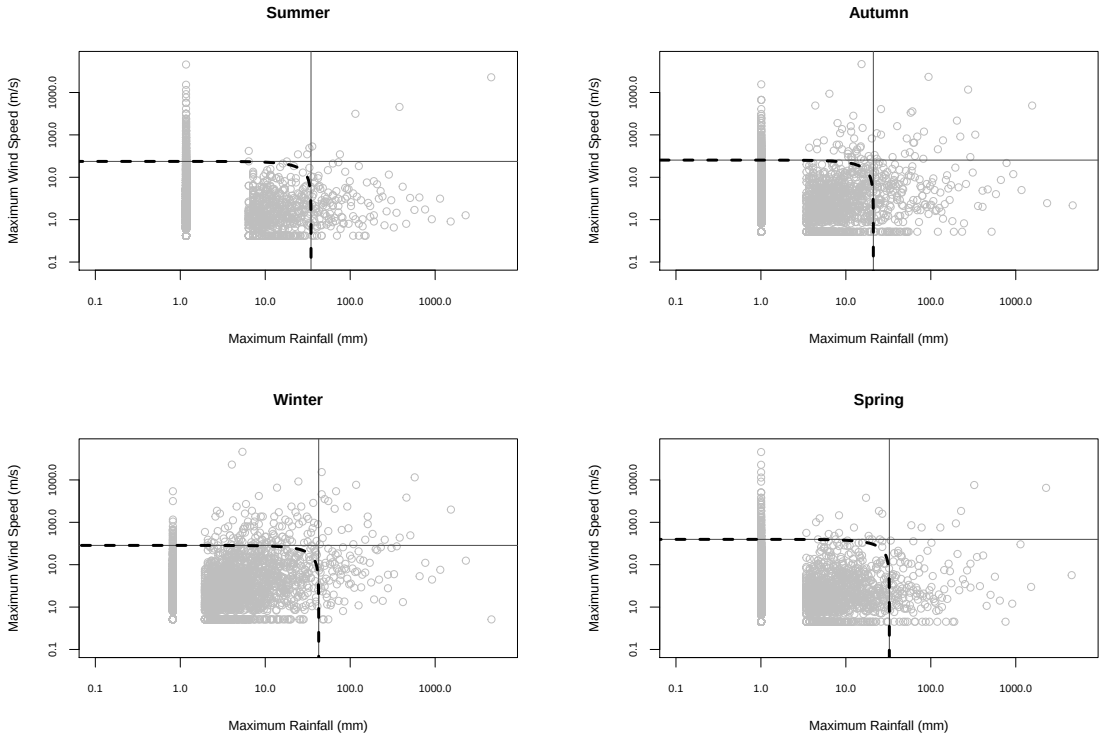


Figure 3.23: Maximum wind speed and maximum rainfall broken up into seasons for Cape Town International Airport.

In terms of the linear threshold, Figure 3.23 shows that for summer at Cape Town International Airport there are very few observations in the interested upper right quadrant i.e. 5 during summer, 27, 24 and 7 during autumn, winter and spring respectively. The small number of observations may lead to inaccuracy in the dependence estimates and standard errors. A reason for this is that in the summer season there are not many days which experience high levels of rainfall with wind speed. Autumn and spring have more observations although not to a great degree, with winter showing the most extremes for the given threshold for wind speed and rainfall. The curve threshold (- - -) allows for more observations to be taken into analyses, i.e. maxima from one and both margins are modelled. This is in comparison to the threshold excess approach which can only handle situations where there is exceedance in both margins as seen by the linearity of the threshold (—). There is some evidence of symmetry around the line $x = y$ which may allude to a logistic model to describe the dependence structure and is a starting point when deciding on which dependence model to use.

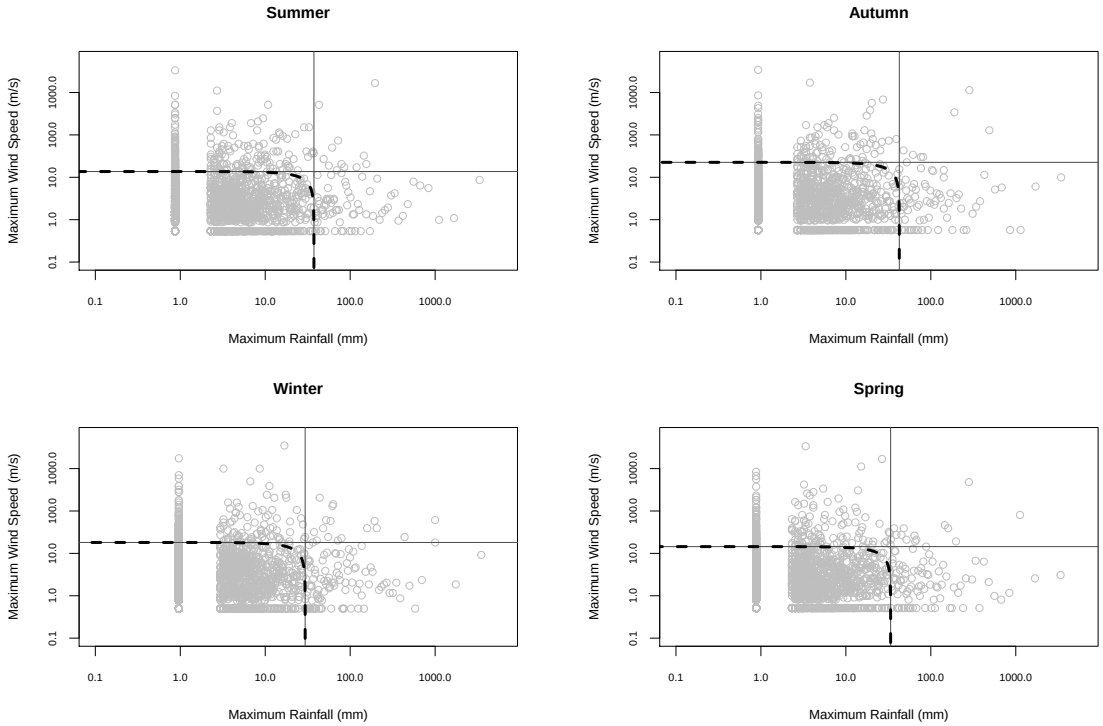


Figure 3.24: Maximum wind speed and maximum rainfall broken up into seasons for George Airport.

The exploratory plots for George Airport in Figure 3.24 show that there are few observations in the interested quadrant. In particular, there are 12, 7, 16 and 14 joint exceedances for summer, autumn, winter and spring respectively. The curved threshold shows many more observations that are able to be taken into the analysis.

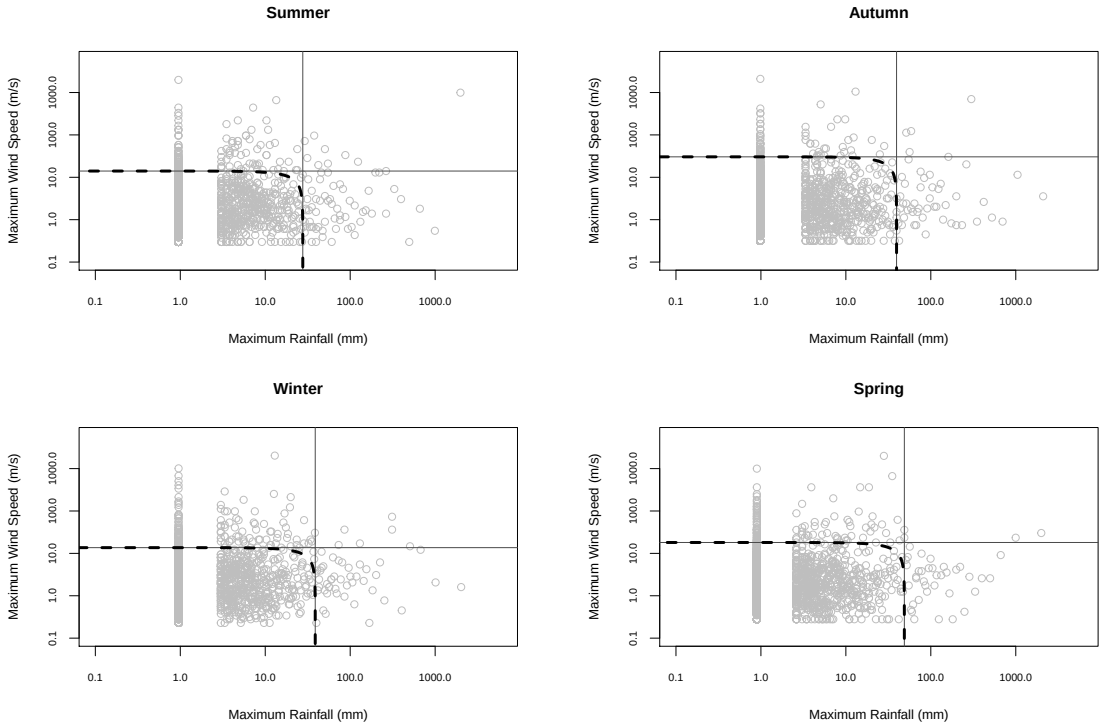


Figure 3.25: Maximum wind speed and maximum rainfall broken up into seasons for Plettenberg Bay.

Figure 3.25 shows that there are very few observations recorded in the interested quadrant for spring at Plettenberg Bay when looking at the linear threshold. There are 9 joint exceedances during summer, 4 during autumn, 6 during winter and 2 during spring. This may suggest that it is unlikely to observe extreme levels of wind speed and rainfall during spring at this station. More observations are above the threshold during summer, autumn and winter, but an even better option would be to use all the observations above the curved threshold.

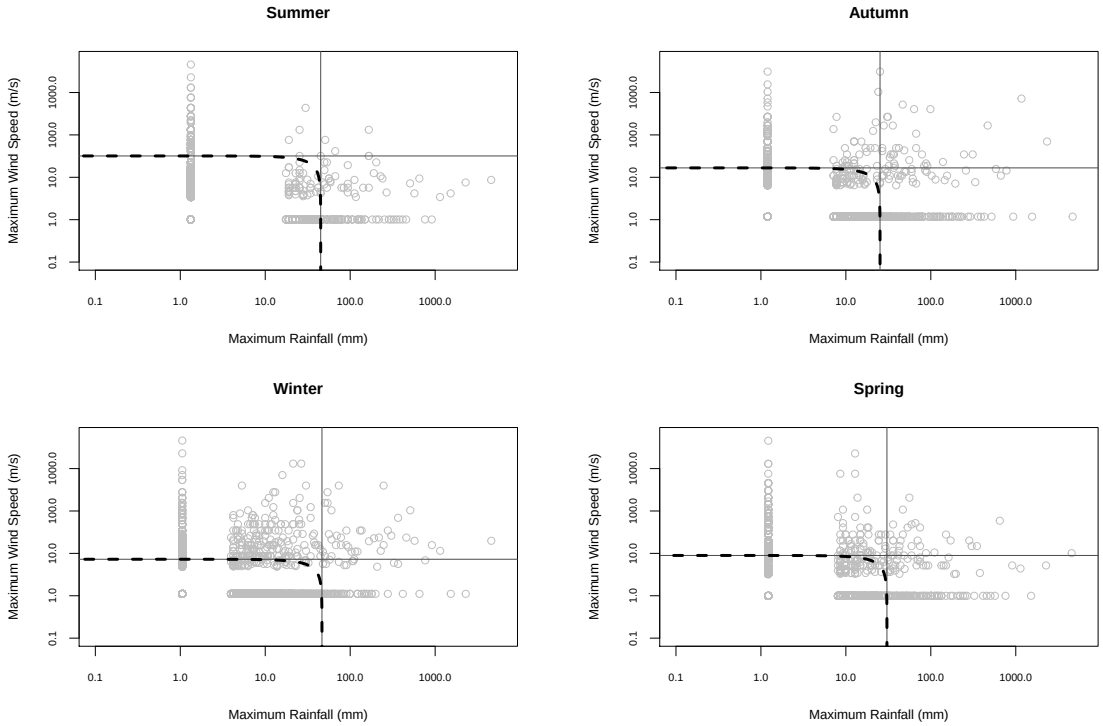


Figure 3.26: Maximum wind speed and maximum rainfall broken up into seasons for Vredendal.

Summer at Vredendal, in Figure 3.26, shows that most of the observations for wind speed and rainfall are zero with 3 joint exceedances. A similar pattern is seen during autumn and spring (32 and 27 joint exceedances, respectively) but more noticeably during the summer season. There are few observations during summer and spring above the linear threshold with many more wind speed and rainfall observations recorded during winter (37 joint exceedances) and autumn. Again the curved threshold enables the maxima of wind speed and rainfall from both and individual components to be incorporated into the analysis for improved estimates.

Maximum Temperature and Maximum Wind Speed

Figures 3.27 - 3.30 show the transformed wind speed and temperature maxima on log scale with point process (- - -) and threshold excess thresholds (—).

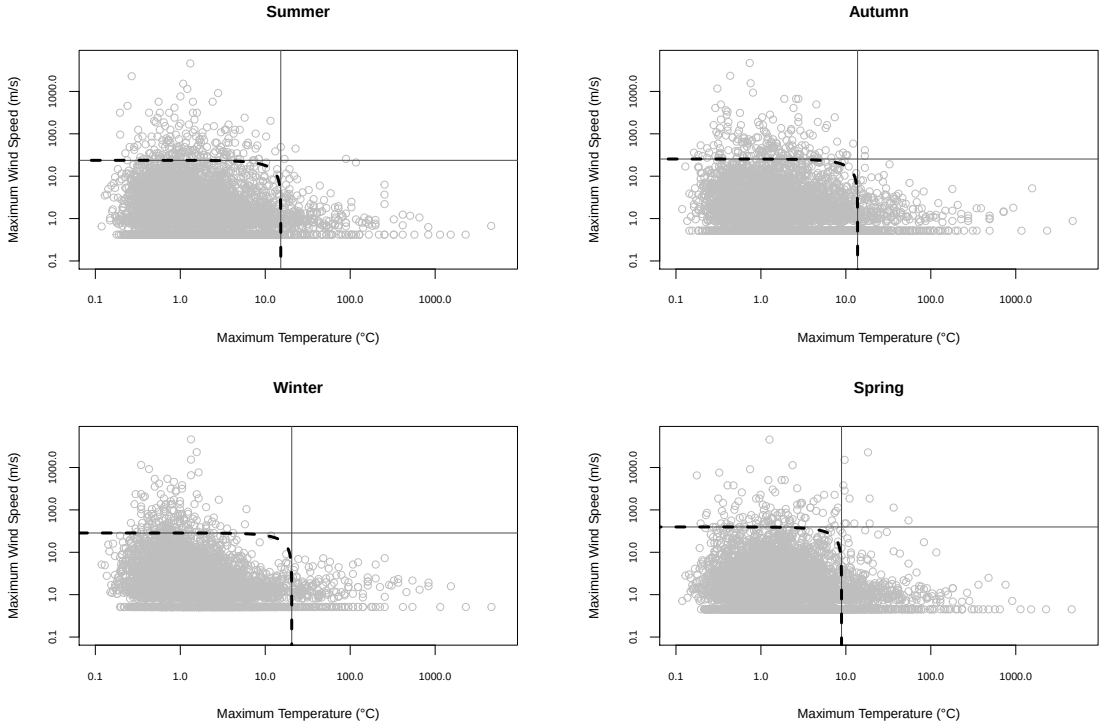


Figure 3.27: Maximum temperature and maximum wind speed broken up into seasons for Cape Town International Airport.

It can be seen from Figure 3.27 that the winter season at Cape Town International Airport has no observations in the interested quadrant which is characteristic of the climate conditions at this station. Spring has 12 joint exceedances with only a few (2 and 5 respectively during the summer and autumn seasons). The extreme levels of wind speed and temperature that have occurred, can be captured by the point process models by using the curved threshold.

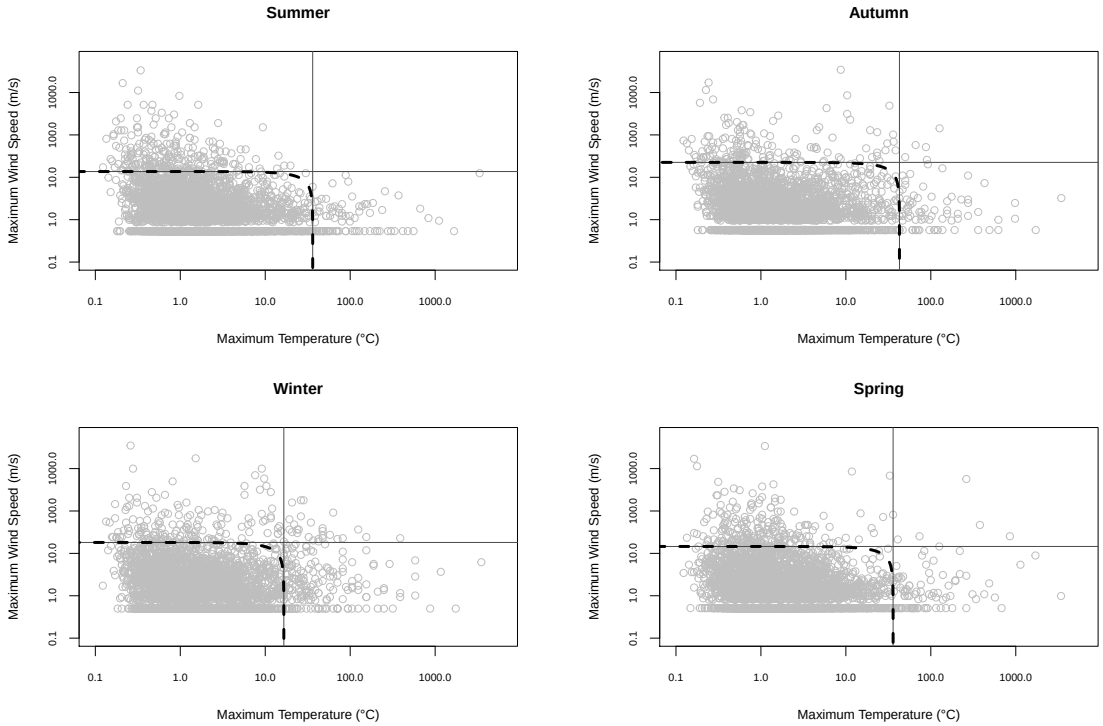


Figure 3.28: Maximum temperature and maximum wind speed broken up into seasons for George Airport.

There are no observations in the quadrant of interest during summer with marginally more during autumn with 5 observations and spring with 5 observations as shown in Figure 3.28. Winter has the most - 27 - observations above the linear thresholds compared to the other 3 seasons at George Airport. Summer exhibits high readings of temperature for low levels of wind speed and high wind speed observations for low values of temperature which can be modelled by the point process approach and not by the threshold excess approach.

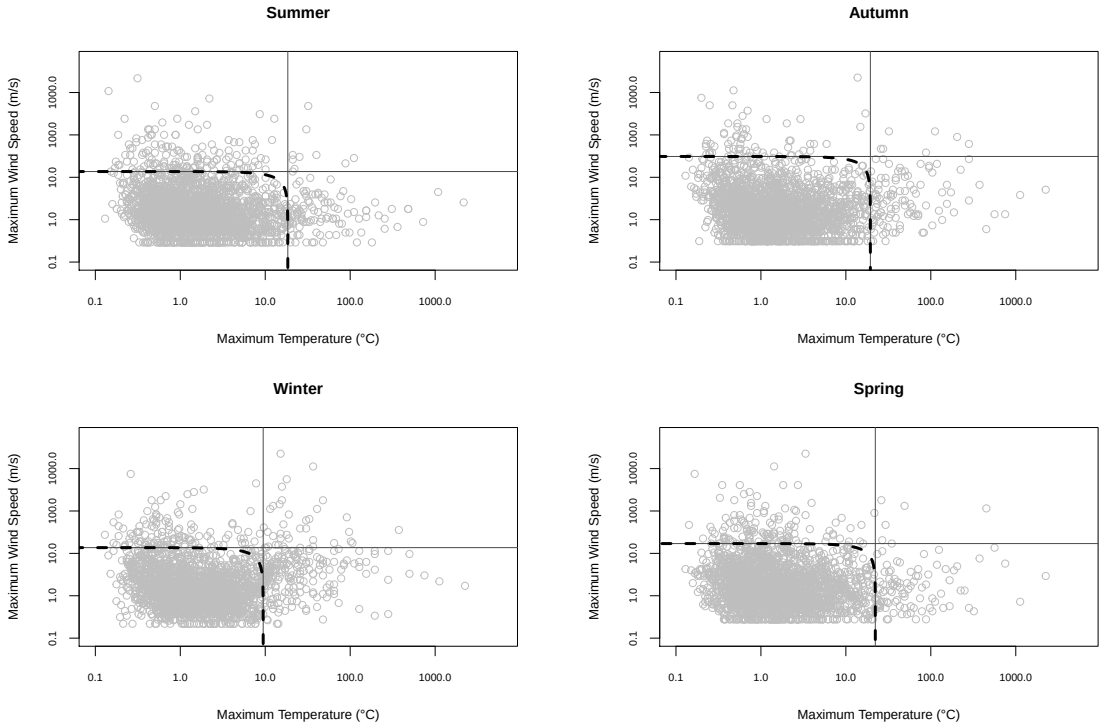


Figure 3.29: Maximum temperature and maximum wind speed broken up into seasons for Plettenberg Bay.

There are 9, 7 and 5 joint exceedances above the linear threshold for summer, autumn and spring with relatively more, i.e. 35, during winter as seen in Figure 3.29. By using the curved threshold, the maxima that are considered extreme in both and individual components can also be taken into the analysis.

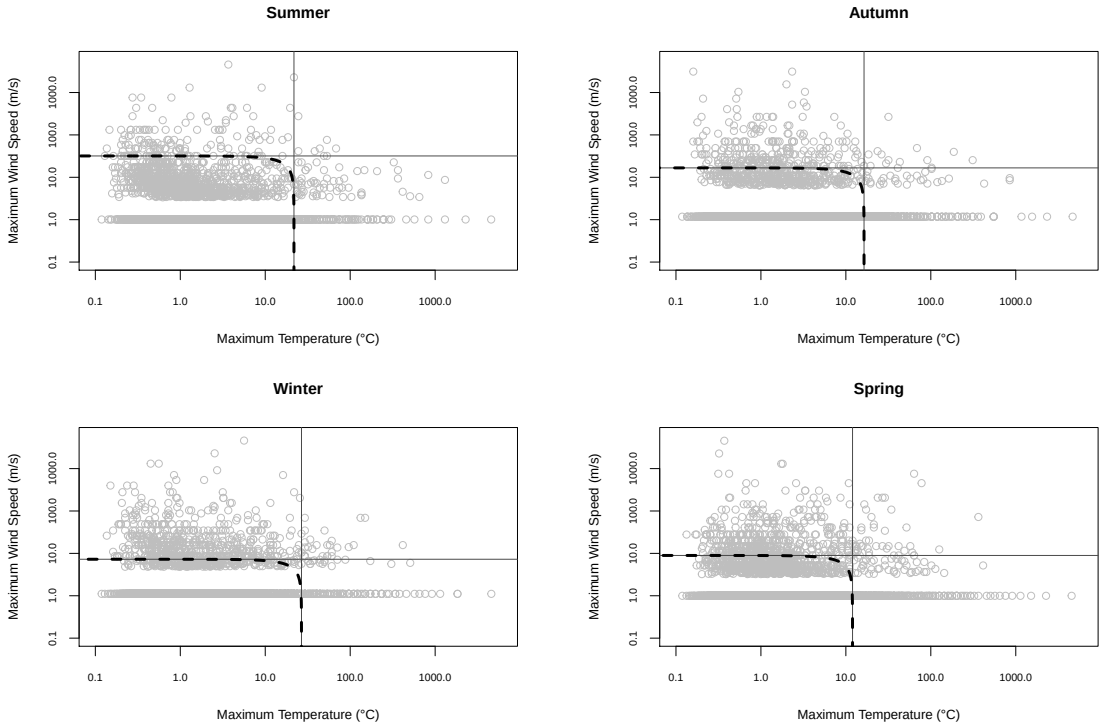


Figure 3.30: Maximum temperature and maximum wind speed broken up into seasons for Vredendal.

Figure 3.30 shows that there are 6 joint exceedances during summer, 13 during autumn, 28 during winter and 39 during spring for Vredendal. A better estimation of the occurrence of extreme events can be found using an increased number of observations which is provided by the curved threshold.

3.5.2 Threshold Excess

Parameter estimates using the threshold excess approach are shown in for Table 3.21 as well as negative log-likelihood (N-LL) which aids in models selection for nested and non-nested models, respectively. Standard errors are shown in parentheses.

Table 3.21: Parameter estimates of the threshold excess approach using the log model.

Station	Season	Wind Speed and Rainfall		Wind Speed and Temperature	
		α	N-LL	α	N-LL
CT Airport	Summer	0.9865 (0.009)	3010.68	0.9996 (2 x 10^{-6})	4136.17
	Autumn	0.9248 (0.018)	3660.60	0.9993 (2 x 10^{-6})	4406.24
	Winter	0.8772 (0.025)	2574.53	0.9992 (2 x 10^{-6})	3448.48
	Spring	0.9997 (2 x 10^{-6})	2484.75	0.9992 (2 x 10^{-6})	4681.41
George Airport	Summer	0.9738 (0.017)	2764.50	0.9995 (2 x 10^{-6})	2774.09
	Autumn	0.9713 (0.017)	2147.08	0.9948 (0.015)	2143.80
	Winter	0.9532 (0.019)	2733.62	0.9497 (0.019)	3340.43
	Spring	0.9667 (0.018)	2782.95	0.9947 (0.009)	2752.16
Plettenberg Bay	Summer	0.9661 (0.022)	1765.27	0.9991 (2 x 10^{-6})	2212.12
	Autumn	0.9687 (0.024)	1166.84	0.9698 (0.021)	1679.68
	Winter	0.9845 (0.021)	1659.23	0.9237 (0.022)	2953.31
	Spring	0.9997 (2 x 10^{-6})	1315.75	0.9957 (0.012)	1974.20
Vredendal	Summer	0.9996 (2 x 10^{-6})	2217.60	0.9991 (2 x 10^{-6})	3010.26
	Autumn	0.9262 (0.018)	3980.06	0.9992 (2 x 10^{-6})	4657.56
	Winter	0.9055 (0.021)	5114.58	0.9992 (2 x 10^{-6})	5682.05
	Spring	0.9721 (0.015)	4793.15	0.9939 (0.009)	6197.13

Rainfall and wind speed maxima parameter estimates show very weak dependence for the stations across all seasons with the exception of winter at Cape Town International Airport and winter at Vredendal under this threshold excess approach. While the estimates in these two instances have dependence estimates that are smaller compared to the other stations, the dependence between the variables is weak. Winter at these two stations showed more observations above the chosen threshold (top right quadrant) of the winter plot in Figures 3.23 and 3.26. The cases where $\hat{\alpha} \geq 0.9996$ demonstrates extremely weak dependence as in the respective plots for spring at Plettenberg Bay and Vredendal shown in Figures 3.25 and 3.26 there are very few observations above the thresholds. Attached to the dependence estimates of 0.9996 and above are extremely small standard errors (2 x 10^{-6}) which may indicate some biased estimates.

The dependence estimates for wind speed and temperature during all the seasons at Cape Town International Airport and 0.9992 and above with extremely small standard errors (2 x 10^{-6}) . Summer, winter and autumn at Vredendal as exhibit extremely weak dependence with less weak dependence during spring. Summer at George Airport and at Plettenberg Bay show the same result viz. very weak dependence. Overall, the α estimates for wind speed and temperature for the remaining seasons at George Airport and Plettenberg Bay show very weak dependence between variables.

Figures 3.23 - 3.30 provided a glimpse into the dependence of the structure between the variables across the seasons for the different stations and is confirmed by the parameter estimates in Table 3.21. These plots suggest that the threshold excess approach may not be an appropriate choice to model wind speed versus rainfall maxima and wind speed versus temperature maxima. The reasoning being that in many of the cases, there were very few observations in the interested quadrant and many more observations that are extreme in the top left and bottom right quadrants. The threshold excess approach used in this study is not able to handle the extreme observations that are not in the top

right quadrant and as a result, inference based on these observations cannot be made. In some instances the information matrix is near-singular when including other dependence models into the analysis and thus the estimates are unreliable. Hence, the analysis using threshold excess approach is fitted using only the logistic dependence model.

The point process approach on the other hand is able to handle scenarios where there is exceedance in one component only. Since a larger bounded region is used in the point process models (top left and bottom right quadrants), further analyses is performed using this approach with an extension to the negative logistic, bilogistic, negative bilogistic and dirichlet dependence models.

3.5.3 Point Process

Maximum Wind Speed and Maximum Rainfall

The models used here are the common and most used dependence models in MEVT studies. The parameter estimates with standard errors in parentheses, N-LL and AIC using the point process approach for wind speed and rainfall are shown in Tables 3.22 - 3.25. The - under the β heading indicates that this parameter does not form part of the logistic and negative logistic models.

Table 3.22: Point process dependence estimates for wind speed & rainfall at CT Airport.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7846 (0.010)	-	3523.79	7057.59
	Autumn	0.7426 (0.010)	-	4275.20	8560.40
	Winter	0.7025 (0.014)	-	3085.43	6180.87
	Spring	0.7831 (0.011)	-	2919.21	5848.41
Negative logistic	Summer	0.4657 (0.015)	-	3449.06	6908.11
	Autumn	0.546 (0.017)	-	4221.58	8453.16
	Winter	0.6337 (0.027)	-	3062.66	6135.31
	Spring	0.4641 (0.017)	-	2869.31	5748.62
Bilogistic	Summer	0.8429 (0.025)	0.7335 (0.027)	3521.21	7054.42
	Autumn	0.6838 (0.024)	0.7906 (0.018)	4270.78	8553.55
	Winter	0.7667 (0.026)	0.6381 (0.030)	3082.30	6176.60
	Spring	0.7014 (0.028)	0.8475 (0.019)	2912.81	5837.61
Negative bilogistic	Summer	1.4700 (0.186)	3.3200 (0.489)	3443.74	6899.48
	Autumn	2.2040 (0.207)	1.5040 (0.146)	4219.36	8450.72
	Winter	1.2930 (0.149)	1.9570 (0.244)	3060.99	6133.99
	Spring	2.9350 (0.348)	1.5620 (0.183)	2865.42	5742.85
Dirichlet	Summer	0.5233 (0.125)	0.171 (0.047)	3501.42	7014.84
	Autumn	0.3096 (0.036)	0.5568 (0.072)	4253.31	8518.62
	Winter	0.6737 (0.103)	0.3799 (0.060)	3072.34	6156.68
	Spring	0.1993 (0.030)	0.5187 (0.082)	2901.15	5814.30

Since all the dependence models are not nested, model selection based on the N-LL is not appropriate. Looking at the AIC values, the best fitting model (lowest AIC) for all the seasons is the negative bilogistic model followed by the negative logistic model. Complete dependence is reached as α and β tend to 0 while independence is reached when α and β tend to ∞. The estimates using the negative bilogistic model show that α and β tend further away from 0 than towards 0.

Table 3.23: Point process dependence estimates for wind speed & rainfall at George Airport.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7331 (0.012)	-	3314.46	6638.91
	Autumn	0.7512 (0.013)	-	2587.02	5184.05
	Winter	0.7427 (0.012)	-	3228.34	6466.68
	Spring	0.7430 (0.011)	-	3293.61	6597.23
Negative logistic	Summer	0.5664 (0.020)	-	3275.51	6561.02
	Autumn	0.5300 (0.022)	-	2553.70	5117.40
	Winter	0.5460 (0.020)	-	3187.52	6385.04
	Spring	0.5490 (0.020)	-	3249.57	6509.14
Bilogistic	Summer	0.8683 (0.016)	0.5829 (0.028)	3290.36	6592.71
	Autumn	0.8574 (0.020)	0.6298 (0.032)	2575.29	5162.58
	Winter	0.8169 (0.021)	0.6634 (0.027)	3221.83	6455.65
	Spring	0.8753 (0.018)	0.5900 (0.033)	3275.13	6562.27
Negative bilogistic	Summer	0.9851 (0.104)	3.4117 (0.414)	3258.21	6528.42
	Autumn	1.2060 (0.146)	3.0610 (0.414)	2546.21	5104.42
	Winter	1.3280 (0.146)	2.5690 (0.305)	3182.73	6377.47
	Spring	1.0180 (0.118)	3.5190 (0.461)	3234.44	6480.88
Dirichlet	Summer	0.9782 (0.157)	0.1758 (0.025)	3277.69	6567.37
	Autumn	0.7336 (0.124)	0.1946 (0.032)	2565.68	5143.36
	Winter	0.6366 (0.096)	0.2560 (0.038)	3208.21	6428.43
	Spring	0.9979 (0.193)	0.1610 (0.025)	3259.30	6530.61

Table 3.23 shows that the best fitting model is the negative bilogistic model followed by the negative logistic model based on AIC. Complete dependence is reached as α and β tend to 0 while independence is reached when α and β tend to ∞. The estimates using the negative bilogistic model show that α and β tend further away from 0 than towards 0 (perhaps with an exception during summer).

Table 3.24: Point process dependence estimates for wind speed & rainfall at Plettenberg Bay.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7392 (0.014)	-	2084.34	4178.67
	Autumn	0.7676 (0.017)	-	1392.52	2795.03
	Winter	0.7302 (0.015)	-	2014.21	4038.42
	Spring	0.7513 (0.017)	-	1610.78	3231.56
Negative logistic	Summer	0.5551 (0.025)	-	2056.33	4122.66
	Autumn	0.4956 (0.027)	-	1372.42	2754.84
	Winter	0.5743 (0.027)	-	1990.90	3991.79
	Spring	0.5286 (0.028)	-	1589.41	3188.81
Bilogistic	Summer	0.8668 (0.023)	0.5851 (0.043)	2072.77	4157.54
	Autumn	0.8185 (0.032)	0.7154 (0.038)	1391.35	2794.27
	Winter	0.8606 (0.023)	0.5912 (0.039)	2003.34	4018.67
	Spring	0.8686 (0.025)	0.6300 (0.039)	1602.50	3217.00
Negative bilogistic	Summer	1.0120 (0.146)	3.3390 (0.525)	2047.09	4106.17
	Autumn	1.6311 (0.273)	2.5269 (0.457)	1371.53	2755.05
	Winter	1.0100 (0.145)	3.2480 (0.534)	1982.50	3977.00
	Spring	1.1508 (0.177)	3.3831 (0.624)	1583.50	3179.01
Dirichlet	Summer	0.9991 (0.234)	0.1783 (0.033)	2062.56	4137.12
	Autumn	0.4697 (0.102)	0.2537 (0.058)	1385.70	2783.39
	Winter	0.9611 (0.215)	0.1863 (0.036)	1994.72	4001.44
	Spring	0.7793 (0.176)	0.1681 (0.039)	1596.09	3204.18

The best fitting model is the negative bilogistic model at Plettenberg Bay as seen in Table 3.24 with the exception of autumn where the negative logistic model shows an improved model fit over the negative bilogistic model.

Table 3.25: Point process dependence estimates for wind speed & rainfall at Vredendal.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7967 (0.011)	-	2614.97	5239.95
	Autumn	0.7507 (0.010)	-	4567.26	9144.52
	Winter	0.6878 (0.009)	-	6110.14	12230.27
	Spring	0.7691 (0.010)	-	3779.51	7569.02
Negative logistic	Summer	0.4393 (0.017)	-	2560.75	5131.51
	Autumn	0.5317 (0.016)	-	4495.34	9000.68
	Winter	0.6724 (0.018)	-	6048.20	12106.39
	Spring	0.4932 (0.016)	-	3715.45	7438.89
Bilogistic	Summer	0.8041 (0.026)	0.7905 (0.024)	2614.93	5241.87
	Autumn	0.8405 (0.021)	0.6635 (0.027)	4557.78	9127.56
	Winter	0.9085 (0.005)	0.3657 (0.024)	5979.69	11971.38
	Spring	0.7436 (0.022)	0.7977 (0.021)	3778.57	7569.13
Negative bilogistic	Summer	1.9903 (0.264)	2.6369 (0.389)	2560.19	5132.38
	Autumn	1.2090 (0.127)	3.0840 (0.366)	4484.90	8981.81
	Winter	0.5023 (0.046)	5.0275 (0.422)	5955.16	11922.33
	Spring	1.6260 (0.180)	2.5760 (0.318)	3712.25	7436.50
Dirichlet	Summer	0.3137 (0.052)	0.2645 (0.050)	2603.69	5219.37
	Autumn	0.7545 (0.135)	0.1897 (0.032)	4532.03	9076.05
	Winter	2.8185 (0.538)	0.1444 (0.011)	5970.79	11953.59
	Spring	0.4355 (0.063)	0.2704 (0.043)	3760.36	7532.71

Table 3.25 shows that the best fitting model for summer is the negative logistic model while for autumn, winter and spring, the negative bilogistic model provides the best fit.

Maximum Wind Speed and Maximum Temperature

The parameter estimates with standard errors in parentheses, N-LL and AIC using the point process approach for wind speed and rainfall are shown in Tables 3.26 - 3.29.

Table 3.26: Point process dependence estimates for wind speed & temperature at CT Airport.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7839 (0.008)	-	4766.11	9542.22
	Autumn	0.7796 (0.008)	-	5097.06	10204.12
	Winter	0.8013 (0.008)	-	3946.61	7903.22
	Spring	0.7606 (0.008)	-	5364.72	10739.45
Negative logistic	Summer	0.4668 (0.012)	-	4658.85	9327.69
	Autumn	0.4754 (0.012)	-	4983.38	9976.75
	Winter	0.4327 (0.012)	-	3843.74	7697.48
	Spring	0.5146 (0.013)	-	5256.79	10523.57
Bilogistic	Summer	0.7036 (0.025)	0.8665 (0.018)	4756.96	9525.92
	Autumn	0.5581 (0.041)	0.9283 (0.012)	5062.95	10137.91
	Winter	0.6865 (0.031)	0.8931 (0.017)	3933.35	7878.71
	Spring	0.3466 (0.025)	0.9583 (0.003)	5174.22	10360.44
Negative bilogistic	Summer	3.2630 (0.355)	1.4860 (0.138)	4650.22	9312.43
	Autumn	4.2220 (0.472)	1.1310 (0.111)	4959.51	9931.03
	Winter	3.6650 (0.423)	1.4900 (0.157)	3834.29	7680.58
	Spring	3.3070 (0.305)	1.1490 (0.095)	5139.81	10291.63
Dirichlet	Summer	0.1613 (0.027)	0.5507 (0.082)	4729.44	9470.87
	Autumn	0.0890 (0.013)	1.2376 (0.289)	5031.13	10074.25
	Winter	0.1215 (0.022)	0.6315 (0.118)	3910.62	7833.24
	Spring	0.0584 (0.004)	4.1089 (1.161)	5165.81	10343.63

The negative bilogistic model provides the best fitting model across the all the seasons for wind speed and temperature as shown in Table 3.26.

Table 3.27: Point process dependence estimates for wind speed & temperature at George Airport.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7960 (0.009)	-	3152.43	6314.85
	Autumn	0.7960 (0.011)	-	2474.33	4958.66
	Winter	0.7588 (0.010)	-	3790.69	7591.38
	Spring	0.7856 (0.010)	-	3156.45	6322.90
Negative logistic	Summer	0.4409 (0.013)	-	3075.98	6161.96
	Autumn	0.4364 (0.017)	-	2427.16	4864.32
	Winter	0.5114 (0.016)	-	3736.47	7482.94
	Spring	0.4611 (0.015)	-	3088.43	6186.87
Bilogistic	Summer	0.9485 (0.008)	0.5935 (0.034)	3118.23	6248.47
	Autumn	0.8422 (0.023)	0.7625 (0.022)	2472.33	4956.66
	Winter	0.7090 (0.023)	0.7992 (0.017)	3787.29	7596.58
	Spring	0.9310 (0.012)	0.6073 (0.035)	3130.95	6273.90
Negative bilogistic	Summer	1.2038 (0.128)	5.0764 (0.701)	3055.68	6123.36
	Autumn	1.8500 (0.203)	3.0080 (0.424)	2425.08	4862.16
	Winter	2.3920 (0.224)	1.5730 (0.153)	3733.75	7479.50
	Spring	1.2130 (0.133)	4.4400 (0.610)	3072.45	6156.90
Dirichlet	Summer	0.8867 (0.160)	0.0652 (0.011)	3103.42	6218.83
	Autumn	0.3588 (0.048)	0.2035 (0.039)	2462.93	4937.86
	Winter	0.2824 (0.032)	0.4969 (0.065)	3771.11	7554.21
	Spring	0.8253 (0.148)	0.0889 (0.016)	3115.81	6243.63

The best fitting dependence model is the negative bilogistic model for summer, autumn, winter and spring at George Airport.

Table 3.28: Point process dependence estimates for wind speed & temperature at Plettenberg Bay.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7642 (0.012)	-	2541.39	5092.77
	Autumn	0.7653 (0.014)	-	1967.43	3944.86
	Winter	0.7098 (0.012)	-	3359.54	6729.08
	Spring	0.7753 (0.012)	-	2285.05	4580.10
Negative logistic	Summer	0.5040 (0.019)	-	2496.88	5003.76
	Autumn	0.4974 (0.022)	-	1940.03	3890.07
	Winter	0.6136 (0.022)	-	3327.73	6665.46
	Spring	0.4817 (0.019)	-	2240.31	4490.61
Bilogistic	Summer	0.8020 (0.025)	0.7290 (0.026)	2540.09	5092.19
	Autumn	0.6064 (0.035)	0.8711 (0.017)	1951.82	3915.65
	Winter	0.5653 (0.028)	0.7977 (0.015)	3341.06	6694.11
	Spring	0.796 (0.026)	0.7559 (0.027)	2284.70	4581.39
Negative bilogistic	Summer	1.7800 (0.204)	2.2360 (0.282)	2496.39	5004.77
	Autumn	3.6070 (0.492)	1.0800 (0.144)	1928.46	3868.91
	Winter	2.2480 (0.220)	1.0020 (0.103)	3315.96	6643.73
	Spring	1.9310 (0.246)	2.2430 (0.306)	2240.12	4492.24
Dirichlet	Summer	0.4267 (0.063)	0.2921 (0.049)	2527.49	5066.97
	Autumn	0.1665 (0.027)	0.8604 (0.173)	1944.61	3901.22
	Winter	0.3033 (0.031)	0.9771 (0.145)	3330.44	6672.88
	Spring	0.3727 (0.061)	0.2969 (0.053)	2273.11	4558.21

The dependence structure during summer and spring at Plettenberg Bay can best be described by the negative logistic model. Autumn and winter is better described by the negative bilogistic model as seen in Table 3.28.

Table 3.29: Point process dependence estimates for wind speed & temperature at Vredendal.

Model	Season	α	β	N-LL	AIC
Logistic	Summer	0.7830 (0.010)	-	3529.36	7068.72
	Autumn	0.7606 (0.009)	-	5436.53	10875.07
	Winter	0.7296 (0.008)	-	6623.21	13256.41
	Spring	0.7373 (0.007)	-	7068.82	14147.65
Negative logistic	Summer	0.4643 (0.015)	-	3466.16	6942.32
	Autumn	0.5135 (0.013)	-	5333.48	10676.97
	Winter	0.5751 (0.014)	-	6536.14	13082.28
	Spring	0.5624 (0.013)	-	6946.75	13909.49
Bilogistic	Summer	0.6323 (0.032)	0.8970 (0.014)	3507.74	7027.48
	Autumn	0.7262 (0.021)	0.7909 (0.018)	5430.81	10873.62
	Winter	0.8821 (0.010)	0.5768 (0.018)	6561.75	13135.50
	Spring	0.7795 (0.016)	0.6956 (0.018)	7065.13	14142.26
Negative bilogistic	Summer	4.1683 (0.515)	1.1467 (0.128)	3447.80	6907.61
	Autumn	2.4010 (0.220)	1.5650 (0.147)	5330.51	10673.02
	Winter	0.9850 (0.066)	3.5360 (0.291)	6494.73	13001.46
	Spring	1.5710 (0.118)	2.0290 (0.162)	6948.26	13908.52
Dirichlet	Summer	0.1191 (0.020)	0.8537 (0.169)	3490.38	6992.77
	Autumn	0.2851 (0.034)	0.4952 (0.066)	5399.31	10810.63
	Winter	0.9707 (0.098)	0.1598 (0.015)	6534.97	13081.93
	Spring	0.5254 (0.052)	0.3418 (0.035)	7021.96	14055.91

Table 3.29 shows that out of the 5 fitted models, the negative bilogistic model best describes the dependence structure between wind speed and temperature at Vredendal.

Tables 3.22 - 3.29 show that in most instances, the negative bilogistic fit best describes the dependence structure between wind speed and rainfall and wind speed and temperature extreme maxima. If it is not the negative bilogisitic model then the best fitting model is negative logistic model. All of the five different dependence models relay similar information about the dependence structure.

The use of radial and angular components allow for analysis and inference to be made on a greater number of observations because of the curved threshold. The point process approach also has the ability to simultaneously model exceedances in both and one margin which is a pattern seen in the bivariate plots in Figures 3.23 - 3.30. While there were not many joint exceedances at certain stations during specific seasons, there were substantially more exceedances individually from each of the variables. There is a negative associated asymptotic independence between the variables especially between wind speed and temperature maxima where there 0 joint exceedances.

Chapter 4

Concluding Remarks

The same thresholds are used across the different approaches in both the univariate and multivariate analyses to allow for comparability between the models. In the univariate section, Tables 3.6 and 3.14 show that the shape parameters with corresponding intervals for rainfall under the threshold excess and point process approaches are the same. This is not an unexpected result as the shape parameters provide information about the tails of the distributions and both approaches are relaying the same details about the extremes of the maxima. Furthermore, the scale parameter estimates for rainfall under the threshold excess approach are smaller than those under the point process approach when the shape parameter estimates are greater than zero. The exception in this case is when the estimate for the shape parameter is less than zero i.e. during winter at Cape Town International Airport, summer at Langebaanweg and autumn at Plettenberg Bay.

Similarly, the threshold excess estimates for the shape parameter (and confidence intervals) for temperature and wind speed extreme maxima match to the corresponding estimates from the point process approach as information about the tails is provided by the shape parameter. The maxima extremes of the temperature data follow a Weibull distribution. The extreme of the maxima for wind speed also follow a Weibull distribution with the exception of winter at Vredendal which is described by a Fr'echet distribution. For temperature maxima, the scale parameter estimates under the point process approach in Table 3.16 are smaller than those under the threshold excess approach in Table 3.9 for all stations across all seasons. Comparing wind speed scale estimates, Tables 3.12 and 3.18 show that the estimates are larger for threshold excess approach than the point process approach except for when the shape parameter estimate is positive i.e. winter at Plettenberg Bay. Overall, the threshold excess and point process approaches provide the same estimates for ξ with the same confidence intervals. Smaller estimates for σ are obtained for the threshold excess approach when the shape parameter estimates are greater than zero for all the variables, unless the shape parameter estimates are less than zero.

The return levels tend to be overestimated when the threshold excess models are used, while underestimated when using point process models for maximum rainfall, maximum temperature and maximum wind speed at the five stations. The probabilities of experiencing extreme levels of rainfall, temperature and wind speed at the five stations

within the next 20 years are very small. The confidence intervals with the point process approach are also slightly narrower compared to the threshold excess approach. Based on the shape estimates and return levels, there is no compelling reason to favour the threshold excess approach over the point process approach and vice versa.

The quantile-quantile plots show the same fit of the models for each season and stations using both the threshold excess and point process approaches. Wind speed at Vredendal showed many deviations from the linearity suggesting that the models from the two approaches did not fit the data well. Moreover, there may be a problem of over-fitting the data at certain stations where there are chunks of missing data. In such cases, it would be worth investigating covariates in the parameters in future work.

For the bivariate analysis, comparing the component-wise dependence estimates for wind speed and rainfall in Table 3.20 to that of the threshold excess approach in Table 3.21, there are differences in the results. These differences can attributed to the varying number of observations used for each method. With the exception of autumn at Cape Town International Airport and winter at Vredendal, the component-wise approach uses more observations than the threshold excess approach. The near independence estimates with extremely small standard errors for the component-wise approach could be attributed to the maxima occurring far apart (in terms of time) from each other or even asymmetry in the data. The near independence structure coupled with extremely small standard errors for threshold excess models of wind speed and rainfall are seen when there are less than 5 observations in the interested region.

In the same tables, the estimates for wind speed and temperature again show varying dependence strength between the variables for the weather stations. This can be attributed to the difference in number of observations used in the models under the two approaches. The component-wise approach has more observations compared to the threshold excess approach for the stations with the only exception being winter at Plettenberg Bay. The standard errors of the near independence estimates for the threshold excess and component-wise fitted models are the same for summer, autumn and winter at Cape Town International Airport, summer at George Airport and Plettenberg Bay and all seasons at Vredendal.

While the thresholds chosen in the univariate case were deemed sufficient for each variable, the bivariate case of the threshold excess approach only handles situations where both components are extreme. Thus, leading to the small number of observations in the interested quadrant R_{11} as shown in Figures 3.23 - 3.30. The component-wise and threshold excess approaches provide a trade-off between the uncertainty in time-structure with a sufficient amount of observations versus certainty in time-structure and fewer observations, respectively. An issue with the component-wise maxima stems from how the dataset is formed for this approach. Since the maxima is partitioned into blocks for each variable, it can occur that corresponding maxima values for the variables do not occur simultaneously. For instance, when looking at daily maximum wind speed together with daily maximum rainfall, the wind speed maxima for that block may have occurred in June while the maximum rainfall may have occurred in August.

Comparison of the dependence estimates using the logistic model with the point process in Tables 3.22 - 3.29 to the component-wise and threshold excess approaches in Tables 3.20 and 3.21, respectively, shows a noticeable change. The estimates for point process logistic models show α tends further away from 1 compared to those seen in the other two approaches. Weak dependence is still exhibited even though these point process estimates tend further away from 1. A reason for the substantial difference in the dependence estimates in the point process models can be attributed to the increased number of observations used in the analysis. Similarly, the negative associated asymptotic independence between the variables displayed in Figures 3.23 - 3.30 is better captured by the point process approach compared to the other two approaches. The standard errors of $\hat{\alpha}$ are small and are below 0.020 for both wind speed versus rainfall and wind speed versus temperature.

Furthermore, the point process approach is able to adequately capture the dependence structure between wind speed versus rainfall and wind speed versus temperature by extending to asymmetric models. The point process approach found that the negative bilogistic model fits that data the best in most seasons at the weather stations compared to the logistic, bilogistic, negative logistic and dirichlet models as there is asymmetry in the maxima.

To summarise, the component-wise and threshold excess models do not perform well in capturing the dependence between the weather extremes. The point process models are better performing models for jointly capturing the relationship between the maxima of different weather variables at a single location. In terms of estimations of parameters, there will always be uncertainty surrounding the models and results that contain some degree of inaccuracy. However, there are some merits to using the threshold excess and point process approaches over the component-wise approach.

The results of the MEVT approaches are in agreement with the existing studies on climate data. In particular, the component-wise maxima approach should be used in cases where the dataset consists only of maxima. The exclusion of vast amounts of data does not provide an accurate representation of the reality of the process of extreme weather events. The threshold excess approach provided an improvement in the component-wise approach as indicated by the literature. However, for the weather variables used at the five stations, there is a further improvement on the estimates using the point process approach which may differ slightly from some of the studies examined in the literature review. The component-wise approach reduced a series of approximately 4500 observations per season per station to only 51 pairwise observations. This is definitely not an efficient use of data with estimates that indicate near independence between the variables. Further benefits are seen when using the point process approach over the threshold excess and component-wise approaches. For instance, summer at Cape Town International Airport which has a total of 128 exceedances for wind speed and 183 for rainfall, reduces to just 5 joint exceedances modelled using the threshold excess approach. The 5 excesses compared to the 311 exceedances modelled using the point process models, provides better insights and estimates of the occurrence of extreme events.

The threshold excess models may provide unreliable estimates in cases where there are extremely few joint exceedances. In cases, where there are no joint exceedances i.e. winter at Cape Town International Airport and summer at George Airport, excluding these from the analysis would be appropriate. The bivariate extreme value distributions are not suitable in these two cases. In such cases, where there is asymptotic independence within the data, applying asymptotic dependent models can produce inaccurate results and should be excluded from the analysis or instead asymptotic independent models should be used.

What is more, the threshold excess and point process approaches both come with subjectivity in that there is no concrete way to choose a suitable threshold for the models. Therefore, there could be disagreements in terms of the chosen suitable threshold. However, the uncertainty associated with threshold choice is a reasonable trade-off to favour compared to modelling maxima that have not occurred within a reasonable period of each other as seen with the component-wise maxima.

Moreover, the benefits of the bivariate analysis over the univariate case are seen when the joint region R_{11} of the threshold excess and point process approaches. Although, in some cases, there were not many joint exceedances, it is important to note that the possibility of observing those extreme maxima does exist. In the same sense, an extension to an analysis of more than 2 variables may provide further improvement on model fit and dependence structure.

Future work includes an extension to three or more variables in the models or even analysing one variable at different stations with a spatial component added into the analysis. For instance, stations such as George Airport and Plettenberg Bay showed higher levels of rainfall compared to the other stations which may indicate that these two stations experience more of the climate from the neighbouring province and not the Mediterranean climate experienced by the Western Cape. Thus, adding in a spatial component may enhance the value of the estimates and models as it can incorporate more of the conditions surrounding an extreme event.

Additionally, utilising the wind direction data can be useful for inference especially since this type of data is readily available with extracting wind speed data. Regarding all three approaches, there is increased complexity as the number of variables increase which makes the likelihood more difficult to compute. The use of copulas may also prove to be a valuable method of estimating extremes.

Appendix A

A.1 Map of the location of the weather stations in Western Cape province

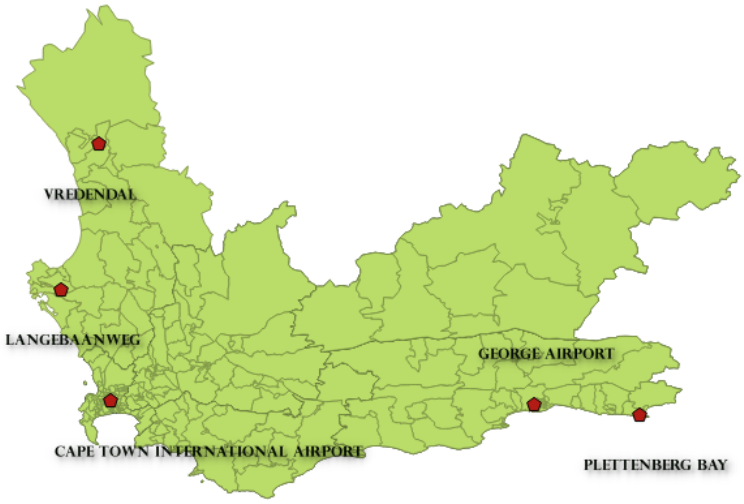


Figure A.1: Map of the five weather stations across the Western Cape province, South Africa.

A.2 Exploratory Plots

A.2.1 Maximum Rainfall

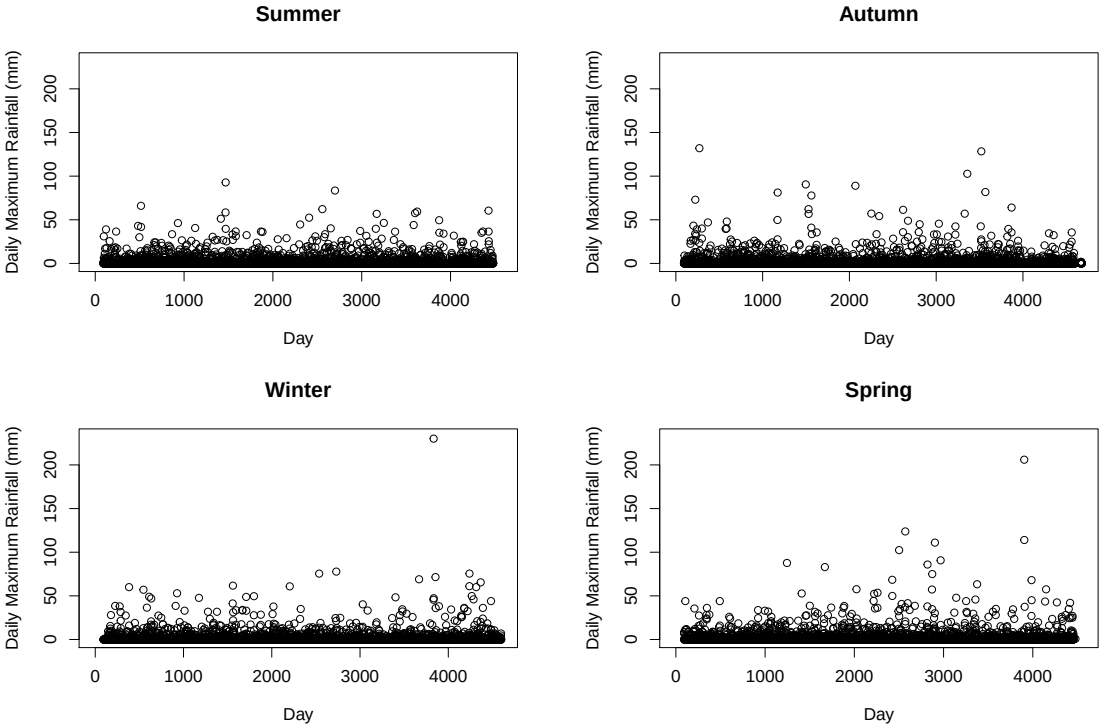


Figure A.2: Daily rainfall broken up into seasons for George Airport

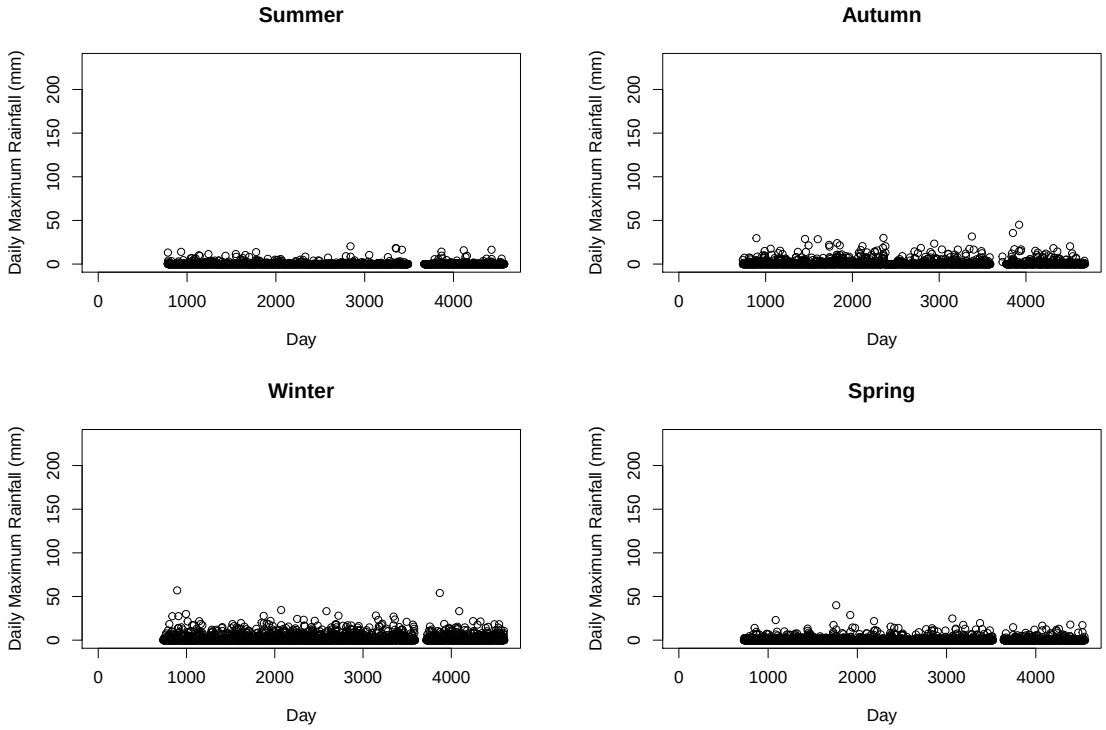


Figure A.3: Daily rainfall broken up into seasons for Langebaanweg

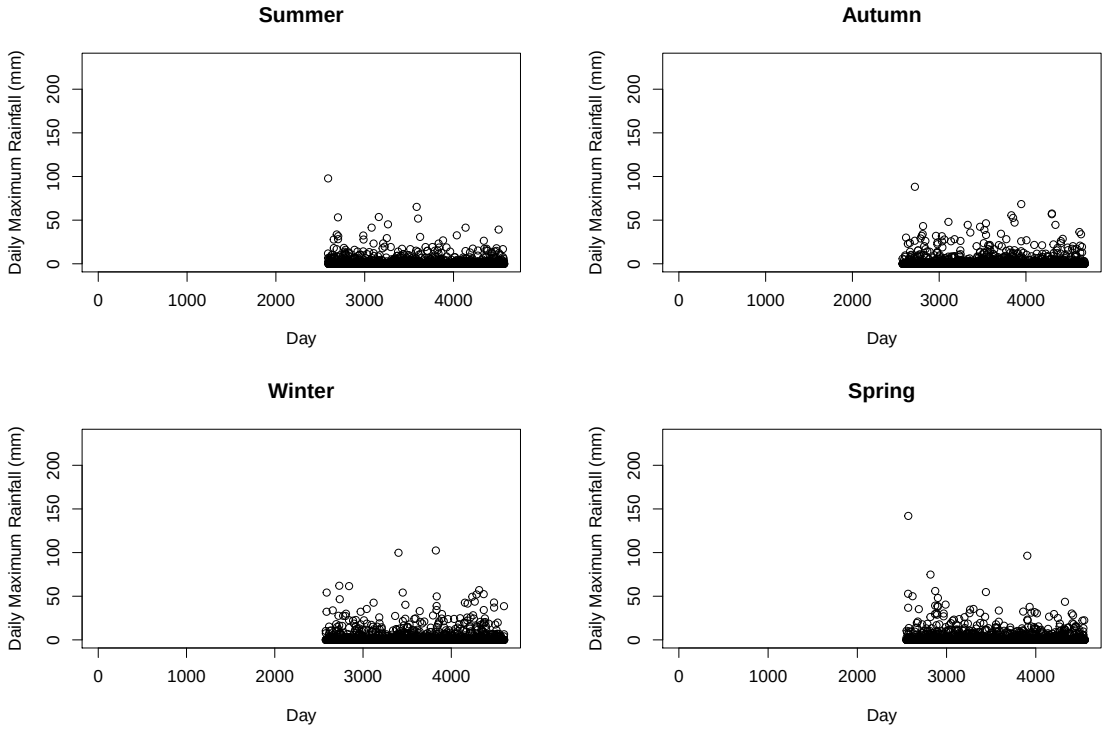


Figure A.4: Daily rainfall broken up into seasons for Plettenberg Bay

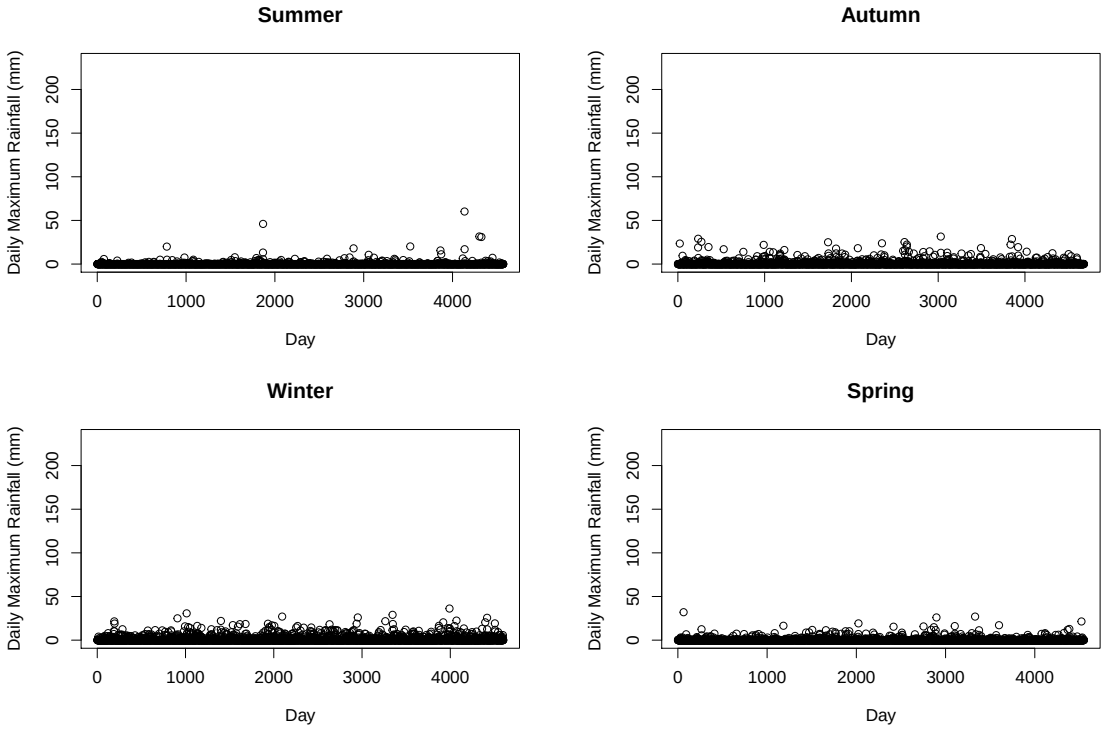


Figure A.5: Daily rainfall broken up into seasons for Vredendal

A.2.2 Maximum Temperature

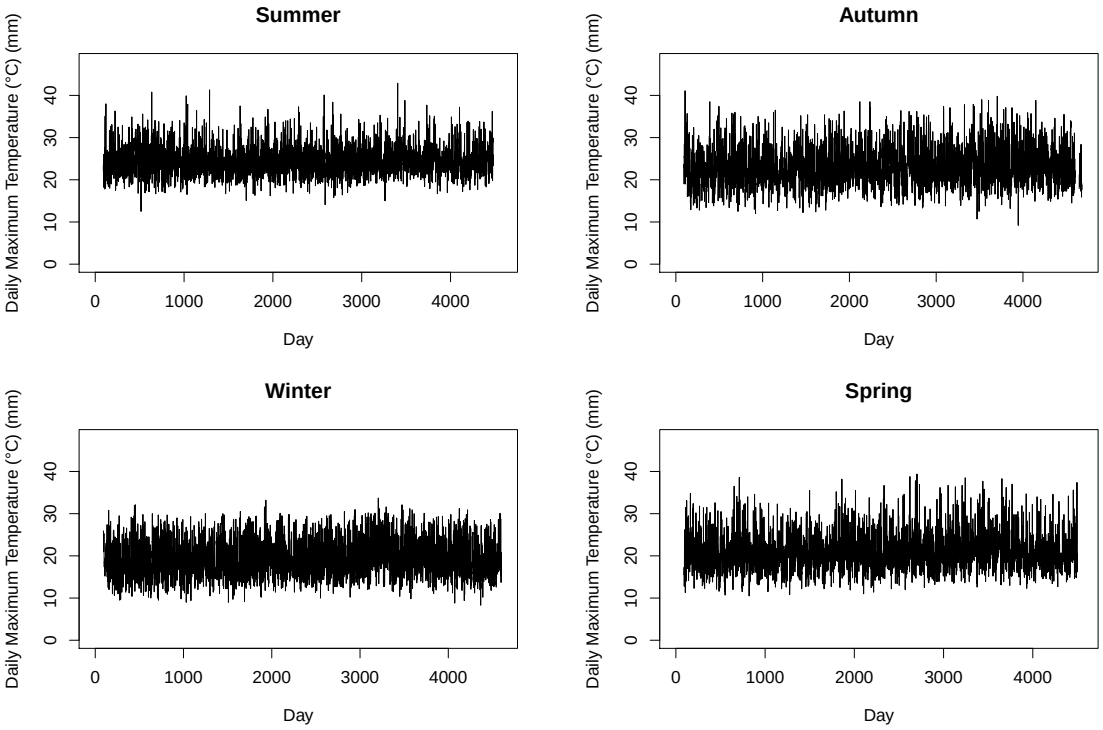


Figure A.6: Daily temperature broken up into seasons for George Airport

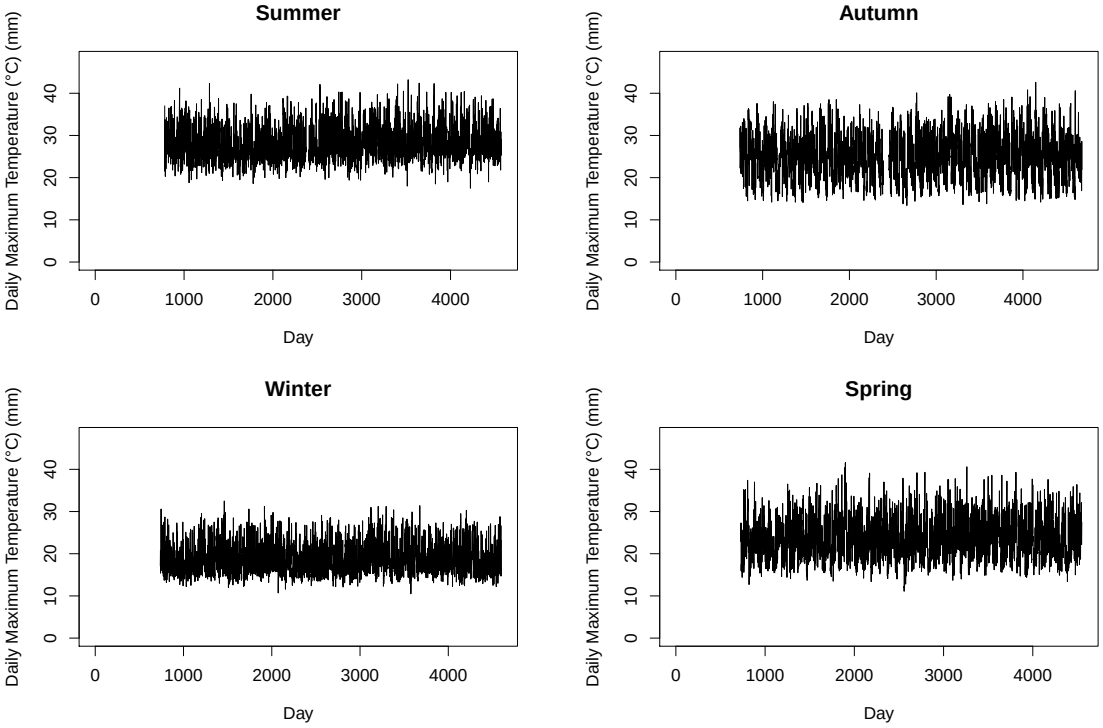


Figure A.7: Daily temperature broken up into seasons for Langebaanweg

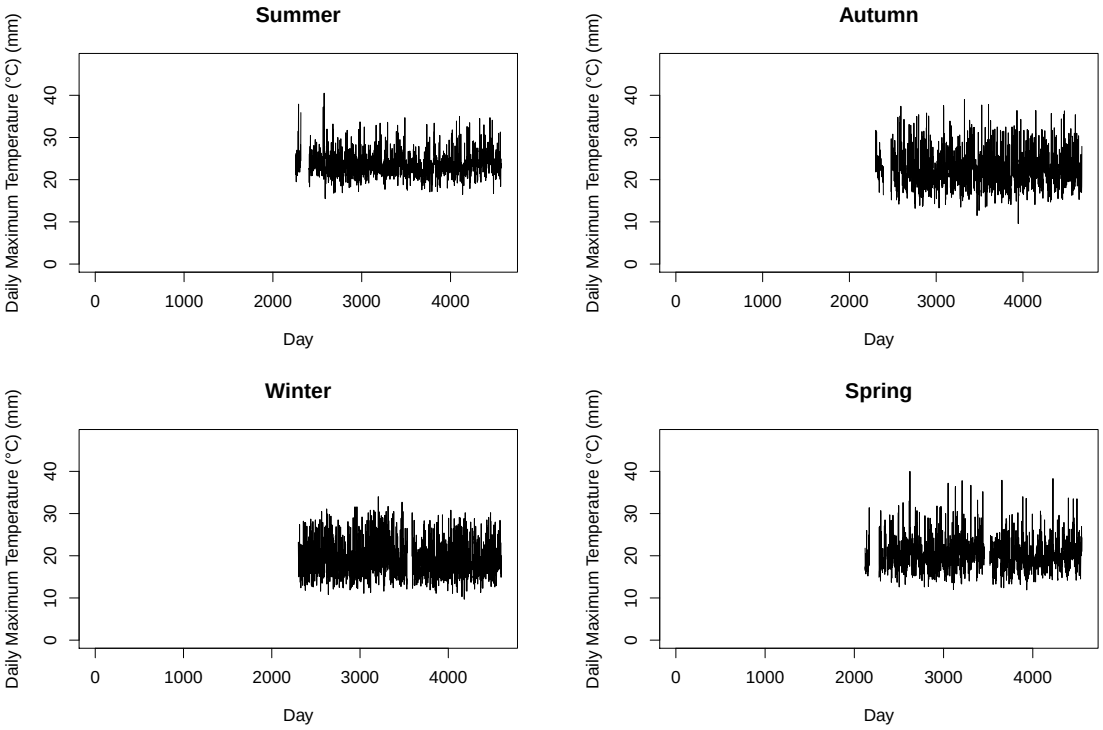


Figure A.8: Daily temperature broken up into seasons for Plettenberg Bay

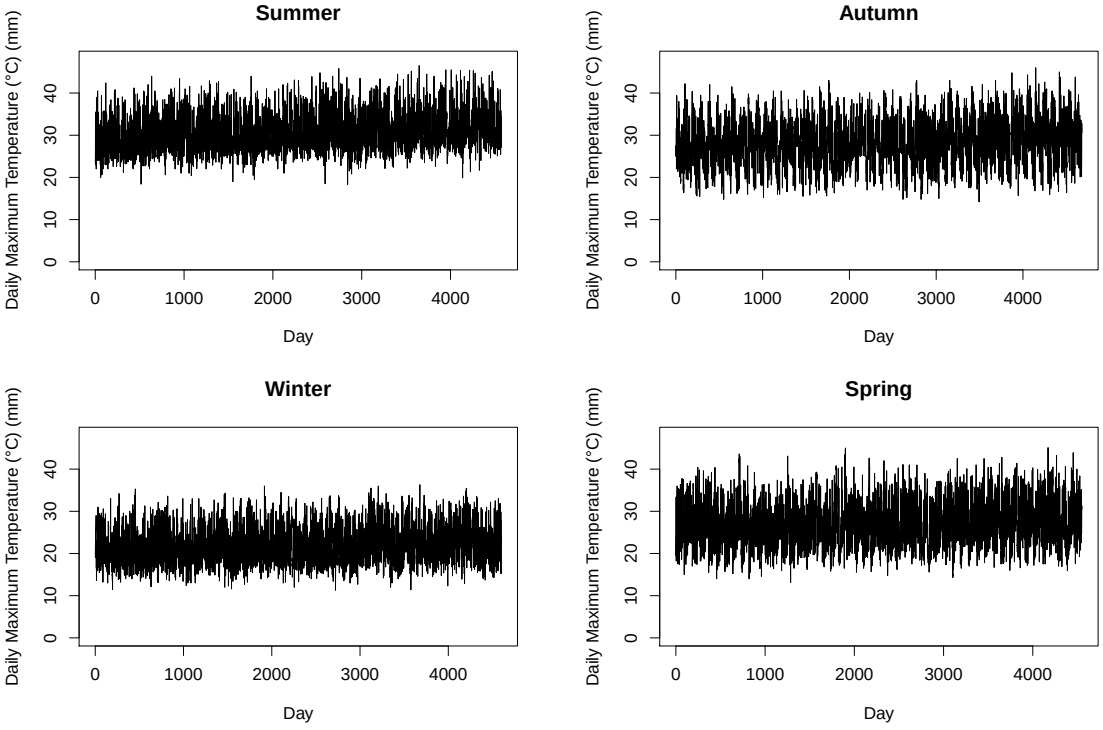


Figure A.9: Daily temperature broken up into seasons for Vredendal

A.2.3 Maximum Wind Speed

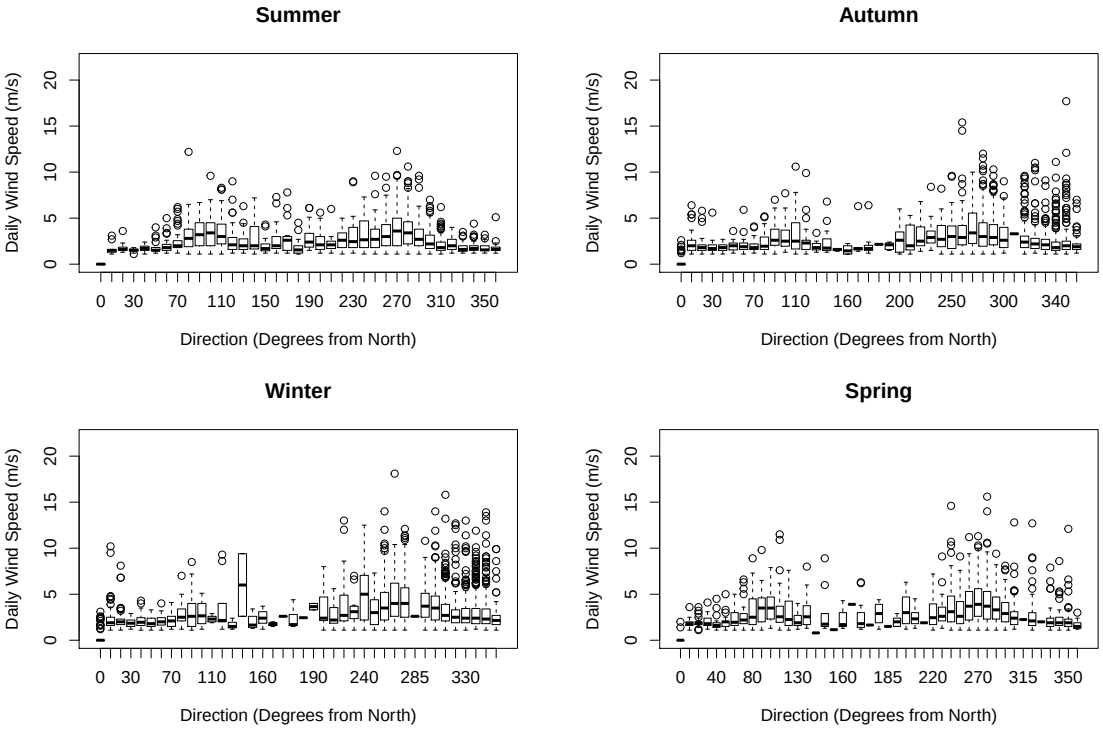


Figure A.10: Daily wind speed broken up into seasons for George Airport

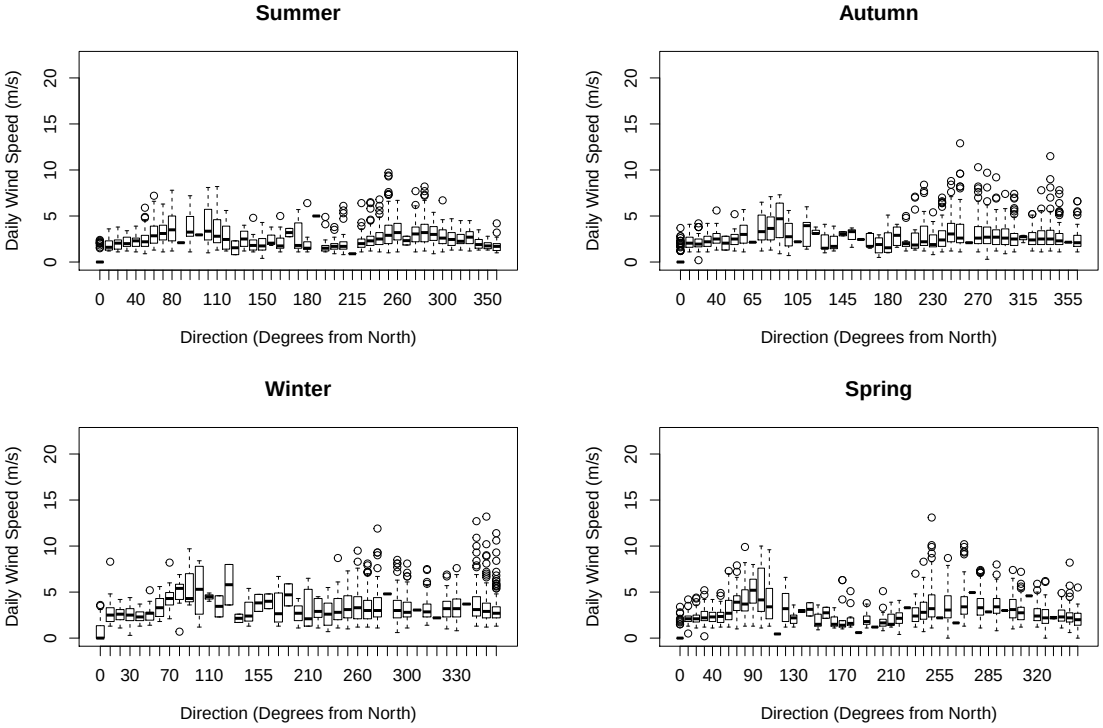


Figure A.11: Daily wind speed broken up into seasons for Plettenberg Bay

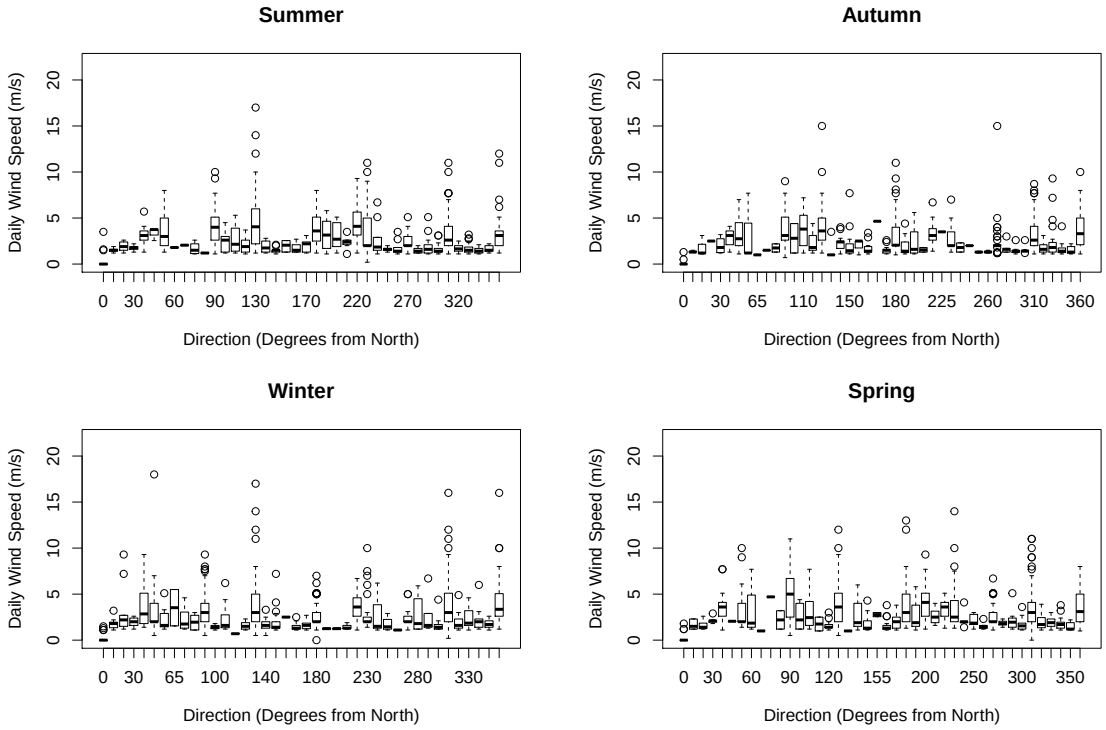


Figure A.12: Daily wind speed broken up into seasons for Vredendal

A.3 Stationarity tests

Table A.1: Phillips-Perron Test: Rainfall

Station	Season	Test Statistic	P-value
CT Airport	Summer	-57.561	<< 0.01
	Autumn	-55.376	<< 0.01
	Winter	-55.634	<< 0.01
	Spring	-56.237	<< 0.01
George Airport	Summer	-54.167	<< 0.01
	Autumn	-53.328	<< 0.01
	Winter	-50.490	<< 0.01
	Spring	-50.636	<< 0.01
Langebaan weg	Summer	-51.937	<< 0.01
	Autumn	-51.522	<< 0.01
	Winter	-50.641	<< 0.01
	Spring	-48.246	<< 0.01
Plettenberg Bay	Summer	-38.986	<< 0.01
	Autumn	-36.410	<< 0.01
	Winter	-36.218	<< 0.01
	Spring	-34.997	<< 0.01
Vredendal	Summer	-55.752	<< 0.01
	Autumn	-53.605	<< 0.01
	Winter	-55.265	<< 0.01
	Spring	-59.890	<< 0.01

Table A.2: Phillips-Perron Test: Maximum Temperature

Station	Season	Test Statistic	P-value
CT Airport	Summer	-42.660	<< 0.01
	Autumn	-34.364	<< 0.01
	Winter	-37.957	<< 0.01
	Spring	-37.971	<< 0.01
George Airport	Summer	-50.256	<< 0.01
	Autumn	-42.576	<< 0.01
	Winter	-43.300	<< 0.01
	Spring	-45.915	<< 0.01
Langebaan weg	Summer	-32.505	<< 0.01
	Autumn	-27.179	<< 0.01
	Winter	-32.869	<< 0.01
	Spring	-33.260	<< 0.01
Plettenberg Bay	Summer	-31.752	<< 0.01
	Autumn	-29.295	<< 0.01
	Winter	-30.943	<< 0.01
	Spring	-30.910	<< 0.01
Vredendal	Summer	-39.509	<< 0.01
	Autumn	-32.148	<< 0.01
	Winter	-37.435	<< 0.01
	Spring	-39.449	<< 0.01

Table A.3: Phillips-Perron Test: Wind Speed

Station	Season	Test Statistic	P-value
CT Airport	Summer	-60.695	<< 0.01
	Autumn	-60.800	<< 0.01
	Winter	-55.156	<< 0.01
	Spring	-62.657	<< 0.01
George Airport	Summer	-53.028	<< 0.01
	Autumn	-49.306	<< 0.01
	Winter	-51.152	<< 0.01
	Spring	-54.364	<< 0.01
Plettenberg Bay	Summer	-42.423	<< 0.01
	Autumn	-41.732	<< 0.01
	Winter	-43.618	<< 0.01
	Spring	-45.789	<< 0.01
Vredendal	Summer	-58.908	<< 0.01
	Autumn	-58.577	<< 0.01
	Winter	-59.792	<< 0.01
	Spring	-59.677	<< 0.01

A.4 MRL Plots

A.4.1 Maximum Rainfall

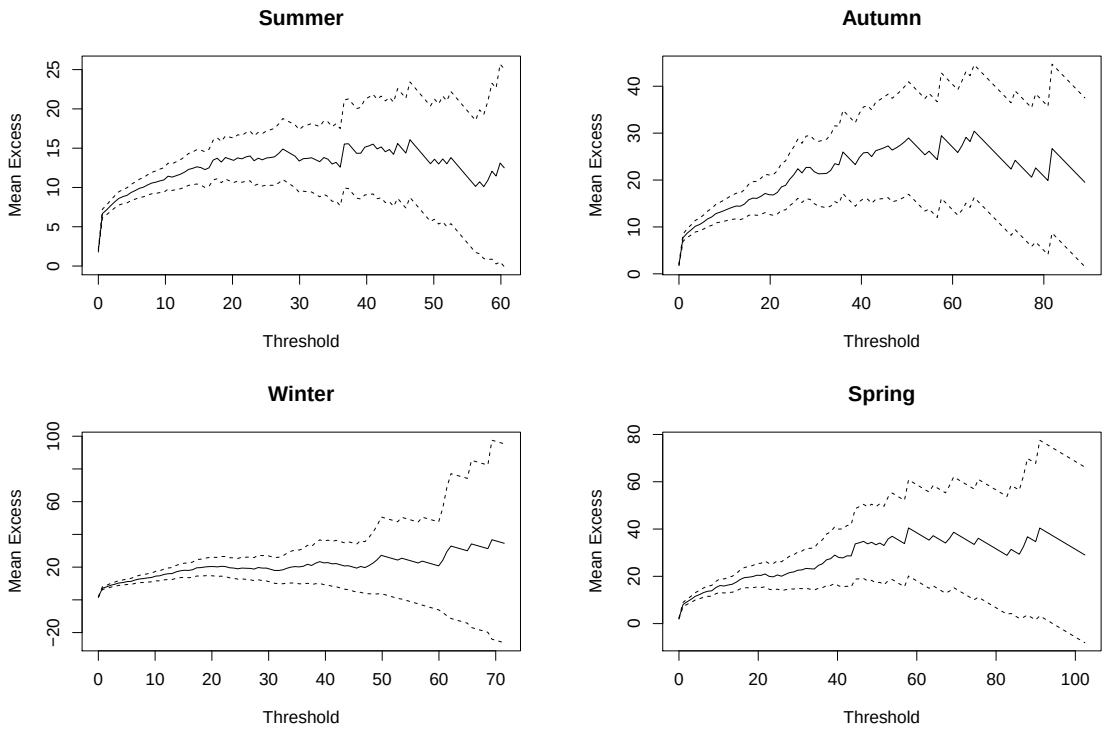


Figure A.13: Mean residual life plots for rainfall at George Airport

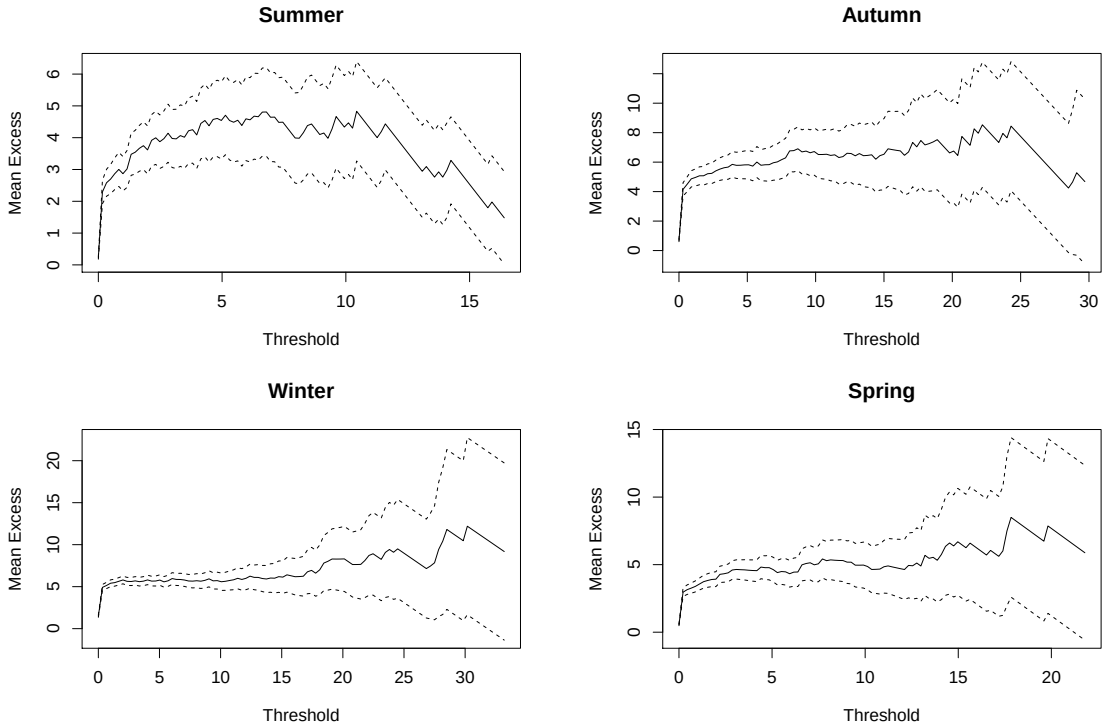


Figure A.14: Mean residual life plots for rainfall at Langebaanweg

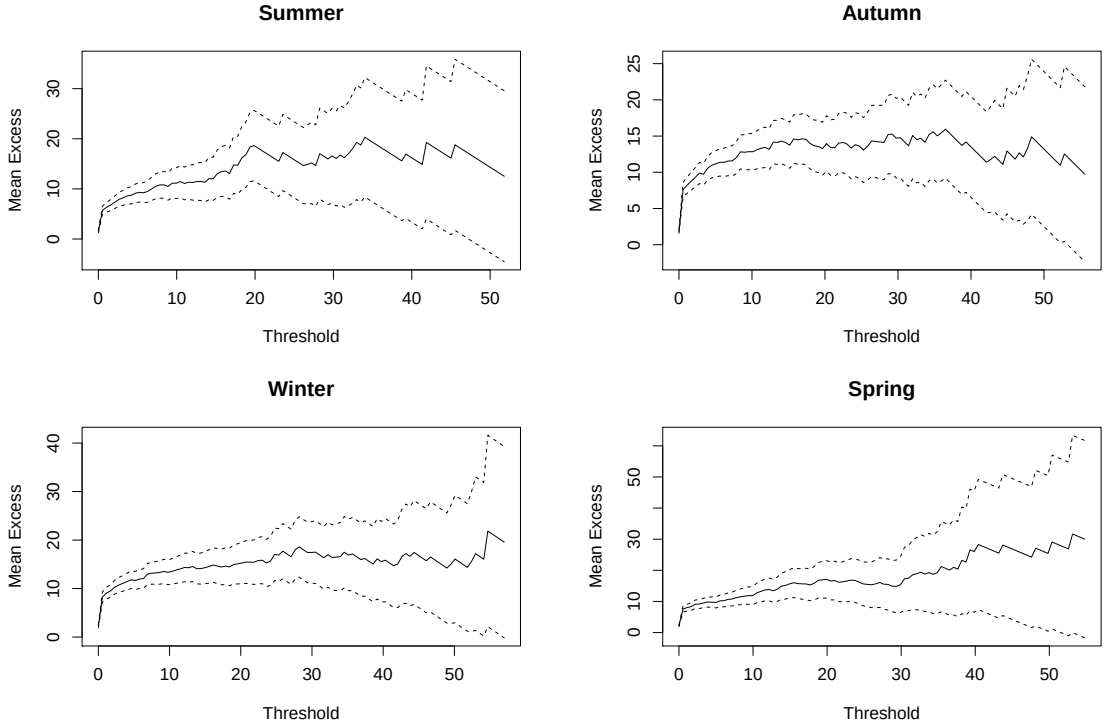


Figure A.15: Mean residual life plots for rainfall at Plettenberg Bay

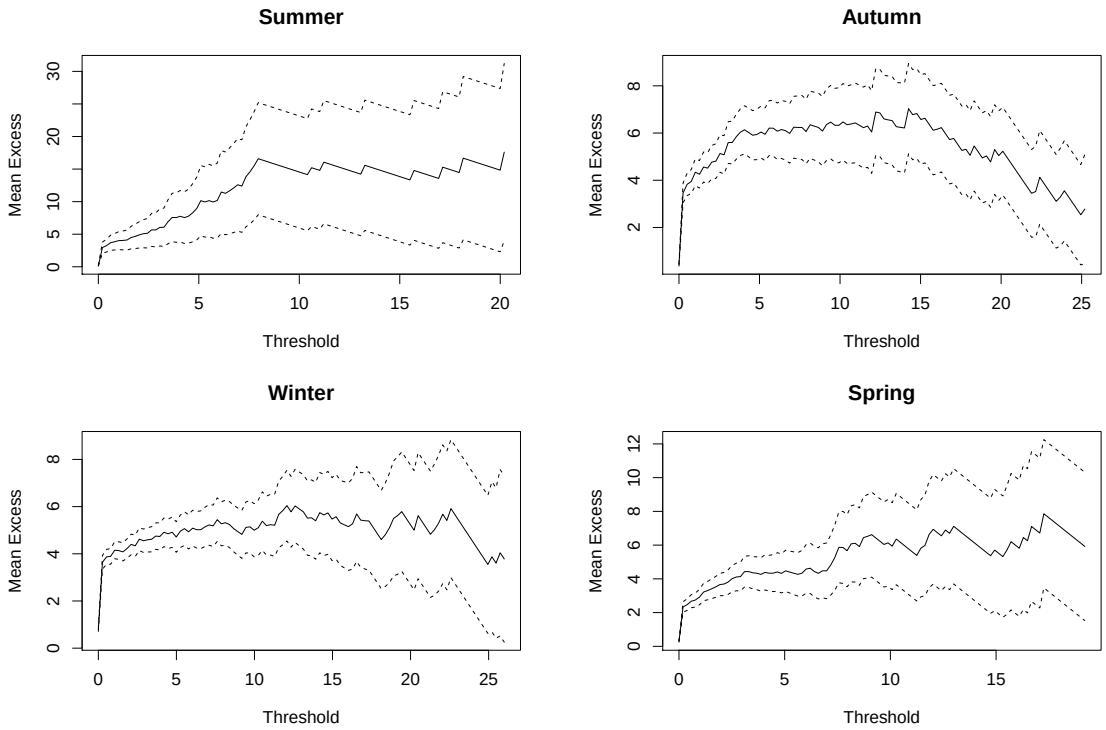


Figure A.16: Mean residual life plots for rainfall at Vredendal

A.4.2 Maximum Temperature

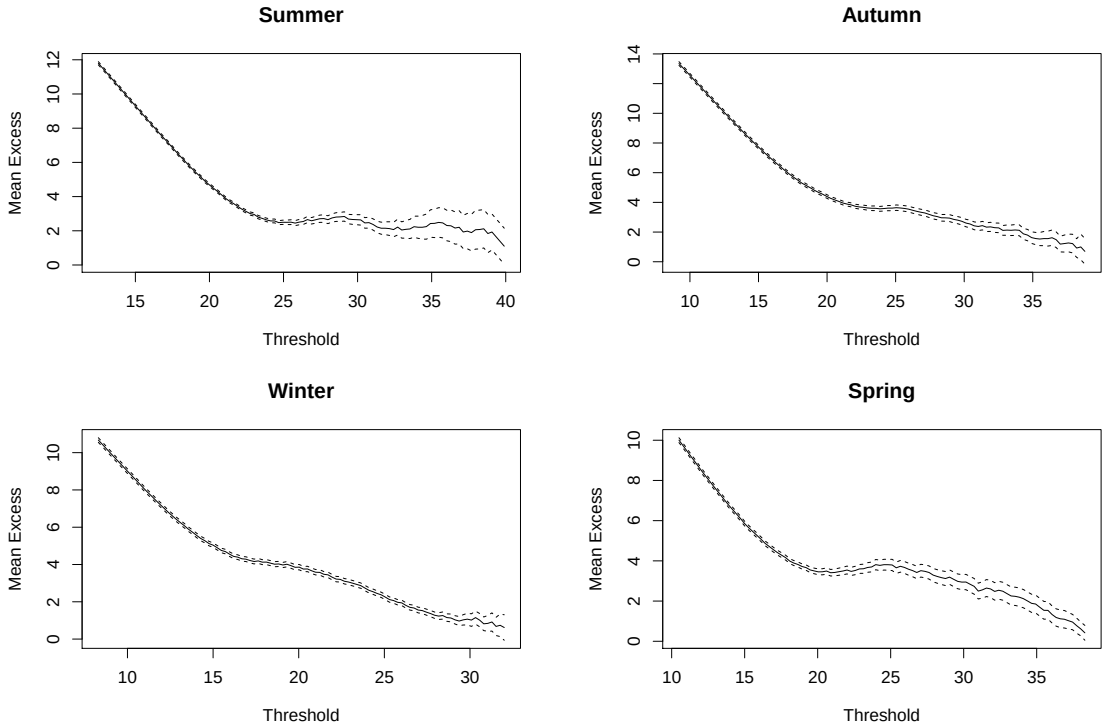


Figure A.17: Mean residual life plots for temperature at George Airport

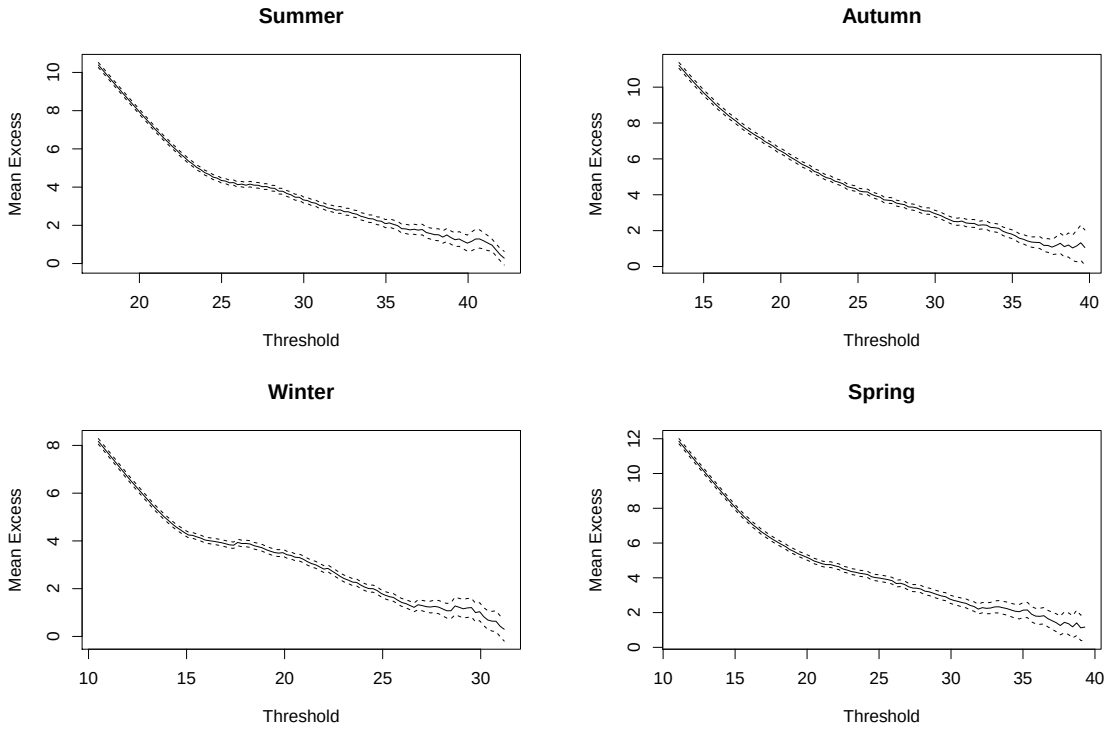


Figure A.18: Mean residual life plots for temperature at Langebaanweg

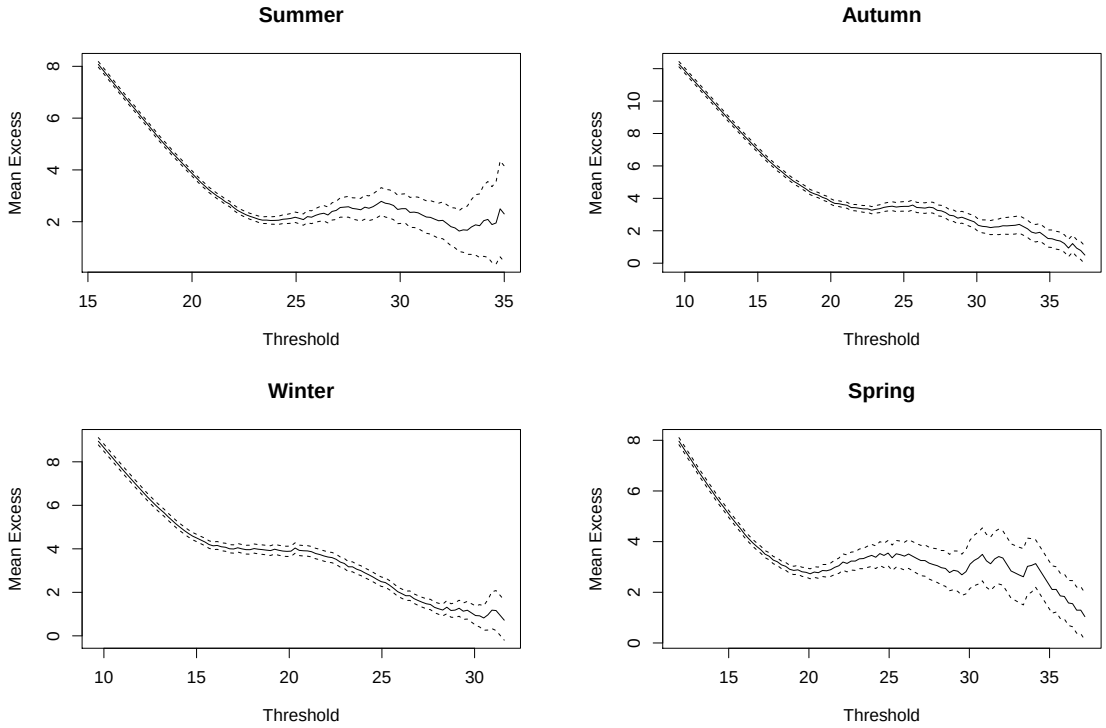


Figure A.19: Mean residual life plots for temperature at Plettenberg Bay

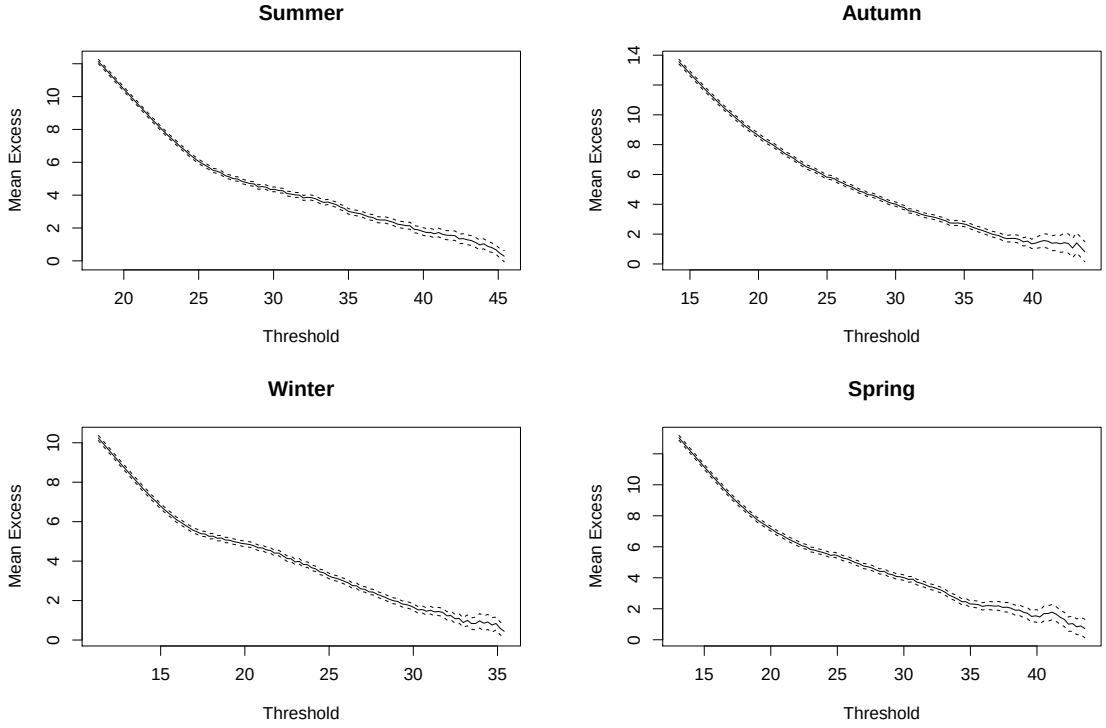


Figure A.20: Mean residual life plots for temperature at Vredendal

A.4.3 Maximum Wind Speed

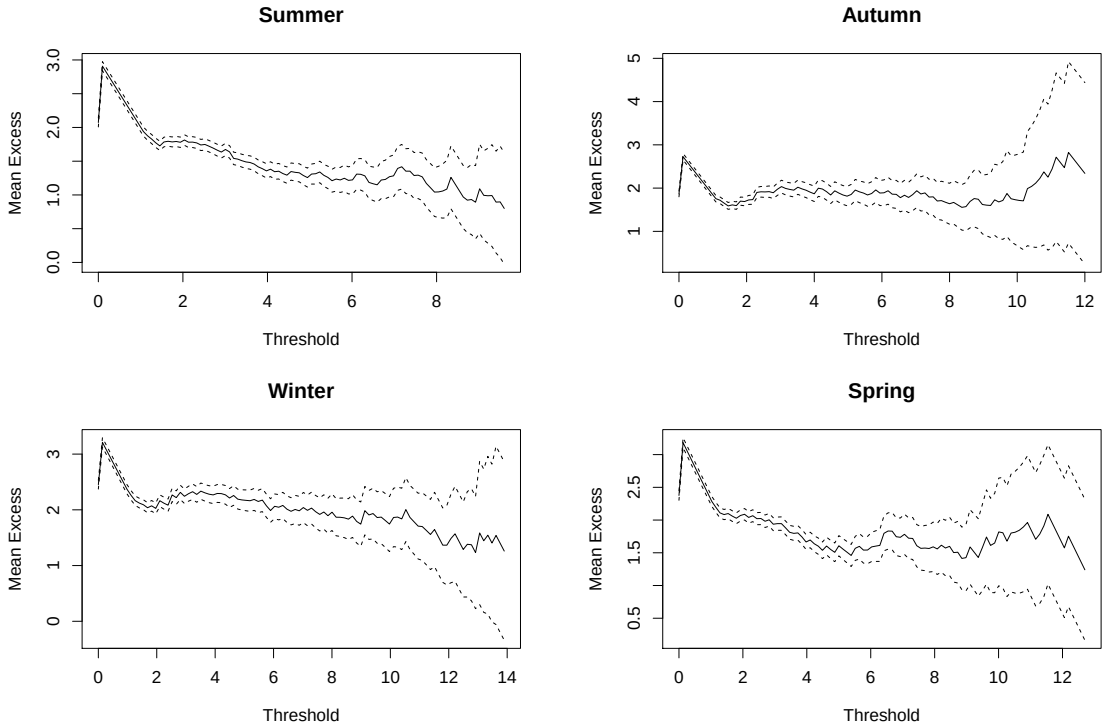


Figure A.21: Mean residual life plots for wind speed at George Airport

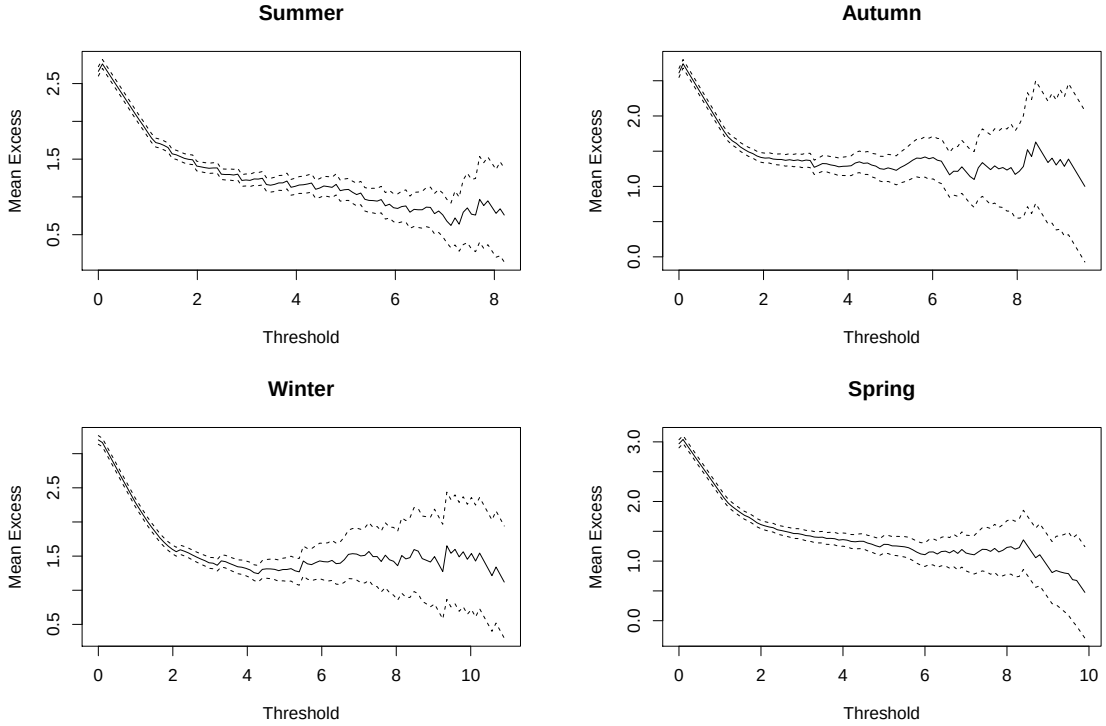


Figure A.22: Mean residual life plots for wind speed at Plettenberg Bay

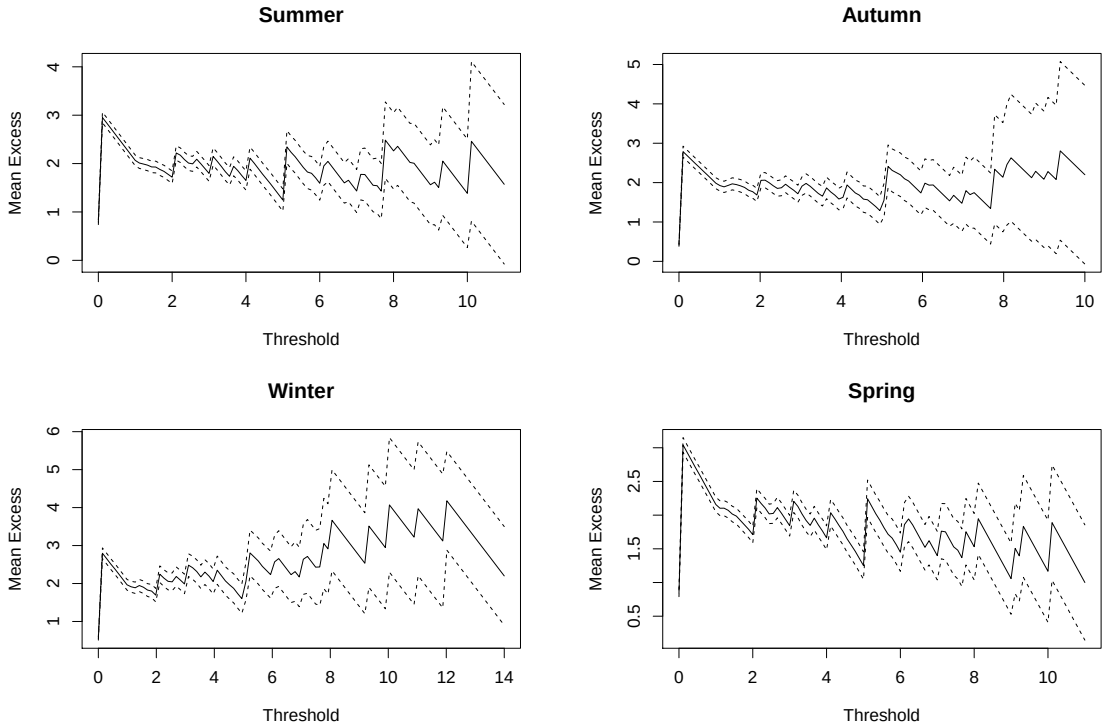


Figure A.23: Mean residual life plots for wind speed at Vredendal

A.5 Extremal Indices

Table A.4: Extremal index estimates for maximum temperature

	Season	Threshold	r=1	r=2	r=3
CT Airport	Summer	32.00	1.000	1.000	1.000
	Autumn	30.00	0.940	1.000	1.000
	Winter	25.00	1.000	1.000	1.000
	Spring	26.90	1.000	1.000	1.000
George Airport	Summer	32.00	1.000	1.000	1.000
	Autumn	33.00	0.883	0.913	0.968
	Winter	27.00	0.937	0.999	1.000
	Spring	30.99	0.907	0.957	0.967
Langebaanweg	Summer	34.00	0.931	0.976	1.000
	Autumn	30.00	1.000	1.000	1.000
	Winter	25.00	0.925	0.979	1.000
	Spring	29.00	1.000	1.000	1.000
Plettenberg Bay	Summer	28.00	0.954	0.996	1.000
	Autumn	30.00	1.000	1.000	1.000
	Winter	25.00	0.925	0.979	1.000
	Spring	26.68	1.000	1.000	1.000
Vredendal	Summer	39.10	0.839	0.885	0.903
	Autumn	37.00	0.878	0.960	1.000
	Winter	30.90	0.904	0.977	1.000
	Spring	34.00	1.000	1.000	1.000

Table A.5: Extremal index estimates for maximum wind speed

	Season	Threshold	r=1	r=2	r=3
CT Airport	Summer	10.20	0.823	0.906	0.925
	Autumn	9.00	1.000	1.000	1.000
	Winter	10.00	0.940	1.000	1.000
	Spring	11.00	0.876	0.924	0.933
George Airport	Summer	5.20	0.847	0.931	1.000
	Autumn	6.00	1.000	1.000	1.000
	Winter	7.00	0.980	0.999	1.000
	Spring	6.00	0.937	0.993	1.000
Plettenberg Bay	Summer	5.00	1.000	1.000	1.000
	Autumn	6.00	0.937	0.968	1.000
	Winter	6.00	0.801	0.834	0.886
	Spring	6.00	0.976	1.000	1.000
Vredendal	Summer	5.10	0.819	0.856	0.988
	Autumn	2.70	0.802	0.884	0.985
	Winter	1.90	0.988	1.000	1.000
	Spring	3.00	0.841	0.986	1.000

A.6 Threshold Excess

A.6.1 Maximum Rainfall

Return Level Plots

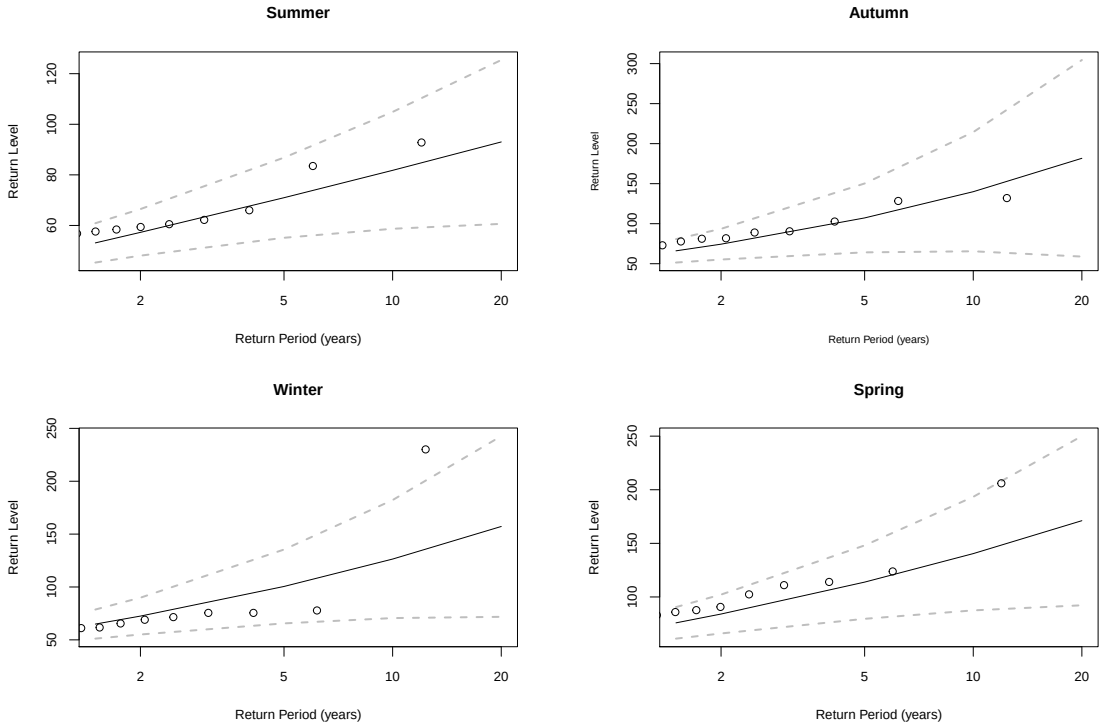


Figure A.24: Return level plots for rainfall at George Airport

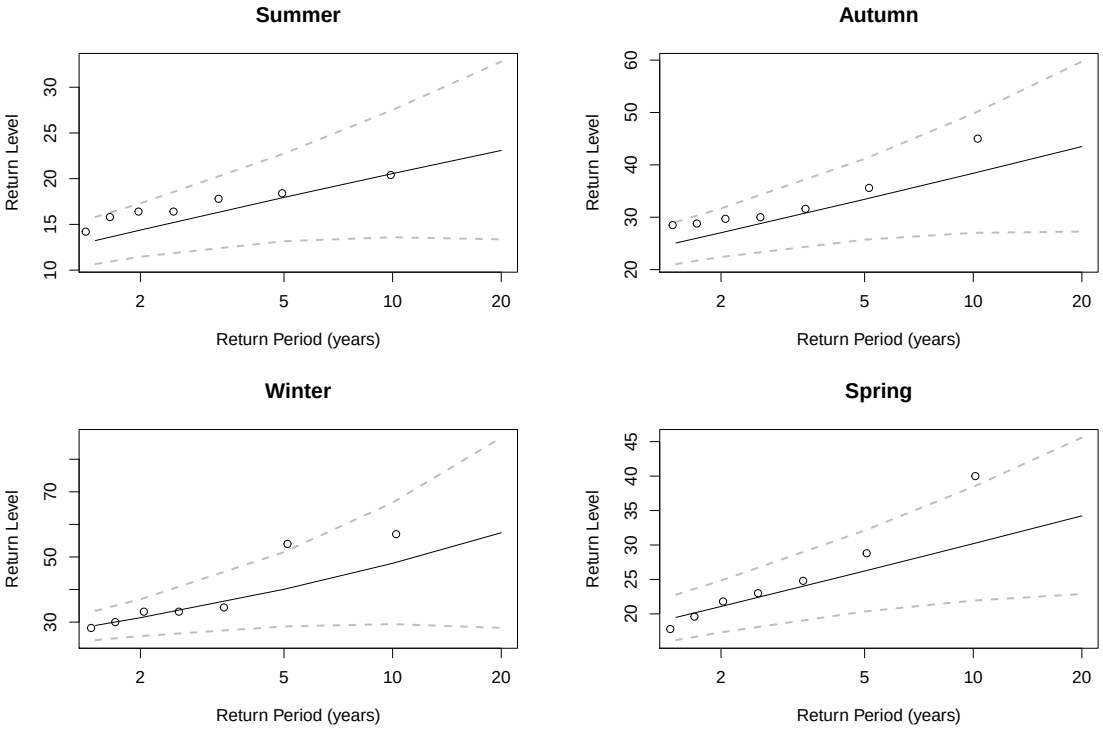


Figure A.25: Return level plots for rainfall at Langebaanweg

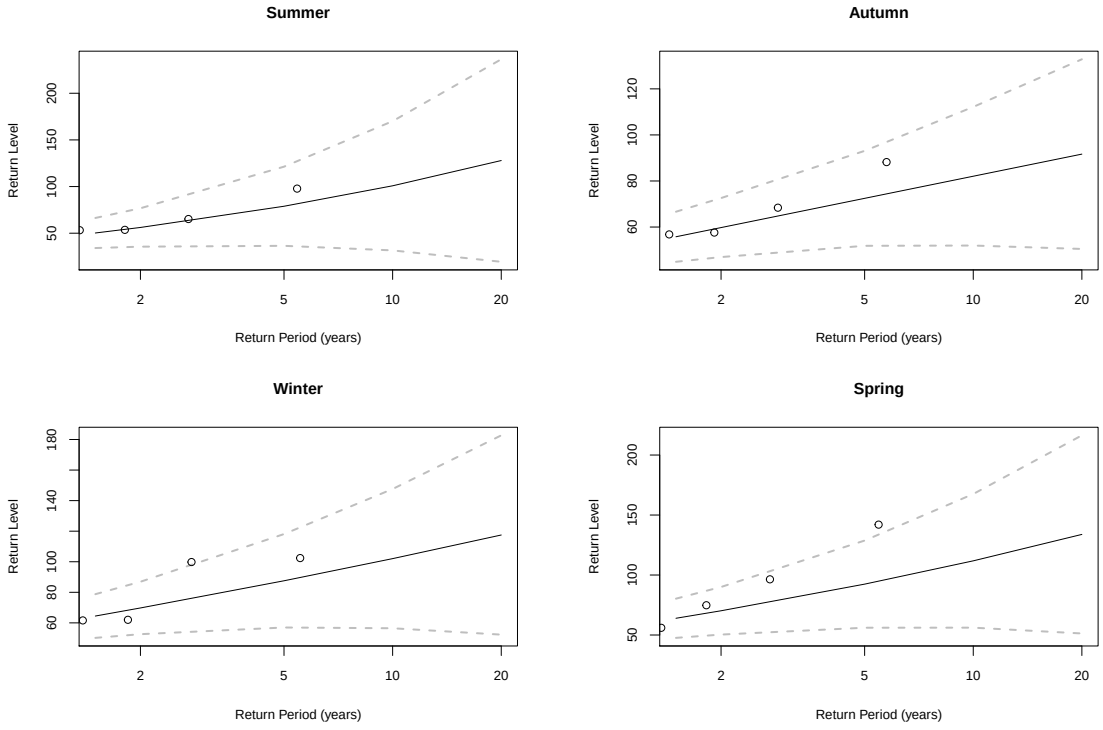


Figure A.26: Return level plots for rainfall at Plettenberg Bay

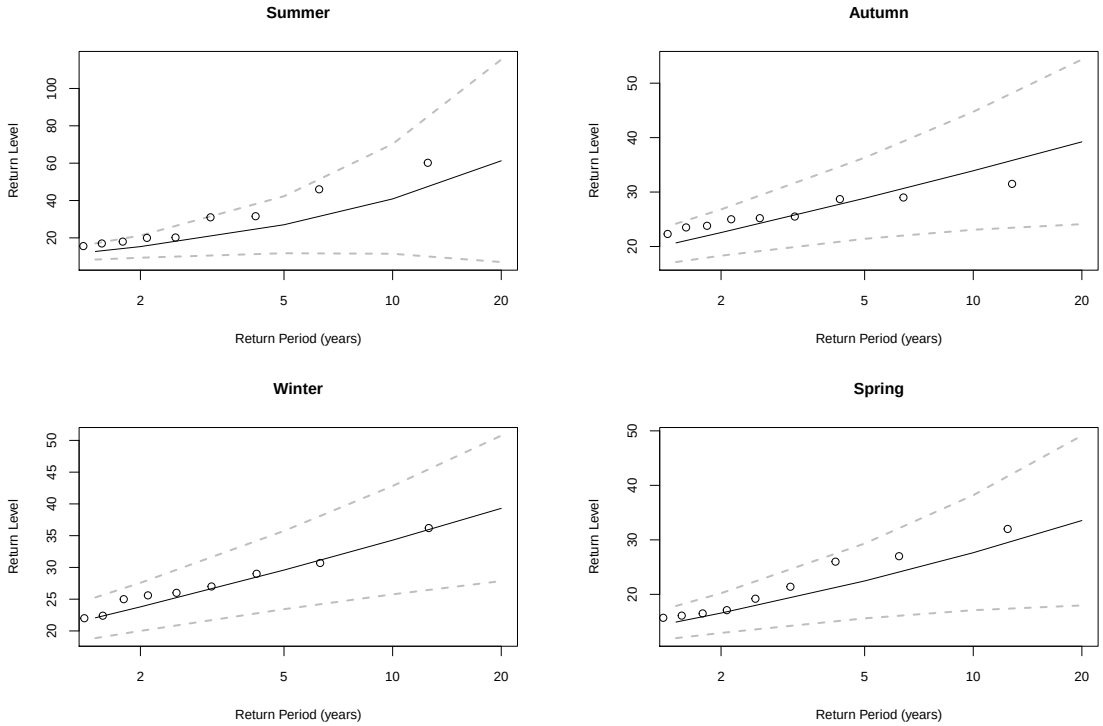


Figure A.27: Return level plots for rainfall at Vredendal

Quantile-Quantile Plots

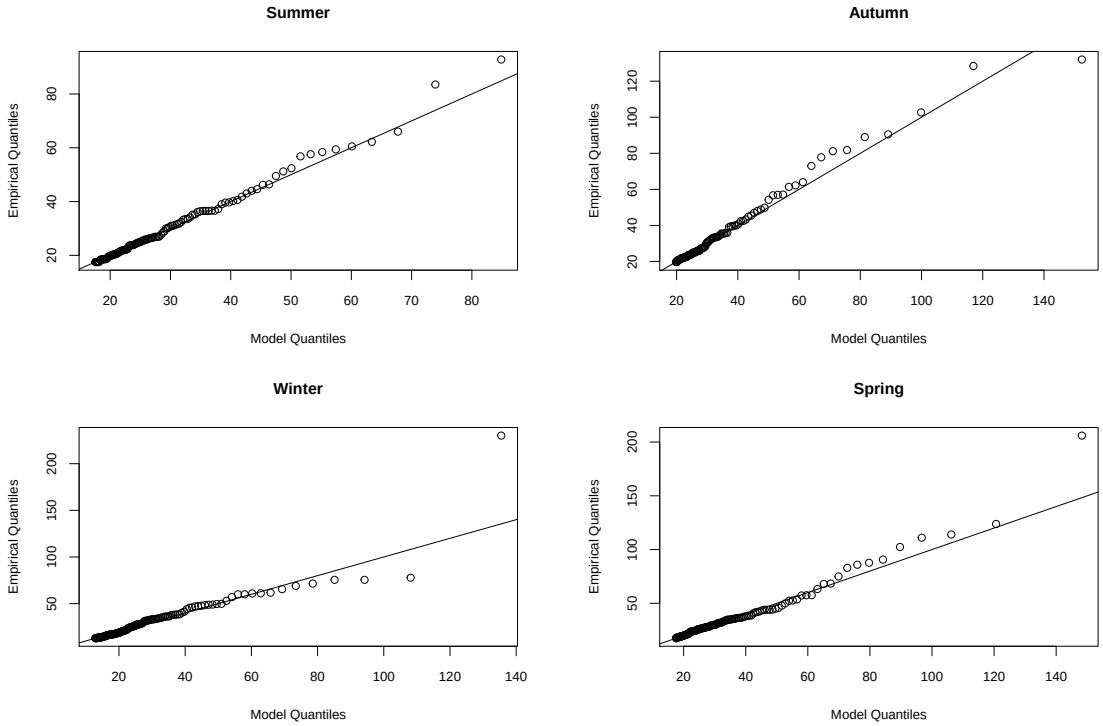


Figure A.28: Quantile-quantile plots for rainfall at George Airport

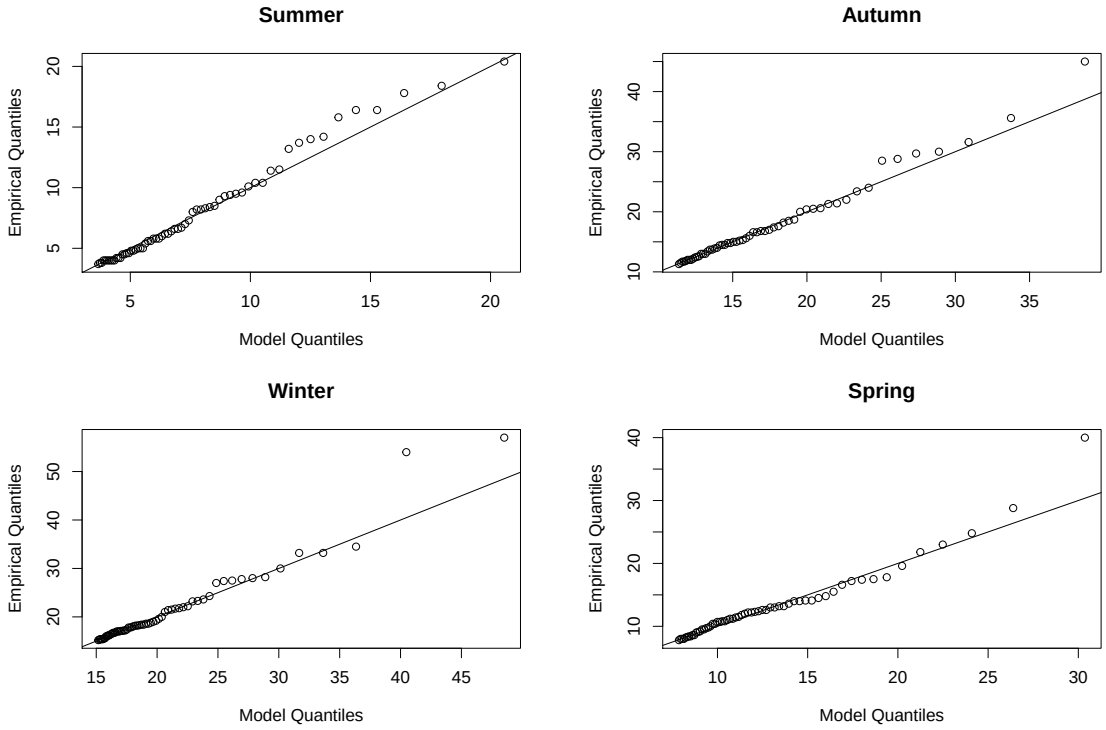


Figure A.29: Quantile-quantile plots for rainfall at Langebaanweg

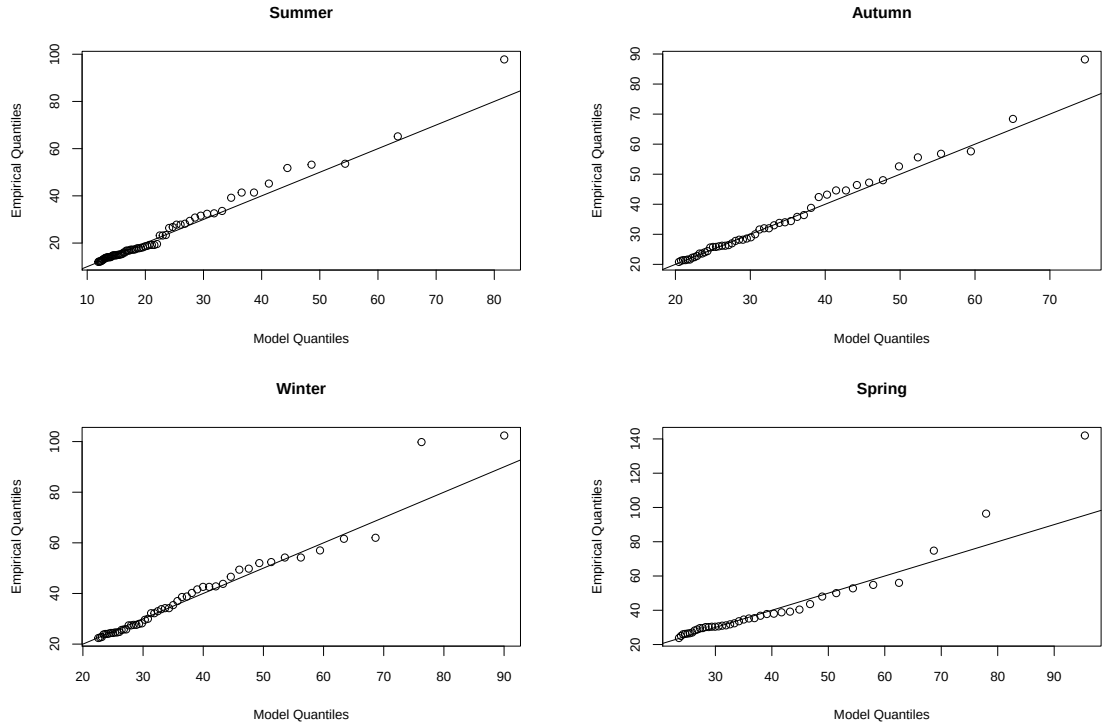


Figure A.30: Quantile-quantile plots for rainfall at Plettenberg Bay

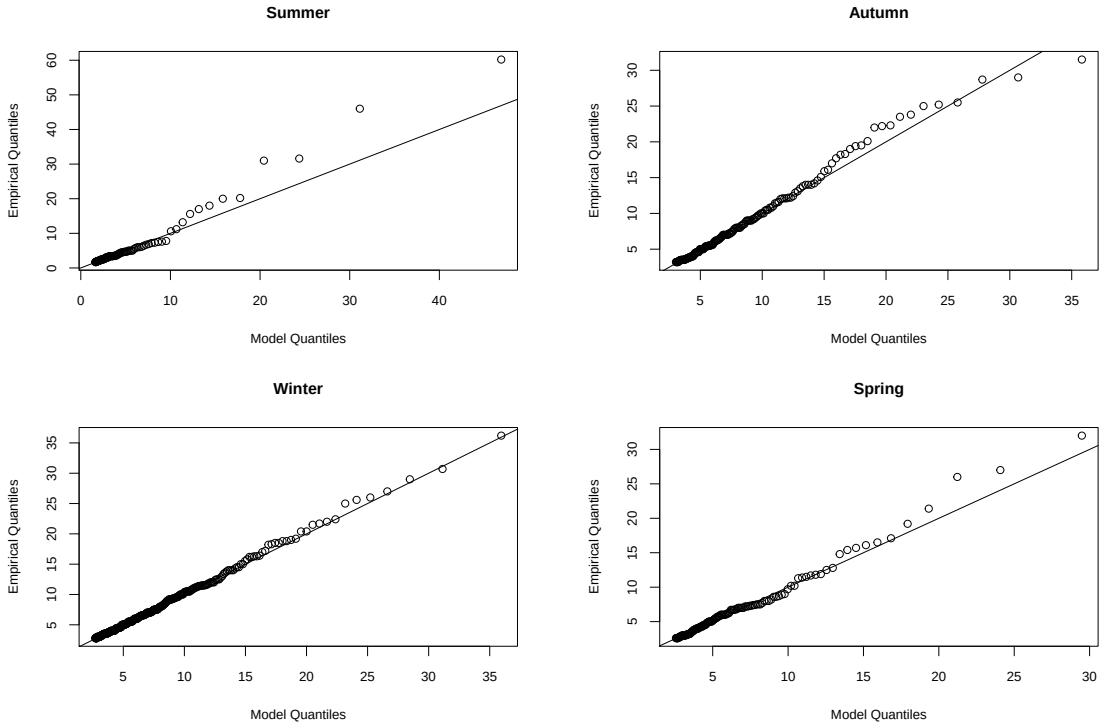


Figure A.31: Quantile-quantile plots for rainfall at Vredendal

A.6.2 Maximum Temperature

Return Level Plots

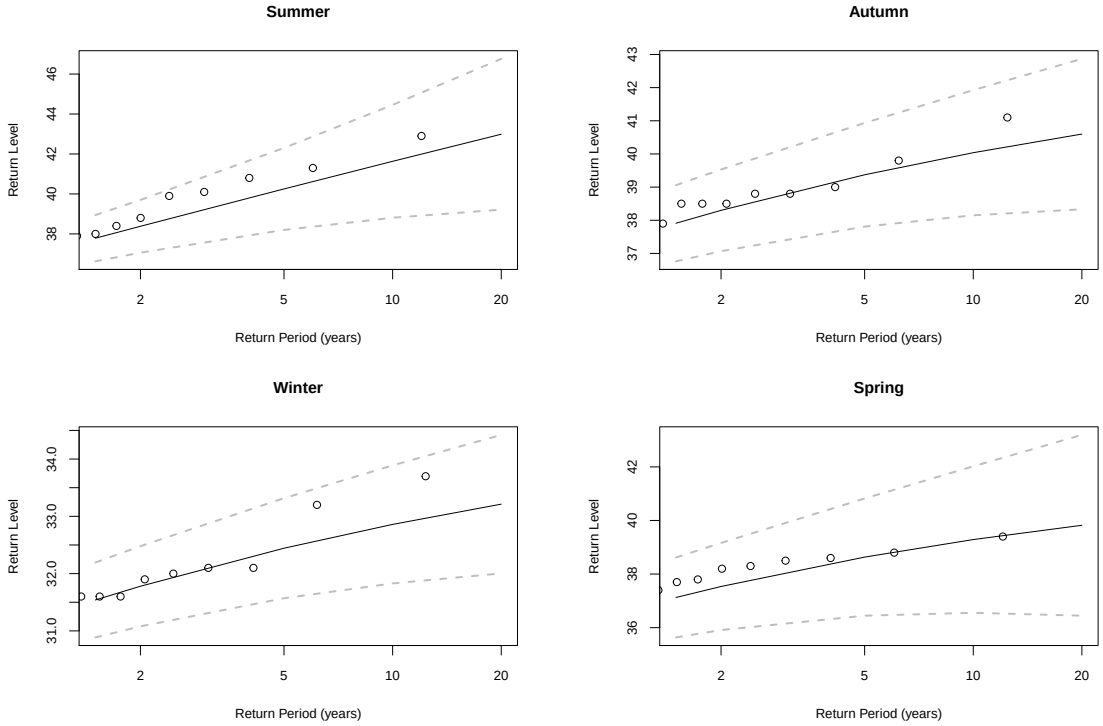


Figure A.32: Return level plots for temperature at George

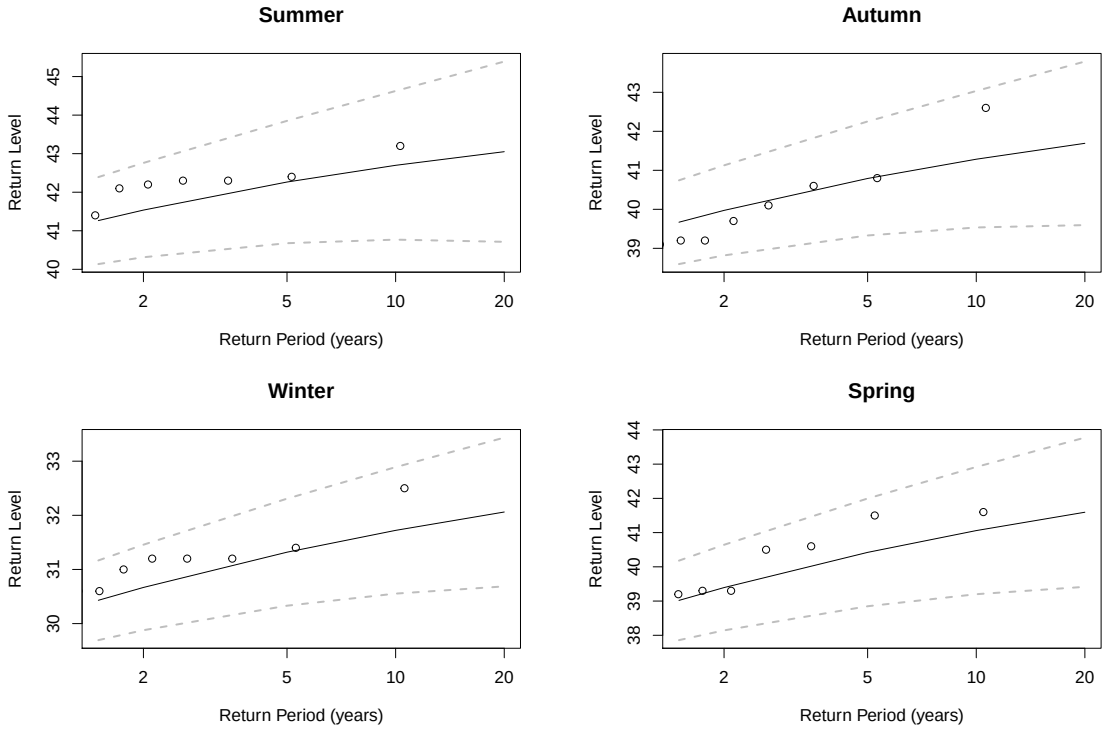


Figure A.33: Return level plots for temperature at Langebaanweg

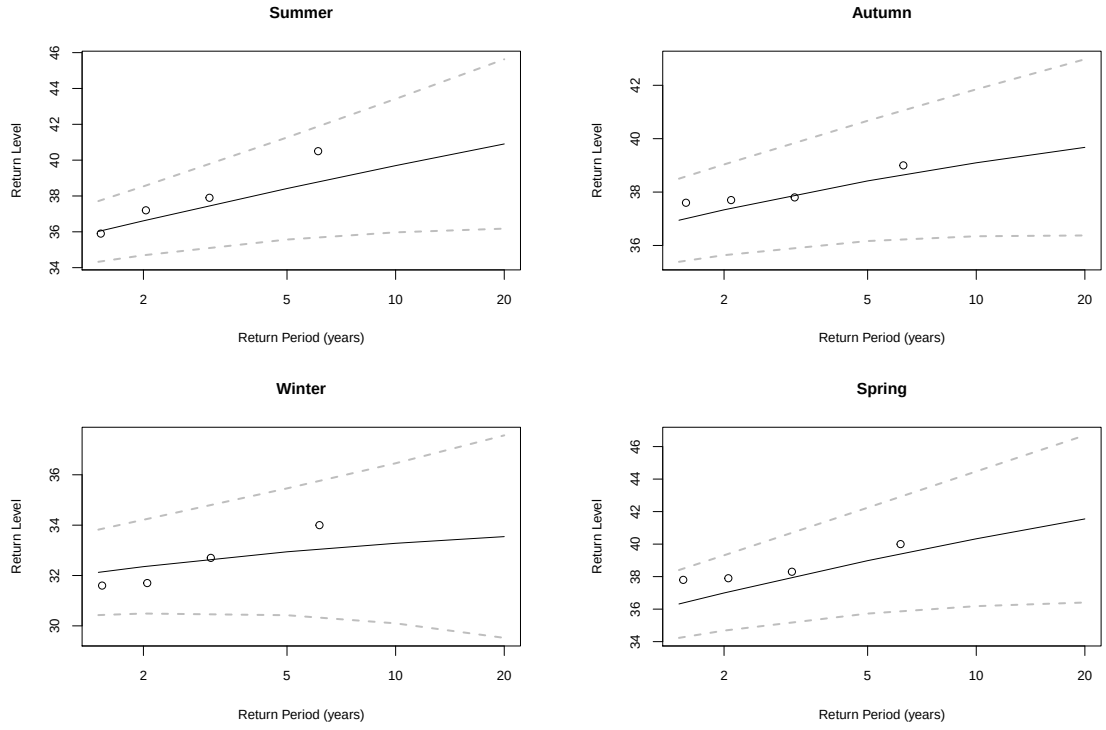


Figure A.34: Return level plots for temperature at Plettenberg Bay

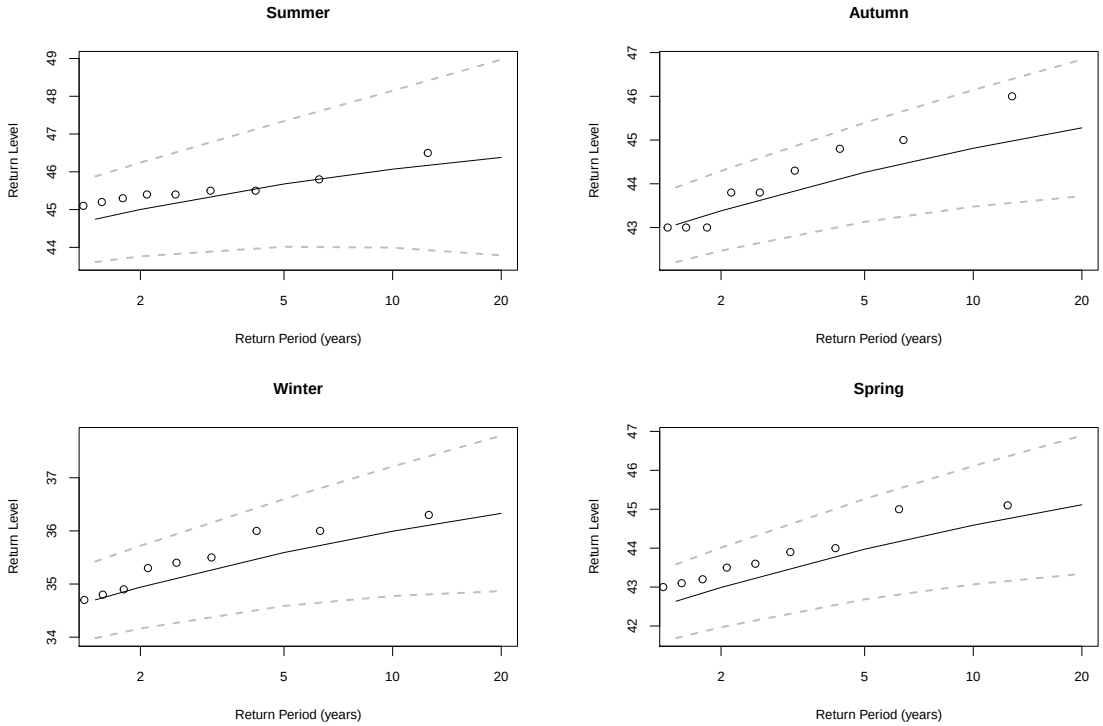


Figure A.35: Return level plots for temperature at Vredendal

Quantile-Quantile Plots

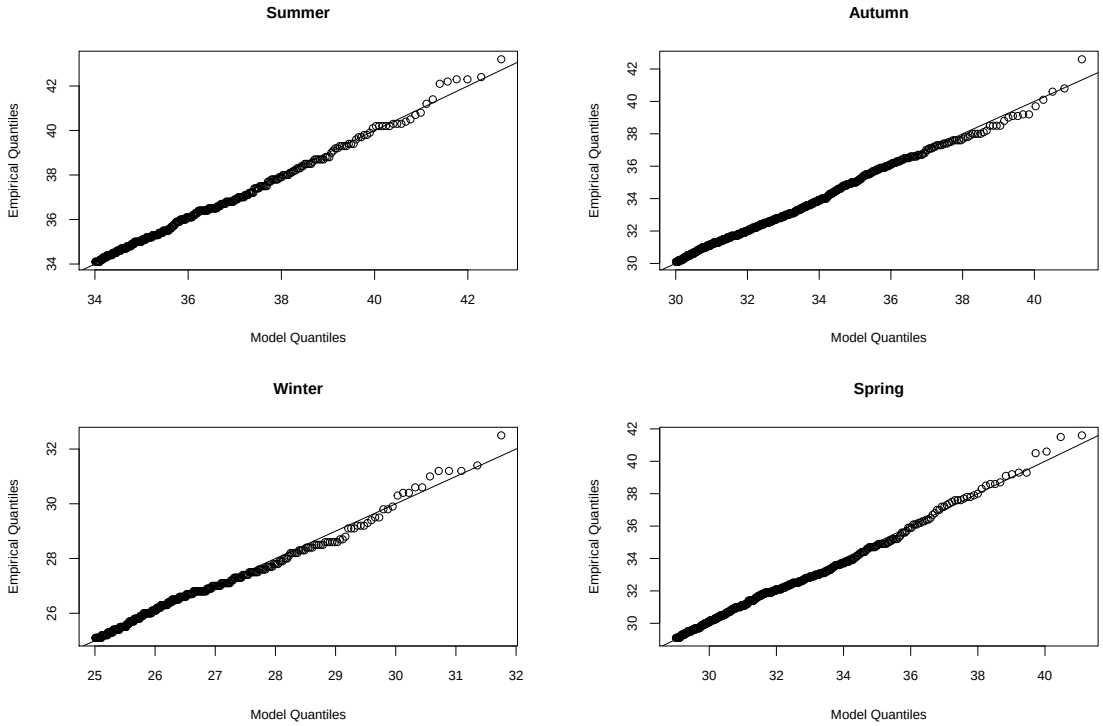


Figure A.36: Quantile-quantile plots for temperature at George Airport

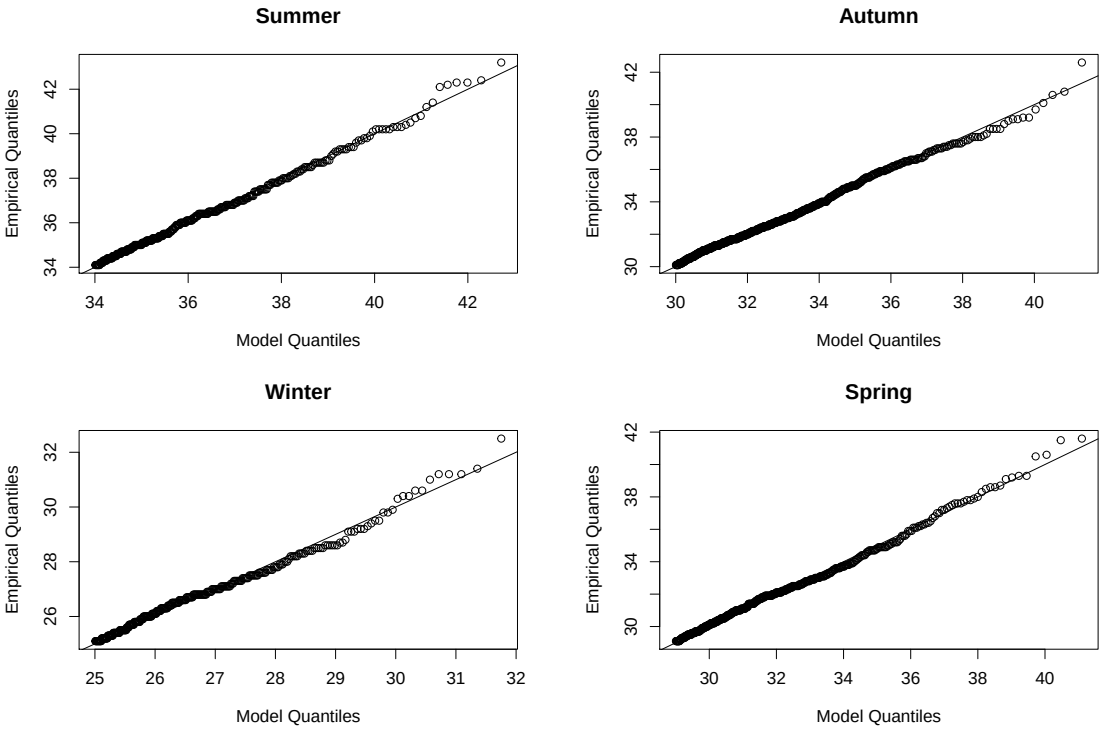


Figure A.37: Quantile-quantile plots for temperature at Langebaanweg

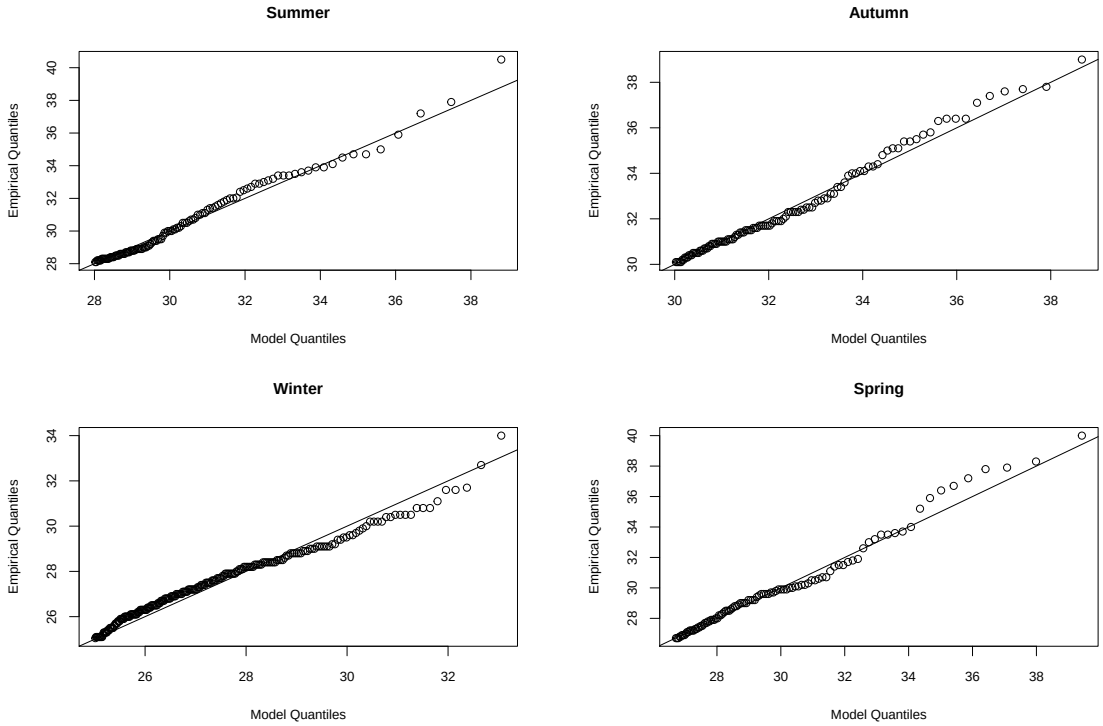


Figure A.38: Quantile-quantile plots for temperature at Plettenberg Bay

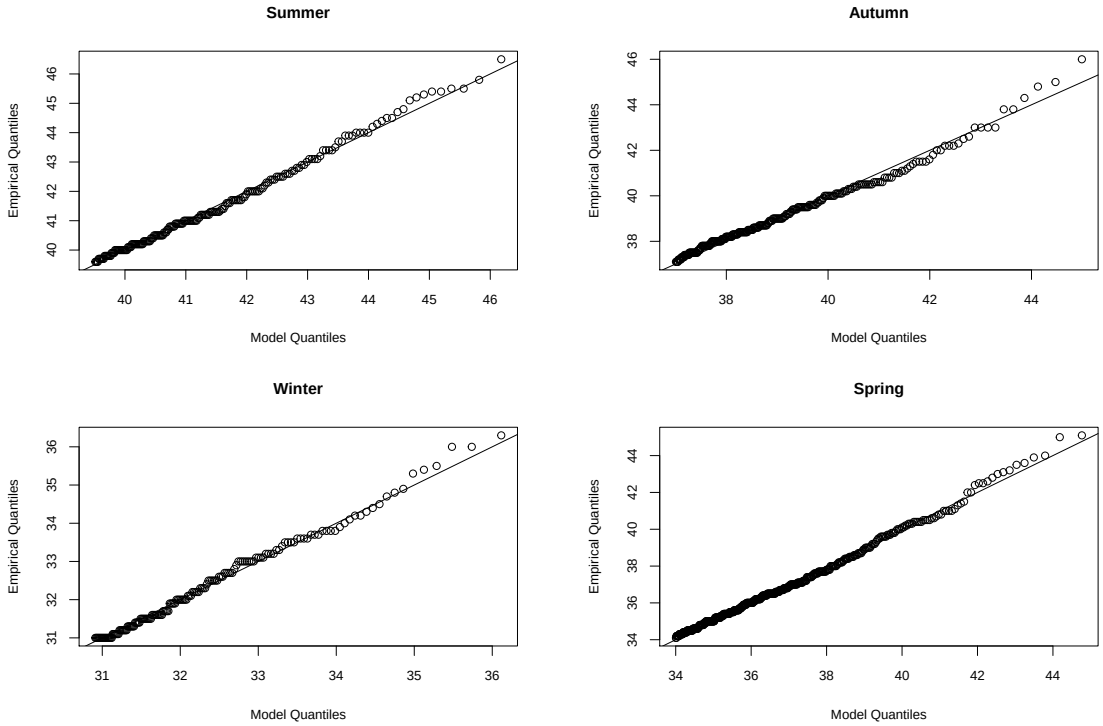


Figure A.39: Quantile-quantile plots for temperature at Vredendal

A.6.3 Maximum Wind Speed

Return Level Plots

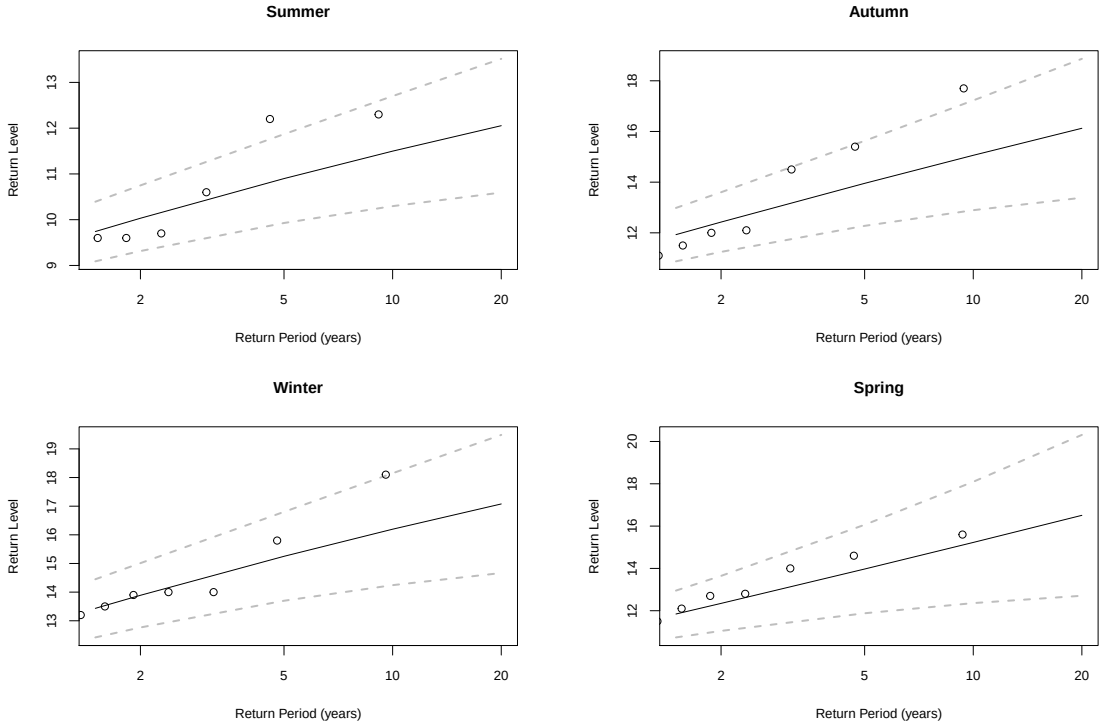


Figure A.40: Return level plots for wind speed at George Airport

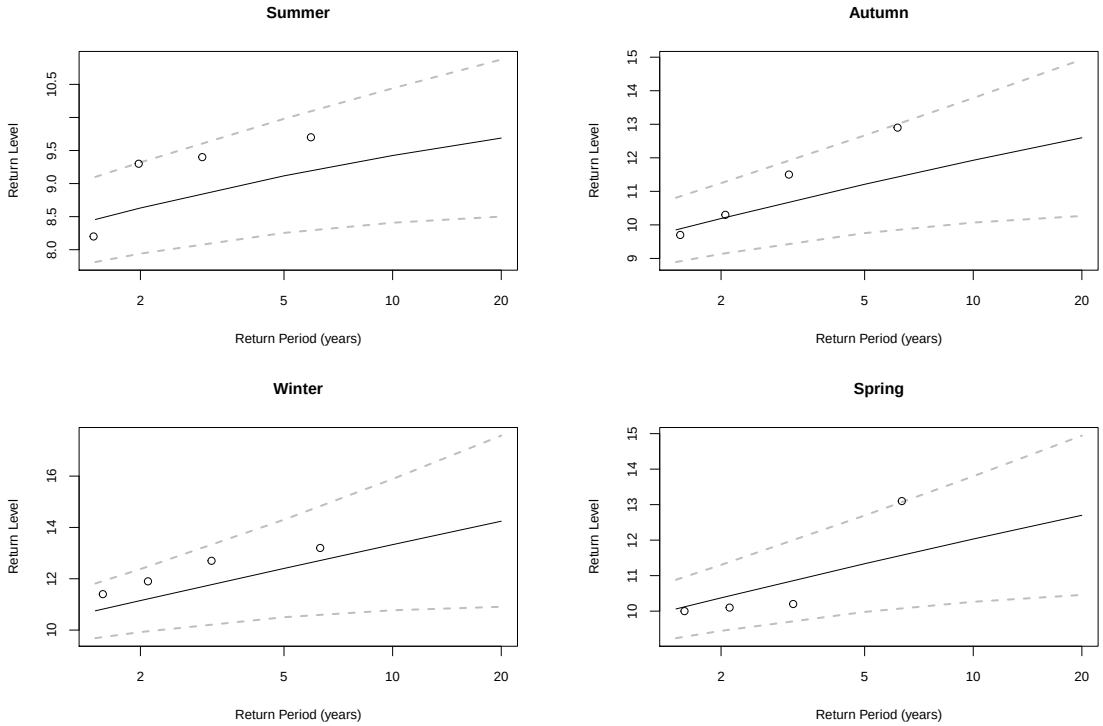


Figure A.41: Return level plots for wind speed at Plettenberg Bay

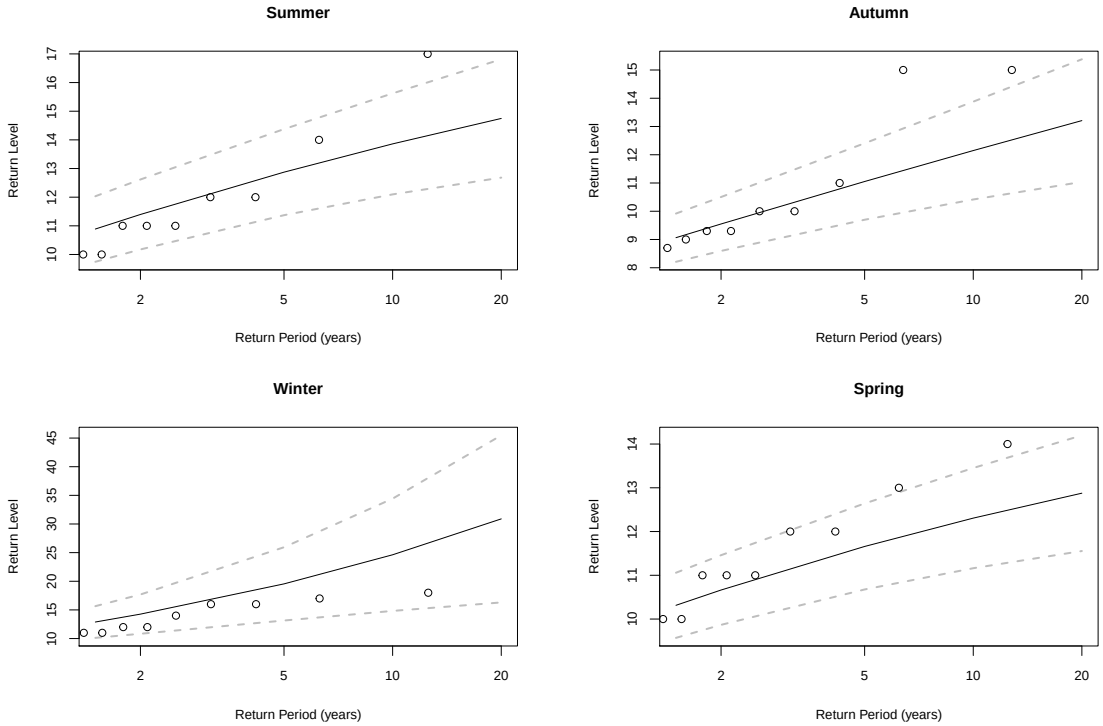


Figure A.42: Return level plots for wind speed at Vredendal

Quantile-Quantile Plots

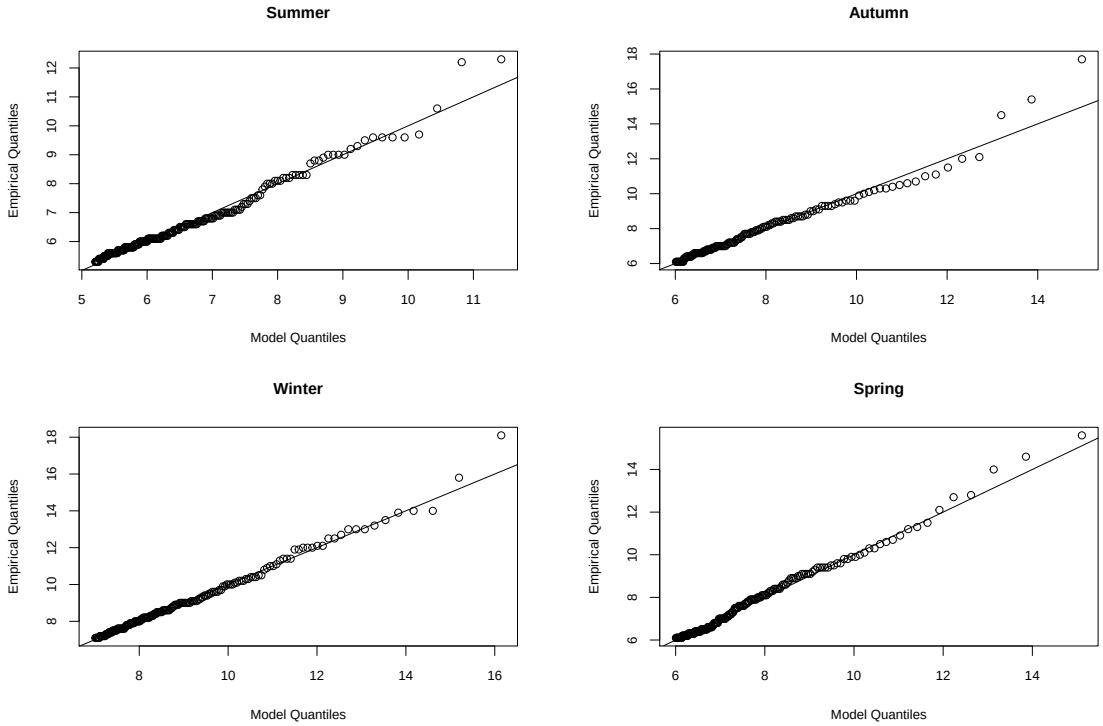


Figure A.43: Quantile-quantile plots for wind speed at George Airport

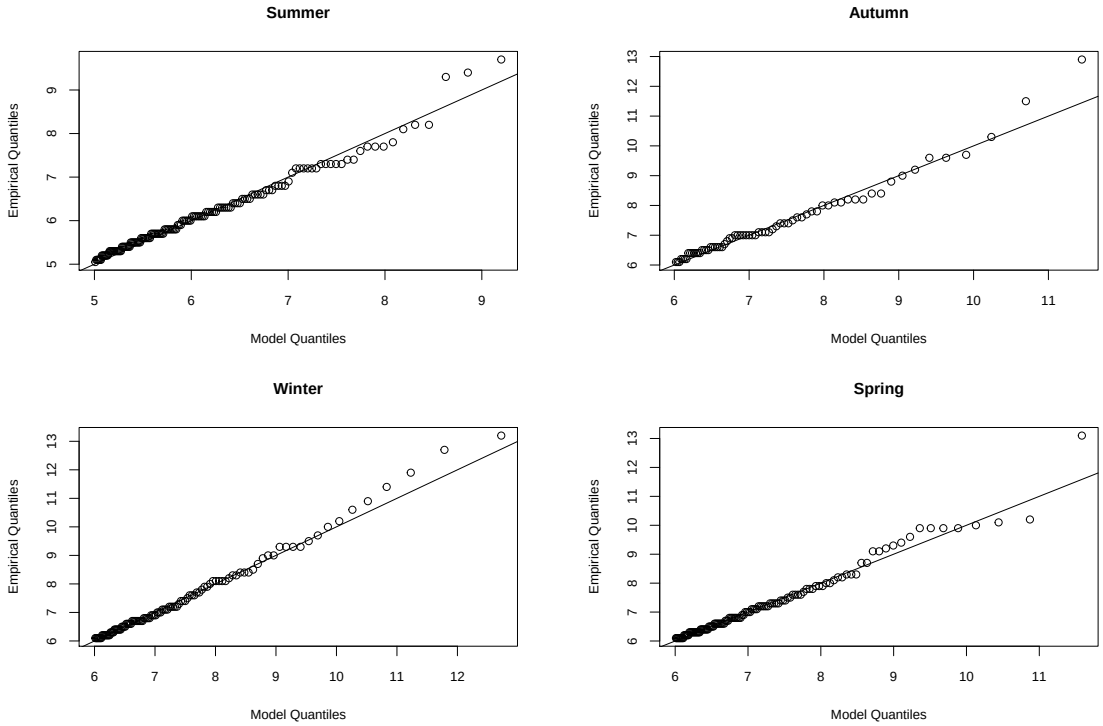


Figure A.44: Quantile-quantile plots for wind speed at Plettenberg Bay

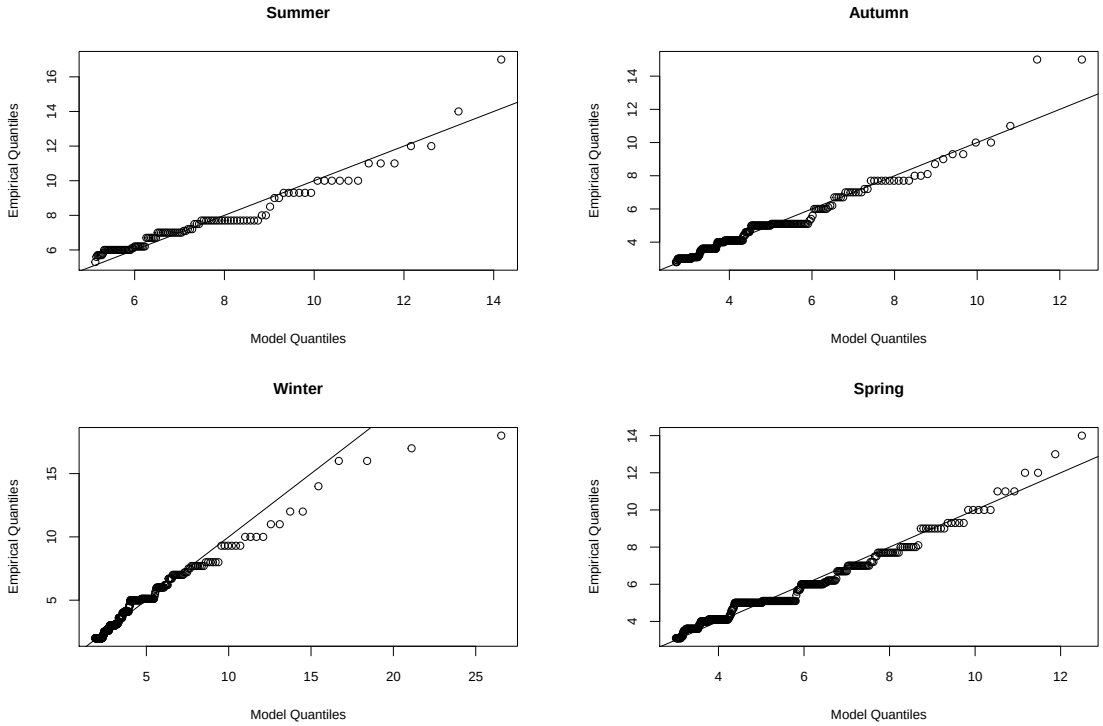


Figure A.45: Quantile-quantile plots for wind speed at Vredendal

A.7 Point Process

A.7.1 Maximum Rainfall

Return Level Plots

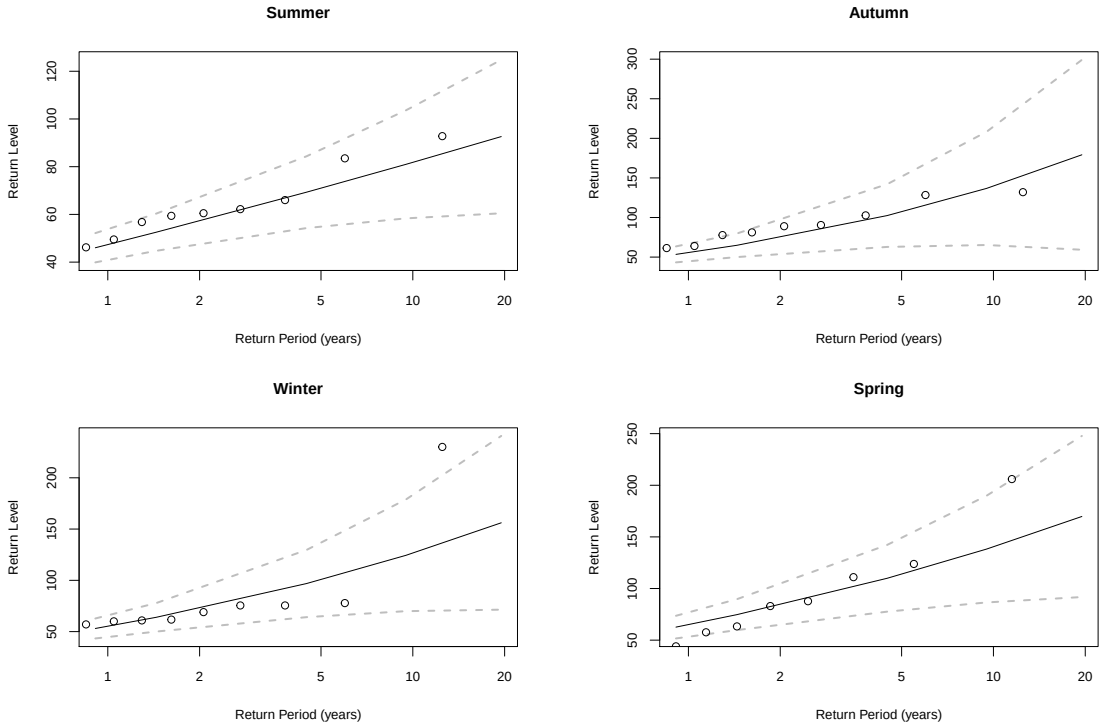


Figure A.46: Return level plots for rainfall at George Airport

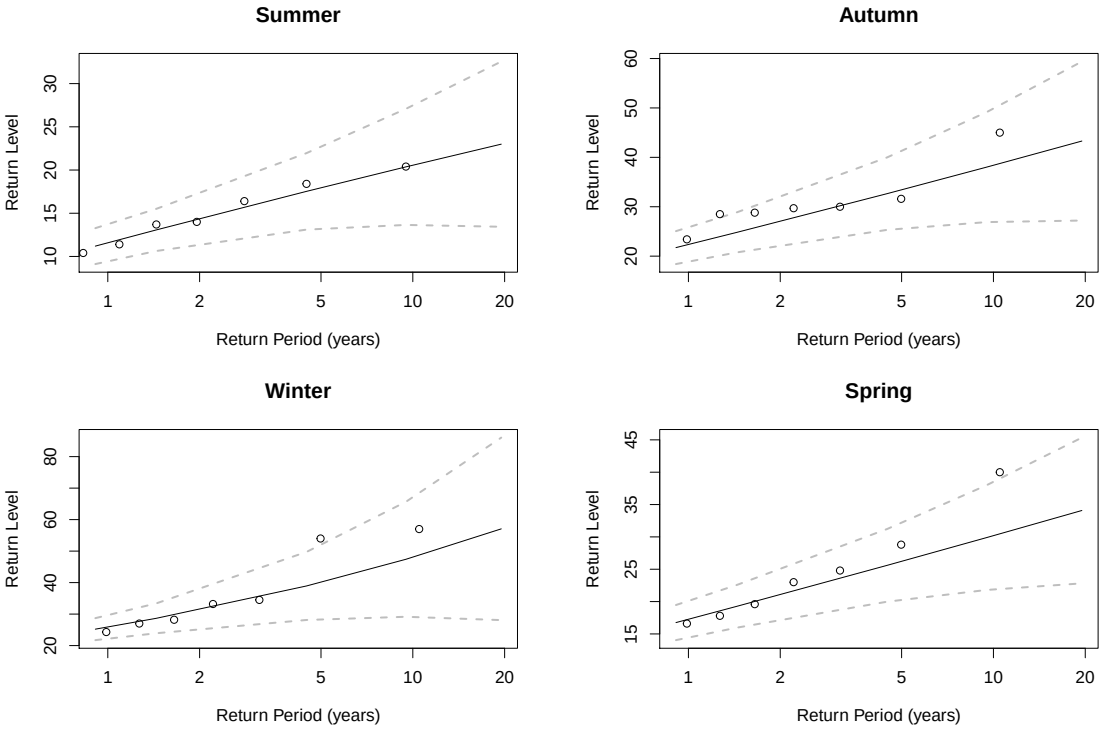


Figure A.47: Return level plots for rainfall at Langebaanweg

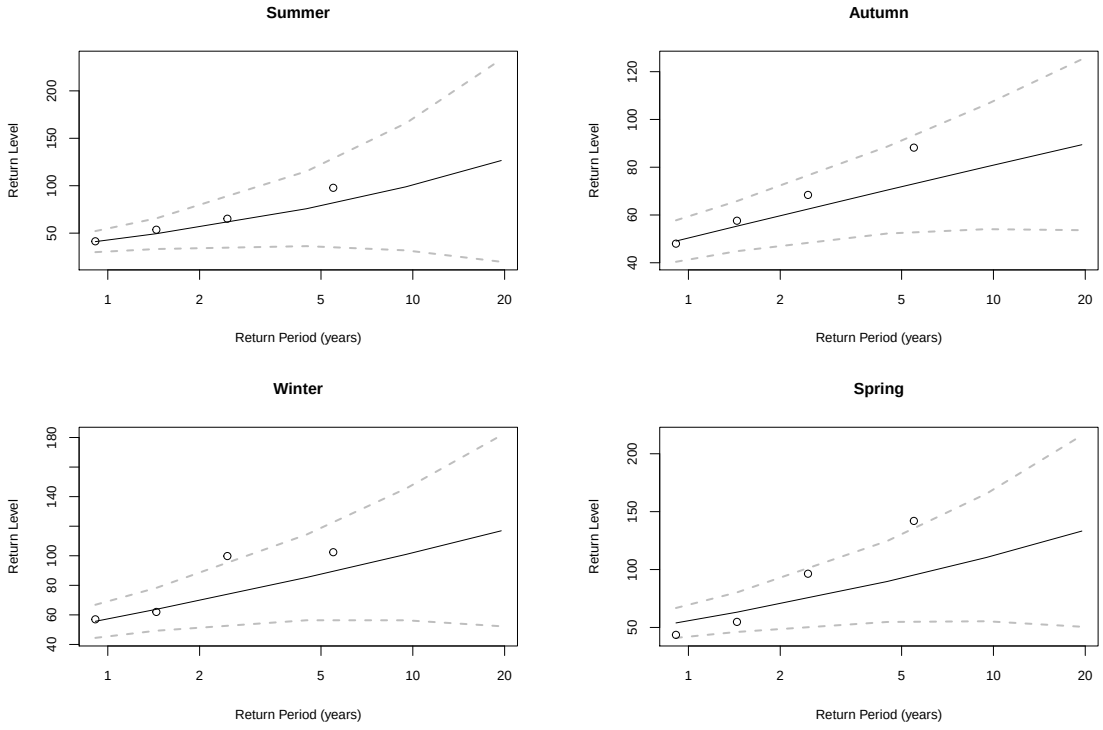


Figure A.48: Return level plots for rainfall at Plettenberg Bay

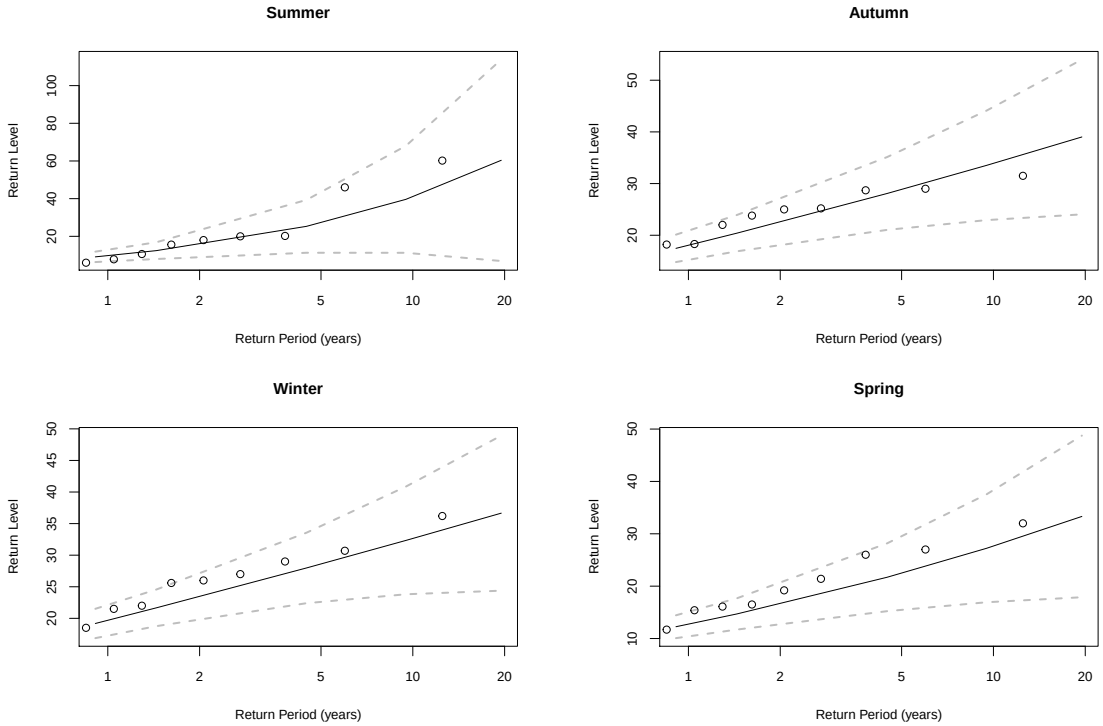


Figure A.49: Return level plots for rainfall at Vredendal

Quantile-Quantile Plots

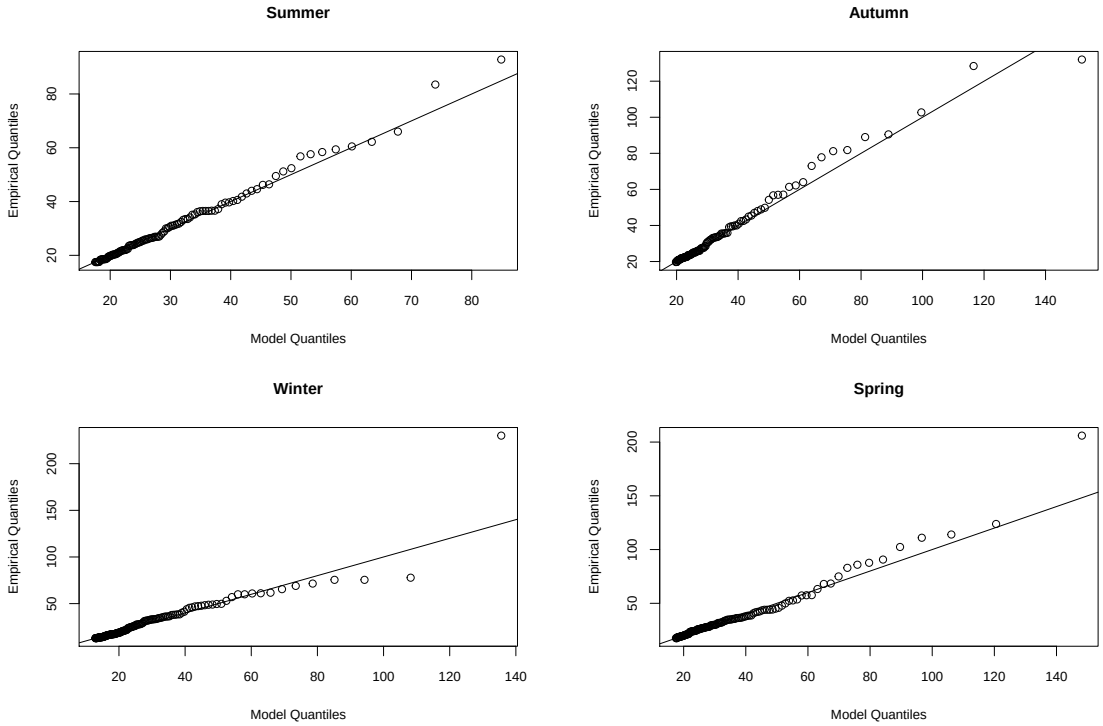


Figure A.50: Quantile-quantile plots for rainfall at George Airport

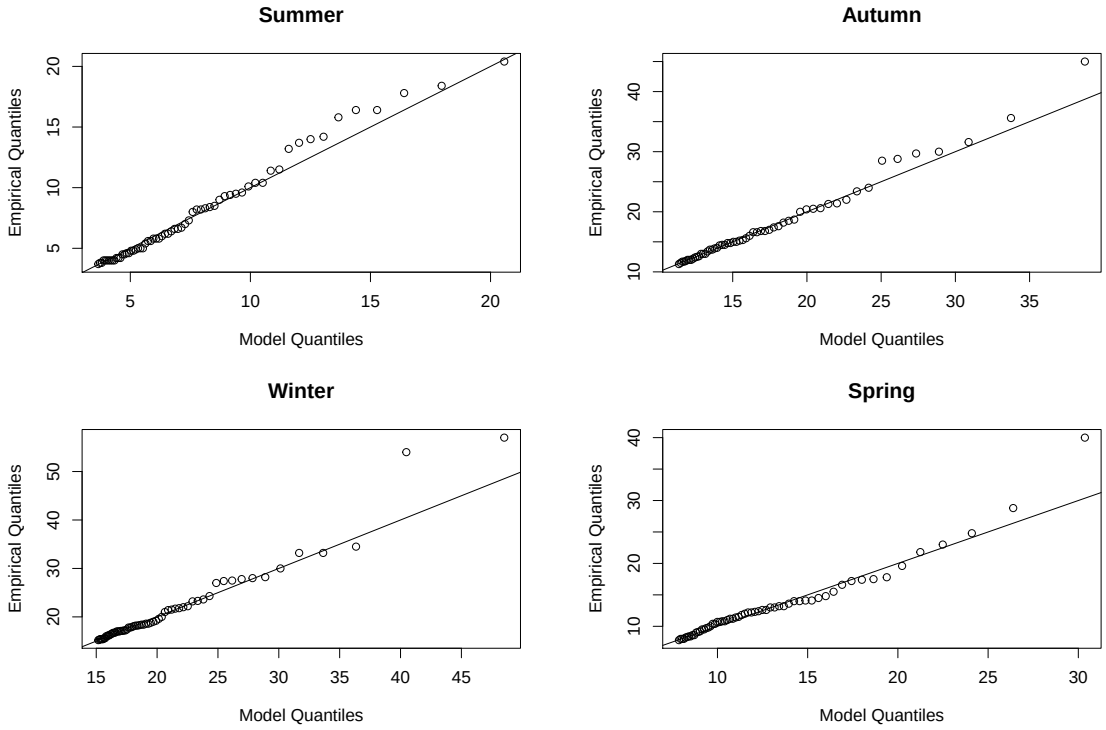


Figure A.51: Quantile-quantile plots for rainfall at Langebaanweg

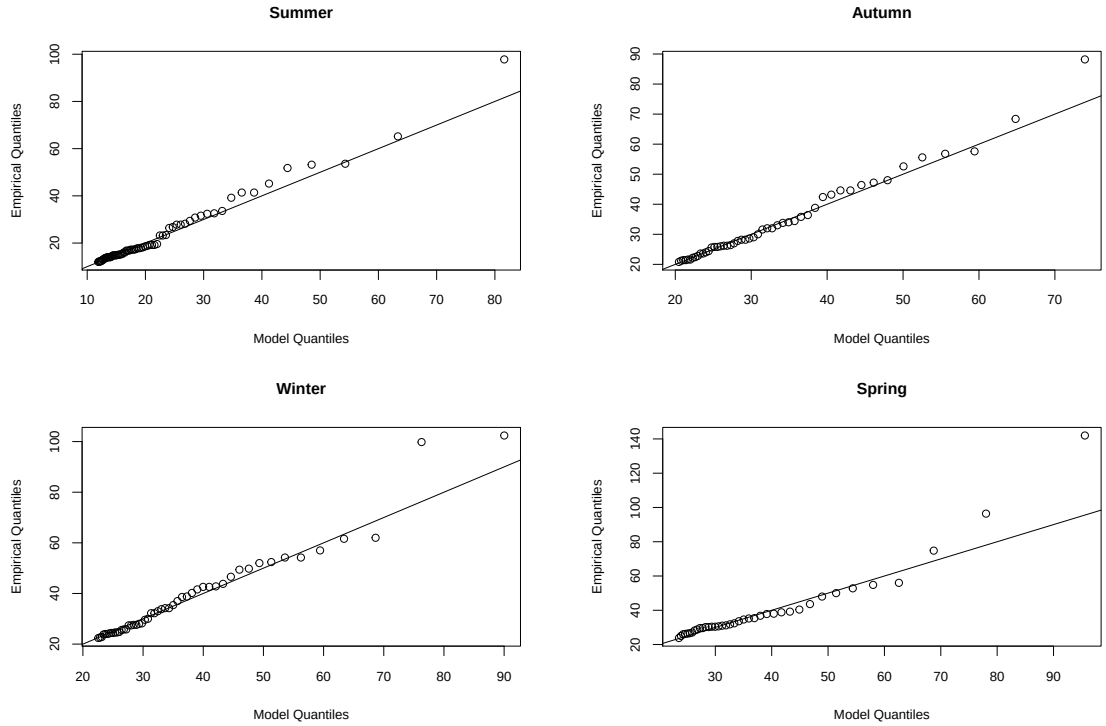


Figure A.52: Quantile-quantile plots for rainfall at Plettenberg Bay

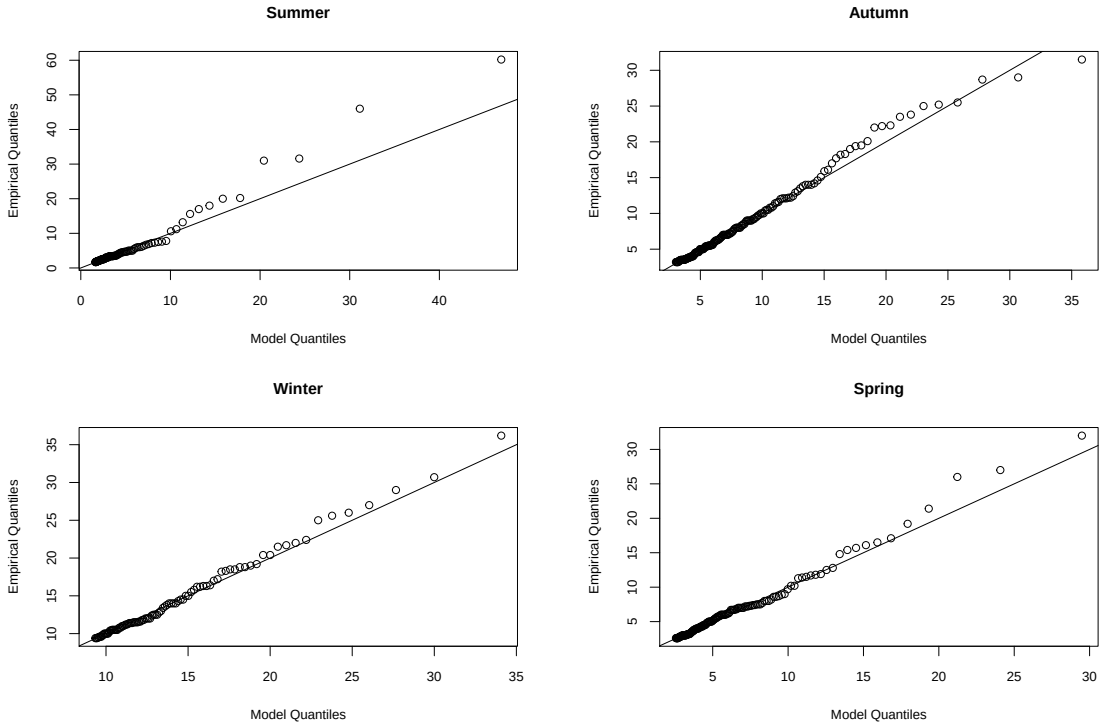


Figure A.53: Quantile-quantile plots for rainfall at Vredendal

A.7.2 Maximum Temperature

Return Level Plots

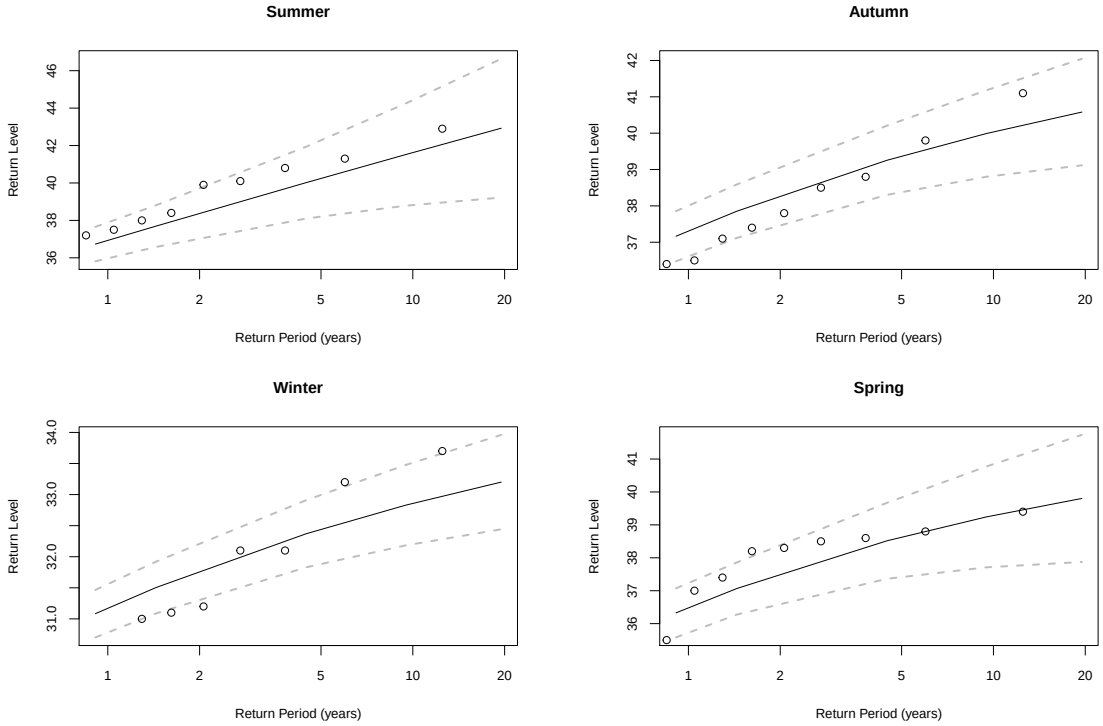


Figure A.54: Return level plots for temperature at George Airport

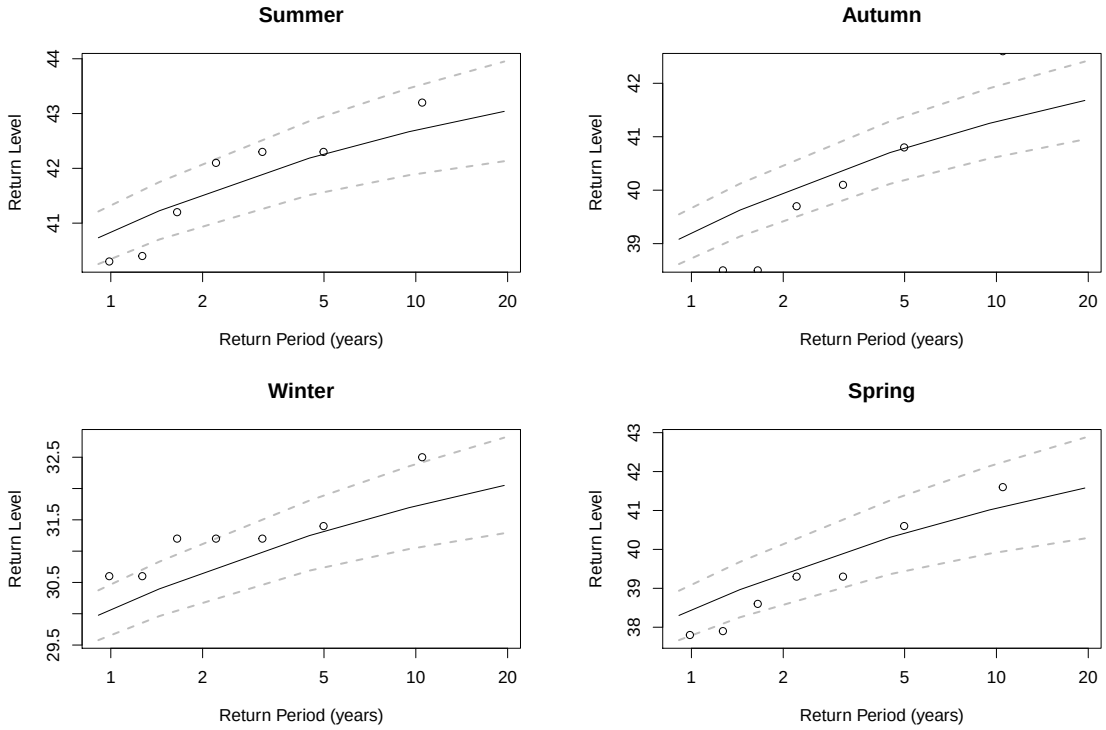


Figure A.55: Return level plots for temperature at Langebaanweg

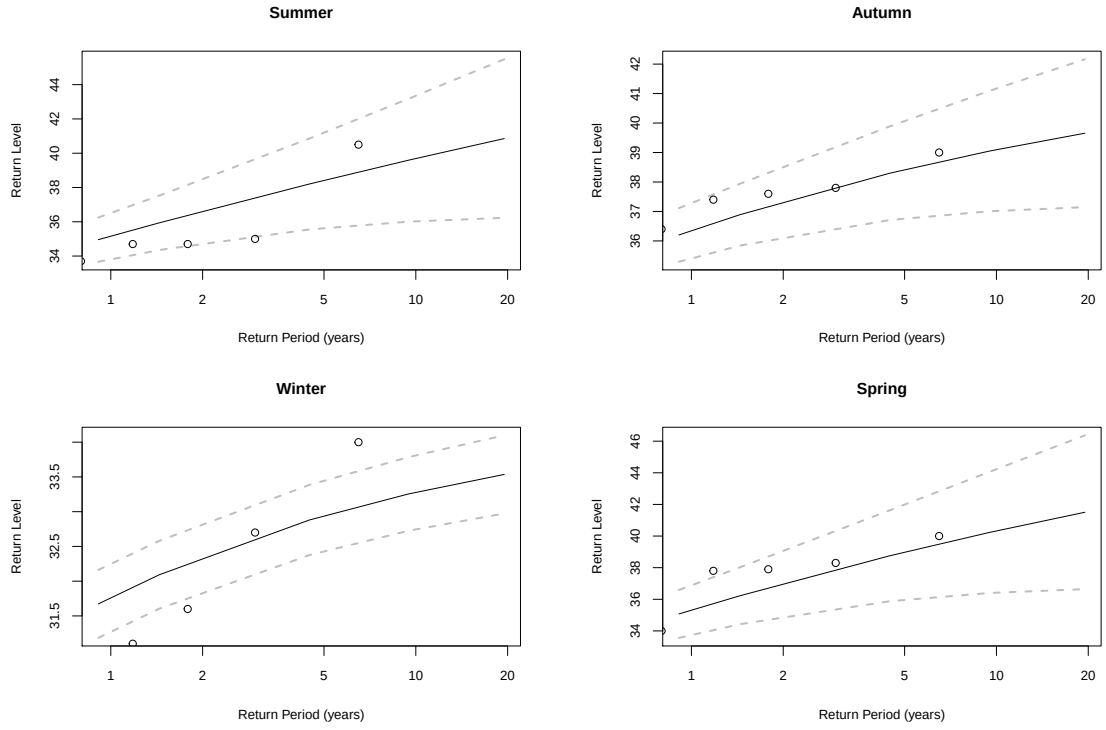


Figure A.56: Return level plots for temperature at Plettenberg Bay

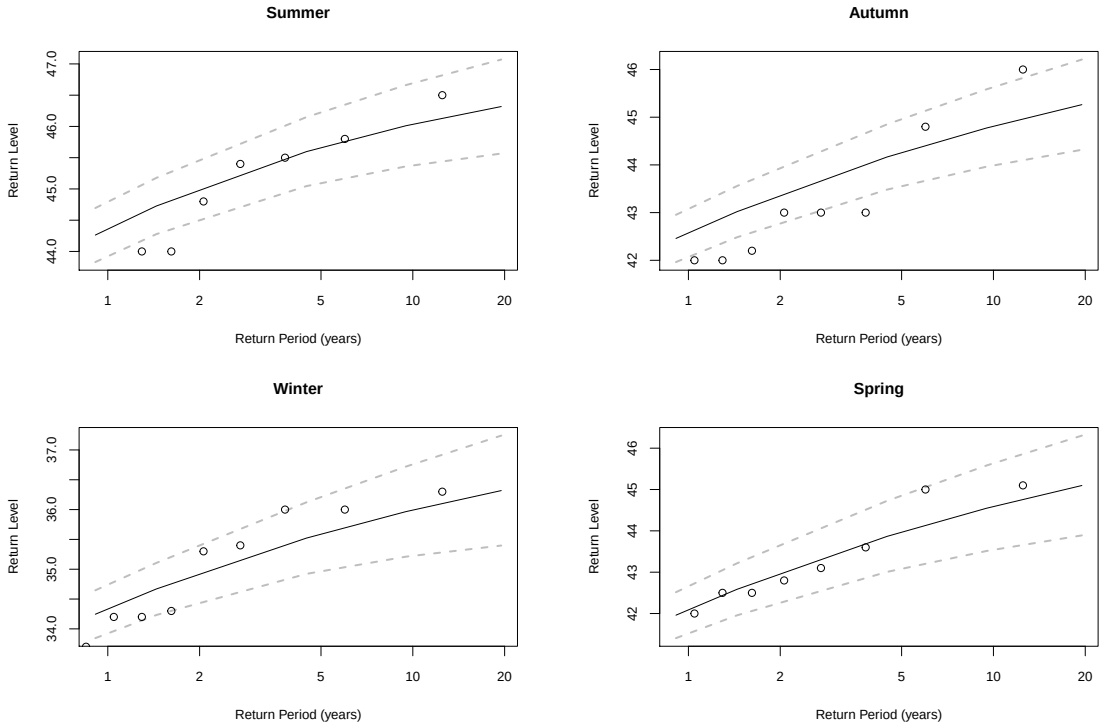


Figure A.57: Return level plots for temperature at Vredendal

Quantile-Quantile Plots

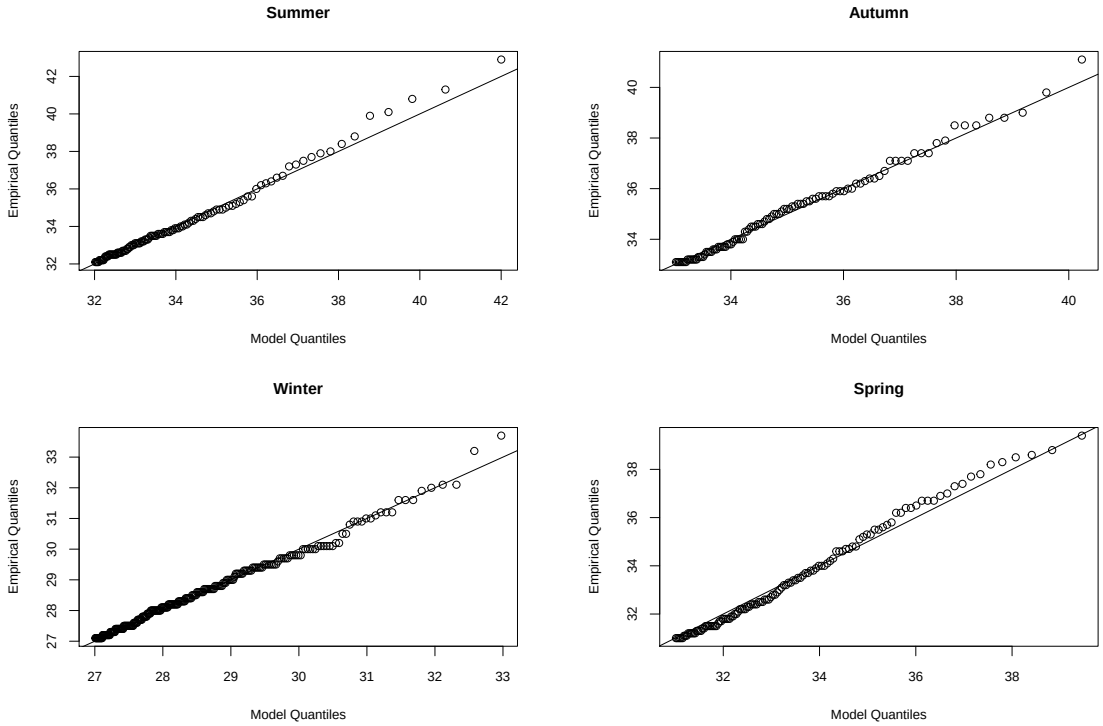


Figure A.58: Quantile-quantile plots for temperature at George Airport

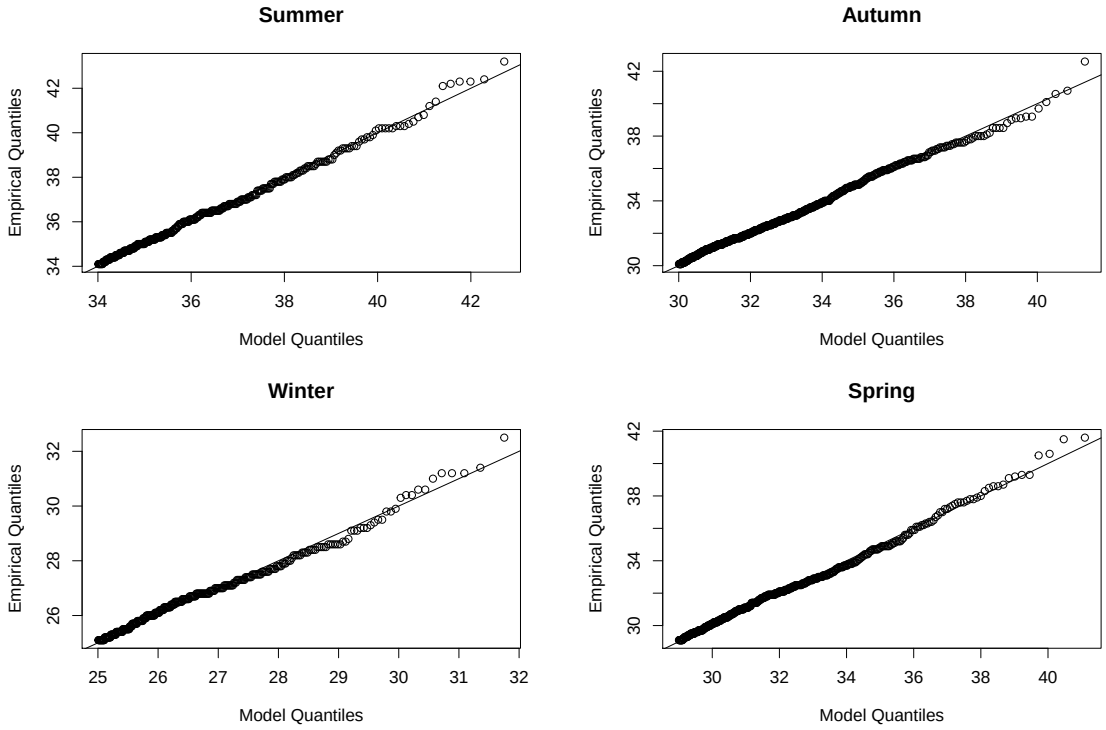


Figure A.59: Quantile-quantile plots for temperature at Langebaanweg

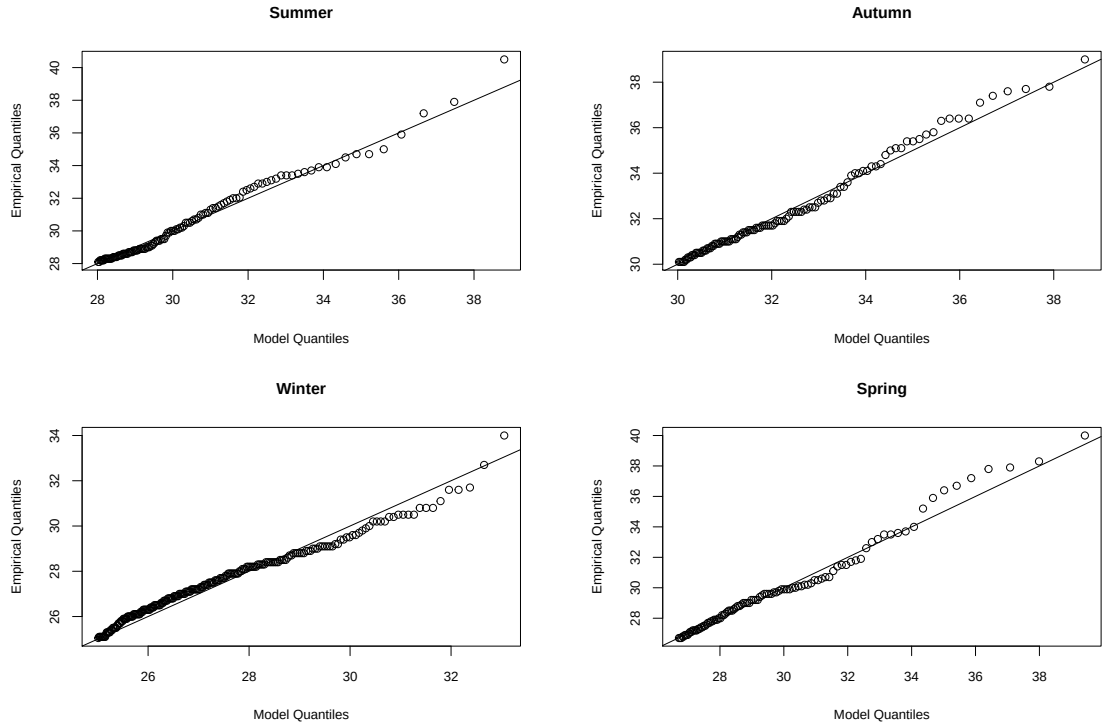


Figure A.60: Quantile-quantile plots for temperature at Plettenberg Bay

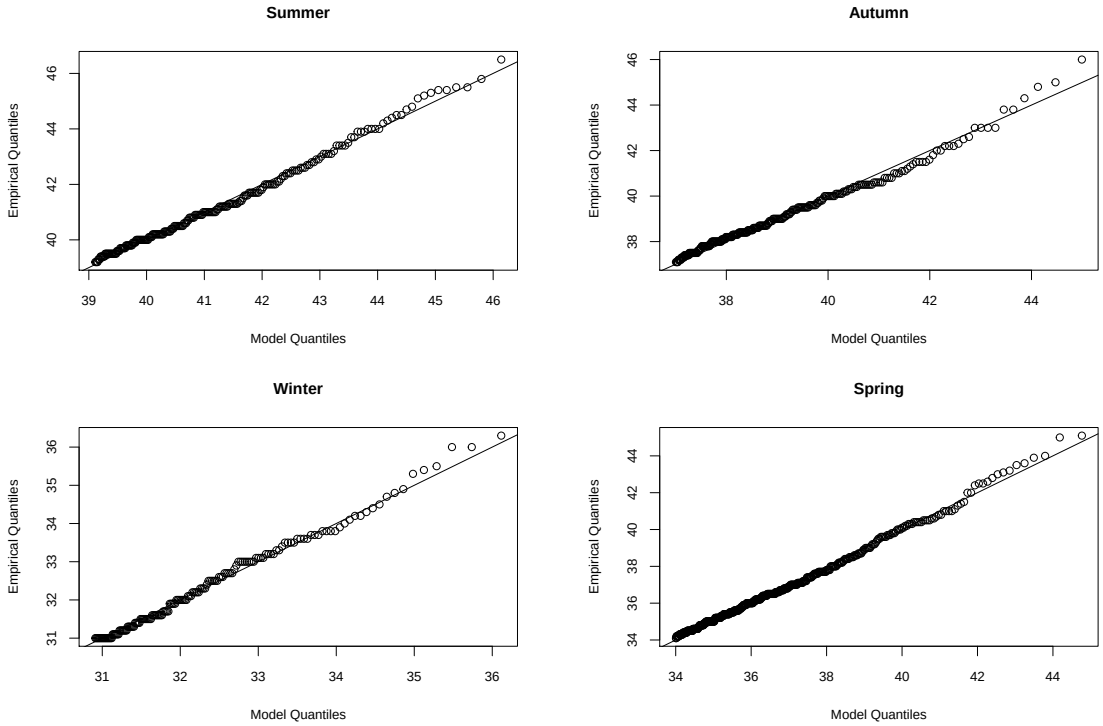


Figure A.61: Quantile-quantile plots for temperature at Vredendal

A.7.3 Maximum Wind Speed

Return Level Plots

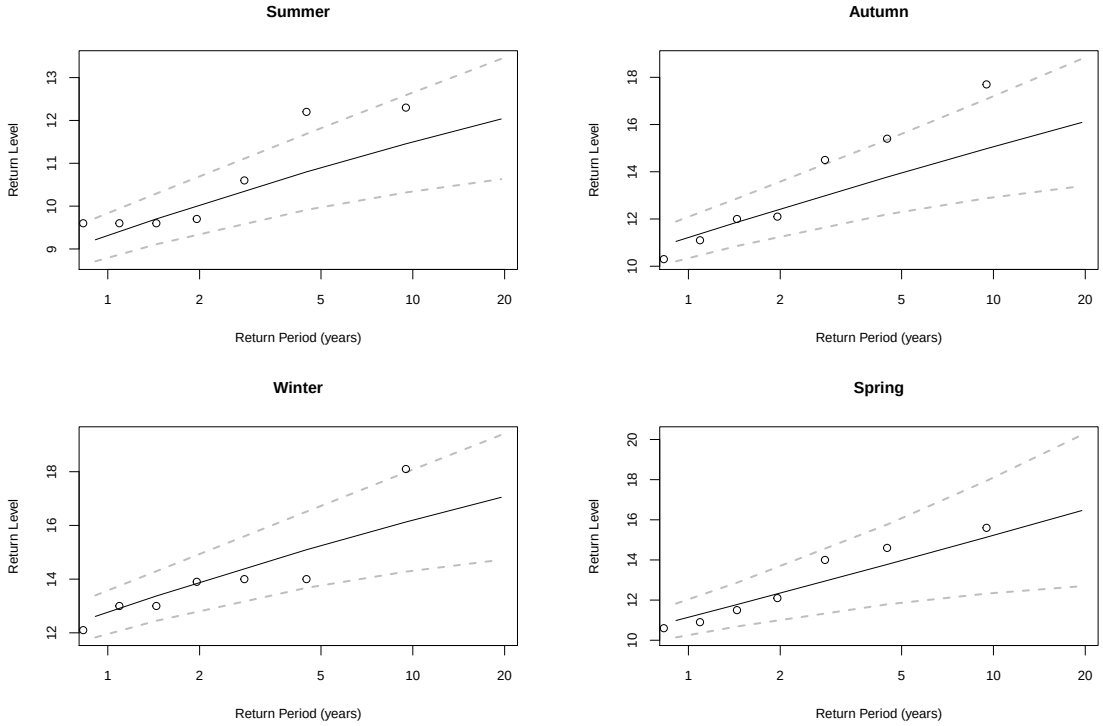


Figure A.62: Return level plots for wind speed at George Airport

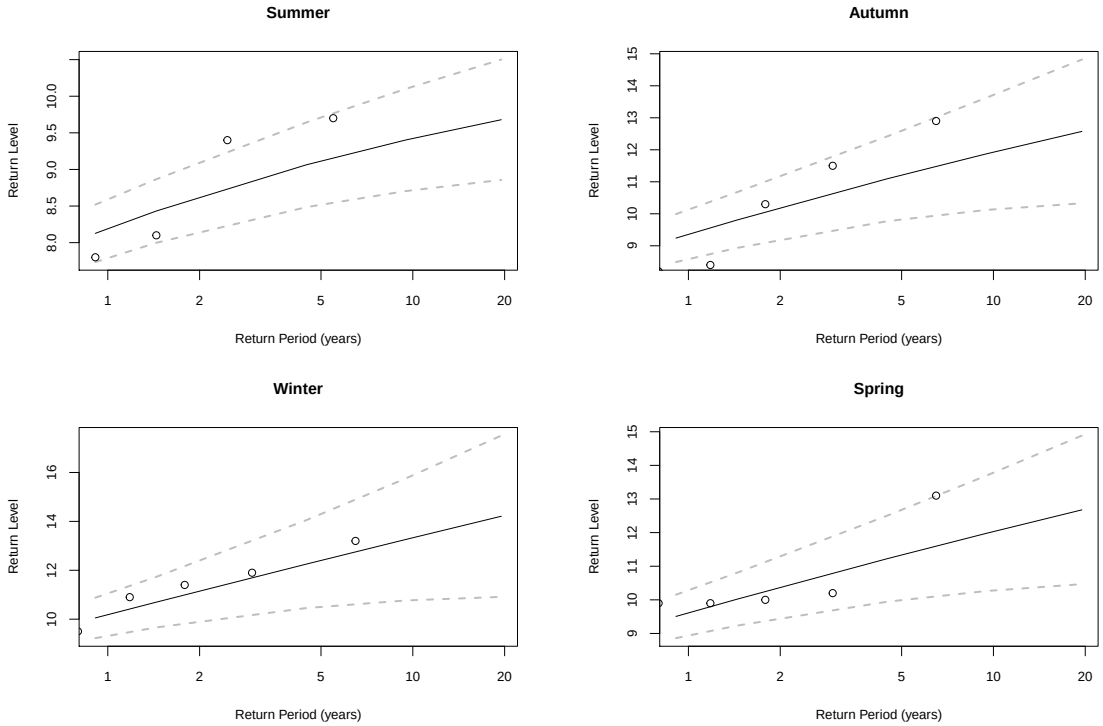


Figure A.63: Return level plots for wind speed at Plettenberg Bay

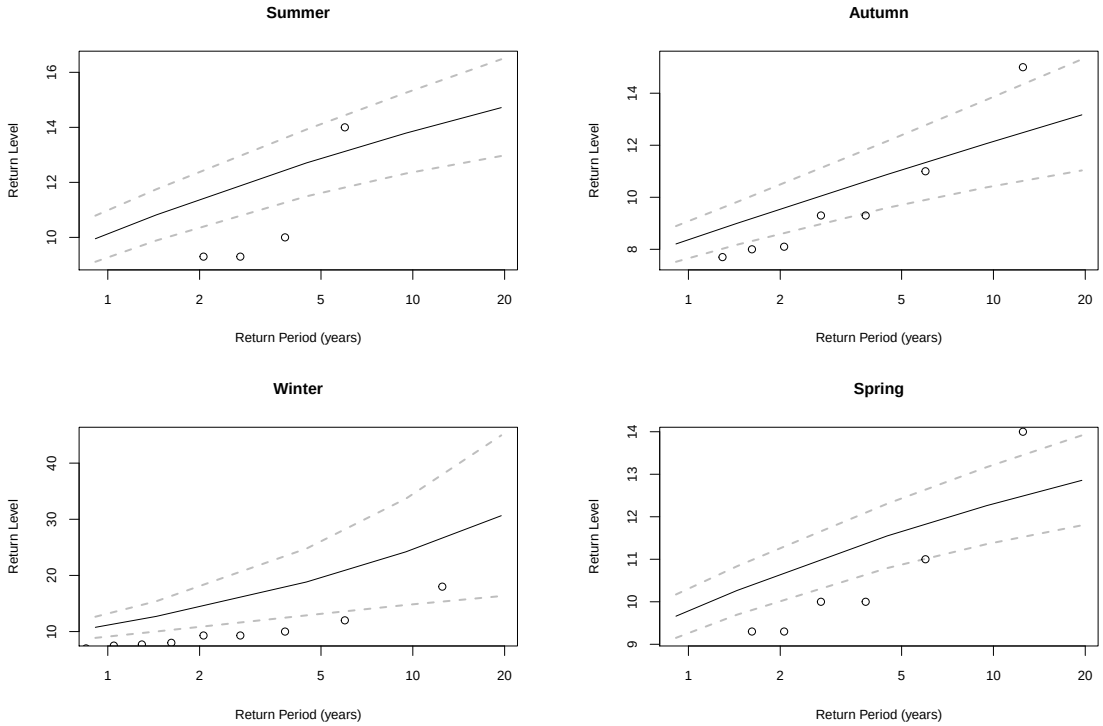


Figure A.64: Return level plots for wind speed at Vredendal

Quantile-Quantile Plots

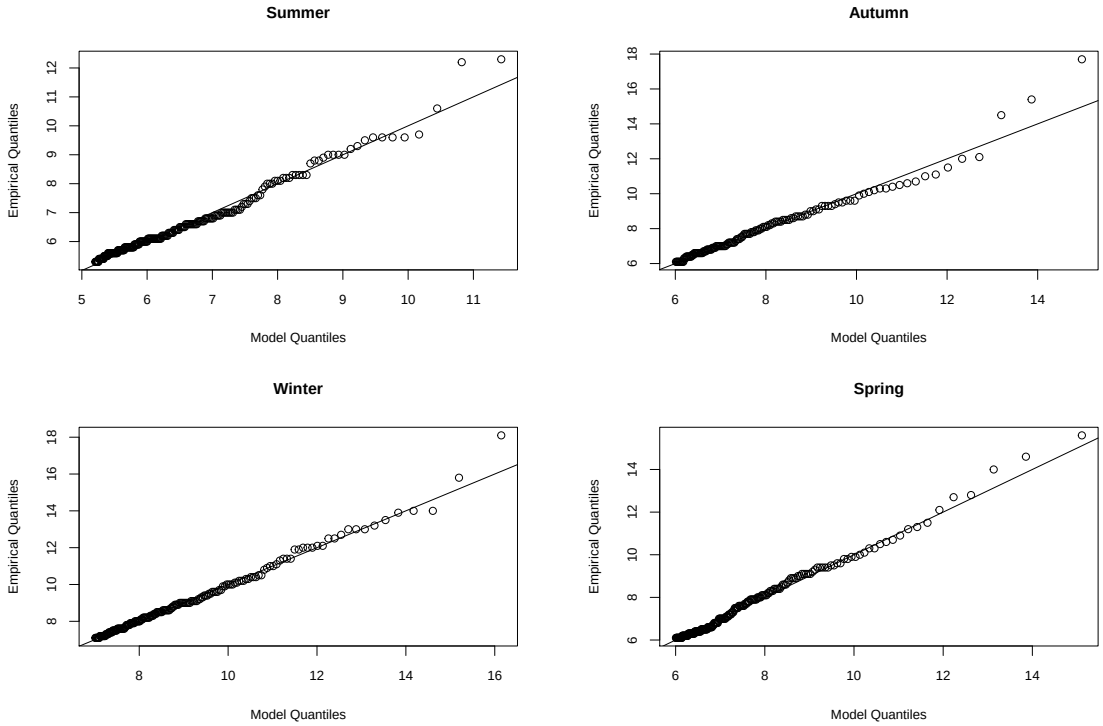


Figure A.65: Quantile-quantile plots for wind speed at George Airport

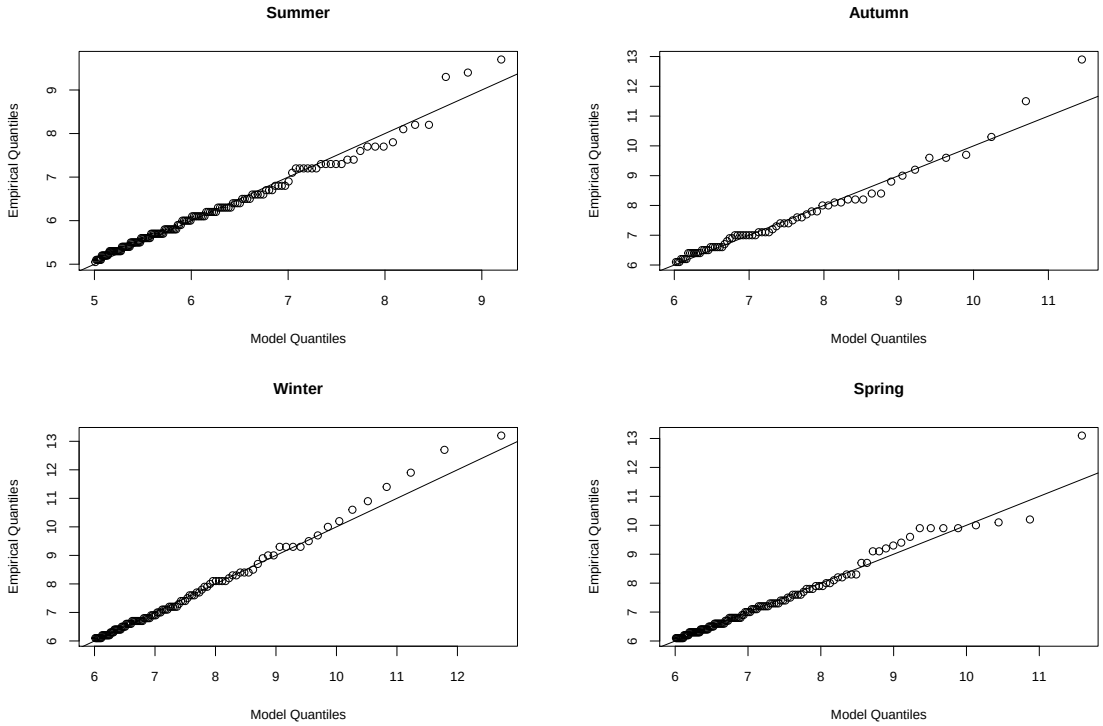


Figure A.66: Quantile-quantile plots for wind speed at Plettenberg Bay

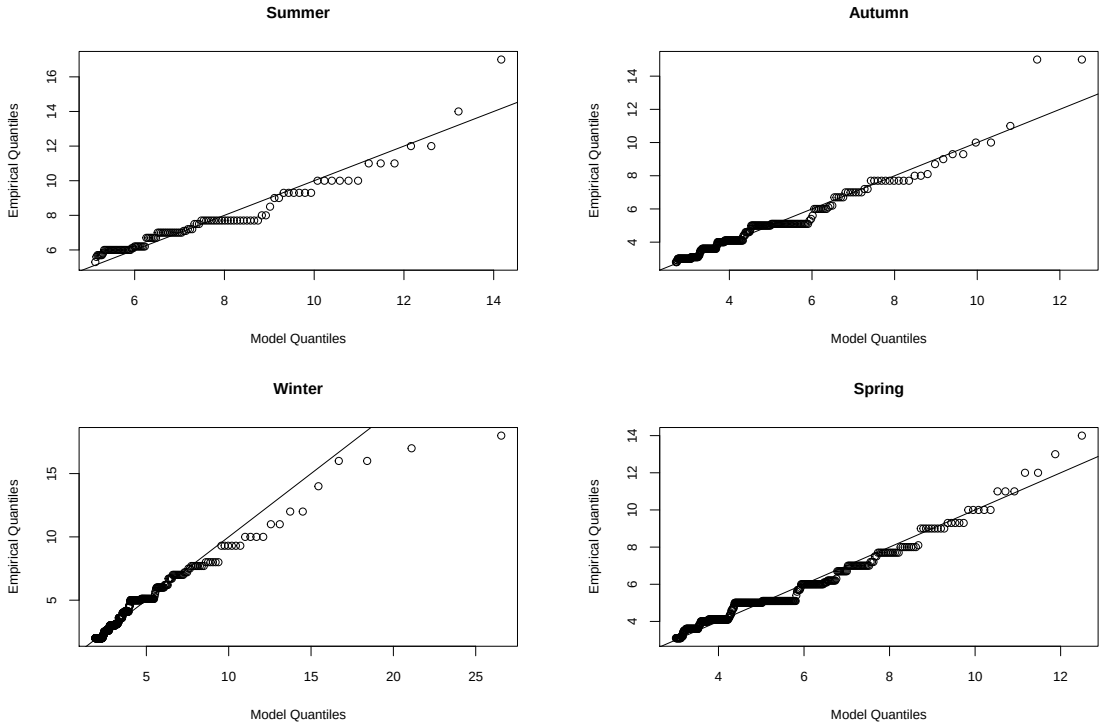


Figure A.67: Quantile-quantile plots for wind speed at Vredendal

Bibliography

Altun, E. and Tatlidil, H. (2015). A comparison of extreme value theory with heavy - tailed distributions in modeling daily var, *Journal of Finance and Investment Analysis* **4**(2).

Anderson, C. and Nadarajah, S. (1993). Environmental factors affecting reservoir safety, *Statistics for the Environment* pp. 163–182.

Araujo, J. A., Abiodun, B. J. and Crespo, O. (2016). Impacts of drought on grape yields in western cape, south africa, *Theoretical and Applied Climatology* **123**(1): 117–130.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.* **2**(5): 792–804.

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D. and Ferro, C. (2004). *Statistics of Extremes*, John Wiley & Sons.

Beirlant, J., Matthys, G. and Dierckx, G. (2001). Heavy-tailed distributions and rating, *ASTIN Bulletin* **31**(1): 3758.

Bortot, P., Coles, S. G. and Tawn, J. A. (2000). The multivariate Gaussian tail model: an application to oceanographic data., pp. 31–49.

Castillo, E., Hadi, A., Balakrishnan, N. and Sarabia, J. (2005). *Extreme value and related models with applications in engineering and science.*

Chikobvu, D. and Chifurira, R. (2015). Modelling of extreme minimum rainfall using generalised extreme value distribution for zimbabwe, *South African Journal of Science* **111**(9-10): 01–08.

Chikobvu, D., Sigauke, C. and Verster, A. (2012). Winter peak electricity load forecasting in south africa using extreme value theory with a bayesian flavour, *Journal of Business and Economics* **3**(5): 380–389.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, 4th edn, Springer London, London.

Coles, S. G. and Tawn, J. A. (1990). Statistics of coastal flood prevention, *Philosophical Transactions: Physical Sciences and Engineering* **332**(1627): 457–476.

Coles, S. G. and Tawn, J. A. (1991). Modelling Extreme Multivariate Events, *Journal of the Royal Statistical Society. Series B* **53**(2): 377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes - an application to structural design., **43**(1): 1–48.

Coles, S. G. and Walshaw, D. (1994). Directional Modelling of Extreme Wind Speeds, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **43**(1): 139–157.

Coles, S., Heffernan, J. and Tawn, J. (1999). Dependence Measures for Extreme Value Analyses, *Extremes* **2**(4): 339–365.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with comments), *Journal of the Royal Statistical Society, Series B: Methodological* **52**: 393–442.

de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*, Mathematical Centre tracts, Mathematisch Centrum.

de Haan, L. (1971). A form of regular variation and its application to the domain of attraction of the double exponential distribution, *Probability Theory and Related Fields* **17**(3): 241–258.

de Haan, L. (1976). *Sample Extremes: An Elementary Introduction*, Report: Econometrisch Instituut, Econometric Institute, Erasmus University.

de Haan, L. (1985). Extremes in higher dimensions: The model and some statistics (stma v27 891), pp. 1–15.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **40**(4): 317–337.

Dixon, M. J. and Tawn, J. A. (1995). A semi-parametric model for multivariate extreme values, *Statistics and Computing* **5**(3): 215–225.

Dobson, A. J. and Barnett, A. (2008). *An introduction to generalized linear models*, CRC press.

Enders, W. (2004). *Applied Econometric Time Series*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, J. Wiley.

Falk, M. and Marohn, F. (1993). Von mises conditions revisited, *Annals of Probability* **21**: 1310–1328.

Finkenstadt, B. and Rootzen, H. (2003). *Extreme Values in Finance, Telecommunications, and the Environment*, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample, *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(02): 180–190.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum, *Annales de la societe Polonaise de Mathematique* **6**: 93–116.

Friederichs, P. (2010). Statistical downscaling of extreme precipitation events using extreme value theory, *Extremes* **13**(2): 109–132.

Galambos, J. (1975). Order statistics of samples from multivariate distributions, *Journal of the American Statistical Association* **70**(351): 674–680.

Galambos, J., Lechner, J. and Simiu, E. (1993). *Extreme Value Theory and Applications*, KLUWER ACADEMIC PUBLISHERS.

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, *The Annals of Mathematics* **44**(3): 423+.

Gumbel, E. (1958). *Statistics of Extremes*, New York, Columbia University Press.

Gumbel, E. J. (1960). Bivariate exponential distributions, *Journal of the American Statistical Association* **55**(292): 698–707.

Gumbel, E. J. and Goldstein, N. (1964). Analysis of empirical bivariate extremal distributions, *Journal of the American Statistical Association* **59**(307): pp. 794–816.

Hasan, H. B., Fadhilah, N., Radi, B. A. and Kassim, S. B. (2012). Modeling of Extreme Temperature Using Generalized Extreme Value (GEV) Distribution : A Case Study of Penang, **I**: 4–9.

Hasan, H., Salam, N. and Adam, M. B. (2013). Modelling Extreme Temperature in Malaysia Using Generalized Extreme Value Distribution, **7**(6): 435–441.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution, *Technometrics* **29**(3): 339–349.

Hugueny, S., Clifton, D. a. and Tarassenko, L. (2009). Novelty detection with multivariate extreme value theory, part II: An analytical approach to unimodal estimation, *Machine Learning for Signal Processing XIX - Proceedings of the 2009 IEEE Signal Processing Society Workshop, MLSP 2009* .

Joe, H. (1987). Estimation of quantiles of the maximum of n observations, *Biometrika* **74**: 347–354.

Joe, H. (1989). On fitting multivariate extreme value distributions, with applications to environmental data, *Technical report*, SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY.

Joe, H., Smith, R. L. and Weissman, I. (1992). Bivariate threshold methods for extremes, *Journal of the Royal Statistical Society, Series B: Methodological* **54**: 171–183.

Katz, R. W. and Grotjahn, R. (2011). Statistical methods for relating temperature extremes to Large-Scale Meteorological Patterns, pp. 4–7.

Katz, R. W., Parlange, M. B. and Naveau, P. (2002). Statistics of extremes in hydrology, *Advances in Water Resources* **25**(8-12): 1287–1304.

Khuluse, S. (2010). Modelling heavy rainfall over time and space.

Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions*, Published By Imperial College Press and Distributed By World Scientific Publishing Co.

Leadbetter, M., Lindgren, G. and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*, Springer series in statistics, Springer-Verlag.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values., pp. 169–187.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of empirical finance* **7**(3): 271–300.

Mikosch, T. V., Resnick, S. I. and Robinson, S. M. (2006). *Springer Series in Operations Research and Financial Engineering*.

Morton, I. and Bowers, J. (1996). Extreme value analysis in a multivariate offshore environment, *Applied Ocean Research* **18**(6): 303–317.

Nadarajah, S. and Choi, D. (2007). Maximum daily rainfall in South Korea, *Journal of Earth System Science* **116**(4): 311–320.

Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.* **3**(1): 119–131.

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika* **67**: 723–724.

Qin, X., Smith, R. L. and Ren, R. (2009). A new class of multivariate survival distributions.

Rakonczai, P. (2009). On Modeling and Prediction of Multivariate Extremes, *Mathematical Statistics Centre for Mathematical* p. 109.

Reiss, R. and Thomas, M. (2007). *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*, Birkhäuser Basel.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, Springer-Verlag.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT - Statistical Journal* **10**(1): 33–60.

Sibuya, M. (1960). Bivariate extreme statistics, i, *Annals of the Institute of Statistical Mathematics* **11**(2): 195–210.

Singh, A. K., Allen, D. E. and Robert, P. J. (2013). Extreme market risk and extreme value theory, *Mathematics and Computers in Simulation* **94**: 310–328.

Smith, R. L. (1987). Approximations in extreme value theory., *Technical report*, DTIC Document.

Smith, R. L. (1989). Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone, *Statistical Science* **4**(4): 367–377.

Smith, R. L. (2004). Statistics of extremes, with applications in environment, insurance, and finance, *Monographs on Statistics and Applied Probability* (March): 1–62.

Smith, R., Tawn, J. and Yuen, H. (1990). Statistics of multivariate extremes, *International Statistical Review/Revue Internationale de Statistique* pp. 47–58.

Southern African Development Community, S. (2012). Climate change adaptation.

Statistics, M. and Wiley, J. (1950). * Reproduced from "Contributions to Mathematical Statistics" (1950) by permission of John Wiley and Sons, Inc.

Stephenson, A., Tawn, J. and Stephenson, B. Y. A. (2005). Biometrika trust exploiting occurrence times in likelihood inference for componentwise maxima, **92**(1): 213–227.

Tawn, J. A. (1988). Bivariate extreme value theory: Models and estimation, **75**(3): 397–415.

Tawn, J. A. (1990). Modelling multivariate extreme value distributions, *Biometrika* **77**(2): 245–253.

Tawn, J. A. and Heffernan, J. E. (2004). A conditional approach to modelling multivariate extreme values (with discussion)., pp. 497–546.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica: Journal of the Econometric Society* pp. 307–333.

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

Zhang, L. and Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean copulas, *Journal of Hydrology* **332**(1-2): 93–109.

Zheng, F., Westra, S., Leonard, M. and Sisson, S. A. (2014). Modeling dependence between extreme rainfall and storm surge to estimate coastal flooding risk, *Water Resources Research* pp. 2050–2071.

Ziervogel, G., New, M., Archer van Garderen, E., Midgley, G., Taylor, A., Hamann, R., Stuart-Hill, S., Myers, J. and Warburton, M. (2014). Climate change impacts and adaptation in south africa, *Wiley Interdisciplinary Reviews: Climate Change* **5**(5): 605–620.