The Influence of Climate Change on Short Duration Rainfall in the Western Cape

by

Gysbertus Johannes Burger

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Engineering

at the

Faculty of Engineering, Stellenbosch University

Supervisor: Dr J.A. du Plessis
Department of Civil Engineering

December 2012

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature:..... Date:.....

Copyright © 2012 Stellenbosch University
All rights reserved

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Opsomming

The Wes-Kaap het die afgelope dekade deur uiterse storms deurgeloop wat miljone Rande se skade aan publieke infrastruktuur en nedersettings veroorsaak het. Huidige klimaatsmodel resultate wat klimaatsverandering inkorporeer, beraam dat die Wes-Kaap hoër reënvalintensiteite kan verwag, waar die grootte van die en reënvalgebeurtenisse toeneem, maar die frekwensie van die reënvalgebeurtenisse afneem. Die doel van hierdie navorsing is om te bepaal of historiese kortduurtereënvaldata in die Wes-Kaap enige toenemende reënvalintensiteite aandui.

Kortduurtereënvaldata ($<$ 24 uur) wat uit versyferde autografiese- en automatiese weerstasie (AWS) data bestaan, was uit die Suid Afrikaanse Weerdienste se databasis gekies vir stasies in die Wes-Kaap en die wyer Suid Afrikaanse gebied, sodoende om die effektiewe rekordlengte te verleng. Redigering- en kwaliteitsbeheerprosudures moes toegepas word na ’n aantal probleme met die data ontdek was. Die versyferde autografiese data het baie foute bevat en oor die algemeen swak vergelyk met die standaard daaglikse totaalmetings van dieselfde periode. Na die redigering- en kwaliteitsbeheerprosudures toegepas was, was sewe stasies gekies en hul autografiese en AWS data was saamgestel vir verdere analise.

Die analise het uit twee afdelings bestaan: die analise van die grootte van die reënvalgebeurtenisse, en die frekwensie van die reënvalgebeurtenisse. Nie-stasionêre ekstreemwaardeteorie was toegepas op die grootte van die reënvalgebeurtenisse deur gebruik te maak van parametriese en nie-parametriese nie-stasionêre metodiek. Die veralgemeende ekstreemwaardeverdeling (AEW) met ’n jaarlikse maksimum reeks, en die veralgemeende Paretoverdeling met ’n pieke bo drempelwaarde reeks was gebruik. Die parametriese metode het bestaan uit die passing van lineêre modelle aan die parameters van die estreemwaardeverdelings. Die nie-parametriese metode het getoets of die terugkeervlakke van die verdelings konstant gebly het oor ’n bewegende vensterperiode. The frekwensie analise het bestaan uit die aantekening van die jaarlikse reënvalgebeurtenisse wat bepaalde drempelwaardes oorskry.

Die parametriese metode het net twee beduidende nie-stasionêre staties opgelewer, uitsluitlik vir die AEW. Die nie-parametriese metode het resultate gegenereer wat moontlike nie-stasionarieit aandui, maar dit is heel moontlik as gevolg van die samestelling van die autografies- en AWS data. Die frekwensie analise het óf geen sigbare neiging vir die resultate vir ván stasies gegenereer nie, óf het resultate gegenereer wat aandui op die samestelling van die autografies- en AWS data.

Dit het tot die gevolgtrekking gelei dat daar geen aanduiding in die kortduurtereënvaldata bestaan wat toenemende reënvalintensiteite ondersteun nie vir beide die grootte en frekwensie van reënvalgebeurtenisse.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Abstract

The Western Cape has been subject to extreme rainfall storms in the last decade, which has been responsible for millions of Rands worth of damage to public infrastructure and human settlements. Current climate model projections under climate change suggest that the Western Cape can expect greater rainfall intensities, with magnitude of rainfall events increasing, but the frequency of occurrence of events decreasing. The aim of this research is to assess if historical short duration rainfall data provides any evidence to support increasing intensities in rainfall in the Western Cape.

Short duration rainfall data ($<$ 24 hours), consisting of digitised autographic- and automatic weather station data (AWS), was selected from the South African Weather Services' database for the Western Cape and wider South African region, in order to combine the data for an extension of the effective record length. Numerous difficulties were encountered with the data that required the application of editing and quality control procedures. The digitised autographic data contained many errors and generally compared poorly to standard gauge daily rainfall totals of the same period. After the application of editing and quality control procedures, seven stations were selected and their autographic and AWS data was combined for further detailed analysis.

Analysis was divided into two sections, one for the analysis of the magnitude of the rainfall, the other for the frequency of occurrence of rainfall events. For the magnitude analysis, non-stationary extreme value theory was applied by implementing a parametric and non-parametric non-stationary approach to both the generalised extreme value distribution (GEV) with an annual maximum series, and a generalised Pareto distribution with a peaks over threshold series. The parametric approach entailed fitting linear models to the parameters of the extreme value distributions, and the non-parametric approach tested if return levels of the distributions remained constant over a moving window period. The frequency analysis entailed the recording of the annual number of rainfall events exceeding predetermined threshold values.

For the magnitude analysis, the parametric approach only produced two stations with significant non-stationarity and only for the GEV, whilst the non-stationary approach produced results that can indicate non-parametric behaviour, but this was most likely because of the combination of autographic and AWS data. The frequency analysis produced results that indicated no observable results for some stations, while others produced trends that are consistent with a change from the autographic to AWS data.

This lead to the conclusion that with the available data, no clear or significant evidence supporting increasing intensities or any other change in short duration rainfall was found regarding the magnitude and frequency of occurrence of rainfall events.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Acknowledgements

I would like to thank the following people and institutions for their aid in the completion of this thesis:

- My supervisor, dr Kobus du Plessis, for all his patience, advice and support over the last three years.
- Western Cape Disaster Management for their financial support for my studies.
- My parents for their continual love, support (both financially and emotionally) and encouragement that made studying at Stellenbosch possible.
- The South African Weather Services for supplying me with data that was essential for this thesis. I would like to thank Elsa de Jager in particular, who had to deal with my constant barrage of emails.
- Prof Jeff Smithers, who helped me obtain the digitised autographic data and who patiently dealt with my requests and questions.
- Dr Tom Berning, who made the complex world of extreme value theory a bit more comprehensible.
- Victoria Goodall from SAEON, who suggested that I use the *R* programming environment for analyses.
- My good friends Wessel Venter and Bertus Basson, who patiently and graciously helped me with many programming issues I encountered.
- The vast number of unknown contributors on online forums and websites who answer questions for the sake of answering. Without the use of the internet, this thesis would not have been possible.
- God for giving me breath and showing me grace.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Contents

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

List of Tables

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

List of Figures

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

List of Symbols

α	-	confidence level
β_i	-	approximating parameters for a linear model
D	-	deviance statistic
$\exp()$	-	natural exponent of an amount
$F(x)$	-	cumulative density function for a variable x
k	-	difference in parameters between a more and less complicated distribution
l_i	-	log-maximum likelihood i
M_i	-	extreme value model i
μ	-	location parameter
$\mu(t)$	-	time dependent location parameter
n	-	record length
ξ	-	shape parameter
$\xi(t)$	-	time dependent shape parameter
R^2	-	coefficient of determination
σ	-	scale parameter
$\sigma(t)$	-	time dependent scale parameter
t	-	time variable
$t_{\alpha/2,n-2}$	-	t-distribution quantile for α confidence level and $n-2$ degrees of freedom
t_0	-	test statistic for a t-test
u	-	threshold value of generalised Pareto distribution
u_d	-	Du Plessis (1992) threshold values
$u_{1.5d}$	-	$1.5\times$ Du Plessis (1992) threshold values
u_{stb}	-	stable shape parameter threshold values
$\chi^2_{\alpha,k}$	-	α quantile for a chi-square distribution with k degrees of freedom

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Abbreviations

AES	-	annual exceedance series
AMS	-	annual maxima series
AWS	-	automatic weather station
EBM	-	energy balance model
ENSO	-	El Niño-Southern Oscillation
EVT	-	extreme value theory
GCM	-	global circulation model
GEV	-	generalised extreme value distribution
GPD	-	generalised Pareto distribution
IDF	-	intensity duration frequency
ML	-	maximum likelihood
NPNS	-	non-parametric non-stationary
PDS	-	partial duration series
PNS	-	parametric non-stationary
POT	-	peaks over threshold
PWM	-	probability weighted moments
RCM	-	regional circulation model
SAWS	-	South African Weather Services
SDR	-	short duration rainfall
SG	-	standard gauge daily rainfall

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Chapter 1

Introduction

The Western Cape has been host to a number of intense storms in the last ten years, causing deaths, displacement and widespread destruction to property and infrastructure (Holloway, Fortune & Chasi, 2010). The damage inflicted can be attributed to the failure of the infrastructure to deal with large amounts of runoff generated by the storms. Stormwater systems play an important role in draining storm-induced runoff, but are often designed on the basis of the homogeneity of the climate.

Developments in climate sciences since the 1990s suggest that the occurrence of extreme and variable weather phenomena like droughts and floods could be the result of climate change. Climate change refers to the increasing concentration of greenhouse gases in the atmosphere because of the burning of fossil fuels, mostly as a result of rapid industrial development. Climate change is expected to cause changes to the hydrological cycle, pushing the characteristics of rainfall, like the frequency of occurrence and the amount or magnitude of rainfall towards extremes. South Africa is not expected to be exempt from the effects of extreme weather, but can expect prevailing variable conditions to become more variable.

These conditions are a matter of concern for engineering design, as existing infrastructure could very likely be under-designed. The aim with this research is to investigate the possible influence of climate change on rainfall in South Africa, focussing on the Western Cape. As the design of stromwater systems may require designs for shorter storm durations, short duration rainfall is used for analysis. Digitised autographic- and five-minute rainfall data are combined to extend the effective record length, and analysed to assess if any trends in the magnitude and frequency of occurrence of rainfall events can be noted.

The research question of this thesis is:

- Are there any indications of increasing rainfall intensities in current short duration rainfall in the Western Cape and wider South Africa that support corresponding climate change projections?

Stellenbosch University http://scholar.sun.ac.za

This question will be addressed as follows in this thesis:

In Chapter 2 the concept of climate change and the methods involved in making projections regarding the future climate will be discussed. Studies on global and South African rainfall patterns are also explored, both for observed and projected data, so as to assess the possible impacts of climate change on rainfall. The state of short duration rainfall in South Africa is also investigated, including measurement methods and current research on the subject. Finally, developments in extreme value theory, commonly used in engineering design, are presented, including different distribution types and methods to analyse non-stationary data.

In Chapter 3 the methodology followed when analysing short duration rainfall for evidence of changing intensities, including the initial editing of rainfall stations and the analysis procedures to investigate changes in the magnitude and frequency of occurrence of rainfall events will be outlined, while in Chapter 4 the results of the application of the procedures in Chapter 3 will be explored.

Finally, in Chapter 5 the results of the analysis will be summarised and conclusions from the results will be drawn and recommendations for future research made.

Stellenbosch University http://scholar.sun.ac.za

Chapter 2

Literature review

As this thesis is concerned with investigating possible climate change effects on South African rainfall, it is essential to first understand what climate change is, secondly, the impacts it is expected to have on rainfall, and lastly, approaches or methods that can measure these changes.

In this chapter literature will be reviewed. The main objective is to illustrate the following:

1. That climate change is a very broad term that is not limited to global warming and greenhouse gas emissions only. This is done by first looking at the earth system and how it affects the climate, then at the various factors that cause the climate to change.

2. That the expected impacts of climate change on rainfall are largely uncertain, specifically in South Africa and the Western Cape. Climate and circulation models are explored as they play an important role in the making of future projections regarding the climate. Literature on rainfall is investigated, looking at the theoretical impacts of climate change, and then investigating global and local scale studies for both observed and projected data. An overview of short duration rainfall is given, focussing on different measurement types and current research on short duration rainfall.

3. That developments in extreme value theory enable the evaluation of trends and changes in rainfall. Alternative methodologies to those commonly used in South African storm rainfall design are explored, including peaks over threshold methodology and non-stationary statistics.

2.1 The earth system and climate

Climate is defined as the long term average or prevailing conditions in weather for a specific region or area (*The Concise Oxford Dictionary*, 1989). The climate is driven by a complex

Stellenbosch University http://scholar.sun.ac.za

array of chemical, physical and biological processes and cycles within the earth system that are in constant interaction with each other.

The most important input of energy, driving the climate and making the earth habitable for life, is the sun. Solar energy enters the earth in the form of short-wave solar radiation and is absorbed by the ground surface and atmosphere and then transformed into heat. Heat is transferred and exchanged in the climate system as thermal long-wave radiation which heats up the earth's surface (O'Hare, Sweeny & Wilby, 2005:58).

The atmosphere aids in heating the earth's surface by preventing some of the thermal radiation from escaping into space. Gases in the atmosphere like water vapour, carbon-dioxide and methane, absorb thermal radiation and re-radiate it to the surface. Thus the atmosphere acts as an insulator of heat, keeping the earth warm. This process of absorption and re-radiation of the atmosphere is called the greenhouse effect (O'Hare et al., 2005:59).

The interaction and exchanges of energy in the earth is illustrated in Figure 2.1, showing the earth's energy balance. Any perturbations in this system will cause readjustments in order to accommodate the changes which will inevitably cause some form of change in the climate.

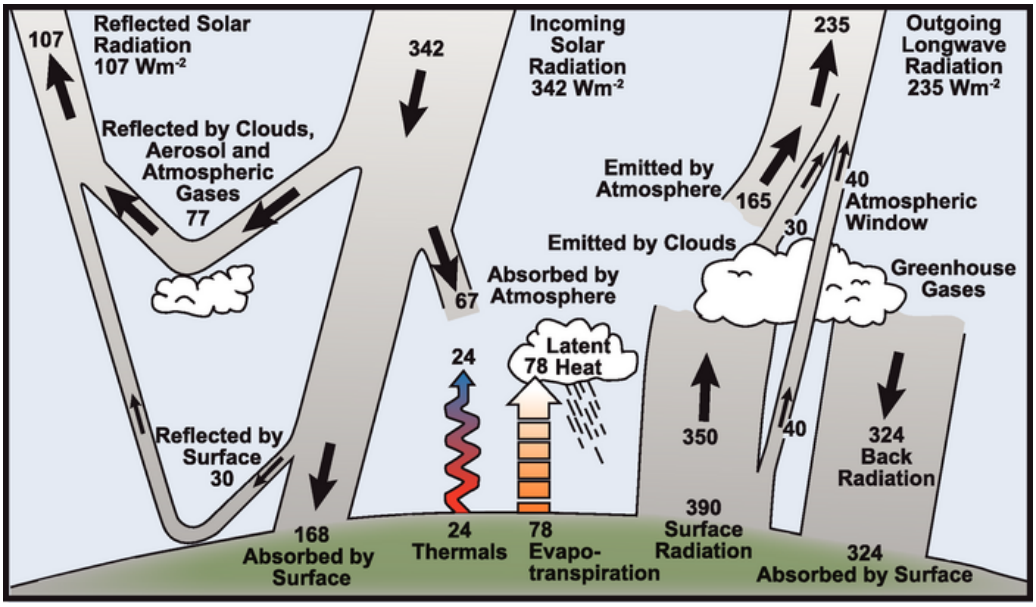


Figure 2.1: Energy balance of the earth (Solomon, Qin, Manning, Chen, Marquis, Averyt, Tignor & Miller, 2007:94).

2.2 Climate change

Given the definition of climate discussed in Section 2.1, climate change implies any deviance from prevailing weather conditions in a region. This definition is problematic, as paleo records of the earth's history show that the earth's climate has never remained constant, but has undergone large scale and periodical changes, like the ice ages and inter-galcial periods. However,

Stellenbosch University http://scholar.sun.ac.za

trends in the deviations from the general climate conditions during the last century have brought into question whether the change in climate is occurring at a normal or accelerated rate.

Over the last hundred years, global mean temperatures have increased by 0.74°C, with the rate of warming over the last 50 years being almost double that of the last hundred years. During the twelve years from 1995 to 2006, eleven years have been reported to be the warmest since recording started in 1850 (Solomon et al., 2007:237).

Sea levels have also been rising and ice sheets in the Northern Hemisphere have been shrinking. Furthermore, increases in extreme weather events such as droughts, heavy precipitation and heat waves have been observed. Some of the largest decreases in precipitation were found to occur over Western Africa, while a drying trend has been observed in Southern Africa since 1901 (Solomon et al., 2007:238). These observed changes can pose threats for current infrastructure development which will have severe socio-economic implications on the population.

What factors are probably responsible for causing changes like this in the climate?

2.2.1 Causes of climate change

Changes in climate are directly related to changes in energy. When the available energy at the earth's surface undergoes change, the climate consequently changes either by internal or external means. External means refer to the input of solar radiation and how it is distributed, while internal inputs consist of how atmospheric composition, the land/ocean/ice surface, biosphere and human activities store, distribute and release energy.

2.2.1.1 Changes in solar energy

Changes in solar energy are a result of the physical energy output of the sun and how the earth interacts with the sun. Solar energy changes are the most important factors that cause changes in the climate, since the sun is the primary source of energy for all activity on earth.

One of the ways in which solar variation can influence the climate directly is through changes in the energy output from the sun. The sun's energy output does not remain constant but shows some variability. On the surface of the sun, large, cool patches, called sunspots, are found which cause variability in the output of solar energy. The number of sunspots show large variability over time, with periods of almost no sunspots during the years 1650-1700, to as many as 300 during the 1950s. The number of sunspots seems to vary in 11, 22 and 80-90 year cycles (O'Hare et al., 2005:77).

The effect of these sunspot cycles on global climate has been investigated in a number of studies, some showing correlation between sunspot activity and temperature variation, but there is currently fierce debate about these matters (Legras, Mestre, Bard & Yiou, 2010:745) and conclusions are often very complex or reductionistic (Aly, 2010:1040,1041). There have also

Stellenbosch University http://scholar.sun.ac.za

been a number of studies that have indicated some connection between rainfall and river flow and sunspot cycles. Hiremath and Mandi (2004) analysed Indian monsoon rainfall and found that the variability in the rainfall correlated positively with 2.7, 16 and 22 year periods of solar activity. Stager, Ryves, Cumming, Meeker and Beer (2005) analysed East African high lake levels and found some synchrony between these and prolonged sunspot minima. It was found that the levels of Lake Victoria rose during the peak of the 11-year sunspot cycle since the late 19th century.

In South Africa, Alexander, Bailey, Bredenkamp, Van der Merwe and Willemse (2007) found that there is strong evidence for the influence of double sunspot cycles (which are two sunspot cycles of approximately 21 year duration) on South African river flow, as an increase in sunspot activity caused temperature and precipitation increases, implying that rainfall and river flow is periodic or non-random, and influenced by solar activity. It is also mentioned that high flood events are associated with sunspot minima which is in agreement with the findings of Stager et al. (2005).

The influence of sunspot cycles on the climate are often questioned on the basis that solar output during the 20th century has only increased by 0.1 % of the solar constant which is seen to be too small to cause the current rise in the world's temperature. Rather, the increases in solar output is only judged to account for a quarter or third of observed global temperature increase (O'Hare et al., 2005:78).

Another reason why the effect of solar radiation on the earth could vary is the earth's orbital eccentricity around the sun. Orbital variations, known as Milankovitch cycles, are thought to be the driver of ice ages in the earth's history which have caused major changes in the average surface temperature. However the conditions that could drive the next ice age are only expected to be experienced 30 000 years from now (Solomon et al., 2007:112).

2.2.1.2 Land surface change and biogeochemical cycling

Land surface change includes long term processes like continental drift, the formation of mountains, and changes in ice cover and vegetation. Light coloured surfaces that reflect solar energy have a cooling effect on the earth's surface, while darker surfaces absorb solar energy which in turn warms the earth surface. Thus any changes in the land cover can either have warming or cooling effects. An example of this is the melting of polar ice caps like the Greenland ice sheet which causes a reduction in the highly reflective ice surface which can lead to more solar energy being absorbed by the earth's surface (O'Hare et al., 2005:76).

Biogeochemical cycling involves the cyclic exchange of materials between living organisms and non-living organisms. An example is carbon rich soils that release greenhouse gases under organic decomposition which can further enhance the warming of the earth's surface (O'Hare et al., 2005:76).

Stellenbosch University http://scholar.sun.ac.za

2.2.1.3 Internal energy cycling

The earth has many internal energy mechanisms that are in constant flux, including cycles in oceanic-atmospheric circulation like El Niño-Southern Oscillation (ENSO), which can change regional and even global climate.

ENSO describes the periodic switching between warm El Niño and cold phase La Niña events. It is usual for the Pacific to experience high pressure and low rainfall in the central and eastern regions, while warm water along with low pressure and wet conditions prevail in the western regions. El Niño describes the change of the usual condition by the development of warm water over the central and eastern Pacific. The warm water along with low pressure bring wet conditions to countries like Peru, Ecuador and Tahiti, while cool water and high pressure over the western part cause low rainfall in Australia and Indonesia. La Niña is the converse of El Niño, when very cold water, strong high pressure and very dry conditions prevail in the eastern Pacific while very warm water, low pressure and wet conditions occur in the western Pacific.

As the Pacific warms, large amounts of heat and moisture are added to the atmosphere. By means of wind transportation, energy and moisture can travel over large distances which can affect distant weather patterns, like intensifying tropical cyclones and thunderstorms (O'Hare et al., 2005:131-141).

2.2.1.4 Modifications in air chemistry

Modification in air chemistry can be caused by both natural and human factors. A good example of a natural factor is volcanic activity. When a volcano erupts, large amounts of particulate matter and gases enter the atmosphere. Included in the particulate matter are sulphate aerosols which are very effective at scattering incoming solar radiation, and this consequently has a cooling effect. For example, on 28 August 1883, the eruption of Krakatoa caused large amounts of debris to be ejected into the atmosphere which caused a decline in solar radiation strength for about 4 years (O'Hare et al., 2005:75,76,165,167).

Human induced changes in air chemistry lie at the heart of the attribution of global warming. The greenhouse effect helps to heat the earth's surface by trapping and re-radiating heat energy. Rapid industrial development since the Industrial age has caused large quantities of greenhouse gases to be emitted into the atmosphere. Paleo-climate data taken from ice cores in the Antarctic show that carbon dioxide and methane concentrations in the atmosphere have greatly exceeded any pre-industrial levels of the last 600 000 years, as illustrated in Figure 2.2. With the increase of greenhouse gases in the atmosphere, heat energy is more effectively trapped which has a warming effect. This is known as the enhanced greenhouse effect.

Stellenbosch University http://scholar.sun.ac.za

2.3 Climate and circulation models

Climate models are also used to attribute climate change to anthropogenic forcing. Energy inputs into the climate system are referred to as forcing. Figure 2.3 illustrates climate model simulations of natural and human induced/anthropogenic forcing scenarios of global temperature. The solid line refers to observed data, the blue line to models that only simulated natural forcing scenarios and the pink line to scenarios that incorporate both natural and anthropogenic forcing in the simulations. When the models included anthropogenic forcing into scenarios, they more accurately simulated the observed data than the models with only natural forcing scenarios. This shows that current climate conditions can be attributed to additional anthropogenic effects and not only to natural effects.

2.3.1 Types of climate models

There are a number of different types of climate models, all varying in complexity and scale. These include energy balance models, global and regional circulation models and statistical downscaling methods. These models will be discussed briefly in the following sections:

2.3.1.1 Energy balance models

Energy balance models (EBMs) are given a forcing input and the results are calculated when an equilibrium condition is reached. These models vary from the very basic zero-dimensional model which only incorporates incoming and outgoing energy flows of the globe, to one-dimensional models which take the horizontal energy transfer through heat transfer into account. Finally, more complicated models like upwelling-diffusion EBMs add further dimensions like the global atmosphere, surface ocean layers and a deep ocean (O'Hare et al., 2005:181-185).

The greatest advantage of EBMs lie in their simplicity, as they describe physical processes in idealised ways. Practically, more simulations under different scenarios can be run on EBMs than on more complex models since they are not as computationally demanding. However, simple EBMs only give global averaged results and give no indication on how the climate will respond in different parts of the world (O'Hare et al., 2005:184-186).

2.3.1.2 Global circulation models

Global circulation models (GCMs) address the shortcomings of EBMs by describing the climate system more realistically. GCMs represent a three-dimensional climate by using equations describing the movement of energy, momentum, conservation of mass and ideal gas laws (O'Hare et al., 2005:186).

Stellenbosch University http://scholar.sun.ac.za

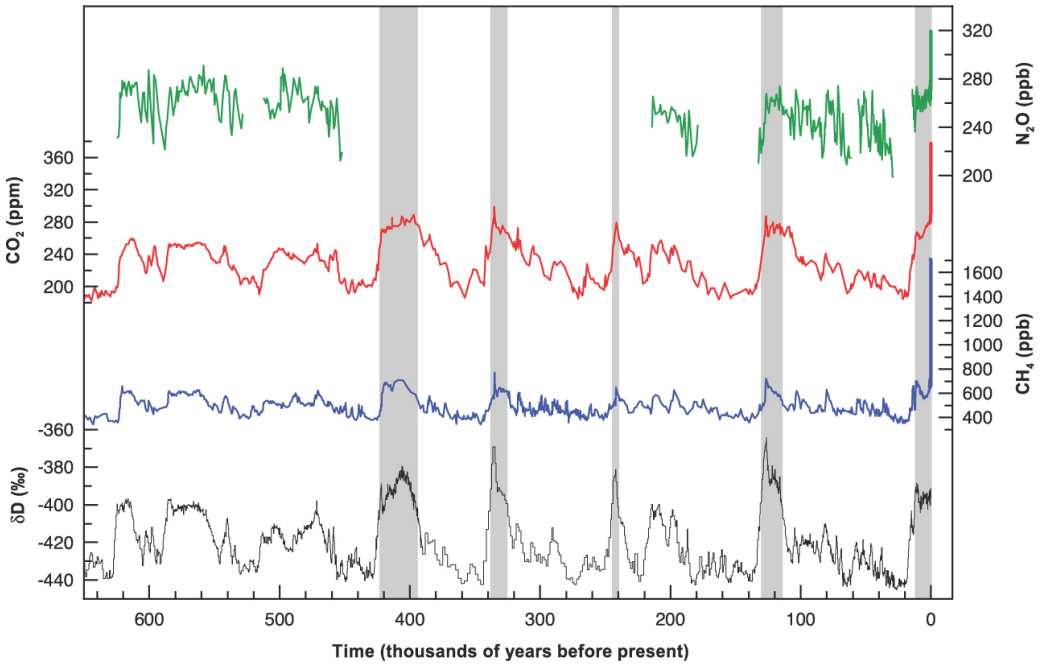


Figure 2.2: Glacial-interglacial ice core data for various greenhouse-gases over the last 600 000 years (Solomon et al., 2007:24).

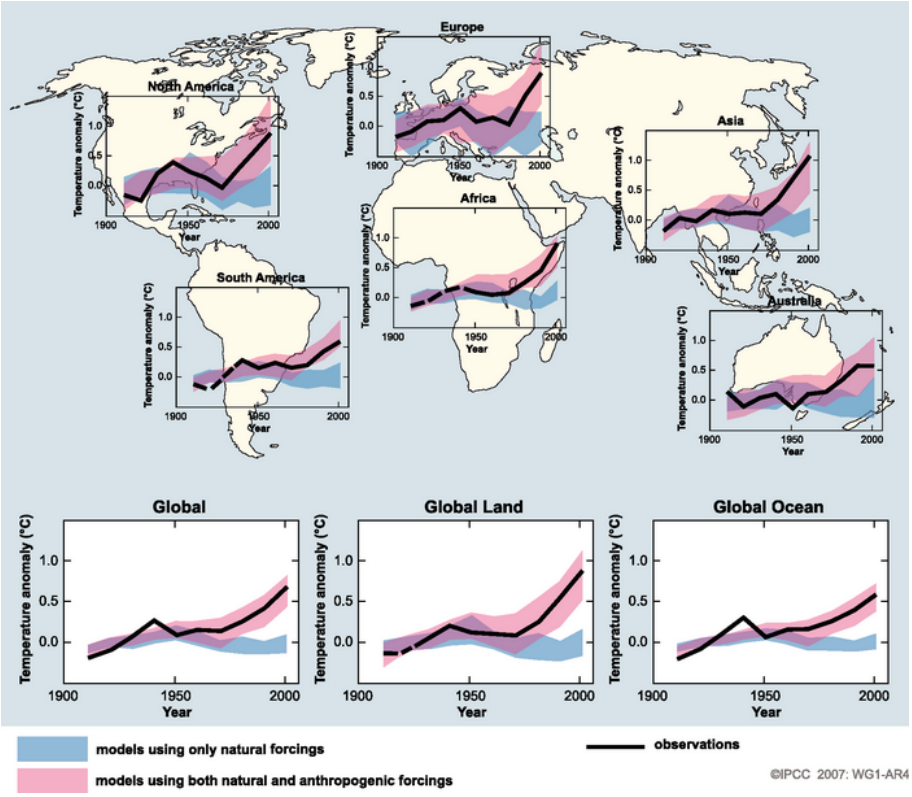


Figure 2.3: Comparison between climate model simulations that incorporate natural and additional anthropogenic forcing on the climate with actual observed data (Solomon et al., 2007:61).

Stellenbosch University http://scholar.sun.ac.za

These equations are solved at different points on the earth's surface for a fixed time interval over several atmospheric layers. GCMs compute the energy transfer through the atmosphere, effects of aerosols, changes in snow and ice cover, heat storage of oceans and soils, surface fluxes of heat, and large scale transport of heat and water by the atmosphere and the ocean. The increase in realism also increases complexity and thus computational demand. In order to reduce the computational demand, the horizontal resolution of GCMs needs to be coarse, but this comes at a cost as many components of the climate scale happen at smaller scales than GCM resolution (O'Hare et al., 2005:187).

GCMs can be generally divided into two different types: equilibrium and transient models. Equilibrium models provide a certain climatology input that is comparable to current climate conditions and runs until equilibrium conditions are reached. Afterwards, a new climate condition is implimented, for example, the doubling of atmospheric carbon dioxide concentration and this runs until a new equilibrium is reached. Because the equilibrium conditions are of interest, response times like those of oceans are not considered. The results can be compared to see what the effect of a certain forcing scenarios will be compared to prevailing conditions. A disadvantage of equilibrium models is the inability to indicate time dependent- and regional responses to different forcing conditions (O'Hare et al., 2005:187).

Transient models incorporate the effects of oceanic-atmospheric coupling to represent the effect of time more realistically. Oceanic response to the climate is a lot slower than the atmosphere's response as energy is distributed differently. However, the oceanic conditions also influence atmospheric conditions, creating a complex feedback effect between these systems. Transient models attempt to incorporate this complex system into the results that are time dependent. Coupling the ocean with the atmosphere is however, very difficult, as the initial conditions of the ocean already incorporate some effects of climate change from the past (O'Hare et al., 2005:188).

2.3.1.3 Regional climate models

Regional climate models (RCMs) attempt to address the shortcomings of GCMs by simulating finer scale regional climate effects. RCMs are more realistic than GCMs as they model regional climate responses to changes in land-surface vegetation or atmospheric chemistry. RCMs are dependent on GCM output data and model finer scale effects accordingly. Because of the even greater physical complexity of RCMs, they are as computationally demanding as GCMs. In addition, RCMs suffer from a dependency on GCM data as an initial condition , because errors in GCM input data can lead to great misrepresentations of regional climate by RCMs (O'Hare et al., 2005:188-191).

Stellenbosch University http://scholar.sun.ac.za

2.3.1.4 Statistical downscaling

Statistical downscaling methods use GCM output data and derive more detailed climate scenarios by statistical techniques. These can yield fine scale results and are able to achieve regional climate effects, while not being as computationally demanding as RCMs. However, these models are dependent on the calibration of large amounts of data and assume scaling relationships remains the same under changing climate conditions (O'Hare et al., 2005:192-195).

2.3.2 Performance of climate models

The performance of climate models is measured by their ability to simulate real climatic events like atmosphere-oceanic interactions like ENSO or climate feedback effects. This is typically achieved by simulating the climate from the past and comparing the results with the observed data. Climate models have performed poorly in the area of precipitation simulation which can be attributed to their representation of clouds. If clouds are not correctly represented, this can influence the effect of incoming solar radiation and lead to uncertainties regarding the magnitude and timing of precipitation events. Climate models are also prone to be biased like being too wet or too cold. One way to overcome model bias is to test a specific model under a number of slightly different boundary conditions and/or to test a number of different models with the same initial conditions. This helps to give a general overview of climate models and helps to show individual model bias (O'Hare et al., 2005:186,199).

The shortcomings of climate models' representation of rainfall events is important from a flood hydrology and disaster management viewpoint, since it creates uncertainty over the expected changes in rainfall.

2.4 Rainfall studies

This section will analyse both observed and climate models' rainfall studies, both on a global and regional scale. According to Solomon et al. (2007:238), Southern Africa can expect a warmer, drier and more variable climate. However, the Southern African region has very different regional climate regimes (Nicholson, 2000:145), and a more detailed analysis is required for South Africa, especially with regards to variability in rainfall.

2.4.1 Impacts of current climate change on rainfall

Solar energy plays an important role in precipitation, as it provides the energy required to fuel the hydrological cycle. The evaporation of water requires energy that transforms it into water vapour that is stored in the atmosphere. This energy, called latent heat, is used to break the

Stellenbosch University http://scholar.sun.ac.za

molecular bonds in water in its liquid form. When this process is reversed, latent heat returns to the environment as sensible heat which is heat felt by the sensory organs. The released latent heat helps to balance the earth's energy and water budget by exporting the excess heat of the low latitudes to fill the energy-deficient areas of the high latitudes. The amount of water vapour that the atmosphere holds is dependent on the temperature. From the Clausius-Clapeyron relationship, the higher the temperature, the more moisture the atmosphere can hold. A precipitation event occurs when water vapour grows on clouds to such an extent that it grows too large and falls back to the surface (Trenberth, Dai, Rasmussen & Parsons 2003:1212, O'Hare et al. 2005:36,37).

The hydrological cycle is expected to be enhanced by the addition of more water vapour in the atmosphere as a result of the increasing temperature. In addition, the characteristics of rainfall are also expected to change as a result of the change in climate. Trenberth et al. (2003:1212) argue that heavy rainfall rates greatly exceed evaporation rates and depend on low level moisture convergence, therefore the rainfall intensity should also increase at about the same rate as moisture increases which is dependent on temperature increases. Consequently, fewer but more intense rainfall events can be expected.

2.4.2 Global rainfall studies

Global studies on observed precipitation have shown some agreement with global climate model studies. In Groisman, Karl, Easterling, Knight, Jamason, Hennessy, Suppiah, Page, Wibig, Fortuniak, Razuvaev, Douglas, Førland and Zhai (1999), daily precipitation data was analysed for a number of mid- and high latitude countries. Results indicated that summer precipitation increased by 5% in the past century and the frequency of summer precipitation events also increased in some countries. Groisman, Knight, Easterling, Karl, Hegerl and Razuvaev (2004) found increasing probabilities of intense precipitation events for many areas outside the tropics. Alexander, Zhang, Peterson, Ceaser, Gleason, Tank, Haylock, Collins, Trewin, Rahimzadeh, Tagipour, Ambeje, Kumar, Revadekar, Griffiths, Vincent, Stephenson, Burn, Aguilar, Brunett, Taylor, New, Zhai, Rusticucci and Vazquez-Aguirre (2006) analysed daily precipitation data from various parts of the globe, finding that precipitation indices showed tendencies towards wetter conditions. The latter part of the 20th century appeared to be wetter and warmer, and global mean total annual precipitation showed a significant increase. A number of other countries have also shown increases in climate variability and extreme rainfall events (such as Greece, Portugal and India) (Philandras, Nastos, Paliatsos & Repapis, 2010, Costa & Soares, 2009, Krishnamurthy, Lall & Kwon, 2009). Solomon et al. (2007:238) indicated changes in global annual precipitation trends, but these varied largely on a regional basis. However, regarding extreme events, a general increase in the intensity of high rainfall events was found.

Stellenbosch University http://scholar.sun.ac.za

Hennessy, Gregory and Mitchell (1997) used two climate models for daily precipitation under a doubling of atmospheric carbon dioxide concentrations. Increases in global average precipitation were found, along with increasing precipitation intensity. More recently, Solomon et al. (2007:750), using the results from a number of climate models and scenarios, reported increasing global mean precipitation with global warming, and large seasonal and spatial variation. In the higher latitudes there seemed to be increases in precipitation while in the mid latitudes there was a decrease in summer precipitation. Precipitation events showed more variability, with an increase in drought and extreme events in most regions. There was a greater increase in precipitation extremes than in mean precipitation.

2.4.3 Regional and local rainfall studies

2.4.3.1 Climate in South Africa and the Western Cape

South Africa experiences some of the most variable rainfall and stream flow in the world. According to Schulze (2003), South Africa's rainfall has high inter-annual variability, where the coefficient of variation for inter-annual rainfall is $>40\%$ which is high by global standards. Run-off shows even greater variability, ranging from 50 to 300%. Furthermore, South Africa to a large extent receives most of its rainfall during a specific season, as illustrated in Figure 2.4 which indicate regions of maximum rainfall periods within a year for South Africa. The concentration of rainfall in certain parts of the year creates seasonal variability, and consequently, more variable run-off. In addition, rainfall is extremely sensitive to location and topography, resulting in large variance of rainfall within the same climatic regions (Kruger, 2007:3). Schulze (2003) further explains that changing land surface use influences the amount of run-off generated. Rising temperatures in South Africa can also influence changes in hydrology like evaporation, irrigation, heat waves, droughts and soil moisture conditions (Schulze, Hewitson, Barichievy, Tadross, Kunz, Horan & Lumsden, 2011:15).

In a large proportion of the Western Cape rainfall occurs mainly in the winter months, while summers are typically dry (Kruger, 2004:17), due to the position of the Western Cape in relation to a band of westerly winds and associated low pressure systems which bring rainfall to the South-Western Cape by means of cold fronts. During winter months, high pressure systems which block low pressure systems move northward. Variation in the position of these pressure systems and westerly winds consequently cause seasonal and annual variation in rainfall which can lead to drought conditions. The Western Cape is also susceptible to cut-off lows which are mid latitude cyclones that are "cut-off" from the main planetary circulations. As the cyclones are cut off from the westerly winds to the south, they lose momentum and can remain stationary for days. These cut-off lows are unstable atmospheric systems which generally occurs in autumn and spring and are one of the main causes of extreme rainfall and flooding events in the Western Cape (Midgley, Chapman, Hewitson, Johnston, De Wit, Ziervogel, Mukheibir, Van Niekerk,

Stellenbosch University http://scholar.sun.ac.za

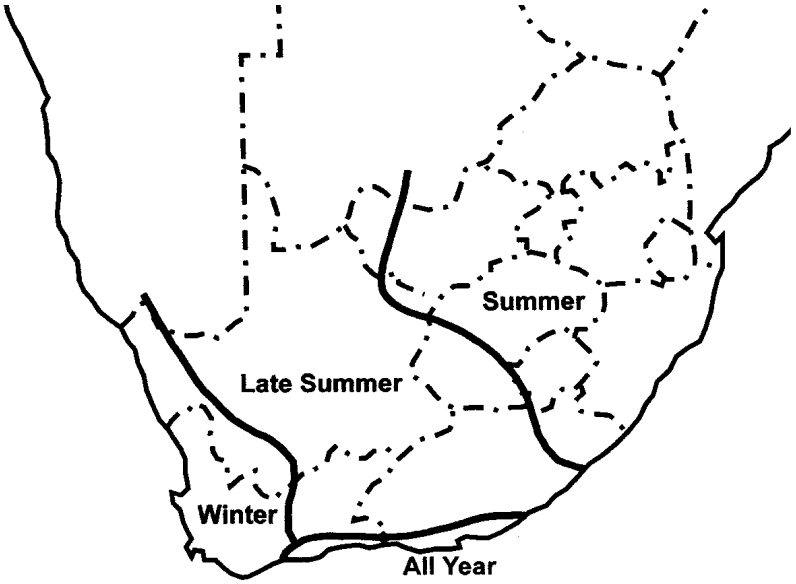


Figure 2.4: Concentration of maximum rainfall periods in South Africa (Kruger, 2007:14).

Tadross, Van Wilgen, Kgope, Morant, Theron, Scholes & Forsyth 2005:19,Holloway et al. 2010:18).

2.4.3.2 Rainfall studies

Tennant and Hewitson (2002) sought to understand rainfall dynamics in South Africa, finding that seasons with high total rainfall have a higher number of heavy rainfall days rather than an increase in light rain days. Furthermore, these high total rainfall seasons also contain prolonged dry periods. Rainfall in drier regions is also heavier and short lived, while wet regions show increases in the number of rainfall events. Kruger (1999) found $\pm$18-20 year oscillations for the summer rainfall region in South Africa. Periods of above and below normal late summer rainfall were identified. It was also found that extreme events of rainfall occur when El Niño occurs during an period of below normal rainfall, and when a La Niña occurs during an epoch of above normal rainfall. The correlation between ENSO and rainfall is confirmed by Rouault, Fauchereau, Pohl, Penven, Richard, Reason, Pegram, Phillippon, Siedler and Murgia (2010) which found that ENSO was a major driver of interannual rainfall, especially during summer. A number of other climatic factors, including sunspot numbers, were analysed, but either weaker or no correlations were found. On a decadal level, no correlation between sunspot numbers and rainfall was found, and no coherent periodicity was found in the 80 years of analysed data. On the other hand, (Alexander et al., 2006:114) using area averaged rainfall figures, found statistically significant 18-21 year periodicity and discernible periodicity for a number of regions corresponding to double sunspot cycles.

Analysing daily precipitation data from 1931-1990 for 316 sites, Mason, Waylen, Mimmack, Rajaratnam and Harrison (1999) found that in most of South Africa there was an increase in

Stellenbosch University http://scholar.sun.ac.za

intensities of high rainfall events during this period, except in parts of east, north-west and the winter rainfall region in the south-west, where decreases in intensities were found. In the east coast however, there was a sharp increase in intensities. The researchers did however conclude that there is little evidence of significant trends in annual totals in the country, except in the eastern part.

Fauchereau, Trzaska and Richard (2003), also analysing daily precipitation for Southern Africa, found shifts to more extreme rainfall events and more intense and widespread droughts, thus indicating increasing variability in annual rainfall, especially in the latter part of the 20th century. New, Hewitson, Stephenson, Tsiga, Kruger, Manhique, Gomez, Coelho, Masisi, Kululanga, Mbambalala, Adesina, Saleh, Kanyanga, Adosi, Bulane, Fortunata, Mdoka and Lajoie (2006), analysing daily rainfall for Southern Africa from 1961-2000, found total rainfall and number of days with heavy rainfall had decreased, but that the amount of rainfall on extreme days had increased. An increase in dry spells, that indicate the number of consecutive days without rain, was observed. Groisman et al. (2004) analysed 100 years of daily rainfall for the eastern half of South Africa and found no change in the annual and summer precipitation, but did find an increase in annual frequency of very heavy precipitation.

According to Midgley et al. (2005), noticeable changes in daily atmospheric circulation patterns were found over the Western Cape during 1958-2001 .The frequency of strong low-pressure systems increased significantly during March-May and decreased during June-August. Mountainous regions in the Western Cape experienced either little change in rainfall or increases in rainfall, while lower areas experienced decreases in rainfall. Seasonal trends were also found, where 10% of the highest daily rainfall events during March, April and May (MAM) experienced an increase in intensity, especially in the mountainous interior, but also in some coastal regions. During June, July and August (JJA), coastal regions drying trends, while mountainous regions in the interiors experienced wetting trends. The 10% of the highest daily rainfall events of JJA also experienced increasing intensities at higher elevations, while showing decreasing intensities at lower elevations.

Kruger (2006) analysed daily precipitation data for 138 stations using precipitation measurement indices for the period of 1910-2004. Total annual precipitation experienced a decrease in an area in the Northern part of Limpopo, an area comprising the southern part of Mpumalanga, and an area in the north-eastern part of the Free State and KwaZulu-Natal, an area in the south Eastern Cape and an area along the south coast. Areas showing significant increases in total annual precipitation were in the northern part of the North-West and an area covering parts of the Western Cape, Northern Cape and Eastern Cape. Areas showing positive trends in the longest annual number of consecutive wet days included the Free State and North-West, while areas showing negative trends include areas overlapping Limpopo, Gauteng, Mpumalanga, eastern Free State and the Eastern Cape. Positive trends in indices showing the amount of annual precipitation that is the results of extreme daily rainfall events, were found in areas including the

Stellenbosch University http://scholar.sun.ac.za

Free State, Eastern Cape and KwaZulu-Natal. Most stations in South Africa did not show trends in the annual number of days that exceeded 30mm of rain.

Thomas, Twyman, Osbahr and Hewitson (2007:301), analysisng daily rainfall for the period 1950-1999, found that the Limpopo Province experienced an increase in dry season length which causes a delay of the wet season, while in the north-west of KwaZulu Natal, trends toward higher rainfall, increases in heavier rainfall events and growing variability in the start and end of rainfall seasons were found. Hoffman, Cramer, Gillson and Wallace (2011:437), analysing the winter rainfall region in the Western Cape for the period 1974-2005, found that monthly rainfall figures experienced no significant changes at any of the stations analysed, although there were signs of considerable inter-annual variability.

With regard to climate model studies, findings in Solomon et al. (2007) on Africa and southern Africa are not very detailed, although a drying trend is expected over southern Africa. The extreme south-west, including the south Western Cape area, shows drying in winter, while a general increase in the intensity of high rainfall events are also noted (Solomon et al., 2007:868,871). Remaining in Southern Africa, Hudson and Jones (2002:10,12), found drying over the central tropical and subtropical land areas in summer, while equatorial regions experienced more rainfall along with more intense and extreme rainfall events. Large parts of Southern Africa did however show significant decreases in rainfall. The South-Western Cape experienced a significant decrease in winter rainfall of roughly 10-20%, and showed tendencies towards more extreme rainfall.

Groisman et al. (2004:1340) analysed future daily rainfall data in South Africa. In the event of doubling of atmospheric carbon dioxide content, there could be an expectation of a decrease in heavy precipitation. Hewitson and Crane (2006:1336) used empirical downscaling with three climate models for daily precipitation, finding that South Africa could experience increased summer rainfall over the convective region of the central and eastern plateau and the Drakensburg Mountains. Little change in rainfall in the Western Cape was projected, but a slight drying in the summer and a slight decrease in winter rainfall can be expected.

Midgley et al. (2005:28,32) states that results of GCM simulations show an increase in total annual rainfall over the eastern regions of the Western Cape and a decrease in rainfall, especially in the western region of the Western Cape during early winter, with some indication of winter drying. The frequency of days with rainfall $>$20mm is projected to remain the same during late summer, while the highest 10% of events are increasing. This indicates that changes in late summer rainfall will mostly be due to increases in the intensity of rainfall.

In a more recent study, Lumsden, Schulze and Hewitson (2009) used six climate models and statistical downscaling methods for the future periods 2046-2065 and 2081-2100. They project that the eastern part of South Africa can expect more rainfall in the form of more rain days and more days with bigger rainfall events. For the west coast and adjacent interior, less rainfall, strong drying tendencies and more inter-annual variability can be expected. A small part of the

Stellenbosch University http://scholar.sun.ac.za

south Western Cape can expect more rainless days.

Schulze et al. (2011:15) summarised the results of GCM grid cell response and downscaled data, stating that the climate, and especially rainfall, is too complex to reduce to a single response. There were some indications of a wetter east coast, a drying west coast and a drier south-west for both seasons and increasing rainfall intensities, not necessarily as a result of increasing rainfall, but due to larger rainfall events.

The results of the observed and climate model studies are summarised in Tables 2.1 and 2.2. To summarise, there is little agreement amongst different sources and between observed and climate model data for South Africa and the Western Cape specifically. This is most likely because of the high spatial and topographic sensitivity of rainfall, although there seems to be some agreement on the increased nature of variability and increasing intensities of rainfall in South Africa. However, all of the studies above were focussed only on rainfall up until a daily scale which is inadequate for assessing rainfall events of a shorter duration, especially in the light of expected increases in intensity. An area might receive 10mm of rain on an average rain day, but this mainly occurs during the morning, making daily rainfall totals inadequate to assess changes in the intensities of shorter storm durations. Kruger (2004:30) shows that rainfall at many stations is more likely to occur during a certain time of the day. In addition to these shortcomings of daily rainfall data, stormwater design accommodates for storm events on a sub-daily basis. Therefore, assessing any climate change impacts on stormwater design requires shorter time scales than daily scales.

2.5 Short duration rainfall

Studies on short duration rainfall (SDR) in South Africa are limited, especially with regards to climate change, most likely because of the limited availability of reliable SDR data.

Up until the 1990s, SDR was mostly recorded with the use of analogue autographic gauges. Rainfall would be recorded cumulatively in real time up until a certain depth, where the stored water would be flushed out, known as a siphon event, and the process would start over. Figure 2.5 illustrates a typical rainfall event recorded on autographic paper. Smithers and Schulze (2000) did a comprehensive, nationwide analysis of SDR in South Africa, including the digitising of autographic rainfall data. Data from 412 stations were analysed, of which data from 334 were provided by the South African Weather Service (SAWS). However, the quality of the data of many of rainfall stations were poor, with numerous gaps and errors in the raw autographic records, while fewer than half of the stations had record lengths greater than 10 years. Although the digitising process aimed to address some of the problems, the digitised records compared poorly with standard daily rainfall totals, generally underestimating the daily rainfall totals. Consequently, this led to the conclusion that most of the digitised autographic rainfall

Table 2.1: Summary of rainfall studies on observed data

Author	Data Type	Analysis Period	Study Area	Summary of findings
Tennant and Hewitson (2002)	daily totals	1936-1999	National	High seasonal rainfall contains greater heavy rainfall days than increases in light rainfall
Kruger (1999)	monthly totals	1930-1995	Summer rainfall area	18-20 year oscillations in summer rainfall region Extreme events tend to correlate with ENSO
Rouault et al. (2010)	monthly totals	1921-2007	National	ENSO major driver of interannual rainfall
Alexander et al. (2006)	area averaged rainfall	1921-1992	National	18-21 year periodicity that corresponds to double sunspot cycles
Mason et al. (1999)	daily totals	1931-1990	National	Most parts showed increasing intensities of high rainfall events. Decreasing intensities for parts of the east, north-west and winter rainfall region. Little evidence ofsignificant trends in annual totals, except in east.
Fauchereau et al. (2003)	daily totals	1901-1998	Southern Africa	Shifts toward more extreme rainfall events and more variable rainfall.
New et al. (2006)	daily totals	1961-2000	National	Total rainfall and number of days with heavy rainfall decreased. Amount of rainfall on extreme days increased.
Groisman et al. (2004)	annual and seasonal totals	1906-1997	East of SA	No change in annual and summer rainfall. Increase in annual frequency of heavy precipitation.
Midgley et al. (2005)	monthly totals	1950-1999	Western Cape	Mountainous areas: • Either little change or increases in rainfall. • Increasing intensities for heavy rainfall events. Low lying areas and coast: • Decrease in rainfall. • Winter rainfall showed drying trends. • Decreasing intensities for heavy rainfall events.
Kruger (2006)	daily totals	1910-2004	National	Results were spatially sensitive with no single response.
Thomas et al. (2007)	daily totals	1950-1999	National	Limpopo: Delay of wet season. KwaZulu-Natal: Higher and heavier rainfall with more variability
Hoffman et al. (2011)	monthly totals	1974-2005	Western Cape	No signs of change, but greater inter-annual variability

Stellenbosch University http://scholar.sun.ac.za

Table 2.2: Summary of rainfall studies on climate model data. Please note that some entries do not include specific details about the model types, scenarios etc. as the studies consisted of reviews from various sources.

Author	Model Type	Scenario	Period	Study Area	Summary of findings
Solomon et al. (2007)[1]	-	-	-	Southern Africa	Drying for Southern Africa. South Western Cape drying in winter and heavy rainfall intensity increase.
Hudson and Jones (2002)	HadAM3H[2]	SRES[10] A2[11] SRES[10] B2[12]	2071-2100	Southern Africa	Large parts of Southern Afica shows decreases in rainfall. South Western Cape decrease in winter rainfall and more extreme rainfall.
Groisman et al. (2004)	HadCM3[3], CGCM1[4]	2 $\times$ CO_2 atm. conc.	2040-2060	South Africa	Expected decrease in heavy precipitation
Hewitson and Crane (2006)	HadAM3[2], ECHAM[5] 4.5, CSIRO[6] Mk2	SRES[10] A2[11]	-	South Africa	Central and eastern plateau: Increased summer rainfall. Western Cape: Slight drying in summer and decrease in winter rainfall
Midgley et al. (2005)	-	-	-	Western Cape	Decrease in rainfall during winter. Late summer rainfall intensity increase.
Lumsden et al. (2009)	CSIRO[6], ECHAM[5], HadAM[2], GFDL[7], MIROC[8], MRI-CGCM[9]	SRES[10] A2[11]	2046-2065, 2070-2100	South Africa	East: More rain with heavier rainfall days. West Coast: Drying and greater inter-annual variability.
Schulze et al. (2011)	-	-	-	South Africa	Indications of wetter east coast, drying west coast and increasing rainfall intensities

[1] The IPCC's Fourth Assessment (Working Group I) summarises a large number of climate model studies.
[2] Hadley Centre Atmospheric Model
[3] Hadley Centre Coupled Model
[4] Coupled Global Climate Model
[5] Max Planck Institute for Meteorology, Germany
[6] Commonwealth Scientific and Industrial Research Organisation
[7] Geophysical Fluid Dynamics Laboratory
[8] Model for Interdisciplinary Research On Climate
[9] Meteorological Research Institute, Japan
[10] Special Report on Emission Scenarios
[11] The A2 scenario describes economically driven human activity in the future
[12] The B2 scenario describes environmentally driven human activity in the future

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

was not suitable for the estimation of design rainfalls (Smithers & Schulze, 2000:86-155) in South Africa.

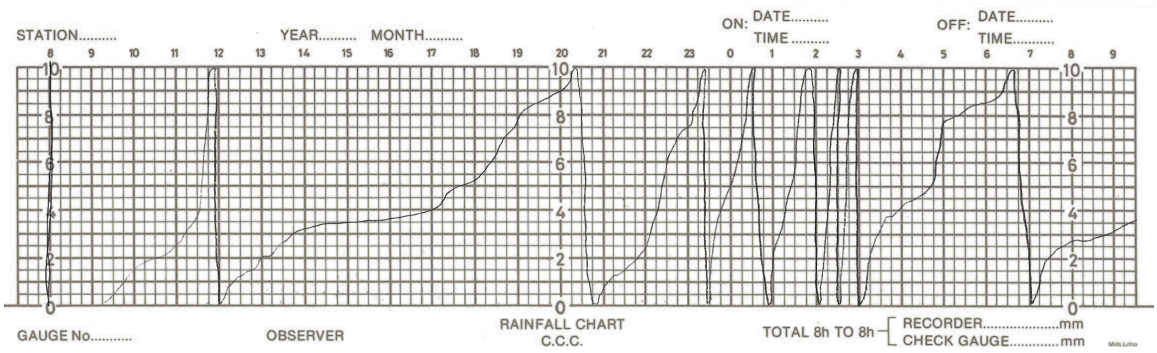


Figure 2.5: A typical autographic logged rainfall day.

More recently, SDR has been recorded with a tipping bucket system by which rainfall is collected in a funnel and the water is then captured alternately by two buckets which hold an equal and specific volume. When one bucket is filled, it tips and the other bucket then fills up. The number of tippings is recorded automatically and logged digitally (Kruger, 2004:1,2). Figure 2.6 shows a typical tipping bucket device. Since the work of Smithers and Schulze (2000), no similar studies have been found gauging the performance of the tipping bucket system in South Africa, although there are indications that the tipping bucket system underestimates standard gauge daily totals as a result of external factors like wind (De Jager, 2012).

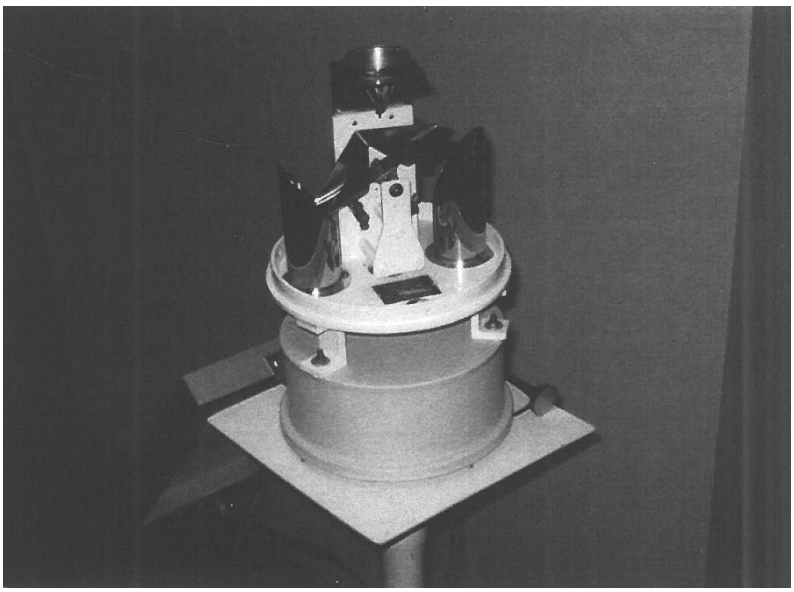


Figure 2.6: A typical tipping bucket measuring device (Kruger, 2007:3).

Obtaining SDR data is difficult, as akthough there are a number of sources available apart from the SAWS, these sources are less known and either have very short record lengths or are unable to provide the data. Combining the two different sources of data to extend the record length presents its own problems, as each source estimates rainfall differently, which leads to

Stellenbosch University http://scholar.sun.ac.za

difficulties (Changnon & Kunkel, 2005). Given these conditions, it comes as no surprise that so little work has been done on sub daily rainfall in South Africa.

A recent study on short duration rainfall in South Africa, (Van Wageningen & Du Plessis, 2007), analysed short duration at the Molteno rainfall station in Cape Town with the intention of comparing the findings to current climate change expectations. For the observed period of 1961-2003, the average rainfall amount per 5 minutes experienced an increase in the last 6 years, while the amount of 5-minute events experienced a decrease over the same period, indicating more rainfall occurring over fewer events. This was reinforced by the fact that even though total annual rainfall was below average, rainfall events decreased more rapidly than the total annual precipitation. These findings seemed to be in agreement with climate change expectations of fewer but more intense rainfall events, but there was uncertainty if the effect was local or occurred over a wider region.

The work done by Van Wageningen and Du Plessis (2007) showed the need to further investigate possible effects of climate change in SDR by analysing the data of a wider range of stations in the Western Cape region and this forms the basis for this thesis.

2.6 Extreme value theory

Stormwater design typically implements intensity duration frequency (IDF) curves in order to determine the intensity of the storm which will be designed for. IDF curves are constructed in the following way:

1. Set up storm duration magnitudes from the SDR database for durations 5 minutes to 24 hours
2. For each duration, use an appropriate data selection method and fit a statistical distribution through the data.
3. Obtain return levels for the desired return periods from the statistical distribution.
4. For each return period, plot the return levels of all the durations against the storm durations.

Figure 2.7 graphically displays steps 2 and 3. For a specific storm duration, the return levels are plotted against the return period. The return period represents the period or time that an event of a certain magnitude would occur once on average. Return periods can also be interpreted as the probability that a certain event will occur. For example, an event with a return period of 10 years is more likely to occur than an event with a 100 year return period. The return level is the corresponding event that would occur at the specific return period. The solid line in Figure

Stellenbosch University http://scholar.sun.ac.za

2.7 represents the statistical distribution that is fitted through the storm duration data, indicated by solid points, and is used to obtain the return levels for the return periods indicated by the grey vertical dashed lines. The statistical distribution is necessary when return levels for events beyond observed levels are required, like for a return period of a 100 years.

Figure 2.8 graphically displays step 4. Each series of points represents return levels for all storm durations for a specific return period. For example, the rectangular points represents return levels for a return period of 2 years. The solid lines represent regression lines that are fitted to the point series for functionality. From this, the design intensity can be obtained for a specific return period at a certain storm duration.

A vital component in the construction of the IDF curves are the statistical distributions mentioned in steps 3, and typically require the implementation of a special branch of statistics called extreme value theory (EVT). Whilst conventional statistics tend to describe average behaviour, extreme value theory (EVT) is focussed on unusual and extreme phenomena. EVT has been thoroughly developed in the last century, especially since the 1950s and is primarily applied in the field of civil engineering and hydrology where the soundness of infrastructure under extreme conditions is vital (Gumbel 1960:4,Coles 2001:vii).

EVT is based on the derivation of limiting distributions for a series of extreme values, which entails the asymptotic behaviour of the underlying values (Coles, 2001:45,46). This assumption is physically problematic, as it implies that distributions are boundless, in other words, the distributions theoretically allow for infinitely large or infinitesimally small values to occur (Gumbel, 1960:1,2). In addition, extreme value distributions are often used to extrapolate for probabilities far beyond observed levels. However, since no alternative and accepted theory exists, EVT remains the best approach to model extreme phenomena (Coles, 2001:vii).

Classic EVT is divided into two over-arching methodologies, event maxima and threshold exceedance models, although other methods like point process characterisation, multivariate and Bayesian models have been developed.

2.6.1 Event maxima models

Event maxima models are distributions that are derived for maximum values over a certain block size. In hydrological settings, the block size is typically chosen on an annual basis to avoid seasonal effects and is called an annual maximum series (AMS). Following the derivation in Coles (2001:45-49), the three possible limiting distributions for a series of maxima are the Gumbel, Fréchet and Weibull distributions, and are combined in a single model known as the generalised extreme value distribution (GEV[1]) with the following distribution function (Coles,

[1]Please note that all occurrences of the abbreviation GEV will refer to "generalised extreme value distribution" in its entirety.

Stellenbosch University http://scholar.sun.ac.za

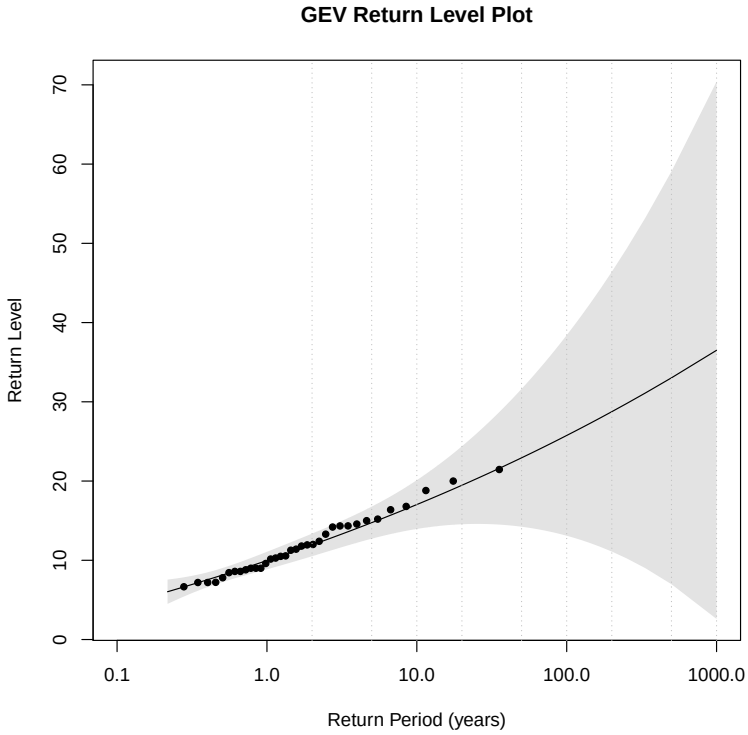


Figure 2.7: Return level plot. Note that the shaded area represents the 95% confidence interval for the statistical distribution.

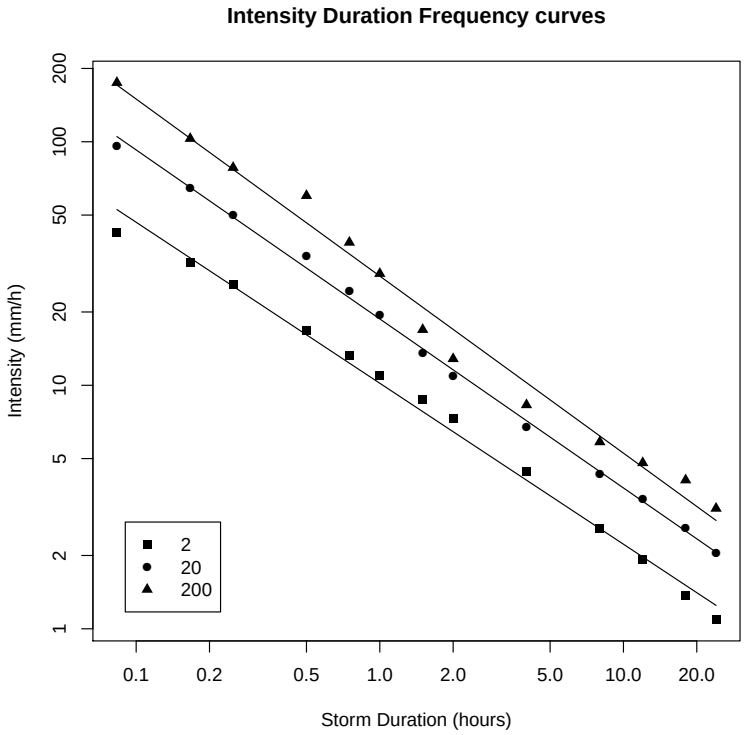


Figure 2.8: IDF curves

Stellenbosch University http://scholar.sun.ac.za

2001:47,48)[2]:

$$F(x) = \begin{cases} \exp\left(-\left[1-\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right) & \text{if } \xi \neq 0 \\ \exp\left(-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right) & \text{if } \xi = 0 \end{cases} \tag{2.1}$$

where:

- x is an independent identically distributed variable of a block maxima sample
- $F(x)$ is the cumulative distribution function of the block maxima sample
- ξ is the shape parameter
- μ is the location parameter
- σ is the scale parameter

The effect of each parameter is displayed in Figure 2.9. The location, scale and shape can be seen as analogous to the mean, variance and skewness respectively, used in regular statistics. Of all the parameters in the GEV, the shape parameter is the most critical and sensitive parameter and determines boundary conditions for the distribution. For $\xi < 0$, the distribution has an upper limit of $(\mu - \sigma)/\xi$, for $\xi > 0$, the distribution has a lower bound of the same value, while $\xi = 0$ is boundless.

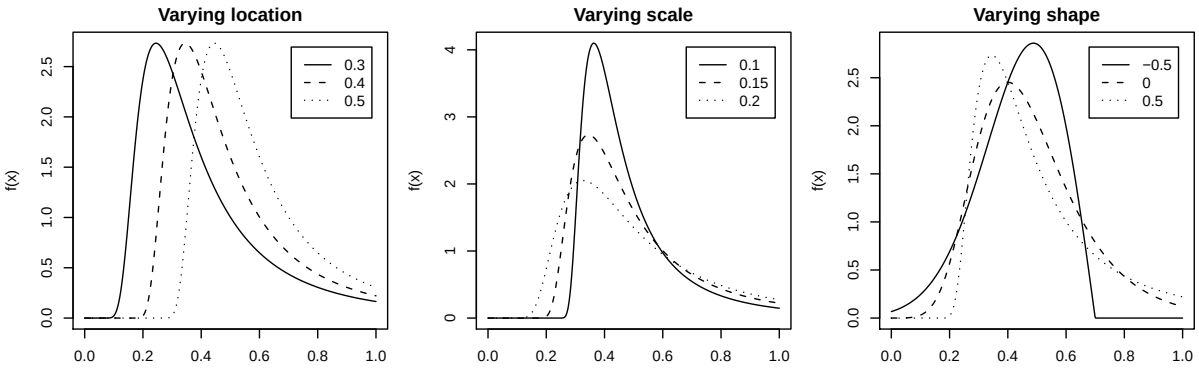


Figure 2.9: Density plots showing the effect of a varying parameter on the GEV, while the other parameters remain constant.

There are a number of distributions available for event maxima modelling, but the GEV remains the most basic derivation for event maxima data and is recommended for use in South Africa (Smithers 1996:216,Smithers & Schulze 2000:200,Van Bladeren, Zawada & Mahlangu 2007:6).

[2]Symbols and notations are taken from Coles (2001). Unfortunately, there are many different symbols and notations in the literature, which can be extremely confusing (see for example Hosking, Wallis and Wood (1985:252) for different symbol usage). Caution is advised when working with a number of different sources.

Stellenbosch University http://scholar.sun.ac.za

2.6.2 Threshold exceedance models

A drawback of event maxima models, especially those with AMS, is the loss of information, as a number of large peak values are ignored if they occur within the same block, while lower values occurring in a different block are selected (Coles 2001:74, Katz, Brush & Parlange 2005:1127). Threshold exceedance models incorporate all data above a threshold level and are commonly referred to as a peaks over threshold (POT) series. POT methodology is part of the partial duration series (PDS) family. Another instance of a PDS is to limit the amount of peak values to a certain value, often to multiples of the record length.

Previously, the exponential distribution was used to model POT, but more recently, the generalised Pareto distribuion (GPD) was shown to be the limiting distribution of a series of independent exceedances over a high threshold following a Poisson procees (Cunane 1973, Wang 1991:264). The GPD has the following distribution(Coles, 2001:75,76):

$$F(x) = \begin{cases} 1 - \left[1 - \xi\left(\dfrac{x-u}{\sigma}\right)\right]^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left[1 - \left(\dfrac{x-u}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \tag{2.2}$$

where the x is an independent identically distributed variable exceeding a threshold u, while ξ and σ are the same as those in Equation 2.1. The most critical parameter is the threshold value u, which is selected manually. The resulting parameters and values are extremely sensitive to the choice of threshold (White, Corney, Grose, Holz, Bennett & Bindoff, 2010:4). If the threshold is too low, it violates the asymptotic basis of the model, while a too high threshold leads to high variance in estimates (Coles, 2001:78). The threshold value remains the greatest weakness of the GPD, because of the subjectivity involved in its selection. As of yet, no reliable automated method exists to determine the threshold (Katz et al., 2005:1128).

A number of methods exist to make an appropriate choice of threshold. Interpretative plots like the Hill and mean residual life plot have been suggested, but these are often difficult to interpret and also suffer from a degree of subjectivity (Coles 2001:78-80, Tsourti & Panaretos 2001:5). Claps and Laio (2003) and Ben-Zvi (2009) have suggested raising the threshold from a very low level until some goodness of fit criteria are met. There are also a number of proposed methods for the selection of independent peaks for the POT sample. These typically involve an empirical rule to select the peaks before the sample parameters are estimated. Suggestions in the hydrological field include that peaks have to be separated by some time unit, like the storm duration (Trefry, Watkins & Johnson 2005:438-439, Van Wageningen 2006:69-70) or periods of zero rainfall/low flow (Ben-Zvi, 2009:110), or that the average amount of peaks be limited to a certain amount (see discussion in Claps and Laio (2003:4)). Coles (2001:98- 100) suggests

Stellenbosch University http://scholar.sun.ac.za

a method of declustering on the basis that peak values tend to cluster and also suggests raising the threshold until a certain stability in the return levels is obtained.

2.6.3 Comparison between models

In a number of studies the performance of the AMS and POT methodology are compared (Cunane 1973, Tavares & Da Silva 1983, Wang 1991). In terms of distributions, the shape parameter, ξ , of the GEV and GPD should theoretically be equal (Coles, 2001:75-76), in practice however, the literature shows that the return levels of the AMS/POT methodologies do not compare well (Ghahraman & Khalili 2004, Hugo 2010), while the comparison between the shape parameter of the GEV and GPD also shows little correlation (Ben-Zvi, 2009:107, 110-111).

In South Africa, the most widely used method is the AMS as this is an easy method to apply and there is confidence in the independence of maxima (Smithers & Schulze 2000:9,Van Bladeren et al. 2007:5). Van Bladeren et al. (2007:5) recommends that POT methodology only be applied to observation periods less than 14 years and Smithers (1996) and Smithers and Schulze (2000) used AMS exclusively for the analysis of short duration rainfall. The GEV is also the recommended distribution to be applied in South Africa (Smithers 1996:216, Smithers & Schulze 2000:200, Van Bladeren et al. 2007:6). The GPD performed poorly according to the various tests applied in Smithers (1996:216) and Smithers and Schulze (2000:200), but it seems likely that the studies applied AMS methodology to the GPD only. This application is invalid, as the GPD is the limiting distribution for POT data and not for AMS data, as mentioned in Section 2.6.2.

Although South African literature strongly supports the use of AMS/GEV methodology, work done by Coles, Pericchi and Sisson (2003) has shown that AMS/GEV methodology can at times be insufficient to estimate extreme events. An extreme rainfall event occurred on the central coast of Venezuela during 1999, with approximately 410mm falling within a day. The magnitude of this event was nearly three times the size of the previous maximum event over a 50 year period. Prior to the event, a simple Gumbel distribution would have estimated an event of that magnitude to be nearly impossible. A variety of statistical methods were analysed to test their predictive power for the extreme event by both including and excluding it into/from the analysed daily rainfall record. The methods included GEV/AMS and GPD/POT methodology, but also more advanced methods like Bayesian inference and seasonal modelling. It was found that GPD/POT methodology performed better in terms of predictive power than the GEV/AMS methodology, by giving a greater probability to the 410mm event. In light of the importance of being able to anticipate extreme events, the study by Coles et al. (2003) shows that there is room for the application of GPD/POT methodology in spite of South African recommendations.

Stellenbosch University http://scholar.sun.ac.za

2.6.4 Parameter estimation

In order to fit a distribution to the measured data, a method to estimate the distribution parameters is required. A variety of methods are available, including the method of moments, maximum likelihood (ML) procedures, probability weighted moments (PWM)/L-Moments and Bayesian inference amongst others. PWM/L-Moments have gained popularity, being simple to calculate. This method performs well for small samples and has low variance and bias (Hosking et al. 1985:251, Smithers & Schulze 2000:18-19,Katz, Parlange & Naveau 2002:1287, Van Bladeren et al. 2007:6). The ML method is not readily applied, most likely because it requires numerical optimisation techniques (Alexander 2001:451, Smithers & Schulze 2000:19) and can perform poorly with small samples, but has the advantage of incorporating covariates into the parameters and estimates the variance of the parameters and distribution values (Katz et al. 2002:1287,1293, Katz et al. 2005:1127, Gilleland & Katz 2006:4). The ML technique works by taking a range of parameter values and calculating the values that will give the highest probability for the likelihood function of the specific distribution, hence the name maximum likelihood[3].

Whilst there seems to be awareness that uncertainties in extreme value models increases for greater return periods, these are rarely incorporated into the leading South African hydrological guidelines (see Alexander 2001, SANRAL 2007). To show the importance of measuring uncertainty in statistics, consider Figure 2.10. Referring to the figure on the left, the GEV distribution fits the data well, making it possible to obtain return levels for high return periods like 50, 100, 200, 500 and even a 1000 years. However, when considering the figure on the right, the confidence in these return levels decreases. For example, one is 95% confident that the 100 year return level lies anywhere between approximately 13 and 38. The question then arises, what value will be used for design? By the 1000 year return period, the return level can lie anywhere between approximately 5 and 70. The ML technique is advantageous in this regard as it can easily incorporate these uncertainties into the results.

2.6.5 Non-stationarity

In one of the earliest textbooks on EVT, *Statistics of Extremes*, Gumbel (1960:1) states:

> *The distribution from which the extremes have been drawn and its parameters must remain constant in time (or space), or the influence that time (or space) exercised upon them must be taken into account or eliminated. Another limitation of the theory is the condition that the observation from which the extremes are taken should be independent. This assumption, made in most statistical work, is hardly ever realised.*

[3]The theoretical background of ML and application to the GEV and GPD is found in Coles (2001:30-36, 55-57, 80-83).

Stellenbosch University http://scholar.sun.ac.za

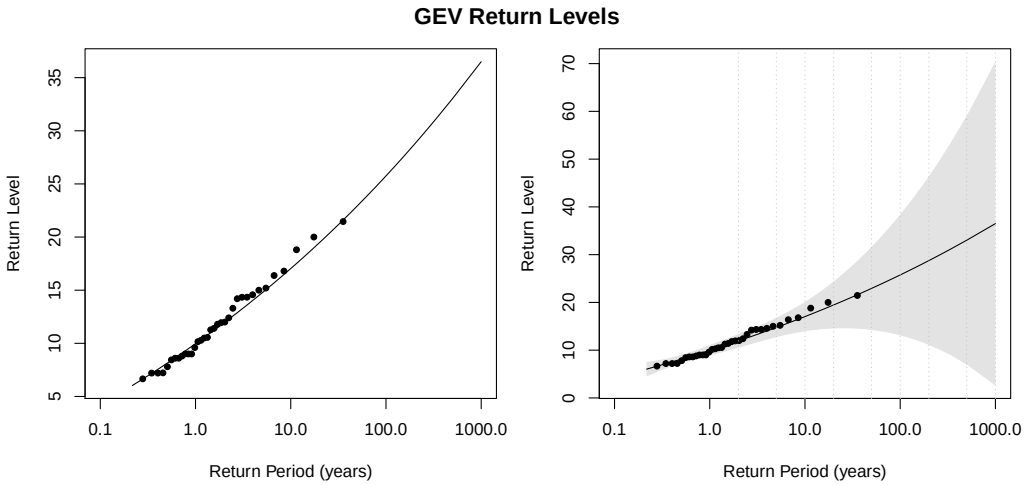


Figure 2.10: Return level plots for a GEV for Langebaanweg (1973-2009) for a storm duration of 60 minutes using ML parameter estimation. Both figures represent the same data and fitted GEV distribution, but the figure on the right includes the 95% confidence bands for the distribution values.

Basic EVT assumes that the data used is independent and identically distributed. This implies that the data points used for the distributions do not influence each other and have exactly the same parent distribution, essentially implying that the data remains homogeneous over time. In light of the effect of various cyclic phenomena on the climate, the assumption of independence in datasets is not accurate.

Stationarity in statistics refers to data that, though dependent on each other, exhibits homogeneous behaviour. A simple example is a sine wave, where the individual points are dependent on the amount degrees/radians, but the average long term behaviour remains the same. Typically, no adjustments need to be made to distributions assuming independence if the period of the cyclic behaviour is identified. That is why AMS are mainly used in hydrology, as the effect of seasonality is avoided by using a period that includes all four seasons. However, annual periods in hydrology are not enough to avoid dependence, as cyclic effects like ENSO transcend annual periods. In addition, stationary sequences assume no underlying trends in the data. If the climate is to become more extreme under climate change, with rainfall intensities expected to increase over time, stationary statistics are insufficient to describe this.

Non-stationary statistics accommodate data undergoing some systematic change in time, by assuming that the distribution changes over time. Non-stationary EVT can roughly be divided into two branches: parametric and non-parametric non-stationarity, though other approaches to non-stationarity are also used.

2.6.5.1 Parametric

The parametric non-stationary model, as the name implies, describes the parameters of a distribution according to a time effect or a covariate. A time effect can be a simple linear relation of a

Stellenbosch University http://scholar.sun.ac.za

parameter to the time of the data points, while a covariate describes a phenomenon or effect that can correspond with a parameter. Coles (2001:113,114) gives an example where high values of the Southern Oscillation Index correspond to high annual maximum sea-level values and are incorporated as a covariate.

The advantage of the ML procedure is the ability to easily incorporate time effects and covariates into the parameters of the distributions. For example, a linear change in the location parameter of the GEV over time can be modelled by:

$$\mu(t) = \beta_0 + \beta_1 t$$

where t is the corresponding time units of the data series, and are incorporated for the parameters β_0 and β_1. Higher order approximations, like a quadratic or exponential model, or even some external effect like sea surface temperatures or an ENSO index could be incorporated as a covariate. One can add a variety of approximations to the parameters, but infinitely many parameters can be added until it fits the data exactly, which takes away the explanatory power of the model. The desired model is the simplest model that can explain the greatest amount of variation in the data. The maximum likelihood ratio test, based on the deviance statistic, is used to compare more complex models to simpler models and is defined as (Coles, 2001:109)[4]

$$D = -2\{l_1(M_1) - l_0(M_0)\} \tag{2.3}$$

where

- D is the deviance statistic
- M_1 and M_0 are the more and less complex models respectively
- l_1 and l_0 are the log-maximum likelihoods for each model

The more complex model, M_1, is accepted if $D > \chi^2_{\alpha,k}$, where $\chi^2_{\alpha,k}$ is the 95% quantile for the chi-square distribution, where k is the difference between the amount of parameters of the models. For example, a model with a linear time related non-stationary model can be tested against the stationary counterpart. In addition, each parameter can be tested for significance by applying a t-test for some α level of significance, similar to regression techniques.

2.6.5.2 Non-parametric

The assumption regarding stationarity in EVT is that the underlying distribution and consequently, the calculated values remain the same over time. The non-parametric approach is to

[4]Please note that in the definition of Coles (2001), the $-$ sign in front of the 2 is missing.

Stellenbosch University http://scholar.sun.ac.za

divide the data into periods, either distinct or overlapping, fit a stationary distribution to each period and obtain the respective values. If the series of values remain constant over time, taking sample variability into account, one can conclude that the data is stationary. If the data shows any trends or changes, the data has non-stationary properties.

The disadvantage of the non-parametric approach is that it is not possible to use it to quantify the relationship between the parameters of the distribution and time.

2.6.5.3 Non-stationary application

The application of non-stationary EVT is rare in hydrology (Katz et al. 2002:1294, Beguería, Angulo-Matínez, Vicente-Serrano, López-Moreno & El-Kenawy 2011:2103, Jakob, Karoly & Seed 2011:2264), though a number of international researchers have investigated non-stationary EVT using either the parametric or non-parametric approach for both the GEV and GPD (Brath, Castellarin & Montanari 2001, Li, Cai & Campbell 2005, Katz et al. 2005, Park, Kang, Lee & Kim 2011, Beguería et al. 2011, Jakob et al. 2011). No literature was found applying non-stationary EVT to hydrological data in South Africa.

2.6.6 Summary of extreme value theory

EVT is a helpful way to describe the behaviour of extremes. The classical approach to EVT is by using either an event maxima or the POT model. Whilst event maxima distributions, like the GEV are typically applied using an AMS, there is a great loss of information. The POT method used in conjunction with the GPD, though rarely applied, incorporates more data which is helpful in assessing general trends in extremes. The determination of the threshold remains the greatest weakness of the GPD, as the model is very sensitive to the choice of threshold and no objective method exists to determine it.

The use of PWM/L-moments to estimate the parameters of EVT distributions is gaining popularity as they perform well for small samples and are simple to calculate, but the determination of variance in the estimates and the incorporation of covariates into the parameters are more complex and rarely applied in a South African context. Variance and covariates are incorporated more simply in ML techniques but the performance for small samples is poor and numerical optimisation is required, though an optimal solution might not always be reached.

Most hydrological design approaches assume stationarity within the data, but are unable to assess any changes of the data over time, like potential impacts of climate change. Non-stationary EVT is a helpful way to assesses the influence of time on the data using either a parameteric or non-parametric approach. Non-stationary applications to hydrological data are rare and no literature was found applying non-stationary EVT to a South African context.

Stellenbosch University http://scholar.sun.ac.za

2.7 Summary of literature reviewed

The earth's climate consists of a complex array of interactions and processes with the sun as the primary source of energy. Climate change refers to any indication of deviations from the expected climate. Various factors can cause changes in climate, like variation in the sun's output energy or oceanic-atmospheric phenomena like ENSO, but the enhanced greenhouse effect, fuelled by anthropogenic fossil fuel emissions has been blamed for increasing global temperatures and the occurrence of extreme weather phenomena. Climate models which simulate the climate by complex mathematical and physical models, are used to assess projections of increasing greenhouse gas emissions in the world and the results from these models indicate that the globe could expect a more extreme climate.

Under an enhanced greenhouse effect with increasing temperatures, the hydrological cycle is also expected to be enhanced, with the atmosphere storing more moisture and this could give rise to more intense rainfall. Though there is general agreement that temperatures are increasing in South Africa, both observed and projected data from studies show less agreement, most likely because of the complex nature of rainfall. However, there seems to be some agreement on increasing rainfall intensities. Whilst studies on short duration rainfall are more able to assess changing intensities than studies on daily rainfall, literature investigating possible changes in short duration rainfall is extremely limited, most likely because of the lack of reliable and long records. The findings of one recent study did however indicate possible increasing rainfall intensities.

The design of storm water systems typically makes use of IDF curves which implement extreme value theory. There have been a number of developments in EVT, including assessing non-stationarity in data, which is vital considering the possibility of increasing intensities under climate change. Application of non-stationary statistics to extremes remains rare in South Africa, indicating that there is room for further research in this field.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Chapter 3

Methodology

From the literature reviewed in Chapter 2 it is evident that observed and projected climate change impacts on rainfall in South Africa and the Western Cape varied largely based on locality and geography. Determining a single response from all the literature sources was difficult, although there were indications of greater variability, with rainfall expected to become more extreme and intense. Furthermore, increasing intensities are not only the consequence of increasing magnitudes, but a change in frequency of occurrence as well.

The lack of studies based on short duration rainfall (SDR) in South Africa, especially related to climate change, show the need for greater investigation into SDR, since engineering problems like the design of storm water systems require the consideration of storm durations shorter than 24 hours. Studies of SDR can give greater insight into changing characteristics of rainfall. However, it became clear that SDR was problematic, with many of the digitised autographic records containing numerous errors. The aforementioned highlights the need for careful selection of SDR stations for further analysis.

The current practice of setting up intensity duration frequency (IDF) curves using extreme value theory (EVT) has shortcomings, as the stationarity of the data is assumed, and it is therefore not possible to assess changes and/or impacts over time. In the light of possible changes in rainfall intensity, non-stationary EVT addresses the shortcoming by allowing time dependent factors into the EVT distributions and will be incorporated into the analyses. Finally, in the light of the importance of analysing changing frequency of occurrence in SDR, the annual frequency of exceedance of a threshold value was determined for each station. The following methodology was followed:

1. Reliable SDR data for the Western Cape and South Africa was obtained (Section 3.1)

2. The autographic and tipping bucket data was processed and combined for analyses (Section 3.2)

Stellenbosch University http://scholar.sun.ac.za

3. SDR was analysed for possible trends in the magnitude and frequency changes or trends (Section 3.3)

As a result of the sheer amount of data involved, the *awk*[1] and *R*[2] programming environments were used to automate most of processing and analyses.

3.1 Data acquisition

According to Smithers and Schulze (2000), the majority SDR stations situated in South Africa are owned by the South African Weather Services (SAWS). Of the 412 stations analysed in this report, 334 were provided by the SAWS. Being aware of the shortcomings of the SAWS' data, some of the other data providers were approached, but the data was either insufficient for the purposes of the analysis or no appropriate response was given to the requests. With this in mind, the SAWS' SDR database was used.

Prior to 1994, the SAWS' SDR was mechanically recorded as autographic rainfall and as mentioned in Section 2.4.3.2, the data underwent a digitising process. Since 1994 however, the SAWS' SDR rainfall data has been recorded using automated weather stations (AWS), where a station is equipped with a tipping bucket system that records rainfall every 5 minutes electronically.

Stations were initially selected according to the following criteria:

- A focus on the Western Cape
- Continuity between the autographic and AWS records
- A long record length (at least 25 years of near continual data) with as little missing data as possible

The selection was done with the aim to synchronise the two available record types to extend the available record length of the data.

3.2 Data processing

After data acquisition, data had to be processed into a usable format and checked for errors and quality control procedures had to be applied. Since two types of data sets were used, different

[1] *awk* is a simple text editing programming language. For more information, see `http://awk.info` [accessed 2012, September 17]

[2] *R* is an high-level open source statistical programming language based on the *S* language. For more information, see `http://cran.r-project.org` [accessed 2012, September 17]

Stellenbosch University http://scholar.sun.ac.za

methods were followed to process each type. In order to combine the two datasets, it was aimed to convert the digitised autographic data into 5 minute intervals to be combined with the AWS.

The data processing routine that was followed is summarised in Figure 3.1. The specific procedure and the associated file name is given in each block. The various procedures will be discussed below.

3.2.1 Digitised autographic data

Digitised autographic data was received from the SAWS which represents rainfall up until 1992. The autographic data was originally recorded on recording paper on a cumulative scale from 0mm rain until a total of 10mm was reached, at which point the logger would undergo a siphoning event and start from 0 again. A new chart would be placed into the gauge every morning at about 08:00.

3.2.1.1 Process in readable format

The raw data came in a compact digital format that required further editing for usability. The `autogr.raw.edit.awk` script was used to edit the raw data into a more readable format. Table 3.1 shows a sample of the raw compacted data, while Table 3.2 shows the edited format of the boxed section in Table 3.1. Notice that the autographic rainfall value occurs after the space that separates the columns in Table 3.1.

Table 3.1: Sample of raw digitised autographic data

```
00104258601010750 38030737  0031045  0031347  3 031615  3031620  6031625  8
00104258601031823  8031909  9031926  9032005 10032018 10032045 11032102 12
00104258601032109 15032113 19032125 23032133 25032143 27032155 27032209 29
00104258601032217 31032226 33032237 36032246 41032256 45032313 49032323 50
```

Table 3.2: Sample of edited digitised autographic data (Please note that the "time in seconds" column refers to the time in seconds since 1970/01/01 00:00:00).

year	month	day	hour	minute	second	time in seconds	autographic rainfall (mm)
1986	1	1	7	50	0	504942600	3.8
1986	1	3	7	37	0	505114620	0.0
1986	1	3	10	45	0	505125900	0.0
1986	1	3	13	47	0	505136820	0.3

3.2.1.2 Removing the cumulative effect

The digitised autographic data is given in a cumulated format, where each consecutive rainfall value is added to the previous value, until a siphon event occurs or a meter reader resets the

Stellenbosch University http://scholar.sun.ac.za

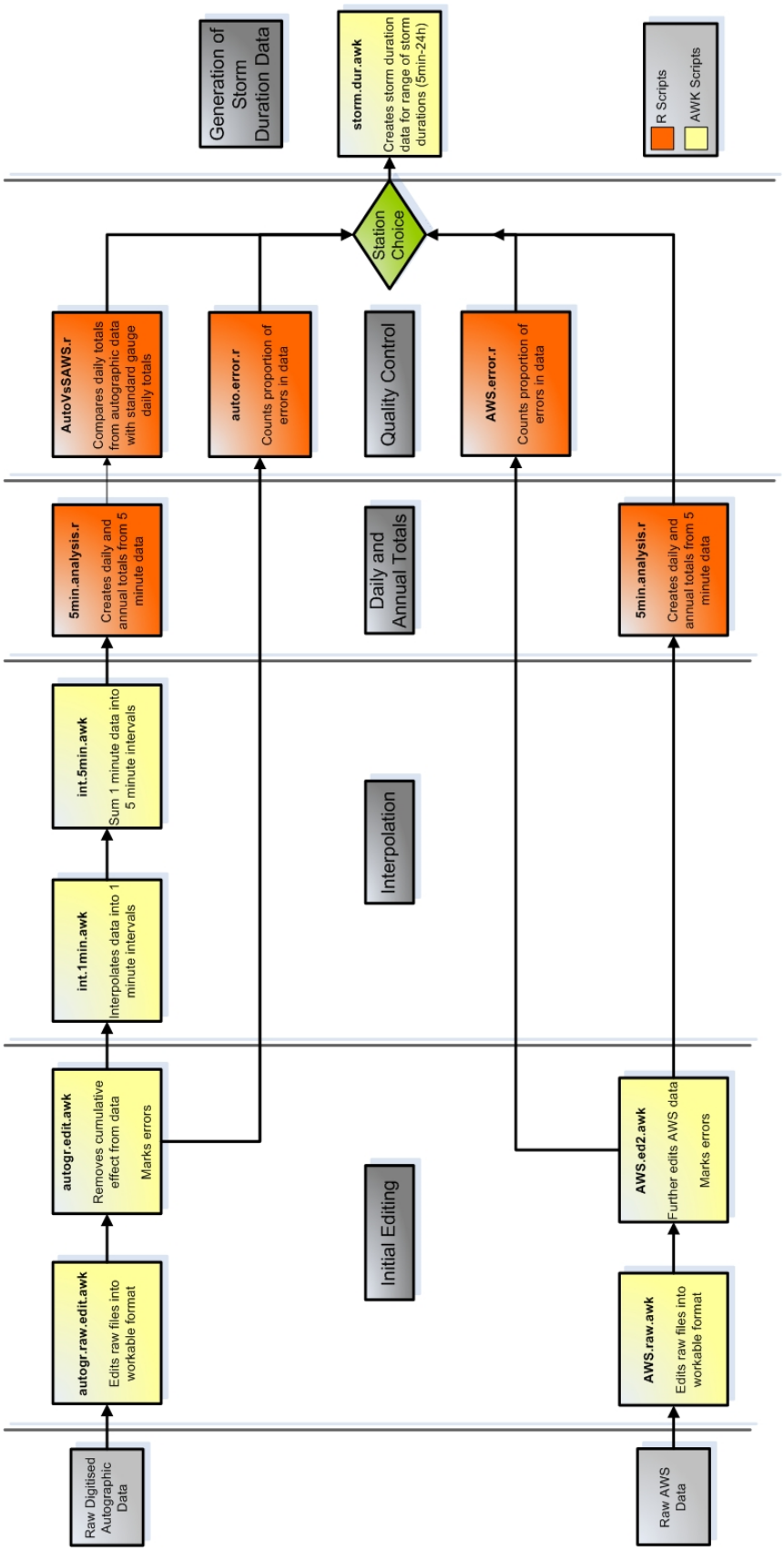


Figure 3.1: Data Processing Flow Scheme

Stellenbosch University http://scholar.sun.ac.za

gauge. A siphon event occurs when the accumulated rainfall reaches a peak of approximately 10mm and automatically resets to 0mm.

The `autogr.edit.awk` script was used to remove the cumulative effect of the digitised autographic rainfall by calculating the difference between consecutive rainfall values, with the exception of siphon events. A summary of the typical procedures can be seen in Table 3.3. Theoretically, the rainfall reading can decrease during either a siphon event or a meter reading reset.

In spite of the digitising process removing a number of errors and impossibilities, a number of inconsistencies were found while trying to remove the cumulative effect. These included small decreases in the cumulative rainfall and false and improper siphon events. A false siphon event occurs when the meter reader resets the gauge and this is erroneously recorded as either a siphon event, or a near siphon event when the decrease in rainfall does not reset to exactly zero. An improper siphon event occurs when the gauge did not reset the rainfall to exactly zero during a siphon event. In addition, the autographic record contained flagged errors from the digitising process. A number of causes could be attributed to these error flags, like ink drying up in the instrument, ink smudges occurring, or a general difficulty to digitally log the event from the rain chart. However no additional detail was provided to indicate the specific type of problem when a flag event occured. Digitising errors were labelled as two successive "-999" values on the raw records.

A flagging method to detect and correct errors was also applied to the `autogr.edit.awk` script. Digitising error flags were recorded as **e1**, while small decreases in accumulated rainfall was recorded as **e2**, and large decreases were recorded as **e3**. Type **e1** errors were removed by removing one of the digitising values and making the difference between both the previous value and the value after the digitising error zero.

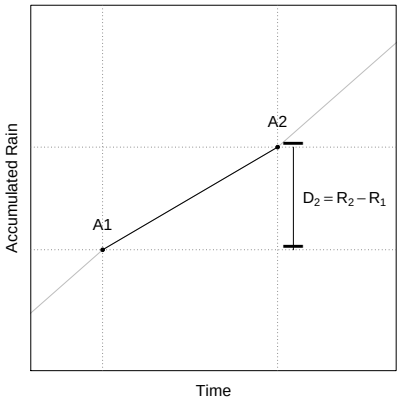
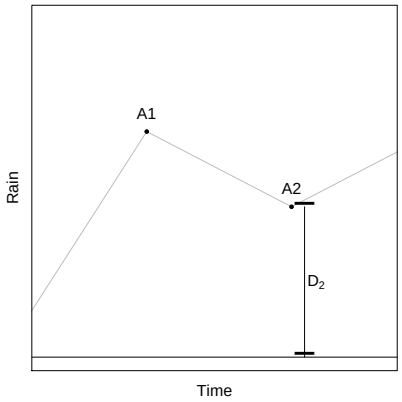
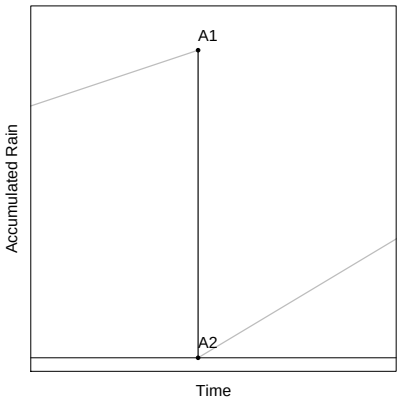
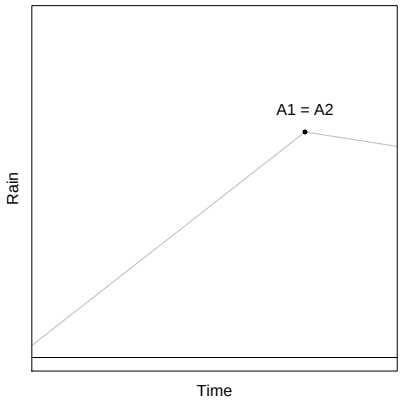
Type **e2** events were considered as decreases smaller than 0.3mm, and the resulting rainfall was recorded as if no rainfall occurred. Type **e3** events were logged to prevent false siphon events and included all decreases greater than 0.3mm, with the two previous rainfall events recorded as 0mm events. The value of 0.3mm was chosen to include as many events as possible without being flagged as a false siphon event. Table 3.4 summaries the errors and resulting actions taken. Please note that the "Actual Rain" column refers to the rainfall after the cumulative effect is removed.

3.2.1.3 Transformation into 5-minute data

Autographic data was converted into five-minute data for comparison with the AWS data. The first step was to interpolate all the data into one minute intervals, as one minute was the smallest recorded interval found in the data, and the next step was to sum the one minute data into specific 5 five minute intervals. The `int.1min.awk` and `int.5min.awk` scripts were used for

Stellenbosch University http://scholar.sun.ac.za

Table 3.3: Typical procedure to remove cumulative effect from the autographic record.

Type	Action	Actual Rain
Normal Event 	Take the difference between consecutive accumulated rainfall values (A_1, A_2) to obtain the actual rainfall (D_2)	
Siphon Event 	Difference is zero with no change in actual rainfall	

these procedures respectively. Some of the recorded autographic readings were very small, <0.2mm, and occurred over long time periods. As a result, some interpolated values contained very low rainfall measurements per 5 minutes.

3.2.1.4 Quality control

Standard gauge daily rainfall (SG[3]) data was requested from the SAWS for the corresponding autographic stations in order to assess the performance of the autographic data.

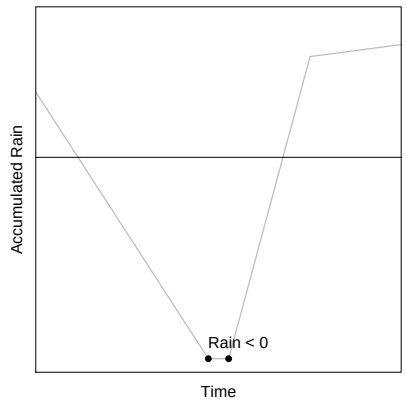
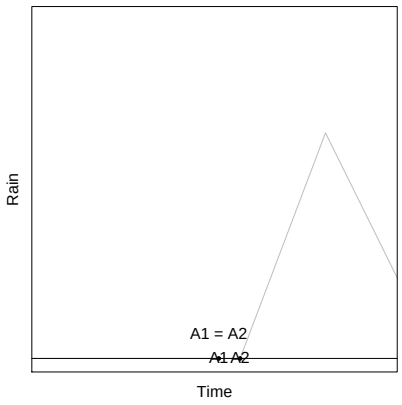
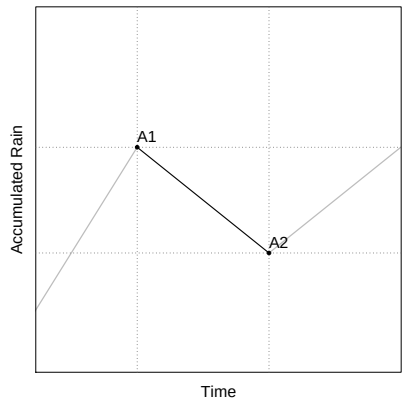
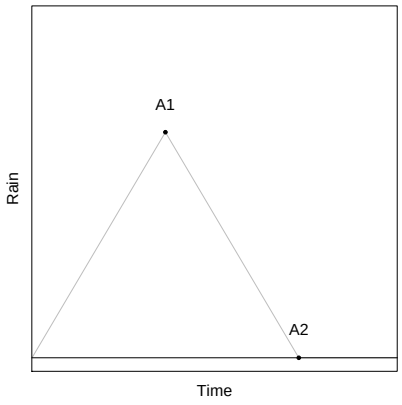
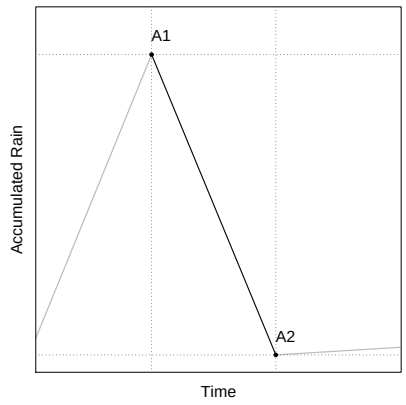
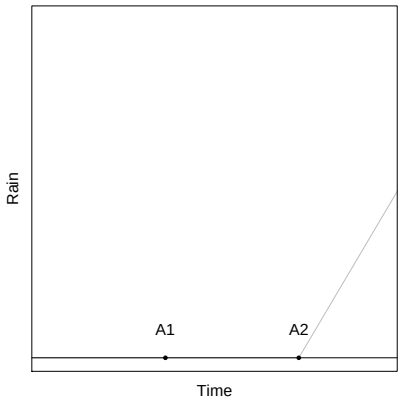
The `5min.analysis.r` script was used to generate daily and annual rainfall totals, and these were compared to the corresponding SG data with the `AutoVsSAWS.r` script. The following aspects were investigated for both daily and annual totals:

- Cumulative rainfall

[3]Please note that all occurrences of the abbreviation SG will refer to "standard gauge daily rainfall" in its entirety.

Stellenbosch University http://scholar.sun.ac.za

Table 3.4: Error types in digitised autographic record

Type	Label and Action	Actual Rain
Error Flags 	**e1** Remove two successive -999 error flags from record and flag first rainfall event after the error. Differences before and after errors are zero	
Small decrease in rainfall (< 0.3mm) 	**e2** Current difference is zero resulting in no change in rainfall	
Large decrease in rainfall (> 0.3mm) 	**e3** Current difference and previous difference are zero	

- Difference between rainfall (for annual totals)
- Ratio of autographic/SG

Stellenbosch University http://scholar.sun.ac.za

The rejection criteria of the stations was applied on an arbitrary basis, but stations showing large deviations in annual totals (>20%) were not used for analysis. In Du Plessis (1992), where autographic data of the City of Cape Town was compared to SG data, it was found that the autographic rainfall was generally lower than the SG. The underestimation found in Du Plessis (1992) was considered insignificant, especially for smaller durations of 5 to 60 minutes. The same assumption was made for the data in this thesis.

3.2.2 AWS data

AWS data was supplied by the SAWS and represents data recorded since approximately 1994, where the total rainfall over 5 minute intervals is recorded in millimetres.

3.2.2.1 Processing the data

Raw data received from the SAWS was in a format listing the rainfall of each five minute interval. Table 3.5 shows a sample of the raw data. Processing involved two steps with the `AWS.raw.awk` and `AWS.ed2.awk` scripts formatting the data in a more workable format and logging missing data and gaps in the record. Table 3.6 shows a sample of the result of using the formatting scripts.

Table 3.5: Sample of raw 5-minute data

```
"ClimNo" "StasName" "Date" "Time" "Rain"
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:05:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:10:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:15:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:20:00 0.0
```

Table 3.6: Sample of edited AWS data (Please note that the "time in seconds" column refers to the time in seconds since 1970/01/01 00:00:00).

year	month	day	hour	minute	second	time in seconds	5 minute rainfall (mm)
1994	10	19	2	5	0	782525100	0
1994	10	19	2	10	0	782525400	0
1994	10	19	2	15	0	782525700	0
1994	10	19	2	20	0	782526000	0

3.2.2.2 Quality control

Two error types were encountered with the AWS data. Gaps in the records were found, where consecutive readings would have a difference of more than 5 minutes. Some readings also had

Stellenbosch University http://scholar.sun.ac.za

blank rainfall entries. Quality control was applied by measuring the number of these errors per year for each station.

3.2.3 Selection of stations and combining data

After the processing of the data from the stations and applying quality control, stations were selected according to the original criteria of a usable and reliable record length of at least 25 years. Attention was given to specific stations undergoing location changes as rainfall is topographic and location sensitive. Large spatial changes in height and/or distance with a station relocation were rejected for further analysis. Details regarding the selected stations will be discussed in Sections 4.1 and 4.2.

3.2.4 Processing into storm durations

The `storm.dur.awk` script was used to calculate storm rainfall for the final list of stations for durations 5, 10, 15, 30, 45, 60, 90 minutes and 2, 4, 8, 12, 18, 24 hours, typically used for IDF curves. Storm rainfall was obtained by using a moving summation over the rainfall record that is equal to the particular storm duration. Two outputs types were generated, one including all storm rainfall events used in the frequency analysis (Section 3.3.1), and a second that selects the maximum event per storm, used in the magnitude analysis (Section 3.3.2). The methodology as proposed by Ben-Zvi (2009:110) (also see Section 2.6.2), where a rainfall event is defined as a series of events separated by at least 24 hours of zero rainfall from another series of events, was used to select peak events for the POT/GPD methodology.

3.3 Data analysis

The main purpose of the analysis of the data is to detect the possibility of cycles or trends in the SDR stations. The analysis consists of two parts: analysing the magnitude of the SDR using non-stationary EVT, and analysing the frequency of occurence of SDR by using a modified approach to that implemented by Van Wageningen (2006). Trends that were found to indicate increasing magnitudes of SDR or any changing frequency are considered to be supporting a possible change in the climate, based on expected increases in the intensities of rainfall.

3.3.1 Magnitude analysis

The magnitude analysis entailed the application of non-stationary EVT on the SDR. The `EVT.Analysis.r` script was used for the calculations in this section.

Stellenbosch University http://scholar.sun.ac.za

3.3.1.1 Choice of EVT distributions

From the literature reviewed in Section 2.6, the GEV and GPD along with AMS and POT datasets were selected for analysis. Whilst the GPD/POT methodology is unpopular in South Africa, POT methods have the advantage of incorporating a greater amount of data points compared to the AMS methodology of only one peak per year. This is especially useful with the short SDR records available in South Africa.

The maximum likelihood (ML) method was used to estimate the parameters of the distributions. The ML method has the advantage of easily calculating the confidence intervals of the quantiles and easily incorporating the parametric non-stationary approach. Since the ML method requires numerical optimisation, the `ismev`[4] and `evir`[5] packages for the R environment were used and modified to apply the method and obtain the parameters and return levels.

A critical factor in the GPD is selecting the threshold value, as a too low threshold undermines the independence of events assumption, while a too high threshold leads to a loss of data and large variance (Coles, 2001:78). Proposed methods to determine the threshold are very difficult to interpret and the threshold choice remains subjective (see Section 2.6.2). In light of this, three threshold types were used in the analysis:

1. Cut-off values used by Du Plessis (1992)
2. 1.5 times the Du Plessis (1992) values
3. The stable shape parameter method

The values used by Du Plessis (1992), seen in Table 3.7, are based on SDR for the City of Cape Town and are used to ensure that approximately the ten worst storms be analysed every year (Du Plessis, 1992:4). The du Plessis values were also multiplied by 1.5 to ensure that only the highest events be selected, especially for areas experiencing greater rainfall than the Cape Town area.

Table 3.7: Thresholds based on Du Plessis (1992) cut-offs. Values in millimetres.

Duration	minutes							hours					
	5	10	15	30	45	60	90	2	4	8	12	18	24
Du Plessis (1992)	2	3	4	5	6	8	10	12	12	12	12	12	12
1.5 × Du Plessis (1992)	3	4.5	6	7.5	9	12	15	18	18	18	18	18	18

[4]This package is strongly based on the theory in Coles (2001). See `http://cran.r-project.org/web/packages/ismev/` [accessed 2012, September 17] for more information on this package.

[5]See `http://cran.r-project.org/web/packages/evir/index.html` [accessed 2012, September 17] for more information on this package

Stellenbosch University http://scholar.sun.ac.za

The stable shape parameter method was developed with the support of the Department Statistics and Actuarial Science at the University of Stellenbosch. The following process is followed in the determination:

1. Create a series of exceedances based on the specific storm duration data
2. Fit a GPD for each number of exceedances and obtain the corresponding shape parameter
3. Select the most stable shape parameter region
4. Return the closest threshold value that would produce the stable shape parameter

The number of exceedances is inversely related to the threshold, the higher the number of exceedances, the lower the threshold. A range of exceedances corresponding to a range of thresholds is applied to the data and a GPD is fitted to each instance of the number of exceedances or threshold.

Since the shape parameter, which strongly influences the estimated quantiles, is very sensitive to the threshold choice, a region is selected that will produce the lowest amount of variance in the shape parameter. Any threshold chosen in that region will produce similar quantiles.

The stable region is selected by using a moving window representing 25% of the number of exceedances and applying the instability measure. The window region with the smallest instability is selected and and a corresponding average shape parameter is calculated. The threshold that produces the closest shape parameter to the average shape parameteris selected . The instability is calculated as:

$$\theta^2 = \text{var}(y) + b^2 \tag{3.1}$$

where:

- θ^2 is the instability measure
- $\text{var}(y)$ is the variance of the shape parameter window
- b is the slope of a fitted linear regression to the shape parameter window

Figure 3.2 displays the selection procedure graphically. For each number of exceedances, a corresponding shape parameter is obtained and plotted, indicated by the solid line. The instability measure identifies the most stable 25% or quarter region, indicated by the two dashed vertical lines. The average shape parameter of the 25% region is indicated by the solid horizontal line, and the value closest to the average value is encircled. The corresponding number of exceedances is returned and the corresponding threshold is used for calculations.

Stellenbosch University http://scholar.sun.ac.za

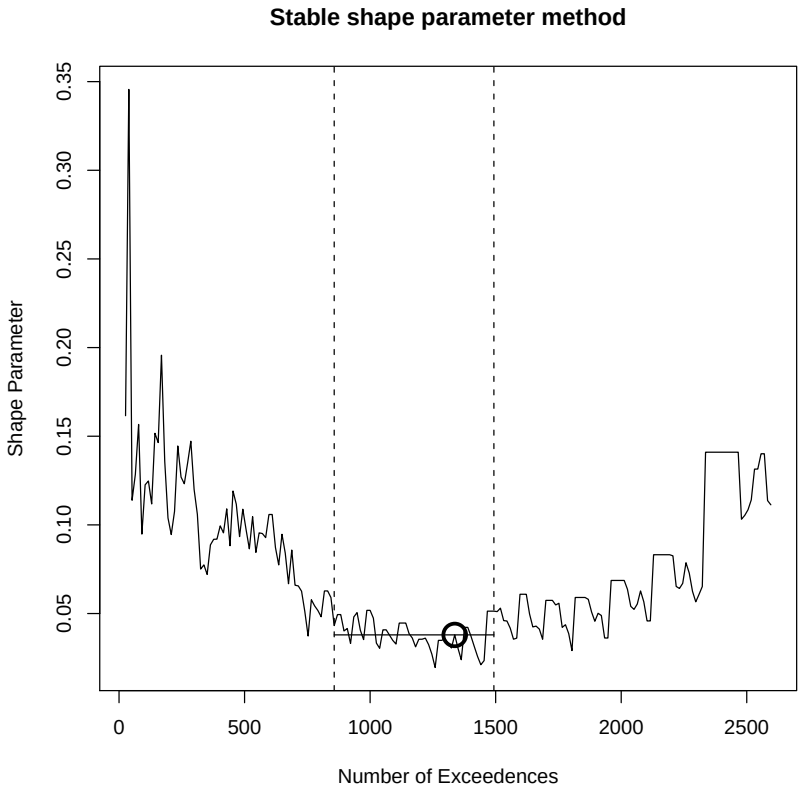


Figure 3.2: Stable shape parameter plot.

3.3.1.2 Parametric non-stationarity

The parametric non-stationary (PNS) analysis adopted followed the process outlined in Coles (2001:105-123). Analysis was limited to a linear time model:

$$\text{par}(t) = \beta_0 + \beta_1 t$$

where

- par is a parameter of the EVT distribution used
- β_0 is the intercept of the fitted line
- β_1 is the slope of the fitted line
- t is the time

Each parameter's significance was measured by applying the t-test to the slope of the line with an $\alpha = 0.05$ level of significance. Each parameter was tested individually and in order. If a parameter obtained a significant fit, it was incorporated into the next parameter's test. If any of the parameters showed a significant fit, the non-stationary model was compared to a stationary

Stellenbosch University http://scholar.sun.ac.za

model using the likelihood test, discussed in Section 2.6.4, to assess if the more complicated model was justified. If a non-stationary model was justified, then the non-stationary model was deemed significant. For the GEV model, if the stationary model was justified, a non-stationary Gumbel model was analysed in addition. If the non-stationary model was justified, the estimated quantiles were plotted against the original dataset to assess the change over time.

A general weakness of the parametric approach adopted for this analysis is the exclusion of higher order models and the use of covariates like ENSO or sunspot numbers which can possibly describe extreme events more appropriately.

3.3.1.3 Non-parametric non-stationarity

The non-parametric non-stationarity (NPNS) analysis is based on the idea that a stationary sequence of data, which is usually assumed in EVT analyses, should produce stable return levels over time. Therefore, two different data sections with an equal length in time should produce similar quantiles. Any changes over time can be considered to be violating of the stationary assumption of the data.

The NPNS analysis is based on fitting EVT distributions to a moving window and obtaining the return levels for the corresponding return period, similar to a moving average calculation. Three window periods, 15, 20 and 25 years were used and return levels for return periods 5, 10 and 20 years were calculated along with the corresponding 95% confidence intervals for each return level. The range of window and return periods was used to determine if results are generally consistent. Return periods were kept low since the variance for large return periods became extremely large.

Figure 3.3 shows a graphical example of the procedure followed for the AMS/GEV approach. AMS values are illustrated on the top graph. A window period of 20 years, starting at 1956, is selected and a GEV is fitted to it. The return level for a return period of 10 years is plotted on the bottom graph. The next window period moves on one year, and the process is repeated for the entire record. As seen in the bottom graph, the return levels do not remain stable, but increase over time, indicating possible non-stationarity.

Each storm duration generates nine graphs per distribution, as three window periods are used combined with three return periods. For each return level and window period, linear regression is applied to the series of return levels and the significance of the slope is measured according to the t-test (for 95% level of significance). The number of significant positive and negative slopes is recorded for all distributions.

Stellenbosch University http://scholar.sun.ac.za

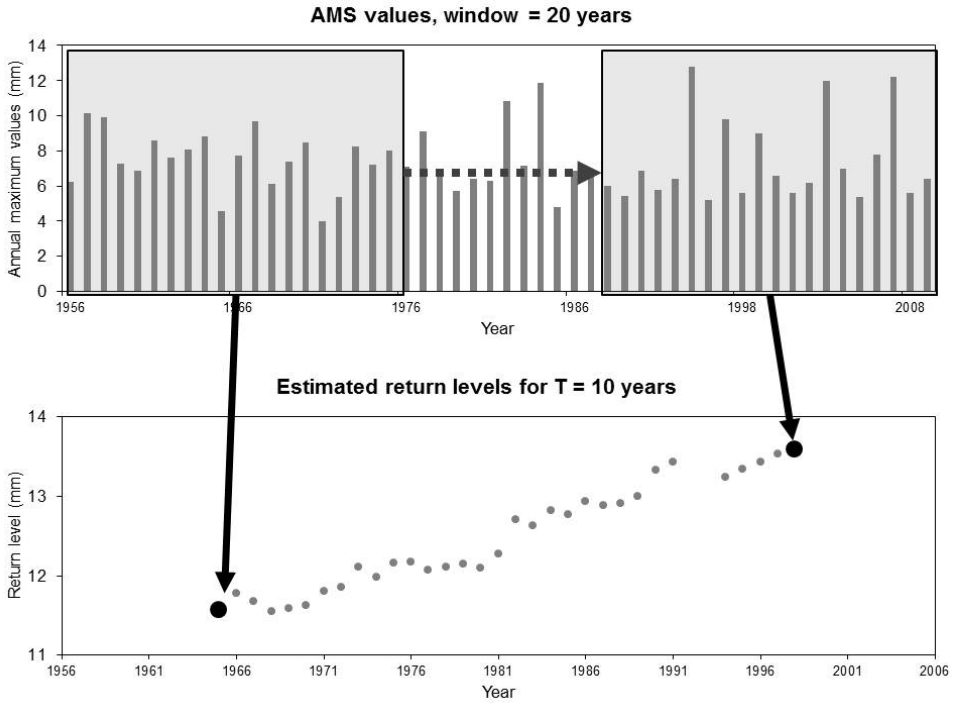


Figure 3.3: NPNS example for an AMS.

3.3.2 Frequency analysis

The frequency analysis used in this thesis is a modified approach of that found in Van Wageningen (2006:72-77). All 5-minute events exceeding threshold values 0.5, 1 and 2mm for each year were plotted for any visible trends. Rainfall above a threshold value was used instead of using all the data as in Van Wageningen (2006:72), since the interpolation technique for the autographic data includes low rainfall values (< 0.2mm) interpolated over long durations, resulting in extremely low and arguably unrealistic rainfall values that will be counted as an event. Higher thresholds were not used since no events exceeding the threshold occur during some years.

3.4 Limitations to analyses

The results of the analyses are to be interpreted with the following in mind:

- Two different SDR record types with varying quality are incorporated in this thesis
- Less than 60 years of total SDR data are available
- The climate signal is complex and spatially sensitive
- There is no single trend or projection for climate change in the Western Cape and larger South Africa

Stellenbosch University http://scholar.sun.ac.za

Consequently, the results of this thesis should not be interpreted as positively proving or disproving climate change, but rather as showing that there is a lack of or enough evidence available in the SDR to support climate change projections.

To illustrate the above mentioned difficulties, consider the following fictional scenario presented in Figure 3.4. In the figure on the left, a series of sine waves were presented, each with a different wavelength and assigned with a weight (or factor of influence) with shorter wavelengths having a smaller weight than longer ones. A weighted total was taken for all the sine waves, indicated by the solid bold line. The line effectively shows the total overall signal of when all the sine waves are incorporated. In the middle of the record, another effect, indicated by the dashed line, is introduced, which has the greatest overall impact on the total signal.

This illustration can be compared to the climate, where various effects influence the overall climate signal. The trend introduced in the middle of the record represents climate change, where a new effect has a strong influence on the overall signal. The figure on the right is a short extract from the figure on the left for the range 650-720°. With only this section available, it becomes extremely difficult to distinguish between trends, cyclic events and overall noise in the signal. This represents the current situation for climate change studies and this thesis, where only a small portion of observed data is available from the past. Whilst this illustration is very crude and oversimplified, it highlights the difficulties faced with climate studies.

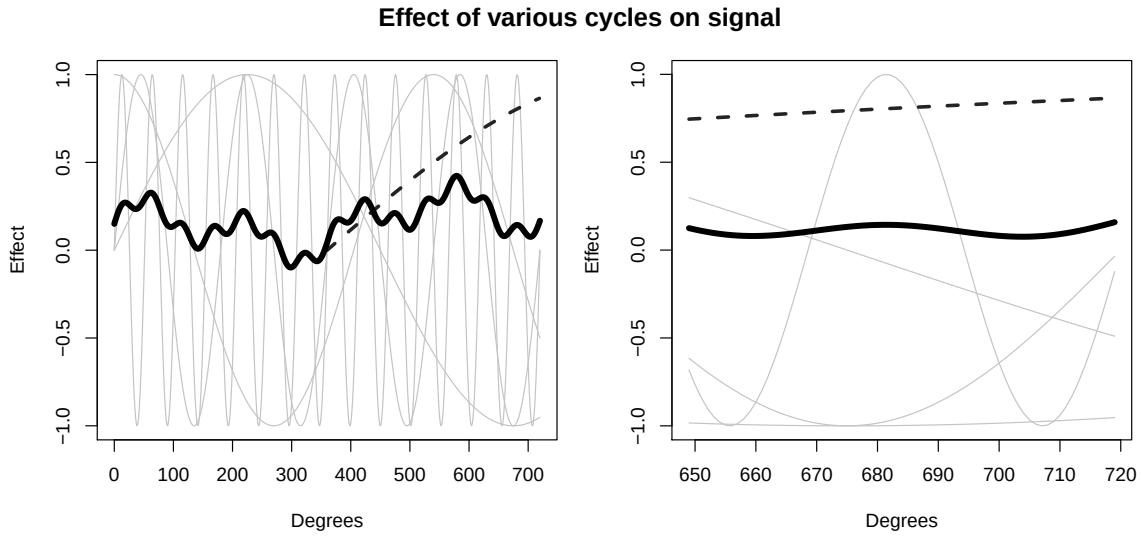


Figure 3.4: The effect of various sinus waves of different weights on the general signal. Cyclic events are indicated in grey, a strong additional effect is indicated by the dashed line, while the solid bold line represents the total effect of all the cycles and trends. The figure on the right shows a small portion of the figure on the right.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Chapter 4

Results

Results of the data acquisition, processing and analysis are presented in this chapter. Most of the processing and analyses were done using automation techniques, as the amount of data and subsequent results are large. Only a few examples of the results are presented here, with the complete results presented in the appendices.

4.1 Data acquisition

Figure 4.1 shows the initial autographic and AWS stations selected. Table A.1 and A.2 display the full list of stations for the autographic and AWS data respectively.

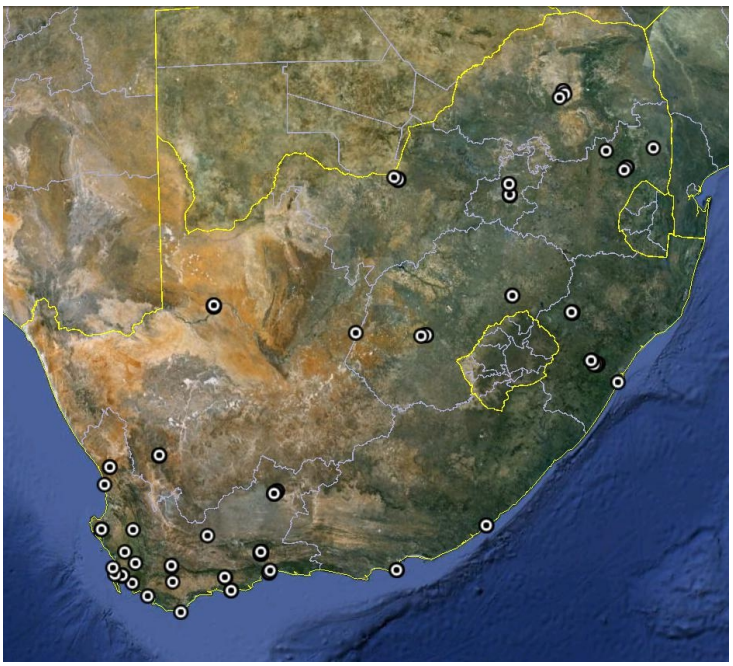


Figure 4.1: Location of initial autographic and AWS stations.

Stellenbosch University http://scholar.sun.ac.za

4.2 Data processing

Selected stations underwent an editing process described in the methodology.

4.2.1 Digitised autographic data

After initial editing, interpolation and the generation of daily and annual totals, the digitised autographic data underwent a two-part quality control procedure for the type of errors found in the removal of the cumulative effect, and for the comparison of daily and annual totals with standard gauge daily rainfall (SG) data of the same stations.

4.2.1.1 Error flagging

Error types **e1**, **e2** and **e3**, representing digitising error readings, small decreases in autographic rainfall and large decreases in autographic rainfall respectively, were logged and graphically presented for a single station in Figure 4.2. The top figure illustrates the annual proportion, in %, of autographic readings in the record that were flagged as either **e2** or **e3**. In the bottom figure the amount of time that **e1** readings occur in a year, indicated by days per year, were plotted, since **e1** errors covered a span of time.

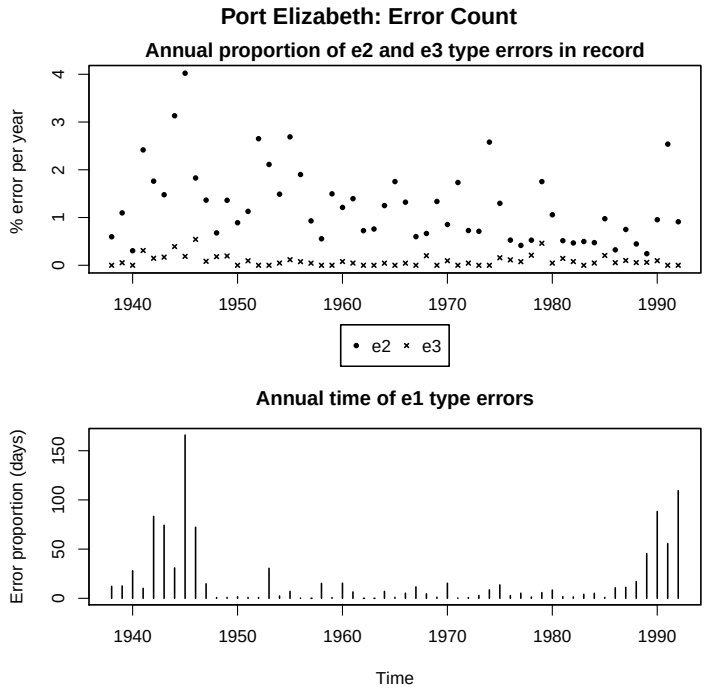


Figure 4.2: Annual proportion of errors in autographic record for Port Elizabeth (0035179). The upper figure illustrates the number of **e2** and **e3** recordings over the total amount of recordings per year. The lower figure illustrates the number of times **e1** error flags occur in days per year.

Stellenbosch University http://scholar.sun.ac.za

Figures 4.3 and 4.4 summarises the error reading proportions for all stations. Figure 4.3 presents the total proportion of time, in %, that **e1** errors occur in the entire record for each station, while Figure 4.4 presents the total proportion of **e2** and **e3** readings in the entire record for each station. Figure 4.3 illustrates that **e1** logging errors occur for a large part of the time in each station, with most stations having more than 10% of the total record time marked as digital errors and many exceeding 20%. As illustrated in Figure 4.4, type **e2** readings occurred more frequently than **e3** readings, with most stations ranging between 0.5 and 2% of total readings, while **e3** readings rarely exceeded 0.5%.

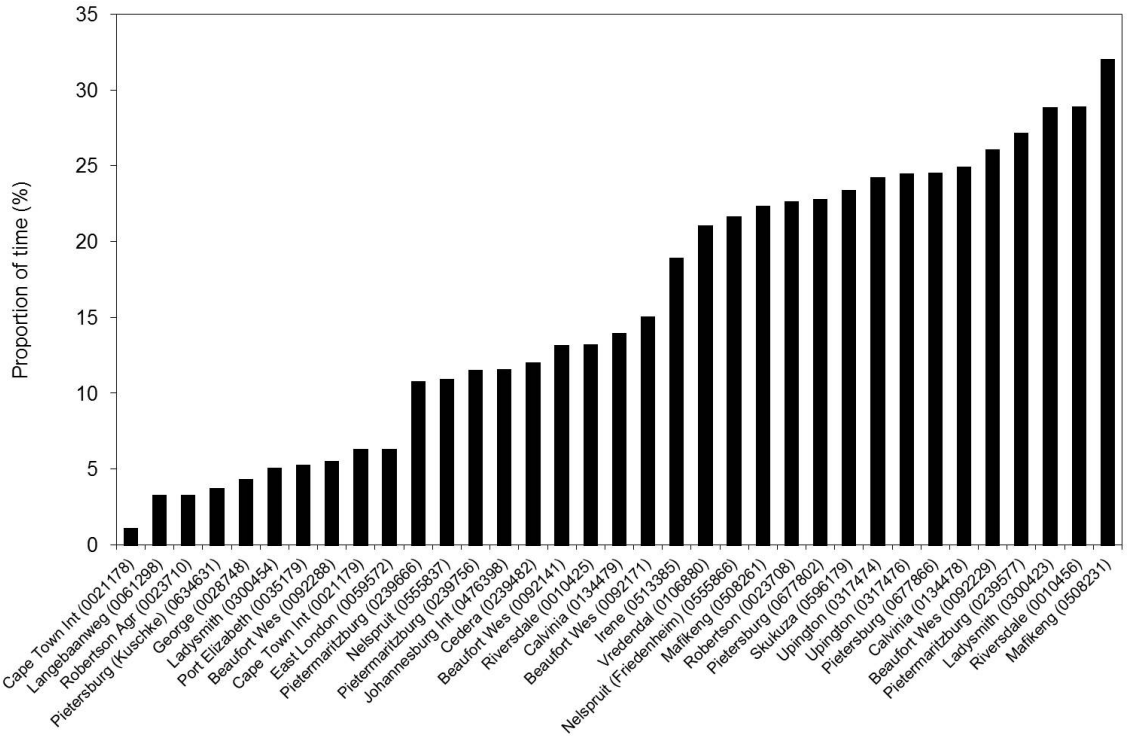


Figure 4.3: Proportion of **e1** readings in the autographic record in terms of time. The proportion was calculated as the total time **e1** readings occur over the total record time per station.

4.2.1.2 Quality control

Daily and annual[1] totals were generated from the autographic data and compared with SG data of the same stations. Figures 4.5 - 4.7 illustrate typical output generated from the comparison. Figure 4.5 illustrates SG data plotted against daily totals of the autographic data. In an ideal plot all the points fall on the 45°line. Figure 4.6 illustrates the ratio of autographic daily totals to SG data on a daily basis. Figure 4.7 illustrates the cumulative plots for both the autographic and SG data. Ideally, the two lines should not deviate from each other. Figure 4.8 compares annual totals of the autographic and SG data in the top graph, while the bottom graph presents the difference between the autographic and SG data. Ideally, the difference should be zero.

[1] In hydrological years, that is from October to September

Stellenbosch University http://scholar.sun.ac.za

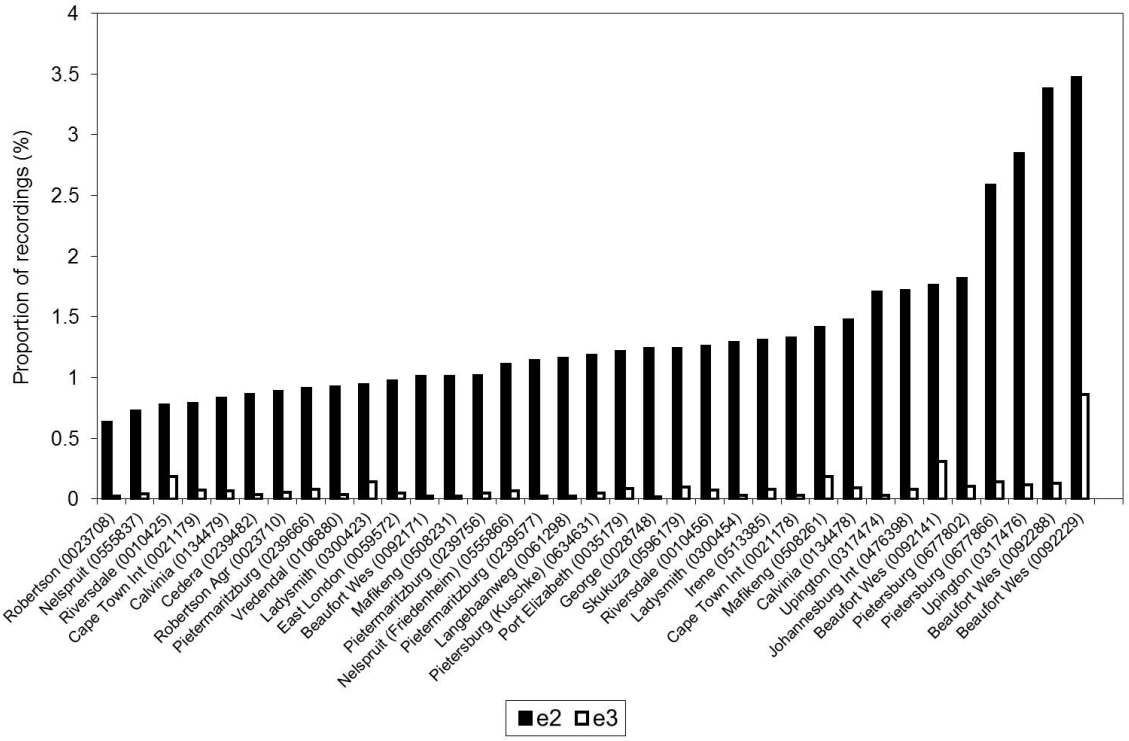


Figure 4.4: Proportion of autographic readings **e2** and **e3** occupy for autographic stations. The proportion was calculated as the total amount of **e2/e3** readings over the total amount of events per station.

A summary of all the stations is presented in Figure 4.9. Figure 4.9 illustrates the autographic/SG ratio of the total rainfall at each station, with 100% indicating no difference between the autographic and SG data. Most stations indicated that the autographic data underestimated SG totals, with only six stations exceeding the SG daily totals, which is in agreement with the findings of underestimation in Smithers and Schulze (2000:86-155).

4.2.1.3 Summary of quality control procedure

Results of the quality control procedure indicate that the autographic data contains large amounts of data that were flagged as error readings by the digitising team, which possibly accounts for the differences from SG daily totals, although manual inspection for certain stations indicated that some differences could not be accounted for by digitising errors only. In addtition, some of the stations indicated that some very large storm events were unfortunately not recorded, one such example is the storm of September 1st, 1968 in Port Elizabeth, where over 400mm of rain fell in four hours (SAWB 1968:16, Alexander 2001:298.)[2]. This inability of the autographic data to log extreme events could be attributed to the logger being unable to keep up with the storm event and it is consequently very difficult to interpret the data. A variety of other effects also caused error flags in the record, like ink depletion or smudges in the logger.

[2]Somehow this storm was successfully recorded using the autographic logger, but not included in the digitised results. The reasons for the omission was not found.

Stellenbosch University http://scholar.sun.ac.za

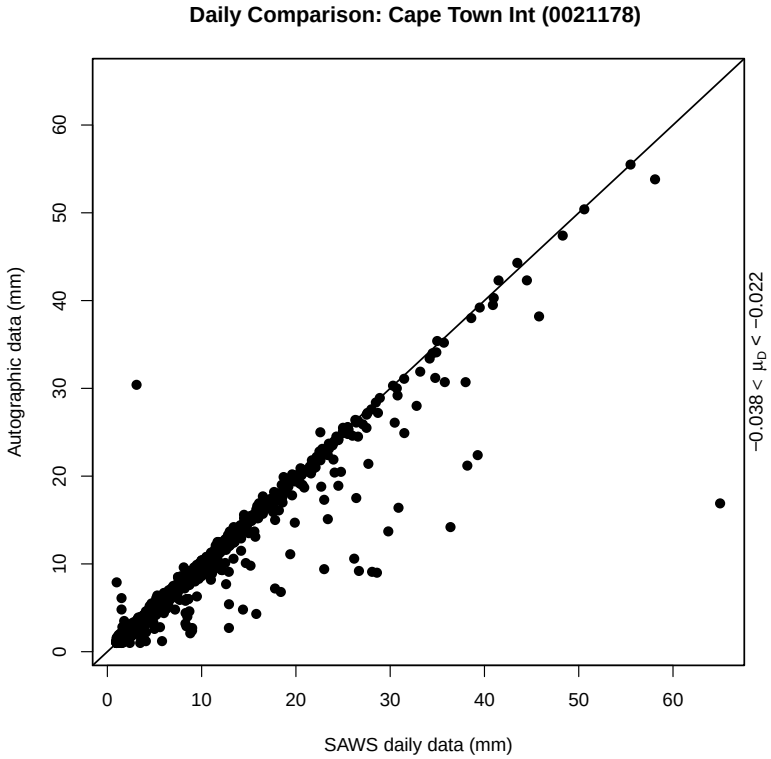


Figure 4.5: Cape Town International (0021778) SG data plotted against autographic daily totals

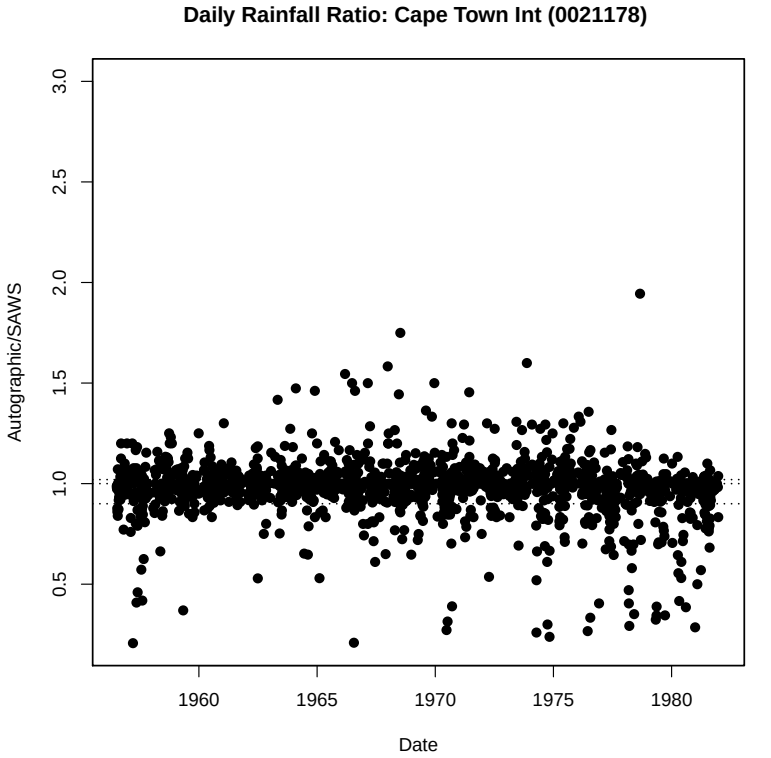


Figure 4.6: Cape Town International (0021778) ratio: Autographic/SG daily totals

Stellenbosch University http://scholar.sun.ac.za

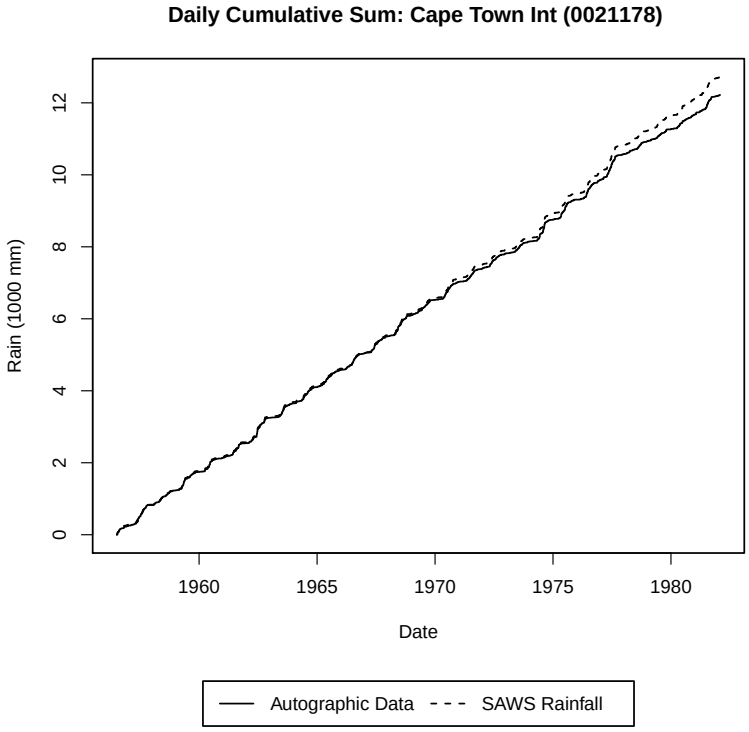


Figure 4.7: Cape Town International (0021778) cumulative sum comparison

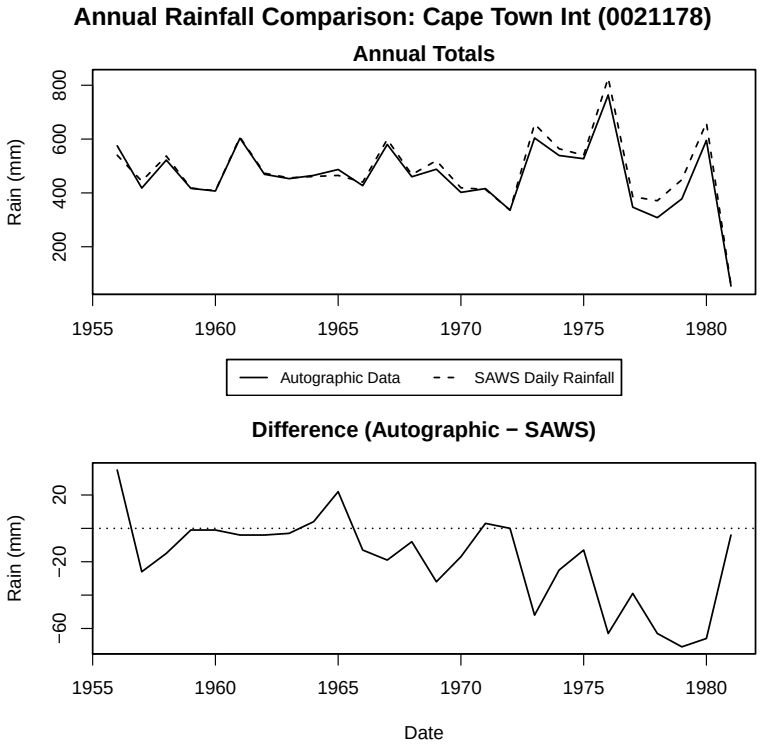


Figure 4.8: Summary

Stellenbosch University http://scholar.sun.ac.za

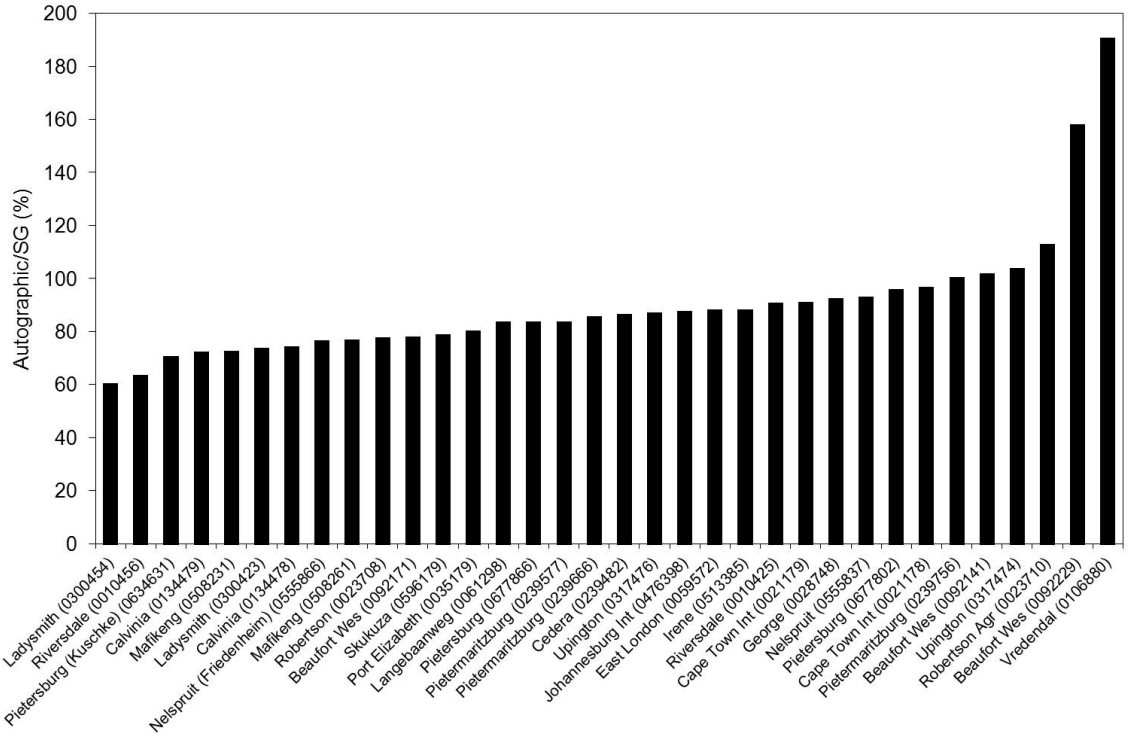


Figure 4.9: Summary of autographic/SG station rainfall totals.

There was particular difficulty in discerning between false siphons and improper siphons, resulting in improper siphon events being flagged as false siphon events which in turn resulted in some loss of information as the flagging method makes the previous difference zero. A investigation of three stations indicated that **e3** events took place infrequently and often in clusters. Therefore it was decided that fixing each station's improper siphon events manually would be too time consuming and that the overall loss of information would be minimal, compared to the effect of **e1** readings.

No consistent pattern in the comparison of the autographic with SG data was found and results varied per event and per station, which ruled out applying a scaling factor to the autographic data. In addition, the large number of times **e1** errors occurred in the record would cause unrealistic scaling. From the results of the comparison between digitised autographic and SG data, greater significance was placed on analysis results with shorter durations, especially $<$ 60 minutes, under the assumption that the underestimation of the autographic data would be negligible for shorter durations (see Section 3.2.1.4).

4.2.2 AWS data

The procedure followed for the AWS was generally simpler as the data was already in a 5-minute format. After initial editing, quality control procedures were applied to the data.

Stellenbosch University http://scholar.sun.ac.za

4.2.2.1 Quality control

The quality control procedure recorded readings with blank rainfall entries and gaps in the rainfall records for each station. Figure 4.10 illustrates a graphical representation of a station's error readings in terms of the amount of time in days lost due to blank entries and gaps per year. A summary of all the data of the stations is found in Figure 4.11, illustrating the total proportion of errors for each station. Figure 4.11 illustrates that most stations presented an error proportion of less than 3% in total, with only a few disproportionally large error counts in some stations. In some stations the percentage of error readings in the data was as high as 5%.

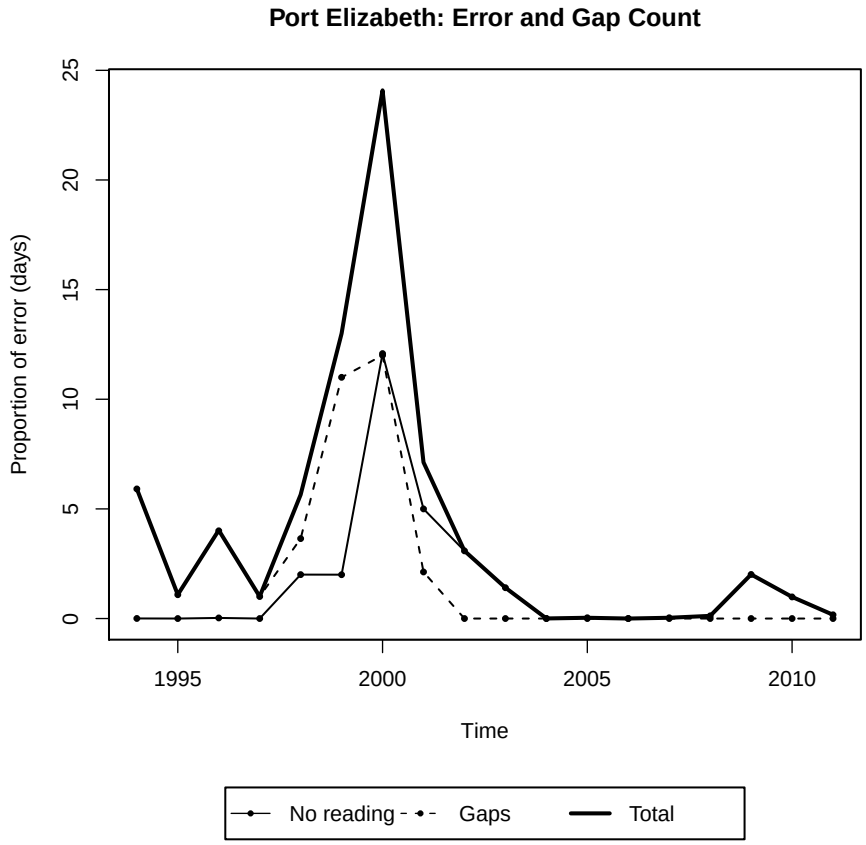


Figure 4.10: Number of errors in AWS record in terms of days per year for Port Elizabeth.

No other quality procedures were applied to the AWS data as it was assumed that the data was used for generating daily and annual totals.

4.2.3 Selection of stations and combining of data

Stations were selected on the basis of the above editing and quality control procedures and original station criteria and the two different datasets were combined for further analysis. There are particular issues associated with the combining of data recorded from different gauge types

Stellenbosch University http://scholar.sun.ac.za

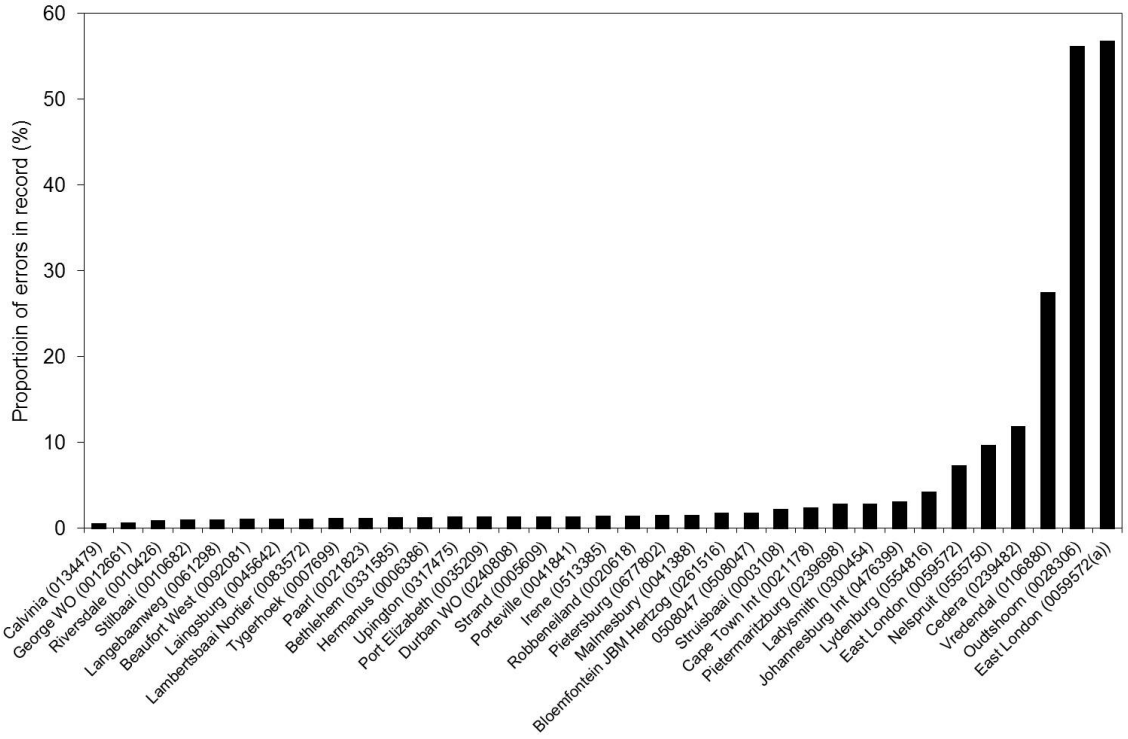


Figure 4.11: Summary of the total proportion of errors in AWS staions.

and relocated sites. Changnon and Kunkel (2005) highlights these issues by showing that large shifts in short duration rainfall occurred when the rain gauge was changed from a tipping bucket mechanism to a weight bucket mechanism. This further highlights the caution taken when interpreting the data, as the combined data could skew the data in a certain direction because of gauge changes.

After the station selection criteria and quality control procedures had been applied, the final stations were limited to only seven, as a result of the poor digitised autographic data quality. The seven stations are listed in Table 4.1 and illustrated spatially in Figure 4.12. The composition of the seven stations is discussed in greater detail in Section B.2.

Table 4.1: List of chosen stations used for analysis

Station name	Autographic		AWS (5 minute)		Total record length (years)
	Station number	Period	Station number	Period	
Cape Town	0021178 0021179	1956-1982 1982-1992	0021178	1994-2010	52
Langegbaanweg	0061298	1973-1992	00612988	1994-2010	35
George	0028690	1978-1992	00126617	1994-2010	30
East London	0559572	1955-1992	0059572	1994-2010	53
Port Elizabeth	0035179	1951-1992	0035209	1994-2010	57
Irene	0513385	1975-1993	0513385	1994-2010	34
Polokwane	0677802	1954-1992	0667802	1994-2010	54

Three of the seven selected stations are situated in the Western Cape, while the other four stations were selected to assess whether the trends found in the data for the Western Cape are

Stellenbosch University http://scholar.sun.ac.za

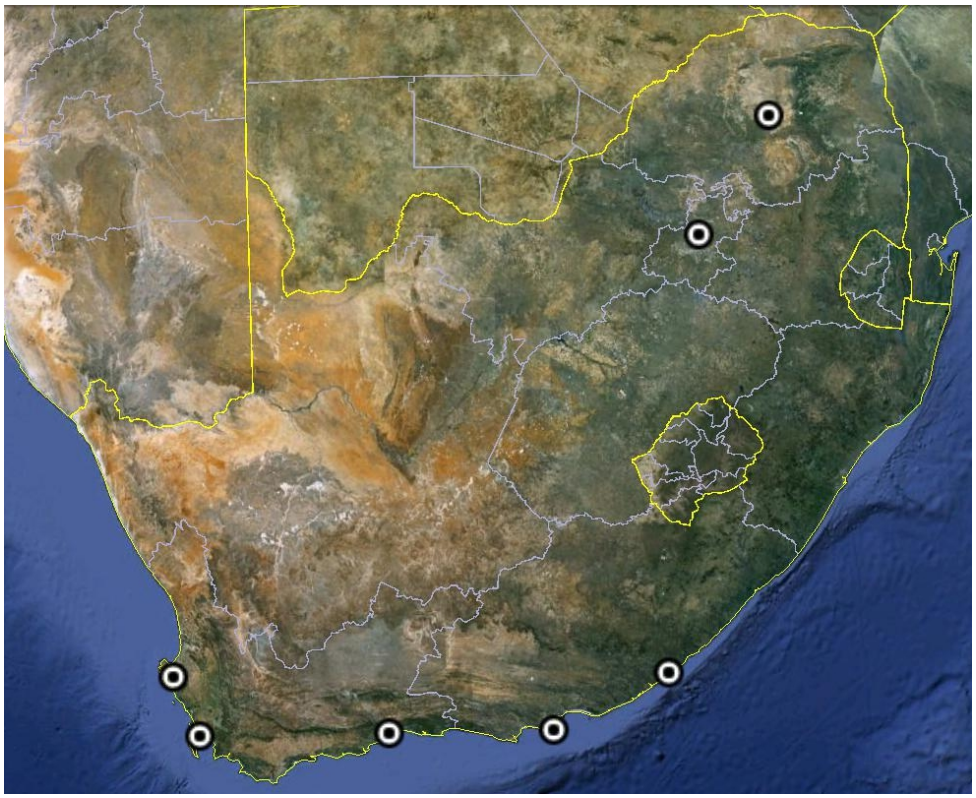


Figure 4.12: Location of chosen stations used for analysis.

unique or widespread around South Africa. As can be observed in Table 4.1, there is a gap in the data of all the stations for the period 1992 to 1994 as the SAWS changed rainfall gauges during that time, resulting in a two year gap.

It was established right at the end of the study that daily rainfall was measured independently from AWS data for a number of stations, hence SG data was available for further quality control. A quality control procedure similar to that for the autographic/SG data was applied to the AWS data for the seven selected stations. Results indicated that differences between the AWS and corresponding SG data also exist, with the AWS underestimating certain events. De Jager (2012) indicated that strong winds influencing the tipping bucket mechanism would be the most likely reason for this difference. The results of the AWS/SG procedure indicated that there was no reason to exclude any of the seven stations. The results are presented in Section B.1.

4.2.4 Summary of data processing

The editing and application of quality control procedures indicated that many digitised autographic stations were problematic, both regarding the number of errors and the comparison with SG data. There was a lower occurrence of error readings in the AWS data than the digitised autographic data. Taking these factors into consideration, seven stations were selected for further analysis.

Stellenbosch University http://scholar.sun.ac.za

4.3 Data analysis

Data analysis was divided into an analysis of the magnitudes and frequency of the SDR for the seven stations. Chapter D presents a worked example for the magnitude and frequency analysis.

4.3.1 Magnitude analysis

Storm duration data generated from the combined autographic and AWS data was analysed using parametric and non-parametric non-stationary extreme value theory (EVT) for the generelised extreme value distribution (GEV) and the generelised Pareto distribution (GPD). For the GPD, the three threshold types, discussed in Section 3.3.1.1, were used on the special peak dataset discussed in Section 3.2.4.

4.3.1.1 Parametric non-stationary

As discussed in Section 3.3.1.2, each parameter was fitted with a linear time model and the significance was measured with a t-test on the slope of the line. If any parameter was significant, it was compared to a stationary model by applying the likelihood test to assess if the non-stationary model is justified. If a non-stationary model was justified, it was considered a significant fit. Each storm duration for each station underwent this test, and the total number of significant non-stationary models was recorded per station. See Section D.1.1 for a worked example of the PNS analysis for both the GEV and GPD.

Of all the stations and durations for both the GEV and GPD analysis, only the GEV produced significant non-stationarity, with the GPD not producing any significant results for all three thresholds. Table 4.2 shows the total number of significant stations for the GEV distribution. As can be seen, George and Port Elizabeth produced the greatest number of significant storm durations, 9 and 5 respectively, while other stations' significant durations were scattered over the range of durations instead of producing a specific region of durations that is significant. See Table C.1 for a more detailed summary of the results for both the GEV and GPD distribution.

Table 4.2: Total number of significant storm durations for PNS analysis for the GEV distribution.

Station	Total number of significant durations
Cape Town	3
East London	1
George	9
Irene	1
Langebaanweg	0
Polokwane	2
Port Elizabeth	5

Stellenbosch University http://scholar.sun.ac.za

The results of George and Port Elizabeth were investigated further. George's significant non-stationary results were for all storm durations longer than 45 minutes, while those for Port Elizabeth were for storm durations longer than 4 hours.

All non-stationary results for George produced increasing return level estimates, with six of the nine durations producing positive non-stationary slopes for the shape parameter, ξ, indicating that the return levels become more extreme over time. Figure 4.13 illustrates a typical output generated from the analysis. Although the parameters are estimated linearly, the effect of the shape parameter on the return levels has a non-linear effect over time.

In contrast to George's results, all Port Elizabeth's non-stationary results produced decreasing return level estimates, with only the location parameter, μ, indicating non-stationarity. Figure 4.14 illustrates a typical output from Port Elizabeth.

4.3.1.2 Non-parametric non-stationary

The NPNS results were calculated as follows:

For a specific distribution:

1. Using a single storm duration, produce a series of return level values for a specific return period over a range of window periods. A total of nine combinations are generated for the range of window (15, 20, 25 years) and return (T = 5, 10, 20) periods.

2. Fit a linear regression line to the series of return levels and obtain the significance of the slope of the fitted line with a t-test for 95% level of significance.

3. Indicate the point in the series of return levels where the change in record type occurs.

4. For all nine combinations of return and window periods, count the number of significant slopes and the corresponding sign of the slope.

5. Repeat the procedure for all storm durations.

Figure 4.15 illustrates the procedure of steps 1-3 for the GEV at Cape Town for a storm duration of 30 minutes. Each graph illustrates the range of return levels produced for one of the nine possible combinations of return and window periods. The return levels are plotted as a series of points. The grey area around the return levels indicates the 95% confidence interval for the return levels. On the right hand side of each graph, details concerning the linear regression are given. The coefficient of determination, R^2 is given as "Rsq", while an indication of whether the slope of the regression line is significant is given as either "slope sign." or "slope not sign.". In this case, seven significant slopes were recorded. The vertical dashed line indicates the point in the return levels where the change of record type occurs.

Stellenbosch University http://scholar.sun.ac.za

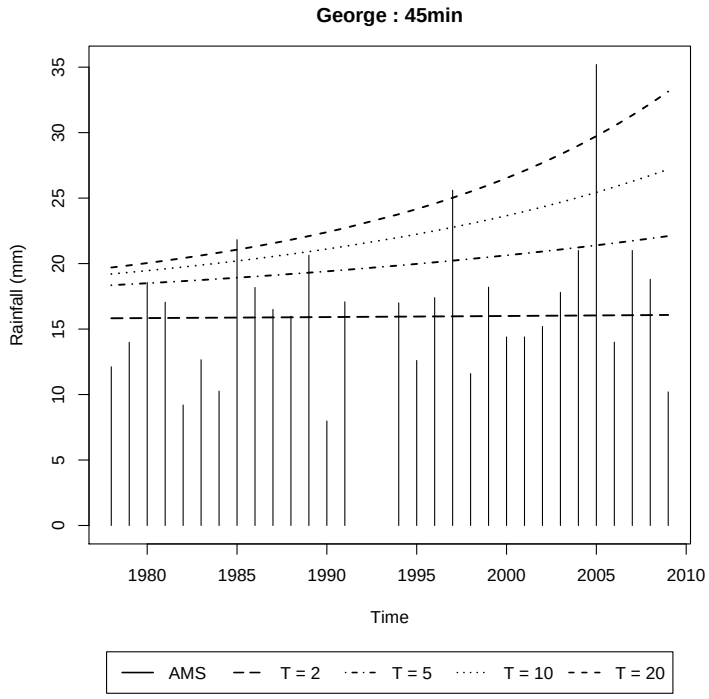


Figure 4.13: PNS output for George for a storm duration of 45 minutes. Solid lines represent AMS values, while the dashed lines represent the estimated return levels from the non-stationary GEV distribution.

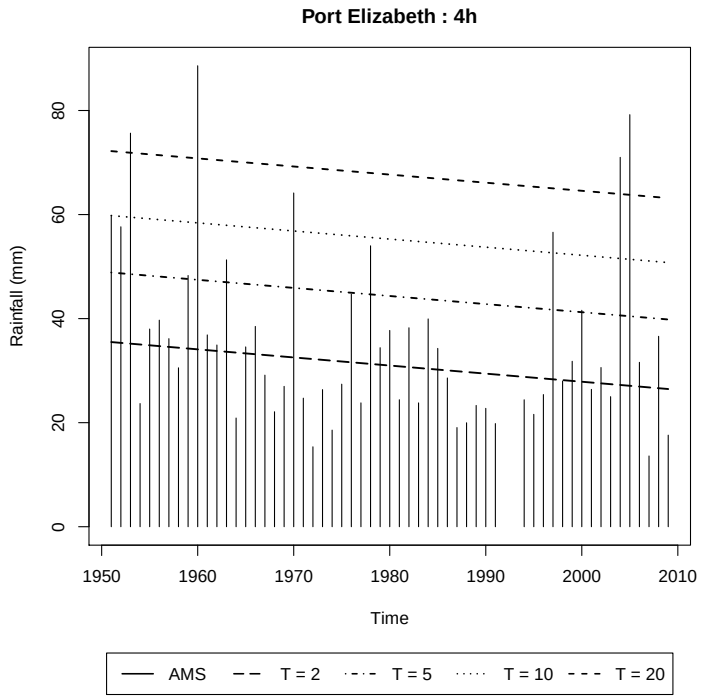


Figure 4.14: PNS output for Port Elizabeth for a storm duration of 4 hours. Solid lines represent AMS values, while the dashed lines represent the estimated return levels from the non-stationary GEV distribution.

Stellenbosch University http://scholar.sun.ac.za

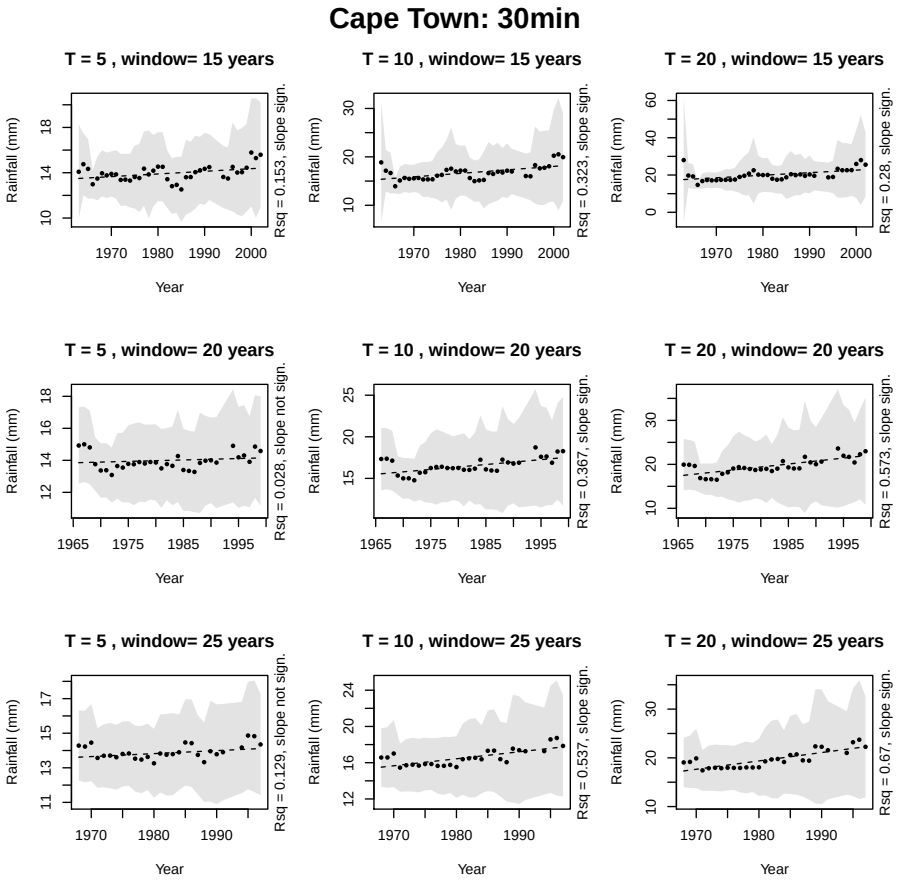


Figure 4.15: NPNS results for a GEV at a storm duration of 30 minutes for Cape Town.

Stellenbosch University http://scholar.sun.ac.za

Table 4.3 is a summary of the total number of significant slopes for all durations for the GEV at Cape Town. The number of positive and negative slopes are indicated by "pos" and "neg", while the total number of significant slopes are indicated by "total". The maximum possible total is nine, as there are nine combinations of return and window periods per storm duration. Figure 4.16 illustrates the results in Table 4.3 as a bar plot. The majority of significant slopes are positive, especially between 10 and 45 minutes, indicating the possibility of increasing magnitudes over time. The full range of the total number of significant slopes for both the GEV and GPD are found in Tables C.2 and C.3. See Section D.1.2 for a worked example of the NPNS analysis for both the GEV and GPD.

Table 4.3: Significant NPNS slopes for GEV at Cape Town

	total	pos	neg
5min	7	4	3
10min	8	6	2
15min	6	6	0
30min	7	7	0
45min	6	6	0
60min	3	3	0
90min	5	5	0
2h	3	3	0
4h	4	4	0
8h	5	4	1
12h	5	3	2
18h	6	5	1
24h	6	3	3

Results for the GEV indicated that George and Polokwane produced strong positive significant results, Port Elizabeth produced negative significant results for durations greater than 30 minutes, while the other stations produced varying signs for the slopes. Cape Town generally produced positive slopes, especially for durations between 15 and 45 minutes, East London produced positive slopes, especially for durations between 30 minutes and 8 hours. Langebaanweg produced negative slopes for durations between 30 minutes and 8 hours and very short and very long durations produced positive slopes. The results for Irene were varied.

Results of the GPD were more complicated as three different threshold values were involved but were generally in agreement with those of the GEV, although more amplified. Many stations produced up to nine significant slopes for a number of durations. The sensitivity of the GPD to the threshold was also portrayed, where the three threshold selection methods produced large differences in some cases.

At some stations there were indications of a change in slope when the change in record type, from autographic to AWS at about 1994, was incorporated into the moving window period. Figure 4.17 presents a typical example of this, where all the window and return periods indicate

Stellenbosch University http://scholar.sun.ac.za

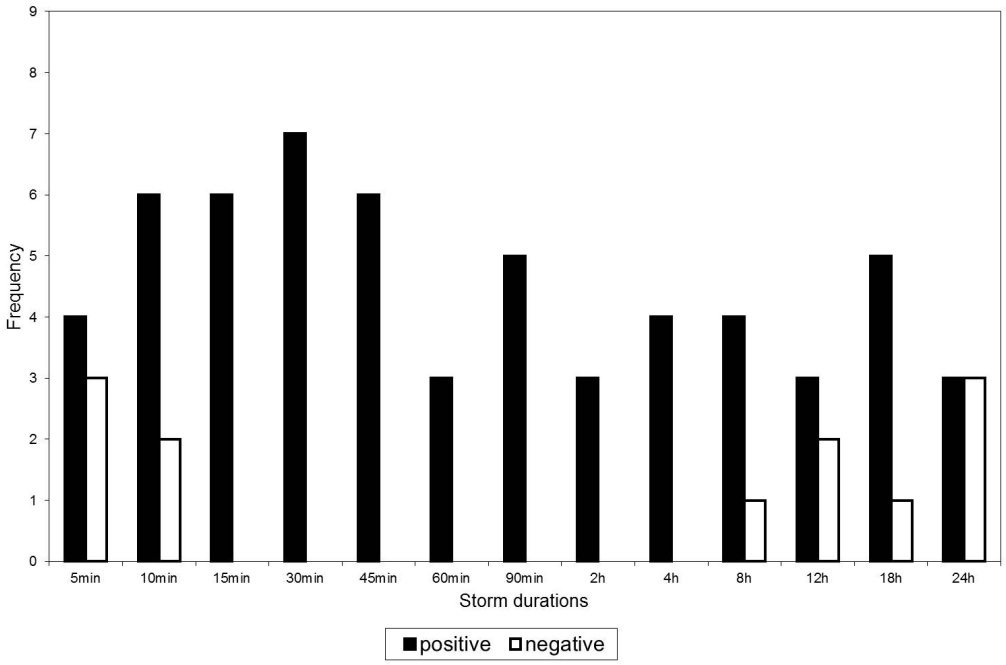


Figure 4.16: Bar plot of significant NPNS slopes for GEV at Cape Town

a decrease in return levels until the vertical dashed line indicating the change in record, where the return levels start to increase. Though the regression indicates that the slope was significant, the actual return levels reflect something different. The resulting output of each storm duration for each distribution was visually investigated to assess if any slope changes occurred with the change in record. Because of the sheer quantity of figures generated from the NPNS procedure[3], the figures used for visual inspection were not included in this thesis, but are available on the accompanying disc.

For both the GEV and GPD distributions, East London, Polokwane and Port Elizabeth showed changes of slope sign at the change of record, while a change at Irene was noted for the GPD distribution for all three thresholds. Since George used a shorter record length than the other stations, this influence could not be assessed with the graphs. Cape Town and most of the durations at Langebaanweg did not produce evidence of change of slopes at the change of the record type.

In summary, the results were difficult to interpret and with the exception of the stations that indicated clear record type changes, no clear overall trend was found except at Cape Town, where return levels produced some evidence of increasing over time and at Langebaanweg, where very short and very long durations indicated increasing trends, while the middle durations produced some evidence of decreasing return levels. Further investigation of stations like Port Elizabeth and East London, both producing decreasing return levels before the record change, is warranted. In addition, the use of linear regression was not always desirable, as the changes

[3]For all 7 stations, 13 figures were generated for the GEV and 3 cases of the GPD. $7 \times 13 \times 4 = 364$ figures in total.

Stellenbosch University http://scholar.sun.ac.za

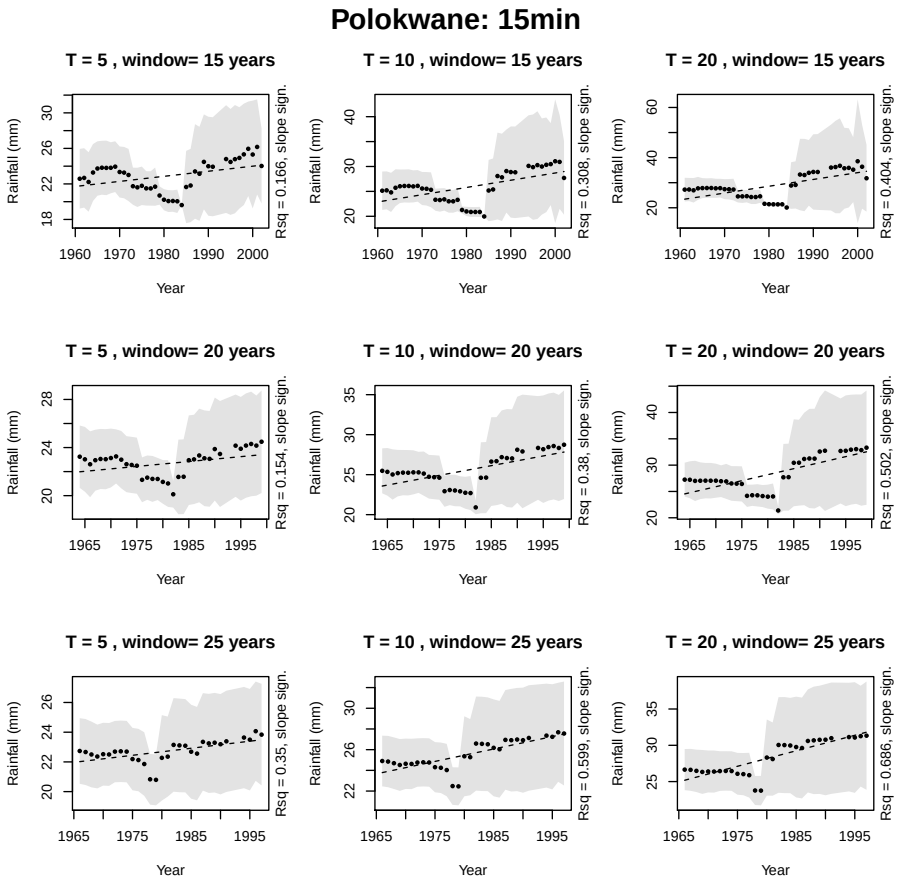


Figure 4.17: NPNS results for a GEV at a storm duration of 15 minutes for Polokwane.

in the slopes were not always linear and sometimes changed direction.

4.3.1.3 Summary of magnitude analysis

The results of the PNS analysis were not significant, as only two stations from the GEV indicated significant linear non-stationarity. The NPNS highlighted the problems associated with combining the autographic and AWS data, as changes in the slopes of the return levels were observed for some stations that coincided with the change in record. Overall, the lack of constant results throughout for the non-stationary analyses indicates that there is a lack of strong evidence for non-stationarity, especially for increasing intensities in rainfall and for stations in the Western Cape.

4.3.2 Frequency analysis

The frequency analysis recorded the number of exceedances of the thresholds 0.5, 1 and 2mm per year for the complete 5-minute dataset for all stations. Moving averages of 5, 10 and 20 years were fitted to the exceedances and plotted graphically, as illustrated in Figure 4.18 for Cape Town. Figures C.1a to C.4 presents the results of the rest of the stations. The light grey

Stellenbosch University http://scholar.sun.ac.za

line represents the number of exceedances of the threshold per year, while the darker dashed lines represents the 5, 10 and 20 year moving averages. The dotted horizontal line represent the mean number of exceedances. Also note the vertical dashed line in light grey at approximately 1994, representing the change in record type.

As the threshold increases, the number of exceedances decreases. In Figure 4.18, the number of exceedances decrease from an average of about 150 per year for a 0.5mm threshold, to an average of 10 for a 2mm threshold.

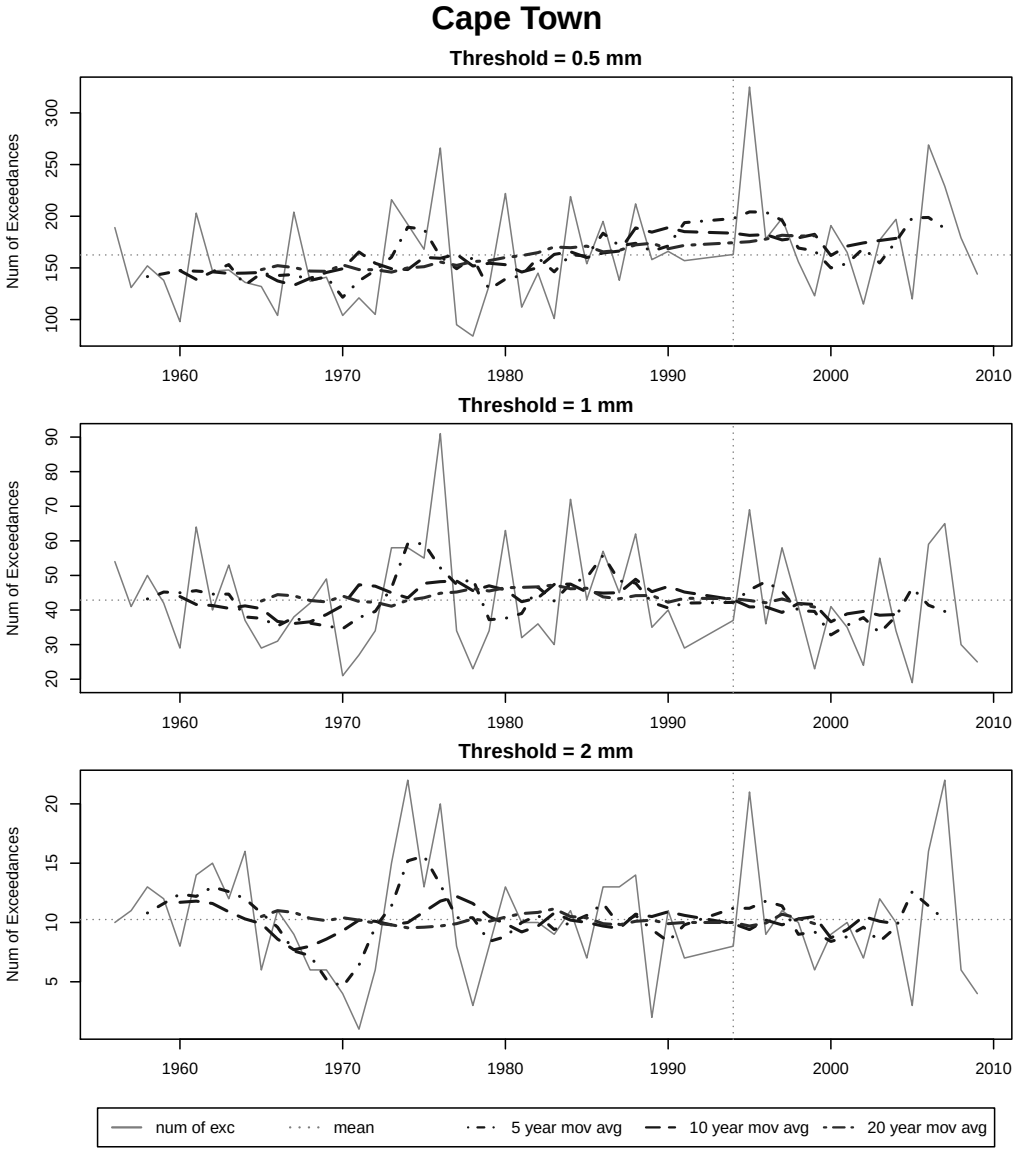


Figure 4.18: Frequency of exceedance plot for Cape Town for thresholds of 0.5, 1 and 2mm along with central moving averages for 5, 10 and 20 years and the mean exceedance plotted as a dashed line.

For a low threshold of 0.5mm, exceedances increased from the 1990s at most stations, while thresholds of 1 and 2mm generally produced similar increases for the stations George, Polokwane and Port Elizabeth, or produced little change or smaller increasing trends at the other stations. The most likely reason for the increasing trends, especially for a low threshold, is the

Stellenbosch University http://scholar.sun.ac.za

change of record type during the 1990s, and this is possibly because the autographic data is unable to measure very low rainfall values with the same degree of accuracy as the AWS data. This is supported by some stations producing little or slight changes for higher threshold values. Consequently, more weight is given to the higher threshold values.

The importance of the threshold values is illustrated in Figure 4.19, illustrating the frequency of exceedance of no threshold to a threshold of 0.4mm. If no threshold is applied, a clear drop in the number of exceedances is noticed from the 1990s, which is similar to the results of Van Wageningen (2006:74). However, with a slight adjustment of 0.1mm, the number of exceedances increased strongly since the 1990s. When the threshold is raised to 0.2mm, a drop is again noticed from the 1990s. This pattern repeats itself for each slight adjustment of the threshold, although becoming weaker with each increase of the threshold. The sensitivity of the number of exceedances to the slight increments of the threshold is explained by the nature of the autographic and AWS data. As mentioned in Section 3.3.2, the autographic data consisted of interpolated values that were often very low (< 0.2mm) and can be counted as events for the frequency analysis. In addition, the AWS data only measures data in 0.2mm increments, thereby rendering zero to low thresholds' exceedances incomparable between the two datasets.

The data used in Van Wageningen (2006) for Molteno station in Cape Town, was obtained for comparison[4], and a number of very low rainfall values (< 0.2mm), were observed in this record for data before 1994. This seems to indicate that interpolation techniques similar to those used in this project were applied for the data supplied to Van Wageningen (2006). The similarity between the results of Van Wageningen (2006) and the zero threshold results in Figure 4.19 indicate that the move from autographic to tipping bucket gauges is most likely the cause of apparent changes in frequency and not any eccentricity in natural cycles (Van Wageningen 2006:77, Van Wageningen & Du Plessis 2007:573).

4.3.2.1 Summary of frequency analysis

The results of the frequency analysis indicated that the change in record was the most likely explanation for increases in exceedances of low threshold values. A few stations produced increases of exceedances for higher thresholds, but these usually occurred at the point of a change in record. The rest of the stations produced slight or no change in the frequency of exceedances. This indicates that there is little evidence of changing frequency of occurrence in rainfall for the stations analysed.

[4]The data used in Van Wageningen (2006) was obtained from the City of Cape Town. The SDR data at Molteno changed from an autographic to a tipping bucket system in 1992 (Van Wageningen, 2006:68).

Stellenbosch University http://scholar.sun.ac.za

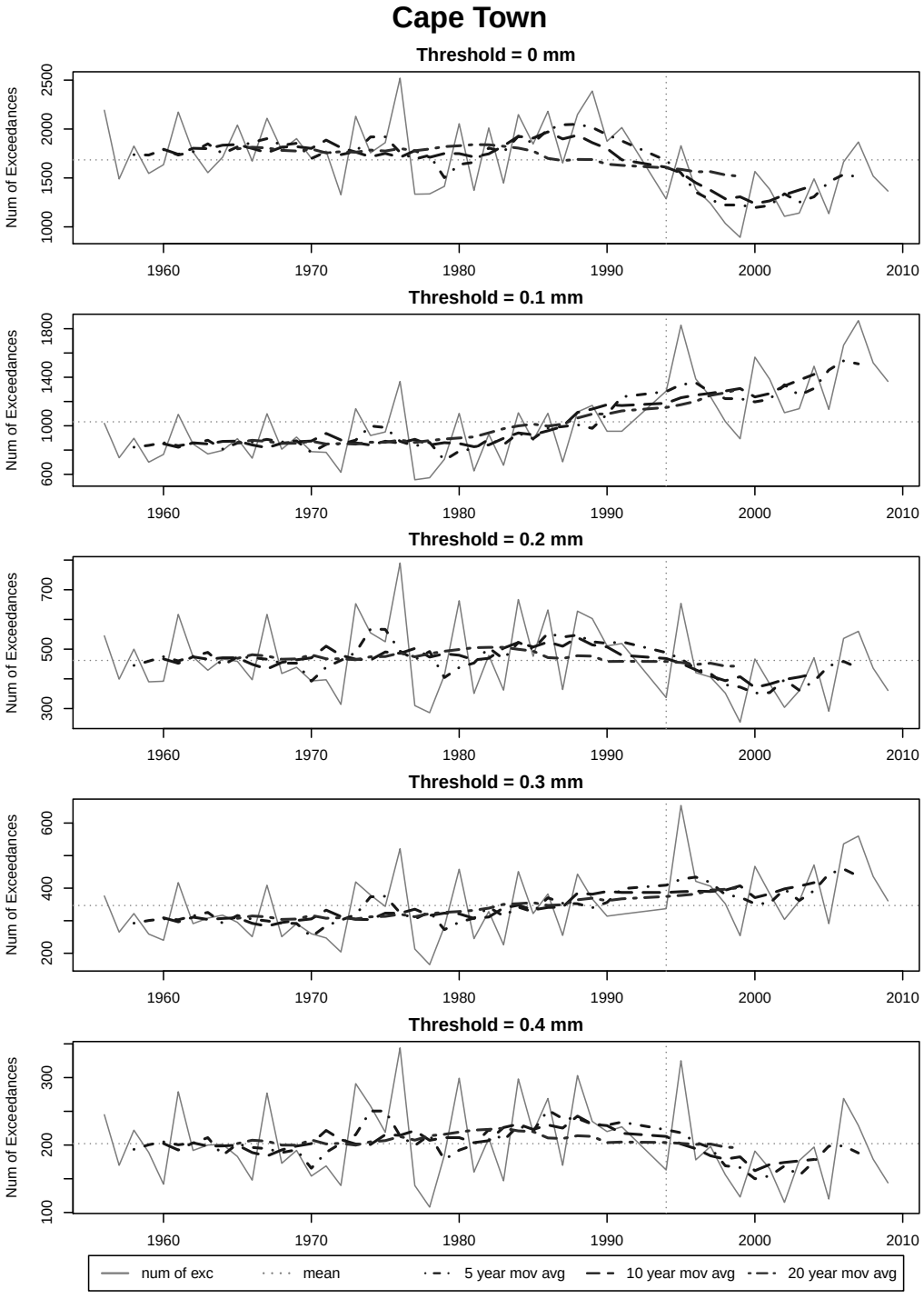


Figure 4.19: Frequency of exceedance plot for Cape Town for thresholds from 0 to 0.4mm.

Stellenbosch University http://scholar.sun.ac.za

4.3.3 Summary of data analysis

Both the magnitude and frequency analysis indicated that the data quality and record change had a strong influence on many of the results. The only stations producing significant results for both magnitude and frequency analyses were George and Port Elizabeth, but there are strong indications that the record change had a influence on the results. Of the remaining stations, those from the Western Cape, Cape Town and Langebaanweg, produced no significant PNS results, while showing some indication of changing return levels over time with the NPNS. The frequency analysis for these stations did not indicate any visible trends in frequency of events either.

This leads to the conclusion that with the available SDR records and methods used, there is not sufficient and comprehensive evidence supporting non-stationary effects on SDR, or evidence of changes in the magnitude, frequency and/or intensities of rainfall.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Chapter 5

Conclusions and recommendations

5.1 Summary of findings

The aim of the researcher was to assess if historical rainfall data provides any evidence of change according to climate change projections for the Western Cape. Although climate change projections do not include a single projection for the Western Cape, there is some consensus that rainfall intensity could increase. Changes in rainfall intensity are important, as they can have serious implications for stormwater design. To this end, short duration rainfall (SDR) was analysed to assess if any changes in both the magnitude and frequency of occurrence of rainfall could be observed. This was done by combining digitised autographic data and modern automatic weather station (AWS) rainfall data to extend the effective record length. A number of difficulties were encountered with the data that necessitated applying editing and quality control procedures. The autographic data performed poorly compared to standard gauge daily rainfall totals of the same period. After the application of editing and quality control procedures, seven stations were selected for detailed analysis.

A parametric and non-parametric non-stationary approach was applied to assess the possibility of trends in the magnitude of the SDR for both the generalised extreme value and generalised Pareto distributions. Only two stations showed significant non-stationarity for the parametric approach, and only for the GEV. When the non-parametric approach was used, some stations showing non-stationary behaviour, but this was most likely as a result of the combination of the autographic and AWS data.

When using the frequency of exceedance of the threshold test, based on a method used by Van Wageningen (2006:74) to analyse changes in the frequency of occurence of rainfall events, either no observable trends were observed, or trends that are consistent with a change of autographic to AWS data were observed.

Stellenbosch University http://scholar.sun.ac.za

5.2 Conclusions

The primary finding as a result of this research is:

- With the use of available data, no clear or significant evidence supporting increasing intensities or any other change in rainfall was found regarding both the magnitude and frequency of occurrence.

Regarding the data editing process, it was found that:

- The digitised autographic data generally compared poorly to the standard gauge daily totals, most likely because of the digitising process containing numerous errors often missing large storm events. This is generally in agreement with the findings of Smithers and Schulze (2000).

5.3 Recommendations

The following recommendations are made. based on the conclusions:

- Midgley et al. (2005:32) and Schulze et al. (2011:15) mention that climate change is more likely to cause seasonal shifts than changes in annual behaviour. As this project did not investigate seasonal effects in detail, possible seasonal trends for both magnitudes and frequency should be studied in the future.
- The limited amount of reliable digitised autographic data available in South Africa makes current studies extremely difficult. Similar studies should be conducted in the future for AWS only, as current records are considered to be too short to detect long term trends.
- The non-stationary extreme value theory used in this thesis can be implemented in future research.
- The application of the parametric non-stationary approach was limited to linear models. More complex models like quadratic models and covariates, like ENSO and sunspot cycles, should be further investigated.
- The difficulties encountered when tyring to obtain SDR for this project highlighted the need to have a publicly accessible database of available rainfall stations and their corresponding suppliers of the data to enable easier access to data.

Stellenbosch University http://scholar.sun.ac.za

Bibliography

Alexander, L., Zhang, X., Peterson, T. C., Ceaser, J., Gleason, B., Tank, A. L., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Ambeje, P., Kumar, K. R., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D., Burn, J., Aguilar, E., Brunett, M., Taylor, M., New, M., Zhai, P., Rusticucci, M. & Vazquez-Aguirre, J. 2006. Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, 111(5): 1–22.

Alexander, W. J. R. 2001. *Flood Risk Reduction Measures*. edition no. 2. Pretoria: University of Pretoria.

Alexander, W. J. R., Bailey, F., Bredenkamp, D. B., Van der Merwe, A. & Willemse, N. 2007. Linkages between solar activity, climate predictability and water resource development. *Journal of South African Institution of Civil Engineering*, 49(2): 32–44.

Aly, N. E. 2010. Spectral analysis of solar variability and their possible role on global warming. *International Journal of Physical Sciences*, 5(7): 1040–1049.

Beguería, S., Angulo-Matínez, M., Vicente-Serrano, S. M., López-Moreno, J. I. & El-Kenawy, A. 2011. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in Northeast Spain from 1930 to 2006. *International Journal of Climatology*, 31(14): 2102–2114.

Ben-Zvi, A. 2009. Rainfall intensity-duration-frequency relationships derived from large partial duration series. *Journal of Hydrology*, 367(1-2): 104–114.

Brath, A., Castellarin, A. & Montanari, A. 2001. At-site and regional assessment of the possible presence of non-stationarity in extreme rainfall in northern Italy. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 26(9): 705–710.

Changnon, S. T. & Kunkel, K. E. 2005. Changes in instruments and sites affecting historical weather records: a case study. *Journal of Atmospheric and Oceanic Technology*, 23(6): 825–828.

Stellenbosch University http://scholar.sun.ac.za

Claps, P. & Laio, F. 2003. Can continuous streamflow data support flood frequency analysis? An alternative to the partial duration series approach. *Water Resource Research*, 39(8): 1216–1227.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics: Springer.

Coles, S., Pericchi, L. R. & Sisson, S. 2003. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1-4): 35–50.

Costa, A. C. & Soares, A. 2009. Trends in extreme precipitation indices derived from a daily rainfall database for the south of Portugal. *Journal of Climatology*, 29(13): 1956–1975.

Cunane, C. 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18(3-4): 257–271.

De Jager, E. 2012. Correspondence [Online], 18 June.

Du Plessis, J. A. 1992. *Rainfall intensity-duration-frequency-curves for the Cape Peninsula*. Cape Town: City of Cape Town.

Fauchereau, N., Trzaska, S. & Richard, Y. 2003. Rainfall variability and changes in Southern Africa during the 20th century in the global warming context. *Natural Hazards*, 29(2): 139–154.

Ghahraman, B. & Khalili, D. 2004. A re-visit to partial duration series of short duration rainfalls. *Iranian Journal of Science & Technology*, 28(B5): 547–558.

Gilleland, E. & Katz, R. W. 2006. Analyzing seasonal to interannual extreme weather and climate variability with the Extremes toolkit. in *86th AMS Annual Meeting*. 86th AMS Annual Meeting.

Groisman, P. Y., Karl, T. R., Easterling, D., Knight, R. W., Jamason, P. F., Hennessy, K. J., Suppiah, R., Page, C., Wibig, J., Fortuniak, K., Razuvaev, V. N., Douglas, A., Førland, E. & Zhai, P. 1999. Changes in the probability of heavy precipitation: Important indicators of climate change. *Climatic Change*, 42: 243–283.

Groisman, P. Y., Knight, R. W., Easterling, D., Karl, T. R., Hegerl, G. C. & Razuvaev, V. N. 2004. Trends in intense precipitation in the climate record. *Journal of Climate*, 18(9): 1326–1350.

Gumbel, E. J. 1960. *Statistics of Extremes*. edition no. 2. New York: Columbia University Press.

Hennessy, K. J., Gregory, J. M. & Mitchell, J. F. B. 1997. Changes in daily precipitation under enhanced greenhouse conditions. *Climate Dynamics*, 13(9): 667–680.

Stellenbosch University http://scholar.sun.ac.za

Hewitson, B. C. & Crane, R. G. 2006. Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *International Journal of Climatology*, 26(10): 1315–1337.

Hiremath, K. M. & Mandi, P. I. 2004. Influence of the solar activity on the Indian Monsoon rainfall. *New Astronomy*, 9(8): 651–662.

Hoffman, M. T., Cramer, M. D., Gillson, L. & Wallace, M. 2011. Pan evaporation and wind run decline in the Cape Floristic Region of South Africa (1974-2005): Implications for vegetation responses to climate change. *Climatic Change*, 109(3-4): 437–452.

Holloway, A., Fortune, G. & Chasi, V. 2010. *RADAR Western Cape 2010*. Risk and Development Annual Review: Crede Communications.

Hosking, J. R. M., Wallis, J. R. & Wood, E. F. 1985. Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics*, 27(3): 251–261.

Hudson, D. & Jones, R. G. 2002. *Simulations of present-day and future climate over Southern Africa using HadAM3H*. United Kingdom: Met Office.

Hugo, P. 2010. An analysis of the relationship between the Annual Maximum Series and Partial Duration Series methods applied to flow data in South Africa.

Jakob, D., Karoly, D. J. & Seed, A. 2011. Non-stationarity in daily and sub-daily intense rainfall - part 1: Sydney, Australia. *Natural Hazards and Earth System Science*, 11(8): 2263–2271.

Katz, R. W., Brush, G. S. & Parlange, M. B. 2005. Statistics of extremes: modeling ecological disturbances. *Ecological Society of America*, 86(5): 1124–1134.

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

Krishnamurthy, C. K. B., Lall, U. & Kwon, H. 2009. Changing frequency and intensity of rainfall extremes over India from 1951 to 2003. *Journal of Climate*, 22(18): 4737–4746.

Kruger, A. C. 1999. The influence of the decadal-scale variability of summer rainfall on the impact of El Niño and La Niña events in South Africa. *Journal of Climatology*, 19(1): 59–68.

Kruger, A. C. 2004. *Climate Regions (WS45)*. Climate of South Africa. Pretoria: South African Weather Service.

Kruger, A. C. 2006. Observed trends in daily precipitation indices in South Africa: 1910-2004. *Journal of Climatology*, 26(15): 2275–2285.

Kruger, A. C. 2007. *Precipitation (WS47)*. Climate of South Africa. Pretoria: South African Weather Services.

Stellenbosch University http://scholar.sun.ac.za

Legras, B., Mestre, O., Bard, E. & Yiou, P. 2010. A critical look at solar-climate relationships from long temperature series. *Climate of the Past*, 6(6): 745–758.

Li, Y., Cai, W. & Campbell, E. P. 2005. Statistical modeling of extreme rainfall in Southwest Western Australia. *Journal of Climate*, 18(6): 852–863.

Lumsden, T. G., Schulze, R. E. & Hewitson, B. C. 2009. Evaluation of potential changes in hydrologically relevant statistics of rainfall in Southern Africa under conditions of climate change. *Water SA*, 35(5): 649–656.

Mason, S., Waylen, P. R., Mimmack, G. M., Rajaratnam, B. & Harrison, J. M. 1999. Changes in extreme rainfall events in South Africa. *Climatic Change*, 41(2): 249–257.

Midgley, G. F., Chapman, R. A., Hewitson, B. C., Johnston, P., De Wit, M., Ziervogel, G., Mukheibir, P., Van Niekerk, L., Tadross, M. A., Van Wilgen, B. W., Kgope, B., Morant, P. D., Theron, A., Scholes, R. J. & Forsyth, G. G. 2005. *A status quo, vulnerability and adaptation assessment of the physical and socio-economic effects of climate change in the Western Cape*. Stellenbosch: Council for Scientific and Industrial Research.

Montgomery, D. C. & Runger, G. C. 2007. *Applied statistics and probability for engineers*. edition no. 4. New Jersey: John Wiley & Sons, Inc.

New, M., Hewitson, B. C., Stephenson, D. B., Tsiga, A., Kruger, A., Manhique, A., Gomez, B., Coelho, C. A. S., Masisi, D. N., Kululanga, E., Mbambalala, E., Adesina, F., Saleh, H., Kanyanga, J., Adosi, J., Bulane, L., Fortunata, L., Mdoka, M. & Lajoie, R. 2006. Evidence of trends in daily climate extremes over Southern and West Africa. *Journal of Geophysical Research*, 111(14): D14102.

Nicholson, S. E. 2000. The nature of rainfall variability over Africa on time scales of decades to millenia. *Global and Planetary Change*, 26: 137–158.

O'Hare, G., Sweeny, J. & Wilby, R. 2005. *Weather, climate and climate change : human perspectives*. edition no. 1. Essex: Pearson Education Limited.

Park, J., Kang, H., Lee, Y. S. & Kim, M. 2011. Changes in the extreme daily rainfall in South Korea. *International Journal of Climatology*, 31(15): 2290–2299.

Philandras, C., Nastos, P., Paliatsos, A. & Repapis, C. 2010. Study of the rain intensity in Athens and Thessaloniki, Greece. *Advances in Geosciences*, 23: 37–45.

Rouault, M., Fauchereau, N., Pohl, B., Penven, P., Richard, Y., Reason, C. J. R., Pegram, G. G. S., Phillippon, N., Siedler, G. & Murgia, A. 2010. *Multidisciplinary analysis of hydroclimatic variability at the catchment scale*. Pretoria: Water Research Commission.

Stellenbosch University http://scholar.sun.ac.za

SANRAL 2007. *Drainange Manual*. edition no. 5. Pretoria: South African National Roads Angency.

SAWB 1968. *WB 36: Part II*. Climate of South Africa. Pretoria: South African Weather Bureau.

Schulze, R. E. 2003. The present is bad enough... The future is not what it used to be... and where to now? Perspectives in climate variability and climate change over Southern Africa and adapting to these. in *The Thukela Dialogue: Managing Water Related Issues on Climate Variability and Climate Change in South Africa*. University of KwaZulu Natal. Pietermaritzburg.

Schulze, R. E., Hewitson, B. C., Barichievy, K. R., Tadross, M. A., Kunz, R. P., Horan, M. J. C. & Lumsden, T. G. 2011. *Methodological approaches to assessing eco-hydrological responses to climate change in South Africa*. Pretoria: Water Research Commission.

Smithers, J. C. 1996. Short-duration rainfall frequency model selection in Southern Africa. *Water SA*, 22(3): 211–217.

Smithers, J. C. & Schulze, R. E. 2000. *Development and evaluation of techniques for estimating short duration design rainfall in South Africa*. Pretoria: Water Research Commission.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L., (eds.) 2007. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*: Cambridge University Press.

Stager, J. C., Ryves, D., Cumming, B. F., Meeker, L. D. & Beer, J. 2005. Solar variability and the levels of Lake Victoria, East Africa, during the last millenium. *Journal of Paleolimnology*, 33(2): 243–251.

Tavares, L. V. & Da Silva, J. E. 1983. Partial duraton series method revisited. *Journal of Hydrology*, 64(3): 1–14.

Tennant, W. J. & Hewitson, B. C. 2002. Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. *Journal of Climatology*, 22(9): 1033–1048.

The Concise Oxford Dictionary 1989. edition no. 7. Oxford: Oxford University Press.

Thomas, D. S. G., Twyman, C., Osbahr, H. & Hewitson, B. C. 2007. Adaptation to climate change and variability: farmer responses to intra-seasonal precipitation trends in South Africa. *Journal of Climatic Chnage*, 83(3): 301–322.

Trefry, C. M., Watkins, D. W. & Johnson, D. 2005. Regional rainfall frequency analysis for the state of Michigan. *Journal of Hydrlogic Engineering*, 10(6): 437–448.

Stellenbosch University http://scholar.sun.ac.za

Trenberth, K. E., Dai, A., Rasmussen, R. M. & Parsons, D. B. 2003. The changing character of precipitation. *Bulletin of the American Meteorological Society*, 84(9): 1205–1217.

Tsourti, Z. & Panaretos, J. 2001. *Extreme value index estimators and smoothing alternatives: review and simulation comparison*: Athens University of Economics and Business.

Van Bladeren, D., Zawada, P. K. & Mahlangu, D. 2007. *Statistical based regional flood frequency estimation for South Africa using systematic, historical and palaeoflood data*. Pretoria: Water Research Commission.

Van Wageningen, A. 2006. The impact of climate change on hydrological predictions, with specific reference to 24-hour rainfall intensities in the Western Cape. Master's thesis. Stellenbosch: Stellenbosch University.

Van Wageningen, A. & Du Plessis, J. A. 2007. Are rainfall intensities changing, could climate change be blamed and what could be the impact for hydrologists?. *Water SA*, 33(4): 571–574.

Wang, Q. J. 1991. The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, 129(1-4): 263–280.

White, C. J., Corney, S., Grose, M., Holz, G., Bennett, J. & Bindoff, N. L. 2010. Modelling extreme events in a changing climate using regional dynamically-downscaled climate projections. in *International Congress on Environmental Modelling and Software*. International Environmental Modelling and Software Society. Ottowa, Canada.

Stellenbosch University http://scholar.sun.ac.za

Appendix A

Data acquisition

Stellenbosch University http://scholar.sun.ac.za

Table A.1: List of original autographic stations selected for processing.

Station Name	Name / Location	Province	Start Date	End Date
0059572	EAST LONDON - WK	Eastern Cape	1940/01/01	1992/11/26
0035179	PORT ELIZABETH - WK	Eastern Cape	1938/01/01	1992/09/30
0261516	BLOEMFONTEIN JBM HER	Free State	1962/01/04	1992/11/20
0331585	BETHLEHEM - WO	Free State	1980/01/04	1992/11/30
0261307	BLOEMFONTEIN - WK	Free State	1938/01/05	1961/12/17
0476398	JAN SMUTS - WK	Gauteng	1953/10/10	1989/05/17
0476399	JAN SMUTS - WK 30L	Gauteng	1989/06/01	1992/11/25
0513385	IRENE -WK	Gauteng	1975/01/06	1993/03/01
0239577	PIETERMARITZBURG-PUR	KwaZulu Natal	1947/08/01	1960/01/01
0239666	PIETERMARITZBURG-MUN	KwaZulu Natal	1960/01/02	1969/03/31
0239756	PIETERMARITZBURG-PUR	KwaZulu Natal	1969/04/02	1987/11/01
0300423	LADYSMITH - MUN	KwaZulu Natal	1967/01/01	1979/12/29
0300454	LADYSMITH - WK	KwaZulu Natal	1949/01/01	1992/01/01
0300483	LADYSMITH - CON	KwaZulu Natal	1956/01/01	1965/12/29
0239482	CEDARA	KwaZulu Natal	1951/01/01	1996/05/26
0677866	PIETERSBURG - WK	Limpopo	1938/01/05	1951/09/03
0677802	PIETERSBURG - WK	Limpopo	1954/01/03	1992/11/25
0634631	PIETERSBURG-KUSCHKE	Limpopo	1985/08/31	1988/08/31
0596179	SKUKUZA	Mpumalanga	1959/11/01	1996/05/29
0555837	NELSPRUIT - AGR	Mpumalanga	1960/01/01	1973/04/27
0555866	NELSPRUIT-FRIEDENHEI	Mpumalanga	1973/05/02	1992/10/30
0508231	MAFIKENG	North West	1966/06/04	1969/12/31
0508261	MAFIKENG - TNK	North West	1970/01/01	1980/06/29
0508232	MAFIKENG - WK	North West	1958/06/01	1966/05/13
0134478	Calvinia-TNK	Northern Cape	1962/01/18	1987/01/22
0134479	Calvinia WO	Northern Cape	1986/08/03	1992/11/21
0317476	UPINGTON - WK	Northern Cape	1951/05/01	1968/10/23
0317474	UPINGTON - WK	Northern Cape	1968/11/11	1992/08/28
0317475	UPINGTON WO	Northern Cape	1992/06/24	1992/11/08
0258213	DRIEPLOTTE - AGR	Northern Cape	1961/01/01	1993/11/23
0092288	BEAUFORT WEST	Western Cape	1950/07/05	1985/09/12
0092141	BEAUFORT WEST - TNK	Western Cape	1964/01/01	1979/12/26
0092171	BEAUFORT WEST - WK	Western Cape	1985/10/27	1992/07/23
0092229	BEAUFORT WEST - WK	Western Cape	1940/01/16	1950/05/24
0021178	CAPE TOWN D F MALAN	Western Cape	1956/07/06	1992/11/25
0021179	CAPE TOWN D.F.MALAN	Western Cape	1982/02/04	1992/08/31
0028690	GEORGE - P.W. BOTHA	Western Cape	1978/01/04	1992/08/26
0028748	GEORGE (WK)	Western Cape	1949/04/01	1977/01/01
0061298	LANGEBAANWEG WO	Western Cape	1973/03/01	1992/11/24
0028337	OUDTSHOORN	Western Cape	1992/04/10	1996/06/01
0028364	OUDTSHOORN	Western Cape	1983/12/06	1988/11/30
0028305	OUDTSHOORN - MIL	Western Cape	1977/08/01	1979/07/30
0028334	OUTSHOORN	Western Cape	1989/10/01	1992/03/31
0028335	OUTSHOORN - TNK	Western Cape	1988/12/01	1989/09/23
0028336	OUTSHOORN - WK	Western Cape	1979/08/01	1983/06/28
0010425	RIVERSDALE	Western Cape	1985/07/01	1996/06/01
0010456	RIVERSDALE	Western Cape	1959/01/01	1985/05/22
0023708	ROBERTSON	Western Cape	1986/12/15	1996/05/31
0023710	ROBERTSON (AGR)	Western Cape	1962/01/18	1986/04/30
0106880	VREDENDAL	Western Cape	1962/01/26	1996/06/01

Stellenbosch University http://scholar.sun.ac.za

Table A.2: List of original AWS stations selected for processing.

Station Name	Name / Location	Province	Start Date
0059572B8	East London WO	Eastern Cape	2001
0035209B1	Port Elizabeth	Eastern Cape	1994
03315859	Bethleheom WO	Free State	1994
0261516B0	Bloemfontein WO	Free State	1994
0513385A2	Irene WO	Gauteng	1994
04763990	Johannesburg Int WO	Gauteng	1994
02394820	Cedra	KwaZulu Natal	2005
0240808A2	Durban WO	KwaZulu Natal	1994
03004543	Ladysmith	KwaZulu Natal	1994
02396985	Pietermaritzburg	KwaZulu Natal	1994
0677802BX	Pietersburg WO	Limpopo	1994
0554816A7	Lydenburg	Mpumalanga	1994
05557509	Nelspruit	Mpumalanga	1994
05080470	Mafikeng WO	North West	1996
0134479A3	Calvinia WO	Northern Cape	1995
0317475A8	Upington WO	Northern Cape	1994
00920815	Beaufort-Wes	Western Cape	1994
0021178A3	Cape Town - AWS	Western Cape	2001
00221178B8	Cape Town WO	Western Cape	1994
00126617	George WO	Western Cape	1994
0007699A0	Hermanus	Western Cape	1994
00456420	Laingsburg	Western Cape	1995
00835728	Lambertsbaai Nortier	Western Cape	1994
00612988	Langebaanweg AWS	Western Cape	1994
0041388	Malmesbury	Western Cape	1994
00207466	Molteno Reservoir	Western Cape	2001
00283065	Outshoorn	Western Cape	1999
0021823	Paarl	Western Cape	1994
0041841X	Porteville	Western Cape	1994
00106820	Riversdale	Western Cape	1994
0020618X	Robbeneiland	Western Cape	1994
00106820	Stilbaai	Western Cape	1994
00056098	Strand	Western Cape	1996
0003108A7	Struisbaai	Western Cape	1994
0007699A0	Tygerhoek	Western Cape	1994
0106880A2	Vredendal	Western Cape	2003

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Appendix B

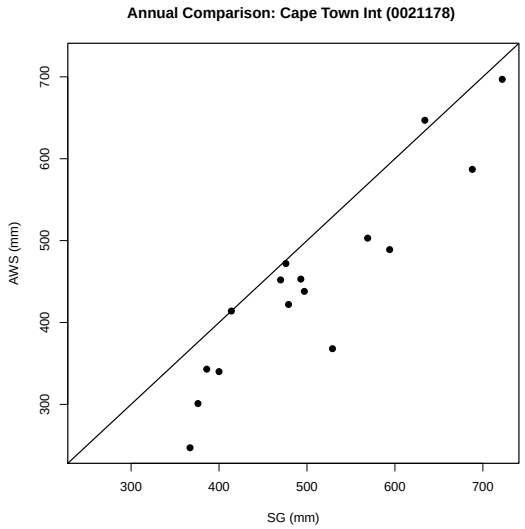
Data processing

Stellenbosch University http://scholar.sun.ac.za

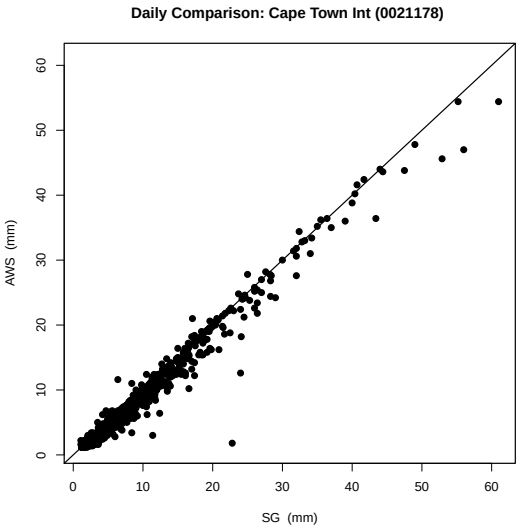
B.1 Comparison between AWS and SG data

Figures B.1 to B.7 illustrate the results of the quality control procedures for the seven stations. Similar to the autographic data, the AWS generally underestimated the SG totals on an annual scale, while daily totals compared much better with a few notable deviations. The reason for the deviations were not investigated due to time constraints, although De Jager (2012) attributes it to weather related factors like wind.

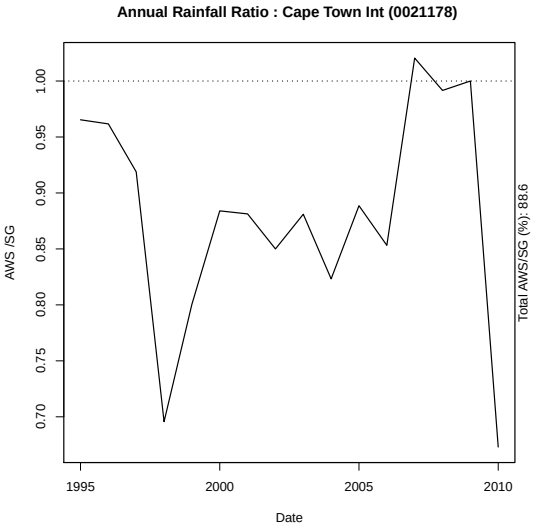
Stellenbosch University http://scholar.sun.ac.za



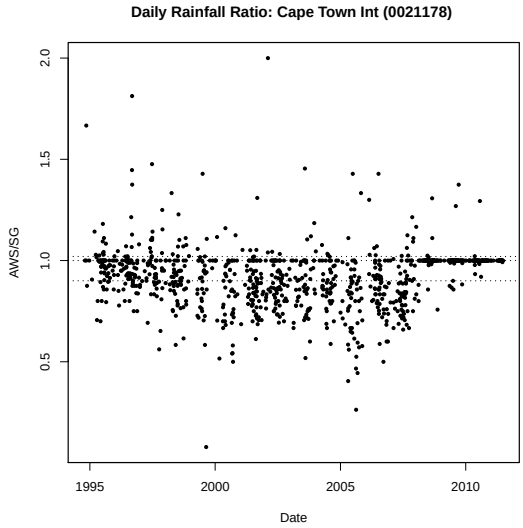
(a) annual comparison



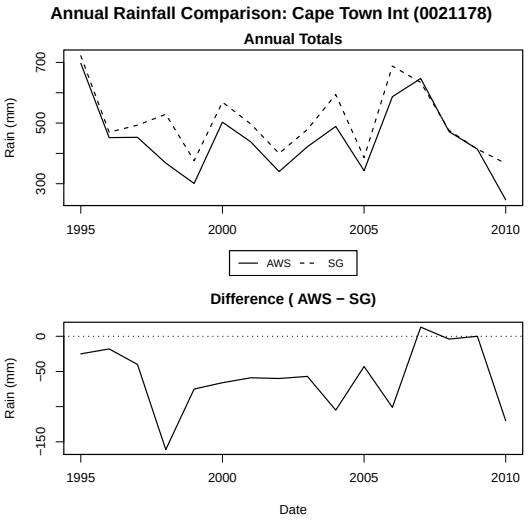
(b) daily comparison



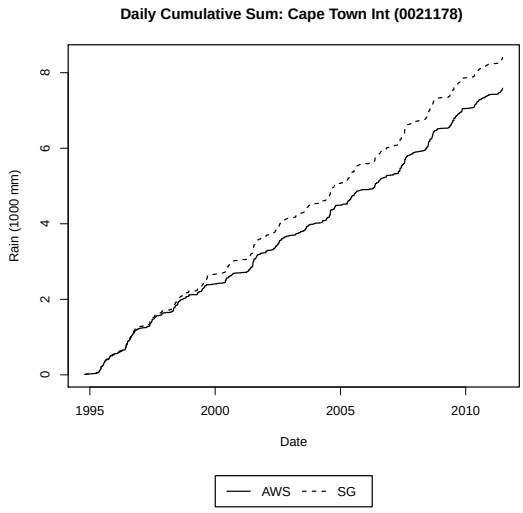
(c) annual ratio



(d) daily ratio



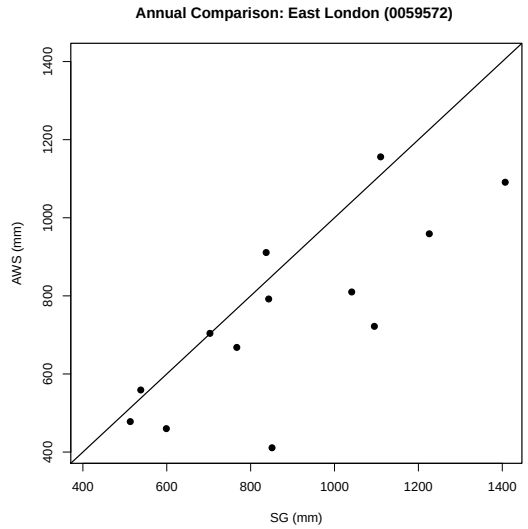
(e) annual difference



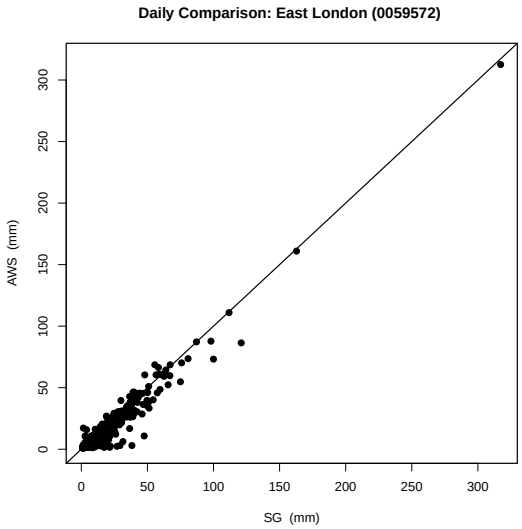
(f) cumulative sum

Figure B.1: Comparison between AWS and SG data for Cape Town

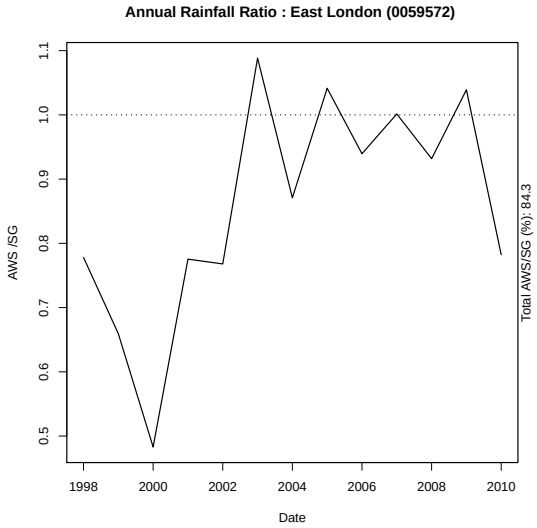
Stellenbosch University http://scholar.sun.ac.za



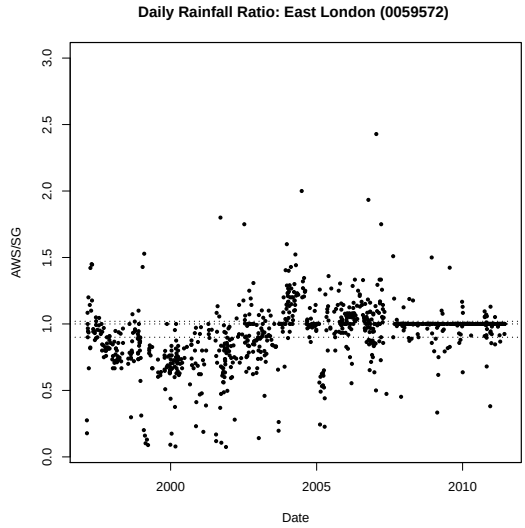
(a) annual comparison



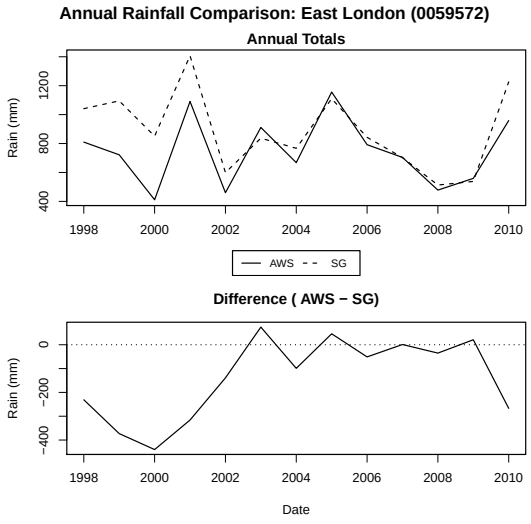
(b) daily comparison



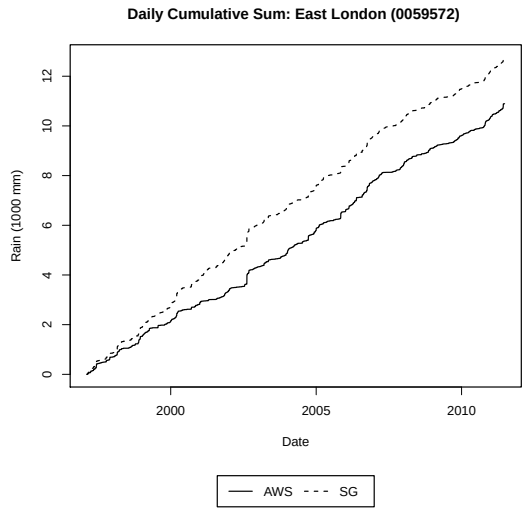
(c) annual ratio



(d) daily ratio



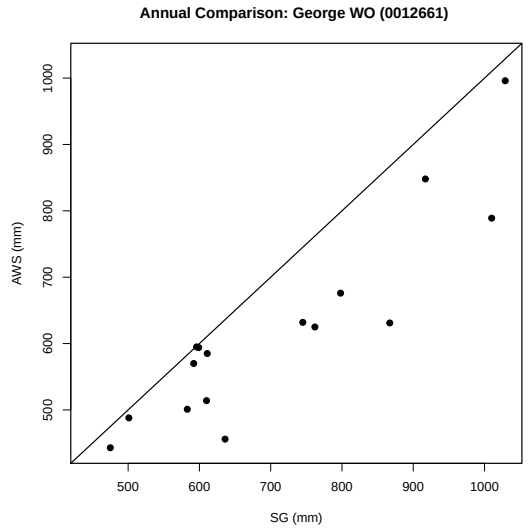
(e) annual difference



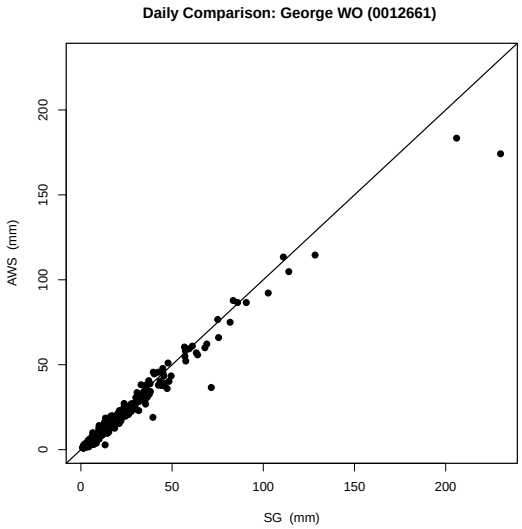
(f) cumulative sum

Figure B.2: Comparison between AWS and SG data for East London

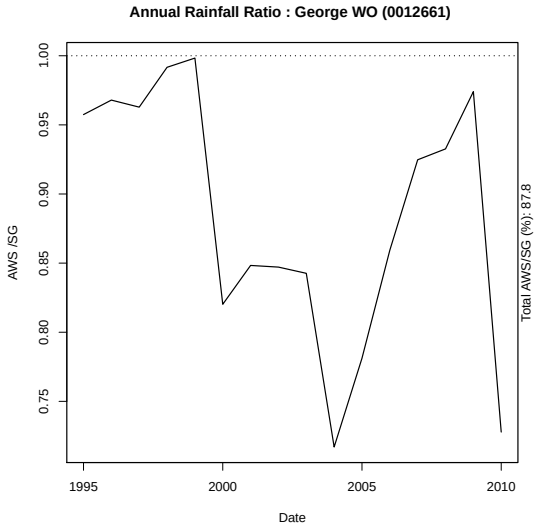
Stellenbosch University http://scholar.sun.ac.za



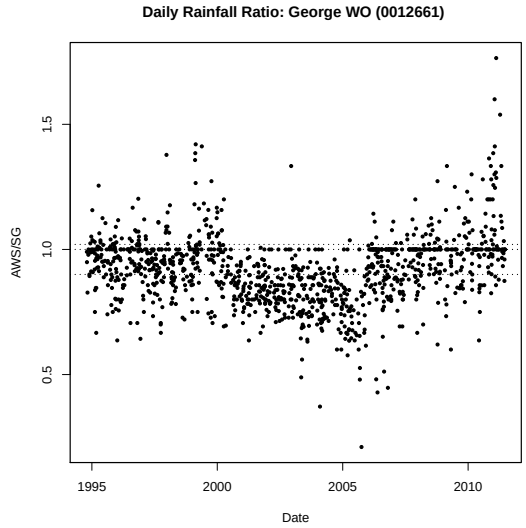
(a) annual comparison



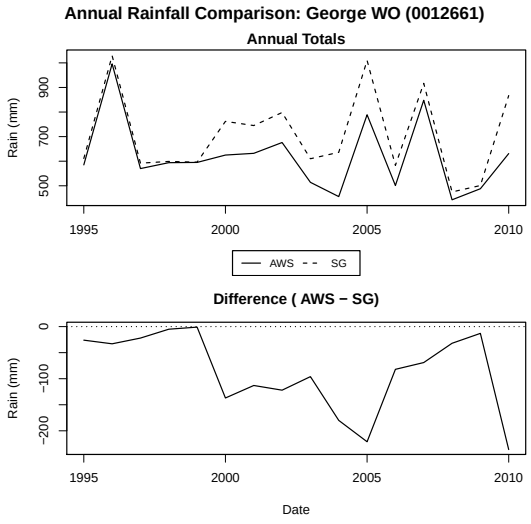
(b) daily comparison



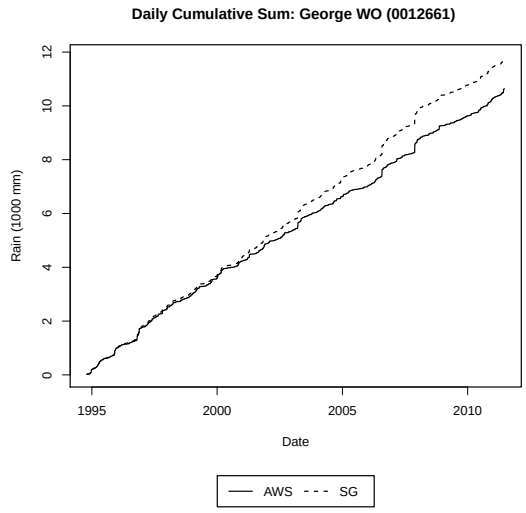
(c) annual ratio



(d) daily ratio



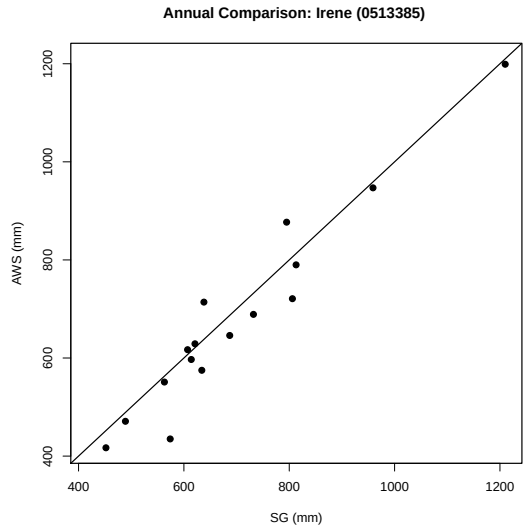
(e) annual difference



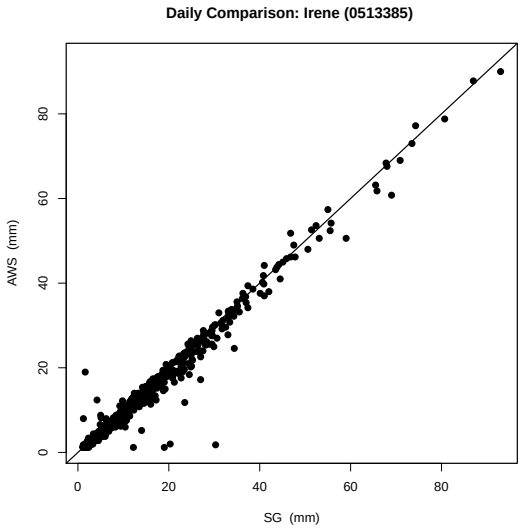
(f) cumulative sum

Figure B.3: Comparison between AWS and SG data for George

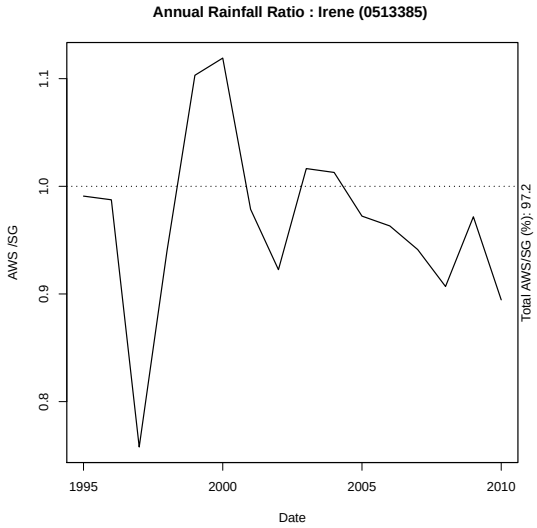
Stellenbosch University http://scholar.sun.ac.za



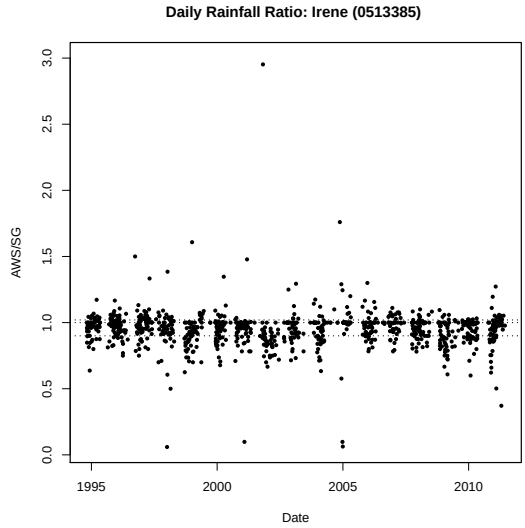
(a) annual comparison



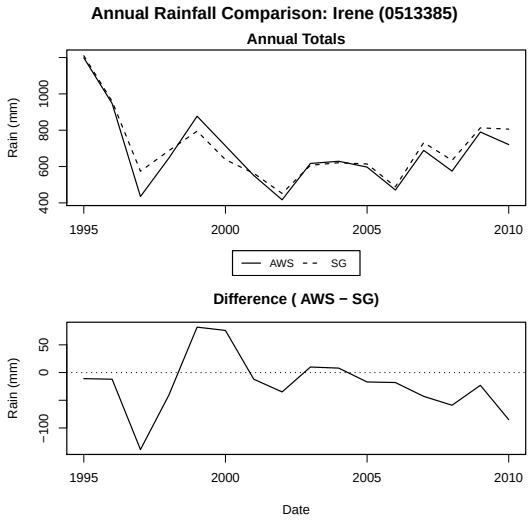
(b) daily comparison



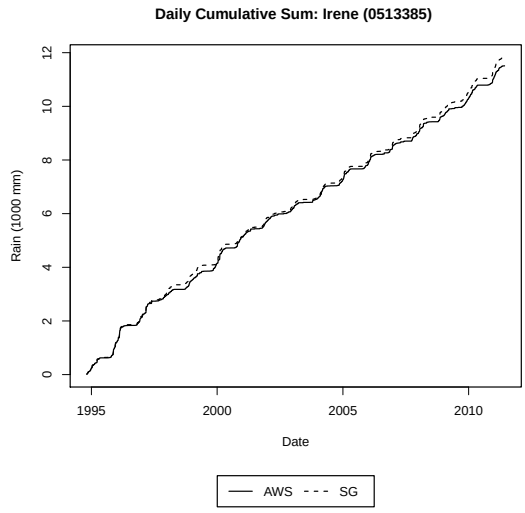
(c) annual ratio



(d) daily ratio



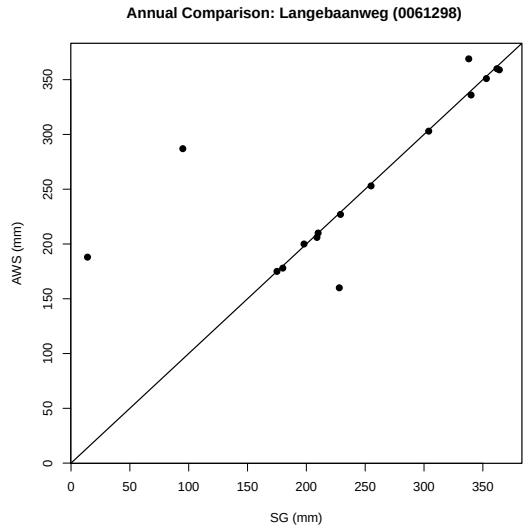
(e) annual difference



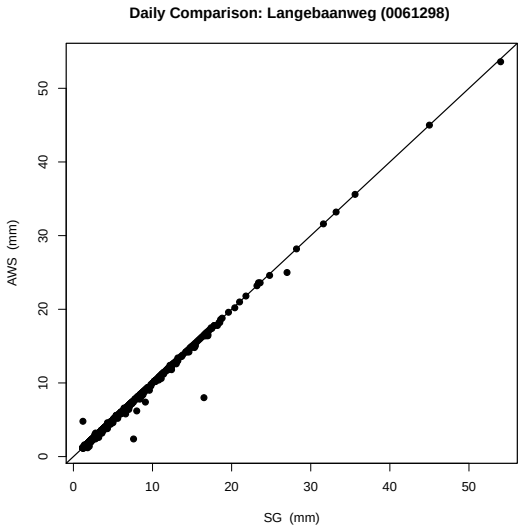
(f) cumulative sum

Figure B.4: Comparison between AWS and SG data for Irene

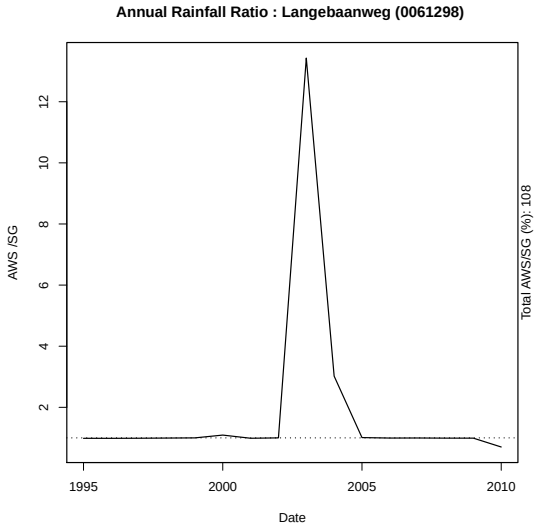
Stellenbosch University http://scholar.sun.ac.za



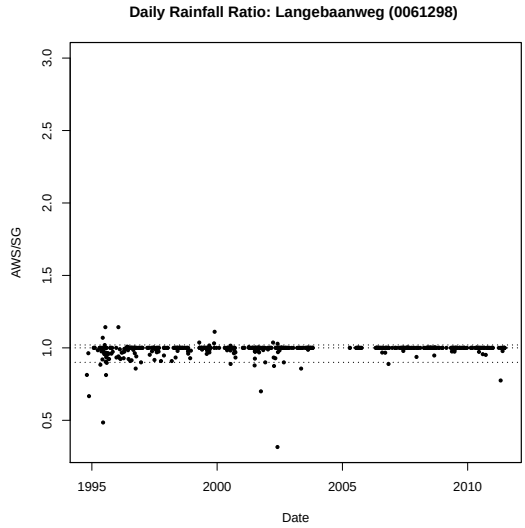
(a) annual comparison



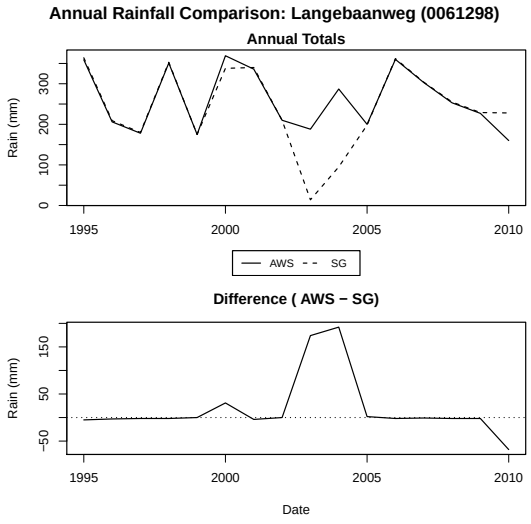
(b) daily comparison



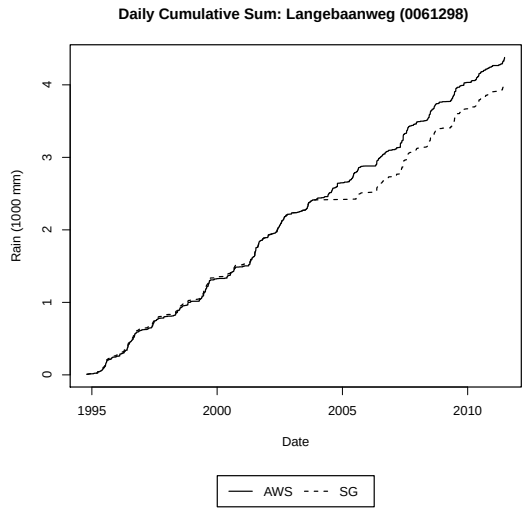
(c) annual ratio



(d) daily ratio



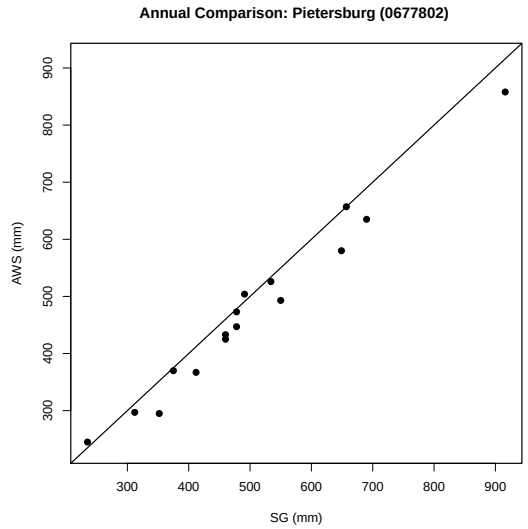
(e) annual difference



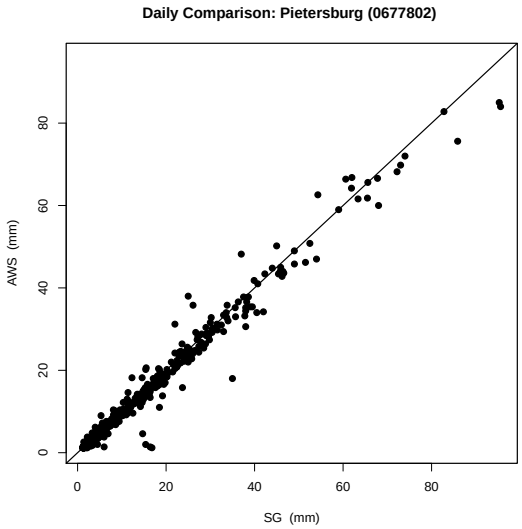
(f) cumulative sum

Figure B.5: Comparison between AWS and SG data for Langebaanweg

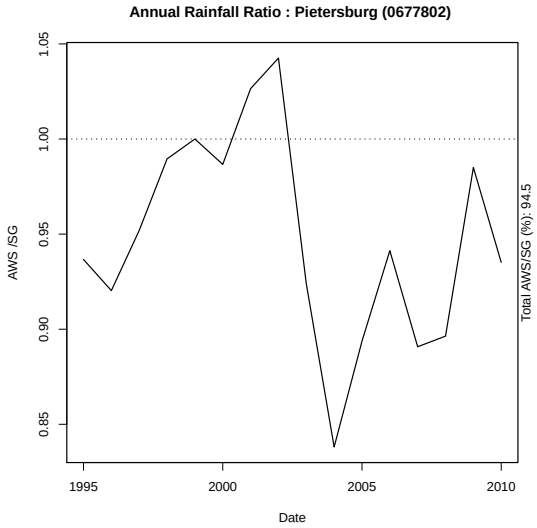
Stellenbosch University http://scholar.sun.ac.za



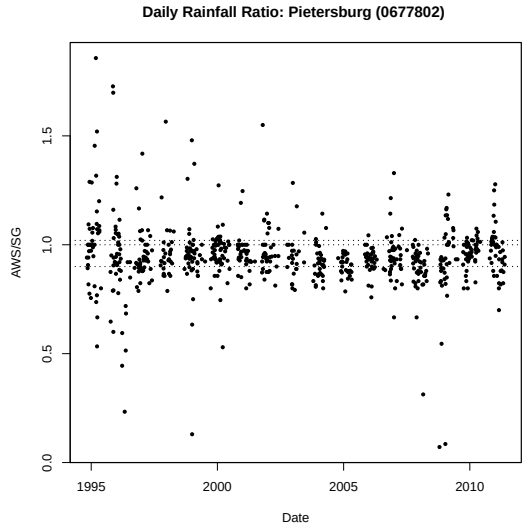
(a) annual comparison



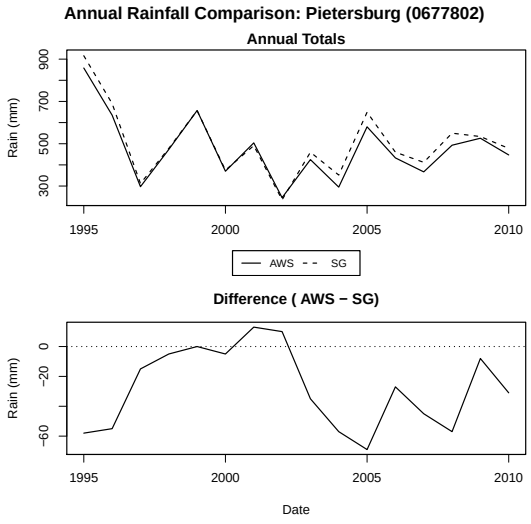
(b) daily comparison



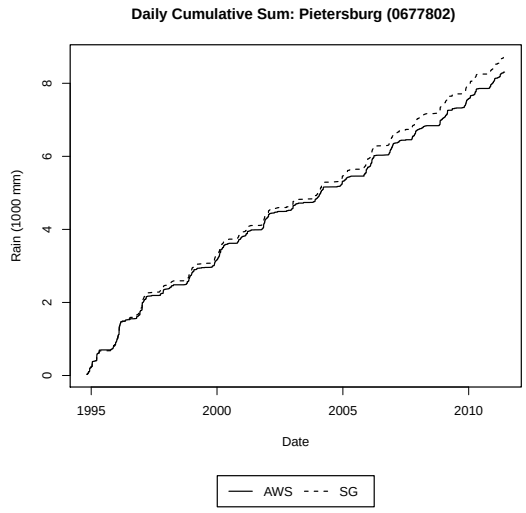
(c) annual ratio



(d) daily ratio



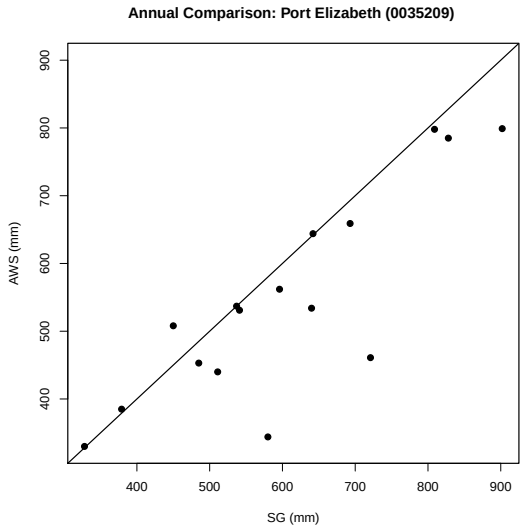
(e) annual difference



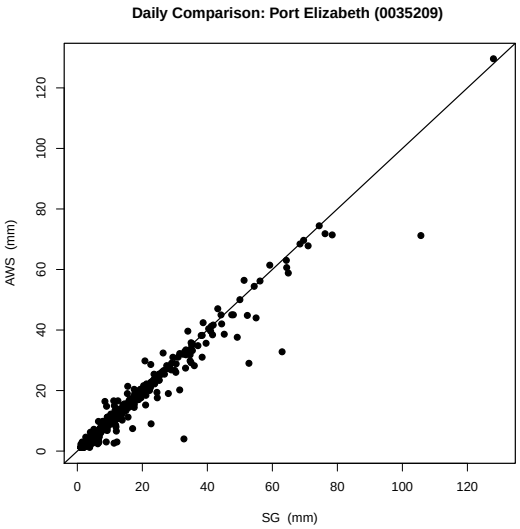
(f) cumulative sum

Figure B.6: Comparison between AWS and SG data for Polokwane

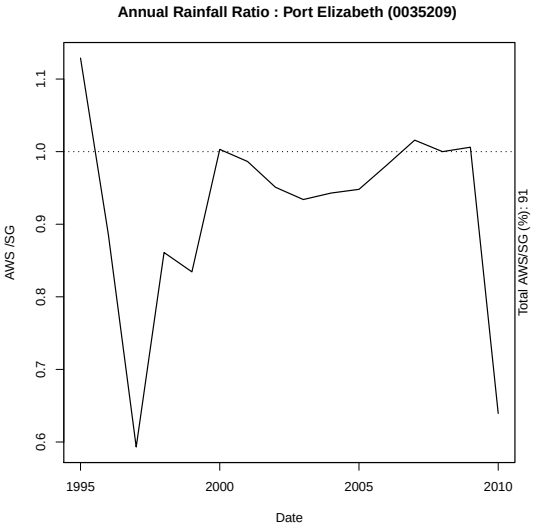
Stellenbosch University http://scholar.sun.ac.za



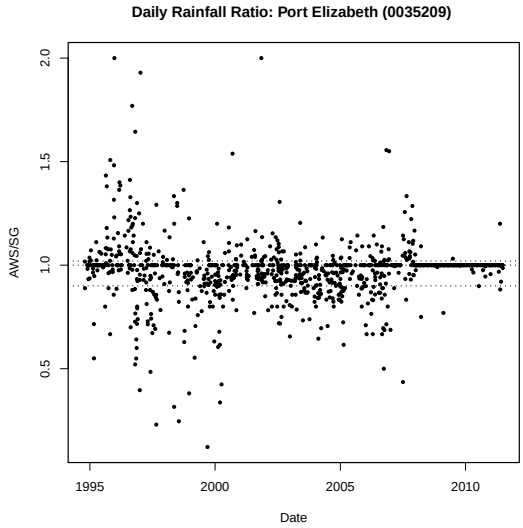
(a) annual comparison



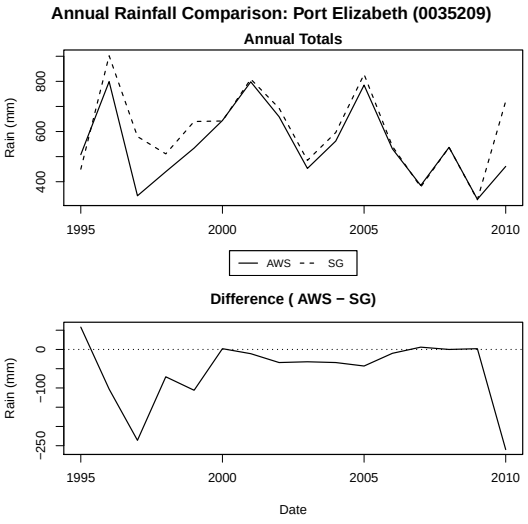
(b) daily comparison



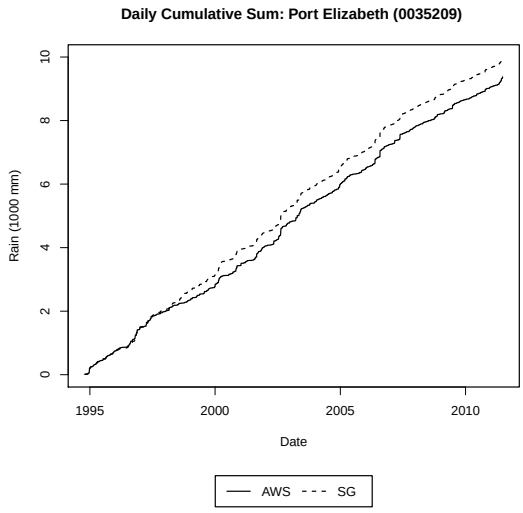
(c) annual ratio



(d) daily ratio



(e) annual difference



(f) cumulative sum

Figure B.7: Comparison between AWS and SG data for Port Elizabeth

Stellenbosch University http://scholar.sun.ac.za

B.2 Station composition

Table B.1 shows the composition of autographic and AWS data of the final 7 stations used in the analyses by using hydrological years. Please note that most digitised autographic records ended at approximately 1992 and all AWS data only started at 1994, leaving a two year gap. The autographic records for Cape Town (0021179) and George (0028690) stopped approximately a month before the hydrological year of 1991 ended. All AWS data only started on the 19th of October 1994, with a loss of nearly a month of data for the hydrological year of 1994. In spite of the loss of approximately a month's data, the hydrological year was included as to maximise the amount of data available for analysis.

Table B.1: Station composition

Station	Period
Cape Town	
0021178	1956/10/01 - 1982/02/03
0021179	1982/02/04 - 1992/08/30
0021178	1994/10/19 - 2010/09/30
Langebaanweg	
0061298	1973/10/01 - 1992/09/30
00612988	1994/10/19 - 2010/09/30
George	
0028690	1978/10/01 - 1992/08/25
00126617	1994/10/19 - 2010/09/30
East London	
0559572	1955/10/01 - 1992/09/30
0059572A3	1994/10/19 - 1997/01/31
0059572B8	1997/02/01 - 2010/09/30
Port Elizabeth	
0035179	1951/10/01 - 1992/09/30
0035209	1994/10/19 - 2010/09/30
Irene	
0513385	1975/10/01 - 1992/09/30
0513385	1994/10/19 - 2010/09/30
Polokwane	
0677802	1954/10/01 - 1992/09/30
0667802	1994/10/19 - 2010/09/30

Stellenbosch University http://scholar.sun.ac.za

Appendix C

Data analysis

Table C.1: PNS analysis for GEV and GPD distributions. Most justifiable distribution is showed per distribution. Each reading is accompanied with either a "S" or "NS" indicating the stationary or non-stationary case for the distribution. GUM represents the Gumbel model. Significant NS readings are in bold.

Dur	5min	10min	15min	30min	45min	60min	90min	2h	4h	8h	12h	18h	24h
GEV													
Cape Town	GUM-S	**GEV-NS**	GUM-S	GEV-S	GEV-S	GEV-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	**GEV-NS**	**GEV-NS**
East London	**GEV-NS**	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GEV-S	GUM-S	GUM-S	GUM-S	GEV-S
George	GUM-S	GUM-S	GUM-S	GUM-S	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**
Irene	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	**GEV-NS**	GUM-S
Langebaanweg	GEV-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S
Polokwane	GUM-S	**GEV-NS**	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GUM-S	GEV-NS	GUM-S	GEV-S
Port Elizabeth	GUM-S	GUM-S	GUM-S	GEV-S	GUM-S	GUM-S	GUM-S	GUM-S	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**	**GEV-NS**
GPD: Threshold 1 (du Plessis values)													
Cape Town	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
East London	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
George	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Irene	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Langebaanweg	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Polokwane	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Port Elizabeth	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
GPD: Threshold 2: 1.5 $\times$ du Plessis values													
Cape Town	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
East London	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
George	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Irene	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Langebaanweg	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Polokwane	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Port Elizabeth	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
GPD: Threshold 3 (Stable shape parameter method)													
Cape Town	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
East London	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
George	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Irene	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Langebaanweg	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Polokwane	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S
Port Elizabeth	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S	GPD-S

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Table C.2: NPNS analysis for GEV and GPD distributions for stations Cape Town, East London, George and Irene. "sig" indicates total number of significant slopes, "pos" indicates the total amount of positive slopes, and "neg" the total amount of negative slopes

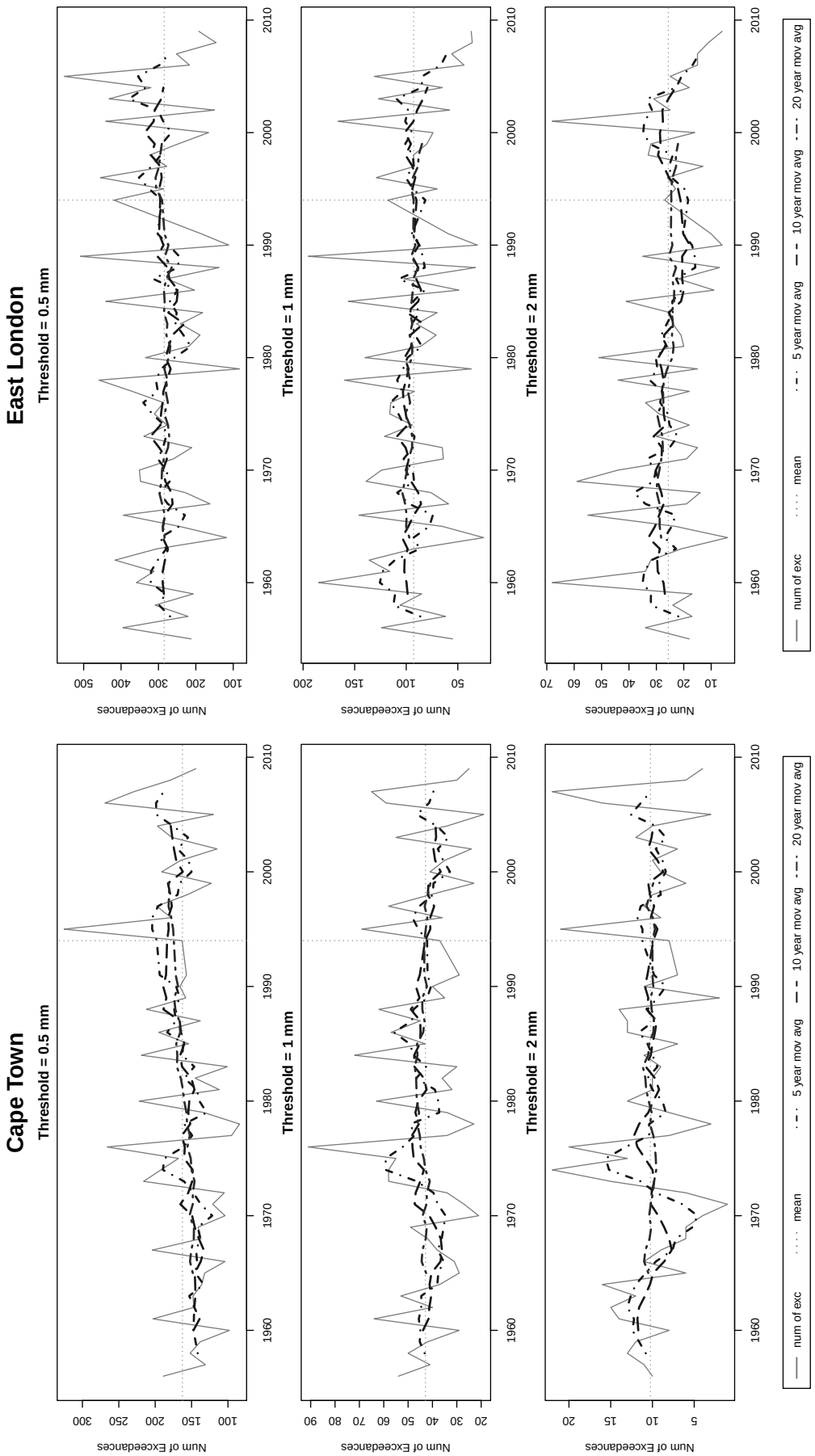
	GEV			**GPD**								
				du Plessis			1.5×du Plessis			stable shape		
Cape Town	sig	pos	neg	sig	pos	neg	sig	pos	neg	sig	pos	neg
5min	7	4	3	3	3	0	4	4	0	5	5	0
10min	8	6	2	9	9	0	8	8	0	9	9	0
15min	6	6	0	9	9	0	9	9	0	0	0	0
30min	7	7	0	8	8	0	9	9	0	8	8	0
45min	6	6	0	7	7	0	7	7	0	0	0	0
60min	3	3	0	6	6	0	5	5	0	1	1	0
90min	5	5	0	6	6	0	4	4	0	6	6	0
2h	3	3	0	6	6	0	0	0	0	6	6	0
4h	4	4	0	2	1	1	1	1	0	5	5	0
8h	5	4	1	1	1	0	8	8	0	6	6	0
12h	5	3	2	7	7	0	9	9	0	9	9	0
18h	6	5	1	4	4	0	9	9	0	9	9	0
24h	6	3	3	4	4	0	9	9	0	9	9	0
East London												
5min	9	0	9	8	0	8	8	0	8	5	0	5
10min	4	0	4	0	0	0	0	0	0	0	0	0
15min	0	0	0	0	0	0	0	0	0	0	0	0
30min	6	6	0	0	0	0	7	7	0	0	0	0
45min	6	6	0	8	8	0	9	9	0	9	9	0
60min	6	6	0	3	3	0	6	6	0	0	0	0
90min	6	6	0	5	5	0	7	7	0	0	0	0
2h	6	6	0	4	4	0	5	5	0	0	0	0
4h	6	6	0	0	0	0	0	0	0	0	0	0
8h	8	8	0	2	2	0	0	0	0	0	0	0
12h	3	3	0	2	2	0	1	0	1	0	0	0
18h	2	2	0	0	0	0	6	0	6	0	0	0
24h	0	0	0	3	3	0	0	0	0	2	2	0
George												
5min	6	6	0	6	6	0	7	6	1	8	8	0
10min	7	7	0	6	6	0	8	8	0	1	1	0
15min	7	7	0	9	9	0	9	9	0	9	9	0
30min	6	6	0	9	9	0	9	9	0	9	9	0
45min	6	6	0	9	9	0	9	9	0	9	9	0
60min	6	6	0	9	9	0	9	9	0	9	9	0
90min	7	7	0	9	9	0	9	9	0	9	9	0
2h	6	6	0	9	9	0	9	9	0	9	9	0
4h	8	8	0	9	9	0	9	9	0	9	9	0
8h	9	9	0	9	9	0	9	9	0	9	9	0
12h	9	9	0	9	9	0	9	9	0	9	9	0
18h	9	9	0	9	9	0	9	9	0	6	6	0
24h	9	9	0	9	9	0	9	9	0	9	9	0
Irene												
5min	6	6	0	6	6	0	9	9	0	7	6	1
10min	7	6	1	6	6	0	6	6	0	6	6	0
15min	6	3	3	6	3	3	6	3	3	6	3	3
30min	6	3	3	6	3	3	8	3	5	6	3	3
45min	8	2	6	9	3	6	8	3	5	6	0	6
60min	6	1	5	8	2	6	7	3	4	8	3	5
90min	5	3	2	6	3	3	6	3	3	6	3	3
2h	5	3	2	6	3	3	6	5	1	9	3	6
4h	6	4	2	6	3	3	6	3	3	6	3	3
8h	6	3	3	9	3	6	5	2	3	3	3	0
12h	8	3	5	8	3	5	6	3	3	6	6	0
18h	8	3	5	7	3	4	6	3	3	6	6	0
24h	8	3	5	8	3	5	5	2	3	8	3	5

Stellenbosch University http://scholar.sun.ac.za

Table C.3: NPNS analysis for GEV and GPD distributions for stations Langebaanweg, Polokwane and Port Elizabeth. “sig” indicates total number of significant slopes, “pos” indicates the total amount of positive slopes, and “neg” the total amount of negative slopes

	GEV			**GPD**								
				du Plessis			1.5×du Plessis			stable shape		
Langebaanweg	sig	pos	neg	sig	pos	neg	sig	pos	neg	sig	pos	neg
5min	7	7	0	8	8	0	8	8	0	9	9	0
10min	6	6	0	6	6	0	6	6	0	9	9	0
15min	6	6	0	6	6	0	6	6	0	9	9	0
30min	6	4	2	7	5	2	8	6	2	9	9	0
45min	6	4	2	7	5	2	7	5	2	9	9	0
60min	4	2	2	7	4	3	7	4	3	6	6	0
90min	3	0	3	5	2	3	5	1	4	7	7	0
2h	6	0	6	5	0	5	7	0	7	6	6	0
4h	9	0	9	9	0	9	9	0	9	7	0	7
8h	4	0	4	6	0	6	5	0	5	6	0	6
12h	6	6	0	3	3	0	3	3	0	3	3	0
18h	8	8	0	6	6	0	6	6	0	5	5	0
24h	7	7	0	3	3	0	5	5	0	6	6	0
Polokwane												
5min	7	7	0	9	9	0	9	9	0	9	9	0
10min	9	9	0	9	9	0	9	9	0	9	9	0
15min	9	9	0	9	9	0	9	9	0	9	9	0
30min	9	9	0	9	9	0	9	9	0	9	9	0
45min	9	9	0	9	9	0	9	9	0	9	9	0
60min	9	9	0	9	9	0	9	9	0	9	9	0
90min	9	9	0	9	9	0	9	9	0	9	9	0
2h	9	9	0	9	9	0	9	9	0	9	9	0
4h	7	7	0	9	9	0	9	9	0	9	9	0
8h	8	8	0	9	9	0	9	9	0	9	9	0
12h	9	9	0	9	9	0	9	9	0	9	9	0
18h	9	9	0	9	9	0	9	9	0	9	9	0
24h	9	9	0	9	9	0	9	9	0	9	9	0
Port Elizabeth												
5min	3	3	0	2	0	2	0	0	0	4	4	0
10min	0	0	0	2	0	2	4	0	4	0	0	0
15min	2	2	0	7	0	7	2	0	2	2	0	2
30min	5	0	5	9	0	9	9	0	9	6	0	6
45min	7	0	7	9	0	9	9	0	9	6	0	6
60min	9	0	9	9	0	9	9	0	9	3	0	3
90min	9	0	9	9	0	9	9	0	9	0	0	0
2h	9	0	9	9	0	9	8	0	8	8	0	8
4h	8	0	8	9	0	9	9	0	9	6	0	6
8h	9	0	9	9	0	9	9	0	9	9	0	9
12h	9	0	9	9	0	9	9	0	9	9	0	9
18h	9	0	9	9	0	9	9	0	9	4	0	4
24h	7	0	7	9	0	9	9	0	9	5	0	5

Stellenbosch University http://scholar.sun.ac.za

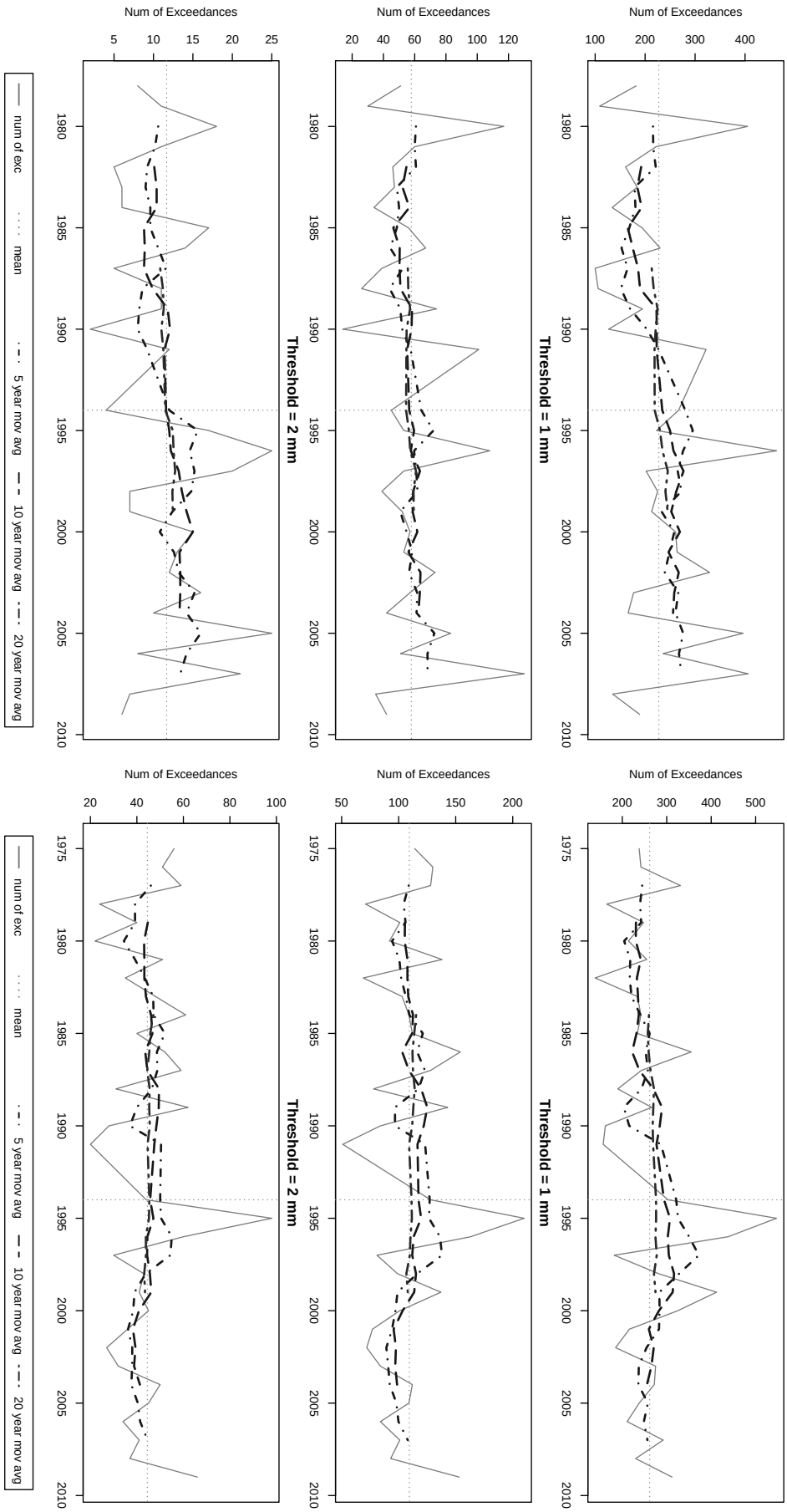


(a) Cape Town

(b) East London

Figure C.1: Frequency of exceedance of threshold plot for Cape Town and East London.

Stellenbosch University http://scholar.sun.ac.za

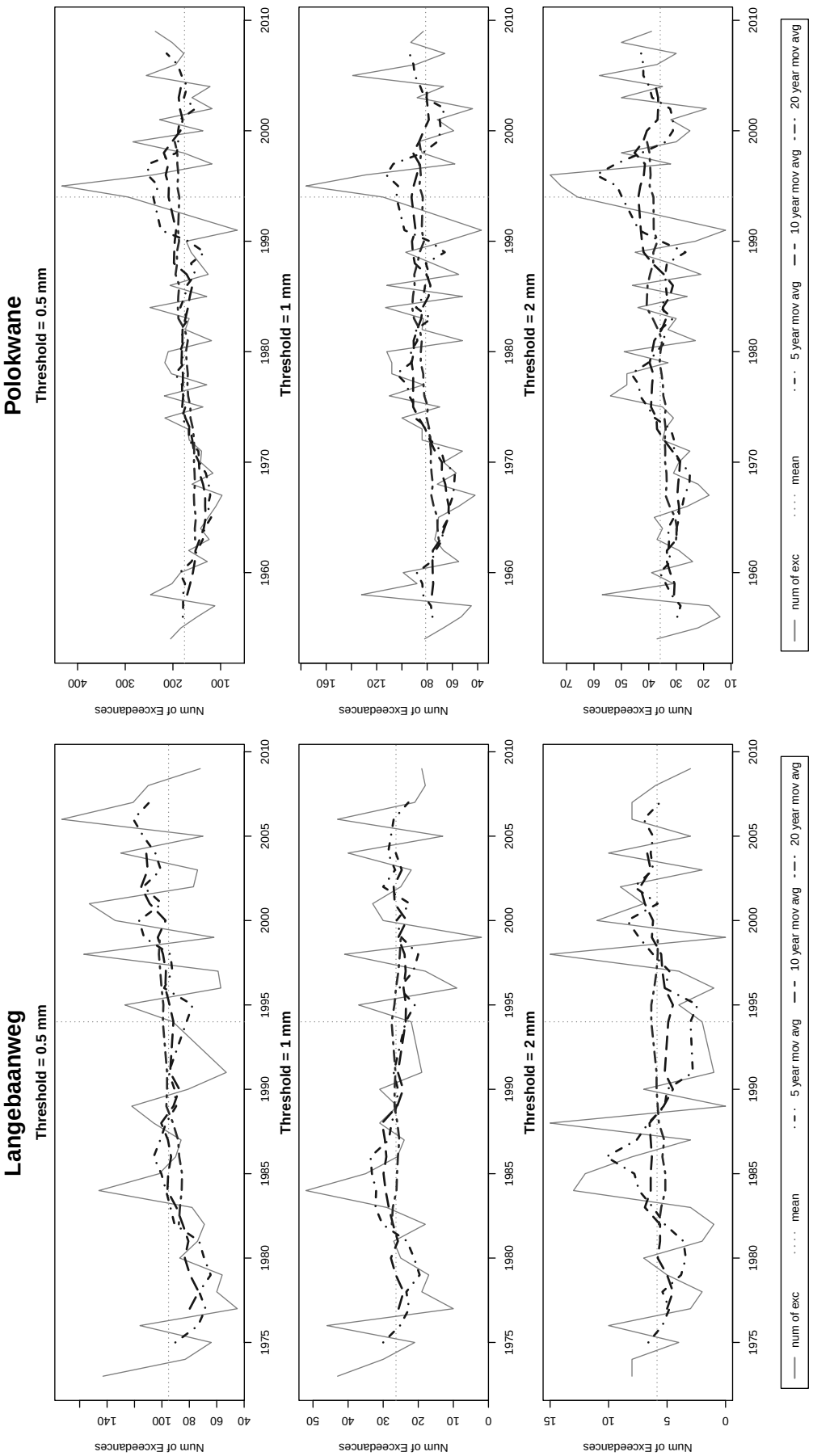


(a) George

(b) Irene

Figure C.2: Frequency of exceedance of threshold plot for George and Irene.

Stellenbosch University http://scholar.sun.ac.za



(a) Langebaanweg

(b) Polokwane

Figure C.3: Frequency of exceedance of threshold plot for Langebaanweg and Polokwane.

Stellenbosch University http://scholar.sun.ac.za

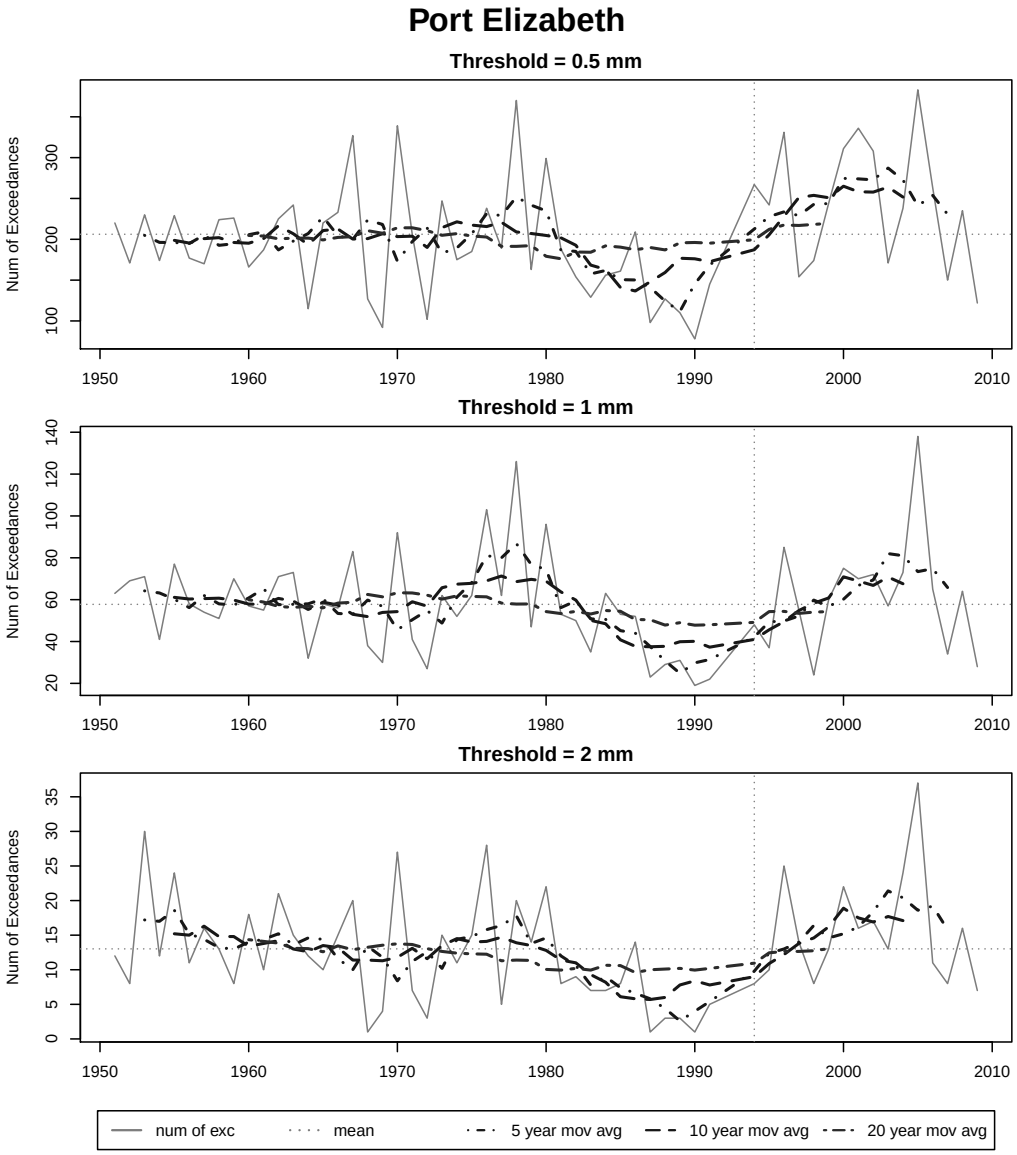


Figure C.4: Frequency of exceedance of threshold plot for Port Elizabeth

Stellenbosch University http://scholar.sun.ac.za

Appendix D

Data analysis example

This chapter provides a detailed example of the data analysis calculations in chapter 4.3, consisting of the magnitude and frequency analyses. The short duration rainfall data of Cape Town will be used as an example. Please note that hydrological years (October 1st to September 30th) were used.

D.1 Magnitude analysis

The magnitude analysis made use of the peak selection method described in Section 3.2.4, where rainfall peaks had to separated by at least 24 hours of no rainfall. Analysis was divided into the parametric non-stationary (PNS) and non-parametric non-stationary (NPNS) methodology for both the generalised extreme value distribution (GEV) and the generalised Pareto distribution (GPD).

D.1.1 Parametric non-stationary

The approaches taken for each extreme value distribution will be discussed separately.

D.1.1.1 Generalised extreme value distribution

The following example will use the peak data series for a storm duration of 10 minutes for Cape Town, with a record length of 52 years.

First, the annual maxima series must be selected, as illustrated in Figure D.1. Please note that the AMS values for 1992 and 1993 are missing as there is a gap in the record between 1992 and 1994 for all stations.

For a standard GEV model

Stellenbosch University http://scholar.sun.ac.za

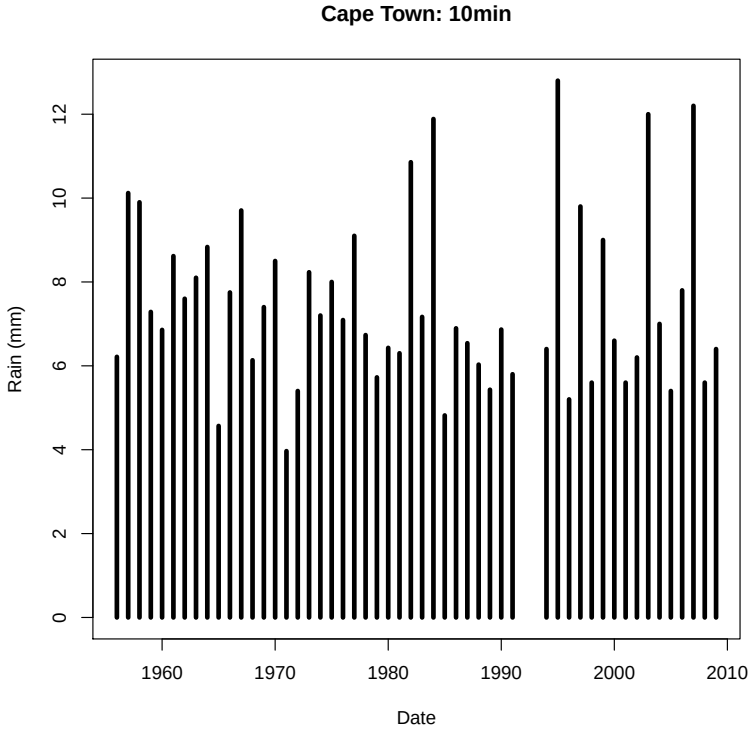


Figure D.1: AMS for Cape Town for a storm duration of 10 minutes.

$$F(x) = \begin{cases} \exp\left(-\left[1-\xi\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right) & \text{if } \xi \neq 0 \\ \exp\left(-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right) & \text{if } \xi = 0 \end{cases} \tag{D.1}$$

a linear time model for each parameter of the GEV is tested, starting with the shape parameter, ξ, then the scale parameter, σ and lastly the location parameter, μ.

Starting with the linear shape parameter, the model:

$$\xi(t) = \beta_0 + \beta_1 t \tag{D.2}$$

is fitted to the GEV, resulting in the following model:

$$F(x,t) = \exp\left(-\left[1-\xi(t)\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi(t)}\right) \tag{D.3}$$

where t is the series of years of the AMS values (1956, 1957...,2009).

The maximum likelihood (ML) method is applied to this model to obtain the values of the parameters, including β_0 and β_1, the intersection and slope of the $\xi(t)$ model. The log-likelihood

Stellenbosch University http://scholar.sun.ac.za

function

$$l(\mu,\sigma,\xi) = -m\log\sigma - (1+1/\xi)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi} \tag{D.4}$$

is maximised by setting the partial derivatives of each parameter for Equation D.4 equal to zero and finding the solution of the derivatives. As there is no analytical solution for the equations, numerical optimisation techniques are required. The `ismev` package in *R* was used to apply the ML technique to non-stationary distributions. For this particular example, the ML method estimated the $\xi(t)$ model parameters as:

$$\begin{array}{llll} \beta_0 & = -0.363 & \beta_1 & = 0.017 \\ se(\beta_0) & = 0.146 & se(\beta_1) & = 0.008 \end{array}$$

where *se()* refers to the standard error of the parameter.

Next, the significance of the model is tested with a t-test, similar to linear regression. The approach followed in Montgomery and Runger (2007:405-407) is applied here. A hypothesis test is applied to the slope of the model:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

H_0 is rejected if:

$$|t_0| > t_{\alpha/2,n-2}$$

where:

- t_0 is the test statistic that is supposed to follow a t distribution
- $t_{\alpha/2,n-2}$ is the $\alpha/2$ percentage point value of a t-distribution with $n-2$ degrees of freedom
- α is a confidence factor
- n is the record length

The value $t_{\alpha/2,n-2}$ can either be read off from a table, (see the table in Montgomery and Runger (2007:715)) or calculated with numerical approximation methods. The t_0 value is calculated as

Stellenbosch University http://scholar.sun.ac.za

$$t_0 = \frac{\beta_1}{se(\beta_1)} \tag{D.5}$$

For this specific example and for a confidence level of $\alpha = 0.05$:

$$\begin{aligned} t_0 &= \frac{0.017}{se(0.008)} \\ &= 2.103 \end{aligned}$$

and

$$\begin{aligned} t_{\alpha/2,n-2} &= t_{0.025,50} \\ &= 2.01 \end{aligned}$$

$t_0 > t_{0.025,50}$, therefore H_0 is rejected. For an $\alpha = 0.05$ level of significance, the slope does not equal zero, therefore the linear model for $\xi(t)$ is significant.

Next, the σ parameter is tested. If the ξ parameter is significant, it will be included in the GEV model as

$$F(x,t) = \exp\left(-\left[1-\xi(t)\left(\frac{x-\mu}{\sigma(t)}\right)\right]^{-1/\xi(t)}\right)$$

otherwise, only the $\sigma(t)$ model will be tested in the GEV. The same applies for the $\mu(t)$ linear model.

Once all the parameters have been tested, if at least one parameter produced a significant linear model, a maximum likelihood ratio test is applied. The maximum likelihood ratio test compares a more complex model to a simpler model in order to assess if the complex model is justified. The maximum likelihood ratio test is:

$$D = -2\{l_1(M_1) - l_0(M_0)\} \tag{D.6}$$

where

- D is the deviance statistic
- M_1 and M_0 are the more and less complex model respectively

Stellenbosch University http://scholar.sun.ac.za

- l_1 and l_0 are the log-maximum likelihoods for each model (maximised value from equation D.4)

The more complex model, M_1, is accepted if

$$D > \chi^2_{\alpha,k}$$

where:

- α is a confidence factor
- $\chi^2_{\alpha,k}$ is the α quantile for the chi-square distribution, where k is the difference between the amount of parameters of the models. Like with the t-distribution, the $\chi^2_{\alpha,k}$ value can either be read off from a table (see Montgomery and Runger (2007:714)) or determined using numerical approximations with a computer.

For this specific example, only the the shape parameter of the GEV produced a significant fit for a linear model, therefore the non-stationary model has one parameter for the σ and μ values each, and two for the $\xi(t)$ parameter, hence

$$\begin{aligned} k &= \text{total number of parameters in non-stationary model} \\ &\quad - \text{total number of parameters for stationary model} \\ &= (1+1+2)-3 \\ &= 4-3 \\ &= 1 \end{aligned}$$

For a confidence factor of $\alpha = 0.05$, $\chi^2_{0.05,1} = 3.84$. From equation D.4, the log-maximum likelihoods for the non-stationary model was: $l_1(M_1) = 102.92$, while for the stationary model: $l_0(M_0) = 105.84$, hence

$$\begin{aligned} D &= -2\{102.92 - 105.84)\} \\ &= 5.82 \end{aligned}$$

$D > \chi^2_{\alpha,k}$ and therefore the non-stationary model is accepted. If the non-stationary model were rejected, the appropriateness of a Gumbel model would be tested. The Gumbel model is a special case of the GEV, where $\xi = 0$, as given in:

Stellenbosch University http://scholar.sun.ac.za

$$F(x) = \exp\left(-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right) \tag{D.7}$$

The appropriateness of the Gumbel model is tested by checking if a stationary GEV model's ξ value includes 0 in its 95% confidence interval. If so, then the Gumbel model is used for another parametric non-stationary test, where the Gumbel model is now used for a PNS test instead of the GEV.

The confidence interval of the ξ is determined by the ML procedure and is illustrated in Figure D.2. The vertical lines indicate the 95% confidence interval, while the turning point of the curve indicates the actual value of ξ. For this specific example, the confidence interval includes 0, so the Gumbel model is appropriate. However, non-stationary behaviour was already observed, so the Gumbel model was not necessary in this example. If the Gumbel model is not appropriate, then it is concluded that no linear parametric non-stationary behaviour is observed.

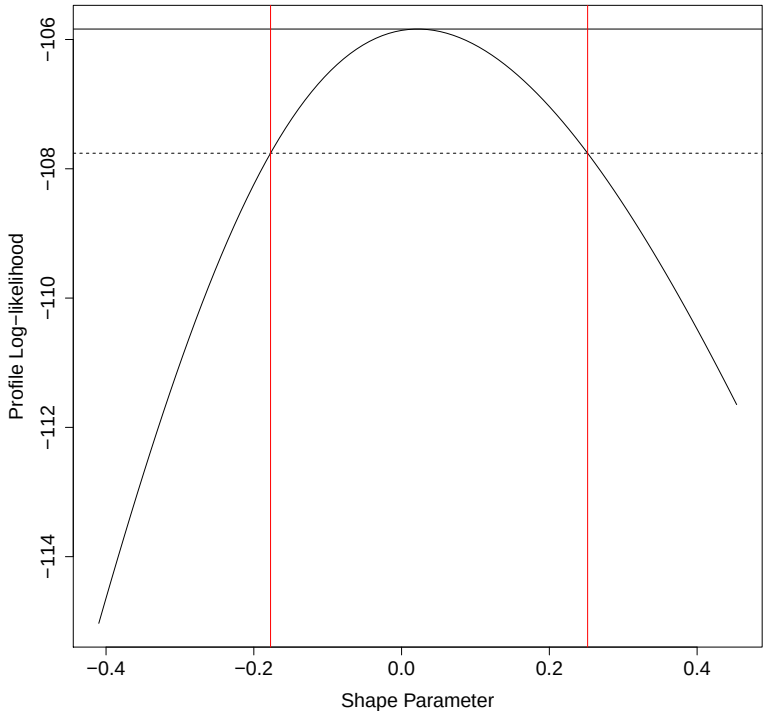


Figure D.2: Profile log-likelihood for the shape parameter ξ with additional 95% confidence intervals.

The same procedure explained above is applied to all storm duration data and all significant non-stationary behaviour is noted. For Cape Town, non-stationary behaviour was only observed for storm durations of 10 minutes, 18 and 24 hours. With only 3 out of 13 storm durations showing linear non-stationarity, it is concluded that there is a lack of evidence to support linear non-stationary behaviour at Cape Town.

Stellenbosch University http://scholar.sun.ac.za

D.1.1.2 Generalised Pareto distribution

For this example, as with the GEV, storm duration data of 10 minutes for Cape Town will be used. The PNS analysis for a GPD is similar to the PNS analysis for the GEV, except that peaks over threshold (POT) data is used. Three threshold types were used:

1. Cut-off values used by Du Plessis (1992): u_d
2. 1.5 $\times$ the Du Plessis (1992) values: $u_{1.5d}$
3. Stable shape parameter method: u_{stb}

For a storm duration of 10 minutes, $u_d = 3$mm, $u_{1.5d} = 4.5$mm. Using the stable shape parameter method described in Section 3.3.1.1, $u_d = 2$mm. The peak data series for 10 minutes and the three threshold types are illustrated in Figure D.3. For the POT data series, all data above the threshold values are used in the GPD.

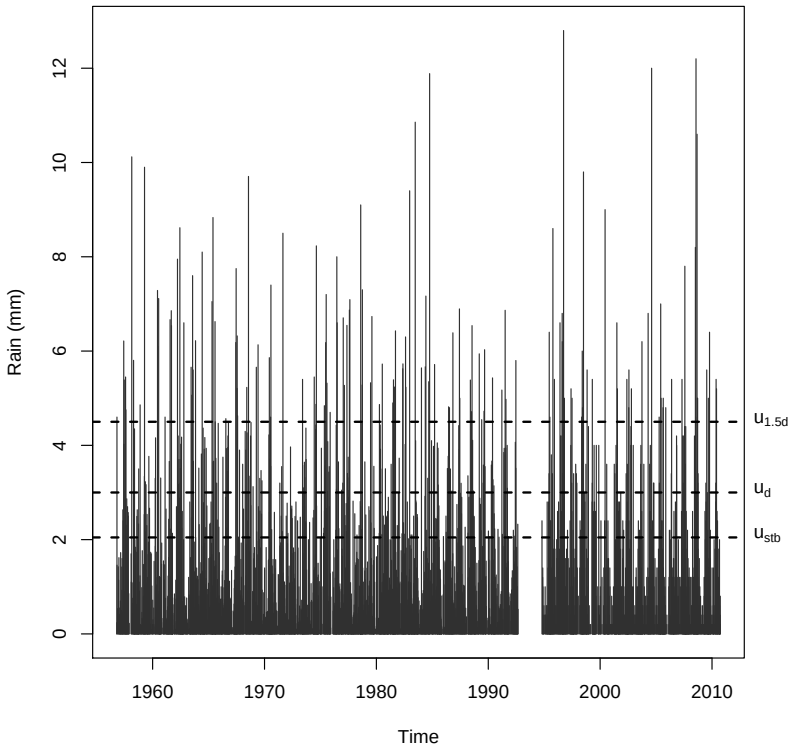


Figure D.3: Peak data series and three threshold types.

Once the POT data series is selected, linear time models, like in Equation D.2 are fitted to the GPD:

Stellenbosch University http://scholar.sun.ac.za

$$F(x) = \begin{cases} 1 - \left[1 - \xi\left(\frac{x-u}{\sigma}\right)\right]^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \exp\left[1 - \left(\frac{x-u}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \tag{D.8}$$

for the two parameters $\xi(t)$ and $\sigma(t)$, where t is the time steps of the POT data, for example for $u_{1.5d}$:

time stamp	peak rainfall value (mm)
1956/10/23 08:45:00	4.6
1957/06/01 20:45:00	6.2
⋮	⋮
2010/05/27 19:55:00	5.2

The parameters are estimated by maximising the function

$$l(\sigma,\xi) = -k\log\sigma - (1+1/\xi)\sum_{i=1}^{k}\log\left[1+\xi\left(\frac{x_i-u}{\sigma}\right)\right] \tag{D.9}$$

in the same way as with the GEV (Coles, 2001:80). The rest of the method proceeds in the same manner as the GEV, with the significance of the linear parameters tested with a t-test, and then if a non-stationary model is attained, it is compared to a stationary model with the likelihood ratio test.

In the case of the threshold $u_{1.5d}$, no linear non-stationarity was observed, nor for any of the other threshold values. None of the other storm duration values produced any significant non-stationarity either.

D.1.2 Non parametric non-stationary

As with the NPS, the procedure for the GEV and GPD will be discussed separately. The NPNS procedure involves plotting the return levels of a moving annual window period for specific return level for a specific distribution. A storm duration data series of 10 minutes for Cape Town will be used as an example.

D.1.2.1 Generalised extreme value distribution

For this example, a window period of 15 years and a return period of 10 years was used. The first group of data equal to the window period, 1956-1970, is selected from the AMS values,

Stellenbosch University http://scholar.sun.ac.za

as illustrated in Figure D.4. The figure on the left illustrates data selected from the AMS series (grey area), while the figure on the right illustrates the selected 15 year data.

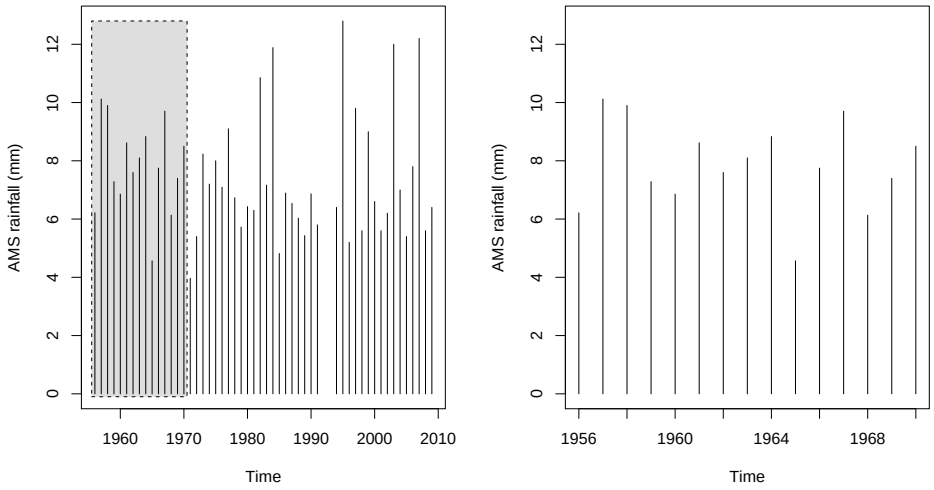


Figure D.4: Window period selection on AMS values for the years 1956-1970.

Next, a stationary GEV model is fitted to the selected set. The return level and corresponding 95% confidence intervals for a return period of 10 years is obtained. Figure D.5 illustrates the return level at the single point with accompanying 95% confidence interval over the original AMS series. Notice that the return level is centred over the window period, that is, the return level is logged for the year $1956 + 15/2 = 1963.5 \approx 1963$.

Next, the moving window period moves to the next year for the period 1957-1971, and the same procedure is applied to this period. The next return level, at 1964, is obtained and illustrated in Figure D.6.

This procedure then applied until the end of record is reached. The whole range of return levels are illustrated in Figure D.7, where the points represent the return levels, and the grey area the 95% confidence intervals for the return levels.

Linear regression is now applied the range of return levels:

$$RL(t) = \beta_0 + \beta_1 t \tag{D.10}$$

where:

- RL is the regression equation for the return levels
- t is the time of occurrence of the return levels
- β_0 is the intersection value of the line

Stellenbosch University http://scholar.sun.ac.za

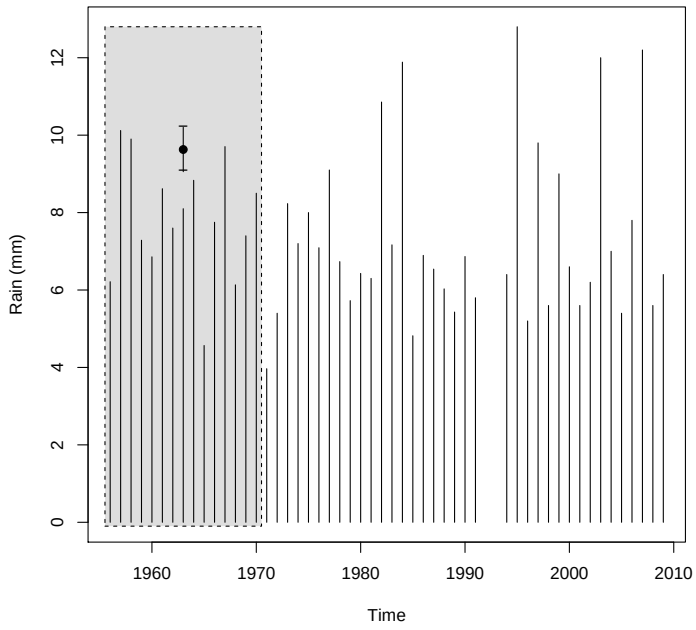


Figure D.5: Return level value for the window period 1956-1970.

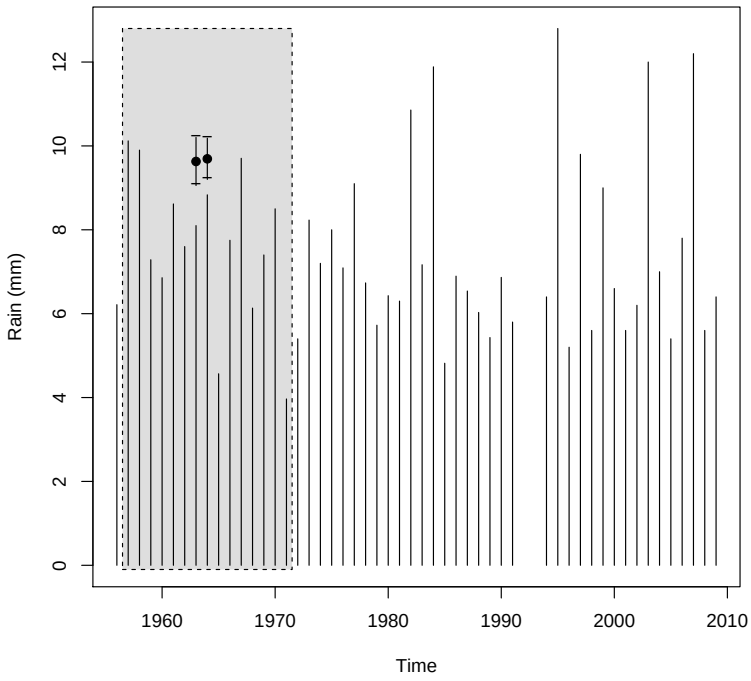


Figure D.6: Return level value for the window period 1957-1971.

Stellenbosch University http://scholar.sun.ac.za

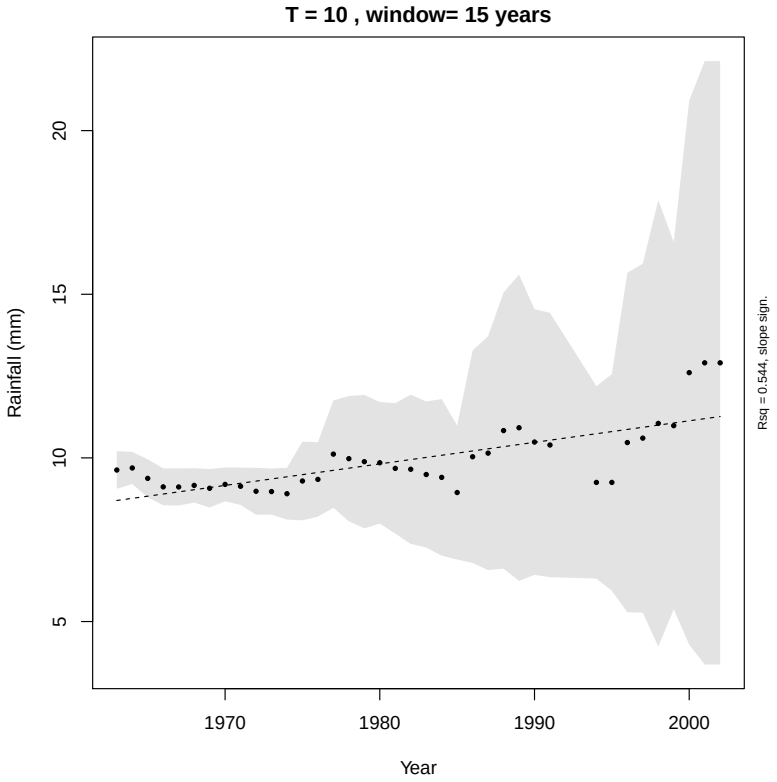


Figure D.7: Return level values for the for the entire record length.

- β_1 is the slope value of the line

For this example, $\beta_0 = -120.46$ and $\beta_1 = 0.066$, and the coefficient of determination is $R^2 = 0.54$. The regression line is included in Figure D.7. A t-test is applied to the slope of the line for $\alpha = 0.05$ and a record length of $n = 38$ years. The standard error of β_1 was determined as $se(\beta_1) = 0.010$. From equation D.5, the t_0 value was determined as:

$$\begin{aligned} t_0 &= \frac{0.066}{se(0.010)} \\ &= 6.54 \end{aligned}$$

Compared to the $t_{\alpha/2,n-2}$ value of 2.03, $t_0 > t_{\alpha/2,n-2}$, therefore the slope is significant.

The same procedure is applied for a range of window (15, 20, 25 years) and return periods (T=5, 10, 20 years). A summary for the whole range of window and return periods is seen in Figure D.8. All significant slopes, and their corresponding sign (either positive or negative) is recorder per duration. A single duration would produce a maximum of nine results. Cape Town produced a total of 8 significant slopes for a storm duration of 10 minutes, where 6 were positive and 2 negative.

The results of all the durations is:

Stellenbosch University http://scholar.sun.ac.za

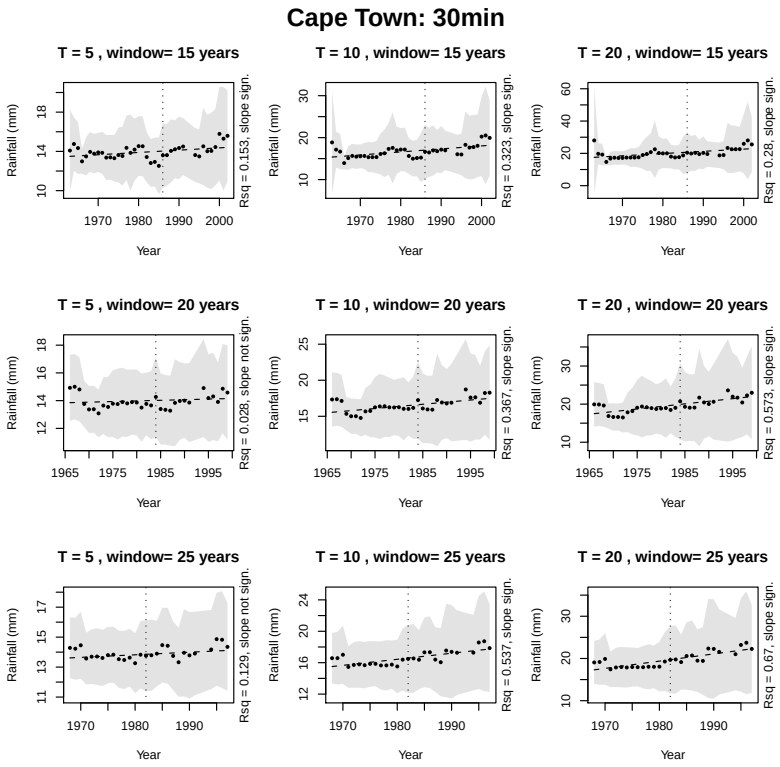


Figure D.8: Return level values the whole range of window and return period for the GEV.

	total	pos	neg
5min	7	4	3
10min	8	6	2
15min	6	6	0
30min	7	7	0
45min	6	6	0
60min	3	3	0
90min	5	5	0
2h	3	3	0
4h	4	4	0
8h	5	4	1
12h	5	3	2
18h	6	5	1
24h	6	3	3

D.1.2.2 Generalised Pareto distribution

As with the GEV, a window period of 15 years and a return period of 10 years is used in this example. For the first window period, peak data above thresholds for the period 1956/10/01 - 1971/09/30 is selected, as illustrated in Figure D.9. The figure on the left illustrates the first window period, highlighted in grey, while the figure on the right illustrates the data set along

Stellenbosch University http://scholar.sun.ac.za

with the thresholds. Data above the threshold value is used fitted to a stationary GPD model.

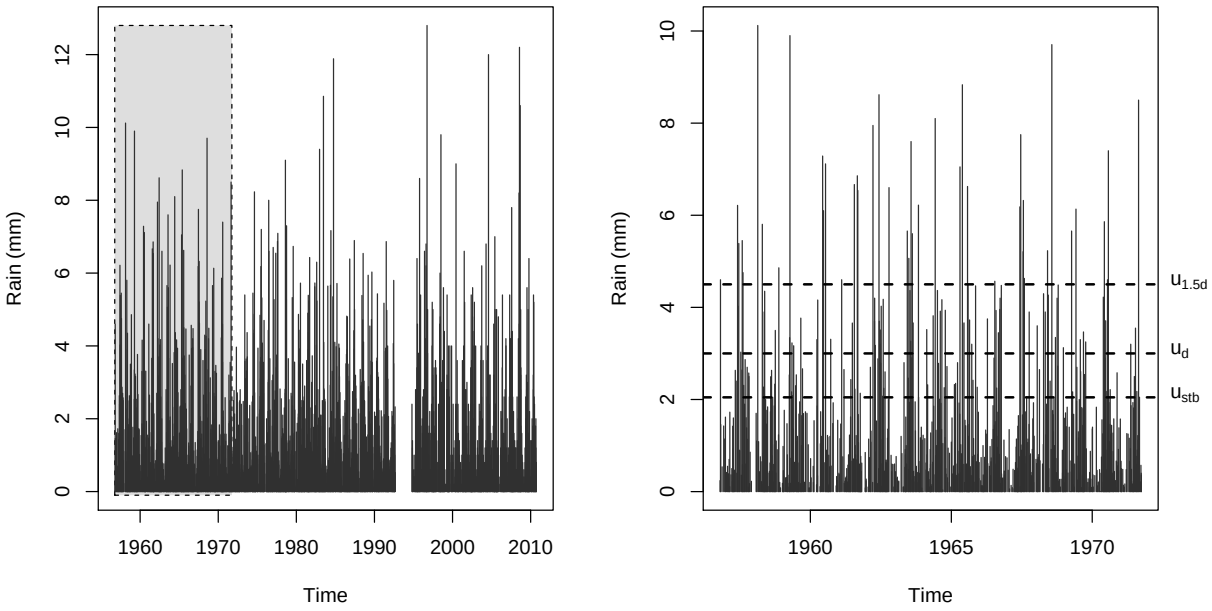


Figure D.9: Window period selection on POT values for the years 1956-1971.

The rest of the procedure is exactly the same as with the GEV procedure. Figure D.10 illustrates the return levels the whole range of the record and the linear regression line fitted to the return levels. The significance of the slope is indicated on the right hand margin as “slope-sign.”.

For the whole range of window and return periods, Figure D.11 is obtained. A summary of all the significant slopes over all the durations for $u_{1.5d}$ is given below:

	total	pos	neg
5min	5	5	0
10min	9	9	0
15min	0	0	0
30min	8	8	0
45min	0	0	0
60min	1	1	0
90min	6	6	0
2h	6	6	0
4h	5	5	0
8h	6	6	0
12h	9	9	0
18h	9	9	0
24h	9	9	0

Stellenbosch University http://scholar.sun.ac.za

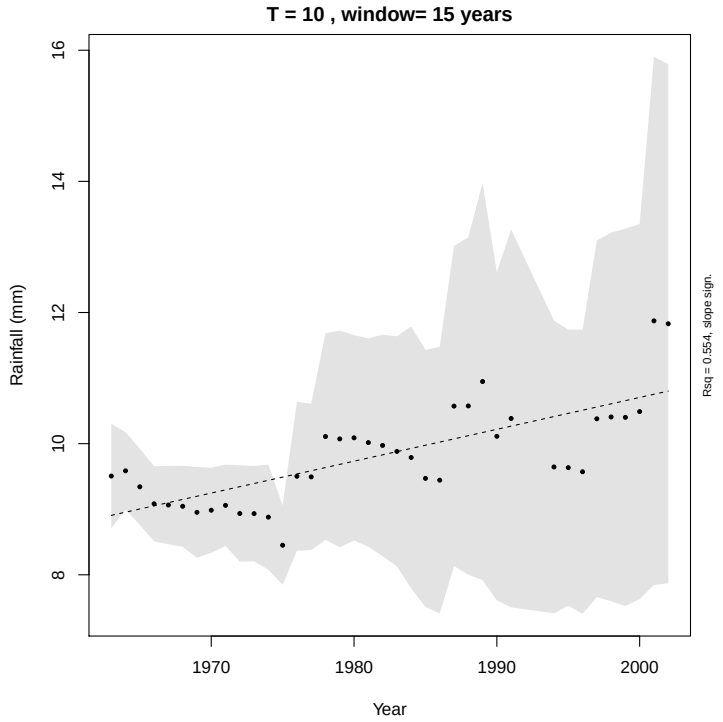


Figure D.10: Return level values the whole range of window and return period for the GPD.

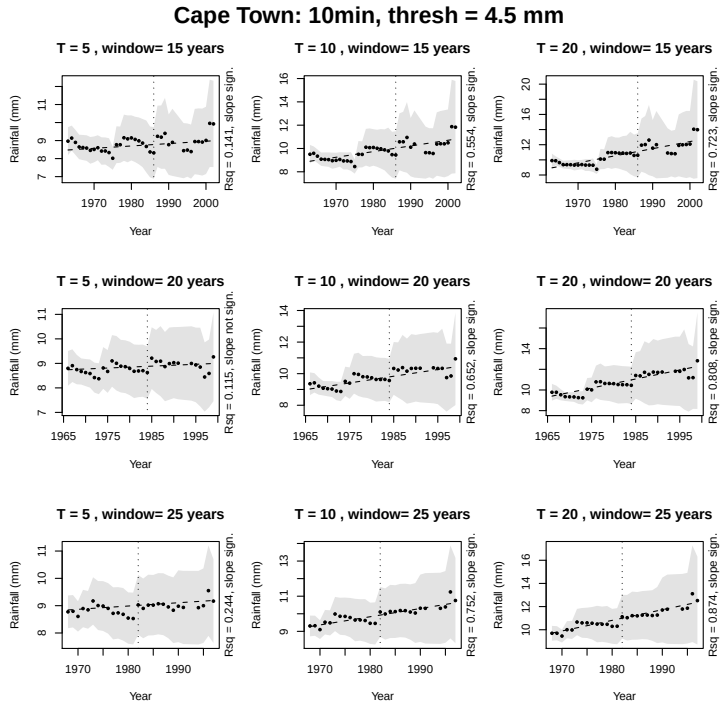


Figure D.11: Return level values the whole range of window and return period for the GPD.

Stellenbosch University http://scholar.sun.ac.za

D.2 Frequency analysis

The frequency analysis entailed the of counting all 5 minute events above a threshold value per year. For this analysis, the 5 minute data set for Cape Town was used that includes all non-zero rainfall events, illustrated in Figure D.12.

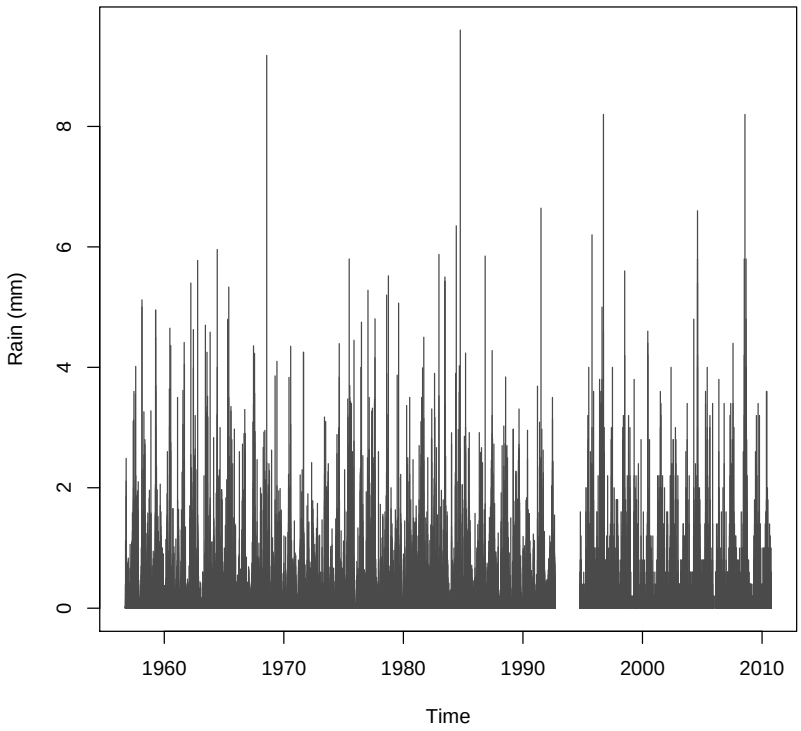


Figure D.12: All 5 minute events for Cape Town.

For the first year (1956), the amount of exceedances of the threshold values 0.5, 1, and 2mm are counted, as illustrated in Figure D.13, showing the 5 minute rainfall events and the threshold values for 1956. 209, 55 and 10 exceedances of the threshold values 0.5, 1 and 2mm respectively, were counted. This process is repeated for all years and the amount of annual exceedances of the threshold is plotted, as illustrated in Figure D.14, where the light grey line represents the annual number of exceedances for a specific threshold. Centred moving average lines were added to detect possible trends. The mean number of exceedances over the entire record is indicated by the dotted line.

Stellenbosch University http://scholar.sun.ac.za

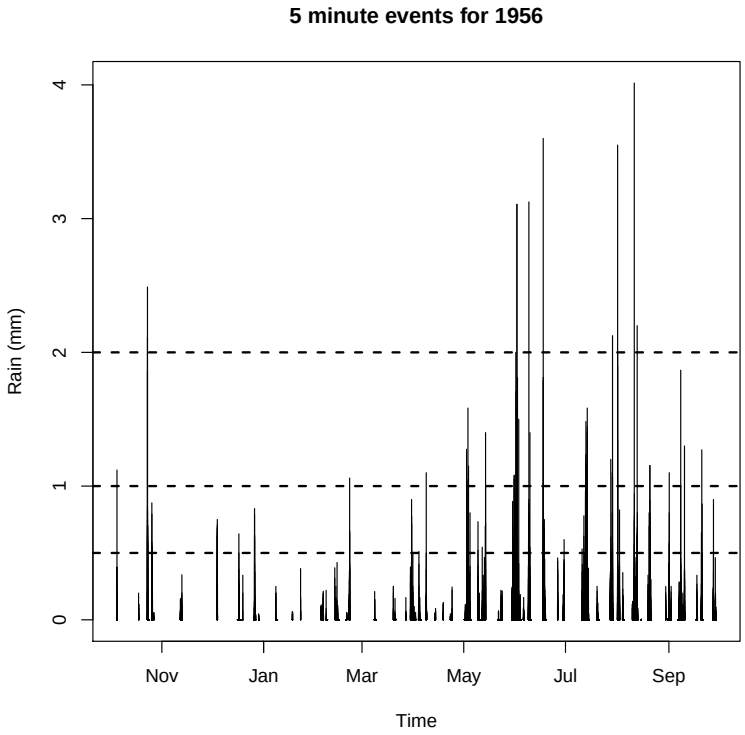


Figure D.13: 5 minute events for the year 1956. Horizontal lines indicate the threshold values.

Stellenbosch University http://scholar.sun.ac.za

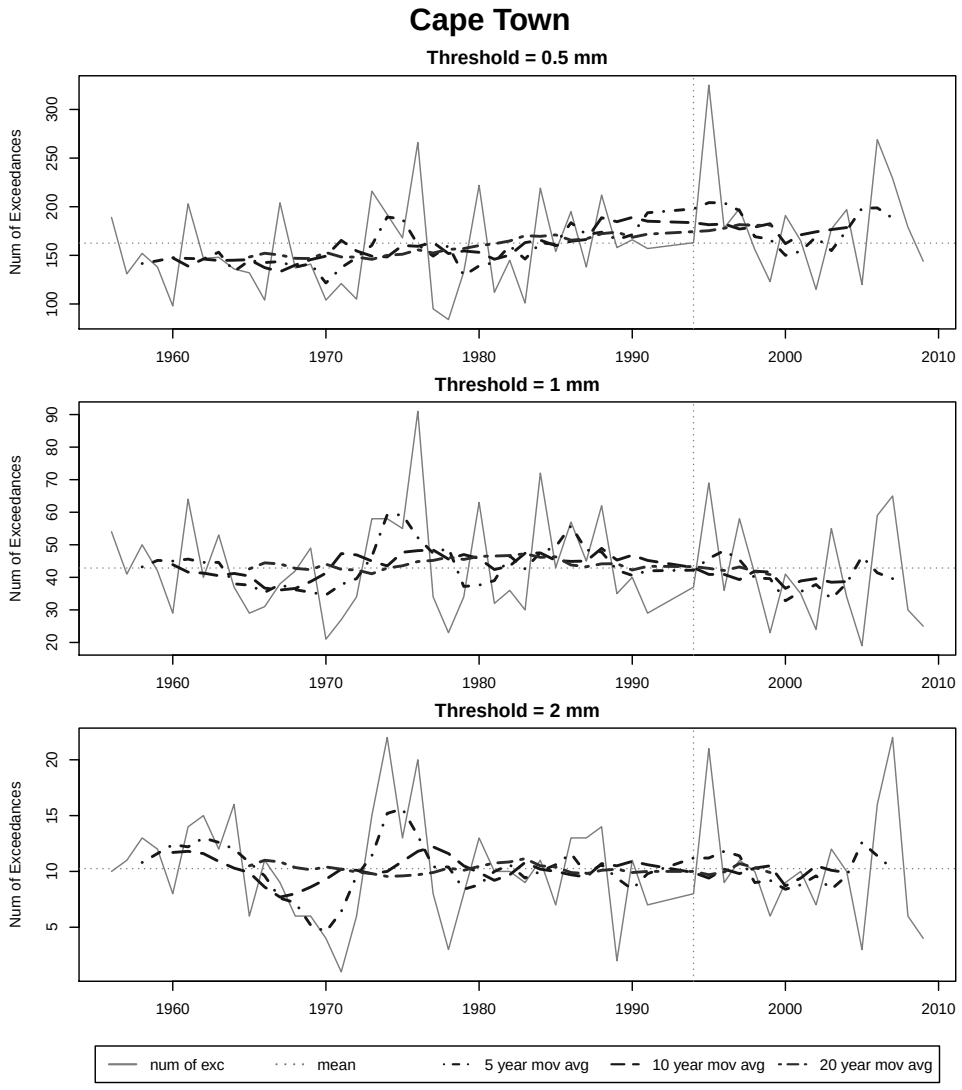


Figure D.14: Annual frequency of exceedance of threshold plot.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Manual for programming procedures

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Contents

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Chapter 1

Introduction

This document describes the automation procedures used in the thesis "The influence of climate change on short duration rainfall" by G.J. Burger (henceforth referred to as Burger (2012)). Automation procedures were applied because of the high resolution of short duration rainfall (SDR) and the large amount of stations that required processing and analysis.

This document is divided into two sections. The *Data Processing* section describes the scripts used to process the digitised autographic and automatic weather station (AWS) SDR data. The *Data Analysis* section describes the scripts used for the analysis of the magnitude and frequency of the SDR. For more information on background theory of the processing and analysis, please consult Burger (2012).

Please note the following:

The *awk*[1] and *R*[2] programming environments were used for the processing and analysis of data. It is assumed that the user is familiar with these programming environments. A vast amount of freely available documents, websites and forums on the mechanics and functionality of the environments are available online.

The development of the scripts and output files were done in a Linux environment. Some of the files may show formatting problems in other environments like Windows. Also note that Linux uses "/" to separate directories, while Windows uses "\". If Linux is not used, please make the appropriate changes to the scripts where necessary.

All required scripts are contained in their correct file locations on the accompanying disc to Burger (2012), under the folder `Scripts`. The following data is provided in the `Data` folder:

[1] *awk* is a simple text editing programming language. For more information, see `http://awk.info` [accessed 2012, September 17]. All script names with the extension `.awk` are associated with the *awk* language.

[2] *R* is an high-level open source statistical programming language based on the *S* language. For more information, see `http://cran.r-project.org` [accessed 2012, September 17]. All script names with the extension `.r` are associated with the *R* language.

Stellenbosch University http://scholar.sun.ac.za

Data type	Folder name
raw autographic	`autogr_raw`
raw AWS	`AWS_raw`
storm duration data for magnitude analysis	`storm_duration`
5 minute data for the frequency analysis	`freq_an`
output results for analyses	`Results`

The user is free to use and customise any of the scripts and to apply it to cases outside the context of the thesis, but does so at his/her own risk. If the user desires to use the scripts, please ensure that the files and scripts, in their respective locations, are copied to read and writeable storage, e.g. a hard drive or flash disk.

If the user wishes to follow the process described in this manual, it is advised to run the script `create.dir.r`, which creates the relevant folders that will be used in the scripts. When opening the file, the user must manually enter the desired working directory in the line marked:

```
# set the home directory:
dir.home <- "[enter working directory here, e.g. C:\My Documents\Analysis]"
```

to the respective working directory of the user, e.g.:

```
# set the home directory:
dir.home        <- "C:\My Documents\Analysis"
```

Please remember to type in the location within the quotation marks. After modifying the `create.dir.r` script, the user must simply run it in the *R* environment, and all the relevant folders will be created. In addition, all *R* scripts contain the same line of the location of the working directory, which must also be manually changed for each script.

Some of the scripts produce very large files, especially for those in the *Data Processing* section. Please ensure that there is enough free space on the storage device to accommodate the generated files.

Both *awk* and *R* scripts generate files with time values in the format: time in seconds since 1970/01/01 00:00:00. The values can either be negative (for values before the 1970/01/01) or positive (for values after the 1970/01/01). This format is used in *R* applications for the analysis of time series data.

Stellenbosch University http://scholar.sun.ac.za

Chapter 2

Data processing

The data processing routine that was followed is summarised in Figure 2.1. The specific procedure and the associated script name is given in each block. Each procedure was executed separately and in the order given in Figure 2.1. Each script will be discussed in the following format:

> **script name**
>
> **Function**
> *Brief description of what the script does*
> **Input**
> *Location of the files the script is programmed to use as input.*
> **Format**
> *File format for input*
> **Sample**
> *A sample of how the input file looks like*
> **Output**
> *Location of the output files*
> **Format**
> *File format for output*
> **Sample**
> *A sample of how the output file looks like*

Please note that the text format comma separated values (CSV) was generally used as an output for text files, while PDFs used for graphics.

Stellenbosch University http://scholar.sun.ac.za

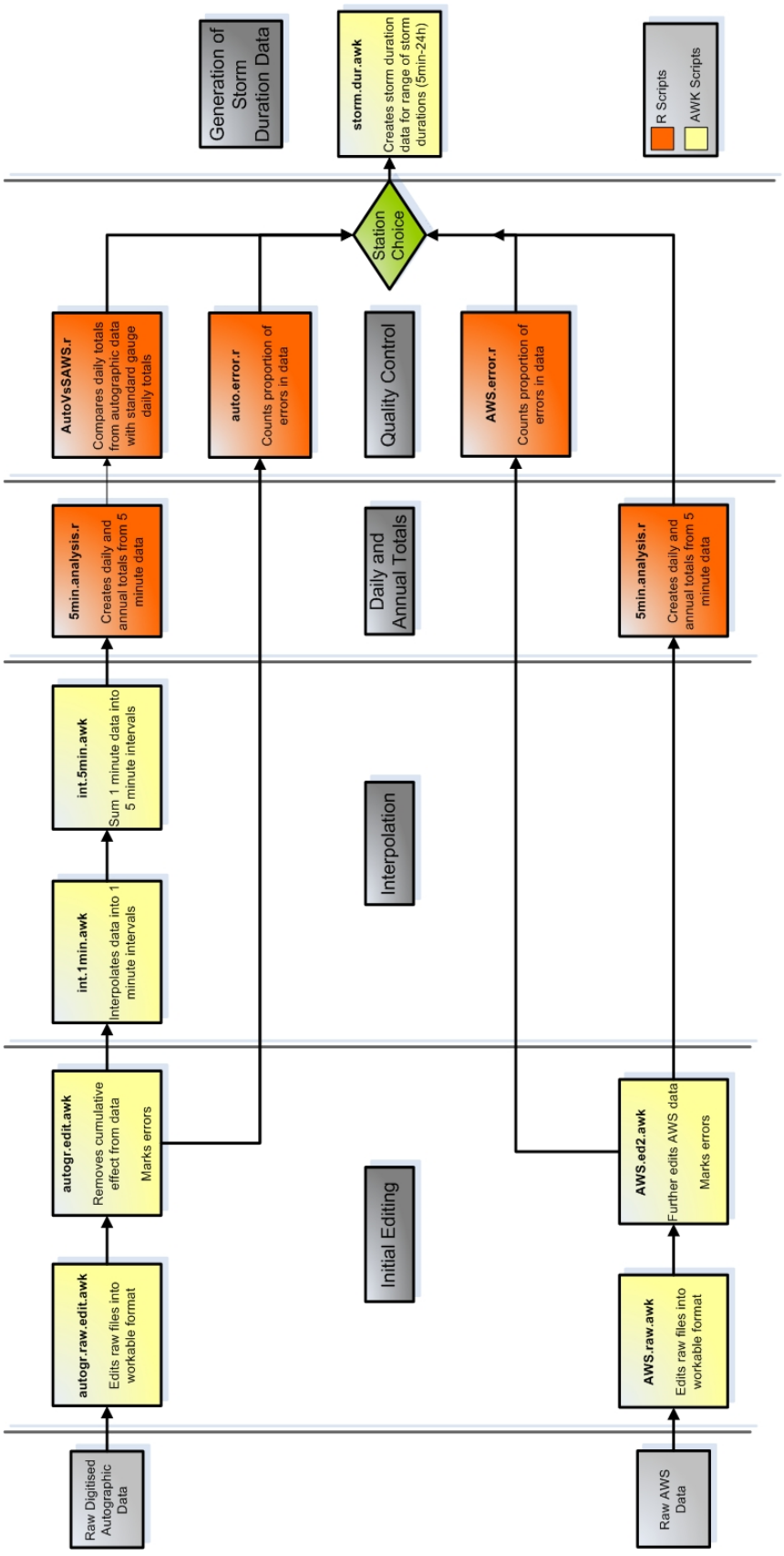


Figure 2.1: Data processing flow scheme

Stellenbosch University http://scholar.sun.ac.za

2.1 auto.raw.edit.awk

Function

Edits raw digitised autographic data into a more workable format.

Input

`.../autogr_raw/[station.number]`

Raw digitised autographic data. It is assumed that the data represents the full record length. Files were originally given in folders per station. Each station folder contained a number of text files that were separated by year.

Format

text: Each line of text represents the autographic readings in a compressed format

Sample

```
00211785607060757   0070408   2070412   3070418   3070443   4070450   7070756   7
00211785607110759   0112127   1112141   2112156   3120112   3120117   5120126  10
00211785607120141  11120151  23120157  26120204  33120210  36120216  40120225  45
00211785607120239  46120246  50120256  51120351  51120356  51120407  52120409  61
```

Output

`.../autogr_ed1/[station.number]`

Workable autographic record.

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), cumulative autographic rainfall reading (mm)

Sample

```
1956,07,06,07,57,00,-425671380,  0.0
1956,07,07,04,08,00,-425598720,  0.2
1956,07,07,04,12,00,-425598480,  0.3
1956,07,07,04,18,00,-425598120,  0.3
```

2.2 auto.edit.awk

Function

Further edits digitised autographic data from `auto.raw.edit.awk`. Removes cumulative effect (up until 10mm) from autographic data. Edits and logs different error types found in record.

Stellenbosch University http://scholar.sun.ac.za

Input

`.../autogr_ed1/[station.number]`

Workable autographic record

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), cumulative autographic rainfall reading (mm)

Sample

```
1956,07,06,07,57,00,-425671380,  0.0
1956,07,07,04,08,00,-425598720,  0.2
1956,07,07,04,12,00,-425598480,  0.3
1956,07,07,04,18,00,-425598120,  0.3
```

Output

`.../autogr_ed2/[station.number]`

Edited autographic record with error flags. Error flags were labelled as:

- **e1** - Digitising errors ("-999"). Usually spanning over a period of time.
- **e2** - Small decrease in the autographic rainfall readings (<0.3mm)
- **e3** - Large decrease in the autographic rainfall readings (>0.3mm). For the detection of false siphon events.

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), rainfall reading (mm), error flag

Error flags are:

Sample

```
1956,07,06,07,57,00,-425671380,0,
1956,07,07,04,08,00,-425598720,0.2,
1956,07,07,04,12,00,-425598480,0.1,
1956,07,07,04,18,00,-425598120,0,
```

2.3 int_1min.awk

Function

Interpolates autographic data from `auto.edit.awk` into 1 minute intervals.

Stellenbosch University http://scholar.sun.ac.za

Input

`.../autogr_ed2/[station.number]`

Workable autographic record

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), rainfall reading (mm), error flag

Sample

```
1956,07,06,07,57,00,-425671380,0,
1956,07,07,04,08,00,-425598720,0.2,
1956,07,07,04,12,00,-425598480,0.1,
1956,07,07,04,18,00,-425598120,0,
```

Output

`.../autogr_int_1min/[station.number]`

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), one minute rainfall reading (mm)

Sample

```
1956,07,06,07,57,00,-425671380,0.000165153,
1956,07,06,07,58,00,-425671320,0.000165153,
1956,07,06,07,59,00,-425671260,0.000165153,
1956,07,06,08,00,00,-425671200,0.000165153,
```

2.4 int_5min.awk

Function

Sums data from int.1min.awk into 5 minute intervals.

Input

`.../autogr_int_1min/[station.number]`

One minute autographic data

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), one minute rainfall reading (mm)

Sample

Stellenbosch University http://scholar.sun.ac.za

```
1956,07,06,07,57,00,-425671380,0.000165153,
1956,07,06,07,58,00,-425671320,0.000165153,
1956,07,06,07,59,00,-425671260,0.000165153,
1956,07,06,08,00,00,-425671200,0.000165153,
```

Output

`... autogr_5min/[station.number]`

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), five minute rainfall reading (mm)

Sample

```
1956,07,06,08,00,00,-425671200,0.000660612,
1956,07,06,08,05,00,-425670900,0.000825765,
1956,07,06,08,10,00,-425670600,0.000825765,
1956,07,06,08,15,00,-425670300,0.000825765,
```

2.5 5min.analysis.r

Function

Creates daily and annual totals from 5 minute data.

Input

Uses either autographic or AWS 5 minute data.

`.../autogr_5min/[station.number]`

or

`.../AWS_ed2/[station.number]`

The user is prompted to choose between autographic or AWS data. Please ensure that the input files are in the correct location.

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), five minute rainfall reading (mm)

Sample

```
1956,07,06,08,00,00,-425671200,0.000660612,
1956,07,06,08,05,00,-425670900,0.000825765,
1956,07,06,08,10,00,-425670600,0.000825765,
```

Stellenbosch University http://scholar.sun.ac.za


```
1956,07,06,08,15,00,-425670300,0.000825765,
```

Output

`.../Results/[rainfall.type]/[station.name]`

`rainfall.type` represents either autographic or AWS data.

Files are organised according to station name, resulting in stations of the same name but different station number placed in the same folder.

A number of files are generated per station. Two folders are generated, `daily` and `annual`, each containing text files of the totals and graphs generated from these totals. Figure 2.2 illustrates the output graphs generated by this script.

Please note that annual totals are in hydrological years (Oct – Sep).

Format

CSV and PDF

Sample

`[station.name]_annual`: year, annual total (mm)

```
1957,418
1958,522
1959,417
1960,407
```

`[station.name]_daily`: year, month, day, daily total (mm)

```
1956,7,6,0.7,
1956,7,7,0,
1956,7,8,0,
1956,7,9,0,
```

2.6 AutoVsSAWS.r

Function

Compares daily and annual statistics generated in `5min.analysis.r` to corresponding standard gauge daily totals (SG).

Input

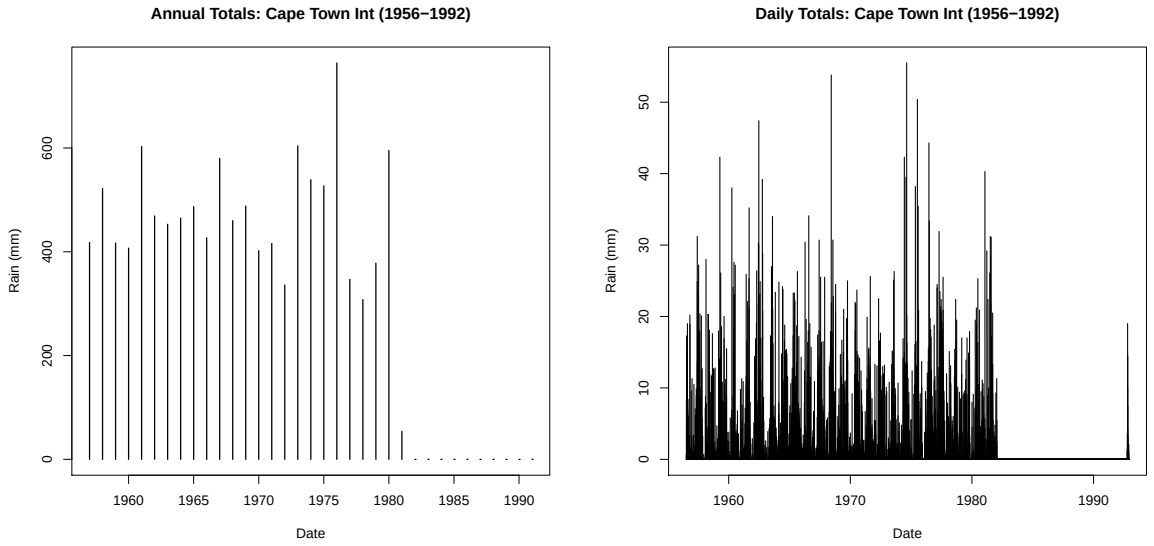
Daily and annual totals of both autographic and SG data.

For autographic data:

`.../Results/autographic/[station.name]`

Which contains both daily and annual rainfall data of the autographic data.

Stellenbosch University http://scholar.sun.ac.za



(a) annual (b) daily

Figure 2.2: Annual and daily totals PDF output

For SG data:

Daily data:

`.../SAWS_daily_ed/[station.number]`

Annual data

`.../SAWS_annual_ed/[station.number]`

Format

CSV

Sample

`[station.name]_annual`: year, annual total (mm)

```
1944,12,31,-788982900,0.0
1945,01,01,-788896500,0.0
1945,01,02,-788810100,0.0
1945,01,03,-788723700,0.0
```

`[station.name]_daily`: year, month, day, daily total (mm)

```
1945,01,15,-787686900,0.3
1945,02,10,-785440500,0.8
1945,02,28,-783885300,0.3
1945,03,01,-783798900,0.5
```

Output

`.../Results/autographic/[station.name]`

All output files are generated in the original input folder using the same output method in `5minute.analysis.r`, whereby files are separated by annual and daily total folders.

Format

CSV

Sample

For annual totals:

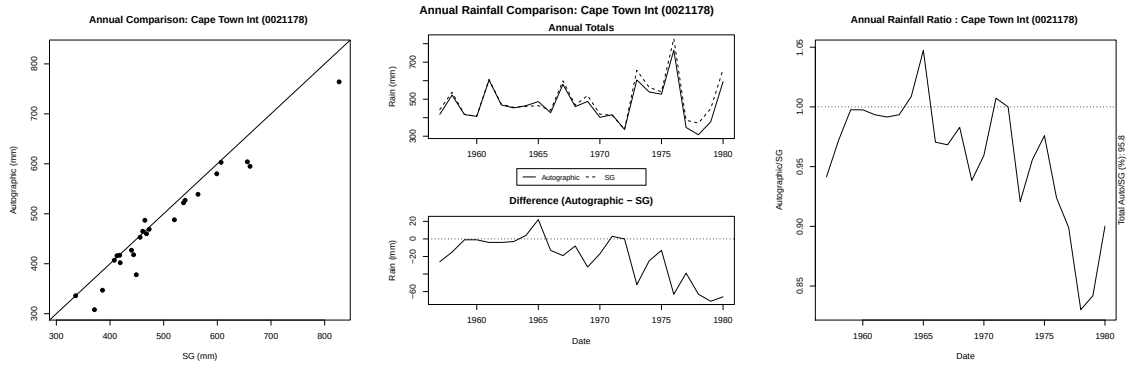
`[station.name]_diff`: year, difference between SDR and SG totals (mm)

```
1957,-26
1958,-15
1959,-1
1960,-1
```

`[station.name]_rat`: year, autographic total (mm), SG total (mm), ratio of SDR/SG

```
year,auto,SG,ratio
1957,418,444,0.941441441441441
1958,522,537,0.972067039106145
1959,417,418,0.997607655502392
1960,407,408,0.997549019607843
```

Figure 2.3 illustrates the output graphics created for the annual totals.



(a) comparison (b) difference (c) ratio

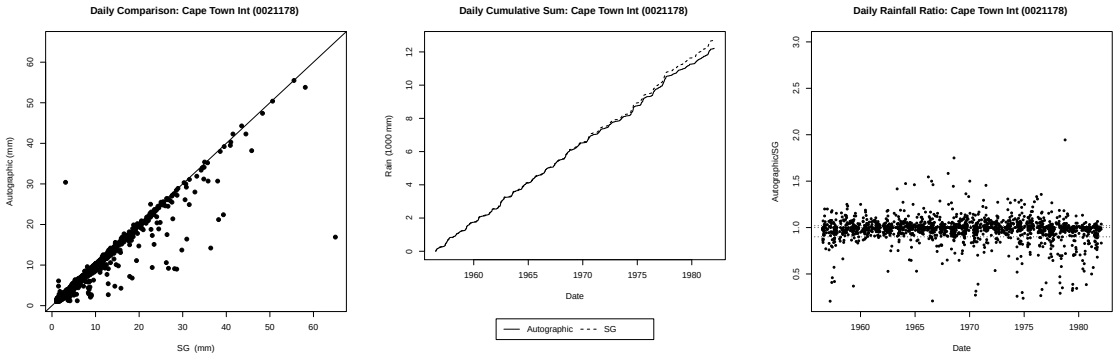
Figure 2.3: Annual totals PDF output

Results for the daily totals:

`[station.name]_rat`: year, month, day, autographic total (mm), SG total (mm), ratio of SDR/SG

```
1956,7,11,8.3,8.4,0.988
1956,7,12,4.8,4.9,0.98
1956,7,15,7.2,7.4,0.973
1956,7,17,17.3,17.5,0.989
```

Stellenbosch University http://scholar.sun.ac.za



(a) comparison (b) cumulative sum (c) ratio

Figure 2.4: SDR comparison with SG for daily total data PDF output graphics

Figure 2.4 illustrates the output graphics created for the annual totals.

Additional packages/functions required

2.7 auto.error.r

Function

Measures the proportion errors in autographic data.

Input

`.../autogr_ed2/[station.name]`

Edited autographic data from `auto.edit.awk`.

Format

CSV: year, month, day, hour, minute, second, time in seconds (since 1970/01/01 00:00:00), rainfall reading (mm), error flag

Sample

```
1956,07,06,07,57,00,-425671380,0,
1956,07,07,04,08,00,-425598720,0.2,
1956,07,07,04,12,00,-425598480,0.1,
1956,07,07,04,18,00,-425598120,0,
```

Output

`.../Results/autographic/[station.name]`

A graph is generated that displays the proportion of different types of errors per year. Three error types are illustrated:

Stellenbosch University http://scholar.sun.ac.za

- **e1** - Digitising errors. Usually spanning over a period of time.
- **e2** - Small decrease in the autographic rainfall readings (<0.3mm)
- **e3** - Large decrease in the autographic rainfall readings (>0.3mm). For the detection of false siphon events.

Format

PDF

Sample

Figure 2.5 illustrated the graphs generated for both **e1**, **e2** and **e3** errors. The top graph illustrates the proportion (in %) of **e2** and **e3** errors in the record per year, while the bottom graph illustrates the amount of time (in days) **e1** errors occupy in the record per year.

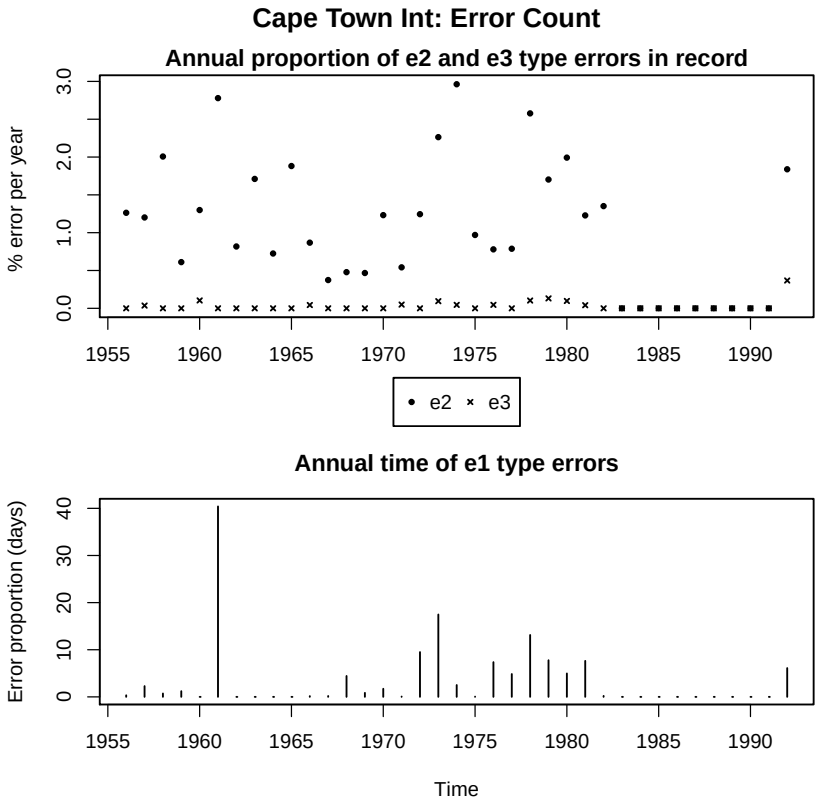


Figure 2.5: Error measuring

2.8 AWS.raw.awk

Function

Edits raw 5 minute AWS data.

Stellenbosch University http://scholar.sun.ac.za

Input

`.../AWS_raw/[station.number]`

Raw digitised autographic data.

Format

It depends, as each request from the SAWS has delivered a different format. Usually tab separated format. Please note that this script will have to be adapted when receiving data in a different format and order.

Sample

```
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:05:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:10:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:15:00 0.0
"0021178A3" "CAPE TOWN WO" 10/19/1994 02:20:00 0.0
```

Output

`.../AWS_ed/[station.number]`

Format

CSV: Output depends on the original format of the data. For this example: month, day, year, hour, minute, second, five minte rainfall (mm)

Sample

```
10,19,1994,02,05,00,0.0
10,19,1994,02,10,00,0.0
10,19,1994,02,15,00,0.0
10,19,1994,02,20,00,0.0
```

2.9 AWS.ed2.awk

Function

Further edits AWS data generated from `AWS.raw.awk` and marks gaps and errors.

Input

`.../AWS_ed/[station.number]`

Format

CSV: Output depends on the original format of the data. For this example: month, day, year, hour, minute, second, five minte rainfall (mm)

Sample

Stellenbosch University http://scholar.sun.ac.za


```
10,19,1994,02,05,00,0.0
10,19,1994,02,10,00,0.0
10,19,1994,02,15,00,0.0
10,19,1994,02,20,00,0.0
```

Output

Two files are generated, one containing the edited AWS data, and another for recording errors in the data. For the error reading file, gaps are recorded as `diff`, and no readings are recorded as `NA`.

Edited AWS data:

`.../AWS_ed2/[station.number]`

Error count file:

`.../error_message/AWS_data/[station.number]`

Format

CSV

Sample

`[station.name]`: year, month, day, hour, minute, seconds, time in seconds (since 1970/01/01 00:00:00), rainfall (mm)

```
1994,10,19,02,05,00,782525100,0.0
1994,10,19,02,10,00,782525400,0.0
1994,10,19,02,15,00,782525700,0.0
1994,10,19,02,20,00,782526000,0.0
```

`[station.name]`: year, month, day, hour, minute, seconds, time in seconds (since 1970/01/01 00:00:00), error flag

```
1994,12,31,02,05,00,788832300,diff
1994,12,31,02,10,00,788832600,diff
1994,12,31,02,15,00,788832900,diff
1994,12,31,02,20,00,788833200,diff
```

2.10 AWS.error.r

Function

Measures the proportion errors in autographic data

Input

`.../error_message/AWS_data/`

Error logged data from `AWS.ed2.awk`.

Stellenbosch University http://scholar.sun.ac.za

Format

CSV: year, month, day, hour, minute, seconds, time in seconds (since 1970/01/01 00:00:00), error flag

Sample

```
1994,12,31,02,05,00,788832300,diff
1994,12,31,02,10,00,788832600,diff
1994,12,31,02,15,00,788832900,diff
1994,12,31,02,20,00,788833200,diff
```

Output

`.../Results/AWS/[station.name]`

A graph is generated that displays the proportion of different types of errors per year.

Format

PDF

Sample

Figure 2.6 illustrates the output graphic. The proportion of the errors are measured by the amount of days error readings occupied each year or the record.

2.11 storm.dur.awk

Function

Creates storm duration files from 5 minute data.

Input

`.../station_analysis/[station.name]`

After choice of stations has been made, a new file is created per station that combines the autographic and AWS data. Please note that these new files were manually combined and edited to hydraulic years.

Format

CSV: year, month, day, hour, minute, seconds, time in seconds (since 1970/01/01 00:00:00), rainfall (mm)

Sample

```
1994,10,19,02,05,00,782525100,0.0
1994,10,19,02,10,00,782525400,0.0
1994,10,19,02,15,00,782525700,0.0
1994,10,19,02,20,00,782526000,0.0
```

Stellenbosch University http://scholar.sun.ac.za

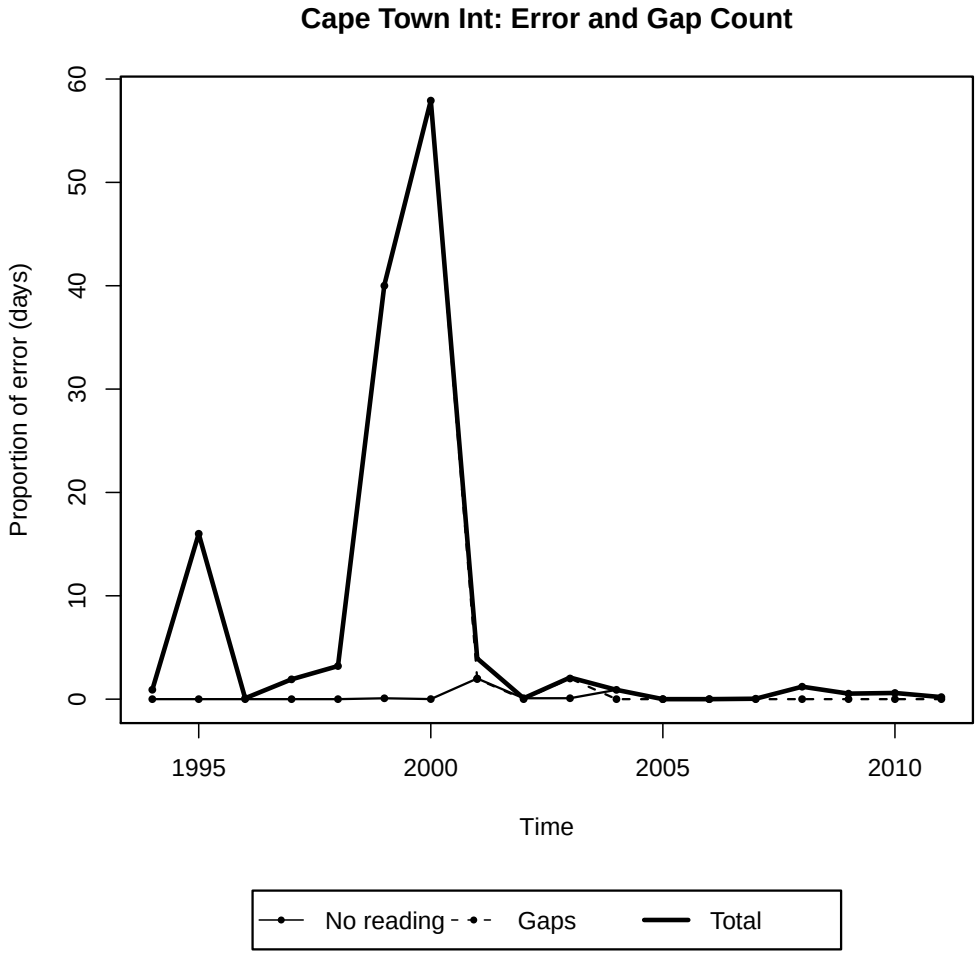


Figure 2.6: Error measuring

Output

Two outputs types are generated. One is storm durations selected according to a 24h hour of zero rainfall between peak events for the purpose of extreme value theory (EVT) analysis; the other records all events per storm duration for the purpose of frequency analysis.

Thirteen storm durations are generated:

Time	minutes							hours					
Amount	5	10	15	30	45	60	90	2	4	8	12	18	24

For EVT analysis:

`.../storm_duration/[station.name]/[duration]`

For frequency analysis:

`.../freq_an/[station.name]/`

Stellenbosch University http://scholar.sun.ac.za

Format

CSV

Sample

Peak data for EVT analysis:

`.../[station.name]/[duration]`: time in seconds since 1970/01/01 00:00:00, storm duration rainfall (mm)

```
-417830100,1.12
-416698500,0.115385
-416250900,2.48889
-416024100,0.875
```

5 minute data for frequency analysis:

`.../[station.name]/[duration]`: time in seconds since 1970/01/01 00:00:00, five minute rainfall (mm)

```
-417835500,0.0142857
-417835200,0.0357143
-417834900,0.0357143
-417834600,0.0342857
```

Stellenbosch University http://scholar.sun.ac.za

Chapter 3

Data analysis

The data analysis section is divided into two parts: magnitude and frequency analysis. The magnitude analysis covers the non-stationary extreme value theory (EVT) analysis. The frequency analysis covers the frequency of exceedance method adopted and modified from Van Wageningen (2006:72-77). Please note that both analyses were heavily customised for the seven stations in Burger (2012). If the user wishes to analyse different stations, appropriate changes to script details need to be made where appropriate.

The *R* packages `ismev` and `evir` are required for the data analysis section. Please ensure that these are installed before using the scripts in this section.

3.1 Magnitude analysis

The magnitude analysis consists of the application of both the parametric non-stationary (PNS) and non-parametric non-stationary methods (NPNS). A single script `EVT.Analysis.r` is used for the most of the magnitude analysis. A number of scripts were developed for functions used within `EVT.Analysis.r`, specifically for PNS and NPNS methods. The most fundamental of these will be discussed in greater detail.

3.1.1 EVT.Analysis.r

Function

Applies PNS and NPNS analysis to storm duration generated data.

Input

`.../storm_duration/[station.name]/[duration]`

Peak storm duration data from Section 2.11.

Stellenbosch University http://scholar.sun.ac.za

The user is prompted to choose what type of distribution should be analysed. Either a generalised extreme value distribution (GEV) or generalised Pareto distribution (GPD) can be selected. If a GPD is selected, the user must make a choice between three threshold selection methods: du Plessis values, 1.5 $\times$ du Plessis values or the stable shape parameter method. The du Plessis values are given in Table 3.1.

Table 3.1: Thresholds based on Du Plessis (1992) cut-offs. Values in millimetres.

Duration	minutes							hours					
	5	10	15	30	45	60	90	2	4	8	12	18	24
Du Plessis (1992)	2	3	4	5	6	8	10	12	12	12	12	12	12
1.5 $\times$ Du Plessis (1992)	3	4.5	6	7.5	9	12	15	18	18	18	18	18	18

Format

CSV: time in seconds since 1970/01/01 00:00:00, storm duration rainfall (mm)

Sample

```
-417830100,1.12
-416698500,0.115385
-416250900,2.48889
-416024100,0.875
```

Output

General output location:

`.../Results/Data.Analysis/[distribution.type]`

For a GEV:

`...GEV/AMS/[station.name]/[storm.duration]`

Annual maximum values (AMS), used in the GEV, are generated in a CSV file and plotted graphically (PDF) for each storm duration. Figure 3.1 illustrates a AMS plot for a specific duration. Please note that hydraulic years (October to September) were used.

Sample

year, annual maximum rain (mm)

```
1956,18.1769
1957,18.2234
1958,39.35
1959,23.2
```

`...GEV/NPNS/[station.name]/[storm.duration]`

Figure 3.2 illustrates the graphic (PDF) generated for a specific storm duration.

`...GEV/NPNS/GEVnpns`

Stellenbosch University http://scholar.sun.ac.za

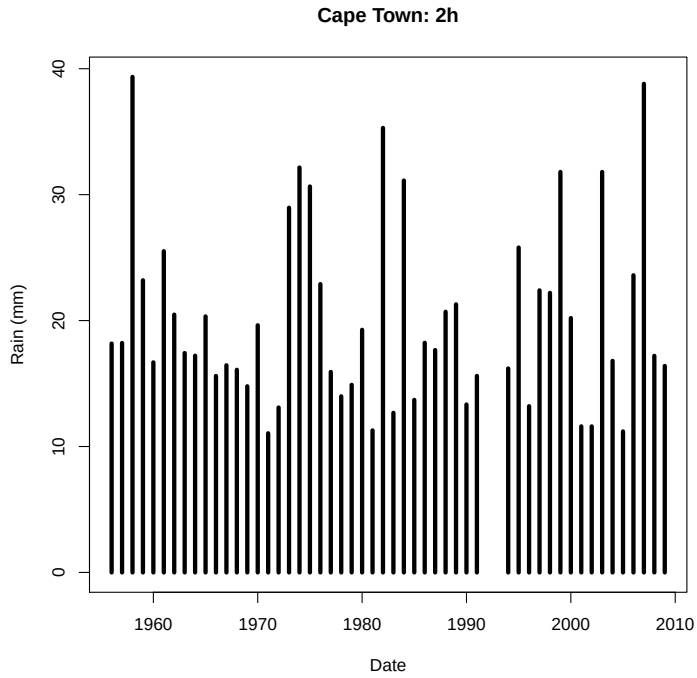


Figure 3.1: AMS graphical plot.

A summary of all the significant regression slopes per distribution for all stations in the file `GEVnpns`.

Sample

duration, total significant slopes, total positive slopes, total negative slopes

```
Cape Town
dur,sig,pos,neg
5min,7,4,3
10min,8,6,2
15min,6,6,0
```

`...GEV/PNS/[station.name]/GEV`

Summaries where significant PNS models were obtained for all distributions and all stations.

Sample

```
Dur,5min,10min,15min,30min,45min...
Cape Town,GUM-S,GEV-NS,GUM-S,GEV-S,GEV-S,...
East London,GEV-NS,GUM-S,GUM-S,GUM-S,GUM-S,...
```

Significant stations are analysed with a separate script. See Section 3.1.2 for more information.

For a GPD:

The structure is mostly similar to the GEV, but separate folders were generated for each of the three threshold selection method, labelled as:

Stellenbosch University http://scholar.sun.ac.za

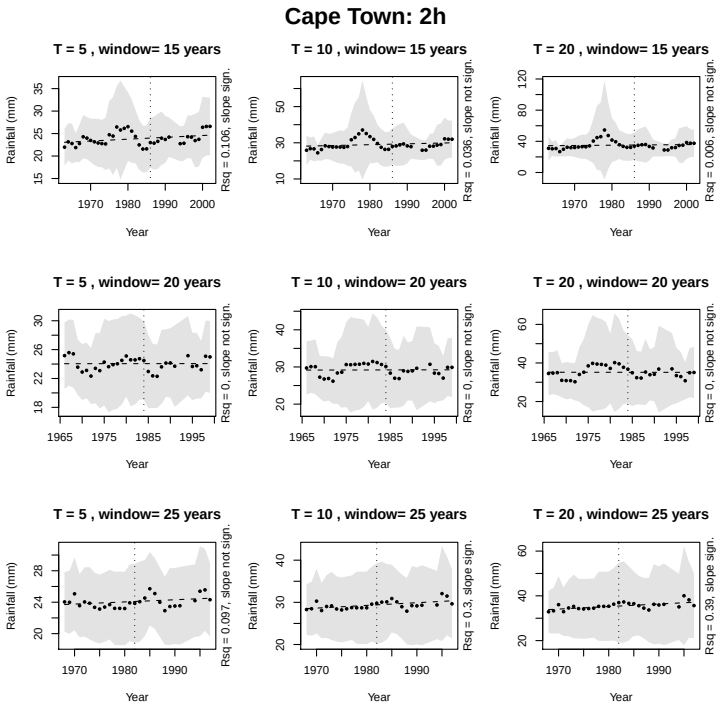


Figure 3.2: NPNS graphical plot.

- `1` - du Plessis values
- `2` - 1.5 $\times$ du Plessis values
- `3` - stable shape parameter method

Therefore, output locations are typically:

`...GPD/[threshold.type]/[nonstationary.method]`

For each threshold selection method, the output structure is exactly the same the GEV.

3.1.2 PNS

This section will discuss the PNS functions used in `EVT.Analysis.r` for the GEV and GPD and the `GEVpns.individual.r` script, specifically designed for significant PNS fits to the data. As mentioned in Section 4.3.1.1 of Burger (2012), the PNS analysis produced no significant results, therefore the further analysis of the significant PNS results were only developed for the GEV.

Stellenbosch University http://scholar.sun.ac.za

3.1.2.1 GEVpns.r

The `GEVpns.r` script was used to analyse AMS values for a specific duration. The function `GEVpns` requires both the AMS values and the corresponding time occurrence as input. An additional input, `alpha`, is the significance parameter of the t-test and maximum likelihood test. By default, it is set to $\alpha = 0.95$. The function makes use of the `ismev` package in *R*.

The function fits a linear non-stationary model

$$\text{par}(t) = \beta_0 + \beta_1 t$$

to each parameter in the GEV in the order ξ, then σ, and lastly μ. Once the linear non-stationary model is fitted to a parameter, the significance of the slope (β_1) is measured with a t-test[1] for α level of significance, similar to significance calculations for regression techniques. If the slope is significant, then the next parameter is tested along with the previous parameter. If not, then only the next parameter is tested. A total of $2^3 = 8$ combinations of linear models are possible for the GEV.

When all the parameters are tested, if at least one parameter had a significant linear model, then a maximum likelihood test is applied to the non-stationary model and compared to a stationary model. The maximum likelihood test is defined as ((Coles, 2001:109)[2]):

$$D = -2\{l_1(M_1) - l_0(M_0)\} \tag{3.1}$$

where

- D is the deviance statistic
- M_1 and M_0 are the more and less complex model respectively
- l_1 and l_0 are the log-maximum likelihoods for each model

The more complex model, M_1, is accepted if $D > \chi^2_{\alpha,k}$, where $\chi^2_{\alpha,k}$ is the α quantile for the chi-square distribution, where k is the difference between the amount of parameters of the models.

If the non-stationary GEV model is rejected, then the confidence interval of the stationary GEV model's shape parameter, ξ, is checked if it includes 0. If so, then a Gumbel model is justified, which is a special case of the GEV, where $\xi = 0$. A new function, `GUMpns`, in the `GUMpns.r` script, is called. This function is exactly the same as the `GEVpns` function, except that a Gumbel distribution is applied to the data.

[1] See Montgomery and Runger (2007:405-407) for more information.
[2] Please note that the definition in Coles (2001) is missing the $-$ sign in front of the 2

Stellenbosch University http://scholar.sun.ac.za

Any non-stationary cases for either the GEV or Gumbel model are logged. A successful non-stationary model will be logged as either "GEV-NS" or "GUM-NS" in the file:

`...GEV/PNS/[station.name]/GEV`

3.1.2.2 GPDpns.r

The `GPDpns.r` script and `GPDpns` function are nearly identical to the GEV counterpart, except that peaks over threshold (POT) data is used instead of AMS values. Only the shape and scale (ξ, σ) are tested as the threshold is already provided. A total of $2^2 = 4$ combinations of linear models are possible for the GPD.

3.1.2.3 GEVpns.individual.r

The `GEVpns.individual.r` script was developed to individually analyse any significant PNS storm durations per stations. Since only the GEV provided significant PNS results, no script was developed for the GPD.

Output is located in:

`.../Results/Data.Analysis/[distribution.type]/PNS/[station.name]`

The PNS output is dependent on the amount of significant non-stationary fits each station gives.

Two figures are generated per storm duration. Figure 3.3 illustrates the probability-probability (PP) and quantile-quantile (QQ) plot for a specific duration, while Figure 3.4 illustrates a time series plot of AMS values and non-stationary return levels. The PP and QQ visually indicate the goodness of fit on the PNS model to the data. The time series plot indicates the influence of the PNS model on return levels over time.

An additional text file is generated giving each significant distribution parameter's slope and intersection values for a linear model.

Sample

paramter, parameter value/slope value, intercept value

```
45min
par,b,a
mmu,14.4834151889067,NA
msig,4.14222178307322,NA
msh,0.0308388030718840,-0.723655916320808
```

Where `45min` is the significant duration, `par` is the parameter type, `b` is the slope value, `a` is the intercept value. In the case of a stationary parameter, the value, `b` is the value of the stationary parameter, while `a` is recorded as `NA`. The characters `mmu`, `msig` and `msh` refer to the location (μ), scale (σ) and shape (ξ) parameters respectively.

Stellenbosch University http://scholar.sun.ac.za

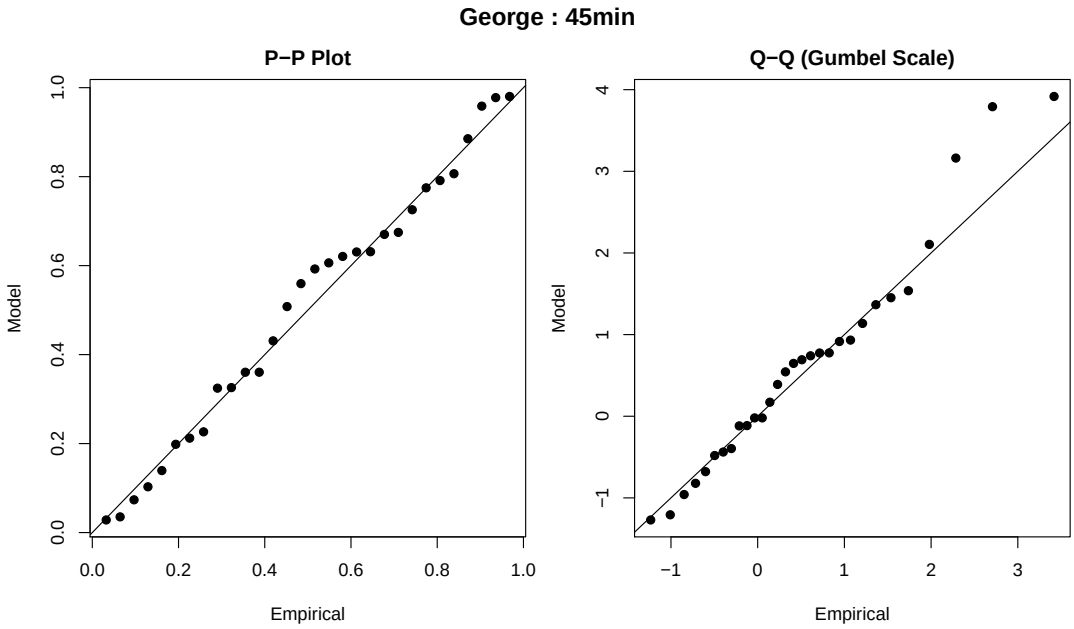


Figure 3.3: NPNS graphical plot.

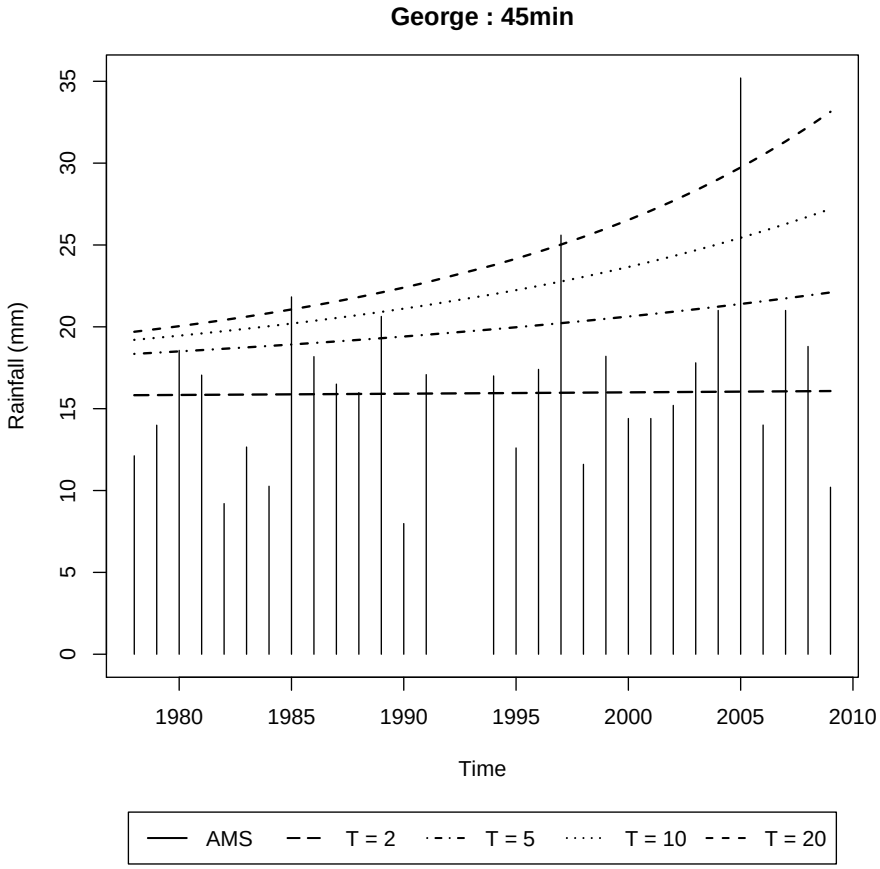


Figure 3.4: NPNS graphical plot.

Stellenbosch University http://scholar.sun.ac.za

3.1.3 NPNS

This section will discuss the NPNS functions used in `EVT.Analysis.r` for the GEV and GPD. The `GEVnpns` and `GPDnpns` functions were used to apply a NPNS test to a single storm duration for a specific station. These functions are contained in the the `GEVpns.r` and `GPDpns.r` scripts. The two functions are very similar, the only difference being that the `GEVnpns` function uses AMS and corresponding time occurrence values as input, while the GPD uses POT and corresponding time occurrence values. In addition, the window and return period(s) are supplied as well, both can either be a single value or a range of values. In Burger (2012), the window period was set to 15, 20 and 25 years, and the return period to 5, 10 and 20 years.

The procedure is as follows: for a specific return period and return period:

- fit a distribution over values within the moving window period and obtain the corresponding return levels for a specific return period.
- apply linear regression to the the return levels, apply a t-test to the slope of the line and record significant slopes and the corresponding sign of the slope.

This procedure is followed for the whole range of window and return periods.

3.2 Frequency analysis

The frequency analysis adopted a method similar to that in van Wageningen, where the frequency of occurrence of rainfall events were measured per year and graphically illustrated. This approach was modified to avoid discrepancies between the autographic and AWS data. In Burger (2012), the frequency of occurrence of exceedances of threshold values were used instead. A single script, `FREQ.Analysis.r`, was mainly used for the frequency analysis.

For a 5 minute SDR dataset:

- count the number of exceedances of a threshold value for each year and plot the annual exceedances graphically
- fit moving averages to range of annual exceedances

The range of threshold and moving average values can be changed, but in Burger (2012), the thresholds were set at 0.5, 1 and 2mm, and the moving averages set to 5, 10 and 20 years.

Input

`.../freq_an/[station.name]/5`

Stellenbosch University http://scholar.sun.ac.za

Format

CSV

Sample

Number of seconds since 1970/01/01 00:00:00, rainfall value (mm)

```
-417835500,0.0142857
-417835200,0.0357143
-417834900,0.0357143
-417834600,0.0342857
```

Output

`.../Results/EV_Analysis/exceedance/[station.name]`

Figure 3.5 illustrates a typical output file for a specific station.

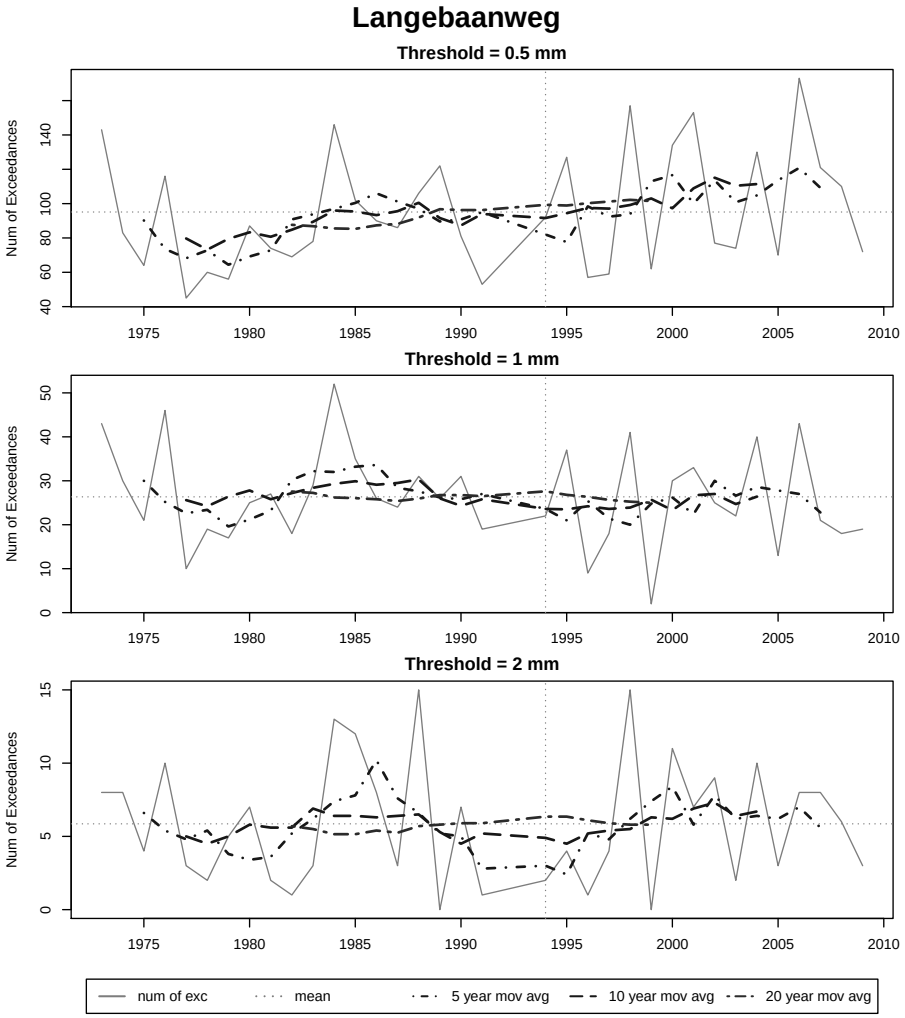


Figure 3.5: Frequency of exceedance of threshold plot.

Stellenbosch University http://scholar.sun.ac.za

Stellenbosch University http://scholar.sun.ac.za

Bibliography

Burger, G. J. 2012. The influence of climate change on short duration rainfall. Master's thesis. Stellenbosch: Stellenbosch University.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics: Springer.

Du Plessis, J. 1992. *Rainfall intensity-duration-frequency-curves for the Cape Peninsula*. Cape Town: City of Cape Town.

Montgomery, D. & Runger, G. 2007. *Applied statistics and probability for engineers*. edition no. 4. New Jersey: John Wiley & Sons, Inc.

Van Wageningen, A. 2006. The impact of climate change on hydrological predictions, with specific reference to 24-hour rainfall intensities in the Western Cape. Master's thesis. Stellenbosch: Stellenbosch University.