ASSESSING THE PERFORMANCE OF REGIONAL FLOOD FREQUENCY ANALYSIS METHODS IN SOUTH AFRICA

JERMAINE JONATHAN NATHANAEL

Submitted in partial fulfilment of the requirements
for the degree of MSc in Hydrology

Centre for Water Resources Research
University of KwaZulu-Natal
Pietermaritzburg
South Africa

December 2015

ABSTRACT

In engineering and flood hydrology, the estimation of a design flood refers to procedures whereby the magnitude of a flood is associated with a level of risk at a given site (Pegram and Parak, 2004). The use of a Regional Flood Frequency Analysis (RFFA) approach improves the accuracy and reliability of estimates of design floods. However, no RFFA method is currently widely used in South Africa, despite a number of RFFA studies having been undertaken, that include South Africa. Hence, the performance of the current RFFA approaches needs to be assessed in order to determine the best approaches to use and to determine if a new RFFA approach needs to be developed for use in South Africa. Through a review of the relevant literature it was found that the Meigh *et al.* (1997) Method, the Mkhandi *et al.* (2000) Method, the Görgens (2007) Joint Peak-Volume (JPV) Method, which uses a K-Region regionalisation, as well as a Veld zone regionalisation, and the Haile (2011) Method are most suitable for application in a nationwide study. Each regional approach was assessed by comparing their design flood estimates with those estimated from an at-site flood frequency analysis of the observed flood data, using both the General Extreme Value (GEV) and Log Pearson Type 3 (LP3) distributions. However, due to the LP3 distribution producing inconsistent design flood estimates, it was removed from further analysis and only the GEV distribution was assessed. Annual Maximum Flood (AMF) data were obtained from the Department of Water and Sanitation (DWS) for 1458 stations across the entire country. In addition to these datasets, 89 synthesised dam inflow records were obtained from the DWS and incorporated into the study. Due to a thorough data screening process, the final number of stations and dam inflow records analysed was reduced to 407 stations. In order to determine the overall accuracy of the RFFA methods, Relative Errors (*RE*) (%) were calculated at each station. Box plots and frequency plots were utilised to represent the distribution of relative errors and the degree of bias was measured using a ratio of the estimated and observed design floods. The results of the study show that the Haile Method generally performs better than the other RFFA methods, however it also consistently under-estimates. The Mkhandi Method generally over-estimates. The Meigh Method generally performs the worst, consistently over-estimating. For the JPV Methods, the K-Region regionalisation generally performs better than the Veld zone regionalisation; however, they both consistently over-estimate design floods. The poor overall performance of the RFFA methods are due to a number of reasons. In the case of the Mkhandi *et al*. (2000) Method, the

tests for homogeneity that were developed were too lenient, which may have incorrectly defined regions as being homogeneous. In the case of the Meigh *et al.* (1997) Method, the regionalisation of homogeneous flood regions were too broad, where only two flood regions have been identified for South Africa. For the Haile (2011) Method, the logarithmic regressions developed for a number of regions were not able to determine index floods for all catchment areas. Therefore, power regressions were developed in this study. In the case of the JPV Methods, the Kovacs K-Regions and Veld zone regions were used, which have not been updated in the past several years. In response to the generally poor performance of the RFFA methods assessed in this study, it has been recommended that a new method be developed for application in design flood practice in South Africa.

DECLARATION

I, Jermaine Nathanael, declare that:

(i) The research reported in this dissertation, except where otherwise indicated, is my original work.

(ii) This dissertation has not been submitted for any degree or examination at any other university.

(iii) This dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv) This dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

(a) their words have been re-written, but the general information attributed to them has been referenced;

(b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated, in detail, which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi) This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the Dissertation and in the References sections.

Signed: ………………..

Jermaine Nathanael

PREFACE

The work described in this dissertation was carried out in the Centre for Water Resources Research, School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Professor JC Smithers and Mr MJC Horan.

The research represents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where, use has been made of the work of others, it is duly acknowledged in the text.

Signed: …………………………..
Professor JC Smithers
Date:

Signed: …………………………
Mr MJC Horan
Date:

ACKNOWLEDGEMENTS

The author would like to thank the following people for their contributions to this study:

- Professor JC Smithers, supervisor of this dissertation, for the opportunity to undertake this study and for your support, patience and guidance. You have indeed been a great mentor to me and I will carry both the skills and work ethic I have learned from you into my future endeavours.
- Mr MJC Horan, co-supervisor of this dissertation, for the countless hours spent tackling ArcGIS problems that were beyond my abilities and for providing intelligent advice and solutions to problems discussed at project meetings.
- The National Research Foundation, for providing the project funds and to the Department of Water and Sanitation, for providing a bursary throughout the duration of this study.
- Mrs Rees, for editing the document so timeously and providing adequate information and guidance to correct my mistakes. It has always been a delight engaging you when I needed assistance.
- Mr D Clark, for providing the Python scripts that were utilised in obtaining catchment area shapefiles for the study.
- TRJ Kjeldsen, for running the R statistical software to validate results in this study.
- My parents, Jonathan and Rebekah Nathanael, for putting me through university and for their support and encouragement throughout the duration of this project and to my sisters, Ronell and Adelle for their motivation and interest in my studies.
- I would also like to thank Dalhia for her support, encouragement and interest in the progress of my study.
- My friends and colleagues in the Centre for Water Resources Research, thank you for all of the time spent conversing about our projects and helping each other out throughout the past two years.
- Finally, I thank God for giving me the strength and the ability to accomplish my academic goals. In times when I have been overwhelmed I have taken solace in your word which says: “I can do all things through Christ who strengthens me” (Philippians 4:13, NKJV). “For with God, nothing will be impossible” (Luke 1:37, NKJV).

TABLE OF CONTENTS

Page

ABSTRACT.......... ii

DECLARATION.......... iv

PREFACE.......... v

ACKNOWLEDGEMENTS.......... vi

LIST OF TABLES.......... xii

LIST OF FIGURES.......... xiii

LIST OF SYMBOLS.......... xvii

LIST OF ABBREVIATIONS.......... xix

1. INTRODUCTION.......... 1

1.1 Aim and Objectives.......... 3

1.2 Research Questions.......... 3

1.3 Delineation and Limitations.......... 3

1.4 Dissertation Structure.......... 4

2. LITERATURE REVIEW.......... 5

2.1 A Review of International Approaches to Regional Flood Frequency Analysis.......... 7

2.1.1 France.......... 7

2.1.2 Australia.......... 8

2.1.3 Iceland.......... 10

2.1.4 United Kingdom.......... 10

2.1.5 Italy.......... 12

2.2 Review of Regional Flood Frequency Analysis Studies in South Africa 13

2.2.1 Van Bladeren Method 13

2.2.2 Meigh Method 16

2.2.2.1 Estimation of index floods 16

2.2.2.2 Development of regional growth curves 17

2.2.2.3 Estimation of design flood peaks 18

2.2.3 Mkhandi Method 18

2.2.3.1 Delineation of homogeneous regions 19

2.2.3.2 Identification of regional distributions 20

2.2.3.3 Estimating the index flood and design flood discharges 21

2.2.4 Kjeldsen Method 22

2.2.4.1 Identification of homogeneous regions 23

2.2.4.2 Clustering of stations 23

2.2.4.3 Regional flood frequency distribution..24

2.2.4.4 Estimating the index flood 25

2.2.5 Joint Peak-Volume Method 26

2.2.5.1 Quantifying flood volume based on flood peak 27

2.2.5.2 Pooling data 28

2.2.5.3 Estimating the index flood 29

2.2.5.4 Estimation of design flood peaks 30

2.2.6 Haile Method 31

2.2.6.1 Developing homogeneous regions 31

2.2.6.2 Fitting regional data to the appropriate frequency distribution........32

2.2.6.3 Developing regional growth curves 33

2.2.6.4 Estimating the index flood 33

2.3 Synthesis of Literature 35

3. STUDY AREA 39

3.1 General Description 39

3.2 Climate and Hydrology...40

4. METHODOLOGY AND DATA COLLATION 41

4.1 Station Selection 41

4.1.1 Record length..42

4.1.2 Rating table exceedence 43

4.1.3 Influence of dams 44

4.1.4 Additional exclusions and final stations 45

4.2 Catchment Parameters...48

4.2.1 Catchment area 49

4.2.2 Mean Annual Runoff (MAR 90) 54

4.2.3 Average main channel slope..58

4.2.4 HRU Veld zones...66

4.2.5 Kovacs K-Regions..68

4.3 New Regression Equations for the Haile Method 70

5. RESULTS 77

5.1 Selection of Probability Distributions 77

5.2 Performance of Methods 80

5.2.1 Slopes of observed versus estimated design floods 80

5.2.2 Relative errors 83

5.2.2.1 Box plots of relative errors 84

5.2.2.2 Frequency plots of absolute relative errors 89

5.2.3 Ratio of the estimated and observed design floods 91

5.2.4 Concluding remarks 95

5.3 Trends in Method Performance 96

5.3.1 Catchment areas 96

5.3.2 Spatial variations 98

6. DISCUSSION AND CONCLUSIONS 100

6.1 Data Issues 100

6.2 Performance of RFFA methods 102

6.3 Concluding Remarks 105

7. RECOMMENDATIONS 109

7.1 Data Issues 109

7.2 Recommendations for the RFFA methods assessed in this study 109

7.3 Future Recommendations for RFFA 110

8. REFERENCES 112

APPENDIX A: COMPLETE INVENTORY OF STATIONS ANALYSED 119

APPENDIX B: STATIONS WHERE QUATERNARIES WERE INCORRECTLY LISTED BY DWS 135

APPENDIX C: DOMINANT UPSTREAM VELD ZONES 137

APPENDIX D: K-REGION AT STATION LOCATION..143

APPENDIX E: FREQUENCY PLOTS OF ABSOLUTE RELATIVE ERRORS 146

APPENDIX F: METHOD RANK 148

LIST OF TABLES

Page

Table 2.1 Parameters for the estimation of design discharges, index floods and regional growth curves for the van Bladeren (1993) Study 15

Table 2.2 Regionalised parameters derived for the 13 homogeneous flood regions identified by Mkhandi *et al.* (2000) 22

Table 2.3 Coefficients for the determination of the index floods (after Görgens, 2007) 29

Table 2.4 Regression models to predict the median values from catchment area (Haile, 2011) 34

Table 4.1 Additional exclusion of stations 46

Table 4.2 Stations where the MAR90 values differed between this study and the JPV study 57

Table 4.3 Quaternary data for station V2H002 58

Table 4.4 Veld zones reported by Görgens (2007), compared to Veld zones determined in ArcGIS 67

Table 4.5 K-Regions reported by Görgens (2007), compared to K-Regions determined in ArcGIS 68

Table 4.6 Regional regression equations for the estimation of the index flood, using the Haile (2011) Method in South Africa 76

Table 5.1 Slope of observed versus estimated design floods for the regional methods, using the GEV distribution 83

Table 5.2 Average relative errors for the lower quartile, median, upper quartile and inter-quartile range 88

Table 5.3 Average and Median *MARE* (%) values for the RFFA methods 91

Table 5.4 Average GU, GO and D ranges for all the return periods 94

LIST OF FIGURES

Page

Figure 2.1 Regions analysed in the Meigh *et al.* (1997) study 16

Figure 2.2 Comparison of regional flood frequency curves developed by Meigh *et al.* (1997) 17

Figure 2.3 Map of homogeneous regions identified for South Africa (after Mkhandi *et al.*, 2000) 21

Figure 2.4 Map of homogeneous regions in KZN and the Kovacs flood regions (Kjeldsen *et al.*, 2002) 24

Figure 2.5 Stations analysed in the Görgens (2007) study 26

Figure 2.6 Veld zone pooling groups utilised in the JPV Method (Görgens, 2007) 27

Figure 2.7 K-Region pooling groups utilised in the JPV Method (Görgens, 2007) 28

Figure 2.8 Map of homogeneous flood regions delineated by Haile (2011) 32

Figure 2.9 Stations in South Africa analysed in the Haile (2011) study 34

Figure 3.1 Map of the main rivers, primary drainage regions and flow gauging stations in South Africa 39

Figure 4.1 Distribution of record lengths for all flow gauging weirs across South Africa 43

Figure 4.2 Rating table exceedence for stations in South Africa 44

Figure 4.3 Location of stations in relation to the location of a dam 45

Figure 4.4 Record lengths for the final stations and dam inflow records analysed in this study 47

Figure 4.5 Rating table exceedence for the final 321 stations analysed in this study 47

Figure 4.6 Distribution of total flow records across South Africa, compared to a map of final stations and dam inflows records used in the analysis 48

Figure 4.7 Example of problem with spatial station location 50

Figure 4.8 Map of the delineation process for sub-catchment area estimation 51

Figure 4.9 Maps of stations located at quaternary outlets and those not located at quaternary outlets 52

Figure 4.10 Comparison of areas published by the DWS and areas estimated, using ArcGIS 53

Figure 4.11 Distribution of catchment areas 54

Figure 4.12 Comparison of MAR90 57

Figure 4.13 Equal area slope method (Gericke and du Plessis, 2012) 59

Figure 4.14 ArcGIS procedure to obtain the river profile using Model Builder 60

Figure 4.15 Equal area slope for Station V2H002 60

Figure 4.16 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Görgens (2007) 61

Figure 4.17 Scatter plot of slopes calculated in this study, compared to slopes calculated by Görgens (2007) 62

Figure 4.18 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Smithers *et al.* (2015) 63

Figure 4.19 Scatter plot of slopes calculated in this study, compared to slopes calculated by Smithers *et al.* (2015) 63

Figure 4.20 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Gericke (2015) 64

Figure 4.21 Scatter plot of slopes calculated in this study, compared to slopes calculated by Gericke (2015) 65

Figure 4.22 ArcGIS procedure to obtain HRU Veld zones 66

Figure 4.23 Map showing the dominant Veld zone for Station A2R006 68

Figure 4.24 Map showing the K-Region at Station A2R005 69

Figure 4.25 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R1 region 71

Figure 4.26 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R2 region 72

Figure 4.27 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R3 region 73

Figure 4.28 Derived equation to estimate an index flood in the ZA_R3 region 74

Figure 4.29 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R5 region 75

Figure 4.30 Derived equation to estimate an index flood in the ZA_R5 region 76

Figure 5.1 Comparison of observed design floods for the GEV and LP3 distribution, for the 50-year Return Period 77

Figure 5.2 Performance of the JPV Method, using the K-Region regionalisation and the LP3 distribution, for the 50-year Return Period 79

Figure 5.3 Performance of the JPV Method, using the Veld zone regionalisation and the LP3 distribution, for the 50-year Return Period 79

Figure 5.4 Performance of the Meigh Method, using the GEV distribution, for the 50-year Return Period 80

Figure 5.5 Performance of the Mkhandi Method, using the GEV distribution, for the 50-year Return Period 81

Figure 5.6 Performance of the JPV Method, using the K-Region regionalisation and the GEV distribution, for the 50-year Return Period 81

Figure 5.7 Performance of the JPV Method, using the Veld zone regionalisation and the GEV distribution for the 50-year Return Period 82

Figure 5.8 Performance of the Haile Method, using the GEV distribution for the 50-year Return Period 82

Figure 5.9 Box plots of relative errors for the 2- to 10-year return periods 85

Figure 5.10 Box plots of relative errors for the 20- to 100-year return periods 86

Figure 5.11 Performance of RFFA methods based on the frequency of absolute relative errors for the 50-year Return Period 89

Figure 5.12 Estimated/observed ratios for the 2- to 10-year return periods 92

Figure 5.13 Estimated/observed ratios for the 20- to 100-year return periods 93

Figure 5.14 Estimated/observed ratios for the 50-year return period at different catchment area ranges 97

Figure 5.15 Method rank for the 50-year return period 98

LIST OF SYMBOLS

B_4	=	bias of the regional average sample L-Kurtosis
DES_i	=	catchment descriptors
g_Q	=	skewness
H_B	=	height at the station (m)
H_i	=	height for a specific contour interval
L_i	=	distance between contours (m)
k	=	curvature of regional curve
K_T	=	frequency factor
P	=	non-exceedence probability
Q_T	=	design flood discharge ($m^3.s^{-1}$)
q	=	standardised flood peak
S_{CH1}	=	average main channel slope
T	=	RI or 1/AEP
$T^4{}_{DIST}$	=	population L-kurtosis of the selected distribution
$t^R{}_4$	=	regional average sample L-kurtosis
u	=	intercept of regional curve
μ_Q	=	mean
μv	=	mean of simulated V values
V	=	variance of the mean L-CV
σv	=	standard deviation of simulated values of V

$W_{(y_1,k)}$ = frequency factor for the GEV distribution

$\bar{x}$ = mean

y = reduced variate

σ = standard deviation

$\sigma 4$ = standard deviation of regional average sample L-kurtosis

α = scale parameter of the regional curve

$>$ = greater than

$<$ = less than

$\leq$ = less than or equal to

Σ = sum

LIST OF ABBREVIATIONS

AEP	=	Annual Exceedence Probability
ALT	=	Altitude
AMF	=	Annual Maximum Flood
AMS	=	Annual Maximum Series
ARI	=	Annual Recurrence Interval
ARR RFFE	=	Australian Rainfall and Runoff Regional Flood Frequency Estimation Model
AWD	=	Area Weighted Depth
BI	=	Bayesian Inference
CONC	=	Rainfall concentration
CS	=	Coefficient of Skewness
CV	=	Coefficient of Variation
DEM	=	Digital Elevation Model
DWA	=	Department of Water Affairs
DWS	=	Department of Water and Sanitation
EV1	=	Extreme Value Type I
EXP	=	Exponential
FEH	=	Flood Estimation Handbook
FFA	=	Flood Frequency Analysis
FSEE	=	Factorial Standard of the Estimate

G2	=	Gamma 2-parameter
GCC	=	Gravelius' Compactness Coefficient
GEV	=	General Extreme Value
GIS	=	Geographic Information System
GLO	=	General Logistic
GLS	=	Generalised Least Squares
GNDCI	=	Group on Hydrogeological Disasters Prevention
GNO	=	General Normal
GPA	=	General Pareto
HOST	=	Hydrology of Soil Types
HRU	=	Hydrological Research Unit
IFD	=	Intensity-Frequency-Duration
JPV	=	Joint Peak-Volume
KZN	=	KwaZulu-Natal
LAT	=	Latitude
L-CK	=	L-Coefficient of Kurtosis
L-CS	=	L-Coefficient of Skewness
L-CV	=	L-Coefficient of Variation
LM	=	L-Moments
LN	=	Log-Normal
LONG	=	Longitude

LP3	=	Log Pearson Type III
MAF	=	Mean Annual Flood
MAP	=	Mean Annual Precipitation
MAR	=	Mean Annual Runoff
MARE	=	Mean Absolute Relative Error
MCS	=	Mean Catchment Slope
MLP	=	Maximum Likelihood Procedure
MOM	=	Method of Moments
P3	=	Pearson Type 3
PRT	=	Parameter Regression Technique
PWM	=	Probability Weighted Moments
QRT	=	Quantile Regression Technique
RFFA	=	Regional Flood Frequency Analysis
RMF	=	Regional Maximum Flood
ROI	=	Region of Influence
RSA	=	Republic of South Africa
SAAR	=	Standard Annual Average Rainfall
STRM	=	Shuttle Radar Topography Mission
UK	=	United Kingdom
VA.PL.	=	"VAlutazione Plene", Flood Estimation
WAK	=	Wakeby

1. INTRODUCTION

On a worldwide scale, flood disasters are considered to be the most predominant and frequently occurring natural hazard (EM-DAT, 2012). The economic loss and loss of life caused by floods can occur at both small and large scales (Hubbart and Jones, 2014). Economic losses resulting from flood events have increased over the past three decades, from an approximate annual median average of R6 billion ($0.5 billion) in the 1980s, to R235 billion ($20 billion) in the first decade of the twentieth century (EM-DAT, 2012). With a growing population, increased urbanisation and climate change, the risks that flood events pose are becoming more severe, which requires researchers to improve the accuracy and reliability of methods used for flood estimation (Wiltshire, 1986; Smithers and Schulze, 2000)

In engineering and flood hydrology, the estimation of a design flood refers to procedures whereby the magnitude of a flood is associated with a level of risk at a given site (Pegram and Parak, 2004). This provides information that is needed in the design, planning and operation of hydraulic structures, such as drainage canals, culverts, dam spillways, bridges, as well as detention and retention ponds (Pegram and Parak, 2004; Chetty and Smithers, 2005; Merz and Blöschl, 2005; Saf, 2010; Haddad and Rahman, 2012). The correct design of such structures will ensure the protection of human life and property in a feasible and pragmatic manner (Pegram and Parak, 2004; Reis and Stedinger, 2005; Saf, 2010). There are numerous methods that can be employed in design flood estimation. In South Africa, many of the methods utilised were developed in the mid- to late-70's; and are in need of being updated. One such method is known as Flood Frequency Analysis (FFA), which involves the interpretation of a past record of hydrological events in terms of the future probability of occurrence. This can be achieved through an at-site FFA or through a regional FFA. In order for an at-site analysis to be achieved, a record of observed flows is required and must be of adequate length and quality (Smithers, 2012). A record of sufficient length is often not available at the site being investigated and thus a regional approach must be considered (GREHYS, 1996a; Viglione *et al.*, 2007). A number of studies have advocated a regional approach for obtaining more reliable design flood estimates (Wiltshire, 1986; Hosking and Wallis, 1997; Saf, 2008; Saf, 2009; Haile, 2011; Smithers, 2012).

Regional Flood Frequency Analysis (RFFA) involves determining homogeneous flood response regions and selecting a suitable frequency distribution for the region (Kachroo *et al.*, 2000). RFFA involves the use of data from sites other than the site under investigation (Hosking and Wallis, 1997). This allows for data from more than one site to be utilised, creating the potential for more accurate estimates of flood quantiles (Hosking and Wallis, 1997). A RFFA is necessary at ungauged sites or at sites where an inadequate length or poor quality of stream flow data is available (Leclerc and Ouarda, 2007). A RFFA can improve flood quantile estimates at gauged sites where the record length is insufficient (Australian Institution of Engineers, 1977). A regional approach will allow for a FFA of short records and annual floods to be implemented, through the determination of the shape of the parent distribution and the estimation of scale to be achieved from data at the site of interest (Bobee and Rasmussen, 1995). Historical data can be pooled from a region that is homogeneous, allowing the estimation of the parameter distribution and subsequently achieving more robust quantile estimates (Kachroo *et al.*, 2000).

The primary assumption made in a RFFA is that at every site within a region under study, the standardised variate will have the same distribution and the data obtained from the sites can be pooled to develop a flood frequency curve that can be utilised anywhere throughout the region (Cunnane, 1989; Hosking and Wallis, 1997; Smithers, 2012). RFFA involves two primary steps, namely, identifying homogeneous regions, which are regions that have similar flood producing characteristics (Mkhandi *et al.*, 2000), and determining an appropriate frequency distribution for the data from the regions (Malekinezhad *et al.*, 2011). In a RFFA approach, a regional growth curve can be produced, which represents the average weighted distribution of a homogeneous region (Haile, 2011). A regional growth curve shows the normalised regional flood quantiles for a given return period (Haile, 2011). The final step in a RFFA approach is to develop a method to apply the RFFA at ungauged sites.

The use of a RFFA approach improves the accuracy and reliability of estimates of design floods (Wiltshire, 1986; Hosking and Wallis, 1997; Saf, 2008; Saf, 2009; Haile, 2011; Smithers, 2012). However, no RFFA method is currently widely used in South Africa, despite a number of RFFA studies having been undertaken, that include South Africa. Hence, the performance of the current RFFA approaches needs to be assessed in order to determine the best approaches to use and to determine if a new RFFA approach needs to be developed for use in South Africa.

1.1 Aim and Objectives

The primary aim of this study is to assess the performance of currently available RFFA methods in South Africa. The objectives of this study are as follows:

a) **Objective 1**: To review the literature pertaining to the current methods employed in RFFA, both locally and internationally, in order to inform the selection of methods to be assessed in this study.
b) **Objective 2**: To apply and assess the performances of the selected RFFA methods in South Africa.
c) **Objective 3**: To identify and compare any variations in the performance of the RFFA methods and the reasons for these variations.
d) **Objective 4**: To select a suitable method, based on its performance, or to recommend the development of a new approach.

1.2 Research Questions

The following research questions need to be addressed:

a) How well do the design flood estimates obtained from the RFFA methods compare to the estimates obtained from an at-site FFA in a nationwide study?
b) What are the advantages and disadvantages of each method?
c) Are there any variations in method performance relating to input parameters, or relating to where the method is applied?
d) Which method is best suited for use in South Africa?
e) Is the development of a new RFFA method warranted, given the performance of the current RFFA methods?

1.3 Delineation and Limitations

This study involves a comparative assessment of the performance of RFFA methods throughout South Africa. While some of the methods assessed have been applied in previous studies in other parts of southern Africa, the results of this study cannot be assumed to hold true in regions outside of South Africa. Design floods will be estimated for the 2-, 5-, 10-, 20-, 50- and 100-year return periods. Design floods beyond the 100-year return period have not been estimated, due to the problems associated with extrapolating design floods beyond the record length of the datasets used (Schulze, 1989). For example, Schulze (1989) explains that an estimate of the 200-year design flood is both statistically and scientifically meaningless when it is being determined from a short record length of 20 years. This assumes that the statiscal pattern observed for a short record length is valid even beyond the observed range of values, which is often not the case in reality. In addition, the amount of data used to obtain the design flood estimates has been limited by the quality of the datasets obtained from the Department of Water and Sanitation (DWS).

1.4 Dissertation Structure

This section outlines the structure of this dissertation and a summary of the chapters that are to follow. Chapter 2 begins with a summary of the common approaches to RFFA. The methods employed in identifying homogeneous flood regions and selecting an appropriate frequency distribution is also discussed. The remainder of this chapter entails a review of the RFFA methods employed internationally and in South Africa. The advantages and disadvantages of these methods are summarized, in order to select those that are most appropriate to be applied and assessed in this study. The chapter ends with a summary of the literature reviewed, leading to the selection of methods to be assessed in this study. Chapter 3 provides a general description of the study area and a brief overview of the climate and hydrology of South Africa. Chapter 4 provides a detailed description of the methodology employed in the assessment of the RFFA methods. The collection and screening of the data utilised is discussed in detail. Chapter 5 contains the results of the study and a description of the evaluation statistics utilised to assess the RFFA methods. Chapter 6 entails a synthesized discussion on the results and other aspects pertaining to the study, as well as the final conclusions. Chapter 7 contains the recommendations of the study and Chapter 8 contains the references, followed by the appendices.

2. LITERATURE REVIEW

This chapter begins with a summary of some of the pertinent aspects involved in the application of a RFFA. Approaches such as the index flood method are addressed, as well as the methods involved in identifying homogeneous flood regions and selecting suitable probability distributions for design flood estimation. Thereafter, the RFFA methods that are currently being applied locally and internationally will be reviewed.

A common approach adopted in RFFA is the index flood method as proposed by Dalrymple (1960). This procedure involves scaling the Annual Maximum Flood (AMF) data (a series containing the largest flood event in each year), by an index flood (e.g. Mean Annual Flood or Median Annual Flood) and the use of regression models to establish a relationship between catchment characteristics and the index flood (Nobert *et al.*, 2014). The primary assumption of the index flood method is that the sites within a homogeneous region will all have similar frequency distributions of flood responses apart from a scaling factor specific to the site being investigated (Hosking and Wallis, 1997). The combination of short records within a region do not produce a longer record, but provide an average of the records that give a better measure of the frequency distributions of the events (Dalrymple, 1960). This procedure allows for the development of regionalised, scaled distribution parameters for each homogeneous region or the development of a growth curve which represents a relationship between a ratio of design flood/index flood and the return period. The first step in the index flood method involves the identification of sites that are similar to the site being investigated (GREHYS, 1996b). The second part involves the use of data from the neighbouring sites, to determine the flood quantiles at the site of interest (GREHYS, 1996b). The index flood method can be applied at a gauged site by determining the index value, used to scale the distribution, from the at-site flood data series, but at an ungauged site, the index must be estimated from the physiographic characteristics of the site being investigated (Ilorme and Griffis, 2013).

A pertinent component of RFFA is the identification of homogeneous regions (Viglione *et al.*, 2007). From the preliminary review of the literature, a number of approaches to regionalisation are evident. These include proximity pooling techniques such as the Region of Influence approach (ROI) (Burn, 1990), hierarchical clustering (Gabriele and Arnell, 1991), which can either be agglomerative or divisive (Crochet, 2012), geographical regionalisation, cluster analysis and mixed procedures (Merz and Blöschl, 2005). The ROI approach involves defining each gauging station with a potentially unique or distinct region (Crochet, 2012). The advantage of this approach is that it does not require regions to be divided by geographic boundaries (Crochet, 2012). Each site can be considered its own region, containing all the catchments that have a similar distribution of flood responses. The hierarchical clustering approach involves manipulating a dataset by dividing it into "mutually excluding and jointly exhaustive" groups (Crochet, 2012). A cluster analysis has the advantage of using the statistics from the observed data independently to evaluate the homogeneity of the region. A problem that is often encountered in regionalisation is discontinuities at the boundaries of regions. Laio *et al.* (2011) proposed an approach to provide an alternative to regionalisation, whereby hydrological information is transferred to ungauged sites assuming no regionalisation of pooling groups.

In order to determine a design event from a flood record, an appropriate frequency distribution must be chosen. A number of methods can be used in fitting distributions. These include the Method of Moments (MOM), Maximum Likelihood Procedure (MLP), L-Moments (LM), Bayesian Inference (BI) and the Non-Parametric Method (Smithers and Schulze, 2000). This review will focus on the use of L-moments and a more detailed explanation is therefore given. The use of L-moments, as proposed by Hosking and Wallis (1997), has been widely applied in a number of regional frequency analysis studies. These include studies by Smithers and Schulze (2000) in South Africa, Kjeldsen *et al.* (2002) in KwaZulu-Natal (KZN), Jingyi and Hall (2004) in the Gan-Ming River basin in China, Kizza *et al.* (2006), in northern Uganda, Saf (2009), in the west Mediterranean region of Turkey, Malekinezhad *et al.* (2011) in the Namak-Lake basin in Iran, Mediero and Kjeldsen (2014) in the Ebro catchment in Spain and Wazneh *et al.* (2015) in northwest Italy. In the case of conventional moments, estimators such as the skewness and kurtosis are determined by cubing or squaring observations (Gordon *et al.*, 2004). This means that conventional moments are influenced more by the very high or low values which may be outliers (Gordon *et al.*, 2004). L-moments are less subject to bias, in comparison to other moments as they are linear functions of the observations (Hosking and Wallis, 1997). L-moments are more robust to the presence of outliers, and are able to characterise a greater range of distributions (Hosking and Wallis, 1997).

2.1 A Review of International Approaches to Regional Flood Frequency Analysis

This section entails a review of the selected methods used in RFFA internationally which have been developed within the past 5 years. The different steps and approaches to a RFFA that have already been discussed above are used in some of the international approaches, as reviewed in the following sections.

2.1.1 France

A RFFA study in France focused on two proposed approaches, as reported by Nguyen *et al.* (2014). The first of the two approaches involves the utilisation of paleo-flood data and historical data, with the aim of achieving a temporal extension of existing data records. The second approach involves the merging of data from statistically homogeneous regions, with the aim of achieving a spatial extension. Gaume *et al.* (2010) proposed a method that combined the above two approaches with the purpose of including extreme flood data from ungauged sites within a region. Nguyen *et al.* (2014) compared the results of the method developed by Gaume *et al.* (2010) (proposed approach) to the Hosking and Wallis (1997) method (standard approach), which is widely used in design hydrology practice. The proposed approach utilised extreme discharge data from ungauged sites obtained from the Hydrometeorological data resources and technologies for effective flash flood forecasting (HYDRATE) project. The HYDRATE database includes estimates from numerous sources, including field surveys based on eye-witness accounts, pictures and films etc. Using these sources of information, estimates have been determined using methods such as hydraulic modelling (1D or 2D) and hydraulic formulae, amongst others.

The inclusion of ungauged extreme flood data involved an index flood relation where the index flood has been related to the catchment area (A) as proposed by Gaume *et al.* (2010). This is expressed mathematically in Equation 2.1.

$$\mu_i = A_i^{\beta} \, and \, \mu_k = A_k^{\beta} \qquad (2.1)$$

where μ_i represents the index flood, A_i and A_k represent the catchment areas at sites i and k and β represents a coefficient that is to be calibrated.

The proposed approach was first compared to the standard approach using only the gauged data to assess the statistical performance before the application of the proposed approach with the incorporation of ungauged data. The results show that on average the standard approach under-estimates quantiles, while the proposed approach over-estimates quantiles when being applied for smaller catchments. A comparison of the proposed and standard approach was then performed in the Ardèche region in France using 5 gauged datasets as well as 18 ungauged extremes taken over 50 years. The results show that the inclusion of ungauged data improved the quantile estimates indicating the value of using ungauged data in a RFFA.

2.1.2 Australia

In Australia, a model has been under development that will enable the estimation of design floods, ranging from the 2- and the 100-year return periods, and for catchment sizes, ranging between 1 km^2 and 1000 km^2. This model is referred to as the Australian Rainfall and Runoff Regional Flood Frequency Estimation Model, or 'ARR RFFE 2012'. Rahman *et al.* (2013) outline the methodology employed in the 'ARR RFFE 2012', from data screening to the final test method. The preliminary data screening in the Rahman *et al.* (2013) study took into account the following criteria, to select the stations that would be suitable for further analysis (Rahman *et al.*, 2013):

a) The catchment size should not exceed 1000 km^2.
b) The record length of the Annual Maximum Series (AMS) data should be 25 years or longer.
c) The catchments should not be influenced by channel activities, such as the existence of a dam.
d) The urbanisation within the catchment should not exceed 10%.
e) The record of data that will be analysed should not have any significant changes in land use occurring throughout the period of record length utilised in the study.
f) The gauging authority must be responsible for the quality of flood data to be used.

The final test utilised 676 catchments to identify six regions and four fringe zones across Australia. The ROI approach was utilised to determine Regions 1, 2, 4 and 5, as well as fringe zones A-D. The remaining regions were determined by using a fixed region approach, where all available sites are included in one fixed region. The data from the catchments were also used to create prediction equations for the ten regions, through a Bayesian generalised least squares (GLS) regression technique. In order to estimate design floods for the 2- to 100-year return periods, a regionalised Log-Pearson Type 3 (LP3) distribution was used. Within each region, design rainfall intensity and catchment area were used in regional prediction equations to estimate the skewness, mean (M) and standard deviation (S) in Equation 2.2 and were used with the LP3 distribution to estimate the flood quantiles. In the development of this model, Rahman *et al.* (2013) compared a Quantile Regression Technique (QRT) to a Parameter Regression Technique (PRT). A QRT involves the estimation of flood quantiles within a region for a large number of gauged catchments. These quantiles are then related to catchment characteristics that influence floods. Similarly, when applying a PRT, the parameters of a selected distribution are related to catchment parameters. Equations were developed to estimate the skewness, mean and standard deviation of the AMS which can then be utilised to estimate these statistics at an ungauged site with the aim of fitting a probability distribution which is then utilised in flood quantile estimation at the ungauged site. The advantage of the PRT, in comparison to a QRT, is that it ensures flood quantiles increase smoothly, with increasing Annual Recurrence Interval (ARI), and it allows for the estimation of quantiles for any ARI (Rahman *et al.*, 2013). The PRT utilised in the Rahman *et al.* (2013) study involved a regionalisation of the skewness, mean and standard deviation and utilised the LP3 distribution in the estimation of flood quantiles, as shown in Equation 2.2:

$$In(Q_T) = M + K_T S \tag{2.2}$$

where

Q_T = the discharge having an Annual Exceedence Probability (AEP) of 1/T (where T is return period in years) ($m^3.s^{-1}$),

M = mean of the natural logarithms of the annual maximum flood series ($m^3.s^{-1}$),

S = standard deviation of the natural logarithms of the annual maximum flood series, and

K_T = frequency factor for the LP3 distribution for AEP of 1/T, which is a function of AEP and skewness.

An application tool has been developed that will automate the 'ARR RFFE 2012' procedures. However, the model requires recalibration with updated Intensity-Frequency-Duration (IFD) data, as well as updated flood data, before it can be applied by design flood practitioners.

2.1.3 Iceland

In a RFFA study conducted in Iceland, Crochet (2012) developed a method using the index flood method as proposed by Dalrymple (1960). Ten stations across Iceland were utilised by Crochet (2012), with catchment areas ranging from 37 km^2 to 1096 km^2. As part of the study, two methods of regionalisation were also assessed, namely the ROI approach and the hierarchical clustering approach, which were compared to a geographical proximity delineation technique.

Crochet (2012) utilised Wards Method (agglomerative technique), which involves defining one cluster for each site and the clusters are merged until the merged cluster becomes non-homogeneous. Following the delineation into homogeneous regions, the groups were tested statistically for homogeneity. This was achieved, using the H-statistic developed by Hosking and Wallis (1993).

The index flood utilised by Crochet (2012) was the mean of the AMS, represented as Q_{index}. The flood frequency distribution at a particular site was determined by rescaling the distribution by Q_{index}. This is expressed mathematically below:

$$Q_i(T) = q_R(T) \, x \, Q_{index} \tag{2.3}$$

where $Q_i(T)$ represents the T-year peak discharge for catchment i and $q_R(T)$ represents the growth factor for a region.

The results of this study indicate that the adopted delineation strategy is important in the overall procedure, as well as for the selection of an appropriate index flood model. This is significant, as an inaccuracy in the estimation of the index flood can result in either an over- or under-estimation of the frequency distribution.

2.1.4 United Kingdom

Following the publication of the Flood Estimation Handbook (FEH) in 1999 (Institute of Hydrology, 1999), a revised version of the index flood method was developed by Kjeldsen *et al.* (2008). The following is a summary of the main technical aspects involved in the Revised

FEH Flood Statistics (ReFS) Method, which has been reported by Castellarin *et al.* (2012). Homogeneous regions were determined by developing pooling groups, using a similarity measure. This measure was based on an index of the extent of upstream flood plains, the Standard Annual Average Rainfall (*SAAR*), an index of Flood Attenuation from Reservoirs and Lakes (*FARL*), the Extent of Flood Plains (*FPEXT*) and catchment Area (*AREA*). The similarity measure (d_{ij}) is expressed mathematically as shown in Equation 2.4, where subscripts *i* and *j* refer to two different sites (Kjeldsen *et al.*, 2008):

$$d_{ij}=\sqrt{3.2\left(\frac{ln[AREA_i]-ln[AREA_j]}{1.28}\right)^2+0.5\left(\frac{ln[SAAR_i]-ln[SAAR_j]}{0.37}\right)^2+0.1\left(\frac{FARL_i-FARL_j}{0.05}\right)^2+0.2\left(\frac{FPEXT_i-FPEXT_j}{0.04}\right)^2} \quad (2.4)$$

The index flood used by Kjeldsen *et al.* (2008) was the median annual maximum flood (*QMED*), which was related to a set of catchment descriptors, using a log-linear regression model developed by Kjeldsen and Jones (2009). The equation developed for the regression model was derived from annual maximum flow data from 602 catchments and is expressed as follows:

$$QMED=8.3062AREA^{0.8510}0.1536^{\left(\frac{1000}{SAAR}\right)}FARL^{3.4451}0.0460^{BFIHOST^2} \quad (2.5)$$

where *QMED* is the median annual maximum flood and *BFIHOST* is an index of base flow, which is defined by HOST (Hydrology of Soil Types) soil classes (Boorman *et al.*, 1995).

The General Logistic (GLO) frequency distribution was utilised by Kjeldsen *et al.* (2008), as it was the distribution used in the FEH (Institute of Hydrology, 1999). The data were scaled using the median as the index flood and the method of L-moments was used to fit the GLO distribution. The catchment data used by Kjeldsen *et al.* (2008) were obtained from a number of datasets, each captured at different scales. Therefore, the catchment descriptors to be used were downscaled to a catchment scale, based on the United Kingdom (UK) river network digitised maps, and a 50 m Digital Terrain Model. The catchment descriptors included data on reservoirs, soils, climate, and land-cover, amongst others. When applying this method in an urban area, adjustments are made to the procedure, which are detailed by Kjeldsen (2010). This study has improved upon flood estimation procedures through updated data and new techniques, as described above. As a result, the reliability of design flood estimation in the UK has been improved.

2.1.5 Italy

The VA.PL. ("VAlutazione Plene", Flood Estimation) project, developed by the National Group on Hydrogeological Disasters Prevention (GNDCI), details a RFFA procedure utilised in Italy. This study has been reported by Castellarin *et al.* (2012). The procedure involves a hierarchical index flood method, whereby three levels have been developed to compartmentalise the regions of Italy. The first level compartmentalises Italy into hydro-climatic macro-regions. These regions are hydrologically homogeneous and the third and higher order moments, such as skewness and kurtosis coefficients, are assumed to remain constant. In the second level, Italy is compartmentalised into homogeneous sub-zones, where the second order moments (Coefficient of variation or L-Coefficient of variation) are assumed to remain constant. The third level consists of regions where the flood formation mechanisms are homogeneous with regards to the first order moments, such as the index flood or the mean of the distribution.

According to Castellarin *et al.* (2012), an advantage of the method is that it covers the entire country and it improves the estimates of higher order moments. However, it is not up-to-date in terms of data and it may cause certain inconsistencies in the development of regional growth curves, such as "abrupt jumps" i.e. sudden changes in the index flood at the boundary between regions. Therefore, Castellarin *et al.* (2012) concluded that there is still room for improvement.

In this chapter, a summary of some of the current methods of RFFA that are utilised internationally has been presented. From the literature reviewed, it is apparent that the index flood method is widely adopted in many studies and has been applied in different ways, depending on the relationship that has been developed between the index flood and catchment descriptors in the respective studies. The following sections will review six RFFA methods that are available for use in South Africa.

2.2 Review of Regional Flood Frequency Analysis Studies in South Africa

This section contains a review of the RFFA approaches that have been developed for use in South Africa. These include the van Bladeren (1993) Method, the Meigh *et al.* (1997) Method, the Mkhandi *et al.* (2000) Method, the Kjeldsen *et al.* (2002) Method, the Görgens (2007) Joint Peak-Volume (JPV) Method, and the Haile (2011) Method

2.2.1 Van Bladeren Method

Van Bladeren (1993) evaluated what were then the current methods of FFA, taking into account historical records, and found that the methods in use at that time had drawbacks and were in need of further improvement. This study was undertaken in the KZN and Transkei regions. The data required to determine the flood peaks of historical events were obtained from a number of sources, such as consultants, local and provincial authorities, as well as literature, newspapers and records from the Department of Water Affairs (DWA). The estimation of the historical flood peaks was achieved by using data from gauging stations and methods such as the Chezy method, slope-area methods and reservoir routing, amongst others.

Following the publication of documentation on historical floods in Natal and Transkei from 1848 – 1989 (van Bladeren, 1992), additional data were collected to develop regional growth curves. The regions delineated by Kovacs (1988) were used in the development of these growth curves and the GEV distribution fitted by Probability Weighted Moments (PWM) was utilised, as used previously by Alexander (1990). Further regionalisation was identified after plotting the skewness of the data sets and comparing them to the Regional Maximum Flood (RMF) regions. The van Bladeren (1993) method can be applied in RMF regions 5.0 and 5.6 in the KZN and Transkei regions. The initial regressions used by van Bladeren (1993) included only the catchment area for each regional growth curve and the *MAF* was determined, using Equation 2.6:

$$MAF = CONSTANT \times AREA^{EXPONENT} \tag{2.6}$$

where

MAF	=	mean annual flood ($m^3.s^{-1}$),
CONSTANT	=	regionalised parameter derived from Table 2.1,
AREA	=	catchment area (km^2), and
EXPONENT	=	regionalised parameter derived from Table 2.1.

Design floods were estimated using the GEV distribution in the van Bladeren (1993) study, using the same approach as used by NERC (1975) and shown in Equation 2.7:

$$Q_T = u + \alpha \times W_{(y_1,k)} \tag{2.7}$$

where

Q_T	=	design discharge for return period, T ($m^3.s^{-1}$) ,
u	=	location parameter for GEV distribution,
α	=	scale parameter for GEV distribution,
k	=	shape parameter for GEV distribution,
$W_{(y_1,k)}$	=	frequency factor for the GEV distribution,
	=	$\dfrac{1 - e^{-k\times}\left[-ln\left(-ln\left[\frac{T-1}{T}\right]\right)\right]}{k}$

Table 2.1 contains the parameters for the estimation of the *MAF*, as well as the parameters for the GEV distribution. Growth factors for the 2- to 200-year return periods are also included in Table 2.1, where CV represents the coefficient of variation and g represents the skewness.

Table 2.1 Parameters for the estimation of design discharges, index floods and regional growth curves for the van Bladeren (1993) Study

Region	Sub-region	*MAF* equation		GEV/PWM Parameters					Growth factors				
									Return Period				
		Constant	Exponent	CV	*g*	*α*	*μ*	*k*	10	20	50	100	200
5.6	Coastal areas of drainage region (≈30 km strip)	1.73	0.72	1.78	5.79	0.35	0.35	-0.57	1.86	2.91	5.13	7.78	11.79
	Interior areas of drainage Region W	29.06	0.41	1.53	3.61	0.42	0.48	-0.39	2.13	3.09	4.83	6.72	9.25
	Region 5.6	2.15	0.71	1.71	5.15	0.37	0.39	-0.52	1.84	2.96	5.05	7.47	11.04
5.4	Mkomazi to Mvoti rivers	9.36	0.52	1.75	4.32	0.31	0.41	-0.54	2.09	3.28	5.69	8.48	12.54
	Drainage regions U6, U7 & U8	6.19	0.61	1.01	2.45	0.66	0.59	0.05	2.34	3.38	4.73	5.95	7.39
	Region 5.4	8.8	0.54	1.35	3.3	0.5	0.51	-0.22	2.23	3.33	5.17	7.1	9.73
5.2		0.93	0.77	1.38	3.75	0.45	0.42	-0.42	1.99	2.88	4.57	6.45	9.06
5	Drakensberg and drainage region V1	2.34	0.74	0.87	2.1	0.65	0.41	-0.18	1.83	2.37	3.26	4.12	5.25
	Drainage region V3 & V6	3.02	0.6	0.74	2.16	0.64	0.42	-0.21	1.83	2.39	3.24	4.01	4.93
	Drakensberg foothills and rest of region 5.0	0.82	0.78	1	3.59	0.55	0.4	-0.35	1.91	2.64	3.9	5.2	6.9
	Region 5.0	6.13	0.53	0.89	2.65	0.61	0.41	-0.25	1.86	2.47	3.49	4.49	5.78

2.2.2 Meigh Method

A study involving a worldwide comparison of RFFA methods under different climatic conditions was conducted by Meigh *et al.* (1997). Figure 2.1 illustrates the regions that were analysed by Meigh *et al.* (1997). In South Africa and Botswana, datasets from 101 flow gauging stations were analysed.

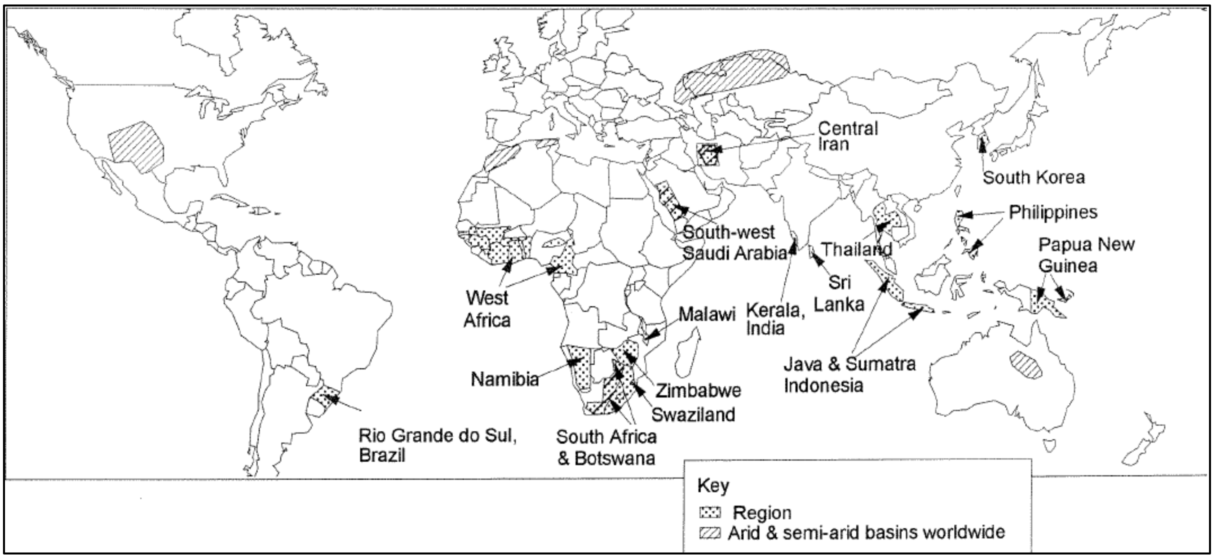

Figure 2.1 Regions analysed in the Meigh *et al.* (1997) study

2.2.2.1 Estimation of index floods

Meigh *et al.* (1997) used the *MAF* as an index value to scale the data. A multiple regression analysis was performed to determine the relationship between the *MAF* and the catchment characteristics, in order to produce prediction equations. For South Africa, the prediction equation was developed based on catchment area and MAP. The prediction equations were assessed using two measures i.e. the coefficient of determination (r^2) and the Factorial Standard Error of the Estimate (FSEE). The r^2 value expresses the proportion of variance of the *MAF* (dependant variable) that is predictable from the catchment area (independent variable), while the FSEE value expresses the degree of deviation of the estimates from the "true" value. The aim in the development of these equations was to maximize the r^2 value and minimize the FSEE value. The prediction equation produced for South Africa had an r^2 value of 0.542 and an FSEE of 2.19 and is expressed mathematically in Equation 2.8:

$$MAF = 6.97 \times AREA^{0.450} \tag{2.8}$$

where

MAF = mean annual flood ($m^3.s^{-1}$), and

$AREA$ = catchment area (km^2).

2.2.2.2 Development of regional growth curves

Regional flood frequency curves were developed by Meigh *et al.* (1997), using the GEV distribution and PWM. These growth curves are illustrated in Figure 2.2.

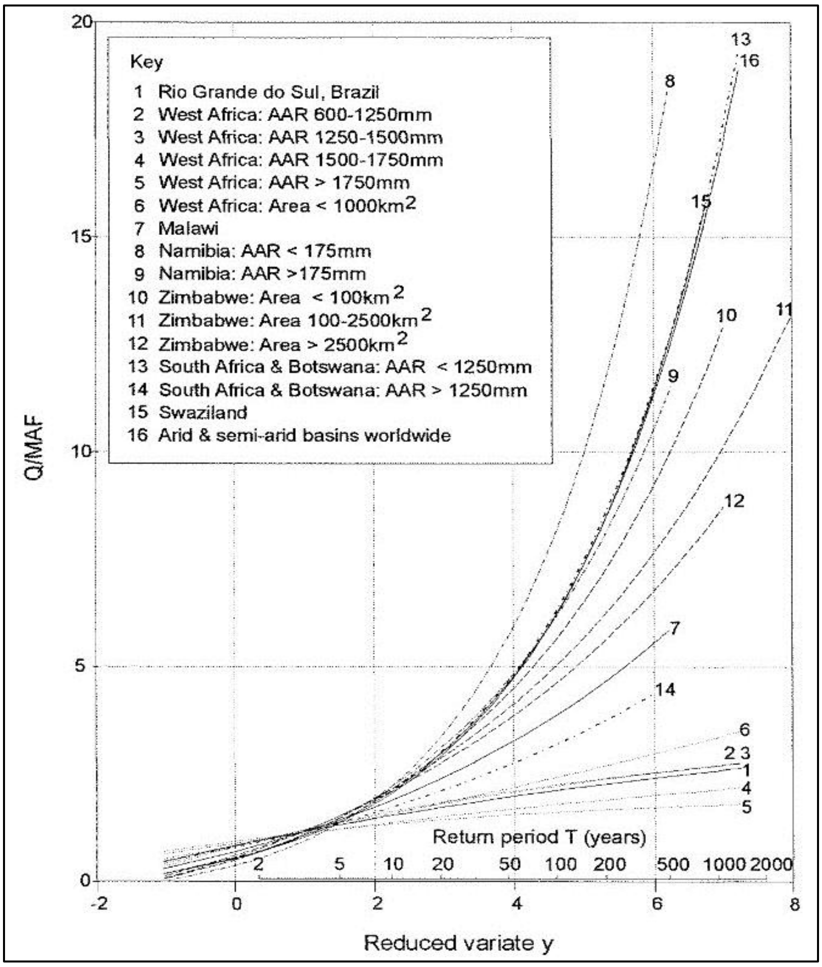

Figure 2.2 Comparison of regional flood frequency curves developed by Meigh *et al.* (1997)

2.2.2.3 Estimation of design flood peaks

In order to determine the design flood (Q_T) for a particular return period (T), the non-exceedence probability (P) must first be determined by using Equation 2.9:

$$P = \left(\frac{T-1}{T}\right) \qquad (2.9)$$

Thereafter the reduced variate (y) is determined using Equation 2.10:

$$y = -\,ln\left[-\,ln\ P\right] \qquad (2.10)$$

The standardised flood peak (q_T) is then determined using Equation 2.11:

$$q_T = u + \frac{\alpha\left(1-e^{-ky}\right)}{k} \qquad (2.11)$$

where

u = the intercept of the regional curve,

α = the scale parameter of the regional curve, and

k = the curvature of the regional curve.

Finally the design flood (Q_T) can be determined for a given return period, using Equation 2.12:

$$Q_T = q_T \times MAF \qquad (2.12)$$

2.2.3 Mkhandi Method

A RFFA developed for southern Africa is reported by Kachroo *et al.* (2000) and Mkhandi *et al.* (2000). Kachroo *et al.* (2000) details the delineation of homogeneous regions in southern Africa, while Mkhandi *et al.* (2000) details the identification of the appropriate regional distributions.

2.2.3.1 Delineation of homogeneous regions

The primary hypothesis adopted by this method is that a region is homogeneous if the gauging sites within the region yield AMF data belonging to a single parent distribution, and the best distribution was selected by the similarity between the L-Coefficient of variation (L-CV) from the historical data at each site and the L-CV from synthetic sequences generated from a parent distribution (Kachroo *et al.*, 2000). This study utilised data from 77 sites in Tanzania and the delineation procedures that were used were applied to several other countries in southern Africa. The following candidate distributions were assessed: Wakeby (WAK), GEV, Extreme Value Type 1 (EV1), Pearson Type 3 (P3), Gamma 2-parameter (G2), LP3, Log-Normal (LN), and the Kappa distribution.

The delineation of homogeneous regions involved three steps (Mkhandi *et al.*, 2000): (a) identifying possible homogeneous regions by geographic information, (b) modifying the regions identified in Step (a) after a check for similarity, using the statistics of observed flood data, and (c) confirming that the delineated regions are homogeneous, using a test for homogeneity, which is detailed below:

(i) Determine the L-CV values using the AMF data in the region for each station and plot these values on EV1 plotting paper.

(ii) Assume a parent distribution P and generate synthetic sequences from this distribution. The length of each synthetic sequence is determined by the product of the number of stations in the region and each stations record length.

(iii) Calculate the L-CV values for each site in the region using the synthetically generated data.

(iv) Determine the lower and upper limit boundaries for each order of L-CV created. Two categories of limits are considered: The first category of limits were defined as the minimum and maximum values of the n^{th} order L-CV values generated from the synthetic sequences and the boundary limits of the second category are equal to the mean plus twice the standard deviation (upper limit) and the mean minus twice the standard deviation (lower limit) (Mkhandi *et al.*, 2000).

(v) If the L-CV values of the AMF data fall within the upper and lower limit boundaries, then the stations within a particular region are considered to be statistically homogeneous with regards to the parent distribution P.

In addition, the Hosking and Wallis (1993) Homogeneity Test was applied to the regions in Tanzania. This test assumes that a region is homogeneous if there is a similarity between the variance of L-CV from the historical dataset and the variance of L-CV from the synthetic sequences, computed by using a kappa distribution. If the variance of the observed L-CV lies within the sampling distribution, then the region is homogeneous. The test assumes that all flood distributions are represented by the Kappa distribution. The following equation is used in the Hosking and Wallis (1993) Homogeneity Test:

$$H = \frac{(V - \mu_v)}{\sigma_v} \tag{2.12}$$

where V is the variance of the observed L-CV values, μ_V is the mean of generated L-CV values and σ_V is the standard deviation of the generated L-CV values.

A region is classified as homogeneous or heterogeneous according to the following criteria: $H < 1$ = acceptably homogeneous, $1 \leq H < 2$ = possibly homogeneous and $H > 2$ = definitely heterogeneous (Hosking and Wallis, 1997).

2.2.3.2 Identification of regional distributions

Mkhandi *et al.* (2000) describe the selection of an appropriate flood frequency distribution, using simulations. A predictive simulation ability test was performed, which involved a comparison of the quantile estimates from generated samples with known population quantiles.

As illustrated in Figure 2.3, 13 homogeneous regions were identified by Mkhandi *et al.* (2000) in South Africa, utilising 316 stations. The LP3 distribution fitted by the MOM was used for Region SAF 13, while the remaining regions utilised the P3 distribution fitted by PWM. The stations utilised in the Mkhandi *et al.* (2000) study have catchment areas that range from 10 km^2 to 337 590 km^2. In addition, it can be seen that there are more stations used in the study from the eastern half of the country than in the western half.

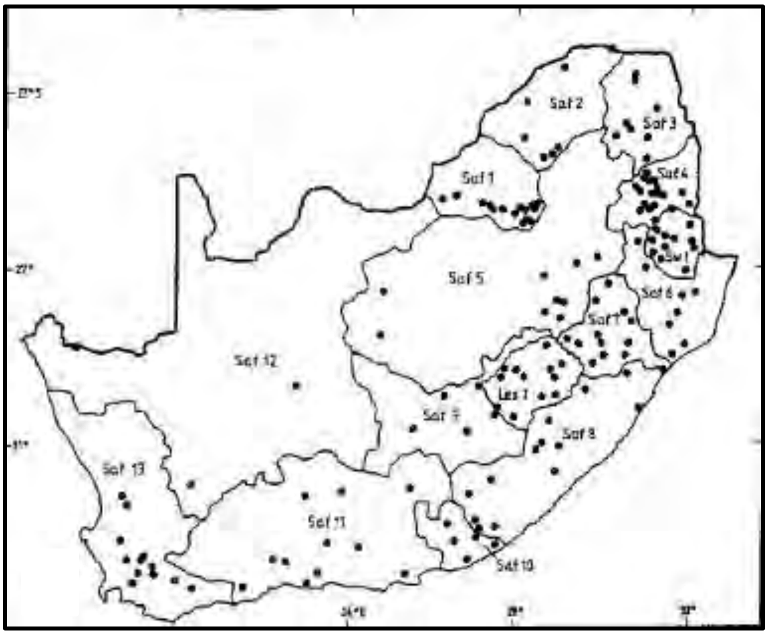

Figure 2.3 Map of homogeneous regions identified for South Africa (after Mkhandi *et al.*, 2000)

2.2.3.3 Estimating the index flood and design flood discharges

Mkhandi *et al.* (2000) derived regionalised parameters for each of the 13 homogeneous regions identified in South Africa. The data in the Mkhandi *et al.* (2000) study was scaled, using the *MAF* as an index. In order to determine the *MAF* at an ungauged site, Equation 2.13 was utilised:

$$MAF = CONSTANT \times AREA^{EXPONENT} \qquad (2.13)$$

where

MAF	=	mean annual flood ($m^3.s^{-1}$),
$CONSTANT$	=	regionalised parameter derived from Table 2.2,
$AREA$	=	catchment area (km^2), and
$EXPONENT$	=	regionalised parameter derived from Table 2.2.

Design floods were calculated in the Mkhandi study, using Equation 2.14:

$$Q_T = \bar{x} + \sigma \times K_T \qquad (2.14)$$

where

Q_T = design flood ($m^3.s^{-1}$),

$\bar{x}$ = mean of the AMS ($m^3.s^{-1}$),

σ = standard deviation of the AMS, and

K_T = Pearson Type 3 frequency factor.

The average coefficients of variation (CV), average coefficients of skewness (CS) and regionalised constants and exponents are presented in Table 2.2.

Table 2.2 Regionalised parameters derived for the 13 homogeneous flood regions identified by Mkhandi *et al.* (2000)

Region	Drainage Basins	Constant	Exponent	CV	CS
SAF1	A1-A3	1.920	0.579	092	2.42
SAF2	A4-A7	0.706	0.601	1.17	2.68
SAF3	A8-A9,B6-B9	10.63	0.354	1.20	2.25
SAF4	X1-X4	0.574	0.766	1.00	2.79
SAF5	B1-B5,C1-C9	5.342	0.445	1.04	2.17
SAF6	W4-W5	0.544	0.903	1.63	4.01
SAF7	V1-V7	4.974	0.540	0.72	2.02
SAF8	R1-R4,S1-S7,T1-T9,U1-U8	2.835	0.618	0.92	2.20
SAF9	D1-D2	5.562	0.560	0.75	1.89
SAF10	P1-P4,Q6-Q9,R5	7.924	0.426	1.76	3.29
SAF11	J1-J4,K1-K9,L1-L7,M1-M3,N1-N4	7.450	0.407	1.16	2.46
SAF12	D4-D8,F1-F3	2.234	0.518	0.88	1.38
SAF13	E1-E4,F4-F6,G1-G4,H1-H9	5.857	0.500	0.48	0.72

The results of this study indicate that the most suitable procedures for estimating design floods using a regional approach are the LP3/MOM and/or the P3/PWM procedures, as the bias produced by these procedures were the lowest for different return periods and sample sizes.

2.2.4 Kjeldsen Method

This method was developed by Kjeldsen *et al.* (2002) in a study conducted in the KZN Province, using flood data from rivers that have not been significantly impacted by anthropogenic activities. The procedure involved the use of an index flood method, to identify two relatively homogeneous regions in the KZN Province.

The AMS from 29 gauging weirs in KZN were used in the study. The site characteristics used were similar to those utilised in a study by Acreman and Sinclair (1986). The catchments were delineated, using ARC/INFO (ESRI, 1991), and a 200 x 200 m Digital Elevation Model (DEM) was created. The site characteristics considered as potential predictor variables of the index flood included the *Gravelius' Compactness Coefficient (GCC)*, *altitude (ALT)*, *rainfall concentration (CONC)*, *Mean Annual Precipitation (MAP)*, *soil characteristics (SOIL)*, *mean catchment slope (MCS)*, as well as *latitude (LAT)* and *longitude (LONG)*.

2.2.4.1 Identification of homogeneous regions

It is recommended by Hosking and Wallis (1997) that methods used in identifying potentially homogeneous regions should utilise site characteristics, which then allows for the independent assessment of the regions for homogeneity, using data statistics from the sites in the region. For this study, the Hosking (1996) Method, which involves the K-means procedure was utilised in the cluster analysis. In order to determine the heterogeneity of the regions under study, the Hosking and Wallis Homogeneity Test (1993) was utilised, which has been described in detail in Section 2.2.3.1.

2.2.4.2 Clustering of stations

The 29 flow gauging stations used by Kjeldsen *et al.* (2002) resulted in an H statistic value of 5.28, indicating that the KZN region is definitely heterogeneous. Therefore, the region was further subdivided into more homogeneous regions. The use of the CONC variable was used to delineate two acceptably homogeneous regions, as illustrated in Figure 2.4, i.e. H = 0.33 for Region 1 and H = -0.19 for Region 2. Region 1, characterized as the coastal and midlands area of KZN, contained 12 catchments, while Region 2, characterized as the mountainous Drakensberg area in the west and north-western regions of KZN, contained 17 stations. Kjeldsen *et al.* (2002) also illustrated that the RMF K-Region 5.0, which has been utilised by van Bladeren (1993) and Görgens (2007) can be classified as being acceptably homogeneous, the RMF K-Regions 5.2 and 5.4 utilised by van Bladeren (1993) and Görgens (2007) can be classified as being possibly homogeneous and regions SAF 7 and SAF 8 utilised by (Kachroo *et al.*, 2000; Mkhandi *et al.*, 2000) are acceptably homogeneous and possibly heterogeneous respectively.

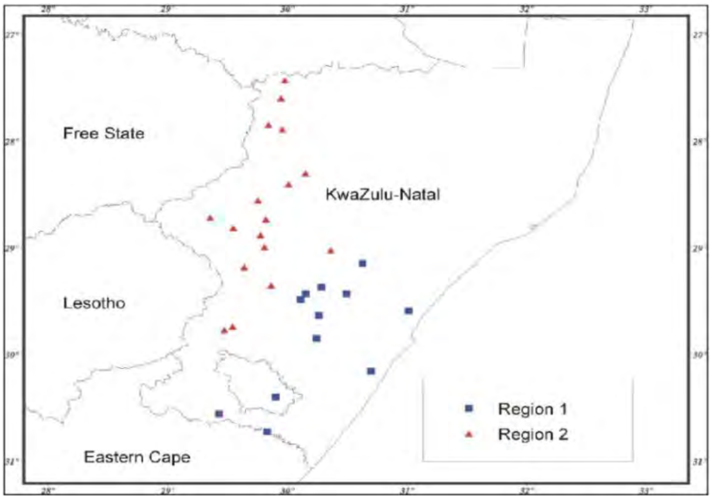

Figure 2.4 Map of homogeneous regions in KZN and the Kovacs flood regions (Kjeldsen *et al.*, 2002)

2.2.4.3 Regional flood frequency distribution

The frequency distribution that was chosen for this study was determined by using a Goodness-of-Fit Test and an L-moment diagram. The L-moment diagram illustrated the sample L-CV plotted against sample L-CS for the homogeneous regions of KZN, as well as the relationship between population L-CV and L-CS, for several candidate distributions. These distributions included the General Normal (GNO), General Pareto (GPA), P3, GEV and GLO.

In addition to the L-moment diagrams, a Z^{DIST}-Goodness-of-Fit Test developed by Hosking and Wallis (1997) was used, as shown in Equation 2.15. This test involves a comparison between the population L-kurtosis and the sample L-kurtosis for the candidate frequency distributions.

$$Z^{DIST} = \frac{T_4^{DIST} - t_4^R + B_4}{\sigma_4} \tag{2.15}$$

where

Z^{DIST} = goodness-of-fit for candidate distribution = *DIST*,

T_4^{DIST} = the population L-kurtosis of the selected distribution,

$t^R{}_4$ = the regional average sample L-kurtosis,

B_4 = the bias of the regional average sample L-Kurtosis, and

σ_4 = the standard deviation of regional average sample L-kurtosis.

It is recommended by Hosking and Wallis (1997) that a distribution should only be considered if $Z^{DIST} \leq 1.64$. The two homogeneous regions within this study were tested, using Equation 2.15 for the following frequency distributions: GNO, GLO, P3, GEV and GPA. It was found that in Region 1, the L-moment diagram suggested that both the GPA and GNO distribution would be suitable for RFFA, while the Goodness-of-Fit Test accepts the GLO and GEV distributions and rejected the P3, GPA and GNO distributions (Kjeldsen *et al.*, 2002). In Region 2, there was a good correlation between the L-moment diagram and the GOF test, as both procedures accepted the P3, GLO and GEV distributions and rejected the GEV and GLO distributions (Kjeldsen *et al.*, 2002).

2.2.4.4 Estimating the index flood

Kjeldsen *et al.* (2002) used the *MAF* as an index, to scale the data. The relationships developed by Kjeldsen *et al.* (2001) to estimate the *MAF* at ungauged sites for Region 1 is expressed by Equation 2.16 and for Region 2 by Equation 2.17.

$$ln(MAF) = 25.3351 + 1.0792 \times ln(AREA \times MAP) - 1.3862 \times ln(GCC) \quad (2.16)$$

$$ln(MAF) = 1.2388 + 0.7295 \times ln(AREA) - 1.2763 \times ln(GCC) \quad (2.17)$$

where

MAP = mean annual precipitation (mm),

AREA = catchment area (km^2), and

GCC = Gravelius' Compactness Coefficient.

In order to assess the regression models for the estimation of *MAF*, the models were first applied using all 29 catchments, therafter they were applied using only the catchments that fell within their region i.e. 12 and 17 catchments for Regions 1 and 2, respectively. It was found that for Region 2, the error variance decreased significantly when using the 17 catchments that fell within the region, in comparison to when all 29 catchments were used. However, Region 1 showed very small improvements when using the 12 catchments within the region, in comparison to when all 29 catchments were used. This may indicate that the current regression for Region 1 is inadequate to determine the *MAF*. Therefore, owing to the aforementioned problems that occurred during modelling, it was suggested by Kjeldsen *et al.* (2001) that the index flood method developed for Region 1 should not be used.

2.2.5 Joint Peak-Volume Method

Görgens (2007) developed procedures that link flood volume exceedence with the flood peak magnitude on a regional scale. The gauging stations used in the study underwent a data screening process, which resulted in 139 gauging stations and dam inflow records being used, as illustrated in Figure 2.5. The catchment areas of these stations range from 45 km^2 to 28 920 km^2.

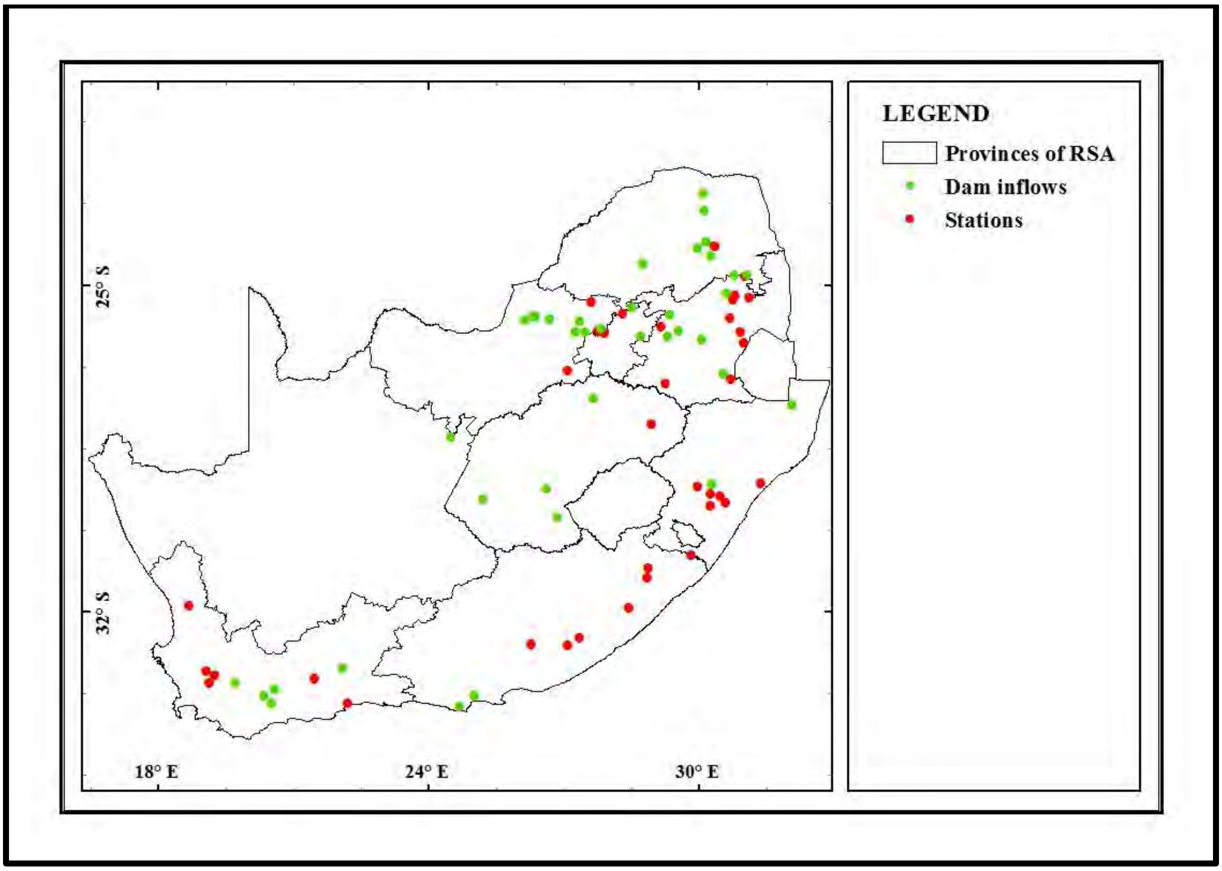

Figure 2.5 Stations analysed in the Görgens (2007) study

2.2.5.1 Quantifying flood volume based on flood peak

According to Görgens (2007), the study takes into account the “comfort zone” within which the South African design flood practice operates. This includes: (a) a focus on flood peaks, and (b) treating the design flood volume as an “annual exceedence probability-neutral” entity, which refers to the probability that the flood volume will be equalled or exceeded in a given year. In identifying these “comfort zones”, the delineation of regions within South Africa utilised regions currently used in practice, i.e. the HRU (1972) Veld zones and the Kovacs (1988) Regional Maximum Flood (RMF) K-Regions. The regions were grouped into three categories, based on either the Veld zone or K-Region. For the Veld zones, the regions were classified as: Groups A (Veld zone 2), B (Veld zones 4, 5, 6, 7) and C (Veld zones 1, 3, 8, 9), as illustrated in Figure 2.6 (Görgens, 2007).

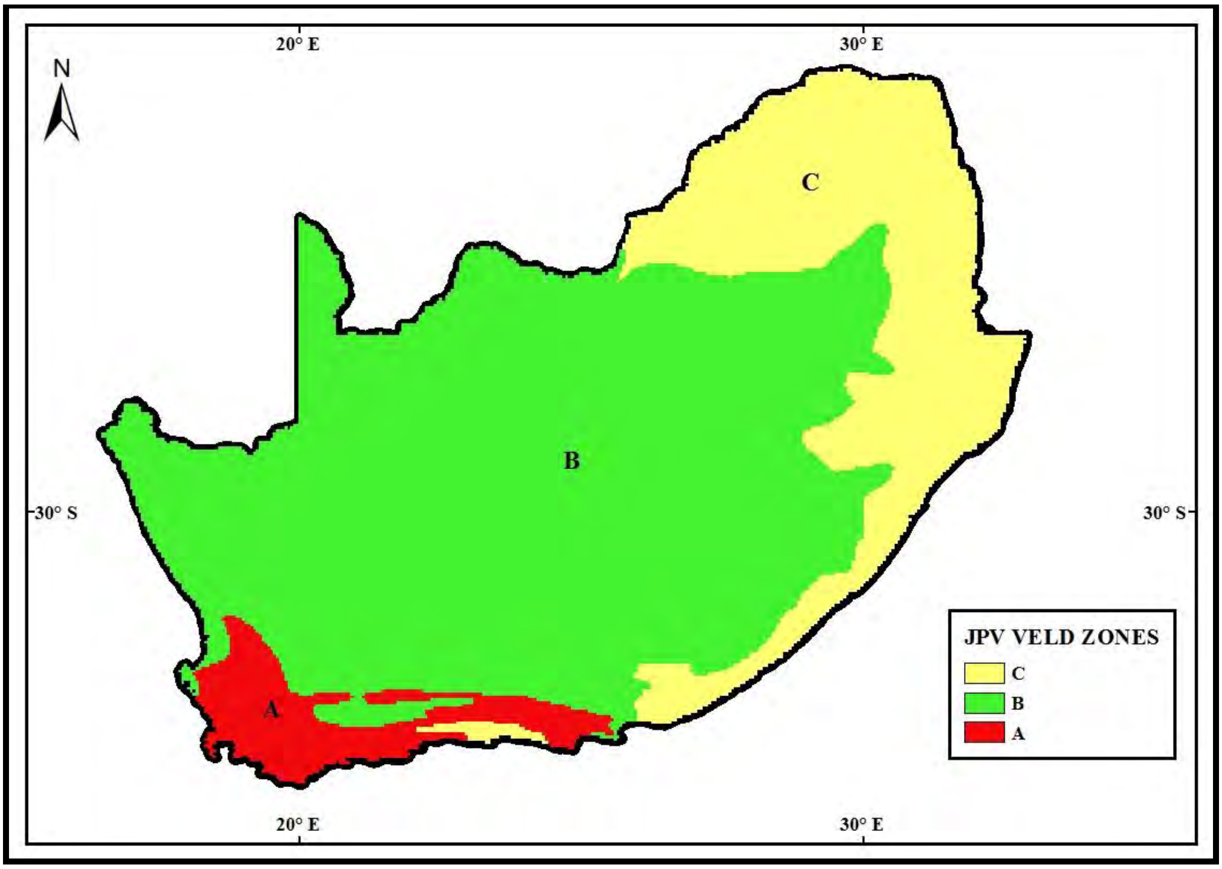

Figure 2.6 Veld zone pooling groups utilised in the JPV Method (Görgens, 2007)

With regards to K-Regions, the regions were classified as: High-K (K>5), Mid-K (K=5) and Low-K (K<5), as illustrated in Figure 2.7 (Görgens, 2007).

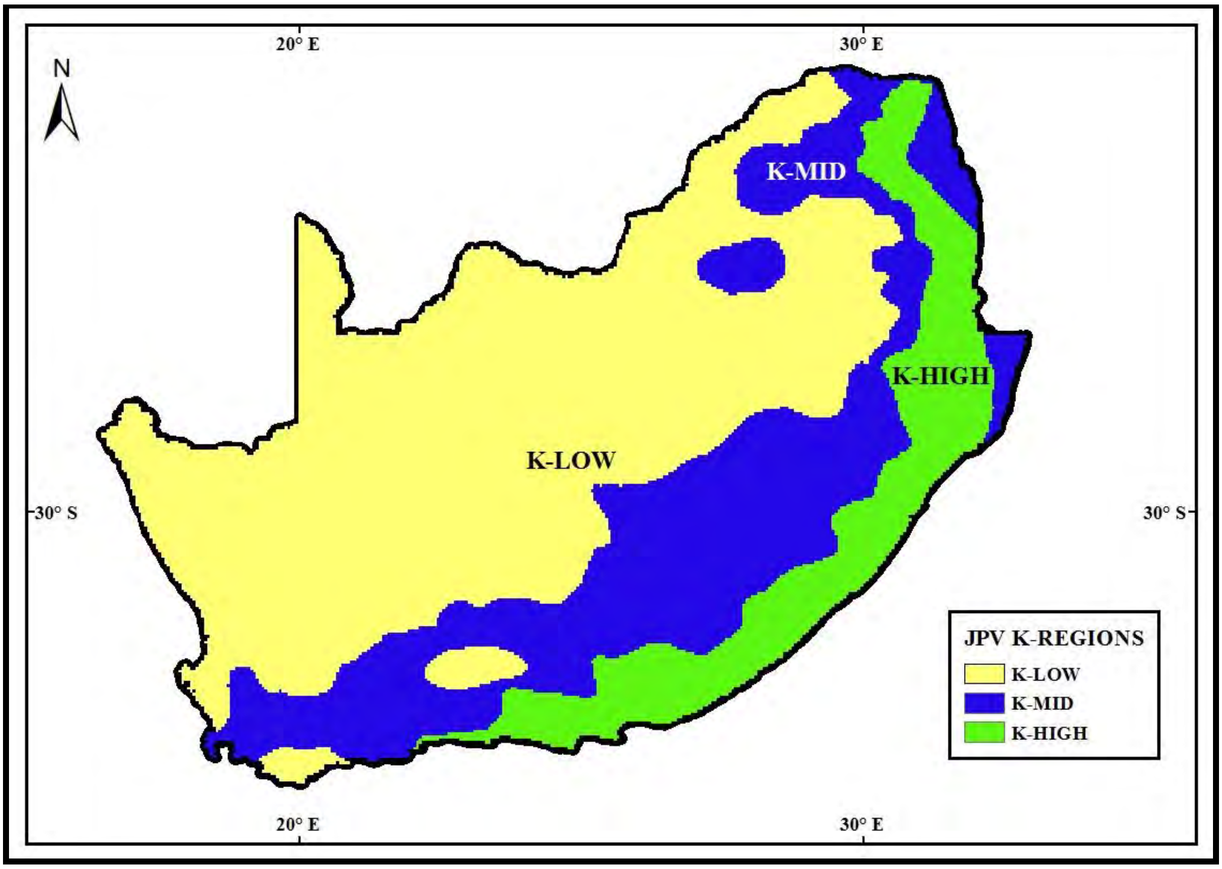

Figure 2.7 K-Region pooling groups utilised in the JPV Method (Görgens, 2007)

2.2.5.2 Pooling data

The pooling groups for the catchments within the study area can fall under either the fixed Pooling-groups ("wide" Pooling) or the adjustable Pooling-Groups ("narrow" Pooling). Fixed pooling groups include the three Veld zone groups and three K-Region groups. In the case of Adjustable Pooling-Groups, a choice can be made from the six fixed groups, mentioned above, to create a more "narrow" pooling group. This choice is made by choosing catchments with the lowest similarity distance value, relative to the site (Görgens, 2007).

The similarity distance measure uses easily quantifiable catchment descriptors to determine the similarity of flood responses from catchments. In the Görgens (2007) study, the following descriptors were taken into account: Catchment area in km^2 (*AREA*), average main channel slope (S) and mean annual runoff in mm extracted from the WR90 National Water Resource information (*MAR90*) study (Midgley *et al.*, 1994). The pooled skewness and coefficient of variation were also determined and calculated as the weighted average of the individual values for skewness and coefficient of variation for the catchments within the different pooling groups (Görgens, 2007).

2.2.5.3 Estimating the index flood

The index flood utilised by Görgens (2007) was the mean of the AMS and the relationships between the catchment descriptors (Des_i) and the index flood was expressed using a multiplicative model. The equation is expressed mathematically below and the coefficients of this regression are presented in Table 2.3:

$$ln(Index\ Flood\ Peak) = B_o + B_1\, ln(Des_1) + B_2\, ln(Des_2) + ... + B_4\, ln(Des_4) \qquad (2.18)$$

Table 2.3 Coefficients for the determination of the index floods (after Görgens, 2007)

Pooling-Group	Constant (B_0)	ln Area (B_1)	S (%) (B_2)	ln MAR90 (B_3)	"Region" (B_4)	R^2
			$ln\ (\mu_Q)$			
Low K-Region	-1.63	0.55	0.05	0.45	0.42	0.76
Mid K-Region	-2.56	0.69	-0.21	0.50	0.38	0.79
High K-Region	-1.14	0.77	0.38	0.04	0.14	0.84
Veld zone A	-1.83	0.52	-0.29	0.89	-0.17	0.89
Veld zone B	0.30	0.51	-0.56	0.28	0.06	0.84
Veld zone C	-1.52	0.68	0.16	0.16	0.27	0.73
			μ_{lnQ}			
Low K-Region	-6.65	0.94	1.69	0.82	-0.07	0.58
Mid K-Region	-5.73	0.84	-0.13	0.77	0.50	0.79
High K-Region	-3.24	0.86	0.31	0.23	0.20	0.88
Veld zone A	-6.08	0.82	-0.21	1.26	-0.17	0.96
Veld zone B	-3.64	0.70	-0.35	0.48	0.44	0.64
Veld zone C	-3.63	0.78	0.16	0.41	0.11	0.72

2.2.5.4 Estimation of design flood peaks

In order to determine a pooled estimation of the design flood peak, the GEV and LP3 distributions were fitted using frequency factors. The design flood equation, to determine the flood peaks, is as follows (Görgens, 2007):

$$Q_T = \mu_Q + K_{g,T} \cdot \sigma_Q \quad (2.19)$$

where

Q_T	=	design flood peak,
T	=	Recurrence Interval (RI) or 1/AEP,
μ_Q	=	mean,
g_Q	=	skewness,
σ_Q	=	standard deviation, and
$K_{g,T}$	=	frequency factors.

In the case of ungauged catchments, values for μ_Q and σ_Q are estimated using the catchment descriptors and pooled statistics. The standard deviation is estimated using the estimated mean flood peak and coefficient of variation, as well as a pooled estimate of the coefficient of skewness. The JPV methodology allows for the estimation of design flood peaks at ungauged sites, using pooled data. This is achieved through the regionalised and standardised hydrographs for South Africa, developed by Görgens (2007). The following considerations were taken into account in the development of these hydrographs: K-Regions and Veld zones, smaller and larger catchments (<1000 km^2 and >1000 km^2, respectively), the magnitude of standardised flood peaks (5 ranges), as well as the identification of several 'typical' hydrograph shapes.

Design hydrographs for the 50-year return period were generated via the JPV approach and compared to a Unit-graph based approach for two differing catchments (Görgens, 2007). The results indicated that the Unit-graph based approach was not conservative enough, producing exceedence frequencies that were greater than 75%, whereas the JPV method produced more conservative and acceptable exceedence percentiles that were between 20% and 50% (Görgens, 2007).

2.2.6 Haile Method

Haile (2011) conducted a regional flood frequency study in southern Africa. The analysis and regionalisation was performed using a combination of the index flood method and L-moments.

2.2.6.1 Developing homogeneous regions

A geographical regionalisation approach was utilised by Haile (2011) to develop homogeneous regions. Catchment characteristics, such as topography, soil and climate, amongst others, were utilised to divide a region into a smaller area. Homogeneous regions were determined by Haile (2011) through the following procedure:

a) Homogeneous regions were identified using geographic information, such as drainage characteristics.
b) Each region identified in Procedure (a) was checked for heterogeneity by its statistical data behaviour.
c) The homogeneity of each region was determined and regions that were not homogeneous were then separated into two or three groups.

The Homogeneity Test recommended by Hosking and Wallis (1997) in Equation 2.12 was applied by Haile (2011). Figure 2.8 illustrates the homogeneous flood regions delineated by Haile (2011).

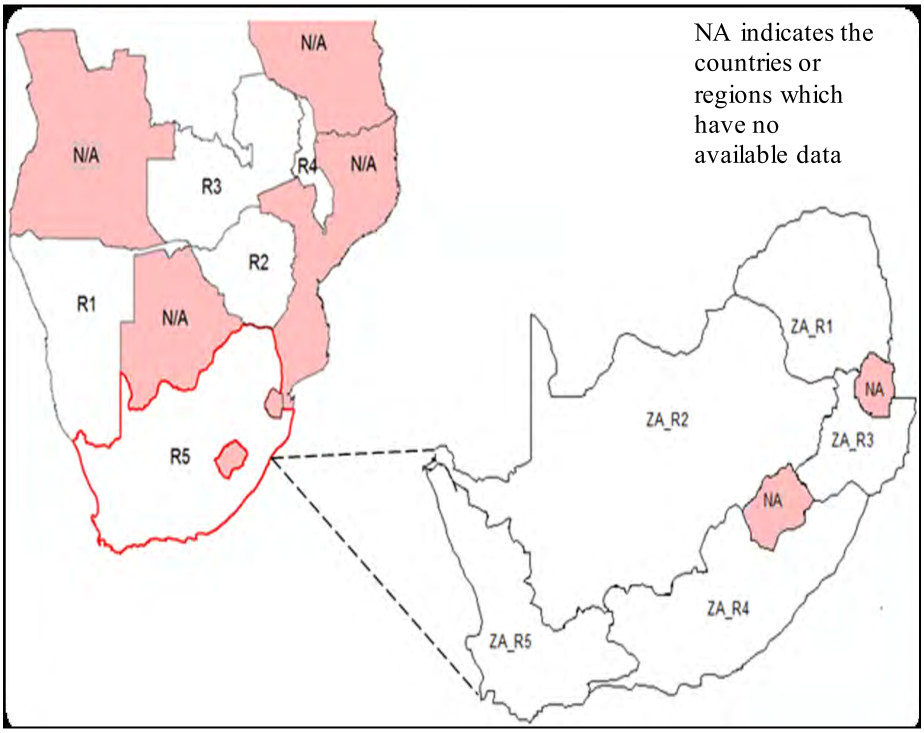

Figure 2.8 Map of homogeneous flood regions delineated by Haile (2011)

2.2.6.2 Fitting regional data to the appropriate frequency distribution

Haile (2011) assessed a number of frequency distributions based, in part, on the use of these distributions in prior studies in southern Africa. These distributions include EV1, Exponential (EXP), LN, P3, GPA, GLO and GEV. The following procedures were carried out to fit the frequency distributions to the observations of the homogeneous regions:

a) The observed series were normalised, using the median as the index flood.
b) An L-moment diagram was utilised.
c) Candidate frequency distributions were chosen, using the L-moment diagram.
d) Statistical Goodness-of-Fit tests were used to confirm the selected distribution for a region.

Having determined the candidate distributions using an L-moment diagram, it was necessary to perform a Goodness-of-Fit Test. When performing a RFFA, two techniques are commonly used to examine whether a given frequency distribution fits the data better than others. These methods are the Anderson-Darling Goodness-of-Fit Test and the Z^{DIST}-Goodness-of-Fit-Test, shown in Equation 2.15. Haile (2011) used both methods as the advantages of the one method complemented the deficiencies of the other. The Z^{DIST}-Goodness-of-Fit-Test was able to choose the best-fitting regional distribution, however it could not test two parameter distributions. On the other hand, the Anderson-Darling Goodness-of-Fit Test could test all distributions, but was not as applicable in determining the best-fitting regional distribution.

2.2.6.3 Developing regional growth curves

Growth curves for southern Africa were developed and applied, using the following procedures (Haile, 2011):

a) Using the regional and theoretical relationships, the scale (μ), shape (k) and location (α) for the best-fitted distributions were determined.
b) From the 2- to 500-year return periods, the standardised quantile estimates were computed for each region, using the model parameters.
c) The standardised at-site quantile floods were then re-scaled from the regional quantile flood, using equation 2.20.

$$Q_{T(i)} = \mu_i \, x \, X_T \tag{2.20}$$

where

$Q_{T(i)}$ = quantile flood ($m^3.s^{-1}$),

μ_i = index flood ($m^3.s^{-1}$), and

X_T = regional quantile for return period (T) ($m^3.s^{-1}$).

2.2.6.4 Estimating the index flood

Haile (2011) developed relationships between catchment characteristics and the index flood i.e. the Median Estimated flood (*MEF*). Catchment characteristics included catchment area, topography and MAP, amongst other potential characteristics that could be related to the index flood. Haile (2011) utilised 459 stations across southern Africa, which, after screening, resulted in 92 stations being used for analysis in South Africa, as illustrated in Figure 2.9. The index flood was determined by the catchment area (A) and median annual flood. The equations developed to calculate the index floods in South Africa are summarised in Table 2.4. As in the case of the Mkhandi *et al.* (2000) study, the stations used by Haile (2011) are also greater in number across the eastern half of the country, as opposed to the western half.

The catchment areas of the stations analysed in South Africa range from 119 km^2 to 850530 km^2.

Table 2.4 Regression models to predict the median values from catchment area (Haile, 2011)

Region	Equation to Estimate Median of AMS	R^2
ZA_R1	MEF = 14.755 x ln(A)-49.338	0.3664
ZA_R2	MEF = 52.664 x ln(A)-340.28	0.7683
ZA_R3	MEF = 66.461 x ln(A)-395.91	0.5218
ZA_R4	$MEF = 0.6089(A)^{0.6639}$	0.5927
ZA_R5	MEF = 42.282 x ln(A) -187.1	0.8890

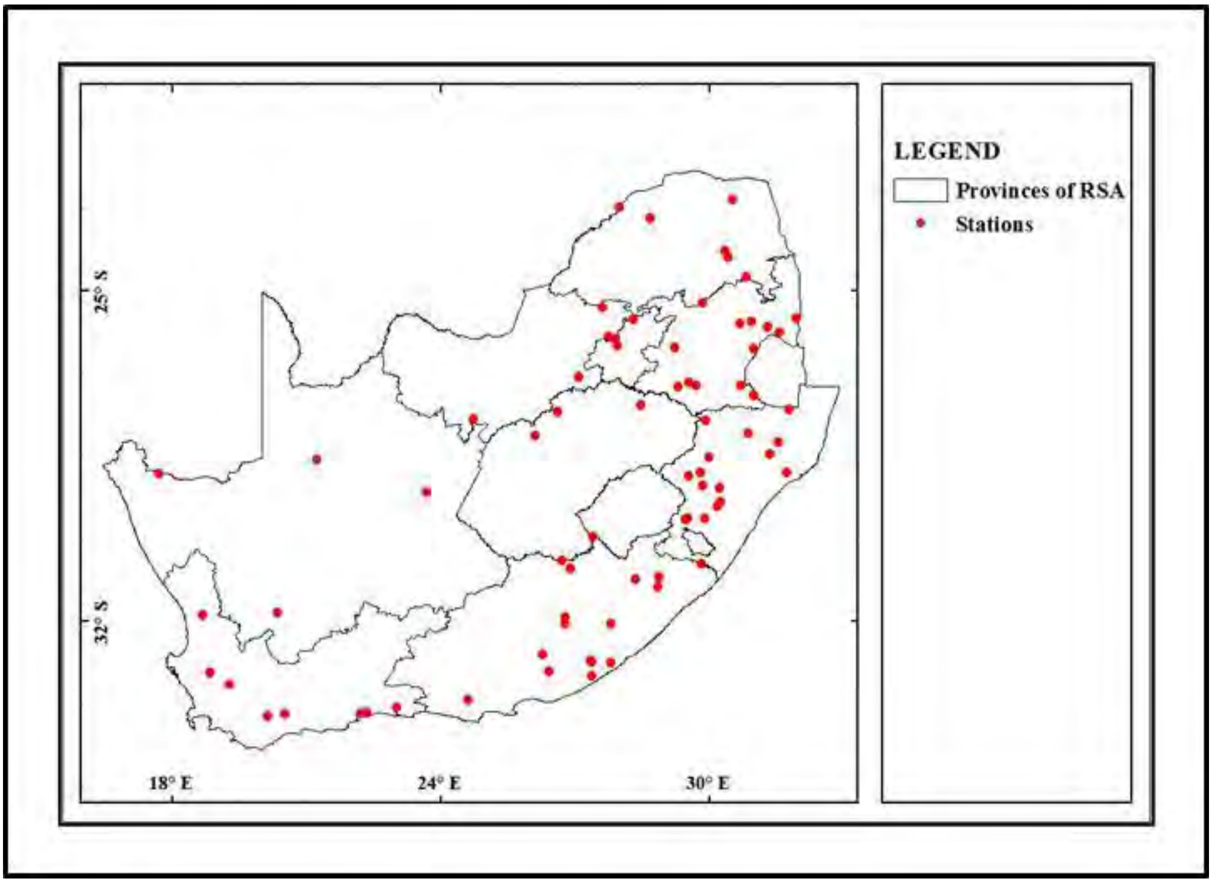

Figure 2.9 Stations in South Africa analysed in the Haile (2011) study

2.3 Synthesis of Literature

Having reviewed the literature pertaining to RFFA methods used locally and internationally, this section contains a summary of the most pertinent aspects that have been reviewed, as well as the methods that will be selected for application in this study and why they have been selected.

Often a record of sufficient length is not available at the site being investigated and thus a regional approach must be considered (GREHYS, 1996a; Viglione *et al.*, 2007). A number of studies have advocated a regional approach for obtaining more reliable design flood estimates (Wiltshire, 1986; Hosking and Wallis, 1997; Saf, 2008; Saf, 2009; Haile, 2011; Smithers, 2012). A regional approach yields more accurate estimates of the parameters of the distribution than an at-site approach, both at gauged and ungauged sites. Having reviewed the approaches to RFFA adopted internationally and locally, it is necessary to assess their performance and discuss the advantages and disadvantages of the various methods in a nationwide study.

The primary success of a RFFA approach stems from the spatial extension provided by additional records from neighbouring sites. Therefore, most regional studies focus only on a spatial extension. Gaume *et al.* (2010) recognised the importance of both a spatial extension, through the merging of data from statistically homogeneous regions, as well as a temporal extension through the utilisation of paleo-flood data and historical data. The results of this study also indicated that the inclusion of ungauged data in a RFFA approach can improve design flood estimates.

In Australia, the procedures undertaken to develop the ARR RFFE 2012 Model are outlined by Rahman *et al.* (2013). The advantage of this approach is that it uses a PRT, rather than a QRT. This technique ensures that flood quantiles increase smoothly with increasing ARI and it allows for the estimation of quantiles for any ARI within the limits of the developed RFFA method (Haddad and Rahman, 2012). The frequency distribution used by Rahman *et al.* (2013) was the LP3 distribution, which has been recommended for use in Australia by the Australian Insititution of Engineers (1987). This is a potential shortcoming as there may be other distributions that may fit the data better or the best distribution to use in the six regions and four fringe zones of Australia may be different. Therefore, this method could be improved by utilising an L-moment diagram and a Goodness-of-Fit Test, to determine the best distribution to use for each homogeneous region or fringe zone in Australia.

Crochet (2012) conducted a study in Iceland, which focuses mainly on the delineation of homogeneous regions, by comparing the ROI approach to a hierarchical clustering approach. This study reported very little on other aspects of RFFA, such as data screening and determining the frequency distribution to be used and hence could be improved by a stronger focus on these aspects.

The index flood method developed by Kjeldsen *et al.* (2008) can be applied to both rural and urban catchments, as adjustments to the procedure have been developed for use in urban areas by Kjeldsen (2010), which makes it a versatile method to use.

The study conducted in Italy, as reported by Castellarin *et al.* (2012), involves a hierarchical index flood method, whereby three levels have been developed to compartmentalise the regions of Italy. An advantage of the method is that it covers the entire country and it improves estimates of moments of higher orders. However, it is not up-to-date in terms of data and it may cause certain inaccuracies in the development of regional growth curves, such as "abrupt jumps" i.e. sudden changes in the index flood at the boundary between regions. Therefore, Castellarin *et al.* (2012) concluded that there is still room for improvement.

Internationally, RFFA is often carried out, using the index flood method. The primary advantage of this approach is that it allows for design flood estimates to be made at ungauged sites, using the regional growth curve and an estimate of the index flood. The methods used internationally in developing homogeneous regions include the region of influence approach and hierarchical clustering.

For South Africa, five RFFA approaches have been reviewed. These include the van Bladeren Method, The Meigh Method, the Mkhandi Method, the Kjeldson Method, the Joint Peak-Volume (JPV) Method, and the Haile Method. In the van Bladeren (1993) Method, RFFA was only carried out in the KZN and Transkei regions. Therefore, it should not be applied outside of KZN and Transkei. The Meigh *et al.* (1997) study involved a worldwide comparison of regional flood estimation methods and delineated two regions in South Africa based on MAP, with regression equations developed for each region for the estimation of index floods throughout South Africa. This broad classification has the potential that the two regions are not homogenous and the method can be refined by more detailed regionalisation. The regionalisation method developed by Mkhandi *et al.* (2000) involved the use of a proposed Homogeneity Test, which was determined to be more lenient than the Hosking and Wallis Homogeneity Test. This may result in areas being classified as acceptably homogenous, when they are not, which may yield further inaccuracies in design flood estimates. Kjeldsen *et al.* (2002) illustrated that regions SAF 7 and SAF 8 utilised by Mkhandi *et al*. (2000) are acceptably homogeneous and possibly heterogeneous, respectively. The Kjeldson Method was only applied in KZN and two regions were delineated (Kjeldsen *et al.*, 2002). However, there were a number of modelling problems that still needed to be addressed. Therefore, Kjeldsen *et al.* (2001) recommended that the index flood method developed for region 1 should not be used. The JPV Method developed by Görgens (2007) has the advantage of being able to determine the exceedence frequency of any design flood hydrograph volume, using any method of determining the design flood peak. The drawback of this method is that, in delineating the regions of South Africa, only the existing K-Regions (Kovacs, 1988) and the HRU (1972) Veld zones were utilised. The Haile (2011) Method, which covered southern Africa, was the most detailed study, in comparison to the other South African studies reviewed. One of the major advantages of this method was that it entailed a thorough data screening process, which is lacking in many South African studies on design flood estimation. The results of this study were satisfactory, but can be improved on by adding the records of another seven years, as the data from 1969 to 2008 was used in the study.

The international approaches that are currently being utilised can provide useful techniques that can be applied in South Africa. However, these approaches have been developed specifically for the climatic and hydrological conditions of the region under investigation and cannot be transferred directly to South Africa. Therefore, the local approaches that have been developed should be considered for application in a nationwide study of South Africa. Owing to the van Bladeren Method being applied only in the KZN and Transkei regions and the Kjeldson Method being applied only in KZN, this comparative study of the performance of RFFA methods in South Africa will only apply the Meigh, Mkhandi, JPV and Haile methods.

The method that produces the best results throughout the country, with acceptable inaccuracies and errors, can be selected as the method to be used in South Africa. However, if none of the methods produce satisfactory results, then a new procedure should be developed, that will include the merits and drawbacks of both the international and local approaches, to produce a RFFA method that is best suited for wide application in South Africa.

3. STUDY AREA

This chapter provides a general description of South Africa as the study area, including the primary drainage regions and the flow gauging structures across the country that record river stage from which discharge is derived. The prevailing climate and hydrological conditions of South Africa are also discussed.

3.1 General Description

As illustrated in Figure 3.1, the Republic of South Africa (RSA) is a country located at the southern end of the African continent. The entire region covers a surface area of 1 219 602 km^2 and extends longitudinally from 17°E to 33°E and latitudinally from 22°S to 35°S (Tibane and Vermeulen, 2014). The country shares boundaries with Zimbabwe, Namibia, Mozambique, Botswana and Swaziland, while the Kingdom of Lesotho is found within South Africa (Tibane and Vermeulen, 2014).

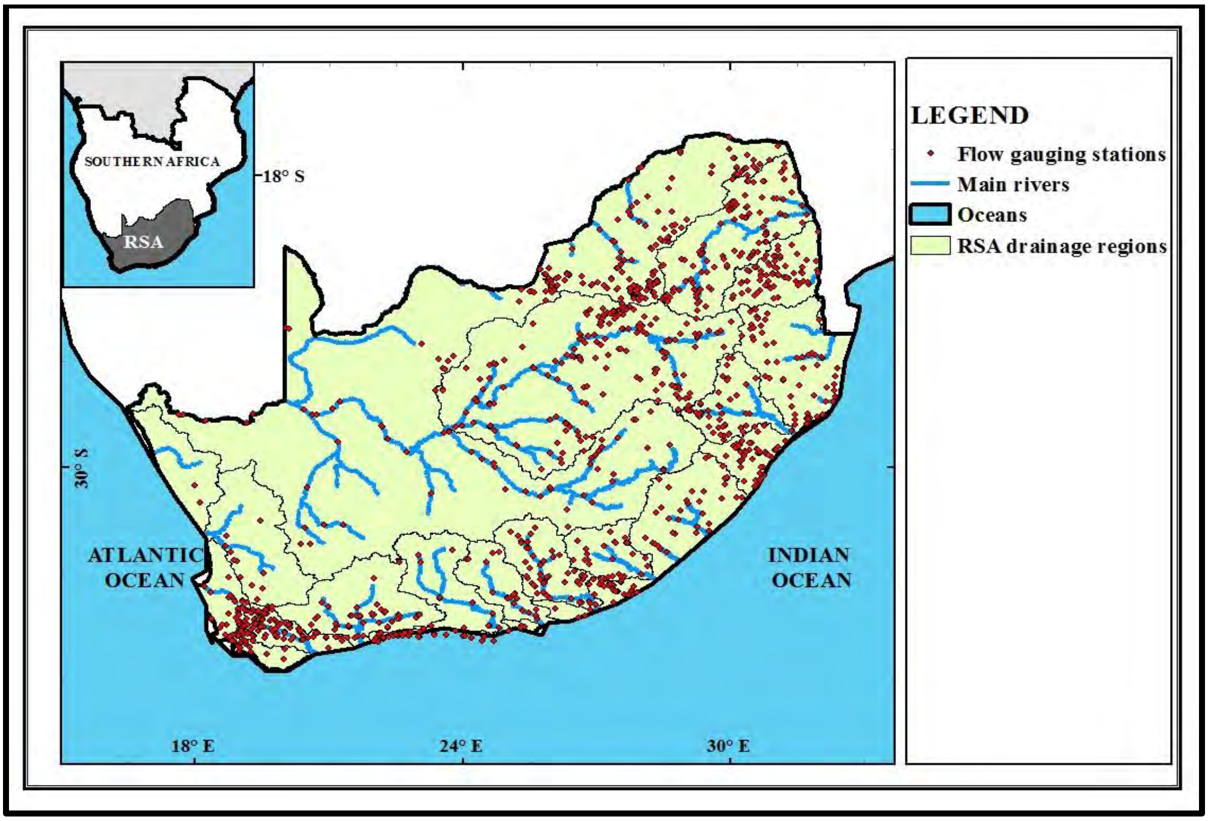

Figure 3.1 Map of the main rivers, primary drainage regions and flow gauging stations in South Africa

There are 22 primary drainage regions in South Africa, as illustrated in Figure 3.1. The primary drainage regions are further subdivided into secondary, tertiary, quaternary and quinary catchments. This study utilised the quaternary catchment level of discretisation. The DWS is the custodian of hydrological flow data in South Africa and 1458 flow gauging stations are represented in Figure 3.1.

3.2 Climate and Hydrology

The climate of South Africa is primarily subtropical, owing to its geographical location (Archer *et al.*, 2010). In addition, the Atlantic and Indian oceans on the west and east coasts of South Africa are responsible for moderating the climate of the country (Tibane and Vermeulen, 2014). The mean annual temperatures across South Africa often exceed 17 ^{o}C, with warmer temperatures experienced in the eastern half and cooler temperatures experienced in the western half of the country (Archer *et al.*, 2010).

South Africa is characterized as a semi-arid country (Mukheibir and Sparks, 2005), with an average MAP of approximately 464 mm, which is a little over half of the world average of approximately 860 mm (Tibane and Vermeulen, 2014). According to Lynch (2004), only 9% of the country receives a MAP greater than 800 mm, while 20% of the country receives a MAP of less than 200 mm. In addition, only 9% of the total rainfall in South Africa is converted into runoff, while the remaining 91% of rainfall is evaporated (Whitmore, 1971).

The estimated MAR in South Africa is approximately 50 x 10^9 m^3 (Pitman, 2011). According to Schulze (1997), the spatial variations in median annual runoff throughout South Africa is a result of a combination of differing soil types, land cover and precipitation characteristics. Low runoff producing regions are identified in the northern and western regions of South Africa, with some areas producing less than 10 mm of runoff (Schulze, 1997). However, there are regions producing much higher runoff, such as the Western Cape Mountains, the Eastern Cape, Mpumalanga and KZN (Schulze, 1997).

4. METHODOLOGY AND DATA COLLATION

This chapter details the procedures undertaken to assess the performance of the Meigh, Mkhandi, JPV and Haile RFFA Methods. Each regional approach was assessed by comparing their design flood estimates with those estimated from a frequency analysis of at-site observed flood data. For South Africa, the LP3 distribution was recommended by (Alexander, 1990; Alexander, 2001) while Görgens (2007) used both the GEV and LP3 distributions and both distributions were advocated by van der Spuy and Rademeyer (2010) for use in South Africa. Hence, in this study design floods were estimated using both the LP3 and GEV distributions.

The outline of the research methodology includes the following:

a) to collate the annual maximum flood series data for all flow gauging stations in South Africa,
b) to screen the data and select the appropriate stations for use in the study,
c) to derive relevant catchment parameters for each station,
d) to apply the Meigh, Mkhandi, JPV and Haile RFFA Methods for all selected stations,
e) to assess the performance of the methods by a comparison of the design floods estimated using the regional methods with design floods estimated using a flood frequency analysis of the observed data,
f) to assess the performance of the methods relating to the input parameters,
g) to assess any spatial variations in the performance of the methods,
h) to select the method that produces the best results for use in the different regions of South Africa, and
i) to recommend if the development of a new/revised RFFA method for use in South Africa is necessary.

4.1 Station Selection

In this study AMF data were used in the at-site frequency analysis of observed data. AMF data were obtained from the DWS for 1458 stations across the entire country. In addition to these datasets, 89 synthesised dam inflow records were obtained from the DWS and incorporated into the study. The dam inflow records were created by the DWS through a back calculation, using the catchment area and dam outflow records. In certain cases, surrounding stations were also utilised to calculate the dam inflows. The following criteria were used to select the stations to be included in the analysis:

a) It must be a river gauging station i.e. not an eye (natural spring), canal or pipeline.
b) Record lengths must be greater than, or equal to, 20 years.
c) The percentage of rating table exceedence should not equal or exceed 20% of the record.
d) The station must not be located at a dam outlet or be significantly influenced by an upstream dam.

The first exclusion criterion (a) resulted in 333 stations being removed from the analysis, due to the data being recorded from an eye, canal or pipeline. An additional 28 stations were excluded, due to the absence of an AMS dataset for those stations. Thereafter, the record lengths of the remaining 1097 stations were further investigated.

4.1.1 Record length

Smithers *et al.* (2015) highlighted the lack of stations in South Africa with record lengths greater than 50 years and Schulze (1989) highlighted the problems that can arise when extrapolating beyond the record length. For this reason, it was necessary to exclude stations with record lengths less than 20 years. Of the 1097 stations analysed, 290 stations had record lengths less than 20 years, with 807 stations having record lengths greater than or equal to 20 years. Figure 4.1 illustrates the distribution of record lengths for all 1097 river stations across South Africa. The record lengths range from a minimum of two years to a maximum of 110 years and the median record length is 33 years. As seen in Figure 4.1, the majority of stations contain data for record lengths between 21 and 40 years, with a small percentage of stations (11%) with record lengths greater than 60 years.

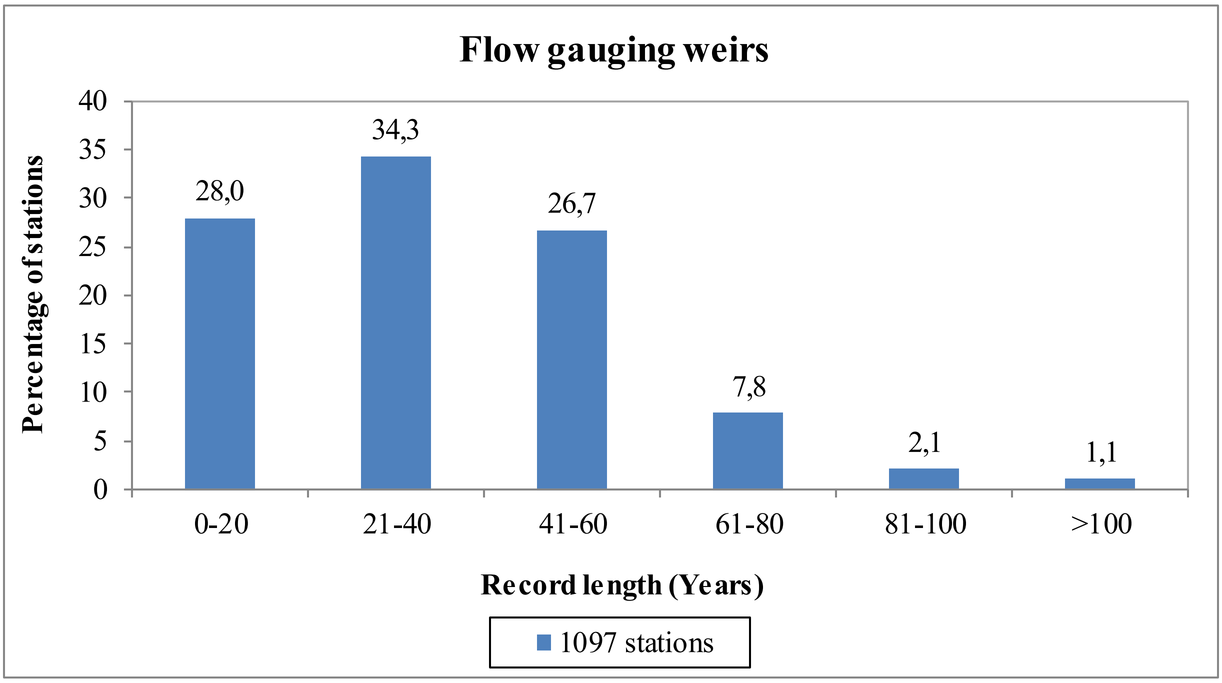

Figure 4.1 Distribution of record lengths for all flow gauging weirs across South Africa

4.1.2 Rating table exceedence

The DWS provide quality codes in each AMF dataset, which provides more information on the individual annual values in the AMF data. The station code “A” represents rating table exceedence, which refers to occurrences where the recorded stage has exceeded the maximum stage of the rating table. Therefore, the discharge for the event is recorded as the maximum discharge of the rating table. This is an inaccurate estimate, as the actual discharge may have been larger.

To analyse the frequency of rating table exceedence, a FORTRAN routine was utilised, which counted the occurrence of the station code “A” within the data set. Figure 4.2 illustrates the rating table exceedence for the total 1097 stations, as well as for the 807 stations remaining, after exclusions were made based on record length. As seen in Figure 4.2, 39% of the 1097 stations contained datasets where the rating table had not been exceeded, while 28% of the stations were exceeded for more than 20% of the record. In the case of the 807 stations with more than 20 years of record, 36% contained datasets where the rating table had never been exceeded, while at 26% of the stations the recorded stage exceeded the maximum stage in the rating table for more than 20% of the record. Of the 807 stations analysed, 217 stations were excluded due to the rating table exceedence being equal to or greater than 20% of the record length, leaving 590 stations remaining for further analysis.

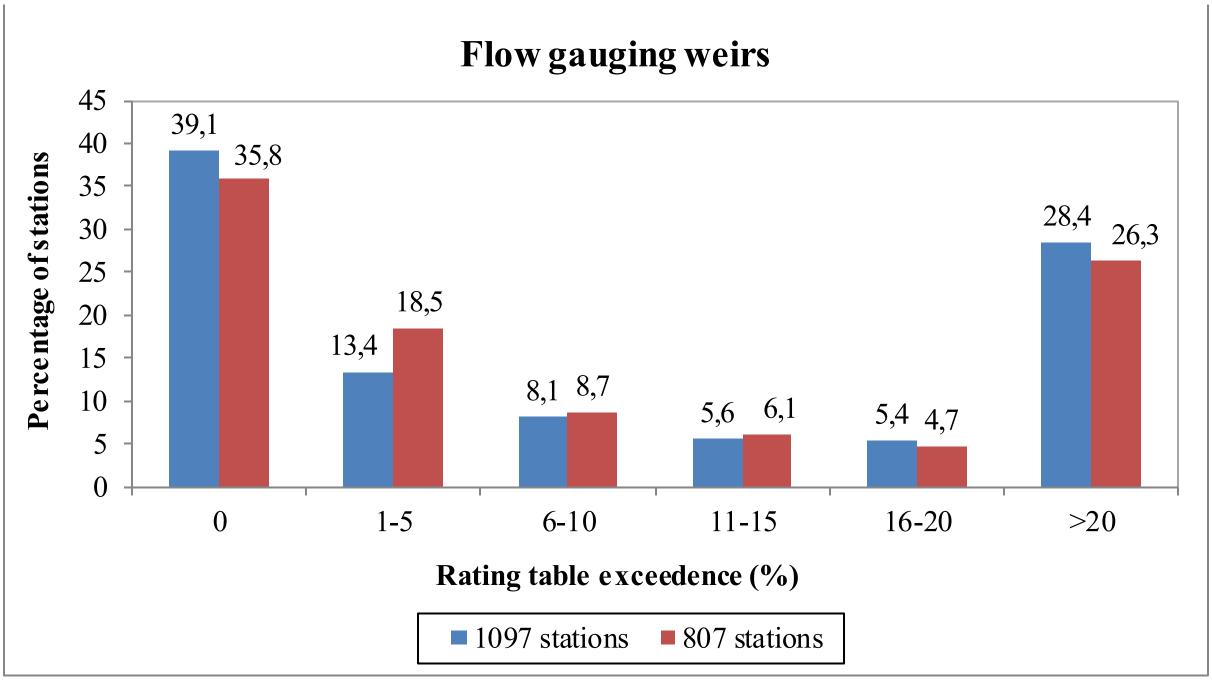

Figure 4.2 Rating table exceedence for stations in South Africa

In addition, the FORTRAN routine calculated the number of times that the maximum value in the record was repeated, as this was an indication of the rating table exceedence. This was indicated by the station code ">". The maximum values that were repeated in the dataset were treated as missing data and removed.

4.1.3 Influence of dams

The final selection criterion involved the exclusion of stations due to the influence of upstream dams; where a catchment would produce lower than expected streamflow, due to the river flow being impounded by the dam. Figure 4.3 illustrates the frequency of stations that were either upstream or downstream of a dam, located at the dam outflow or not affected by a dam at all. This is illustrated for the initial 1097 river stations, as well as for the 590 stations that had thus far been selected for analysis. It can be seen that the majority of stations to be analysed were not affected by a dam (52.9%). In addition, 9.8% of the stations were upstream of dam and were therefore also not affected by a dam. The remaining stations

(36.8%) where either downstream of a dam, or located at the dam outlet, and these 217 stations were therefore excluded.

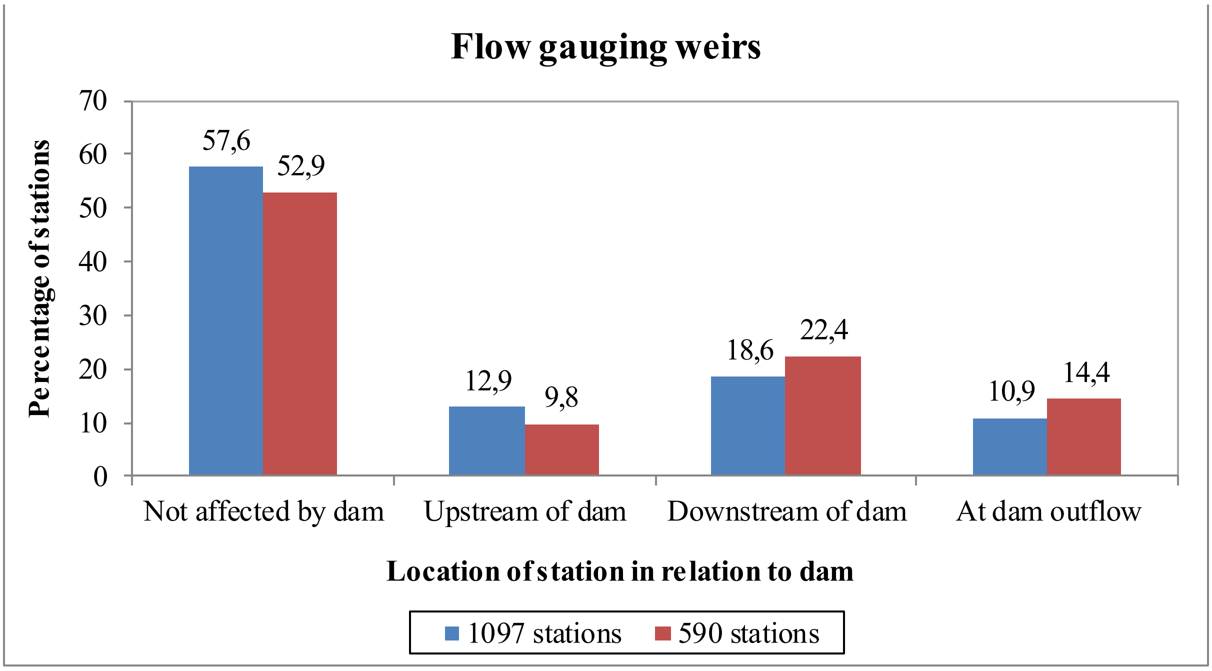

Figure 4.3 Location of stations in relation to the location of a dam

After all selection criteria had been applied, a total of 373 stations remained. As a result of further investigation, it was found that, despite the thorough screening process that had been carried out, there were still stations that were not suitable for analysis. The following section deals with the additional exclusions of stations and the final stations selected for analysis.

4.1.4 Additional exclusions and final stations

A number of stations that had passed the above screening process produced design flood estimates that were not consistent with surrounding gauges and therefore their datasets were further investigated. Table 4.1 contains the stations that were excluded and the reasons for their exclusion. As mentioned in Section 4.1.2, flood records were treated as missing data and removed from the AMS, if the maximum value was repeated in the dataset. Stations were removed if the number of missing data years, based on the above criterion, reduced the record length to below 20 years. Stations, marked as having missing data in Table 4.1, are such stations. Similarly, stations were removed if the exclusion of the zero flow values in the AMF dataset reduced the record length to below 20 years. Stations that were located at an international boundary were removed, because their catchments extended beyond South African borders and catchment parameters required for design flood estimates could not be determined.

Table 4.1 Additional exclusion of stations

Gauge	Reason for exclusion	Gauge	Reason for exclusion
A3H001	Missing data	E3H001	Unrealistic data
A3H016	Missing data	E3H002	Missing data
A3H030	Unrealistic data	F5H001	Zeroes
A5H003	At international boundary	G1H014	Missing data
A5H006	At international boundary	G1H023	No discharge information
A6H002	No discharge information	G1H024	No discharge information
A6H023	Zeroes	G1H042	No discharge information
A7H008	At international boundary	G1H060	Compensation water
B8H001	Missing data	G1H061	Inlet tunnel
C1H017	No discharge information	G1H062	No discharge information
C2H004	Missing data	G1H064	Compensation water
C2H014	Missing data	H3H005	Zeroes
C2H015	Missing data	H4H008	Zeroes
C2H021	Missing data	J2H006	Unrealistic data
C2H032	Unrealistic data	Q1H013	Zeroes
C2H067	Zeroes	Q3H004	Unrealistic data
C5H020	Missing data	R1H007	Missing data
C6H007	Unrealistic data	R1H008	Missing data
C9H005	Missing data	R1H009	Missing data
C9H026	Missing data	R1H012	Missing data
D3H015	Zeroes	T3H008	Missing data
D8H003	At international boundary	T5H002	Missing data
D8H004	At international boundary	U4H002	Missing data
D8H005	At international boundary	U4H004	Missing data
D8H008	At international boundary	W1H018	Missing data
D8H009	At international boundary	X1H024	Missing data

After the 52 stations listed in Table 4.1 were removed, the total number of stations to be analysed was reduced to 321 stations. In addition, the 89 dam inflow datasets that were analysed brought the final number of flow records used in this study to 410. An inventory of these stations is provided in Appendix A. Figure 4.4 illustrates the distribution of record lengths for the final 321 river stations, as well as the 89 dam inflow datasets that have been analysed in this study. The record lengths for the river stations ranged from a minimum of 20 years to a maximum of 110 years, with a median record length of 42 years. The record lengths for the dam inflow datasets ranged from a minimum of 22 years to a maximum of 109 years, with a median record length of 56 years. More than 80% of the stations contain record lengths between 20 to 59 years, with very few stations with record lengths greater than or equal to 100 years (2.5%).

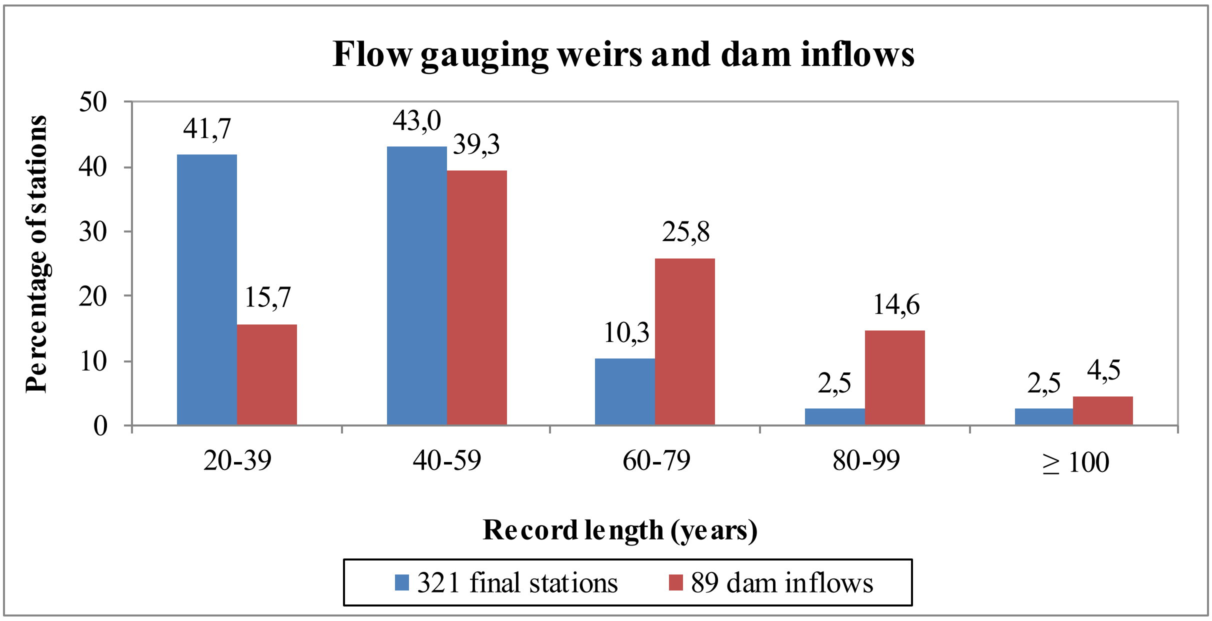

Figure 4.4 Record lengths for the final stations and dam inflow records analysed in this study

Figure 4.5 illustrates the rating table exceedence for the final 321 stations. It can be seen that 52.3% of the stations contained datasets where the rating table was not exceeded.

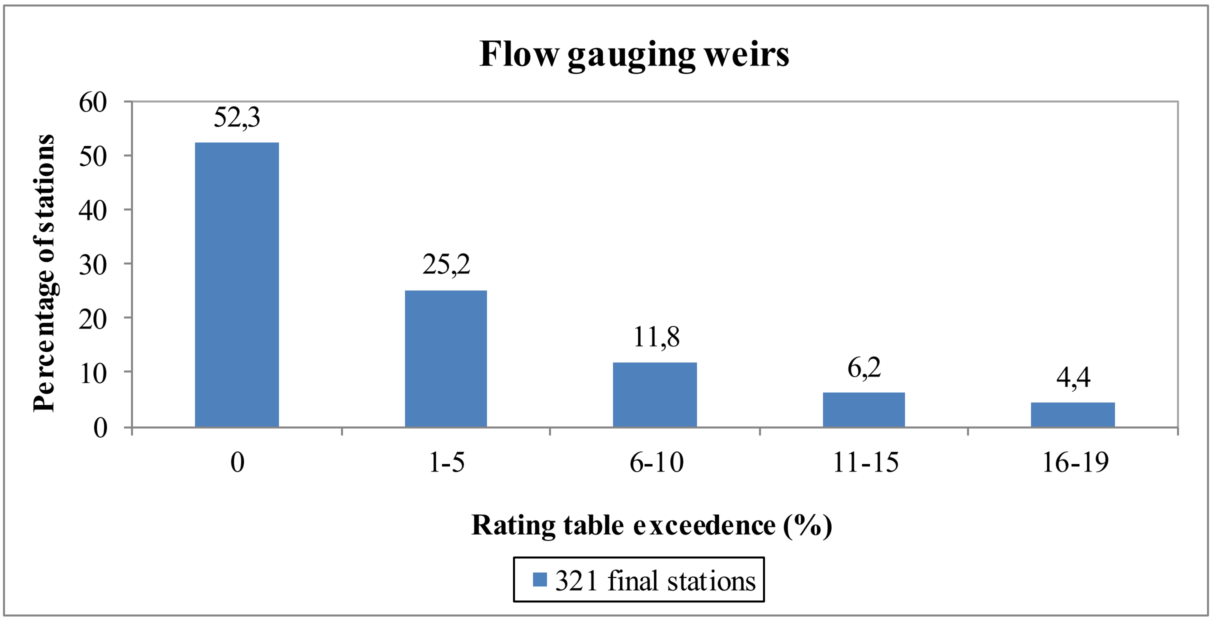

Figure 4.5 Rating table exceedence for the final 321 stations analysed in this study

Figure 4.6 illustrates the distribution across South Africa of the total original 1458 stations received from DWS, as well as a map of the final stations and dam inflow records selected for analysis in this study. The distribution of the final stations is reasonable, covering all regions of the country.

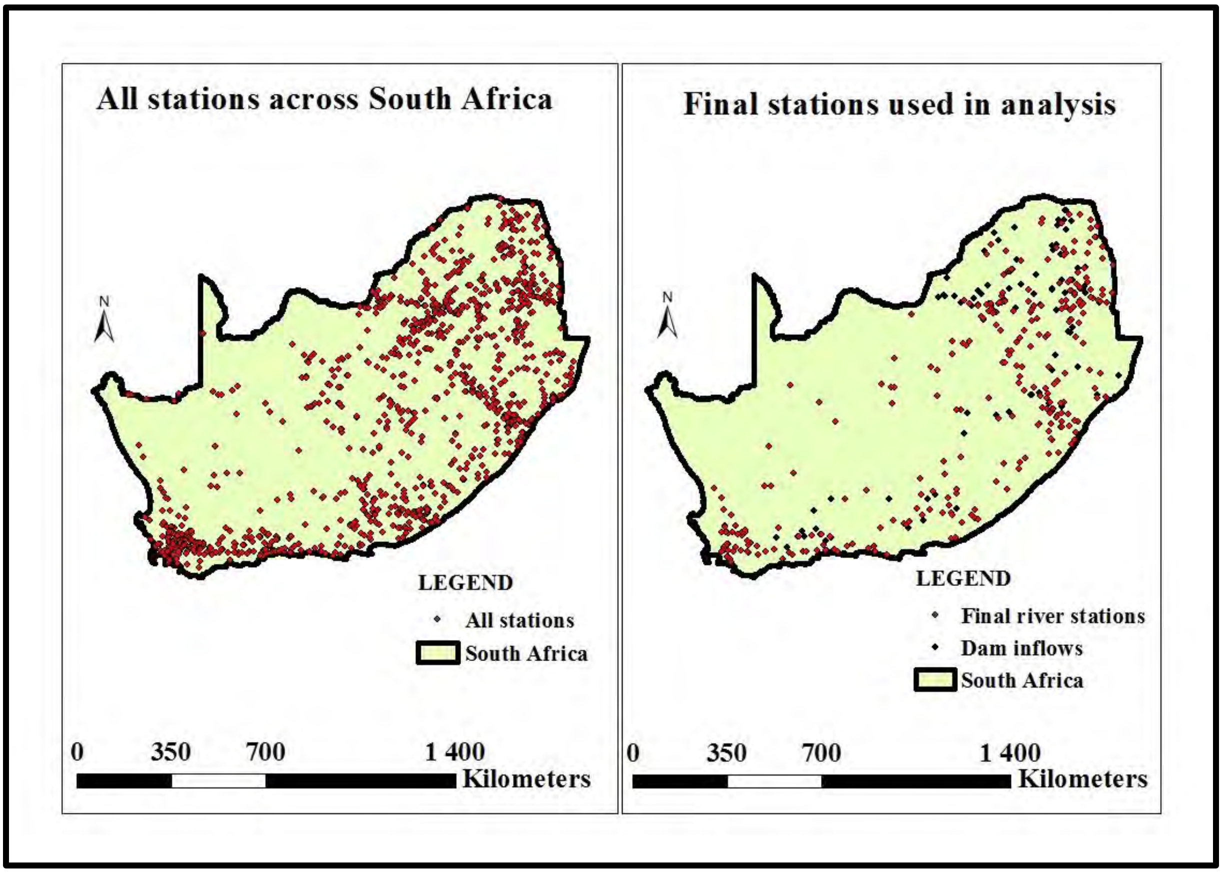

Figure 4.6 Distribution of total flow records across South Africa, compared to a map of final stations and dam inflows records used in the analysis

4.2 Catchment Parameters

Having selected the stations for analysis, the catchment parameters for each station were obtained, which are required to apply each regional method. The Mkhandi, Haile and Meigh Methods only require catchment area (km^2), while the JPV Method requires catchment area (km^2), MAR90 (mm), Kovacs K-Regions, HRU Veld zones and the average main channel slope. These parameters were obtained, using the ArcGIS 10.2.1 software suite (ESRI, 2014), as detailed below.

4.2.1 Catchment area

Catchment areas for each gauging weir were provided by the DWS; however, the shapefiles for these areas do not exist. For this reason, the existing quaternary catchment coverage (after Midgley *et al.*, 1994) for South Africa was utilised to produce catchment shapefiles for each station and to determine the catchment area.

The DWS provided an inventory, which contained the spatial location of each station, given as latitude and longitude coordinates in decimal degrees. In addition, the quaternary catchment, in which each station was located, was also given. Using this information, a Python script coded by Clark (2014) was utilised to generate new catchment shapefiles. This script allows a user to input all of the upstream quaternary catchments contributing to a particular station, which allows the Python program to output a copy of the quaternary catchment shapefile containing only the user-entered quaternaries.

The accuracy of the quaternary catchment data in which a station is located, determines the accuracy of the estimated catchment area. In a number of cases, the location of a particular station was correct spatially; however, it did not accurately reflect the quaternaries contributing to a station's recorded streamflow. For example, according to the information provided by DWS, Station X2H024 is geographically located in Quaternary X23D as shown in Figure 4.6. For the given latitude and longitude for X2H024, it is not located at the outlet of X23D, but just after the boundary between X23C and X23D, indicating that either the quaternary boundaries may not have been accurately digitized, or that the latitude and longitude values were incorrect. From Figure 4.6, it is clear that any streamflow recorded at station X2H024 is being generated in quaternary X23C only and not in X23D. The catchment area given by the DWS for station X2H024 is 80 km^2. If both Quaternaries X23D and X23C were used as inputs in the Python script, the estimated catchment area would be the sum of X23D and X23C, which is 266.24 km^2. The area of Quaternary X23C is 82.22 km^2, which is closer to the DWS area, which indicates that either the latitude and longitude values were incorrect or that the quaternary boundaries have not been accurately digitized.

After a visual inspection of the DWS stations across South Africa, it was found that 55 of the stations used in this study had the same problem as the one described above for Station X2H024. Therefore, the correct quaternaries were noted for these stations and were used to estimate their catchment areas. A list of these stations is given in Appendix B.

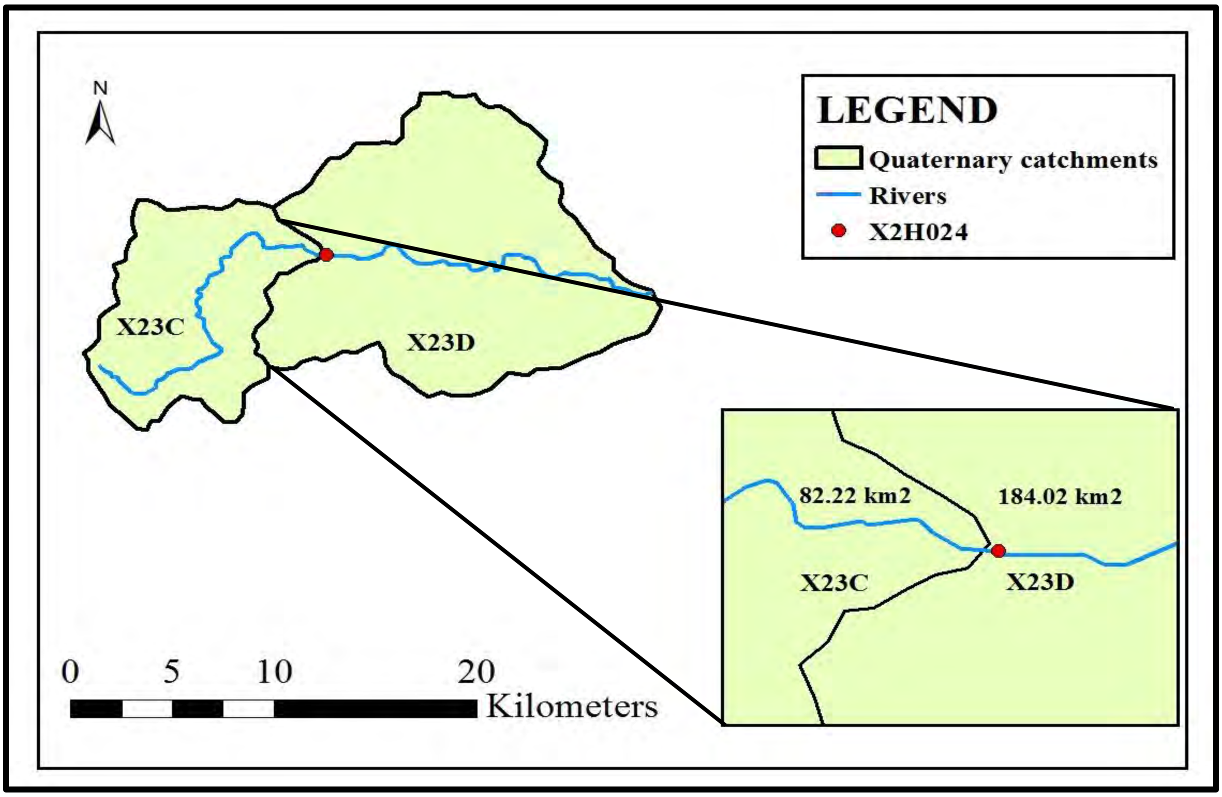

Figure 4.7 Example of problem with spatial station location

A number of stations in this study were located within a quaternary and not at the quaternary outlet. For these stations, sub-catchment boundaries needed to be delineated and catchment areas recalculated. Figure 4.8 is an example of such a case, where Station A6H011 is located in the middle of Quaternary A61A. The area published by the DWS for station A6H011 is 73 km^2 and the area of quaternary A61A is 383.22 km^2. In order to delineate the sub-catchment boundary, the cut polygon tool in the ArcGIS software was utilised with 20 m contour lines. The new area of the delineated sub-catchment was calculated using the Xtools Pro 11.1 software, which is an extension of the ArcGIS software that provides tools for shape conversion, attribute table management and vectospatial anaylsis. In order to calculate the new area in km^2, the shapefile was projected from decimal degrees into meters. Thereafter, the calculate geometry tool of the Xtools pro software was utilised to recalculate the area in m^2 and to convert the area in m^2 to km^2. The area of the new sub-catchment was 73.66 km^2, which corresponds to the DWS area of 73 km^2.

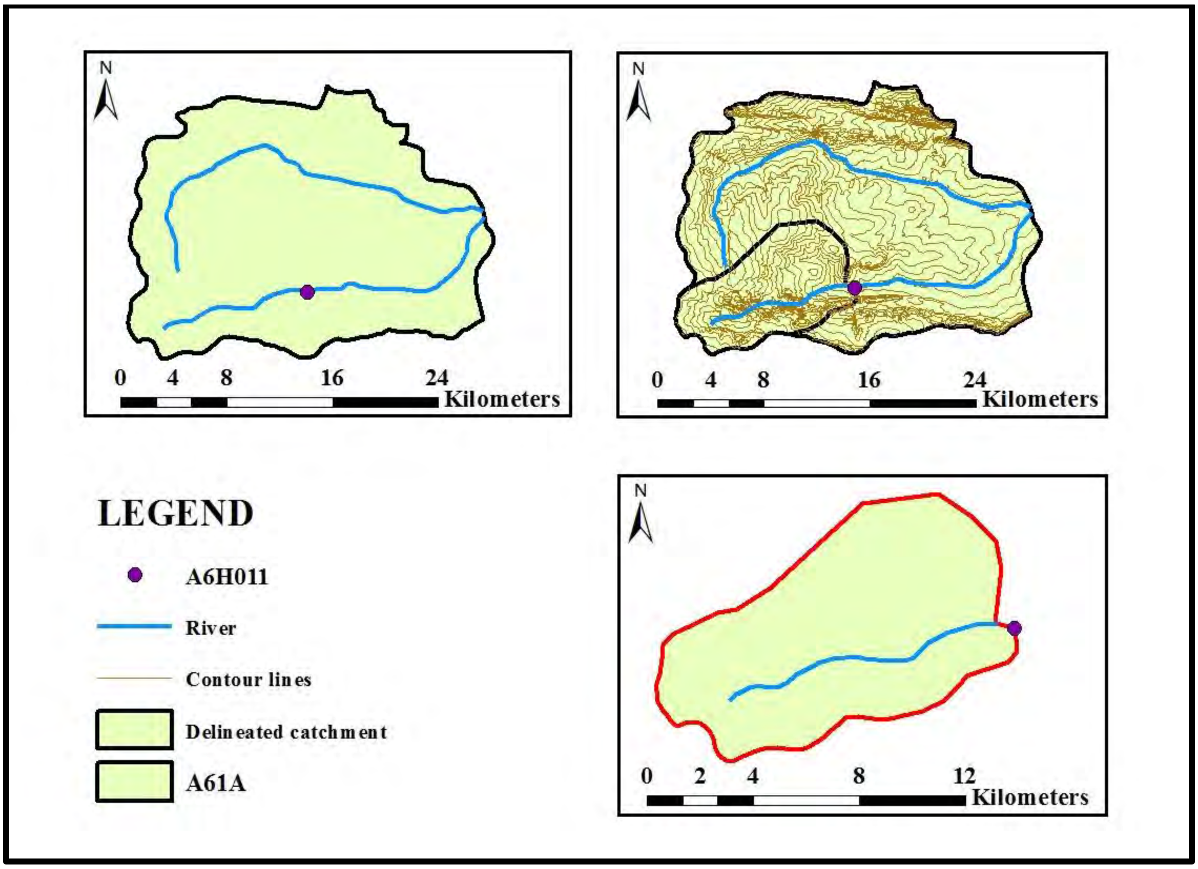

Figure 4.8 Map of the delineation process for sub-catchment area estimation

From the 410 river stations and dam inflow datasets analysed, 332 stations were located at the quaternary outlet and the remaining 78 stations were not found at quaternary outlets, as shown in the example in Figure 4.8. Figure 4.9 displays a map of the stations located at the quaternary outlets and those not located at quaternary outlets.

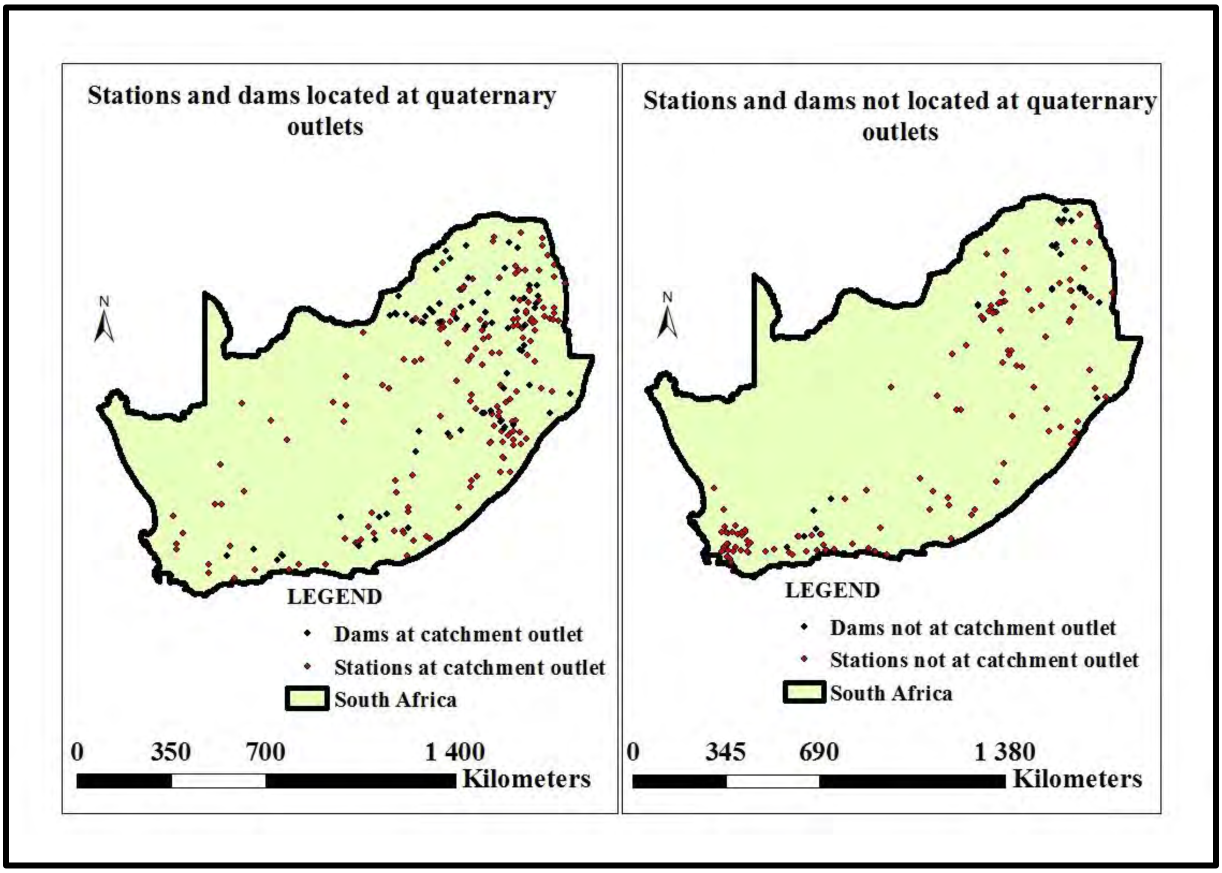

Figure 4.9 Maps of stations located at quaternary outlets and those not located at quaternary outlets

Having produced the catchment shapefiles for all of the stations, using the ArcGIS 10.2.1 software, the GIS estimated catchment areas were compared to the areas published by the DWS. This is illustrated by the scatter plot in Figure 4.10. It can be seen that there is a good correlation between the DWS areas and the GIS estimates by the R^2 value of 0.9998. There is also no scatter above or below the 1:1 line, indicating that there is very little over- or under-estimation of the catchment area using ArcGIS. The catchment areas range from as small as 0.31 km^2 to as large as 370 061.7 km^2.

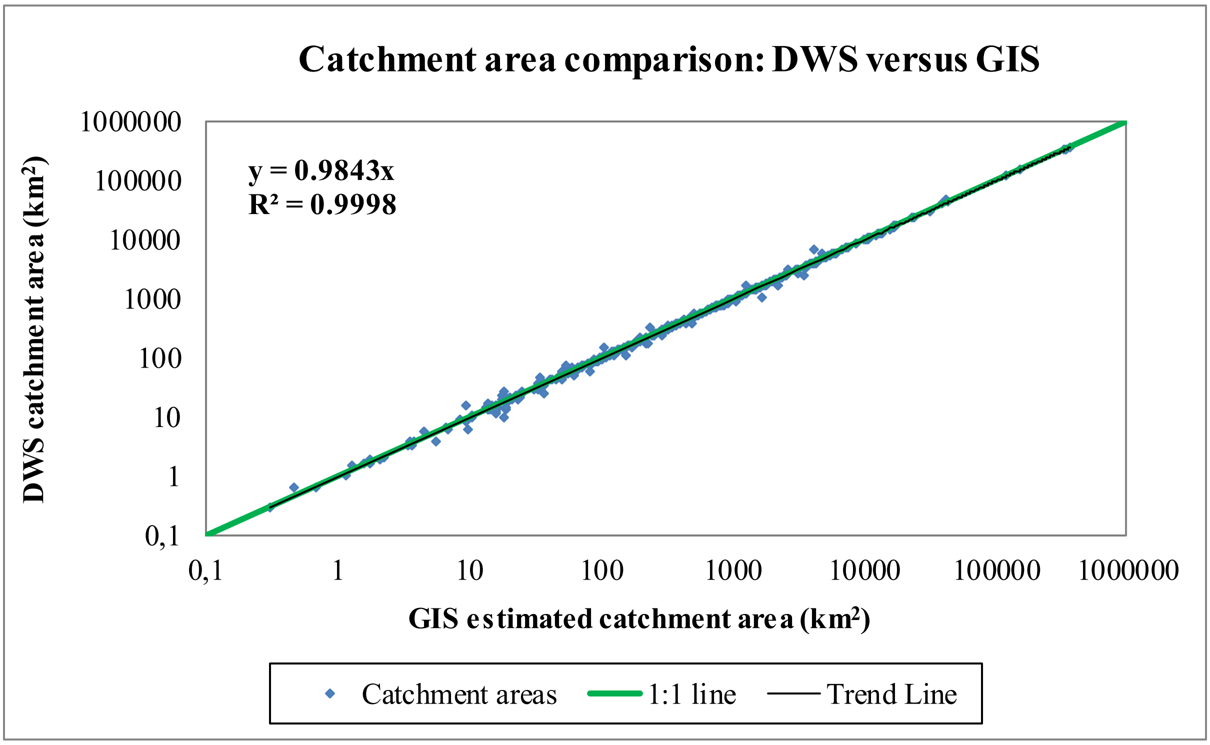

Figure 4.10 Comparison of areas published by the DWS and areas estimated, using ArcGIS

Figure 4.11 shows the distribution of stations that fall into different catchment area ranges and it can be seen that a wide range of catchment areas are being accounted for by the 410 stations in this study. More than 50% of the stations have areas that lie between 101 and 10 000 km^2, while 27.8% of the stations have catchment areas that are 100 km^2 or smaller. Only 7.4% of the stations have catchment areas that are larger than 10 000 km^2.

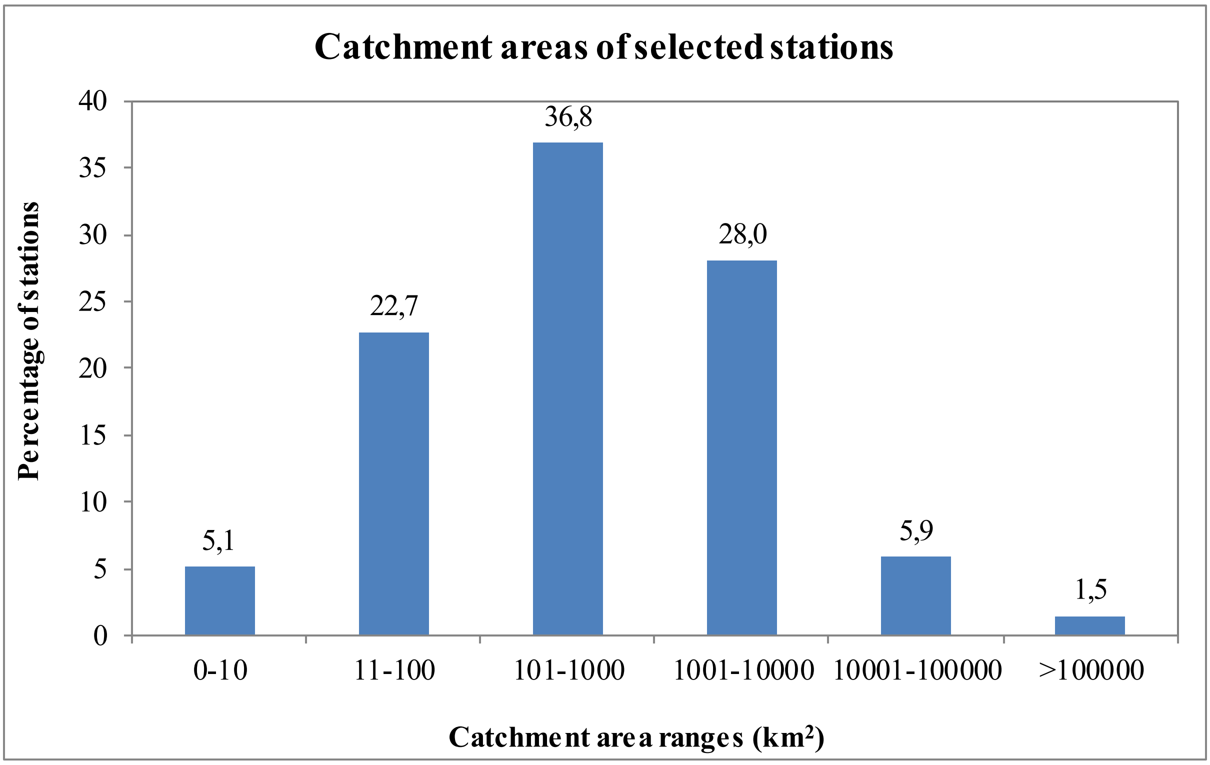

Figure 4.11 Distribution of catchment areas

4.2.2 Mean Annual Runoff (MAR 90)

In 1994, the Water Research Commission published a report entitled: "Surface water resources of South Africa" by Midgley *et al.* (1994). This report aimed at updating and revising the previous Surface Water Resources of South Africa Report by Midgley *et al.* (1981) which involved a national survey of South African water resources where information such as MAP and MAR, amongst others were determined at a quaternary catchment scale. The MAR90 was determined by using the Pitman hydrological model. The simulated monthly flows were reduced during irrigation periods, by multiplying the difference between the catchment rainfall and potential evapotranspiration by the area under irrigation (Midgley *et al.*, 1981). To determine the final MAR90 for each quaternary catchment, the MAR for the tertiary catchments were subdivided in the same proportion as the corresponding quaternary catchments.

Since the Midgley *et al.* (1994) study, there has been more updated documentation; however, due to the JPV Method utilising the MAR90 from the Midgley *et al.* (1994) study, the results from Midgley *et al.* (1994) are used in this study. For each station, the MAR90 had to be determined as an Area Weighted Depth (AWD), expressed mathematically in Equation 4.1.

The MAR90 values provided by the Midgley *et al.* (1994) study were given as a volume in million cubic meters ($m^3 \times 10^6$), while the JPV method requires the MAR90 to be expressed as a depth in mm.

$$AWD = \frac{\sum_{i=1}^{n}(MAR90)}{\sum_{j=1}^{n}(AREA)} \tag{4.1}$$

where

AWD	=	Area Weighted Depth (mm),
MAR90	=	Mean Annual Runoff ($m^3 \times 10^6$), and
AREA	=	Catchment area of quaternary (km^2).

When a gauging station was not located at a quaternary outlet, the MAR90 was calculated by calculating the depth of MAR90 in the entire quaternary catchment and multiplying this depth by the area of the quaternary catchment which contributes to flow at the gauging station. As an example, Table 4.2 provides the original area and corresponding MAR90 values for the quaternaries that contribute to the streamflow recorded at Station T1H001, which is located within Quaternary T11C. The correct MAR90 must be calculated, considering the edited area of Quaternary T11C.

Table 4.2 Catchment area and MAR90 data contributing to station T1H001

Quaternary	Area (km^2)	MAR90 ($m^3 \times 10^6$)
T11A	332.90	34.4
T11B	418.87	46.2
T11C	386.91	66.6

The catchment area of Quaternary T11C which contributes to flow at Station T1H001 is 248.09 km^2. Therefore, the MAR90 is calculated as follows:

a) Calculate the ratio of the area of Quaternary T11C to the area of T11C after being edited.
   - 248.09/386.91 = 0.64

b) Calculate a new MAR90 value for Station T1H011 in Quaternary T11C by multiplying the original MAR90 value of 66.6 $m^3 \times 10^6$ by the quaternary area ratio.

- 66.6 $m^3 \times 10^6 \times 0.64 = 42.7\ m^3 \times 10^6$

c) Finally, calculate the *AWD*, using the new MAR90 of 42.7 $m^3 \times 10^6$ for Quaternary T11C expressing the MAR90 in m^3 and the catchment area in m^2

AWD = [(34.4+46.2+42.7)/ (332.90 +418.87 +248.09)]
AWD = 123300000 m^3/ 999860000 m^2
AWD= 0.123 m

d) Convert from m to mm by multiplying by 1000 to give a final *AWD* of 123.31 mm.

In order to gain confidence in the MAR90 values calculated within this study, a comparison was made with the MAR90 values reported by Görgens (2007). Figure 4.12 illustrates that the majority of stations plot close to the 1:1 line and there is a good correlation (R^2 = 0.9086) between the Görgens (2007) MAR90 values and the MAR90 values calculated in this study. However, there are several stations where the MAR90 values are significantly different from each other. This could be due to the contributing areas in this study being different from those in the JPV study. It is also important to note that other studies, such as Smithers *et al.* (2015), have also identified differences in the values published in the JPV study.

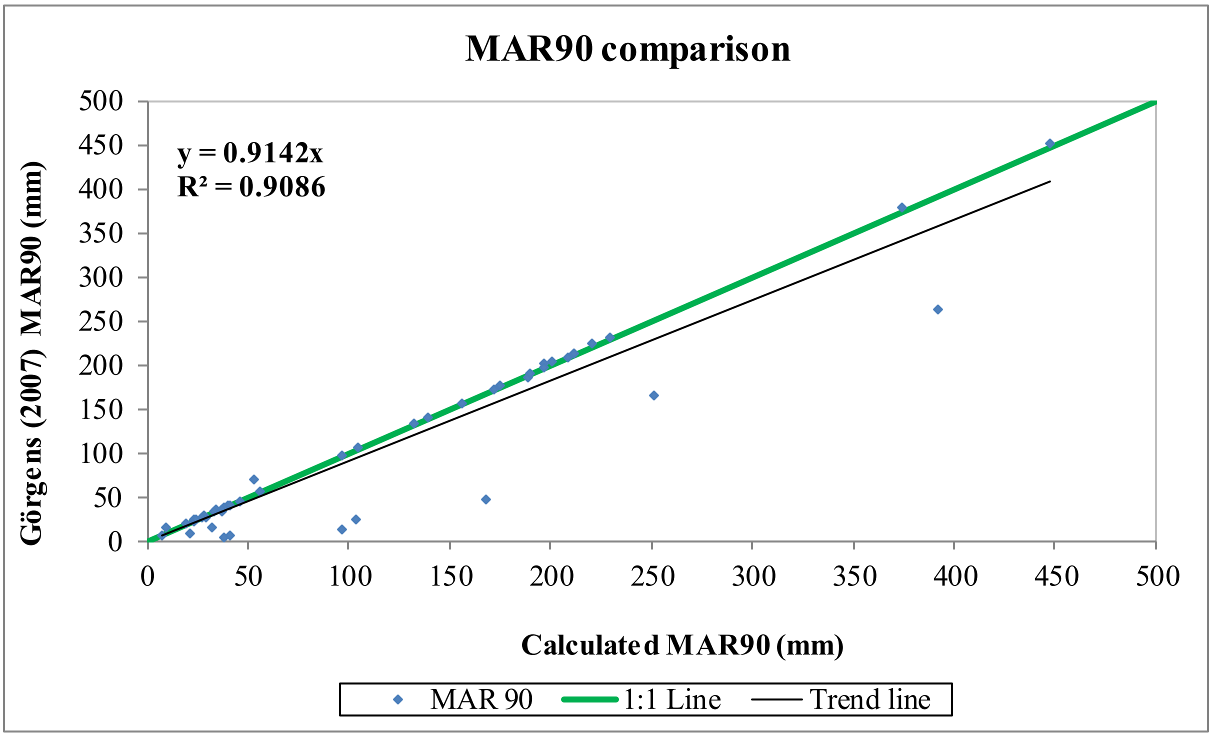

Figure 4.12 Comparison of MAR90

Table 4.3 provides a list of the stations where there are discrepancies in the MAR90 values. A sample calculation is provided for Station V2H002, which illustrates that the MAR90 value calculated in this study is correct, based on the quaternary data used. For this reason, the MAR90 values calculated in this study for the stations in Table 4.3 were used, despite the fact that they were different to the MAR90 values reported by the Görgens (2007) study.

Table 4.2 Stations where the MAR90 values differed between this study and the JPV study

Gauge	Calculated MAR90 (mm)	JPV MAR90 (mm)
B3R002	38	4
B7R003	97	14
A2H012	41	7
X1H001	103	24
B6R003	168	47
E2H003	21	8
A2H006	33	15
A3R003	10	15
V2H002	251	166
X3H006	392	263
C8H003	53	70

A sample calculation for Station V2H002 is provided below and Table 4.4 provides the relevant information. The sum of the quaternary catchment areas gives a total area of 951.87 km^2, while the area reported by the DWS is 937 km^2.

Table 4.3 Quaternary data for station V2H002

Quaternary catchment	MAR90 (m^3 x 10^6)	Quaternary catchment area (km^2)
V20C	51	189.31
V20D	51	301.70
V20B	53	191.73
V20A	84	269.13

A sample calculation of the MAR90 for the V2H002 station is provided below, where the MAR90 values have been summed in m^3 x 10^6 and then converted to m^3 and the catchment areas have been summed in km^2 and then converted to m^2.

$$AWD = \frac{\Sigma_{i=1}^{n}(MAR90)}{\Sigma_{j=1}^{n}(AREA)}$$

$$AWD = \frac{51+51+53+84}{189.31+301.70+191.73+269.13}$$

$$AWD = \frac{239\ 000\ 000\ m^3}{951\ 870\ 000\ m^2}$$

$$AWD = 0.251\ m$$

The final *AWD* in mm is 251.20 mm, not 166 mm, as reported by Görgens (2007). If the DWS catchment area is used (937 km^2) then the *AWD* = 255.1 mm.

4.2.3 Average main channel slope

The JPV Method requires the average main channel slope to be calculated, using the equal area method. The equal area slope is calculated by drawing a slope along the longitudinal profile of the main river channel that equally divides the areas above and below the slope, as illustrated in Figure 4.13. The equation used to calculate the equal area slope is given in Equation 4.2 (Gericke and du Plessis, 2012).

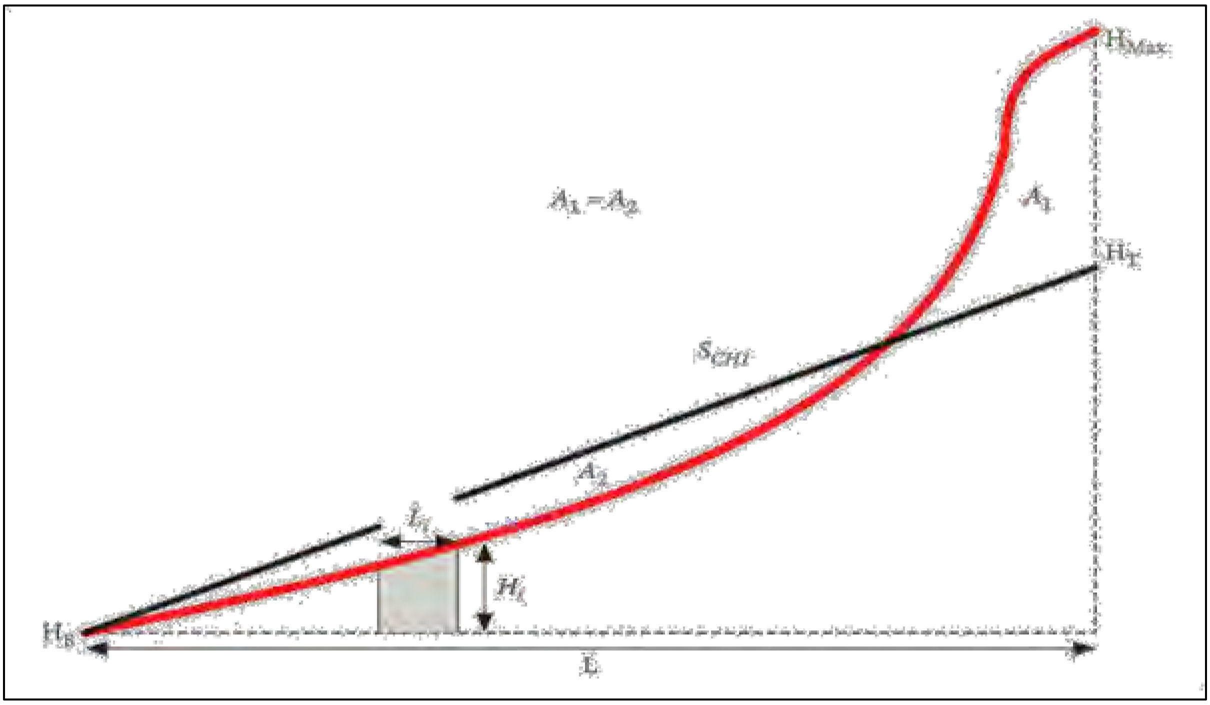

Figure 4.13 Equal area slope method (Gericke and du Plessis, 2012)

$$S_{CH1} = \frac{(H_T - H_B)}{L} \qquad (4.2)$$

where

S_{CH1} = average main channel slope

$$A_i = \left(\frac{(H_i + H_{i+1})}{2} - H_B\right) L_i$$

$$H_T = \frac{(\sum_{i=1}^{n} A_i \times 2)}{L} + H_B$$

where

H_B = height at the station (m),

H_i = height for the specific contour interval (m),

L = length of main channel (m), and

L_i = distance between contours (m).

In order to obtain the longitudinal river profile, the Model Builder application in the ArcGIS 10.2.1 software was utilised. The data necessary to derive the river profile included the station catchment shapefiles created in this study, a shapefile representing the main rivers and tributaries for South Africa, as well as a DEM. In this study, the Shuttle Radar Topography Mission (STRM 90) gap-filled DEM (Weepener *et al.*, 2011) was utilised and projected into meters. The resolution of this DEM is 90 m x 90 m. Figure 4.14 illustrates the procedure that was utilised to attain the river profile.

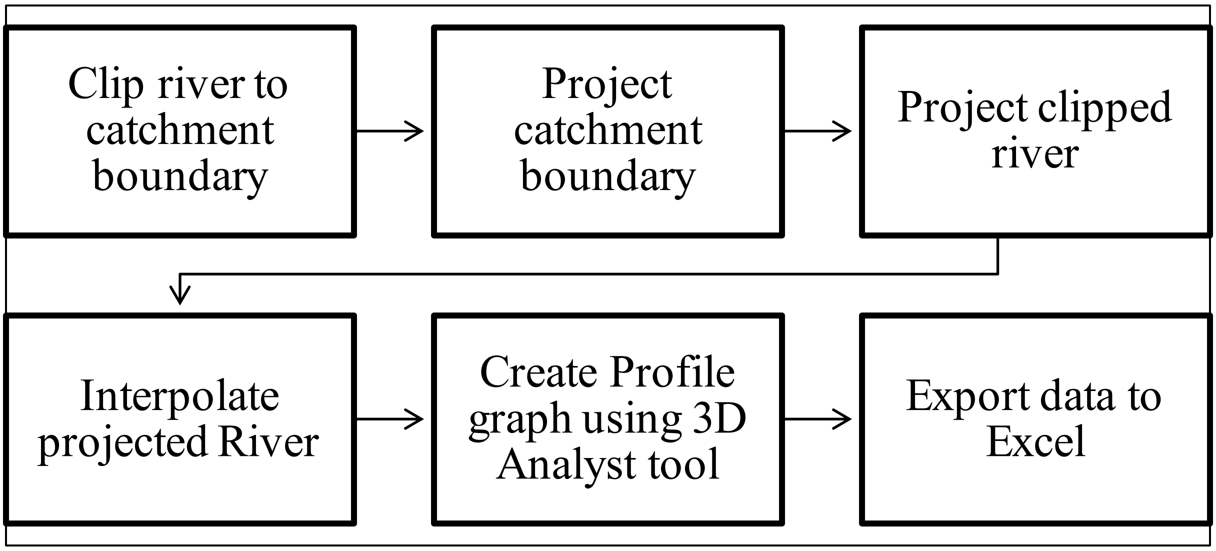

Figure 4.14 ArcGIS procedure to obtain the river profile using Model Builder

The data for the river profile produced by the above procedure was imported into Excel to calculate the equal areas slope using Equation 4.2. Figure 4.15 illustrates the equal areas slope produced for Station V2H002, which allowed the determination of the average main channel slope, namely, 0.34% or 0.0034.

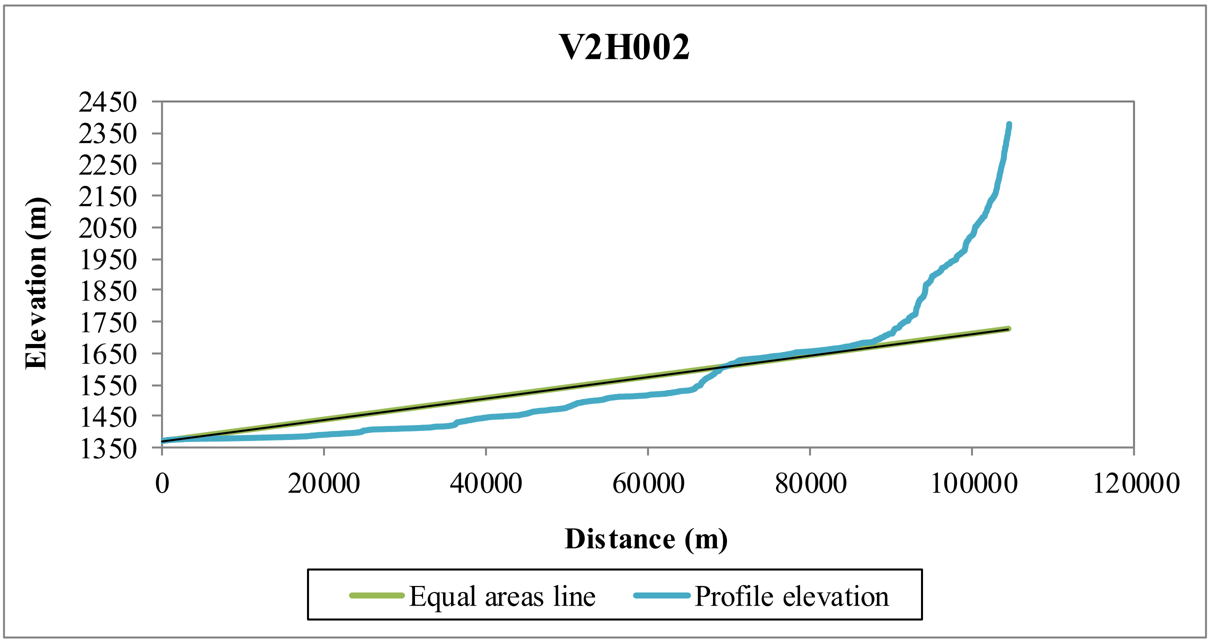

Figure 4.15 Equal area slope for Station V2H002

As in the case of the MAR90 values, the slope values calculated in this study were compared to those reported by Görgens (2007), Smithers *et al.* (2015) and Gericke (2015). It is important to note that the equal area slope method is sensitive to the resolution of the elevation data used, which may account for differences in slope values obtained in the different studies. Figure 4.16 illustrates the distribution of relative errors (%) calculated between the slope values for the 48 stations from the Görgens (2007) study and the slope values determined in this study. The relative errors are calculated using the following equation:

$$RE = \frac{[E\text{-}O]}{O} \times 100 \qquad (4.3)$$

where

RE = Relative Error (%),

E = Estimated slope determined in this study,

O = sloped determined in either the Görgens (2007) study, Gericke (2015) study or the Smithers *et al.* (2015) study.

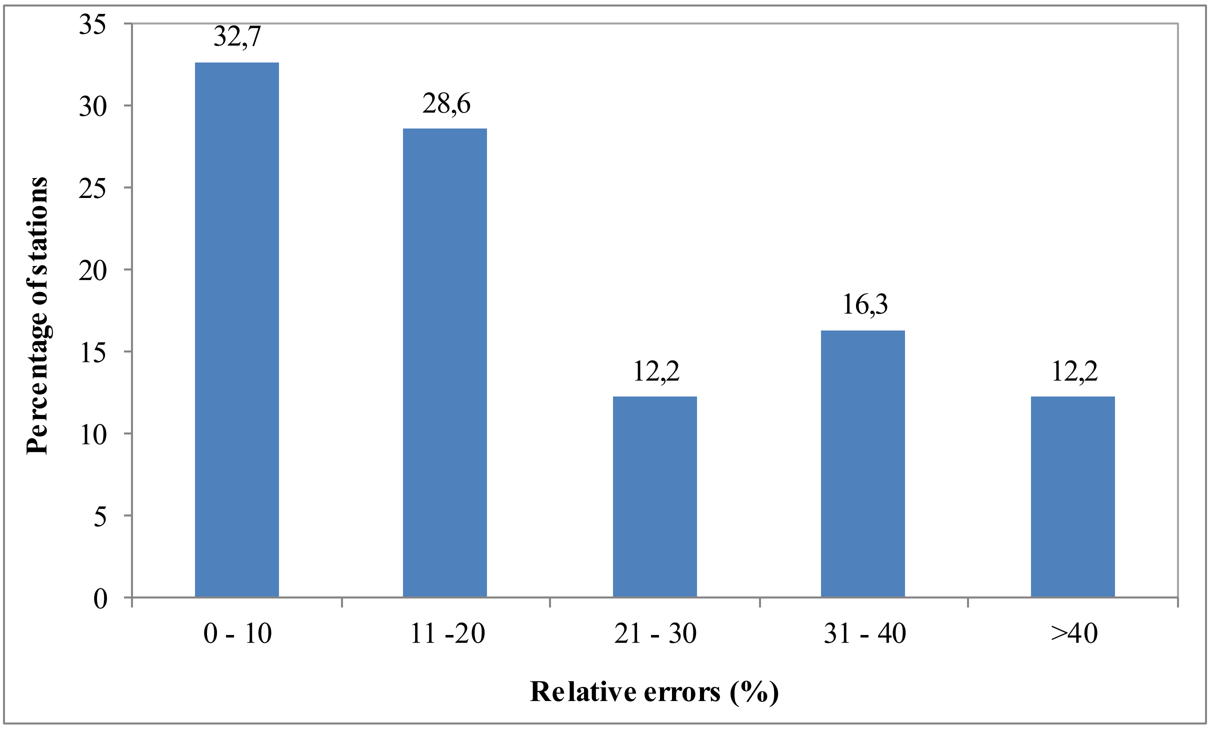

Figure 4.16 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Görgens (2007)

It can be seen that more than 60% of the stations used in this comparison have relative errors less than or equal to 20%. However, 12.2% of the stations contain relative errors that are greater than 40%. Figure 4.17 illustrates a scatter plot of the slope values calculated in this study against the slope values calculated by Görgens (2007). It can be seen that the majority of stations plot close to the 1:1 line, indicating that the slope values for the majority of stations did not differ greatly between this study and the Görgens (2007) study. There are, however, several stations, which do not plot close to the 1:1 line. This could be attributed to the different river coverage that may have been used to calculate each slope, or to a different tributary being selected as the main channel, therefore generating differing river profiles, and subsequently, differing average main channel slopes.

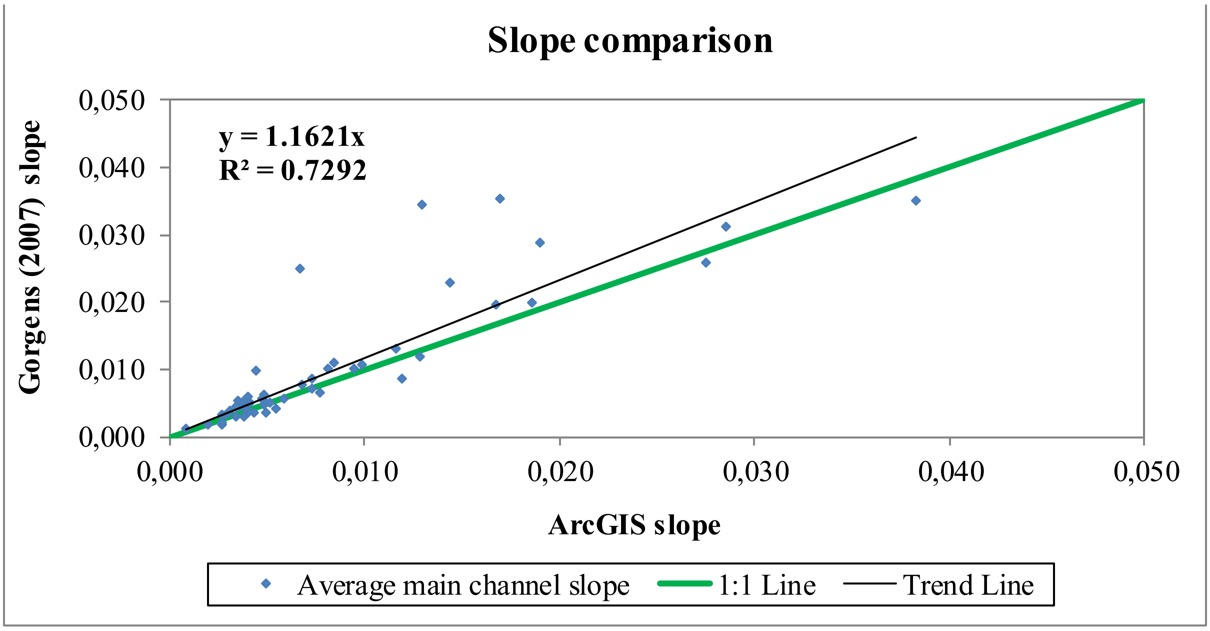

Figure 4.17 Scatter plot of slopes calculated in this study, compared to slopes calculated by Görgens (2007)

The study conducted by Smithers *et al.* (2015) assessed the performance of regional flood frequency analysis methods in KZN. A slope comparison was conducted between 34 stations from the Smithers *et al.* (2015) study and the same stations in this study. Figure 4.18 illustrates the relative errors calculated from the 34 slope values calculated from the Smithers *et al.* (2015) study and the slope values determined in ArcGIS for this study.

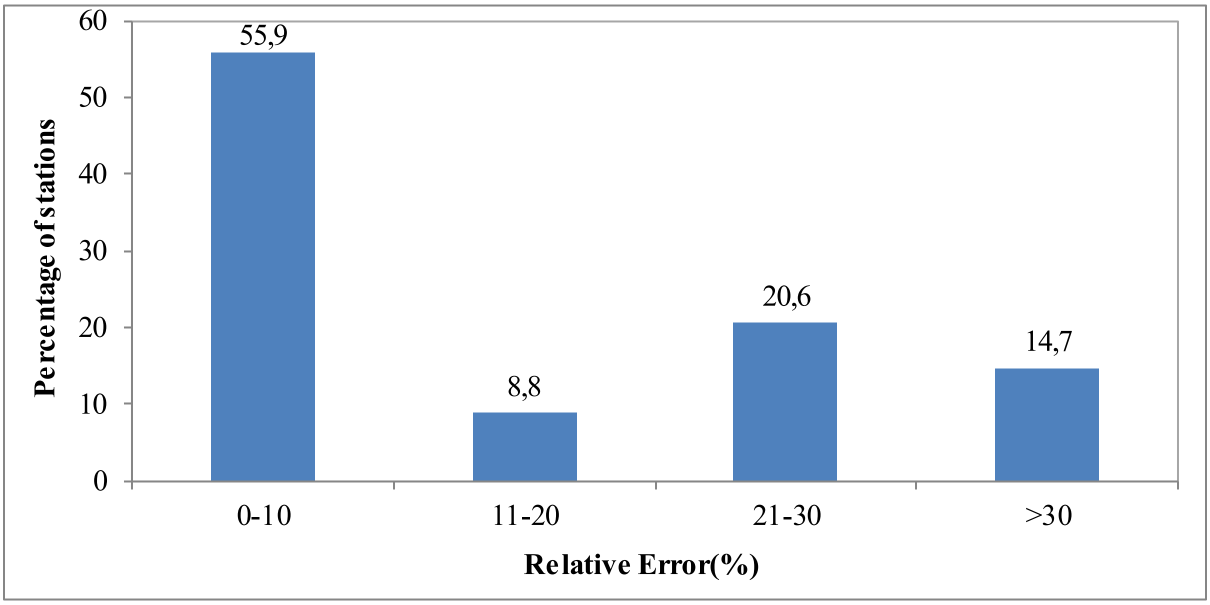

Figure 4.18 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Smithers *et al.* (2015)

It can be seen that 55.9% of the stations used in this comparison contain relative errors that are less than or equal to 10%, while 35.3% of the stations have relative errors greater than 20%. Figure 4.19 illustrates that the majority of stations plot close to the 1:1 line, with several stations that plot away from the 1:1 line.

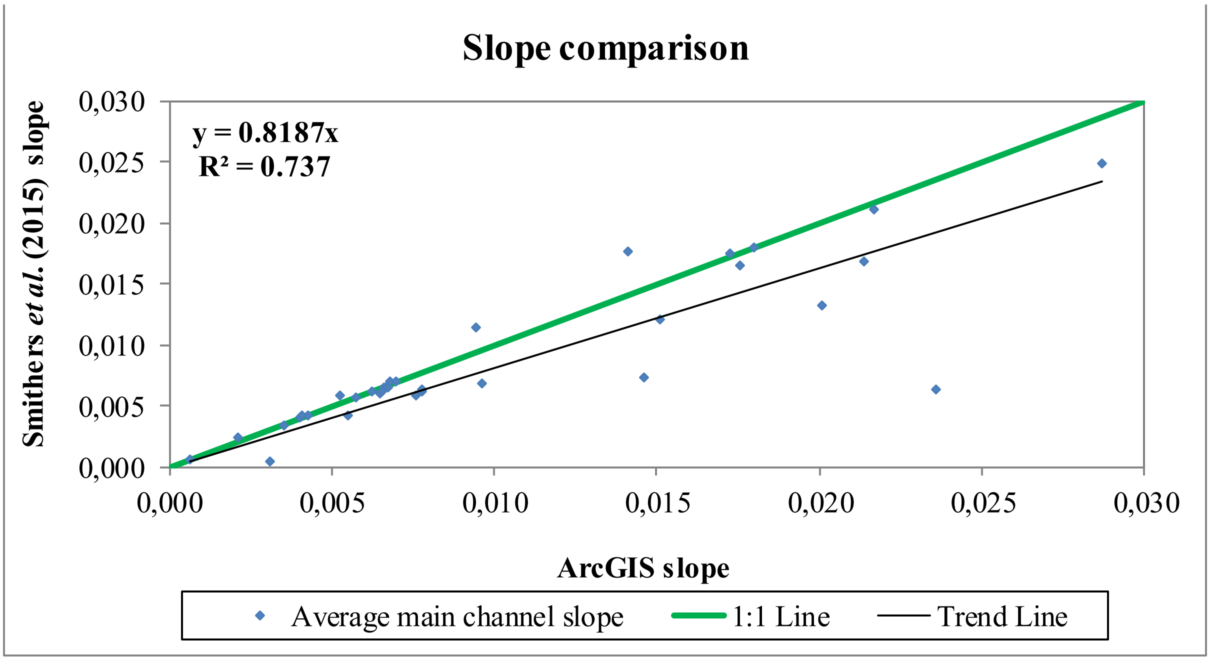

Figure 4.19 Scatter plot of slopes calculated in this study, compared to slopes calculated by Smithers *et al.* (2015)

A slope comparison was also conducted, using 35 stations from a study conducted by Gericke (2015) and the same stations in this study. Figure 4.20 illustrates the relative errors calculated from the 35 slope values determined from the Gericke (2015) study and slope values determined in ArcGIS for this study. It can be seen that more than 48.6% of the stations used in this comparison contain relative errors that are less than or equal to 5%. Only 25.8% of the stations have relative errors greater than 10%. Figure 4.21 shows a scatter plot of slopes calculated in this study, compared to the slopes calculated by Gericke (2015). It can be seen that the majority of stations plot close to the 1:1 line, with only one station plotting away from the 1:1 line. There is a very good correlation between the data indicated by the R^2 value of 0.9961.

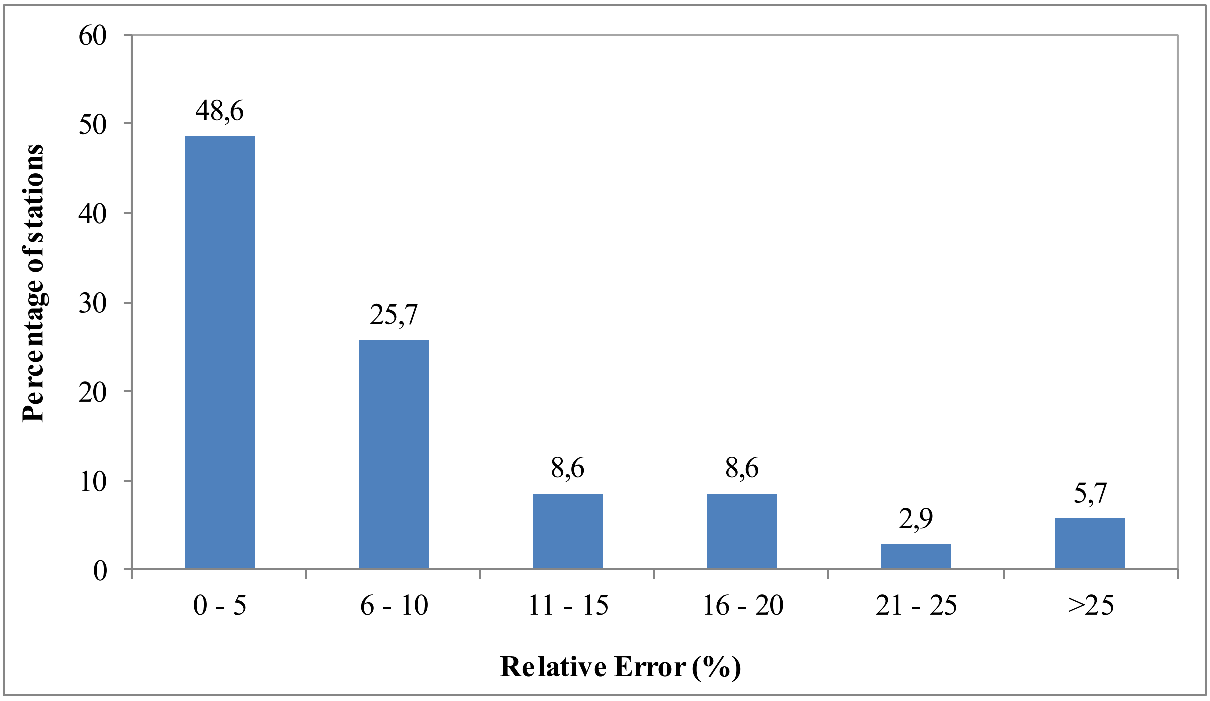

Figure 4.20 Distribution of relative errors calculated from the differences between slopes determined in ArcGIS and slopes determined by Gericke (2015)

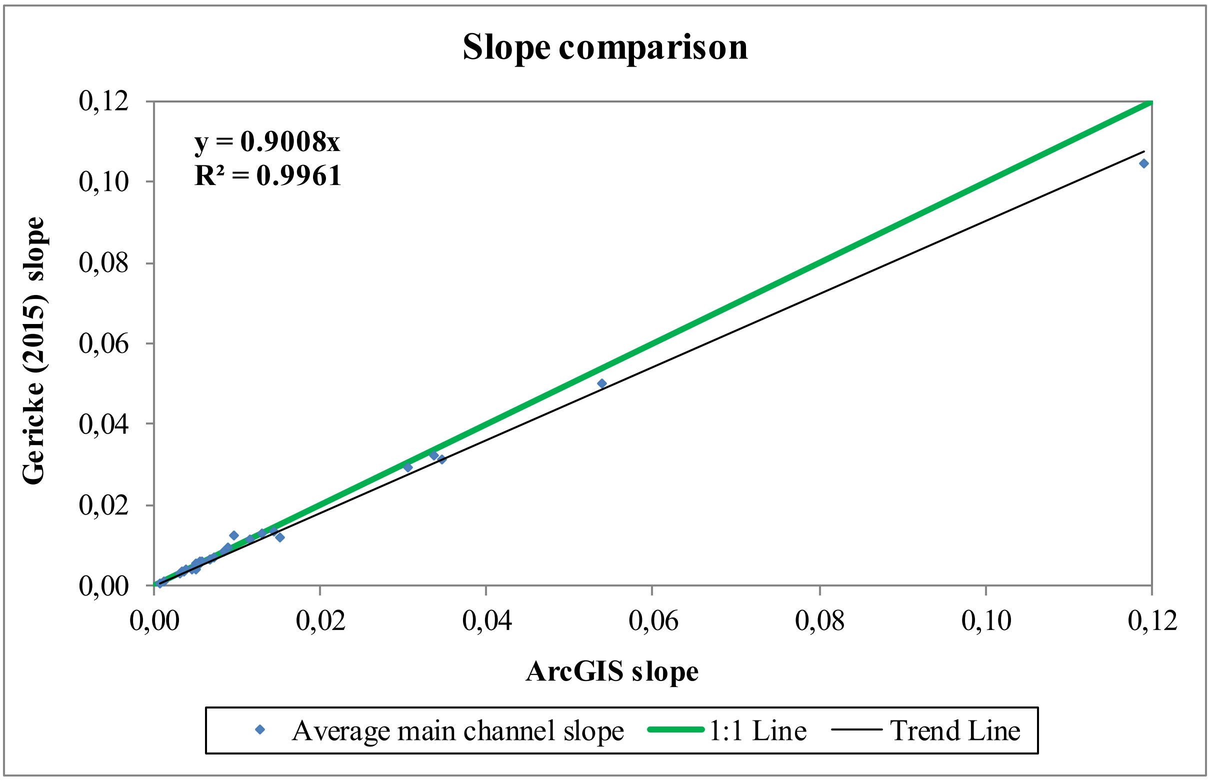

Figure 4.21 Scatter plot of slopes calculated in this study, compared to slopes calculated by Gericke (2015)

Having compared the slopes calculated in this study to those calculated in the Görgens (2007) study, the Smithers *et al.* (2015) and the Gericke (2015) study, it was found that in all three comparisons, the slopes calculated in this study were generally close to those calculated in the aforementioned studies. The frequency plots also indicate that the majority of stations produced relative errors that were less than 10% or 20%. Therefore the slopes calculated in this study were considered to be acceptable and were utilised in the application of the JPV Method.

There are only a few exceptions where the slopes differed significantly. These differences may be attributed to the type of data used in obtaining the river profiles. The source of the DEM, the river shapefile and the catchment shapefiles will all affect the river profile and subsequently the equal area slope.

4.2.4 HRU Veld zones

In order to apply the JPV Method, the pooled HRU Veld zone needs to be determined. The Veld zone that is chosen is the most dominant Veld zone in the catchment upstream of the station. Figure 4.22 illustrates the workflow created in Model Builder, to determine the Veld zones for each station. The data required to determine the Veld zone included a raster shapefile of the Görgens (2007) pooled Veld zone groups (A, B and C), illustrated in Figure 2.6 in Section 2.4.1, as well as the catchment shapefiles created in this study.

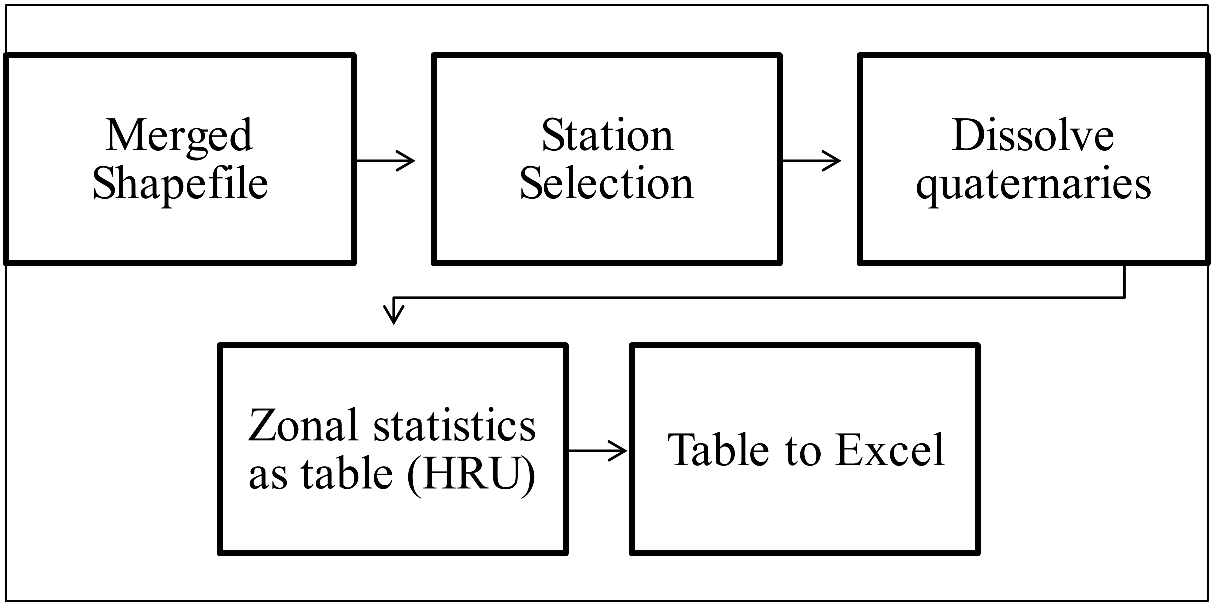

Figure 4.22 ArcGIS procedure to obtain HRU Veld zones

The model begins by selecting the first station of the merged shapefile, containing all stations and selecting all of the quaternaries that contribute to the selected station and dissolves them into one shapefile representing the entire contributing catchment. Thereafter, the zonal statistics tool was utilised, which summarizes raster data within the zones of another feature and produces a table of the results (ESRI, 2014). This allowed the determination of the most dominant Veld zone group in the upstream catchment of the station. The process was repeated automatically until all stations had been analysed.

The Veld zones determined in this study for selected stations were compared to those reported by Görgens (2007). Table 4.5 contains a list of stations where the Veld zones reported by Görgens (2007) differed from those determined in this study. This often occurred when the catchment of a station fell into two different Veld zone groups.

Table 4.4 Veld zones reported by Görgens (2007), compared to Veld zones determined in ArcGIS

Station	Görgens (2007) Veld zones	ArcGIS Veld zones
A2H013	C	B
A2R006	C	B
A2R007	C	B
A3R001	C	B
A3R002	C	B
A3R003	C	B
B1H004	C	B
E2H003	A	B
J1R002	B	A
Q9H002	C	B
V2R001	C	B
X1H001	C	B

The catchment shapefiles of the stations listed in Table 4.5 were edited, so that a line was digitized along the boundary between the two Veld zone groups in which a catchment was found. The areas of the catchment above and below this digitized line were determined, using the Xtools Pro 11.1 software. The final dominant Veld zone group was determined as the Veld zone group within which a greater portion of the catchment was found. An example is given in Figure 4.23, which illustrates a map of Station A2R006, where the catchment covers two Veld zones, with areas of 529.48 km^2 (Veld zone = C) and 547.77 km^2 (Veld zone = B). The portion of the catchment with the larger area (547.77 km^2) was located in Veld zone group B, not Veld zone group C, as reported by Görgens (2007). Similar maps were created for all the stations in Table 4.5 and are provided in Appendix C.

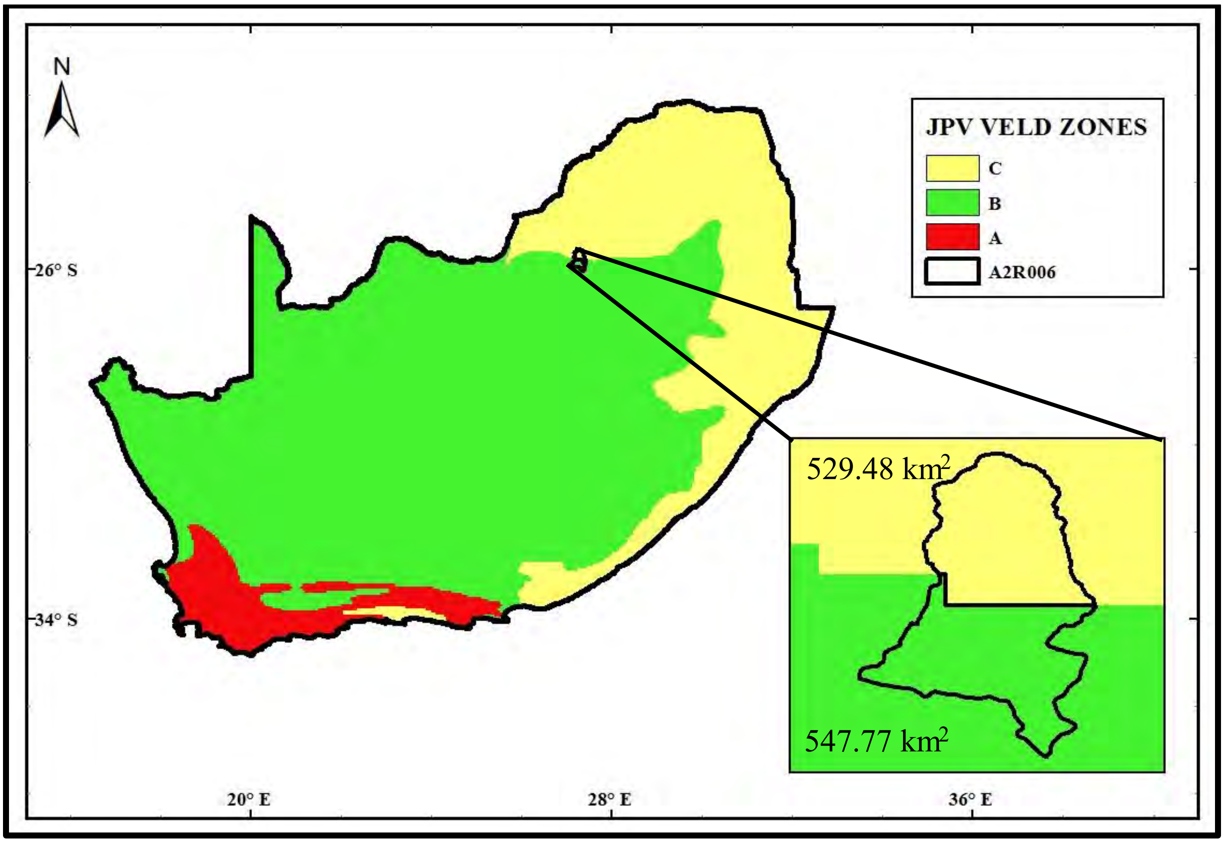

Figure 4.23 Map showing the dominant Veld zone for Station A2R006

4.2.5 Kovacs K-Regions

The pooled Kovacs K-Regions, utilised in the Görgens (2007) study, were chosen as the K-Region at the station location. Table 4.9 below provides a list of stations where the K-Region determined in this study differed from those reported by Görgens (2007).

Table 4.5 K-Regions reported by Görgens (2007), compared to K-Regions determined in ArcGIS

Station	Görgens (2007) K-Region	ArcGIS K-Region
A2R005	K-LOW	K-MID
B6R001	K-MID	K-LOW
U2H006	K-MID	K-HIGH
W5R003	K-LOW	K-MID
X1H001	K-HIGH	K-LOW

Figure 4.24 illustrates that the K-Region at Station A2R005, is the K-MID region and not the K-LOW region as, reported by Görgens (2007). Similar maps were created for all the stations in Table 4.6 and are provided in Appendix D.

Generally the K-Regions reported by Görgens (2007) differed from those determined in this study for stations that were located close to the boundary between two K-Regions. For example, in Figure 4.24, Station A2R005 is close to the boundary between the K-LOW and K-MID K-Regions. The Görgens (2007) report does not indicate how the pooled K-Regions were digitized. Therefore it is possible that the differences in the K-Regions reported by Görgens (2007) to those determined in this study are due to different sources of K-Region boundaries.

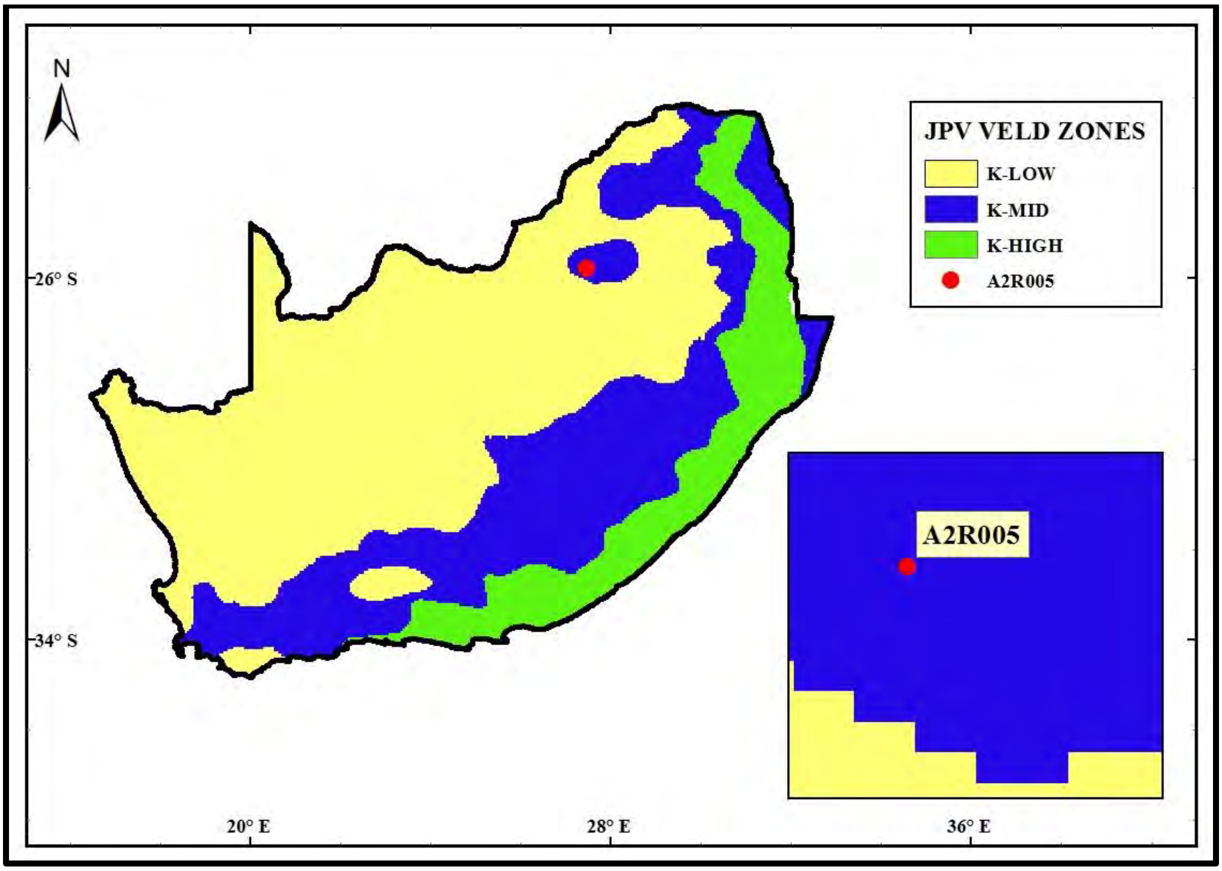

Figure 4.24 Map showing the K-Region at Station A2R005

4.3 New Regression Equations for the Haile Method

This section deals with the regression equations that were derived for the application of the Haile (2011) Method in this study. As summarised in Table 2.4, the original regression equations developed by Haile (2011) for regions ZA_R1, ZA_R2, ZA_R3 and ZA_R5 were derived by fitting a logarithmic trend line to a plot of the catchment area, against the median annual flood for each station within the region. The negative intercept in these regression equations produces a catchment area limit, below which a negative flood index would be calculated. Therefore, new regression equations needed to be determined, for application in this study.

Figure 4.25 illustrates the logarithmic relationship between the *MEF* and catchment area for stations analysed by Haile (2011) in the ZA_R1 region. A relationship has also been illustrated between the *MEF* and catchment area for stations analysed in this study, where the Haile Method cannot be applied. These stations produced a negative index flood when using the Haile logarithmic equation, due to their areas being below 28.3 km^2. A power regression was fitted to these stations as illustrated in Figure 4.25. The *MEF* for these stations were determined using the same period of record that Haile used, which was from 1969 to 2008. The regression was then extrapolated until it intersected the Haile regression at a catchment area of 1145 km^2. Therefore, in this region, the power fitted regression has been applied to all stations with a catchment area that is less than or equal to 1145 km^2 and the Haile logarithmic equation has been applied to all stations with catchment areas greater than 1145 km^2.

It is also important to make mention of the portions of the power and logarithmic regressions that have been highlighted in Figure 4.25 by the black square. The logarithmic regression as developed by Haile (2011) could have been applied to a number of stations that were above the area limit of 28.3 km^2; however, the power regression was used instead, to allow for a smooth transition from the power regression to the logarithmic regression. Out of the 84 stations that were in region ZA-R1, the *MEF* was estimated using the power regression rather than the logarithmic regression for 73 stations (86%). This is a large percentage; therefore if one were to apply the Haile Method in the future in the ZA_R1 region, it may be beneficial to assess the impact of using the power equation rather than the logarithmic regression in these cases.

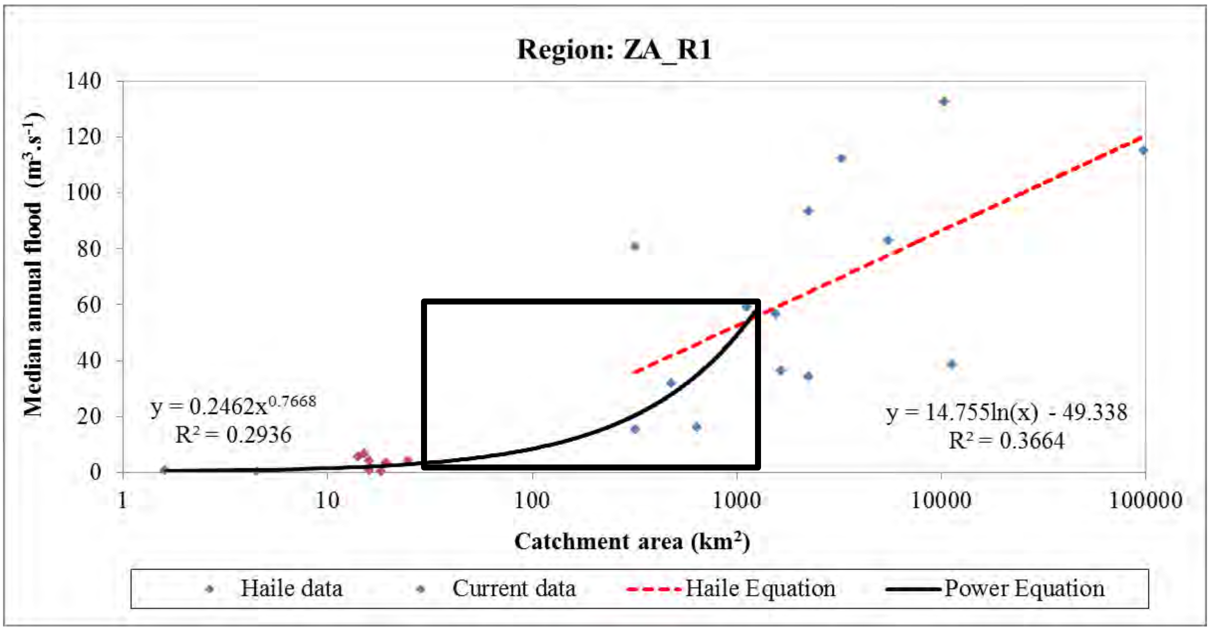

Figure 4.25 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R1 region

It is important to mention the low R^2 value of 0.2936 for the power regression for Region ZA_R1 in Figure 4.25. Considering the detailed data screening process undertaken in this study, the power regression is the best possible regression that can be developed to determine the *MEF*. It is also important to note that the original logarithmic regression developed by Haile (2011) also produced a very low R^2 value of 0.3664. However, this regression has to be applied because the purpose of this study is to assess the performance of the Haile (2011) Method. Therefore the logarithmic regression cannot be changed in any way to improve the low R^2, as this will not be an accurate application of the Haile (2011) Method.

Figure 4.26 illustrates the logarithmic relationship between the *MEF* and catchment area for stations analysed by Haile (2011) in the ZA_R2 region. A relationship has also been illustrated between the *MEF* and catchment area for stations analysed in this study, where the Haile Method cannot be applied. These stations produced a negative index flood when using the Haile logarithmic equation, due to their areas being below 639.9 km^2. A power regression was fitted to these stations as illustrated in Figure 4.26. The regression was then extrapolated until it intersected the Haile regression at a catchment area of 1 076 km^2. Therefore, in this region, the power fitted regression has been applied to all stations with a catchment area that is less than or equal to 1 076 km^2 and the Haile logarithmic equation has been applied to all stations with catchment areas greater than 1 076 km^2.

As in the case of region ZA_R1, portions of the power and logarithmic regressions have been highlighted in Figure 4.26 by the black square, for region ZA_R2. The logarithmic regression as developed by Haile (2011) could have been applied to a number of stations that were above the area limit of 639.9 km^2; however, the power regression was used instead. Out of the 38 stations that were in region ZA_R2, the *MEF* was estimated using the power regression rather than the logarithmic regression for 10 stations (26%). This is a small percentage; however, if one were to apply the Haile Method in the future in the ZA_R2 region, it may be beneficial to assess the impact of using the power regression rather than the logarithmic regression in these cases.

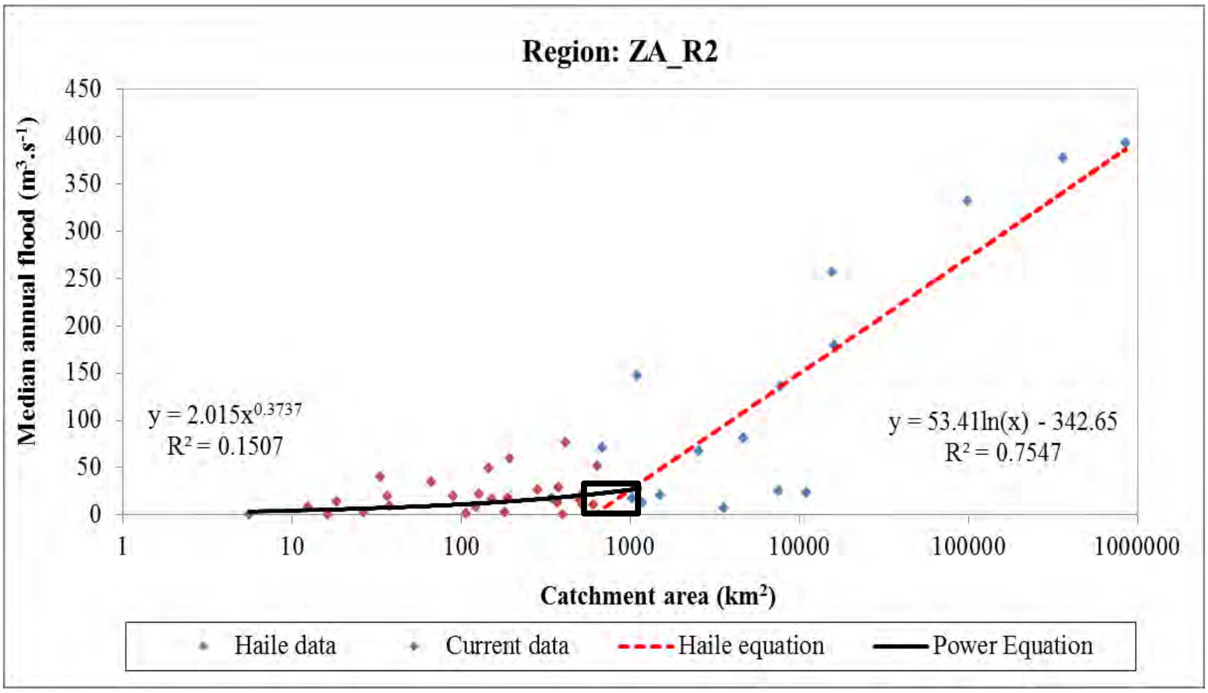

Figure 4.26 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R2 region

As in the case for Region ZA_R1, the power regression in region ZA_R2 also produced a low R^2 value of 0.1507. Based on the available data, this was the best regression that could be produced for the prediction of the *MEF*.

Figure 4.27 illustrates the logarithmic relationship between the *MEF* and catchment area for stations analysed by Haile (2011) in the ZA_R3 region. A relationship has also been illustrated between the *MEF* and catchment area for stations analysed in this study, where the Haile Method cannot be applied. These stations produced a negative index flood when using the Haile logarithmic equation, due to their areas being below 386.5 km^2. A power regression was fitted to these stations as illustrated in Figure 4.27. It can be seen that the power fitted regression and the logarithmic regression do not intersect at any point. Therefore, neither equation was used in the estimation of the index flood in the ZA_R3 region, but rather all of the *MEF* values calculated in this study were used to fit a new power regression, as seen in Figure 4.28. It can be seen from Figure 4.28 that there is a reasonable agreement between the index values calculated using the power equation to the index values calculated using Haile's logarithmic equation for catchment areas greater than 100 km^2 but poor agreement for catchment areas less than 100 km^2.

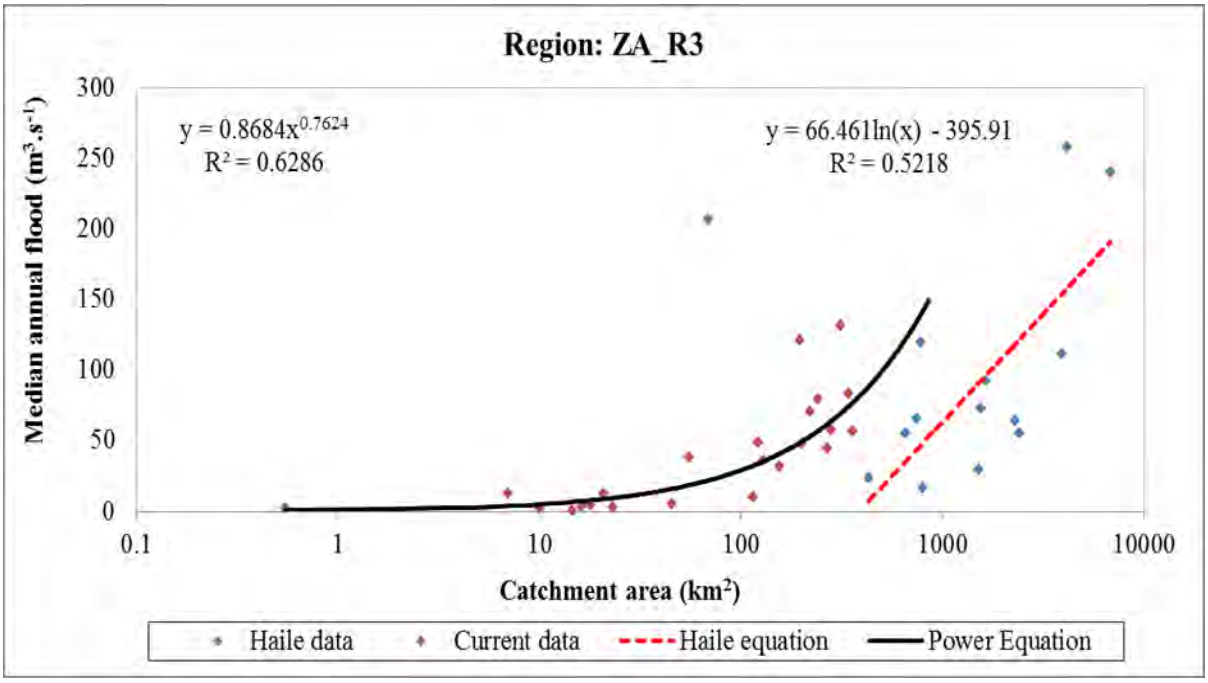

Figure 4.27 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R3 region

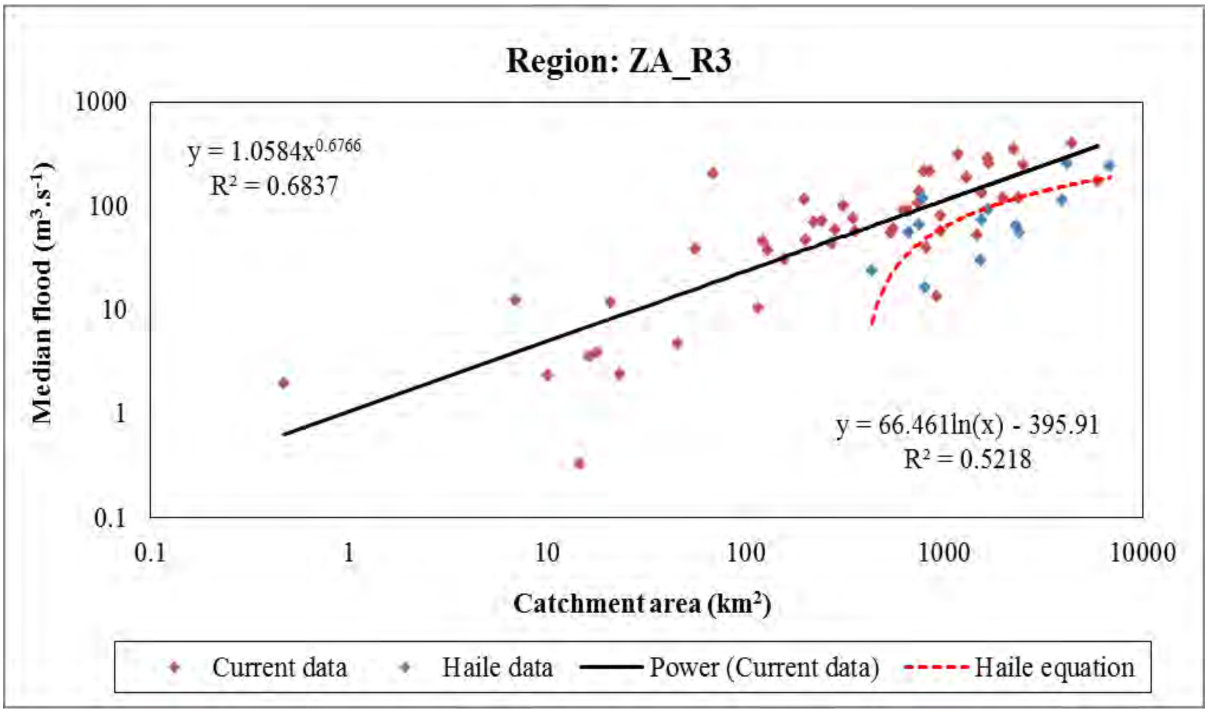

Figure 4.28 Derived equation to estimate an index flood in the ZA_R3 region

Figure 4.29 illustrates the logarithmic relationship between the *MEF* and catchment area for stations analysed by Haile (2011) in the ZA_R5 region. A relationship has also been illustrated between the *MEF* and catchment area for stations analysed in this study, where the Haile Method cannot be applied. These stations produced a negative index flood when using the Haile logarithmic equation, due to their areas being below 83.5 km^2. A power regression was fitted to these stations as illustrated in Figure 4.29. It can be seen that the power fitted regression and the logarithmic regression do not intersect at any point. It is important to note that the data points circled in red in Figures 4.29 were not provided in Appendix A of the Haile (2011) study, however they were estimated from the regional regression graphs provided in Appendix D of the Haile (2011) study. Stations G1H004, H1H017 and H1H033 have also been circled in black in Figure 4.29, as these stations do not fit the general trend of stations with similar areas within the region and have been considered to be outliers, and were subsequently removed from the analysis.

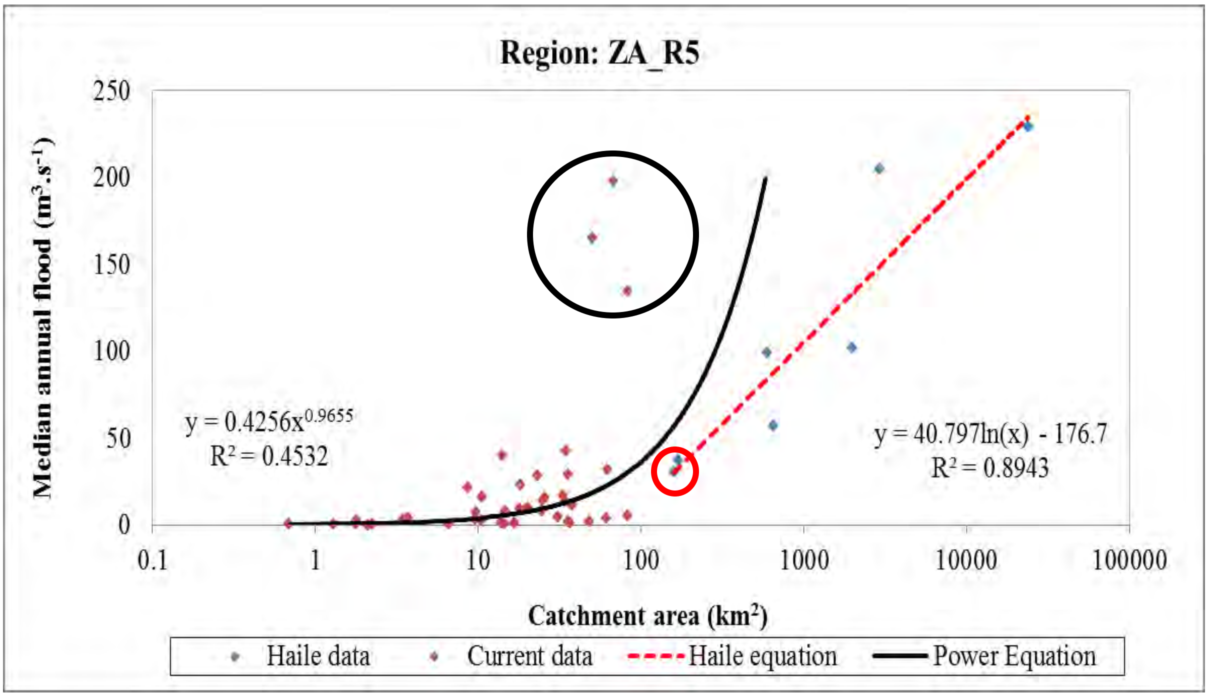

Figure 4.29 Power and logarithmic fitted regressions for the estimation of the index flood in the ZA_R5 region

Figure 4.30 illustrates the relationship between catchment area and the *MEF*, without the outliers identified in Figure 4.29. A power regression was fitted to the stations below the area limit of 83.5 km^2, as illustrated in Figure 4.30. The regression was then extrapolated until it intersected the Haile regression at a catchment area of 234 km^2. Therefore, in this region, the power fitted regression has been applied to all stations with a catchment area that is less than or equal to 234 km^2 and the Haile logarithmic equation has been applied to all stations with catchment areas greater than 234 km^2.

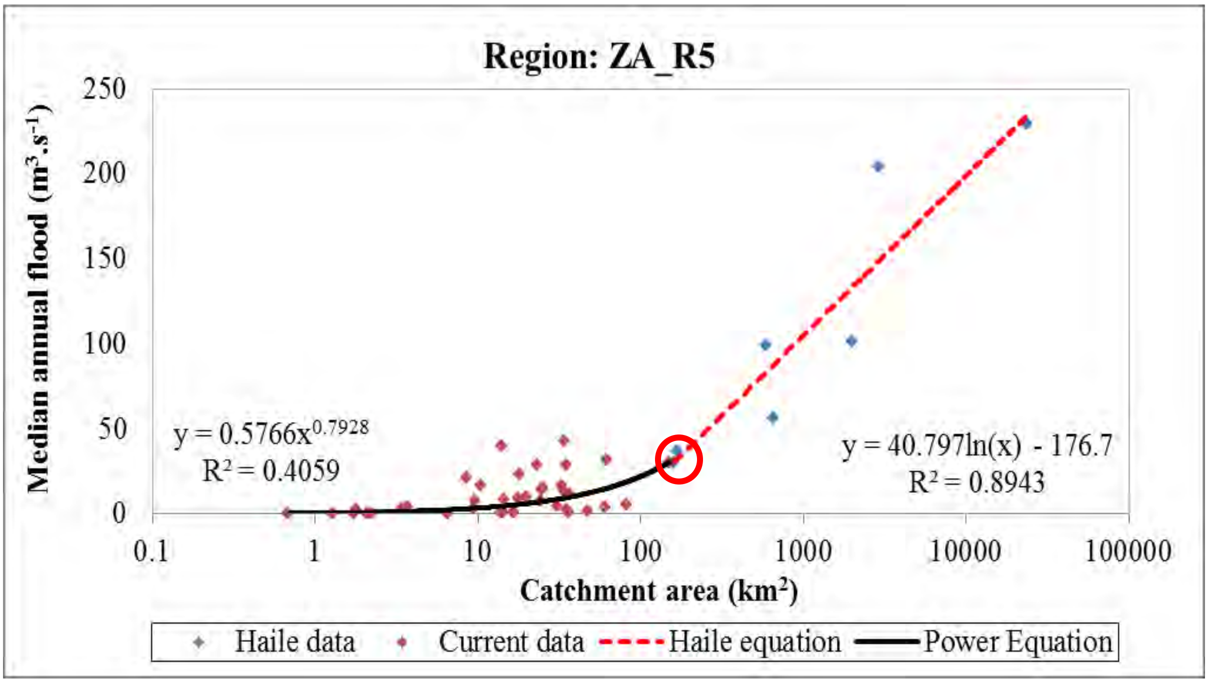

Figure 4.30 Derived equation to estimate an index flood in the ZA_R5 region

Table 4.6 contains the regional logarithmic-fitted regressions that have been provided by Haile (2011), as well as the regional power-fitted regressions that have been derived in this study, to apply the Haile Method in cases where the catchment area was below the area limit of the Haile regressions.

Table 4.6 Regional regression equations for the estimation of the index flood, using the Haile (2011) Method in South Africa

Region	Power regression	R^2	Logarithmic regression (Haile, 2011)	Area limits (km^2)	R^2
ZA_R1	$MEF = 0.2462(A)^{0.7668}$	0.29	$MEF = 14.755\ln(A) - 49.338$	28.30	0.37
ZA_R2	$MEF = 0.2015(A)^{0.3737}$	0.15	$MEF = 52.664\ln(A) - 340.28$	639.9	0.77
ZA_R3	$MEF = 0.8684(A)^{0.7624}$	0.63	$MEF = 66.461\ln(A) - 395.91$	386.5	0.51
ZA_R5	$MEF = 0.5766(A)^{0.7928}$	0.41	$MEF = 42.282\ln(A) - 187.10$	83.50	0.89

This chapter has provided all of the methodological procedures carried out in this study, from the data screening and station selection to the determination of the input catchment parameters for the RFFA methods, as well as the adjustments made to the Haile Method that have been addressed above. Having excluded an additional three stations in this section, the final number of stations to include in further analysis was 407 stations.

5. RESULTS

This chapter contains the results of the performance of the Meigh, Mkhandi, JPV and Haile methods. The 2- to 100-year design floods estimated by these methods were compared to the 2- to 100-year design floods estimated using an at-site frequency analysis of the observed data at the selected stations in this study. Section 5.1 deals with the selection and use of the GEV and LP3 distributions in this study.

5.1 Selection of Probability Distributions

The application of both the GEV and LP3 distributions for design flood estimation in South Africa have been advocated by a number of studies (Görgens, 2007; van der Spuy and Rademeyer, 2010). The Görgens (2007) study developed approaches to estimate design flood using both the GEV and LP3 distributions. Therefore design floods in this study were initially estimated using both the GEV and LP3 distributions, fitted by L-moments (Hosking and Wallis, 1997). Figure 5.1 illustrates a comparison between the observed design floods for both the GEV and LP3 distributions.

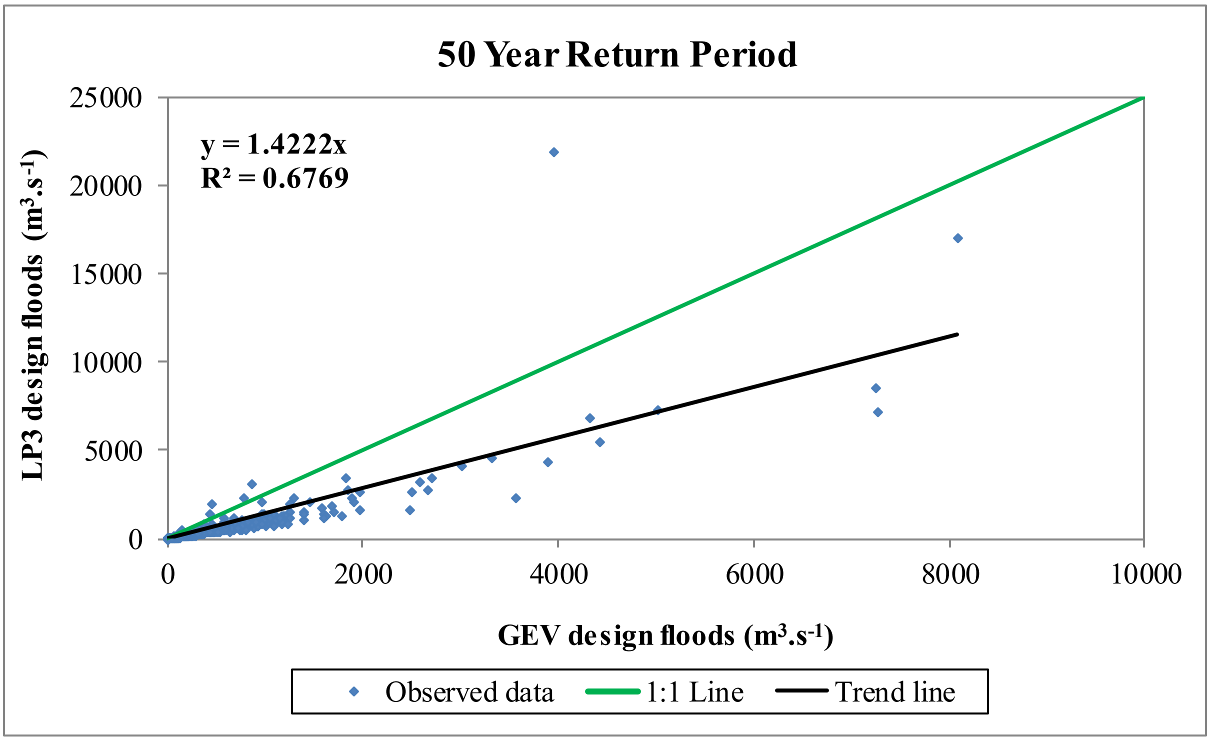

Figure 5.1 Comparison of observed design floods for the GEV and LP3 distribution, for the 50-year Return Period

It can be seen in Figure 5.1 that in certain cases the LP3 distribution produces design floods that are inconsistent with the GEV design floods. In these cases the observed LP3 design floods for the 50-year return period are orders of magnitude larger than the GEV design floods. The observed design floods using L-moments were determined in this study using a FORTRAN routine developed by Smithers (2014), using a FORTRAN library for the application of L-moments developed by Hosking and Wallis (1996). To ensure that the FORTRAN routine was calculating the observed design floods for the LP3 distribution accurately, station D8H005 was selected to be checked using a different program. The AMS for station D8H005 was checked by Kjeldsen (2015) using the R statistical software. The results produced by the R statistical software were the same as those produced using the FORTRAN routine therefore the software used in this study was correct.

In addition, Figures 5.2 and 5.3 which illustrate the performances of the JPV Methods indicate that not only are the observed LP3 design floods inconsistent but there are also cases where the JPV methods, using the LP3 distribution produce design floods that are orders of magnitude greater than both the JPV GEV design floods and design floods computed from observed AMF for the same stations. Therefore the JPV Methods were only assessed using the GEV distribution.

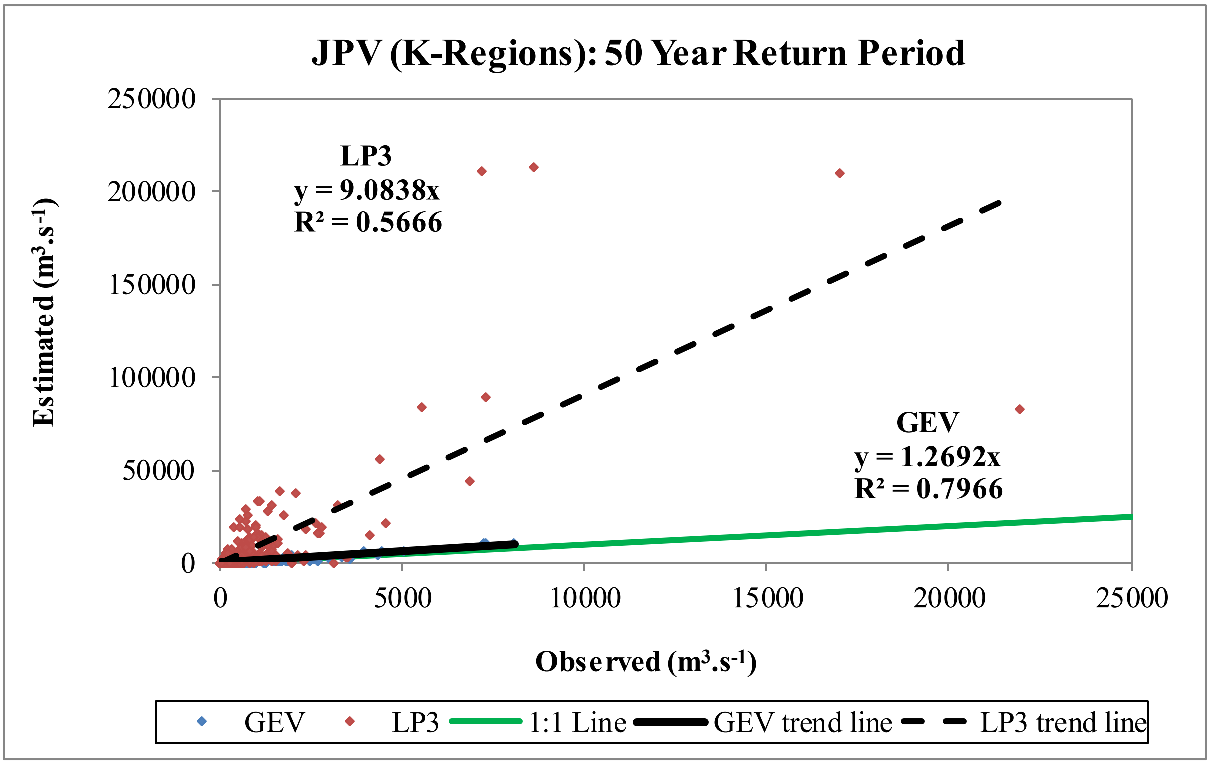

Figure 5.2 Performance of the JPV Method, using the K-Region regionalisation and the LP3 distribution, for the 50-year Return Period

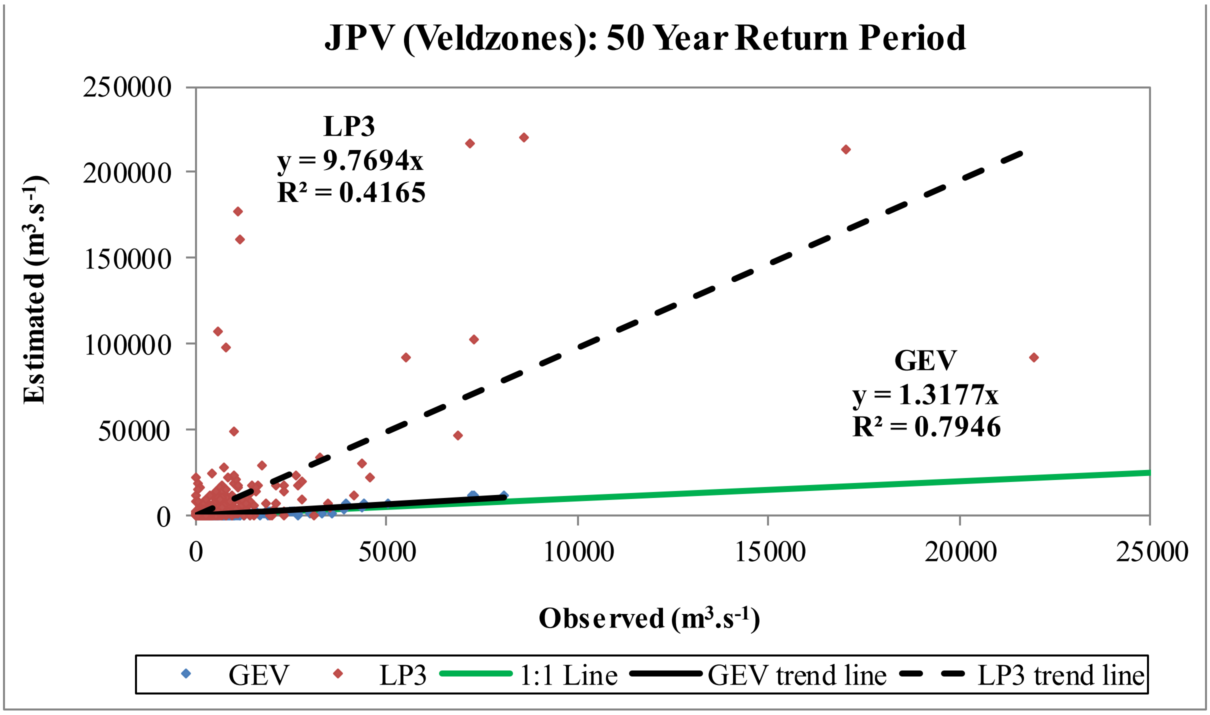

Figure 5.3 Performance of the JPV Method, using the Veld zone regionalisation and the LP3 distribution, for the 50-year Return Period

Due to the inconsistent performance of the LP3 distribution fitted to the observed data, and the JPV method applied using the LP3 option, the use of the LP3 distribution, and the JPV method with the LP3 distribution option, were removed from further analysis in this study and only the GEV distribution was utilised.

5.2 Performance of Methods

The following sections detail all of the assessment criteria utilised to assess the performance of the RFFA methods. These include graphical plots of the observed versus estimated design floods, computation of Relative Errors (RE), and a ratio of the estimated and observed design floods.

5.2.1 Slopes of observed versus estimated design floods

In order to determine whether or not the RFFA methods were generally over- or under-estimating design floods, graphs were produced representing a comparison of the observed design floods (at-site) to the estimated design floods (regional method). These are illustrated in Figures 5.4 to 5.8.

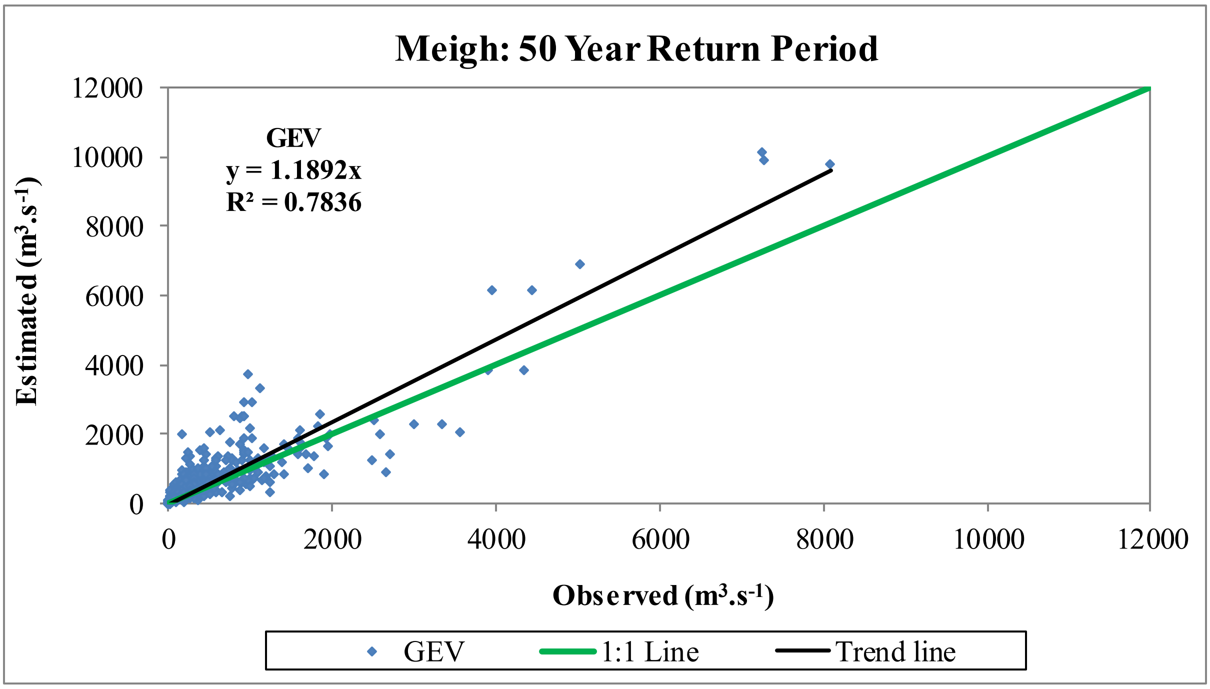

Figure 5.4 Performance of the Meigh Method, using the GEV distribution, for the 50-year Return Period

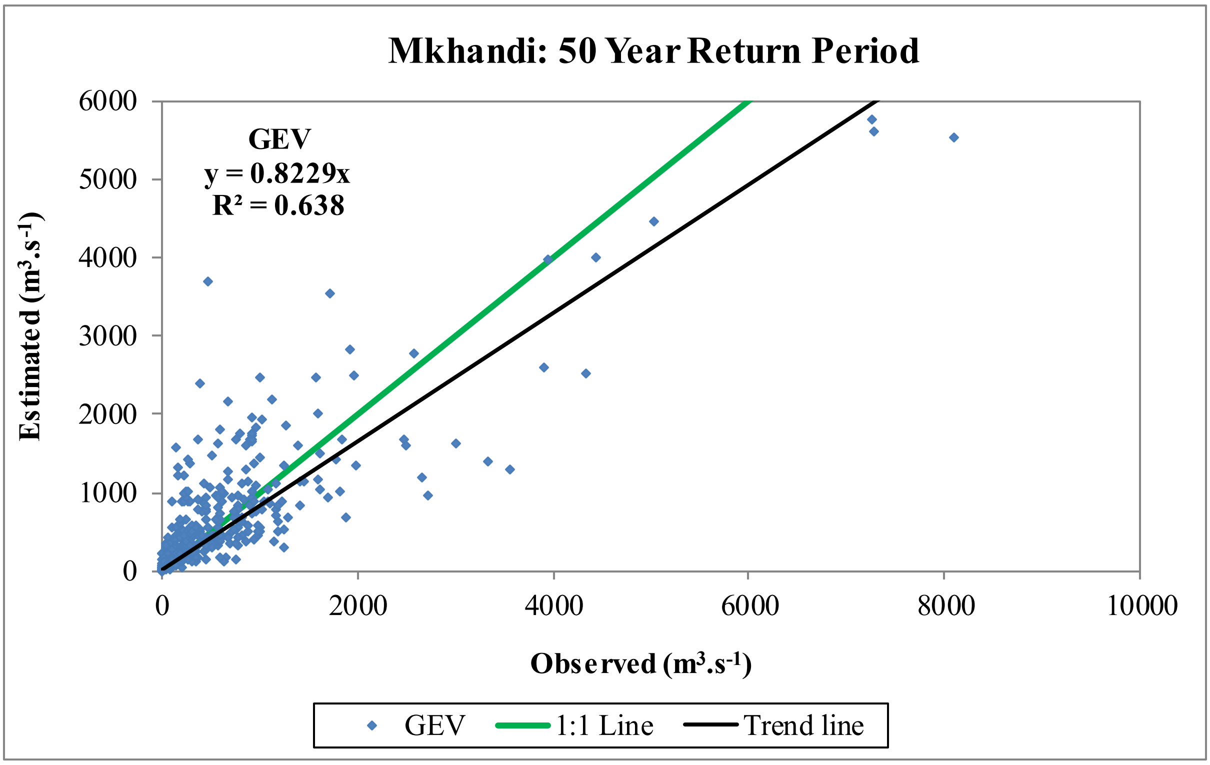

Figure 5.5 Performance of the Mkhandi Method, using the GEV distribution, for the 50-year Return Period

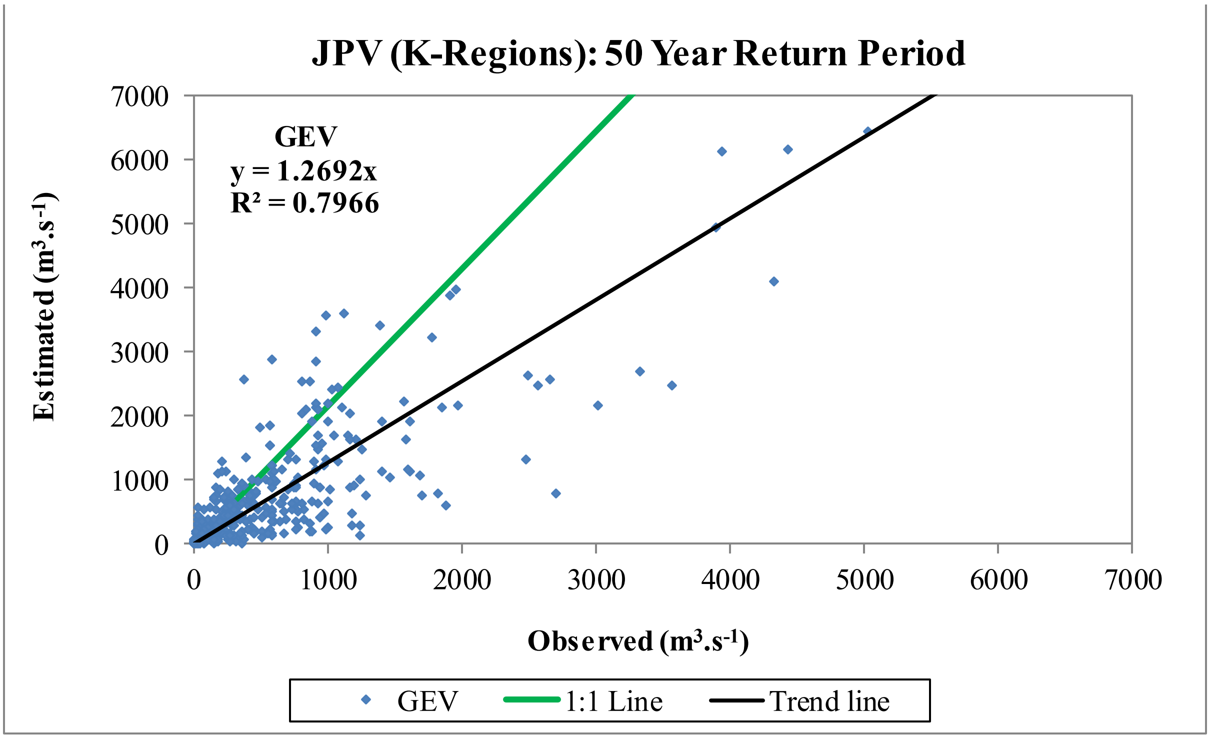

Figure 5.6 Performance of the JPV Method, using the K-Region regionalisation and the GEV distribution, for the 50-year Return Period

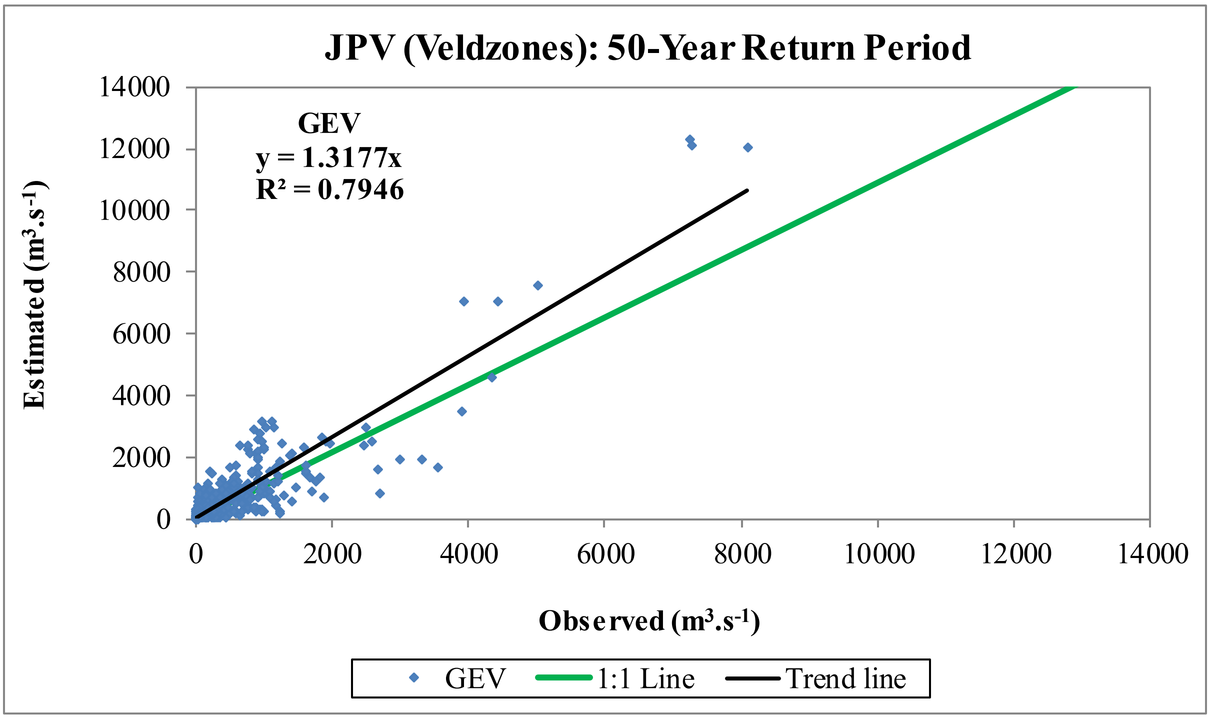

Figure 5.7 Performance of the JPV Method, using the Veld zone regionalisation and the GEV distribution for the 50-year Return Period

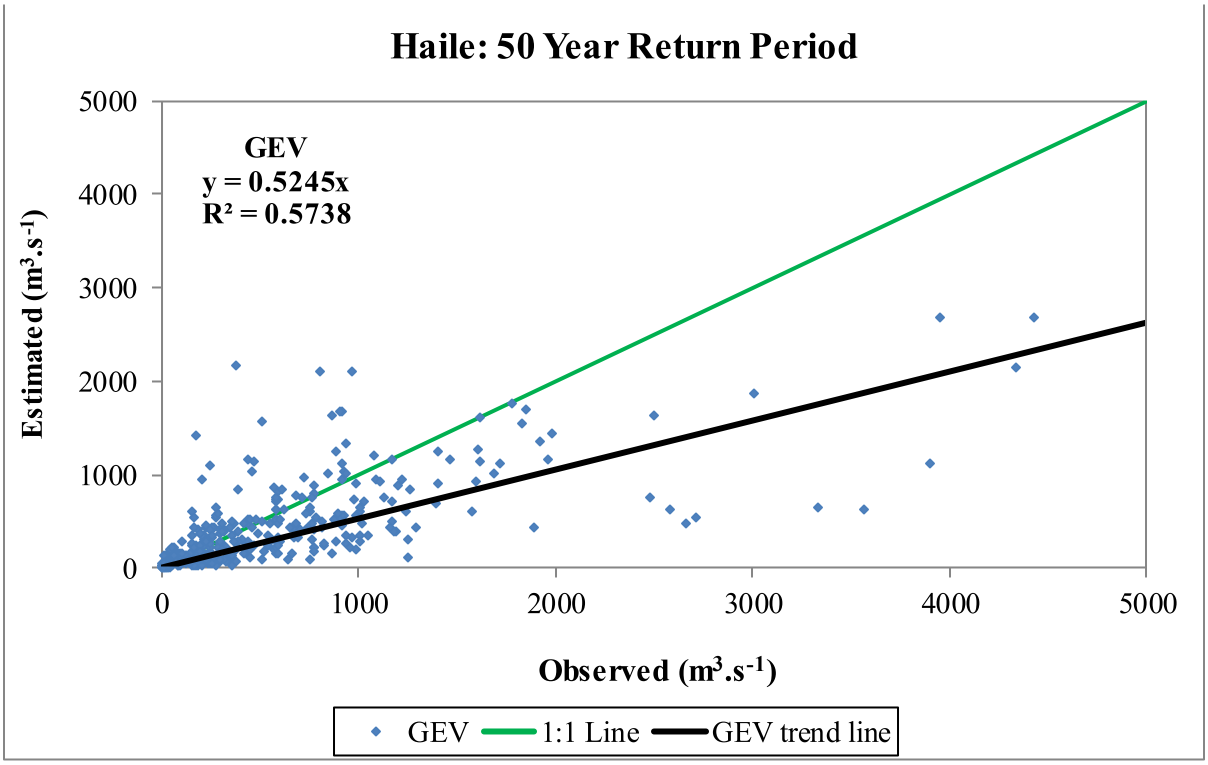

Figure 5.8 Performance of the Haile Method, using the GEV distribution for the 50-year Return Period

Table 5.1 provide the slopes that have been produced by the graphs of observed versus estimated design floods for the GEV distribution for the 2- to 100-year return periods.

Table 5.1 Slope of observed versus estimated design floods for the regional methods, using the GEV distribution

Method	Slope of observed versus estimated design floods					
	2	**5**	**10**	**20**	**50**	**100**
JPV (K-Region)	1.18	1.47	1.48	1.42	1.27	1.13
JPV (Veld zone)	1.23	1.52	1.52	1.47	1.32	1.18
Mkhandi	1.04	1.08	1.04	0.97	0.82	0.70
Haile	0.48	0.58	0.60	0.59	0.52	0.46
Meigh	1.03	1.09	1.14	1.17	1.18	1.17

It can be seen from Table 5.1 that the JPV Methods, using the K-Region and Veld zone regionalisations generally over-estimate design floods, producing a greater over-estimation for the 5- to 20-year return periods. The Meigh Method performs reasonably well with a general over-estimation for all return periods, while the Mkhandi Method over-estimates for the 2- to 10-year return periods and under-estimates for the 20- to 100-year return periods. The Haile method under-estimates for all return periods. The slopes produced in this study only provide a general indication of the method performance across all stations. Therefore, further assessment criteria have been utilised, and are presented in the following sections.

5.2.2 Relative errors

In order to determine the overall accuracy of the RFFA methods, *RE* (%) were calculated at each station. Box plots were utilised to represent the distribution of relative errors, considering both positive and negative errors. Absolute relative errors were represented through frequency plots, and were computed in the same way as the relative errors, however, without taking into consideration whether or not the *RE* is positive or negative. The *RE* and absolute *RE* values are an objective assessment of the degree of bias of the RFFA methods. It gives an absolute magnitude of the degree to which the design floods estimated using the RFFA methods differ from those determined using an at-site frequency analysis. The *RE* is expressed mathematically in Equation 5.1:

$$RE_{M,D} = \frac{[E_{M,T} - O_{D,T}]}{O_{D,T}} \times 100 \qquad (5.1)$$

where, $RE_{M,D}$ represents the relative error (%) for RFFA Method = M and probability distribution = D (GEV), $E_{M,T}$ represents the design flood estimated using RFFA Method = M and for return period = T (2, 5, 10, 20, 50 or 100 years) and $O_{D,T}$ represents the design flood estimated using observed AMS and probability distribution = D (GEV) for return period = T (2, 5, 10, 20, 50 or 100 years).

5.2.2.1 Box plots of relative errors

In this section, box plots have been created which display the relative errors for each method and each return period, considering both positive and negative errors, as shown in Figures 5.9 and 5.10. The black diamonds in each graph indicate the mean relative error for each method. It is important to note that the maximum value is not shown on the y-axis in each box plot, because a number of stations produced relative errors that exceeded 1000%, and were not excluded through the station selection criteria of this study. Instead, labels on each graph have been used to display the maximum values of each return period. Figures 5.9 and 5.10 provide box plots of the relative errors of each RFFA method for different return periods. The labels on the x-axes named K-reg and Veld zone represents the JPV method using the K-Region regionalisation and the Veld zone regionalisation, respectively.

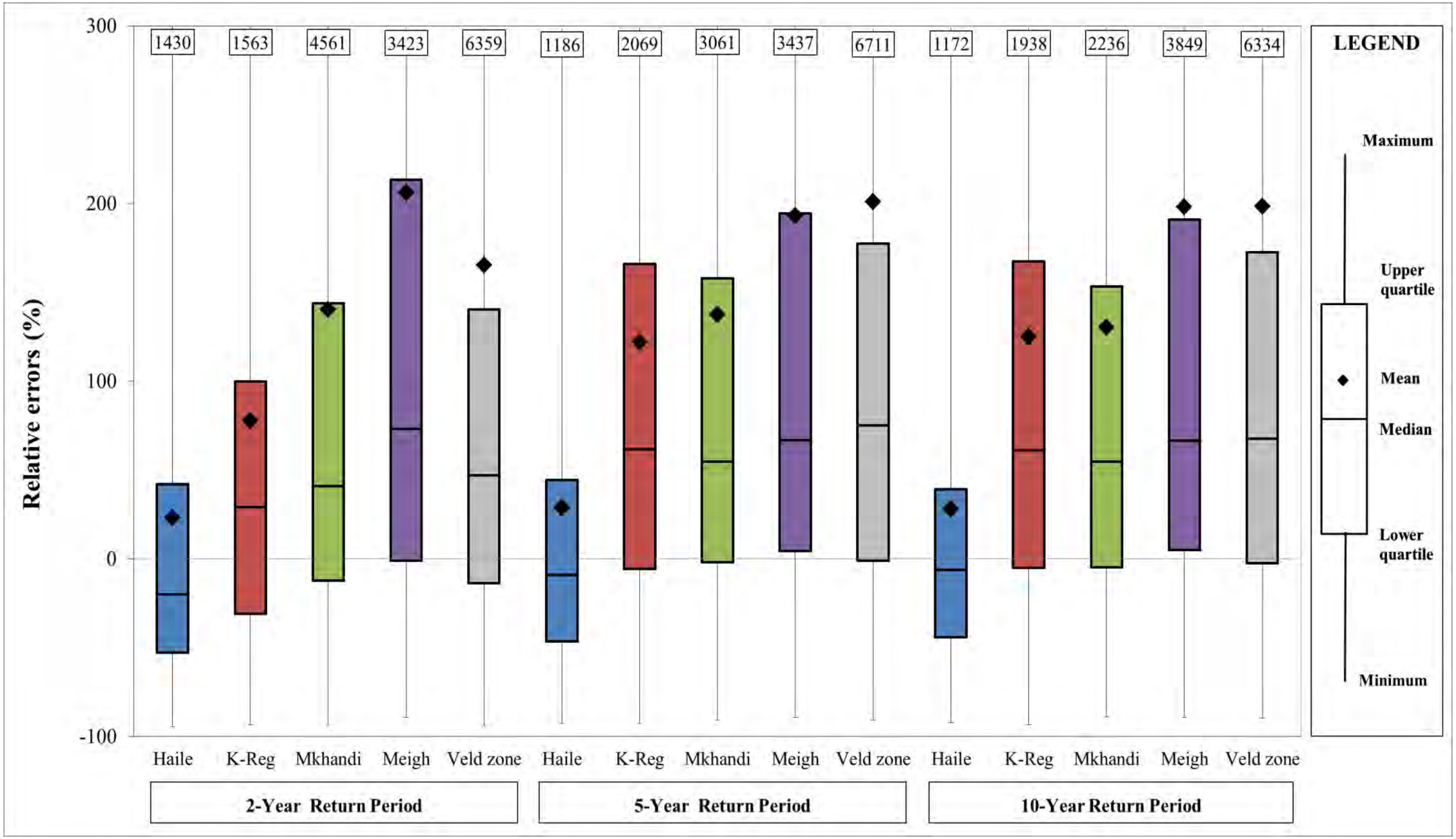

Figure 5.9 Box plots of relative errors for the 2- to 10-year return periods

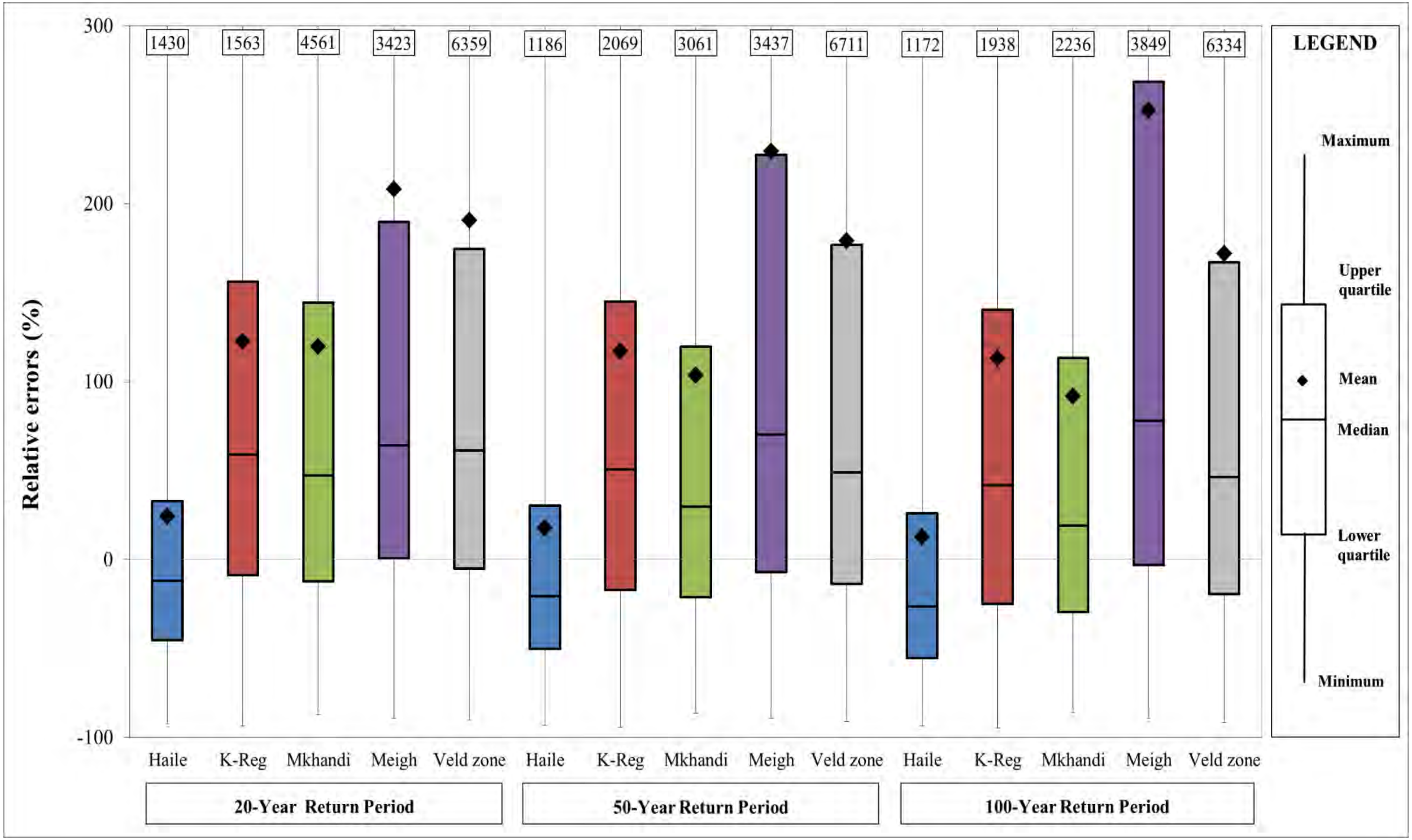

Figure 5.10 Box plots of relative errors for the 20- to 100-year return periods

It can be seen from Figures 5.9 to 5.10, that the Haile Method generally out performs all of the other RFFA methods. The Haile Method consistently produces a median relative error that is closer to zero for all return periods, than any other method. This indicates that there is less bias produced by the Haile Method in comparison to the other methods. The range between the 25th and 50th percentile values (widths) of the boxes of the Haile boxplots are narrower than the other methods indicating that 50% of the stations lay within a smaller range of relative errors. However, it can also be seen that between the 25th and 75th percentiles, there is almost an equal percentage of stations where the design floods are over-estimated as there were stations where the design floods are under-estimated. The Mkhandi Method and the JPV Method, using the K-Region regionalisation do not perform as well as the Haile Method, in terms of the widths of their boxplots, which indicate a larger range of errors; however, these methods generally over-estimate consistently, which may make them more applicable in design flood practice than the Haile Method which under-estimates for a large percentage of stations. For example, in the case of an over-estimation, more costs will be involved in the construction of the hydraulic structure to accommodate the over-estimated design flood. However, in the case of an under-estimation, the hydraulic structure will be inadequately designed resulting in possible failure and subsequently even greater costs will be incurred than in the case of an over-estimation, due to potential loss of life, the costs of repairing the structure and dealing with the damage that its failure has caused. The Meigh Method generally produces the greatest over-estimation, however, the JPV Method, using the Veld zone regionalisation, produced the highest relative errors for all return periods. These methods performed the worst out of all the RFFA methods.

With the exception of the Haile Method it can be seen that all of the RFFA methods are generally conservative and over-estimate for the majority of the stations. This is indicated by the 25th percentile for the box plots of these methods generally being larger than a relative error of 0%. It is important to note that the maximum relative errors exceeding 1000% in Figures 5.9 and 5.10 may be a result of poor method performance or the result of errors in the observed data set. A thorough data screening process has already been carried out in this study and stations have been removed for a number of reasons related to the observed data. Further analysis showed that generally all the methods did not all consistently perform poorly at the same site. Therefore, there is no further justification to exclude stations with large relative errors for some of the methods and all the selected stations were retained in the analysis.

Table 5.2 contains a summary of the relative errors represented in the boxplots in Figures 5.9 and 5.10. The average relative errors (%) across all return periods have been calculated for the lower quartile (25th percentile), median, upper quartile (75th percentile) and the inter-quartile range (upper quartile – lower quartile) and the RFFA methods have been ranked from best to worst (one to five) according to these averages. The methods have been ranked for the lower quartiles from the average relative error which is closest to zero to the average relative error which is furthest away from zero. For the median, upper quartile and inter-quartile range (IQR), the methods have been ranked from the lowest to the highest average relative error.

Table 5.2 Average relative errors for the lower quartile, median, upper quartile and inter-quartile range

Average relative errors (%)								
Method	Lower quartile		Median		Upper quartile		IQR	
	%	rank	%	rank	%	rank	%	Rank
Haile	-48.61	5	-15.29	1	36.24	1	84.85	1
Mkhandi	-13.59	3	41.05	2	138.75	2	152.34	2
JPV K-Region	-15.42	4	50.46	3	145.78	3	161.20	3
JPV Veld zone	-9.21	1	57.77	4	168.20	4	177.41	4
Meigh	10.05	2	79.99	5	224.22	5	214.17	5

It can be seen from the table above that the Haile Method out performs the other RFFA methods, by producing the lowest average median relative error, as well as the smallest IQR. However, the Haile Method ranks the worst when considering the average lower quartile relative error. This is indicative of the consistent under-estimation of the Haile Method. The Mkhandi Method and the JPV Method, using the K-Region regionalisation rank 2nd and 3rd, respectively, and the JPV Method, using the Veld zone regionalisation and the Meigh Method rank 4th and 5th, respectively, according to their average median relative errors. While the JPV Method, using the Veld zone regionalisation and the Meigh Method rank 1st and 2nd, according to the average lower quartile relative error, this is not an indication of these methods being acceptable for application but rather an indication that these methods generally do not under-estimate design floods. Based on the average median relative errors alone, one could consider the Haile Method to be the best RFFA for design flood practice, and the Mkhandi Method and the JPV Method, using the K-Region regionalisation may be considered to be acceptable methods for application. However, due to the under-estimation of the Haile Method, and the over-estimation of the Mkhandi Method and the JPV Method, using the K-Region regionalisation, none of the methods are suitable for application throughout South Africa.

5.2.2.2 Frequency plots of absolute relative errors

This section contains the results of the performance of each regional method, based on the frequency of the absolute relative errors (%) in the following percentage ranges: 0-20, 21-40, 41-60, 61-80, 81-100, 101-500, 501-1000 and >1000. Figure 5.11 illustrates the performance of the RFFA methods, based on the frequency of absolute relative errors for the 50-year return period.

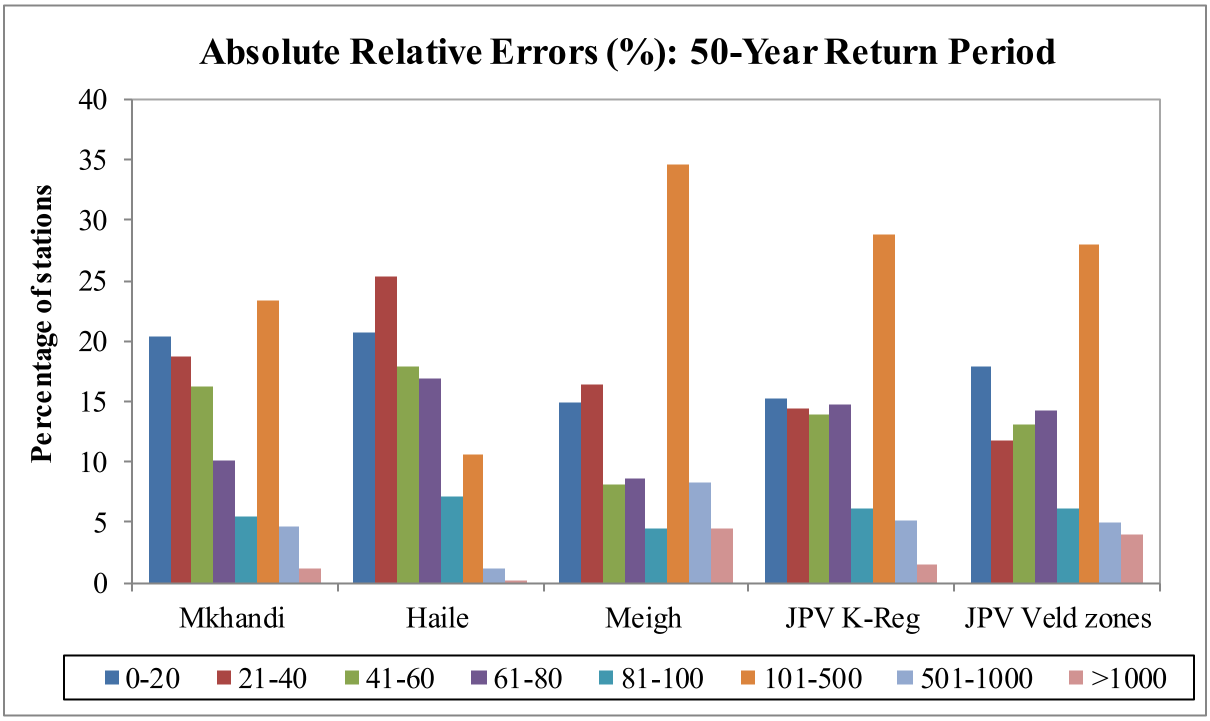

Figure 5.11 Performance of RFFA methods based on the frequency of absolute relative errors for the 50-year Return Period

Figure 5.11 illustrates that the Haile Method produces the highest percentage of stations with relative errors that fall into the lower ranges i.e. 0-20% and 20-40%, as well as the lowest percentage of stations, with errors that fall into the higher ranges i.e. greater than 100%. The Mkhandi Method also performs well; however, it has a much higher percentage of stations with relative errors that fall into the higher ranges. The Meigh Method performs poorly, with almost 50% of the stations producing relative errors greater than 100%. The JPV Method, using the K-Region regionalisation, produces stations with relative errors greater than 100% for approximately 35% of the stations. The JPV Method, using the Veld zone regionalisation, produces stations with relative errors greater than 100% for approximately 37% of the stations.

It is also important to note that the majority of the stations assessed produced relative errors that were greater than 100% and less than 500%, for all of the RFFA methods, with the exception of the Haile Method. Figure 5.11 also illustrates that while there are stations that produce relative errors that are greater than 1000%, these stations account for only a small percentage of the total stations being analysed. Similar trends are followed for the 2-, 10-, 20- and 100-year return periods, which are provided in Appendix E.

Table 5.3 contains the average and median Mean Absolute Relative Errors (*MARE*) for each RFFA method. The average and median *MARE* values gives an indication of the overall performance of the RFFA Methods. It summarizes the *RE* and absolute *RE* values for all return periods across all stations, so that the RFFA Methods can be ranked objectively from best to worst. The *MARE* was computed for each method and for every station, as shown in Equation 5.2 (Smithers *et al.*, 2015).

$$MARE_{M,D} = \frac{100}{407} \times \sum_{n=1}^{407} \sum_{T=1}^{6} \frac{|E_{M,T} - O_{D,T}|}{O_{D,T}} \qquad (5.2)$$

where, $MARE_{M,D}$ represents the mean absolute relative error (%) for RFFA Method = M and probability distribution = D (GEV) for all stations (407) used, $E_{M,T}$ represents the design flood estimated using RFFA Method = M and for return period = T (2, 5, 10, 20, 50 or 100 years), and $O_{D,T}$ represents the design flood estimated using observed AMS and probability distribution = D (GEV) for return period =T (2, 5, 10, 20, 50 or 100 years).

Table 5.3 Average and Median *MARE* (%) values for the RFFA methods

Absolute relative errors (%)				
Methods	**Average *MARE***	**Rank**	**Median *MARE***	**Rank**
Haile	74.00	1	44.72	1
JPV K-Region	138.65	2	65.80	3
Mkhandi	142.42	3	57.48	2
JPV Veld zone	208.42	4	71.82	4
Meigh	227.65	5	85.59	5

The table above indicates that the Haile Method performs the best producing the lowest average *MARE* and median *MARE*, while the Meigh Method performs the worst producing the highest average *MARE* and median *MARE*. It is important to note that the average *MARE* provides an absolute relative error that takes the average of the relative errors for the 2- to 100-year return periods for each station, and then divides the sum of all of these averages by the total number of stations analysed. This is effectively an average of an average. While this provides a general understanding of a method's performance, a more detailed investigation is required to understand whether or not the methods are over- or under-estimating design floods. The following section will deal with the over- or under-estimation by the RFFA methods.

5.2.3 Ratio of the estimated and observed design floods

The systematic over-estimation or under-estimation of a method gives an indication of the degree of bias in the method (Haddad and Rahman, 2012). The ratio statistic used to measure this degree of bias is defined as E/O, where E is the estimated design flood computed using the regional method, and O is the observed design flood computed using the at-site analysis. Haddad and Rahman (2012) considered three limits of this ratio to define the degree of bias produced by the regional methods. An E/O ratio that falls between 0.5 and 2 is an indication of a "desirable estimate (D)". An E/O ratio that is less than 0.5 is considered to be a "gross under-estimation (GU)" and an E/O ratio that is greater than 2 is considered to be a "gross over-estimation (GO)". These limits have been defined by (Haddad and Rahman, 2012) and are subjective; however, they do provide a reasonable indication of the accuracy of a method. Figures 5.12 and 5.13 illustrate the percentage of stations for each method that produced an E/O ratio that fall in either the GU, GO or D ranges for the 2- to 10-year return periods and the 20- to 100-year return periods, respectively.

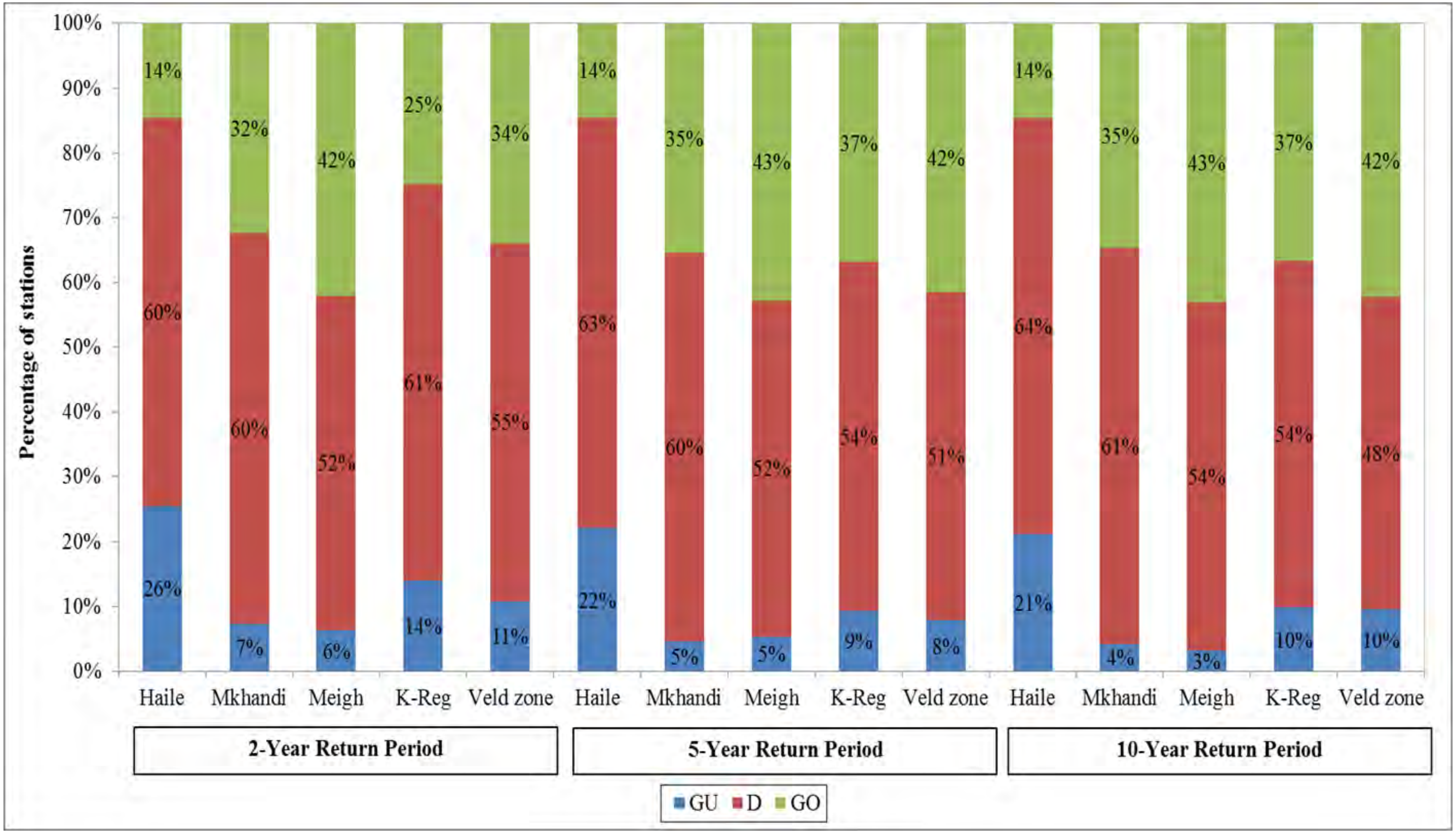

Figure 5.12 Estimated/observed ratios for the 2- to 10-year return periods

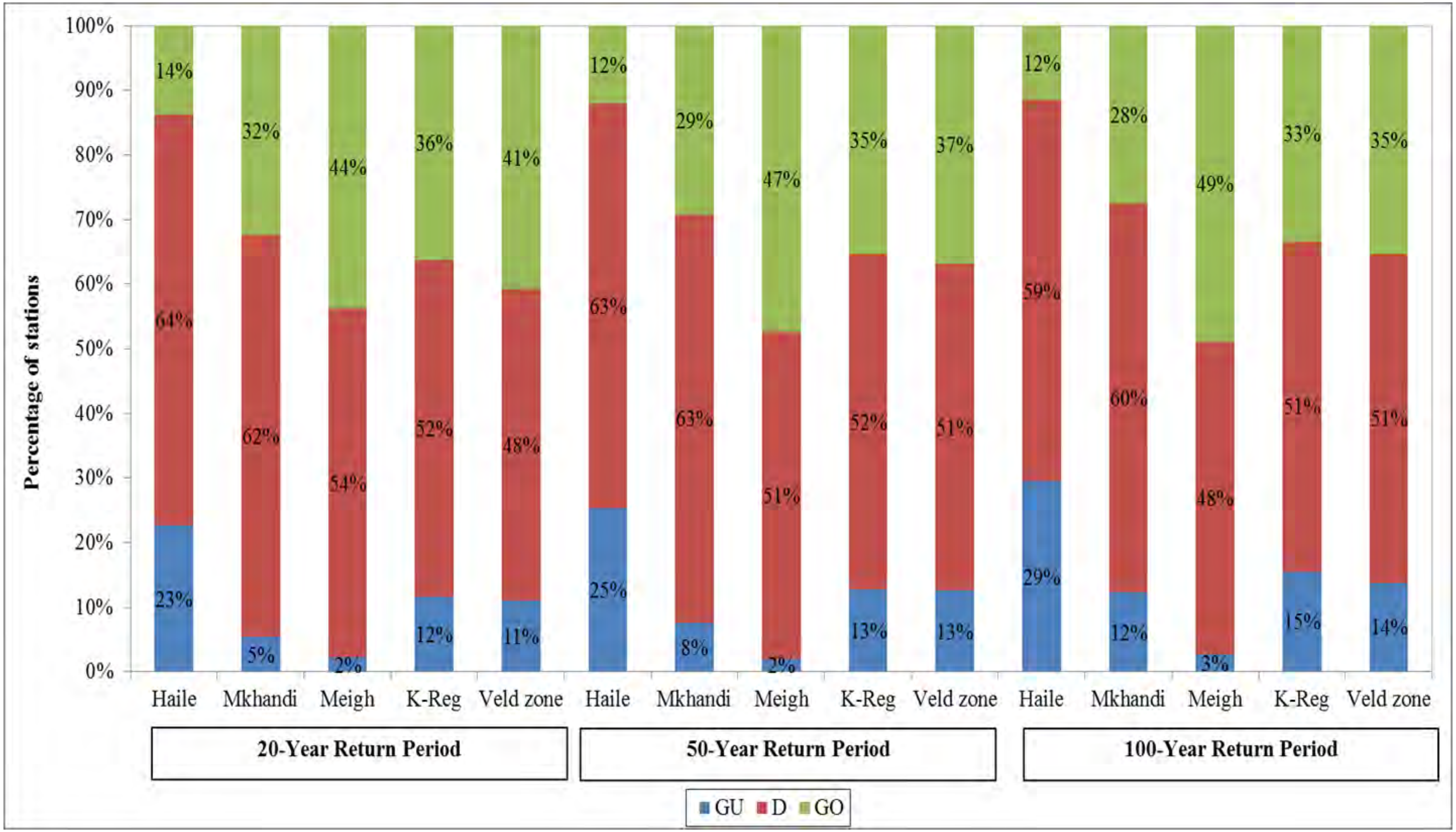

Figure 5.13 Estimated/observed ratios for the 20- to 100-year return periods

It can be seen from Figures 5.12 and 5.13 that the percentage of stations that fall within the desirable estimate range for the Haile Method is generally 60% or greater. However, more than 20% of the stations fall within the gross under-estimation range for all return periods. Similarly, approximately 60% or more of the stations fall within the desirable estimate range for the Mkhandi Method, however close to 30% of the stations fall within the gross over-estimation range. The JPV Method, using the K-Region regionalisation produces 50% or more stations that lie within the desirable estimate range, however approximately 35% of the stations fall within the gross over-estimation range. The JPV Method, using the Veld zone regionalisation produces approximately 48% or more stations that lie within the desirable estimate range, however approximately 35% of the stations fall within the gross over-estimation range. In the case of the Meigh Method almost half of the stations (between 41% and 49%) fall within the gross over-estimation range. This indicates poor performance, as almost half of the stations analysed produced design floods that are grossly over-estimated.

Table 5.4 provides a summary of the results, where the average GU, GO and D ranges have been calculated for each method across all return periods. The methods have been presented from the highest to lowest average ratio for the desirable estimate range.

Table 5.4 Average GU, GO and D ranges for all the return periods

RFFA Method	Average Number of Stations (%)		
	GU	D	GO
Haile	24	62	13
Mkhandi	7	61	32
JPV K-Region	12	54	34
Meigh	4	52	45
JPV Veld zone	11	51	38

It can be seen from Table 5.4 that the Haile Method on average produces the most stations that lie within the desirable estimate range. However, it also produces the highest number of stations that fall within the gross under-estimation range (24%) when compared to the other methods. The Mkhandi Method also produces almost the same percentage of stations as the Haile Method that lie within the desirable estimate range, however it also produces a high percentage of stations that fall within the gross over-estimation range. The JPV Method, using the K-Region regionalisation and the Veld zone regionalisation both perform similarly, producing more than 50% of the stations that fall within the desirable estimate range and more than 33% of the stations falling within the gross over-estimation range. The Meigh

Method performs the worst producing an average of 45% of the stations that fall within the gross over-estimation range.

5.2.4 Concluding remarks

This section contains a summary of the results that have been presented, particularly the average and median relative errors and absolute relative errors, as well as the average percentage of stations that fell within the desirable estimate range for each method.

The Haile Method has been ranked number one throughout the assessment, producing the lowest average median relative error (-15.29%), the lowest median *MARE* (44.72%) and the highest percentage of stations that lay within the desirable estimate range (62%). The Mkhandi Method and the JPV method, using the K-Region regionalisation were ranked 2nd and 3rd, respectively, producing the following percentages of errors and percentages of stations in the desirable estimate range: average median relative errors (41.05% and 50.46%), median *MARE* (57.48% and 65.80%) and the percentage of stations that lay within the desirable estimate range (61% and 54%). The JPV method using the Veld zone regionalisation and the Meigh Method performed the worst, being ranked 4th and 5th, respectively. The percentages of errors for these methods are as follows: the average median relative errors (57.77% and 79.99%) and the median *MARE* (71.82% and 85.59%). The desirable estimate range percentages indicate that the JPV Method, using the Veld zone regionalisation performs the worst (51%), rather than the Meigh Method (52%). From the above mentioned criteria, the average median relative error (%) and the average percentage of stations that lay within the desirable estimate range will be used to rank the methods from best to worst. The average and median *MARE* will not be further used, as these results only give a very general indication of method performance.

Having assessed the overall performance of the RFFA Methods, it is necessary to identify any spatial variations or variations pertaining to the input parameters utilised in each method. This will be dealt with in the following section.

5.3 Trends in Method Performance

The third objective of this study, as mentioned in Section 1.3, is to identify and compare any variations in method performance and to investigate the reasons for these variations. Therefore, this section deals with the variations that arise in method performance due to either the location where the method is applied or due to input parameters of the methods, such as catchment area.

5.3.1 Catchment areas

The 407 stations analysed in this study have been divided up according to three catchment area ranges i.e. stations with catchments that have areas that are less than or equal to 100 km^2 (111 stations), stations with catchment areas that are greater than 100 km^2 and are less than, or equal to, 1000 km^2 (151 stations) and stations with catchment areas that are greater than 1000 km^2 (145 stations). The E/O ratios have been calculated for all of the stations within each catchment area range for the 50-year return period and Figure 5.14 illustrates the percentage of stations that fall into either the GU, GO or D ranges for the different methods.

In general, all of the RFFA methods produce better results for catchments with larger areas, particularly those with areas greater than 1000 km^2. Conversely, the RFFA methods produce the worst results or the least percentage of stations within the desirable estimate range for smaller catchments with areas that are less than or equal to 100 km^2. The Haile Method produced the highest percentage of stations in the desirable estimate range for catchments with areas less than or equal to 100 km^2 and for stations with catchment areas greater than 1000 km^2. However, the Haile Method also produced the highest percentage of stations in the gross under-estimation range for all catchment area ranges. The Mkhandi Method produced the highest percentage of stations that lay within the desirable estimate range for catchments with areas that were greater than 100 km^2, while the Meigh Method performed the worst producing the lowest percentage of stations that lay within the desirable estimate range for all catchment area ranges, with the exception of the stations that were greater than 100 km^2 and less than or equal to 1000 km^2, where the JPV methods using the K-Region regionalisation and the Veld zone regionalisation produced the lowest percentage of stations (52% and 50%, respectively) in the desirable estimate range .

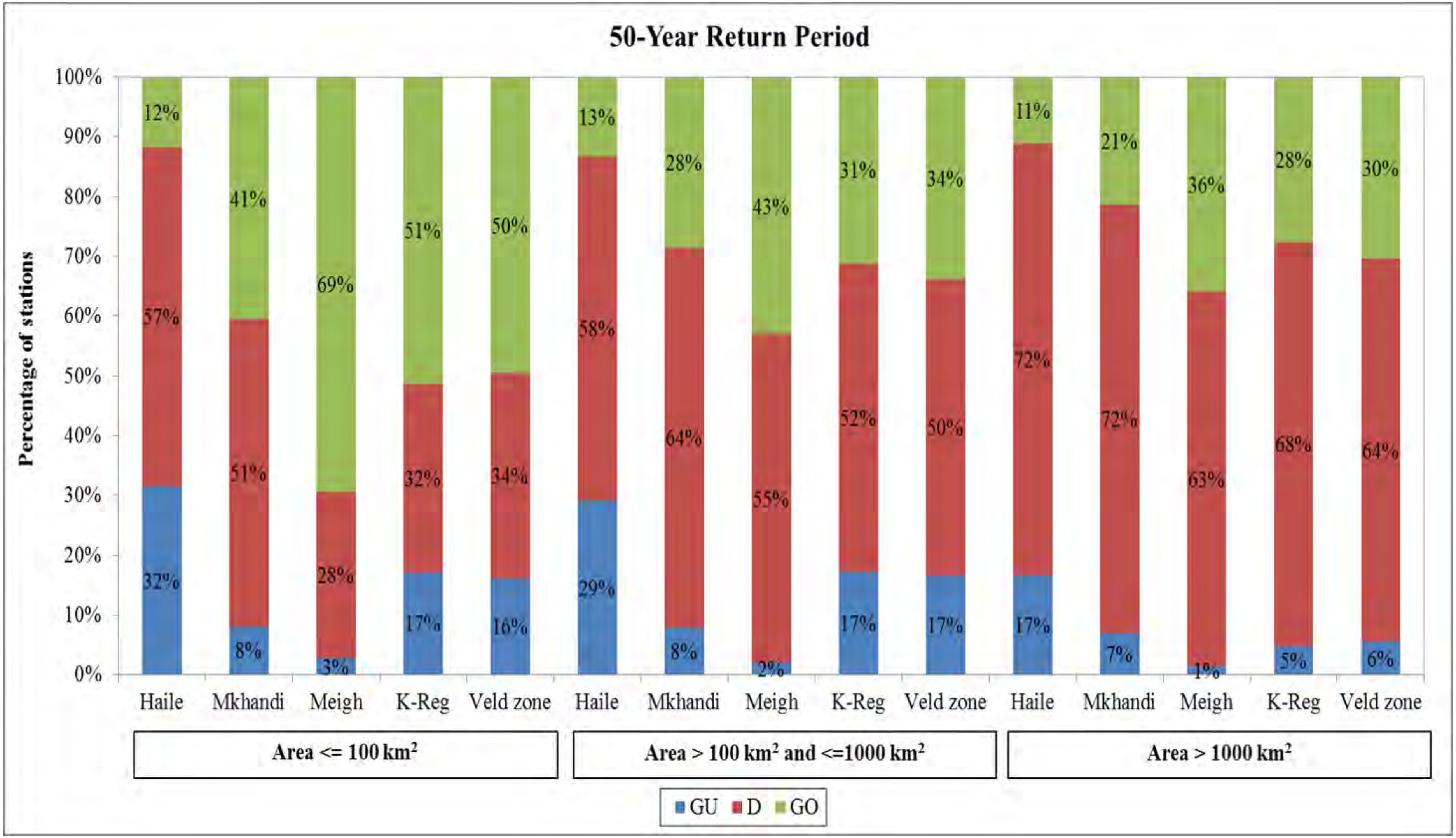

Figure 5.14 Estimated/observed ratios for the 50-year return period at different catchment area ranges

5.3.2 Spatial variations

This section will deal with the variation in method performance, based on where the method is applied in South Africa. In order to determine the spatial variation in method performance, the methods were ranked from one to five (best to worst) according to the method that produced an E/O ratio that was closest to one at a particular station. Thereafter, the location of every station, where a particular method ranked number one, was plotted in ArcGIS to represent that method. This was done for all the RFFA methods to produce the map in Figure 5.15, which represents the 50-year return period. The same procedure has been followed for the 2-, 10-, 20- and 100-year return periods and the maps produced are provided in Appendix F.

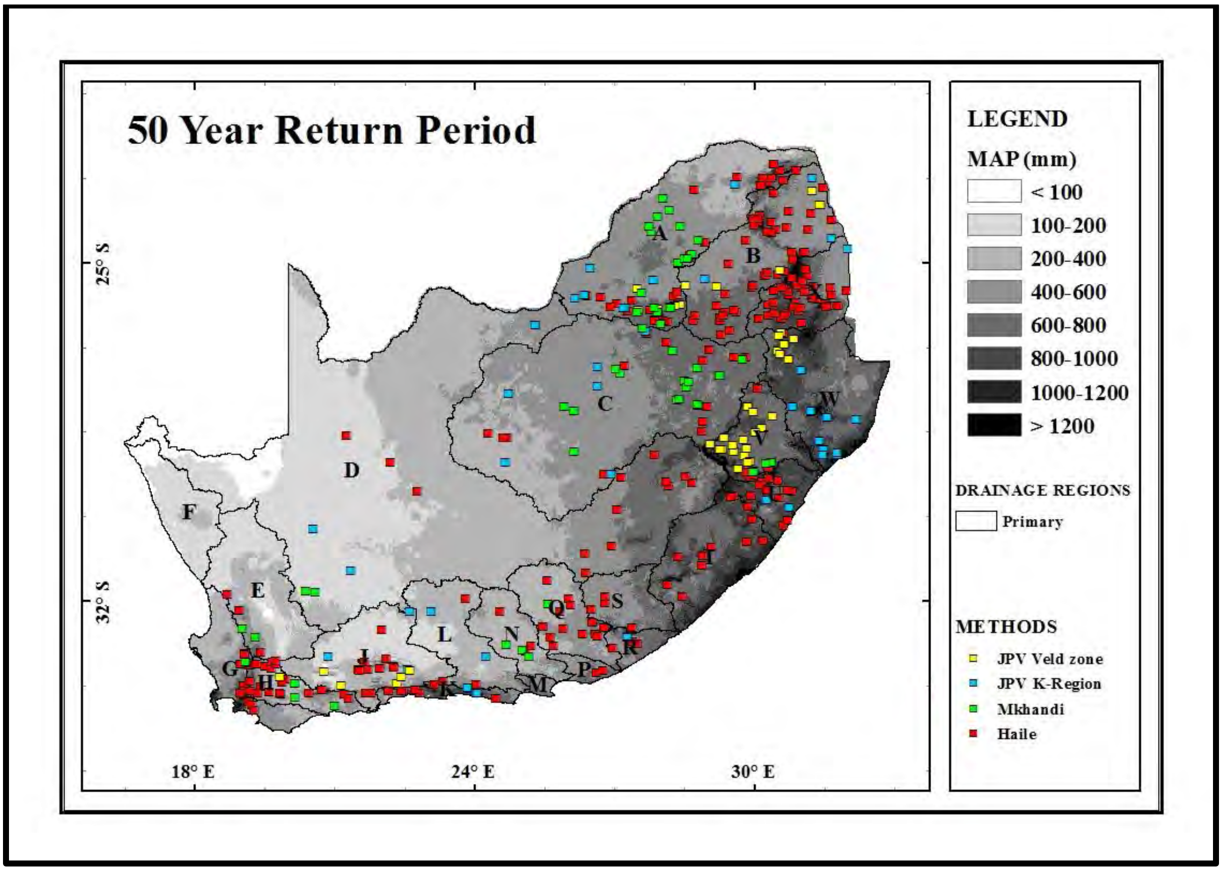

Figure 5.15 Method rank for the 50-year return period

It is evident from Figure 5.15 that there is some clustering of methods being ranked as the best in certain regions. The JPV Method, using the Veld zone regionalisation, represented by yellow squares, is ranked number one, exclusively in drainage region V. The JPV Method, using the K-Region regionalisation, represented by blue squares, is ranked number one, exclusively in drainage region W. However, these are minor spatial trends, as the Haile Method continues to perform the best and is ranked number one throughout the coastal areas of South Africa, as well as for the northern regions of the country. The map illustrates not only the methods ranked number one, but also the MAP and the primary drainage regions across the South Africa. Therefore, it is evident that the primary drainage region and the MAP do not have a significant effect on method performance. For the 50-year return period, the Meigh Method is not ranked as the best in any of the stations, however, for other return periods there are several cases where it does, which can be seen in Appendix F. In general, the Haile Method was ranked as the best method for the highest number of stations throughout the country. However, it is important to note that a method being ranked as the best for a particular station does not indicate that the results are acceptable. The method may produce better results than the other methods but it may still be an unacceptable or inconsistent design flood estimate. It can be concluded from Figure 5.15 that there are no major spatial variations in method performance.

6. DISCUSSION AND CONCLUSIONS

This chapter includes a synthesis of some of the challenges encountered in this study, a summary of the main findings and a discussion on the results obtained. The final concluding remarks of this chapter will relate to the objectives of this study that are presented in Section 1.3.

6.1 Data Issues

This study entailed an analysis of the AMF data for 1458 stations across South Africa. After screening, only 318 flow gauging stations and 89 dam inflow records were used in the final analysis. One of the most common reasons for the exclusion of stations in this study was the result of observed river stages exceeding the maximum rated stage in the rating table. In cases where the rating table exceedence was greater than 20%, these stations were removed from the analysis. Due to the large number of stations (217) that were excluded, based on this criteria, it would be beneficial for a method to be developed to extend the rating tables of the current stations, in order to produce better estimates of extreme flood events in the future. A number of stations were also removed due to the influence of upstream dams. This was done visually in ArcGIS, where any station that was seen to be downstream of a dam was removed from the analysis, if the station produced unusually low streamflow, due to the impoundment of the river flow. Due to the large number of dams across South Africa, a large number of stations had to be removed. It may be beneficial to develop an index that accounts for the attenuation of streamflow due to an upstream dam, such as the index of flood attenuation from reservoirs and lakes (FARL) developed by Kjeldsen *et al.* (2008). This would allow the use of a large number of stations despite their location downstream of a dam. It is important to note that despite the thorough data screening process undertaken in this study, a number of stations still produced inconsistent design flood estimates. Therefore, it is of utmost importance that detailed quality checks are conducted, when using any of the datasets from the DWS.

In addition to the issues encountered in obtaining suitable AMF data, a number of issues were encountered in determining the catchment parameters that were used in the application of the RFFA methods. Catchment area was a parameter that was required for the application of all the RFFA methods. The DWS provided the catchment area for each station; however, the shapefiles for these catchments do not exist. Therefore, these shapefiles were produced in ArcGIS for each station, to be analysed in this study. It may be beneficial for shapefiles to be created for the catchment area of all 1458 stations, so that the catchment area of the shapefile is consistent with the catchment area reported by the DWS. Due to the catchment shapefiles for the DWS areas not being available, it was necessary for these shapefiles to be produced in this study using the ArcGIS 10.2.1 software. A comparison between the areas determined using ArcGIS to the areas provided by the DWS was undertaken and represented using a scatter plot. The scatter plot illustrated that there is a good correlation between the DWS areas and the ArcGIS estimates indicated by an R^2 value of 0.9998. There was also no scatter above or below the 1:1 line, indicating that there was very little over- or under-estimation of the catchment areas estimated using ArcGIS.

The Veld zones determined in this study for selected stations were compared to those reported by Görgens (2007). It was found that for 12 stations the Veld zones reported by Görgens (2007) differed from those determined in this study. This often occurred when the catchment of a station fell into two different Veld zone groups i.e. A/B or B/C. The catchment shapefiles of these 12 stations were edited, so that a line was digitized along the boundary between the two Veld zone groups in which a catchment was found. The areas of the catchment above and below this digitized line were determined, using the Xtools Pro 11.1 software. The final dominant Veld zone group was determined as the Veld zone group within which a greater portion of the catchment was found. The results produced indicated that the Veld zones determined in this study were correct.

Similarly, the K-Regions determined in this study for several stations differed from those reported by Görgens (2007). The dominant K-Region for a catchment was determined as the K-Region at the catchment outlet. Generally the K-Regions reported by Görgens (2007) differed from those determined in this study for stations that were located close to the boundary between two K-Regions. The Görgens (2007) report does not indicate how the pooled K-Regions were digitized. Therefore it is possible that the differences in the K-Regions reported by Görgens (2007) to those determined in this study are due to different sources of K-Region boundaries. The differing Veld zones and K-Regions may also provide a reason for the poor performance of the JPV Methods in certain cases.

6.2 Performance of RFFA methods

Through the literature reviewed in this study, it was found that the GEV and LP3 distributions were most commonly used in design flood practice in South Africa. However, a comparison of the observed design floods produced by the GEV distribution versus the observed design floods produced by the LP3 distribution indicated that the LP3 distribution often produced inconsistent observed design flood estimates. In certain cases the observed LP3 design floods were orders of magnitude larger than the observed GEV design floods. To ensure that the LP3 design floods were being computed correctly by the FORTRAN routine used in this study, the AMS of a selected station was used to compute the LP3 design floods using different software i.e. the R statistical software. This check was carried out by Kjeldsen (2015) and it was concluded that the results produced for both the FORTRAN routine and the R statistical software were the same. In addition, the JPV Methods using the LP3 distribution also produced design flood estimates that were orders of magnitude larger than the JPV Method using the GEV estimate. Due to the LP3 distribution producing inconsistent design flood estimates, it was removed from further analysis and only the GEV distribution was used in the assessment.

The slopes produced by the graphical plots of observed versus estimated design floods using the GEV distribution indicate that the JPV Methods, using the K-Region and Veld zone regionalisations generally over-estimate design floods, producing a greater over-estimation for the 5- to 20-year return periods. The Meigh Method performs reasonably well with a general over-estimation for all return periods, while the Mkhandi Method over-estimates for the 2- to 10-year return periods and under-estimates for the 20- to 100-year return periods. The Haile method generally under-estimates for all return periods. The slopes produced in this study only provide a general indication of the method performance across all stations. Therefore, further criteria were used to assess method performance. This included relative errors and absolute relative errors, which were represented through the use of box plots and frequency plots, as well as a ratio utilised by Haddad and Rahman (2012), which gives an indication of the degree of bias produced by the RFFA method.

The box plots of relative errors illustrated the distribution of errors and gave an indication of the whether or not the method consistently over- or under-estimated design floods. In order to further synthesize the results of the box plots, the average relative errors (%) across all return periods have been calculated for the lower quartile (25th percentile), median, upper quartile (75th percentile) and the inter-quartile range (upper quartile – lower quartile) and the RFFA methods have been ranked from best to worst (one to five) according to these averages. The Haile Method out performs the other RFFA methods, by producing the lowest average median relative error, as well as the smallest IQR. However, the Haile Method ranks the worst when considering the average lower quartile relative error. This is indicative of the consistent under-estimation of the Haile Method. The Mkhandi Method and the JPV Method, using the K-Region regionalisation rank 2nd and 3rd, respectively, and the JPV Method, using the Veld zone regionalisation and the Meigh Method rank 4th and 5th, respectively, according to their average median relative errors. While the JPV Method, using the Veld zone regionalisation and the Meigh Method rank 1st and 2nd, according to the average lower quartile relative error, this is not an indication of these methods being acceptable for application but rather an indication that these methods generally do not under-estimate design floods. Based on the average median relative errors alone, one could consider the Haile Method to be the best RFFA for design flood practice, and the Mkhandi Method and the JPV Method, using the K-Region regionalisation may be considered to be acceptable methods for application. However, due to the under-estimation of the Haile Method, and the over-estimation of the Mkhandi Method and the JPV Method, using the K-Region regionalisation, none of the methods are suitable for wide application throughout South Africa.

The frequency plots of absolute errors indicated the percentage of stations that produced relative errors in the following ranges: 0-20, 21-40, 41-60, 61-80, 81-100, 101-500, 501-1000 and >1000. To further synthesize these results, the average *MARE* and the median *MARE* was calculated for all of the RFFA methods. These results indicated that the Haile Method performs the best producing the lowest average *MARE* and median *MARE*, while the Meigh Method performs the worst producing the highest average *MARE* and median *MARE*. While this provided a general understanding of the performance of the methods, a more detailed investigation is required to understand whether or not the methods are over- or under-estimating design floods. Therefore the ratio of the estimated (E) and observed (O) design floods as defined by Haddad and Rahman (2012) was utilised.

The E/O ratio can fall into one of three categories to define the degree of bias produced by the regional methods. An E/O ratio that falls between 0.5 and 2 is an indication of a "desirable estimate (D)". An E/O ratio that is less than 0.5 is considered to be a "gross under-estimation (GU)" and an E/O ratio that is greater than 2 is considered to be a "gross over-estimation (GO)". The percentage of stations that lay within each of these ranges was calculated for each method and for each return period. To summarise these ratios, the average percentage of stations to fall within the GU, GO and D ranges across all return periods were calculated and the methods were ranked according to the percentage of stations within the desirable estimate range. It was found that the Haile Method on average produces the most stations that lie within the desirable estimate range (62%). However, it also produces the highest number of stations that fall within the gross under-estimation range (24%) when compared to the other methods. The Mkhandi Method also produces similar average percentages of stations as the Haile Method that lie within the desirable estimate range, however it also produces a high average percentage of stations that fall within the gross over-estimation range. The JPV Method, using the K-Region regionalisation and the Veld zone regionalisation both perform similarly, producing more than 50% of the stations that fall within the desirable estimate range and more than 33% of the stations falling within the gross over-estimation range. The Meigh Method performs the worst producing an average of 45% of the stations that fall within the gross over-estimation range.

In order to test the performance of the methods with different catchment sizes, the 407 stations analysed in this study were divided in to three catchment area ranges i.e. stations with catchments with areas less than or equal to 100 km^2, stations with catchment areas that are greater than 100 km^2 and are less than, or equal to, 1000 km^2 and stations with catchment areas that are greater than 1000 km^2. Thereafter, the E/O ratios were calculated for all of the stations within each catchment area range for the 50-year return period. In general, all of the RFFA methods produced better results for catchments with larger areas, particularly those with areas greater than 1000 km^2. Conversely, the RFFA methods produce the worst results or the least percentage of stations within the desirable estimate range for smaller catchments with areas that are less than or equal to 100 km^2.

In order to determine the spatial variation in method performance, the methods were ranked from one to five (best to worst), according to the method that produced an E/O ratio that was closest to one at a particular station. Thereafter, the location of every station, where a particular method ranked number one, was plotted in ArcGIS to represent that method. It was found that neither MAP nor the drainage regions have any discernable effect on the methods' performance. In general, the Haile Method was ranked as the best method for the highest number of stations throughout the country. However, it is important to note that a Method being ranked as the best for a particular station does not indicate that the results are acceptable. The method may produce better results than the other methods but it may still be an unacceptable or inconsistent design flood estimate. It can be concluded that there are no major spatial trends that were found for any of the RFFA methods.

6.3 Concluding Remarks

The four main objectives of this study are as follows: **Objective 1**: To review the literature pertaining to the current methods employed in RFFA, both locally and internationally, in order to advocate the selection of methods to be assessed in this study, **Objective 2**: To apply and assess the performances of the selected RFFA methods in South Africa, **Objective 3**: To identify and compare any variations in method performance and the reasons for these variations, **Objective 4**: To select a suitable method based on its performance or recommend the development of a new approach.

Objective 1 has been met, as the literature has been reviewed and it was found that the Meigh, Mkhandi, JPV and Haile Methods are most suitable for a nationwide study. **Objective 2** has been met as the methods have been applied and their performances have been assessed, using the slope between estimated and observed values, the relative errors and absolute relative errors displayed using box plots and frequency plots, as well as the E/O ratio. **Objective 3** has been met by producing maps to assess the spatial variation in the performance of the methods and through the catchment area analysis, discussed in Section 5.3.1. **Objective 4** will be addressed later in this chapter.

The following research questions were presented in Section 1.4 and can be answered as follows:

a) How well do the design flood estimates obtained from the RFFA methods compare to the estimates obtained from an at-site FFA in a nationwide study?

   The best criteria to assess the methods have been the average median relative error, determined from the box plot of relative errors and the average percentage of stations with E/O ratios that lay within the desirable estimate range. The ranking of the RFFA methods were similar when using both of the aforementioned criteria.

The Haile Method has been ranked number one, producing the lowest average median relative error (-15.29%) and the highest percentage of stations with ratios that lay within the desirable estimate range (62%). The Mkhandi Method and the JPV method, using the K-Region regionalisation were ranked 2nd and 3rd, respectively, producing average median relative errors of 41.05% and 50.46%. The average percentage of stations that produced ratios that lay within the desirable estimate range for these methods were 61% and 54%, respectively. The JPV Method, using the Veld zone regionalisation and the Meigh Method performed the worst, being ranked 4th and 5th respectively. These methods produced average median relative errors of 57.77% and 79.99%, respectively. According to the average percentage of stations that produced ratios that lay within the desirable estimate range, the JPV Method, using the Veld zone regionalisation, is ranked as the worst method (51%) followed by the Meigh Method which produced an average percentage of 52%.

b) What are the advantages and disadvantages of each method?

The Haile Method generally produced better results than the other RFFA methods analysed; however, this method also consistently under-estimates. In addition, the Haile regression equations were not applicable to all stations in South Africa due to the area limits produced by the logarithmic equations. The Mkhandi Method produced reasonable results and produced better results than the other RFFA methods analysed for catchments with areas that are greater than 100 km^2 and are less than or equal to 1000 km^2; however, this method also consistently over-estimated design floods. Similarly, to the Mkhandi Method, the JPV Method, using the K-Region regionalisation produced reasonable results but consistently over-estimated design floods. From a design perspective this may indicate that the Mkhandi Method and the JPV Method, using the K-Region regionalisation are possibly suitable for application in design flood practice. However, due to large errors produced by these methods for a number of stations, they are not recommended for use. The JPV Method, using the Veld zone regionalisation and the Meigh Method both consistently over-estimated design floods for a greater number of stations than any other RFFA method and produced the worst results.

In addition, none of the methods provide adjustments for application in highly urbanised areas as well as adjustments to the methods which would account for climate change scenarios. These are drawbacks that should be accounted for in future methods.

c) Are there any variations in method performance relating to input parameters, or relating to where the method is applied?

The Haile Method generally performed the best at catchment areas that are less than or equal to 100 km^2 and for catchments with areas that are greater than 1000 km^2, however the method continued to produce a high percentage of stations where design floods were under-estimated even at these catchment area ranges. The Mkhandi Method performed the best for catchments with areas that are greater than 100 km^2 and are less than or equal to 1000 km^2. Generally the Meigh Method performed the worst regardless of what catchment area range was analysed. In addition, it was also found that there were no major spatial variations in method performance.

d) Which method is best suited for use in South Africa?

According to the ranking of the RFFA methods which have been discussed in question (b) in this section; the Haile Method, Mkhandi Method and the JPV Method, using the K-Region regionalisation have been ranked as the top three methods, respectively. As far as the overall performance of the methods, the Haile Method consistently outperformed the Mkhandi Method and the JPV Method and should therefore be considered as the best method for use in South Africa. However, the Haile Method consistently under-estimates design floods. In addition, the impact of the adjustments that have been made to the Haile Method in this study, for the estimation of the *MEF* need to be assessed, particularly for regions ZA-R1 and ZA_R2. The Mkhandi and JPV Methods consistently over-estimated design floods, which may indicate that these methods are more suitable for application as an over design may be acceptable in design flood practice. However, it is the extent of over-estimation by these methods that make them unacceptable for use, as both of these methods produce an average percentage of stations that lay within a gross over-estimation range that is approximately 30% or greater.

e) Is the development of a new RFFA method warranted, given the performance of the current RFFA methods?

From the analysis, it was found that the Haile Method generally performed the best. However, it was found through further investigation that this method under-estimates design floods, almost 50% of the time. In addition, the Haile Method produces an average percentage of stations that lay within a gross over-underestimation range that is greater than 20%. Due to the unsatisfactory results produced by all of the RFFA methods, it is necessary for a new method to be developed.

7. RECOMMENDATIONS

This chapter deals with the recommendations that have arisen from this study relating to issues that have been encountered in dealing with streamflow data as well as recommendations for RFFA in South Africa in the future.

7.1 Data Issues

The recommendations regarding data issues are as follows:

- The development a method to extend the rating tables at all gauging stations to enable the estimation of discharges for all observed stages.
- The development of an index that accounts for the attenuation of streamflow, due to an upstream dam, such as the index of flood attenuation by reservoirs and lakes (FARL) as developed by Kjeldsen *et al.* (2008).
- The production of shapefiles in ArcGIS for the catchment areas published by DWS for the 1458 stations across South Africa.
- The use of the LP3 distribution in design flood practice in South Africa needs to be further investigated, in order to identify possible reasons for the inconsistent results produced at a number of stations in the country.

7.2 Recommendations for the RFFA methods assessed in this study

The advantages and disadvantages of the RFFA methods assessed in this study have already been discussed. Therefore, the following adaptations are recommended for the improvements of these methods:

- The Meigh *et al.* (1997) study involved a worldwide comparison of regional flood estimation methods and delineated two regions in South Africa based on MAP, with regression equations developed for each region for the estimation of index floods throughout South Africa. This broad classification has the potential that the two regions are not homogenous and the method can be refined by more detailed regionalisation.
- The regionalisation method developed by Mkhandi *et al.* (2000) involved the use of a proposed Homogeneity Test, which was determined to be more lenient than the

Hosking and Wallis Homogeneity Test. Therefore, the method may be improved by either refining the proposed Homogeneity Test or by using the Hosking and Wallis Homogeneity Test, which will provide more detailed regionalisation.

- The Görgens (2007) JPV Method produced inconsistent design floods when, using the LP3 distribution in comparison to the GEV distribution. Therefore, this needs to be investigated, to determine whether there is a threshold beyond which the LP3 distribution is not applicable in this method.
- The Haile (2011) Method utilised logarithmic regressions to estimate the *MEF*, which produced negative index floods below an area limit due to the negative intercept in the regressions. This has been addressed in this study, by the use of power regressions. The impact of using power regressions as opposed to the logarithmic regressions in certain cases must be investigated.

7.3 Future Recommendations for RFFA

Due to the unsatisfactory results produced by the RFFA methods, it is necessary that a new RFFA method be developed for use in South Africa. In addition, the new method should make use of more data, which could be made possible, if the recommendations of Section 7.1 are met. Through the literature that has been reviewed it is recommended that the new method be developed using an index flood method and L-moments. The following aspects should be considered in the development of a new RFFA method:

- A thorough data screening process must be employed to minimize errors and bias in method development.
- The development of homogeneous flood regions must include more updated databases than the Kovacs K-Regions (Kovacs, 1988) and HRU Veld zones (HRU, 1972) that are commonly used, as they have not been updated for many years. The homogeneous regions developed must be thoroughly tested for homogeneity, using techniques such as the Hosking and Wallis (1997) Homogeneity Test.
- The distribution selected for the new RFFA method does not necessarily need to be the same throughout the entire country. It is recommended that a Goodness-of-Fit Test be performed for several candidate distributions throughout all the homogeneous flood regions developed. This may indicate that different distributions fit data better in different regions of the country, which will improve design flood estimates.

- A multiple regression analysis should be performed to determine the relationships between the index flood and catchment parameters.
- The new RFFA method should provide different approaches to estimate the design flood for different catchment area ranges, focusing particularly on methods for estimating design floods for smaller catchments with areas that are less than or equal to 100 km^2.
- Adjustments must be made to the method in order to be applicable in highly urbanised regions as well as for climate change scenarios.
- With regards to both the RFFA Methods assessed in this study, and any possible RFFA method to be developed in the future, it is recommended that the cost of incorrect design flood estimates be investigated and taken into account the development of the methods.

8. REFERENCES

Acreman, MC and Sinclair, CD. 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *Journal of Hydrology* 84 (3–4): 365-380.

Alexander, WJR. 1990. *Flood Hydrology for Southern Africa*. SANCOLD, Pretoria.

Alexander, WJR. 2001. *Flood Risk Reduction Measures*. University of Pretoria, Pretoria, RSA.

Archer, E, Engelbrecht, F and Landman, W. 2010. *South African Risk and Vulnerability Atlas*. CSIR and DST, Pretoria, RSA.

Australian Insititution of Engineers. 1987. *Australian Rainfall and Runoff: A Guide to Flood Estimation*. EA Books, Crows Nest, Canberra, Australia.

Australian Institution of Engineers. 1977. *Australian Rainfall and Runoff: Guide to Flood Estimation*. EA Books, Crows Nest, Australia.

Bobee, B and Rasmussen, F. 1995. Recent advances in flood frequency analysis. *U.S National Report 1991-1994.Twenty-first General Assembly International Union of Geodesy and Geophysics Boulder,* Colorado, 1111-1116.

Boorman, DB, Hollis, JM and Lilly, A. 1995. *Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom*. Insititute of Hydrology, Report No.126, Insititute of Hydrology, Wallingford, UK.

Burn, DH. 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research* 26 (10): 2257-2265.

Castellarin, A, Kohnova, S, Gaal, L, Fleig, A, Salinas, J, Toumazis, A, Kjeldsen, T and Macdonald, N. 2012. *Review of applied-statistical methods for flood-frequency analysis in europe*. NERC/Centre for Ecology & Hydrology, 122pp. (ESSEM COST Action ES0901).

Chetty, K and Smithers, J. 2005. Continuous simulation modelling for design flood estimation in South Africa: Preliminary investigations in the Thukela catchment. *Physics and Chemistry of the Earth, Parts A/B/C* 30 (11–16): 634-638.

Clark, D. 2014. Personal communication. Department of Bioresources Engineering, School of Engineering, University of Kwazulu-Natal, Pietermaritzburg, RSA.

Crochet, P. 2012. *Evaluation of two delineation methods for regional flood frequency analysis in northern Iceland*. Report No. VI 2012-013, Icelandic Meteorological Office, Iceland.

Cunnane, C. 1989. *Statistical distribution for flood frequency analysis*. World Meteorlogical Organization, Operational Hydrology Report No.33, WMO-No. 718, Geneva, Switzerland.

Dalrymple, T. 1960. *Flood-frequency analyses, Manual of Hydrology: Part 3*. US Geological Survey. Water Supply Paper. 1543-A, United States Government Printing Office,Washinton, USA.

EM-DAT. 2012. The OFDA/CRED International Disaster Database Université Catholique de Louvain, Brussels (Belgium). http://www.emdat.be/.

ESRI. 1991. *Cell Based Modelling With GRID*. Environmental Systems Research Institute, Redlands, California, USA.

ESRI. 2014. *ArcGIS Desktop: Release 10.2.1*. Environmental Systems Research Institute, Redlands, California, USA.

Gabriele, S and Arnell, N. 1991. A hierarchical approach to regional flood frequency analysis. *Water Resources Research* 27 (6): 1281-1289.

Gaume, E, Gaál, L, Viglione, A, Szolgay, J, Kohnová, S and Blöschl, G. 2010. Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology* 394 (1–2): 101-117.

Gericke, OJ. 2015. Personal communication, Central University of Technology, Free State, RSA, 16 April 2015.

Gericke, OJ and du Plessis, JA. 2012. Catchment paramter analysis in flood hydrology using GIS applications. *Journal of the South African Institution of Civil Engineering* 54 (2): 15-26.

Gordon, ND, Finlayson, BL and McMahon, TA. 2004. *Stream Hydrology: An Introduction for Ecologists*. Wiley, Chichester, England.

Görgens, AHM. 2007. *Joint Peak-Volume (JPV) Design Flood Hydrographs for South Africa*. WRC, Report No. 1420/3/07. Water Research Commission, Pretoria, RSA.

GREHYS. 1996a. Inter-comparison of regional flood frequency procedures for Canadian rivers. *Journal of Hydrology* 186 (1–4): 85-103.

GREHYS. 1996b. Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology* 186 (1–4): 63-84.

Haddad, K and Rahman, A. 2012. Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework – Quantile Regression vs. Parameter Regression Technique. *Journal of Hydrology* 430–431 (0): 142-161.

Haile, AT. 2011. Regional Flood Frequency Analysis in Southern Africa. Unpublished thesis, Department of Geosciences, University of Oslo, Oslo.

Hosking, JRM and Wallis, JR. 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* 29 (2): 271-281.

Hosking, JRM and Wallis, JR. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge, UK.

Hosking, JRM. 1996. Fortran routines for use with the method of L-moments, Version 3.03, IBM Research Report RC20525, T.J. Watson Research Center, Yorktown Heights, New York, USA.

HRU. 1972. *Design flood determination in South Africa*. Reports No. 1/72, Hydrological Research Institute, University of the Witwatersrand, Johannesburg, RSA.

Hubbart, JA and Jones, JR. 2014. Floods. *Reference Module in Earth Systems and Environmental Sciences, Encyclopedia of Inland Waters* 88-91.

Ilorme, F and Griffis, VW. 2013. A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology* 492 (0): 151-162.

Institute of Hydrology. 1999. *Flood Estimation Handbook (FEH)*. Institute of Hydrology, Crowmarsh Gifford, UK.

Jingyi, Z and Hall, MJ. 2004. Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology* 296 (1–4): 98-117.

Kachroo, RK, Mkhandi, SH and Parida, BP. 2000. Flood frequency analysis of southern Africa: I. Delineation of homogeneous regions. *Hydrological Sciences Journal* 45 (3): 437-447.

Kizza, M, Ntale, HK, Rugumayo, AI and Kigobe, M. 2006. Regional Flood Frequency Analysis for Northern Uganda Using the L-Moment Approach. In: ed. Taban-Wani, JAM, *Proceedings from the International Conference on Advances in Engineering and Technology*. Elsevier Science Ltd, Oxford, UK.

Kjeldsen, TR. 2010. Modelling the impact of urbanization on flood frequency relationships in the UK. *Hydrology Research* 41 (5): 391-405.

Kjeldsen, TR. 2015. Personal communication, Department of Architecture and Civil Engineering, University of Bath, England, 12 March 2015.

Kjeldsen, TR and Jones, DA. 2009. An exploratory analysis of error components in hydrological regression modeling. *Water Resources Research* 45 (2): 1-13.

Kjeldsen, TR, Jones, DA and Bayliss, AC. 2008. *Improving the FEH statistical procedures for flood frequency estimation*. Final Research Report to the Environmental Agency, R&D Project SC050050, CEH Wallingford, UK.

Kjeldsen, TR, Smithers, JC and Schulze, RE. 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology* 255 (1–4): 194-211.

Kjeldsen, TR, Smithers, JC, Schulze, RE, Kjeldsen, TR, Smithers, JC and Schulze, RE. 2001. Flood frequency analysis at ungauged sites in the KwaZulu-Natal Province, South Africa. *Water S.A.* 27 (3): 315-324.

Kovacs, Z. 1988. *Regional maximum flood peaks in southern Africa*. Report No. TR 137. Department of Water Affairs, Pretoria, RSA.

Laio, F, Ganora, D, Claps, P and Galeati, G. 2011. Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *Journal of Hydrology* 408 (1–2): 67-77.

Leclerc, M and Ouarda, TBMJ. 2007. Non-stationary regional flood frequency analysis at ungauged sites. *Journal of Hydrology* 343 (3–4): 254-265.

Lynch, SD. 2004. *Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa*. WRC Report 1156/1/04. Water Reasearch Commision, Pretoria, RSA. pp 78.

Malekinezhad, H, Nachtnebel, HP and Klik, A. 2011. Comparing the index-flood and multiple-regression methods using L-moments. *Physics and Chemistry of the Earth, Parts A/B/C* 36 (1–4): 54-60.

Mediero, L and Kjeldsen, TR. 2014. Regional flood hydrology in a semi-arid catchment using a GLS regression model. *Journal of Hydrology* 514 158-171.

Meigh, JR, Farquharson, FAK and Sutcliffe, JV. 1997. A worldwide comparison of regional flood estimation and climate. *Hydrological Sciences Journal* 42 (2): 255-244.

Merz, R and Blöschl, G. 2005. Flood frequency regionalisation—spatial proximity vs. catchment attributes. *Journal of Hydrology* 302 (1–4): 283-306.

Midgley, DC, Middleton, B, Pitman, WV and Robertson, RM. 1981. *Surface water resources of South Africa*. Report No. 10/81. Hydrological Research Unit, University of the Witwatersrand, Johannesburg, RSA.

Midgley, DC, Pitman, WV and Middleton, B. 1994. *Surface water resources of South Africa-1990*. Volumes I-VI. WRC Report No. 298/1/94. Water Research Commission, Pretoria, RSA.

Mkhandi, SH, Kachroo, RK and Gunasekara, TAG. 2000. Flood frequency analysis of southern Africa: II. Identification of regional distributions. *Hydrological Sciences Journal* 45 (3): 449-464.

Mukheibir, P and Sparks, D. 2005. *Climate variability, climate change and water resources strategies for small municipalities*. WRC Report No. K5/1500. Water Research Commision, Pretoria, RSA.

NERC. 1975. *Flood Studies Report, Natural Environment Research Council*. London, UK.

Nguyen, CC, Gaume, E and Payrastre, O. 2014. Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology* 508 (0): 385-396.

Nobert, J, Mugo, M and Gadain, H. 2014. Estimation of design floods in ungauged catchments using a regional index flood method. A case study of Lake Victoria Basin in Kenya. *Physics and Chemistry of the Earth, Parts A/B/C* (0): 1-8.

Pegram, G and Parak, M. 2004. A review of the regional maximum flood and rational formula using geomorphological information and observed floods. *Water SA* 30 (3): 377-384.

Pitman, WV. 2011. Overview of water resource assessment in South Africa: Current state and future challenges. *Water SA* 37 (5): 659-664.

Rahman, A, Haddad, K, Zaman, M, Kuczera, G, Weinmann, P, Stensmyr, P and Babister, M.2013. New Regional Flood Frequency Estimation (RFFE) method for the whole of Australia: Overview of Progress. *Proceedings of the 53rd Floodplain Management Association Conference*, FMA, Tweed Heads, Australia.

Reis, DS and Stedinger, JR. 2005. Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology* 313 (1–2): 97-116.

Saf, B. 2008. Application of Index Procedures to Flood Frequency Analysis in Turkey. *JAWRA Journal of the American Water Resources Association* 44 (1): 37-47.

Saf, B. 2009. Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey. *Water Resources Management* 23 (3): 531-551.

Saf, B. 2010. Assessment of the effects of discordant sites on regional flood frequency analysis. *Journal of Hydrology* 380 (3–4): 362-375.

Schulze, R.E. 1989. Non-stationary catchment responses and other problems in determining flood series: A case for a simulation modelling approach. In: Kienzle SW and Maaren H (eds.) *Proceedings of the Fourth South African National Hydrological Symposium*. SANCIAHS, Pretoria, South Africa. 135–157.

Schulze, RE. 1997. *South African Atlas of Agrohydrology and Climatology*. WRC, Report No. TT82/96. Water Research Commission, Pretoria, RSA.

Smithers, JC. 2012. Methods for design flood estimation in South Africa. *Water SA* 38 (4): 633-646.

Smithers, JC and Schulze, RE. 2000. *Development and evaluation of techniques for estimating short duration design rainfall in South Africa*. WRC, Report No. 681/1/100. Water Research Commission, Pretoria, RSA.

Smithers, JC. 2014. Personal Communication. Department of Bioresource Engineering, School of Engineering, University of Kwazulu-Natal, Pietermaritzburg, RSA.

Smithers, JC, Streatfield, J, Gray, RP and Oakes, EGM. 2015. Performance of regional flood frequency analysis methods in South Africa. *Water SA* 41 (3): 390-397.

Tibane, E and Vermeulen, A. 2014. *South Africa Yearbook 2013/14*. Government Communications (GCIS), Pretoria, RSA.

van Bladeren, D. 1992. *Historical Flood Documentation Series. No.1: Natal and Transkei, 1848 - 1989*. Technical Report TR147. Department of Water Affairs and Forestry, Pretoria,RSA.

van Bladeren, D.1993. Application of historical flood data in flood frequency anlysis, for the Natal and Transkei region. *Proceedings of the Sixth South African National Hydrological Symposium*, Department of Agricultural Engineering, University of Natal, Pietermaritzburg, RSA, 359-366.

van der Spuy, D and Rademeyer, PF. 2010. *Flood Frequency Estimation Methods as applied in the Department of water Affairs*. Pretoria, RSA.

Viglione, A, Laio, F and Claps, P. 2007. A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research* 43 (3): 1-10.

Wazneh, H, Chebana, F and Ouarda, TBMJ. 2015. Delineation of homogeneous regions for regional frequency analysis using statistical depth function. *Journal of Hydrology* 521 (0): 232-244.

Weepener, HL, van de Berg, HM, Metz, M and Hamandawana, H. 2011. *The development of a hydrologically improved Digital Elevation Model and derived products for South*

Africa based on the SRTM DEM. WRC, Report No. K5/1908. Water Research Commission, Pretoria, South Africa.

Whitmore, JS. 1971. South Africa's water budget. *South African Journal of Science* 67 (3): 166-176.

Wiltshire, SE. 1986. Regional flood frequency analysis I: Homogeneity statistics. *Hydrological Sciences Journal* 31 (3): 321-333.

APPENDIX A: COMPLETE INVENTORY OF STATIONS ANALYSED

Appendix A provides a complete inventory of the stations analysed in this study. The station name, geographical coordinates and record lengths for each station is provided, as well as all of the catchment parameters that were obtained for the application of each RFFA method.

Table A1 Complete inventory of the 410 stations analysed in this study

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
A2H006	-25.38	28.32	1905-03-01	2013-12-10	110	1051.28	0.0036	32.52	C	K-MID
A2H007	-25.73	28.17	1908-07-01	1951-08-01	44	145.45	0.0096	114.82	C	K-MID
A2H012	-25.81	27.91	1922-10-01	2013-12-11	92	2579.65	0.0049	40.81	B	K-MID
A2H013	-25.78	27.76	1922-10-01	2013-12-11	93	1164.43	0.0049	29.63	B	K-MID
A2H023	-25.95	27.96	1965-10-23	2013-12-09	50	689.85	0.0062	48.82	B	K-MID
A2H024	-26.15	27.59	1965-12-19	2013-12-17	49	16.18	0.0169	21.23	B	K-LOW
A2H027	-25.66	28.35	1962-05-22	2013-12-09	53	377.07	0.0051	42.00	B	K-MID
A2H029	-25.65	28.39	1962-05-21	2013-12-09	53	122.81	0.0087	42.00	C	K-MID
A2H032	-25.64	27.03	1963-09-05	2013-11-26	51	516.10	0.0075	16.08	B	K-MID
A2H038	-25.73	27.21	1970-12-23	2013-11-26	44	26.54	0.0180	23.61	B	K-MID
A2H039	-25.72	27.19	1971-05-18	2013-11-19	43	12.37	0.0454	23.71	B	K-MID
A2H040	-26.03	28.11	1971-07-02	1993-03-15	23	192.86	0.0100	48.82	B	K-MID
A2H042	-26.01	28.03	1971-07-01	1995-11-20	26	416.35	0.0078	48.82	B	K-MID
A2H044	-25.90	27.93	1971-07-18	2013-12-11	43	764.04	0.0055	48.82	B	K-MID
A2H045	-25.89	27.91	1972-05-25	2013-12-11	42	663.72	0.0067	55.45	B	K-MID
A2H047	-26.07	27.97	1971-07-21	2013-12-09	43	66.39	0.0136	48.82	B	K-MID
A2H049	-25.98	27.84	1972-07-04	2013-12-09	43	372.83	0.0077	56.33	B	K-MID
A2H050	-25.99	27.84	1973-04-06	2013-12-09	41	152.63	0.0106	54.32	B	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
A2H053	-25.81	27.48	1973-07-13	2013-11-26	41	89.00	0.0138	36.80	B	K-MID
A2H054	-25.68	28.29	1982-09-02	2013-12-09	33	36.77	0.0096	42.00	C	K-MID
A2H056	-25.73	28.18	1982-09-02	2013-12-10	33	18.29	0.0108	114.82	B	K-MID
A2H058	-25.75	27.91	1982-09-02	2013-12-10	33	107.00	0.0060	36.26	C	K-MID
A2H061	-25.47	28.26	1984-04-13	2013-12-10	31	638.04	0.0045	48.59	C	K-MID
A2H063	-25.70	28.19	1984-05-10	2013-12-10	30	33.28	0.0079	29.03	C	K-MID
A2H077	-25.38	28.32	1905-12-15	1950-03-04	33	1017.79	0.0040	36.32	C	K-MID
A4H005	-24.08	27.77	1962-08-27	2012-09-08	51	3828.28	0.0030	58.02	C	K-MID
A4H007	-23.76	27.91	1962-09-25	2013-08-21	52	399.94	0.0130	26.00	C	K-MID
A5H004	-23.98	28.40	1955-12-01	2013-08-22	59	638.81	0.0055	81.06	C	K-MID
A6H010	-24.57	28.64	1964-08-28	2013-07-09	51	72.45	0.0162	42.73	C	K-MID
A6H011	-24.76	28.34	1966-11-24	2013-08-22	48	73.66	0.0157	48.80	C	K-LOW
A6H012	-24.67	28.48	1966-11-12	2013-08-23	49	117.70	0.0141	69.74	C	K-MID
A6H018	-24.77	28.35	1973-07-27	2013-08-22	42	15.86	0.0203	50.44	C	K-LOW
A6H020	-24.67	28.56	1973-08-10	2013-07-08	42	40.92	0.0172	69.74	C	K-MID
A6H021	-24.63	28.60	1973-06-26	2013-07-09	42	15.90	0.0385	42.73	C	K-MID
A6H022	-24.60	28.61	1973-06-26	1997-10-01	26	1.60	0.0527	42.73	C	K-MID
A6H024	-24.32	28.92	1973-08-24	2013-08-23	41	18.21	0.0138	40.33	C	K-MID
A7H001	-22.91	29.61	1957-07-12	2000-01-27	32	7773.38	0.0035	4.80	C	K-MID
A7H003	-23.07	29.58	1947-10-01	1995-11-08	49	4283.46	0.0032	4.48	C	K-MID
A9H004	-22.77	30.54	1932-07-26	2004-06-22	73	332.40	0.0056	356.50	C	K-HIGH
A9H006	-23.04	30.28	1961-11-13	2013-09-11	53	15.02	0.0912	372.39	C	K-HIGH
A9H012	-22.77	30.89	1987-11-04	2013-09-10	27	2272.52	0.0025	147.34	C	K-HIGH
B1H002	-25.82	29.34	1956-10-09	2013-11-21	55	247.79	0.0055	35.92	B	K-LOW
B1H004	-25.67	29.17	1959-02-08	2013-11-20	56	380.83	0.0059	45.69	B	K-LOW
B1H005	-26.01	29.25	1972-07-13	2013-03-05	42	3234.82	0.0012	34.53	B	K-LOW
B1H012	-25.81	29.59	1978-01-30	2013-11-21	37	1501.96	0.0019	27.86	B	K-LOW

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
B1H017	-26.31	29.27	1989-11-21	2013-09-19	26	388.17	0.0024	32.98	B	K-LOW
B1H018	-26.22	29.46	1989-11-22	2013-11-21	26	952.70	0.0018	38.63	B	K-LOW
B1H019	-25.94	29.26	1990-03-05	2013-11-20	25	88.79	0.0041	35.64	B	K-LOW
B2H007	-26.00	28.66	1985-08-26	2013-11-22	30	323.20	0.0058	36.51	B	K-LOW
B3H007	-25.27	29.18	1980-03-13	2013-11-19	35	973.58	0.0063	25.58	C	K-LOW
B4H005	-25.04	30.22	1960-09-03	2013-12-09	54	190.74	0.0129	100.06	B	K-LOW
B4H007	-25.01	30.50	1968-08-05	2013-12-09	46	156.28	0.0177	221.40	C	K-LOW
B5H002	-24.27	29.80	1948-09-01	1980-01-31	32	31690.89	0.0009	23.45	C	K-MID
B6H001	-24.68	30.80	1909-11-11	2013-12-13	105	517.60	0.0050	382.73	C	K-MID
B6H002	-24.68	30.81	1909-12-19	1939-03-01	30	95.24	0.0165	532.34	C	K-MID
B6H003	-24.69	30.81	1959-08-31	2013-12-13	55	95.24	0.0166	532.34	C	K-MID
B6H006	-24.93	30.55	1968-07-31	2007-09-07	40	42.44	0.0359	45.33	C	K-LOW
B7H003	-24.12	30.36	1948-09-21	1972-11-23	23	82.03	0.0247	165.79	C	K-HIGH
B7H004	-24.56	31.03	1950-11-01	2013-12-10	64	135.12	0.0170	211.41	C	K-HIGH
B7H008	-24.01	30.67	1956-04-24	1999-01-01	44	843.74	0.0036	71.70	C	K-HIGH
B7H010	-24.04	30.43	1960-07-27	2013-12-12	54	323.38	0.0097	96.79	C	K-HIGH
B7H014	-24.12	30.36	1973-08-01	2013-12-12	41	82.03	0.0247	165.79	C	K-HIGH
B7H019	-24.04	31.13	1988-10-21	2013-12-11	26	2365.13	0.0030	33.74	C	K-MID
B7H020	-24.23	31.63	1988-12-01	2013-04-10	21	950.44	0.0022	16.62	C	K-MID
B8H010	-23.89	30.36	1960-01-13	2014-01-23	55	483.54	0.0067	139.18	C	K-HIGH
B8H011	-23.53	31.40	1960-12-09	2013-08-27	54	444.88	0.0031	19.33	C	K-MID
B8H014	-23.88	30.08	1968-05-03	2014-01-20	47	295.46	0.0160	351.38	C	K-HIGH
B8H017	-23.65	30.72	1977-03-15	2013-10-10	38	2618.47	0.0025	129.95	C	K-HIGH
B8H018	-23.84	31.64	1984-02-14	2013-08-27	30	13547.37	0.0015	42.05	C	K-MID
B8H019	-23.53	31.40	1984-01-04	2013-08-27	30	444.88	0.0030	19.33	C	K-MID
B8H034	-23.70	31.21	1988-09-08	2013-07-09	26	10723.42	0.0019	49.55	C	K-MID
B9H002	-23.22	31.22	1983-11-15	2013-08-28	31	828.03	0.0019	23.55	C	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
B9H003	-23.14	31.46	1984-02-01	2013-02-19	30	4651.95	0.0019	15.83	C	K-MID
B9H004	-22.95	31.23	1983-11-15	2013-02-13	31	763.05	0.0021	12.32	C	K-MID
C1H002	-27.17	29.23	1906-10-01	2013-10-08	108	4183.03	0.0017	60.60	B	K-LOW
C1H004	-26.63	29.02	1960-08-12	2013-10-19	54	904.04	0.0012	58.96	B	K-LOW
C1H006	-26.78	29.54	1964-12-11	2013-12-06	50	1111.06	0.0012	72.72	B	K-LOW
C1H007	-26.84	29.72	1972-10-22	2013-10-08	42	4762.74	0.0005	67.48	B	K-LOW
C1H008	-26.86	28.88	1973-12-12	2013-10-29	41	2243.59	0.0008	53.89	B	K-LOW
C1H012	-27.00	28.77	1985-09-23	2013-12-12	29	15696.73	0.0004	57.66	B	K-LOW
C1H015	-27.17	29.24	1906-11-13	2013-10-29	108	4183.03	0.0013	60.60	B	K-LOW
C1H027	-26.78	29.81	1994-11-15	2013-10-29	20	1372.16	0.0007	55.17	B	K-LOW
C2H018	-26.97	27.21	1938-10-03	2013-10-10	76	43250.04	0.0010	44.57	B	K-LOW
C2H024	-26.28	27.68	1957-10-04	1996-05-15	31	181.22	0.0050	29.31	B	K-LOW
C2H027	-26.23	27.65	1957-10-10	1992-12-16	36	5.54	0.0234	29.31	B	K-LOW
C2H070	-26.64	28.23	1977-06-13	1996-02-12	20	2711.06	0.0015	42.62	B	K-LOW
C2H073	-26.98	26.63	1986-08-18	2008-02-25	23	4777.71	0.0011	15.76	B	K-LOW
C2H141	-26.45	28.09	1977-10-12	2010-12-17	27	1290.16	0.0024	31.78	B	K-LOW
C3H004	-27.56	24.71	1923-11-01	1947-03-31	24	10192.77	0.0013	4.46	B	K-LOW
C4H002	-27.84	25.90	1935-12-13	1972-05-31	23	17711.47	0.0003	31.38	B	K-LOW
C4H004	-27.94	26.12	1968-09-05	2013-09-05	47	16798.08	0.0004	32.90	B	K-LOW
C5H015	-28.81	26.11	1949-01-01	1983-11-22	35	6084.42	0.0011	28.00	B	K-LOW
C5H018	-29.04	24.64	1960-01-11	1999-03-15	40	17376.56	0.0007	10.68	B	K-LOW
C5H022	-29.29	26.92	1980-10-14	2013-10-24	34	37.74	0.0129	38.80	B	K-MID
C5H023	-29.29	26.76	1983-06-04	2008-09-23	26	188.91	0.0049	38.80	B	K-MID
C6H003	-27.40	26.63	1967-03-22	2010-02-17	44	7728.93	0.0007	21.82	B	K-LOW
C7H005	-27.12	27.11	1954-03-15	1995-10-09	43	5499.28	0.0011	31.42	B	K-LOW
C7H006	-27.05	27.00	1978-04-18	2013-11-26	37	5781.47	0.0007	31.36	B	K-LOW
C8H003	-27.85	28.96	1954-01-13	2013-06-28	61	859.95	0.0039	53.31	B	K-LOW

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
C8H004	-27.70	28.32	1957-03-01	1996-01-24	33	3541.31	0.0011	33.50	B	K-LOW
C8H005	-28.38	28.86	1963-12-12	2013-07-12	51	691.84	0.0061	129.80	B	K-MID
C8H011	-28.16	28.87	1972-08-02	1997-12-10	27	1488.49	0.0028	87.27	B	K-LOW
C8H014	-27.81	28.78	1973-10-31	2002-07-16	29	7487.00	0.0005	68.88	B	K-LOW
C8H020	-27.69	28.38	1974-10-14	2013-06-13	41	3605.23	0.0010	33.51	B	K-LOW
C8H022	-27.30	28.50	1961-12-11	2008-07-02	47	15766.39	0.0005	52.32	B	K-LOW
C8H026	-27.43	28.53	1985-03-21	2013-06-13	30	4657.59	0.0010	34.53	B	K-LOW
C8H027	-27.30	28.59	1985-06-06	2013-06-28	30	10533.13	0.0005	60.92	B	K-LOW
C8H028	-27.80	28.77	1988-12-02	2013-06-12	27	7498.57	0.0005	68.84	B	K-LOW
C9H003	-28.51	24.70	1909-01-01	2012-12-31	100	123039.19	0.0011	30.71	B	K-LOW
C9H009	-28.52	24.60	1968-08-13	2013-01-15	46	121230.66	0.0008	31.00	B	K-LOW
C9H010	-28.41	24.27	1974-10-30	2013-07-31	40	157999.39	0.0017	24.93	B	K-LOW
D1H001	-31.00	26.35	1912-10-01	2013-09-19	103	2391.33	0.0037	17.15	B	K-MID
D1H004	-31.40	26.37	1925-02-13	1981-07-01	57	324.34	0.0044	20.97	B	K-MID
D1H011	-30.83	26.92	1965-10-06	2013-10-29	50	8697.52	0.0019	74.89	B	K-MID
D1H032	-29.55	28.15	1985-11-23	2013-11-13	29	1084.13	0.0075	257.63	B	K-MID
D1H033	-29.48	28.64	1985-11-21	2012-04-25	28	3204.81	0.0047	231.66	B	K-MID
D2H034	-28.88	27.84	1991-10-17	2013-02-14	24	1096.39	0.0019	63.12	B	K-MID
D4H002	-26.09	25.28	1905-10-01	1964-10-01	39	518.99	0.0035	0.30	B	K-LOW
D4H032	-26.09	25.28	1927-01-01	1964-10-01	39	603.55	0.0035	0.30	B	K-LOW
D5H003	-31.81	20.36	1927-10-01	2013-11-05	87	1507.08	0.0035	18.71	B	K-LOW
D5H011	-31.82	20.58	1958-06-01	2013-11-05	56	1674.13	0.0040	15.77	B	K-LOW
D5H013	-31.37	21.32	1958-06-01	1998-03-31	37	13108.09	0.0020	4.80	B	K-LOW
D5H016	-30.47	20.52	1973-03-07	2013-05-09	38	40042.87	0.0008	5.93	B	K-LOW
D7H002	-29.65	22.75	1959-05-01	2013-11-12	56	341299.94	0.0007	32.20	B	K-LOW
D7H005	-28.46	21.24	1936-10-01	2013-11-15	79	370061.83	0.0007	29.95	B	K-LOW
D7H008	-29.03	22.19	1932-10-01	2013-11-13	83	351032.92	0.0007	31.49	B	K-LOW

Gauge	Latitude (⁰)	longitude (⁰)	Start date	End date	Record length (years)	Area (km²)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
E1H006	-32.21	18.94	1971-03-05	2013-11-05	43	163.03	0.0221	205.48	A	K-LOW
E1H013	-32.60	19.01	1992-06-17	2013-11-05	22	930.14	0.0055	272.23	A	K-LOW
E2H003	-31.86	18.69	1908-05-17	2013-11-05	106	24003.52	0.0031	21.02	B	K-LOW
E2H007	-32.78	19.28	1930-04-01	2013-11-12	84	267.35	0.0024	114.46	A	K-LOW
E2H010	-33.12	19.39	1982-10-25	2013-11-12	32	82.78	0.0046	182.89	A	K-LOW
G1H008	-33.31	19.07	1954-05-01	2013-11-07	60	396.07	0.0144	172.44	A	K-MID
G1H010	-33.39	19.16	1964-05-05	2013-11-07	50	10.40	0.0228	172.44	A	K-MID
G1H011	-33.38	19.15	1964-04-29	2013-11-07	50	18.29	0.0181	172.44	A	K-MID
G1H012	-33.35	19.10	1964-04-20	1996-06-04	33	34.43	0.0398	171.36	A	K-MID
G1H015	-33.82	19.06	1964-06-06	1988-07-18	25	1.80	0.2130	722.69	A	K-MID
G1H016	-33.82	19.06	1964-06-06	2013-08-27	50	3.74	0.0939	722.69	A	K-MID
G1H017	-33.83	19.03	1964-06-06	1988-07-19	25	1.76	0.1848	722.69	A	K-MID
G1H018	-33.82	19.05	1964-06-06	2013-08-27	50	3.49	0.1681	722.69	A	K-MID
G1H028	-33.13	19.06	1972-05-06	2013-11-07	42	186.53	0.0405	666.38	A	K-MID
G1H029	-33.16	19.05	1972-11-30	2013-11-07	42	35.66	0.0728	40.00	A	K-MID
G1H038	-33.94	19.03	1978-09-15	2013-08-28	36	14.12	0.1847	1012.75	A	K-MID
G1H040	-33.36	18.96	1979-08-16	2013-11-06	35	37.62	0.0108	112.12	A	K-MID
G2H008	-33.99	18.96	1947-06-01	1995-04-07	49	23.36	0.0539	867.66	A	K-MID
G4H008	-34.15	19.14	1964-04-11	1992-05-05	29	1.30	0.2162	134.45	A	K-MID
G4H009	-34.17	19.13	1964-04-11	1992-04-28	29	2.11	0.1688	134.45	A	K-MID
G4H010	-34.17	19.13	1964-04-11	1992-05-05	29	6.65	0.0930	134.45	A	K-MID
G4H012	-34.15	19.14	1965-03-19	1992-05-05	28	0.69	0.2737	134.45	A	K-MID
G4H013	-34.16	19.13	1965-03-12	1992-05-05	28	2.25	0.1631	134.45	A	K-MID
G4H014	-34.24	19.22	1967-04-13	2013-08-28	47	248.64	0.0102	134.45	A	K-MID
G4H033	-34.36	19.25	1977-04-28	2013-08-28	37	24.57	0.0254	121.37	A	K-MID
H1H013	-33.36	19.30	1965-02-24	2013-11-04	49	62.63	0.0326	265.32	A	K-MID
H1H016	-33.42	19.48	1966-05-04	1991-04-10	22	10.52	0.1503	285.56	A	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
H1H018	-33.72	19.17	1969-02-26	2013-11-08	45	110.35	0.0305	855.56	A	K-MID
H2H005	-33.46	19.62	1969-09-26	2013-11-12	45	14.81	0.1523	96.54	A	K-MID
H2H008	-33.33	19.64	1982-06-29	2013-11-04	32	9.69	0.1102	44.50	A	K-MID
H3H001	-33.79	20.12	1925-11-01	1948-05-01	23	600.97	0.0053	45.43	A	K-MID
H3H004	-33.70	19.93	1965-03-20	1992-09-08	27	13.92	0.0902	37.64	A	K-MID
H4H005	-33.76	19.85	1950-04-01	1981-12-21	33	18.03	0.1192	51.88	A	K-MID
H4H007	-33.64	19.81	1965-03-30	1992-09-08	28	48.25	0.0173	15.32	A	K-MID
H4H009	-34.01	19.84	1967-04-26	1992-09-07	26	20.18	0.0754	46.03	A	K-MID
H4H012	-33.95	19.59	1969-02-28	1992-05-25	24	14.48	0.1272	66.18	A	K-MID
H4H013	-33.86	19.41	1970-03-06	1991-06-17	22	102.92	0.0156	125.56	A	K-MID
H4H015	-33.99	19.82	1978-05-10	2010-12-02	34	25.04	0.0423	46.03	A	K-MID
H6H007	-33.94	19.17	1964-03-14	1992-09-07	29	34.52	0.1039	561.03	A	K-MID
H6H009	-34.08	20.14	1964-05-09	2013-08-29	50	2019.55	0.0020	222.77	A	K-MID
H6H010	-33.98	19.33	1969-02-17	2013-08-28	45	16.76	0.1136	384.56	A	K-MID
H7H004	-33.91	20.71	1951-05-02	2013-10-16	63	25.60	0.0336	48.60	A	K-MID
H7H005	-33.99	20.42	1960-01-25	2013-08-27	54	8.70	0.2043	268.40	A	K-MID
H8H001	-34.25	20.99	1967-04-21	2013-10-17	47	789.07	0.0055	110.13	A	K-MID
H9H002	-34.01	21.20	1963-04-16	2013-06-12	51	88.59	0.0344	242.46	A	K-MID
H9H005	-34.09	21.29	1969-04-09	2013-10-08	45	202.84	0.0119	206.64	A	K-MID
J1H004	-33.20	20.85	1920-10-01	1955-10-01	37	3087.75	0.0053	8.29	B	K-MID
J1H015	-33.35	19.72	1974-07-05	2013-11-04	40	9.54	0.1816	37.46	A	K-MID
J1H016	-33.29	19.73	1974-06-24	2013-11-04	40	30.88	0.0378	37.46	A	K-MID
J2H005	-33.49	21.49	1955-02-01	2013-09-24	59	290.92	0.0281	34.22	B	K-MID
J2H007	-33.49	21.51	1955-01-31	2013-09-25	59	37.01	0.0597	34.22	B	K-MID
J3H005	-33.78	22.32	1926-03-01	1947-09-30	22	104.47	0.0084	14.42	A	K-MID
J3H012	-33.48	22.55	1964-05-04	1994-07-05	30	687.32	0.0155	21.82	B	K-MID
J3H014	-33.42	22.24	1966-10-19	2013-08-06	48	150.79	0.0257	53.06	B	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
J3H015	-33.43	22.25	1966-04-06	2013-08-06	48	61.58	0.0737	53.06	B	K-MID
J3H020	-33.46	21.96	1974-08-23	2013-06-06	40	35.76	0.0365	49.84	B	K-MID
J4H002	-33.98	21.65	1964-05-01	2013-10-08	50	43542.96	0.0028	13.34	B	K-MID
J4H004	-33.99	21.78	1967-03-30	1996-11-20	31	99.22	0.0201	89.52	A	K-MID
K1H002	-33.94	22.13	1958-07-02	2013-10-02	56	3.83	0.0603	237.43	A	K-MID
K1H018	-33.94	22.13	1963-06-18	2013-10-02	51	3.58	0.0671	237.43	A	K-MID
K3H002	-33.94	22.46	1961-04-01	2013-08-22	53	1.18	0.1767	301.33	C	K-MID
K3H004	-33.95	22.42	1961-04-12	2013-08-22	53	34.04	0.0384	301.33	C	K-MID
K4H001	-33.98	22.80	1959-11-19	1993-05-17	34	111.61	0.0263	240.12	C	K-HIGH
K4H003	-33.91	22.71	1961-05-13	2013-08-20	53	71.15	0.0220	212.52	C	K-MID
K6H001	-33.80	23.14	1961-08-19	2013-08-13	53	161.48	0.0154	86.08	C	K-MID
K8H001	-33.98	24.02	1961-06-20	2013-11-27	54	25.40	0.0373	419.49	C	K-HIGH
K8H002	-33.98	24.05	1961-07-11	2013-11-27	54	35.18	0.0356	419.49	C	K-HIGH
K8H005	-34.10	24.44	1995-06-20	2013-11-26	20	138.48	0.0102	203.89	A	K-HIGH
L1H001	-32.24	23.05	1917-07-01	1977-09-01	32	3937.70	0.0036	6.65	B	K-MID
L2H003	-31.96	23.78	1954-04-01	1993-04-04	40	1155.51	0.0072	17.28	B	K-MID
L6H001	-33.20	24.23	1926-10-01	2013-11-26	89	1294.41	0.0037	5.36	B	K-MID
L8H001	-33.87	23.84	1965-04-03	2013-11-27	50	21.07	0.0678	53.31	C	K-HIGH
L8H002	-33.74	23.30	1970-07-09	2013-08-13	44	51.86	0.0307	55.72	C	K-MID
L8H005	-33.79	24.03	1990-04-06	2013-11-27	25	1626.82	0.0057	65.47	A	K-HIGH
N2H002	-32.95	24.67	1923-11-01	1992-12-07	70	11395.68	0.0032	12.79	B	K-MID
N2H005	-33.08	25.02	1928-09-01	1947-09-30	20	14114.37	0.0023	12.14	B	K-MID
N2H008	-33.08	25.08	1979-06-20	2013-11-26	36	344.13	0.0081	16.27	B	K-MID
N3H001	-32.98	25.19	1928-09-01	1948-07-31	20	1585.96	0.0050	19.23	B	K-MID
P3H001	-33.55	26.60	1969-07-04	2013-12-04	46	579.19	0.0062	32.11	C	K-HIGH
P4H001	-33.51	26.74	1969-07-09	2013-12-04	46	576.81	0.0068	37.79	C	K-HIGH
Q1H012	-31.57	25.54	1977-07-30	2014-01-15	38	1551.94	0.0048	16.53	B	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
Q3H005	-32.09	25.58	1977-04-21	2013-11-22	38	10834.65	0.0026	12.08	B	K-MID
Q4H003	-31.97	26.00	1964-12-11	1992-12-07	29	1263.40	0.0039	11.56	B	K-MID
Q6H003	-32.61	25.88	1980-09-08	2013-12-02	35	813.38	0.0088	24.77	B	K-HIGH
Q8H004	-32.56	25.45	1957-03-19	1987-02-12	31	806.45	0.0092	18.60	B	K-MID
Q8H008	-32.79	25.61	1979-08-07	2013-12-03	36	1505.79	0.0046	22.05	B	K-MID
Q8H010	-32.56	25.45	1987-02-12	2013-12-04	28	806.45	0.0092	18.60	B	K-MID
Q9H002	-32.71	26.30	1928-10-01	2013-12-05	87	1250.60	0.0095	41.34	B	K-HIGH
Q9H008	-32.71	26.58	1921-12-01	1971-09-02	49	754.76	0.0072	72.08	B	K-HIGH
Q9H014	-32.46	26.51	1964-01-30	1986-07-02	23	250.74	0.0196	66.87	B	K-HIGH
Q9H029	-32.76	26.63	1991-10-08	2013-12-06	24	1718.07	0.0045	42.02	B	K-HIGH
Q9H030	-32.47	26.51	1982-01-04	2013-12-05	33	251.00	0.0168	66.87	B	K-HIGH
R1H013	-33.01	26.95	1950-01-01	1986-05-27	33	1526.81	0.0036	64.64	B	K-HIGH
R2H005	-32.88	27.38	1947-10-01	2013-12-02	68	416.17	0.0068	111.49	B	K-HIGH
R2H012	-32.79	27.26	1959-11-07	1997-10-13	38	13.48	0.0257	94.80	B	K-HIGH
R2H015	-32.93	27.47	1988-03-21	2013-12-02	27	202.63	0.0079	79.91	B	K-HIGH
S3H003	-32.20	26.48	1963-03-29	1995-08-17	32	246.18	0.0145	25.25	B	K-MID
S3H004	-32.05	26.79	1964-04-17	2014-01-13	51	1411.33	0.0041	17.01	B	K-MID
S3H006	-31.92	26.79	1964-05-05	2013-11-19	51	2189.50	0.0039	19.14	B	K-MID
S6H001	-32.58	27.37	1947-04-12	2013-12-04	68	91.39	0.0186	156.55	B	K-HIGH
T1H001	-31.67	28.11	1947-06-24	2013-11-25	68	999.86	0.0063	123.34	B	K-HIGH
T1H004	-31.92	28.45	1956-06-04	2007-04-04	27	4940.35	0.0038	132.40	B	K-HIGH
T3H005	-31.03	28.88	1951-09-20	2014-01-17	64	2576.73	0.0050	188.84	B	K-HIGH
T3H006	-31.24	28.85	1951-10-16	2014-01-17	64	4300.87	0.0034	208.82	B	K-HIGH
T3H007	-30.86	29.07	1984-09-20	2013-11-21	31	6938.87	0.0047	108.12	B	K-MID
T3H009	-31.07	28.35	1964-08-15	2013-11-29	51	307.02	0.0035	288.58	B	K-MID
T4H001	-30.73	29.83	1951-09-05	2013-11-26	63	728.22	0.0085	220.68	B	K-HIGH
T5H001	-30.26	29.94	1931-07-19	1979-05-07	46	3664.42	0.0053	261.30	B	K-HIGH

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
T5H003	-29.75	29.54	1949-06-20	2009-11-02	62	142.29	0.0096	435.73	B	K-MID
T5H004	-29.78	29.47	1949-07-01	2013-11-25	65	540.83	0.0071	431.00	B	K-MID
T5H005	-29.99	29.85	1949-07-07	2013-11-26	65	98.17	0.0236	198.04	B	K-HIGH
T5H007	-30.25	29.93	1956-10-11	2013-10-29	58	3664.42	0.0065	261.30	B	K-HIGH
T5H012	-30.72	30.16	1970-09-08	2013-10-29	44	428.53	0.0055	109.44	C	K-HIGH
U1H005	-29.74	29.90	1960-08-14	2013-11-07	54	1753.97	0.0078	376.40	B	K-MID
U2H002	-29.65	30.80	1928-03-04	1975-05-21	36	4082.60	0.0069	169.85	C	K-HIGH
U2H006	-29.38	30.28	1954-01-04	2013-10-29	60	341.26	0.0045	229.74	B	K-HIGH
U2H007	-29.44	30.15	1954-07-16	2013-09-23	60	355.90	0.0096	199.49	B	K-MID
U2H011	-29.65	30.26	1957-12-24	2013-10-29	57	221.59	0.0116	196.31	C	K-HIGH
U2H012	-29.42	30.49	1960-08-11	2013-08-13	54	439.03	0.0068	189.51	C	K-HIGH
U2H013	-29.51	30.09	1960-08-10	2013-09-23	54	295.70	0.0151	287.79	B	K-MID
U2H055	-29.64	30.69	1989-10-26	2013-10-09	25	3505.34	0.0078	176.19	C	K-HIGH
U6H003	-29.80	30.52	1981-11-13	2013-11-29	33	424.40	0.0076	105.33	C	K-HIGH
U7H001	-29.85	30.24	1949-07-09	2013-11-01	65	16.06	0.0614	193.09	C	K-HIGH
U7H004	-29.84	30.27	1955-01-08	1974-01-16	20	0.31	0.1133	96.84	C	K-HIGH
U7H008	-30.01	30.74	1978-12-11	2013-09-24	36	58.34	0.0201	136.89	C	K-HIGH
U8H001	-30.40	30.60	1986-05-20	2013-10-30	28	213.74	0.0146	140.29	C	K-HIGH
U8H003	-30.27	30.70	1987-05-27	2013-11-27	27	503.47	0.0095	110.10	C	K-HIGH
V1H001	-28.74	29.82	1924-11-04	2013-11-04	90	4403.61	0.0021	287.67	C	K-MID
V1H009	-28.89	29.77	1954-01-15	2013-11-04	61	196.69	0.0066	105.24	C	K-MID
V1H010	-28.82	29.55	1964-11-26	2014-01-06	51	786.65	0.0066	315.39	C	K-MID
V1H029	-28.51	29.35	1968-05-07	1993-03-23	26	20.85	0.0368	184.37	C	K-MID
V1H030	-28.51	29.34	1968-04-26	1993-03-23	26	23.14	0.0353	184.37	C	K-MID
V1H032	-28.64	29.03	1974-01-07	1993-03-22	20	68.75	0.0322	402.39	C	K-MID
V1H038	-28.56	29.75	1971-10-19	2013-11-05	43	1660.10	0.0031	137.70	C	K-MID
V2H001	-29.03	30.36	1931-09-14	1976-02-08	46	1967.08	0.0058	174.01	B	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
V2H002	-29.22	29.99	1950-06-12	2013-11-11	64	951.87	0.0035	251.29	B	K-MID
V2H004	-29.07	30.25	1960-05-01	2013-11-14	54	1555.77	0.0043	187.88	B	K-MID
V2H005	-29.36	29.88	1972-09-22	2013-11-11	42	269.13	0.0067	312.12	B	K-MID
V2H007	-29.24	29.79	1972-07-28	2013-11-12	42	115.33	0.0041	266.76	B	K-MID
V3H007	-27.85	29.84	1948-07-01	2013-11-05	66	129.50	0.0088	189.19	C	K-MID
V3H010	-28.06	30.37	1960-04-27	2014-01-08	55	5971.07	0.0007	117.55	C	K-MID
V6H003	-28.31	30.15	1954-01-01	2014-01-08	61	310.62	0.0068	123.62	C	K-MID
V6H004	-28.40	30.01	1954-01-01	2014-01-08	61	663.92	0.0041	130.29	C	K-MID
V7H012	-29.01	29.88	1962-11-17	2013-11-12	52	199.89	0.0067	141.58	C	K-MID
V7H016	-29.19	29.63	1972-10-23	2013-11-07	43	122.12	0.0214	357.84	C	K-MID
V7H017	-29.19	29.64	1972-10-23	2013-11-12	42	282.24	0.0176	418.79	C	K-MID
W1H004	-28.87	31.46	1948-08-03	2014-01-07	67	17.85	0.0070	249.30	C	K-HIGH
W1H005	-28.57	31.39	1948-08-11	2014-01-14	67	45.59	0.0172	87.42	C	K-HIGH
W1H015	-28.88	31.77	1976-11-11	1998-01-10	21	6.96	0.0180	345.74	C	K-HIGH
W1H017	-28.84	31.75	1976-11-11	1998-01-10	22	0.46	0.0287	345.74	C	K-HIGH
W2H006	-28.07	31.55	1963-08-20	2014-01-14	52	2235.08	0.0063	117.00	C	K-HIGH
W2H007	-27.96	31.19	1965-08-03	1993-11-03	30	55.48	0.0141	101.27	C	K-HIGH
W2H028	-27.94	31.21	1987-09-17	2013-11-20	28	241.66	0.0217	142.76	C	K-HIGH
W5H001	-26.26	30.55	1910-04-04	1991-10-28	65	14.59	0.0229	108.13	C	K-MID
W5H005	-26.83	30.73	1950-08-03	2013-10-23	64	811.70	0.0038	104.72	C	K-MID
W5H011	-26.28	30.59	1956-12-11	2013-10-08	58	915.54	0.0031	58.26	B	K-MID
W5H016	-26.32	30.52	1963-09-16	1992-06-29	30	10.02	0.0082	108.13	C	K-LOW
W5H022	-27.07	30.99	1968-08-15	2013-10-23	46	2350.08	0.0040	128.72	C	K-HIGH
W5H024	-26.39	30.84	1976-09-29	2013-10-08	38	1453.69	0.0049	105.41	C	K-MID
X1H001	-26.04	31.00	1909-10-01	2013-12-11	105	5560.31	0.0042	103.21	B	K-MID
X1H003	-25.68	31.78	1939-10-04	2013-09-26	75	8902.81	0.0023	117.61	C	K-HIGH
X1H012	-25.63	31.50	1967-05-12	1991-12-18	26	119.09	0.0167	348.48	C	K-HIGH

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
X1H014	-25.67	31.58	1968-08-02	2013-11-26	46	1134.47	0.0068	246.55	C	K-HIGH
X1H016	-25.95	30.57	1970-08-21	2013-12-11	44	591.16	0.0115	139.56	B	K-MID
X1H017	-25.89	30.28	1971-10-26	2013-10-22	43	2440.84	0.0031	57.23	B	K-LOW
X1H018	-25.84	30.41	1972-07-31	2013-12-10	42	2625.18	0.0046	61.60	B	K-LOW
X1H019	-25.84	30.67	1973-09-07	2013-09-17	41	188.23	0.0185	268.29	C	K-MID
X1H020	-25.84	30.68	1973-09-14	2013-10-23	41	47.87	0.0331	191.98	C	K-MID
X2H008	-25.79	30.92	1948-02-01	2013-12-03	66	182.50	0.0275	201.10	C	K-MID
X2H010	-25.61	30.87	1948-02-11	2013-12-02	66	128.28	0.0128	250.23	C	K-MID
X2H011	-25.65	30.28	1956-10-01	1999-12-11	45	400.52	0.0096	104.61	B	K-MID
X2H013	-25.45	30.71	1959-01-21	2013-12-02	55	1533.95	0.0094	146.81	C	K-MID
X2H014	-25.38	30.70	1958-12-17	2013-12-02	56	254.13	0.0170	281.35	C	K-MID
X2H016	-25.36	31.96	1960-08-24	2013-12-18	54	10380.48	0.0027	118.00	C	K-HIGH
X2H017	-25.44	31.63	1959-08-28	1998-09-01	40	8898.30	0.0044	135.68	C	K-HIGH
X2H018	-25.28	31.62	1960-08-25	1997-03-04	37	628.71	0.0038	18.45	C	K-HIGH
X2H022	-25.54	31.32	1960-08-31	2013-12-03	54	1660.43	0.0069	124.00	C	K-HIGH
X2H024	-25.71	30.84	1964-09-25	2013-12-02	50	82.22	0.0342	308.93	C	K-MID
X2H025	-25.29	30.57	1966-07-21	1992-05-13	27	24.81	0.0861	281.35	C	K-MID
X2H026	-25.29	30.57	1966-07-19	1992-05-13	27	19.21	0.0703	281.35	C	K-MID
X2H027	-25.30	30.60	1966-08-02	1992-05-13	27	73.86	0.0325	281.35	C	K-MID
X2H028	-25.30	30.57	1966-07-19	1992-05-13	27	4.49	0.1021	281.35	C	K-MID
X2H031	-25.73	30.98	1966-06-23	2013-12-03	48	266.24	0.0168	202.45	C	K-MID
X2H032	-25.51	31.22	1968-09-15	2013-11-26	46	5395.49	0.0065	169.07	C	K-MID
X2H035	-25.19	30.88	1982-01-28	2013-09-19	32	15.90	0.0396	288.04	C	K-MID
X2H047	-25.61	30.40	1985-10-24	2013-12-02	29	107.68	0.0305	119.15	B	K-MID
X2H072	-25.27	31.26	1989-12-13	2013-09-19	25	251.77	0.0074	40.91	C	K-HIGH
X3H001	-25.09	30.78	1948-03-15	2013-10-15	66	232.70	0.0190	447.36	C	K-MID
X3H002	-25.09	30.78	1963-11-08	2013-10-15	51	56.25	0.0195	447.36	C	K-MID

Gauge	Latitude (⁰)	longitude (⁰)	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
X3H006	-25.03	31.13	1958-09-04	2000-01-19	43	780.40	0.0129	392.11	C	K-HIGH
X3H011	-24.89	31.09	1978-11-28	2013-09-25	36	216.47	0.0111	482.75	C	K-HIGH
X4H004	-24.45	31.98	1960-11-23	2013-09-17	54	938.69	0.0029	6.50	C	K-MID
A2R001	-25.73	25.73	1923-02-22	2014-02-03	91	37.25	0.0044	37.25	B	K-MID
A2R003	-25.79	25.79	1929-01-23	2011-01-04	82	23.40	0.0081	23.40	B	K-MID
A2R005	-25.78	25.78	1936-02-12	2013-01-19	77	36.80	0.0099	36.80	C	K-MID
A2R006	-25.56	25.56	1928-01-12	2008-03-17	80	23.52	0.0040	23.52	B	K-MID
A2R007	-25.50	25.50	1947-03-30	2013-03-22	66	19.80	0.0055	19.80	B	K-LOW
A2R009	-25.62	25.62	1959-01-25	2014-03-10	55	42.00	0.0048	42.00	B	K-MID
A2R011	-25.70	25.70	1965-02-18	2012-01-26	47	20.05	0.0077	20.05	B	K-LOW
A2R012	-25.13	25.13	1955-02-01	2012-11-29	57	22.11	0.0016	22.11	C	K-LOW
A2R014	-25.31	25.31	1951-05-16	2013-01-18	62	18.01	0.0024	18.01	C	K-MID
A2R015	-25.41	25.41	1951-05-16	2012-11-25	61	33.69	0.0029	33.69	B	K-MID
A3R001	-25.47	25.47	1934-12-11	2012-11-23	78	24.15	0.0049	24.15	B	K-LOW
A3R002	-25.52	25.52	1907-03-04	2011-02-07	104	7.14	0.0073	7.14	B	K-LOW
A3R003	-25.44	25.44	1956-02-26	2012-12-04	56	9.88	0.0052	9.88	B	K-LOW
A3R004	-24.87	24.87	1958-02-17	2014-03-13	56	12.90	0.0018	12.90	C	K-LOW
A4R001	-23.98	23.98	1962-12-06	2014-03-12	52	55.46	0.0029	55.46	C	K-MID
A5R001	-23.63	23.63	1958-01-06	2014-03-12	56	52.19	0.0050	52.19	C	K-MID
A5R002	-23.38	23.38	1958-01-06	2013-01-22	55	38.01	0.0034	38.01	C	K-LOW
A6R001	-24.28	24.28	1938-11-25	2013-01-20	75	38.41	0.0034	38.41	C	K-MID
A6R002	-23.19	23.19	1961-02-17	2013-01-21	52	23.61	0.0013	23.61	C	K-LOW
A8R002	-22.63	22.63	1964-12-16	2013-01-20	49	84.25	0.0323	84.25	C	K-MID
A8R003	-22.63	22.63	1964-12-14	2013-01-20	49	84.25	0.0281	84.25	C	K-HIGH
A8R004	-22.95	22.95	1991-03-26	2013-01-21	22	158.48	0.0176	158.48	C	K-MID
A9R001	-23.11	23.11	1946-01-08	2013-01-20	67	55.69	0.0077	55.69	C	K-HIGH
A9R002	-22.95	22.95	1964-02-10	2013-01-15	49	336.61	0.0277	336.61	C	K-HIGH

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
A9R004	-22.98	22.98	1946-01-08	2014-01-06	68	122.35	0.0037	122.35	C	K-HIGH
B1R001	-25.89	25.89	1904-11-21	2013-12-04	109	34.64	0.0009	34.64	B	K-LOW
B1R002	-25.77	25.77	1958-03-15	2013-12-03	55	27.96	0.0020	27.96	B	K-LOW
B2R001	-25.89	25.89	1905-03-04	2014-03-31	109	37.43	0.0029	37.43	B	K-LOW
B3R001	-25.23	25.23	1933-12-15	2014-03-11	81	27.19	0.0040	27.19	C	K-MID
B3R002	-25.42	25.42	1937-12-25	2014-03-08	77	37.73	0.0027	37.73	B	K-LOW
B3R005	-25.10	25.10	1985-02-11	2014-03-04	29	14.64	0.0024	14.64	C	K-LOW
B4R001	-25.28	25.28	1962-11-10	2013-12-01	51	58.34	0.0294	58.34	B	K-LOW
B4R002	-25.23	25.23	1962-11-18	2014-01-08	52	58.34	0.0489	58.34	B	K-LOW
B4R004	-24.96	24.96	1960-12-21	2013-12-11	53	100.06	0.0135	100.06	B	K-LOW
B5R002	-24.78	24.78	1937-12-25	2014-03-08	77	28.54	0.0019	28.54	B	K-LOW
B6R001	-24.93	24.93	1957-03-12	2014-02-01	57	196.78	0.0285	196.78	C	K-LOW
B6R003	-24.54	24.54	1951-01-17	2014-02-03	63	168.46	0.0073	168.46	C	K-MID
B7R001	-24.52	24.52	1951-04-23	2014-03-05	63	211.41	0.0130	211.41	C	K-HIGH
B7R003	-24.10	24.10	1948-10-23	2013-12-29	65	96.79	0.0383	96.79	C	K-HIGH
B8R001	-23.94	23.94	1948-07-14	2013-01-16	65	374.54	0.0048	374.54	C	K-HIGH
B8R002	-23.75	23.75	1978-01-28	2013-01-20	35	320.28	0.0260	320.28	C	K-HIGH
B8R003	-23.82	23.82	1972-03-24	2013-01-16	41	320.28	0.0234	320.28	C	K-HIGH
B8R006	-23.81	23.81	1978-01-02	2011-12-16	33	374.54	0.0153	374.54	C	K-HIGH
B8R007	-23.27	23.27	1987-02-05	2013-01-20	26	39.52	0.0039	39.52	C	K-HIGH
D1R002	-29.34	29.34	1986-02-04	2014-03-11	28	277.65	0.0036	277.65	B	K-MID
D1R003	-29.46	29.46	1985-12-22	2014-03-11	29	257.63	0.0093	257.63	B	K-MID
D2R001	-30.05	30.05	1938-02-27	2014-03-11	76	51.83	0.0059	51.83	B	K-MID
D2R002	-29.36	29.36	1935-03-21	2014-02-25	79	48.43	0.0027	48.43	B	K-MID
J1R001	-33.52	33.52	1980-03-13	2014-01-08	34	5.90	0.0142	5.90	A	K-MID
J1R004	-33.83	33.83	1977-05-09	2012-10-22	35	7.23	0.0086	7.23	A	K-MID
J2R001	-33.49	33.49	1920-03-21	2013-04-03	93	45.58	0.0067	45.58	B	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
J2R002	-32.62	32.62	1959-03-02	2012-12-02	53	13.93	0.0045	13.93	B	K-MID
J2R003	-33.25	33.25	1931-04-15	2012-10-30	81	32.85	0.0168	32.85	A	K-MID
J2R004	-32.24	32.24	1960-08-19	2012-12-13	52	17.32	0.0361	17.32	B	K-MID
J2R006	-33.31	33.31	1921-12-29	2014-01-08	93	9.59	0.0026	9.59	B	K-MID
J3R001	-33.64	33.64	1912-06-18	2012-10-23	100	33.96	0.0048	33.96	A	K-MID
J3R002	-33.51	33.51	1923-11-14	2013-02-10	90	7.29	0.0032	7.29	A	K-MID
N1R001	-32.24	32.24	1925-02-24	2013-12-10	88	11.94	0.0074	11.94	B	K-MID
N2R001	-33.21	33.21	1923-07-20	2010-04-29	87	12.95	0.0019	12.95	B	K-MID
Q4R001	-32.23	25.82	1926-01-27	1996-11-23	70	13.57	0.0037	13.57	B	K-MID
Q4R002	-32.11	32.11	1926-01-27	2013-11-16	87	15.92	0.0042	15.92	B	K-MID
Q8R001	-32.97	32.97	1981-05-31	2014-01-06	33	20.80	0.0038	20.80	B	K-HIGH
Q9R001	-32.57	32.57	1965-11-03	2013-11-16	48	92.54	0.0138	92.54	B	K-HIGH
V1R001	-28.68	28.68	1931-11-11	2014-03-09	83	361.05	0.0036	361.05	C	K-MID
V1R002	-28.76	28.76	1931-11-11	2008-03-16	77	466.23	0.0099	466.23	C	K-MID
V1R003	-28.76	28.76	1931-11-11	2013-12-29	82	458.10	0.0140	458.10	C	K-MID
V2R001	-29.16	29.16	1963-07-03	2014-03-03	51	174.52	0.0119	174.52	B	K-MID
V2R002	-29.25	29.25	1951-01-22	2013-12-26	62	251.29	0.0039	251.29	B	K-MID
V2R003	-29.32	29.96	1972-12-26	2014-02-05	42	276.44	0.0039	276.44	B	K-MID
V3R001	-27.95	27.95	1962-09-12	2014-01-30	52	119.07	0.0017	119.07	C	K-MID
V3R003	-27.44	27.44	1947-11-07	2014-03-08	67	193.31	0.0017	193.31	B	K-MID
V7R001	-29.04	29.04	1947-02-05	2013-12-26	66	296.73	0.0070	296.73	C	K-MID
W1R001	-28.77	28.77	1956-02-13	2013-12-15	57	102.97	0.0066	102.97	C	K-HIGH
W1R002	-28.87	28.87	1948-04-09	1977-12-20	29	249.30	0.0075	249.30	C	K-HIGH
W2R001	-27.84	27.84	1972-02-25	2013-11-15	41	142.78	0.0073	142.78	C	K-HIGH
W3R001	-28.12	28.12	1963-07-04	2014-03-09	51	51.89	0.0031	51.89	C	K-HIGH
W5R001	-26.66	26.66	1967-02-13	2013-11-26	46	109.17	0.0024	109.17	C	K-LOW
W5R002	-26.51	26.51	1951-12-16	2014-03-08	63	95.09	0.0016	95.09	C	K-MID

Gauge	Latitude (^{0})	longitude (^{0})	Start date	End date	Record length (years)	Area (km^2)	Slope (%)	MAR90 (mm)	Veld Zone	RMF K-region
W5R003	-26.71	26.71	1951-02-17	2014-01-24	63	96.58	0.0027	96.58	C	K-MID
X1R001	-25.95	25.95	1961-03-04	2014-03-07	53	40.03	0.0027	40.03	B	K-LOW
X1R003	-25.88	25.88	1972-01-23	2013-12-13	41	84.85	0.0046	84.85	B	K-MID
X1R004	-25.71	25.71	1969-01-07	2014-03-05	45	250.02	0.0077	250.02	C	K-HIGH
X2R001	-25.28	25.28	1977-01-23	2014-03-06	37	270.32	0.0054	270.32	C	K-MID
X2R002	-25.22	25.22	1977-03-06	2014-03-09	37	270.32	0.0084	270.32	C	K-MID
X2R003	-25.24	25.24	1970-02-01	2013-12-29	43	288.04	0.0275	288.04	C	K-MID
X2R004	-25.39	25.39	1971-03-26	2014-03-06	43	169.26	0.0080	169.26	C	K-MID
X2R005	-25.36	25.36	1959-02-20	2014-03-10	55	132.61	0.0136	132.61	B	K-MID
X3R001	-25.14	25.14	1974-01-14	2014-03-05	40	291.04	0.0187	291.04	C	K-MID
X3R002	-24.88	24.88	1979-03-04	2013-12-30	34	482.75	0.0091	482.75	C	K-HIGH

APPENDIX B: STATIONS WHERE QUATERNARIES WERE INCORRECTLY LISTED BY DWS

Appendix B provides the complete list of stations, where the stations have been reported by the DWS as being located in one quaternary, but after inspection in this study, they were found to be located in another quaternary.

Table B1 List of stations where the quaternary catchments needed to be changed

Gauge	DWS Quaternary	Actual contributing quaternary
A2H044	A21H	A21C
A2H045	A21H	A21E
A2H049	A21E	A21D
A2H063	A23E	A23D
A7H001	A71J	A71H
B6H002	B60D	B60C
B7H003	B72G	B72F
B7H010	B72H	B72E
B7H014	B72G	B72F
B7H020	B73G	B73F
C1H006	C11J	C11H
C1H012	C12H	C12C
C2H018	C23L	C23C
C4H002	C43D	C43C
C4H004	C43C	C43A
C7H006	C70K	C70J
C8H005	C81H	C81F
D1H004	D14C	D14B
D1H032	D17C	D17B
D5H013	D55J	D55H
H1H013	H10C	H10B
H8H001	H80E	H80D
L8H005	L82E	L82D
N2H008	N22E	N22D

Gauge	DWS Quaternary	Actual contributing quaternary
P4H001	P40C	P40B
Q3H005	Q30D	Q30C
R1H013	R10K	R10J
R2H005	R20D	R20B
R2H015	R20E	R20D
S3H004	S32H	S32C
T1H004	T13B	T13A
T3H005	T34J	T34H
T3H006	T35L	T35K
T3H007	T33H	T33G
T4H001	T40D	T40C
T5H003	T51E	T51D
T5H012	T52L	T52K
U2H006	U20E	U20D
U7H004	U70B	U70A
V1H009	V14D	V14C
V1H032	V11C	V11A
V6H003	V60E	V60D
W2H006	W22G	W22H
W2H028	W22C	W22A
W5H024	W55E	W55C
X1H001	X12K	X12H
X1H012	X14G	X14F
X1H014	X14H	X14G
X1H016	X12D	X12C
X1H018	X11G	X11F
X2H008	X23F	X23E
X2H014	X22B	X22A
X2H024	X23D	X23C
X3H006	X31G	X31D
X4H004	X40B	X40A

APPENDIX C: DOMINANT UPSTREAM VELD ZONES

Appendix C provides a series of maps illustrating the dominant Veld zone upstream of a station. The Veld zones for these stations differ from those reported by Görgens (2007).

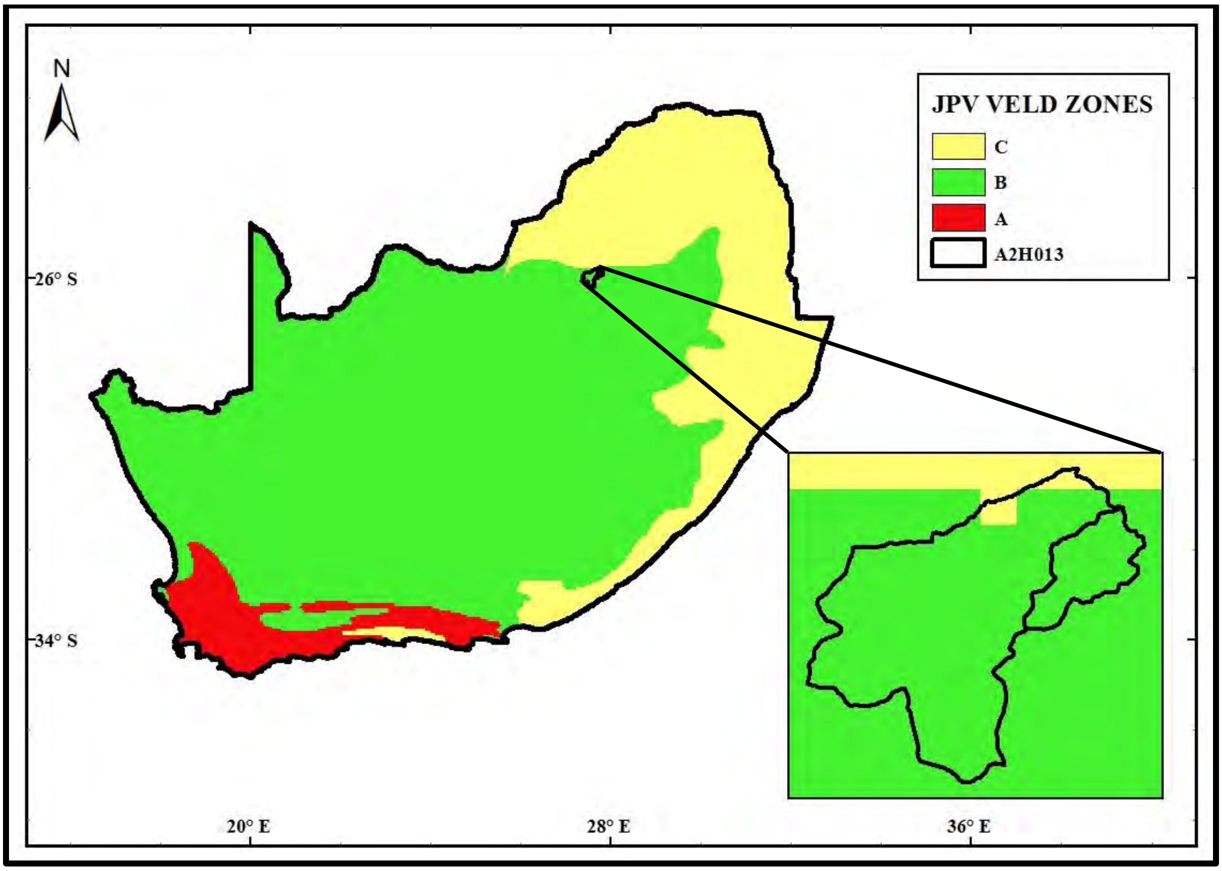

Figure C1 Dominant Veld zone upstream of station A2H013

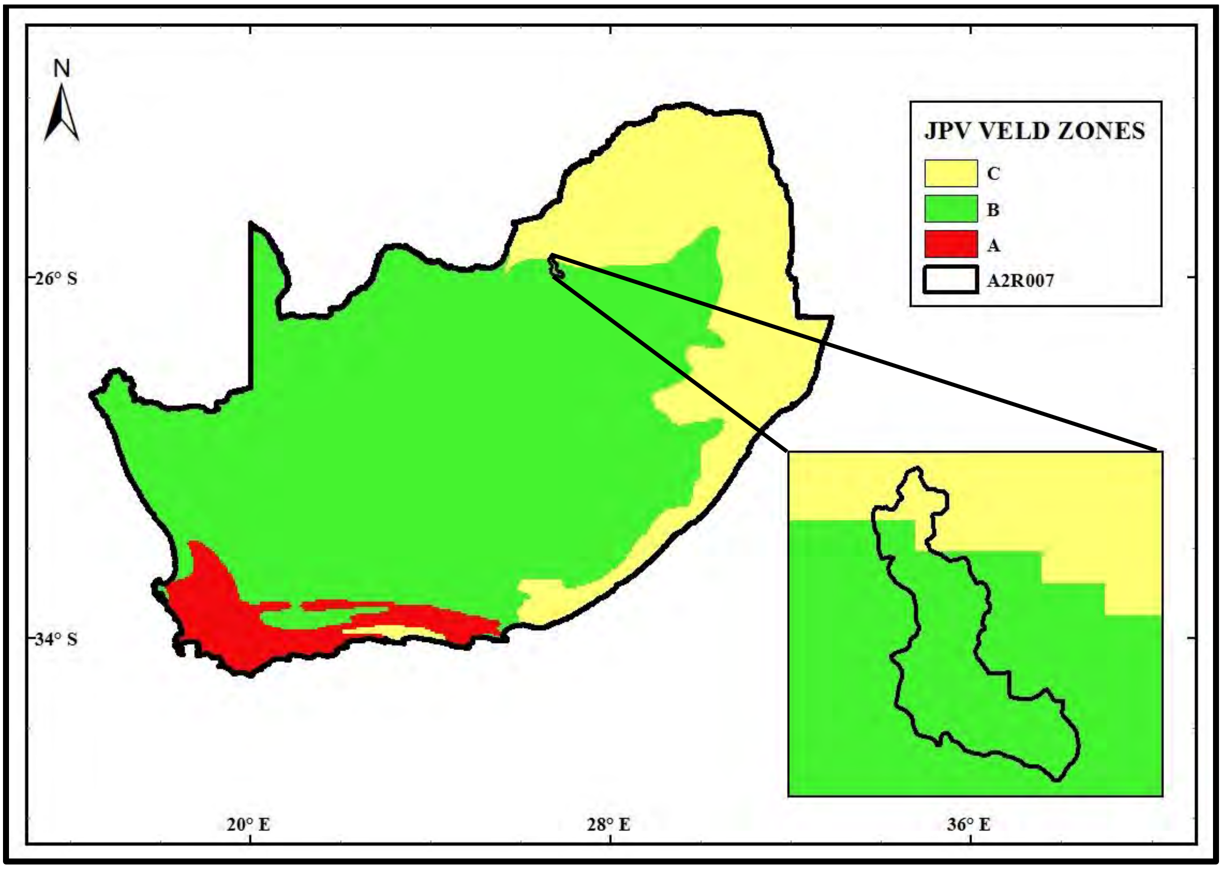

Figure C2 Dominant Veld zone upstream of station A2R007

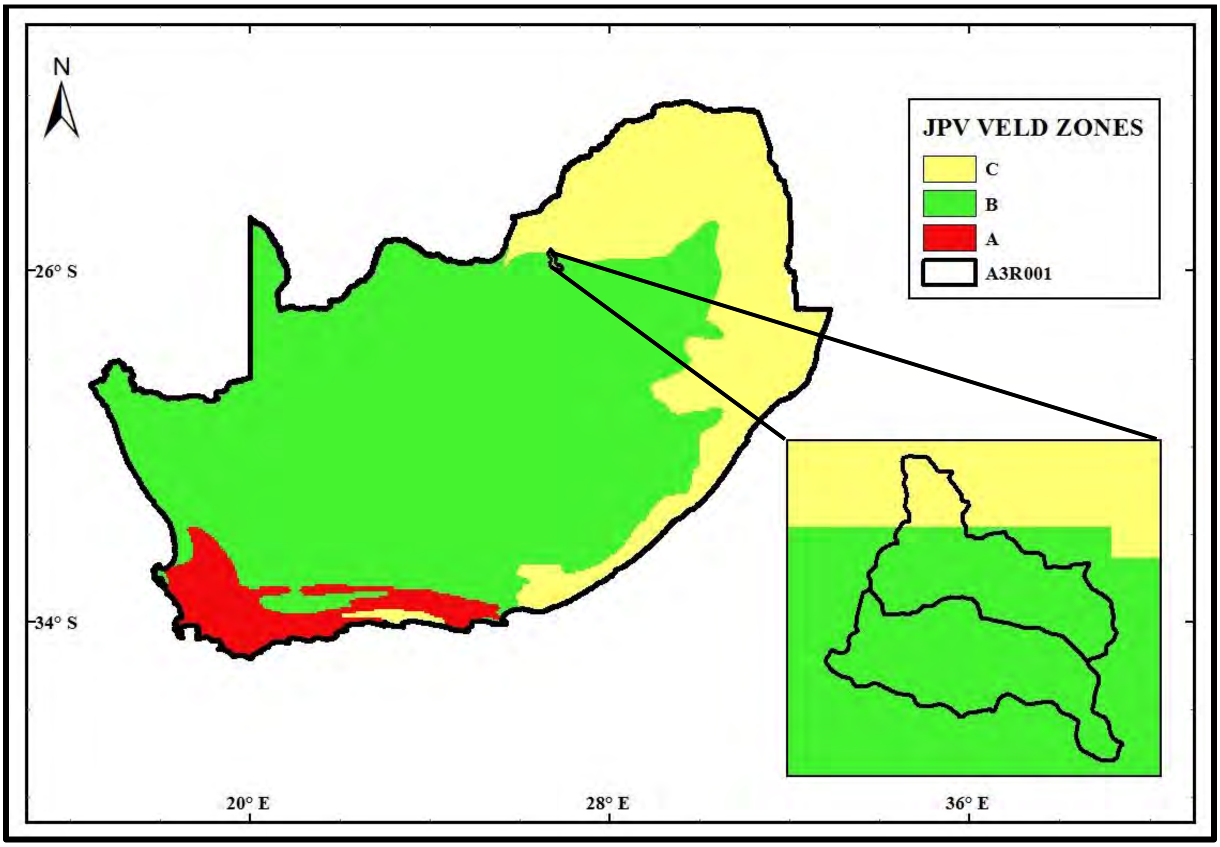

Figure C3 Dominant Veld zone upstream of station A3R001

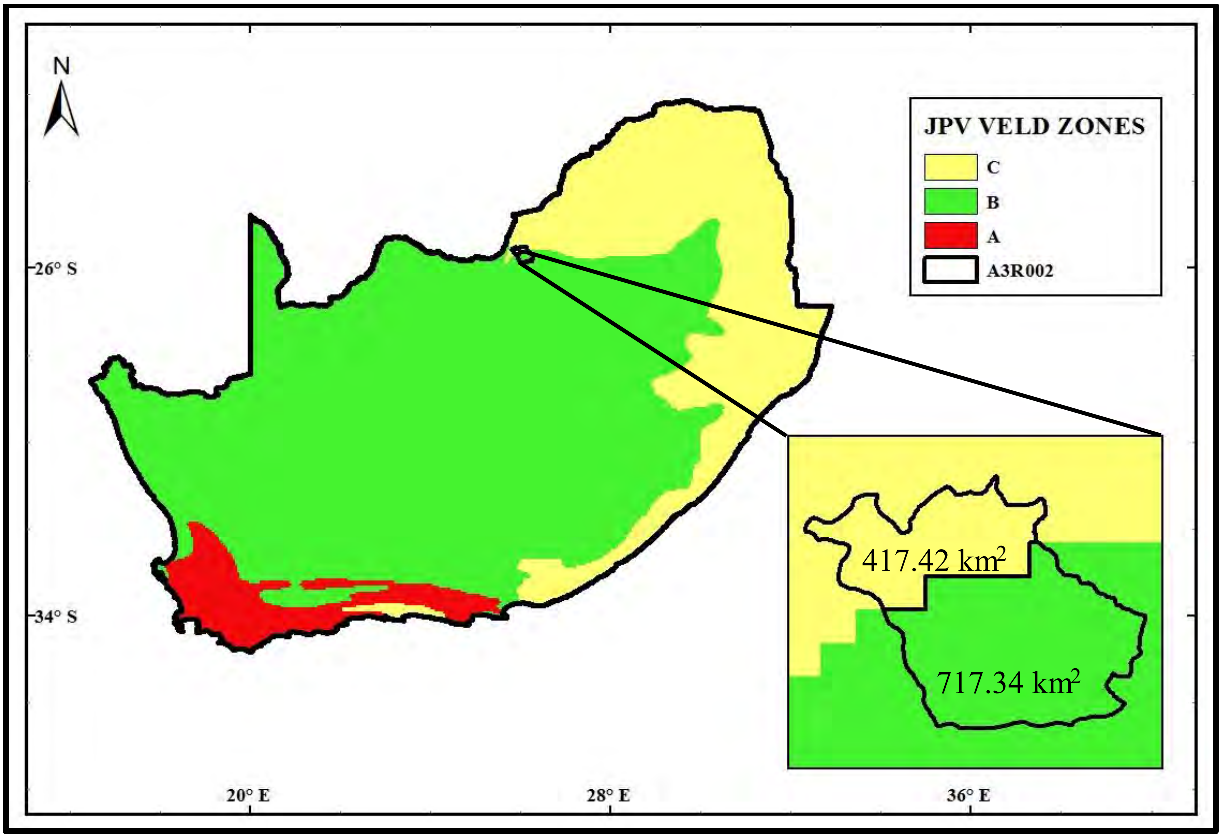

Figure C4 Dominant Veld zone upstream of station A3R002

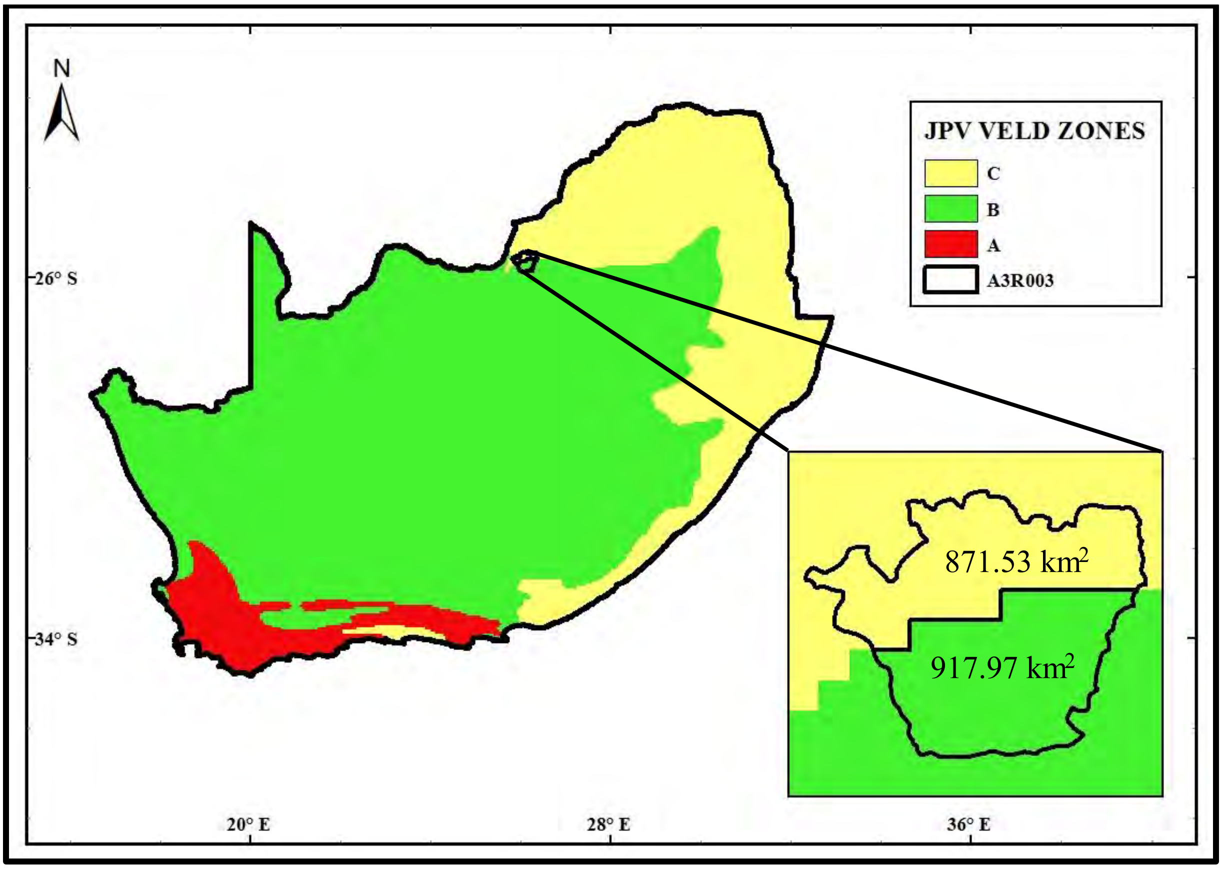

Figure C5 Dominant Veld zone upstream of station A3R003

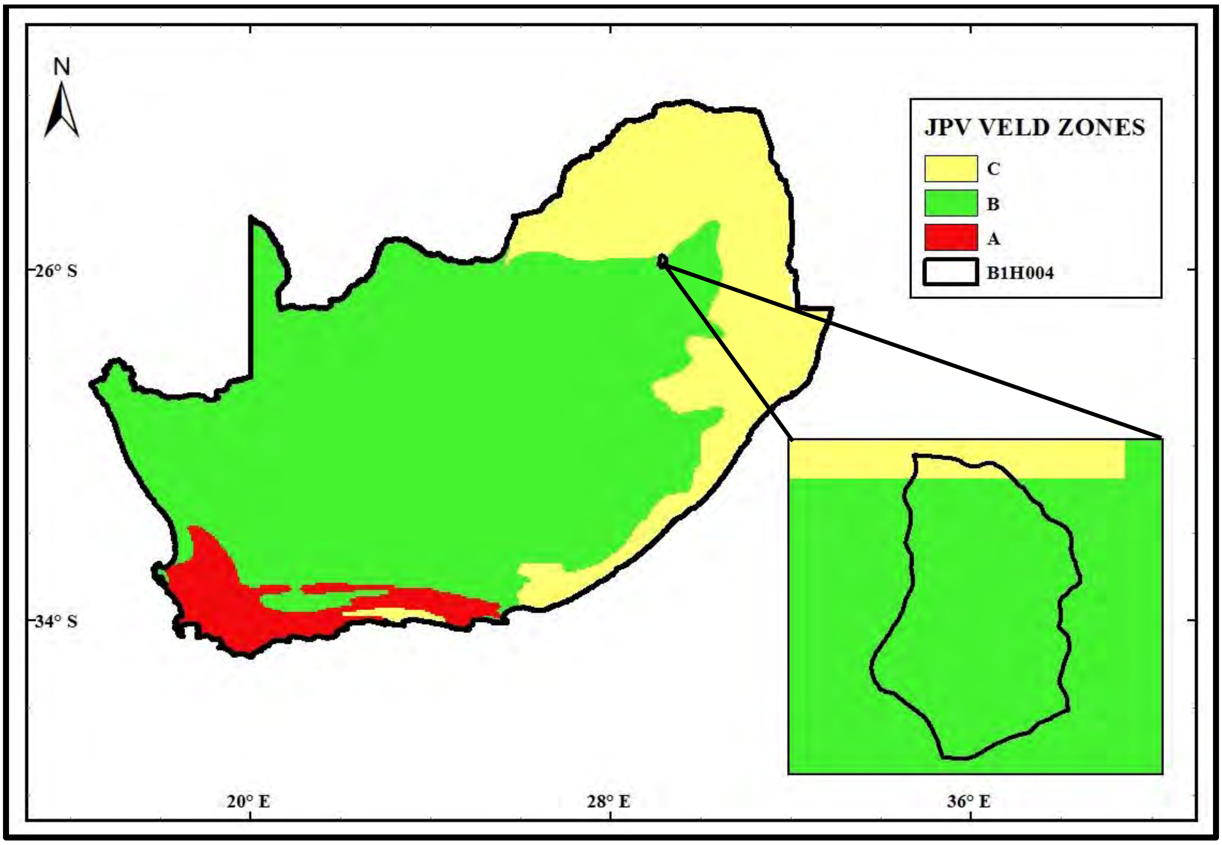

Figure C6 Dominant Veld zone upstream of station B1H004

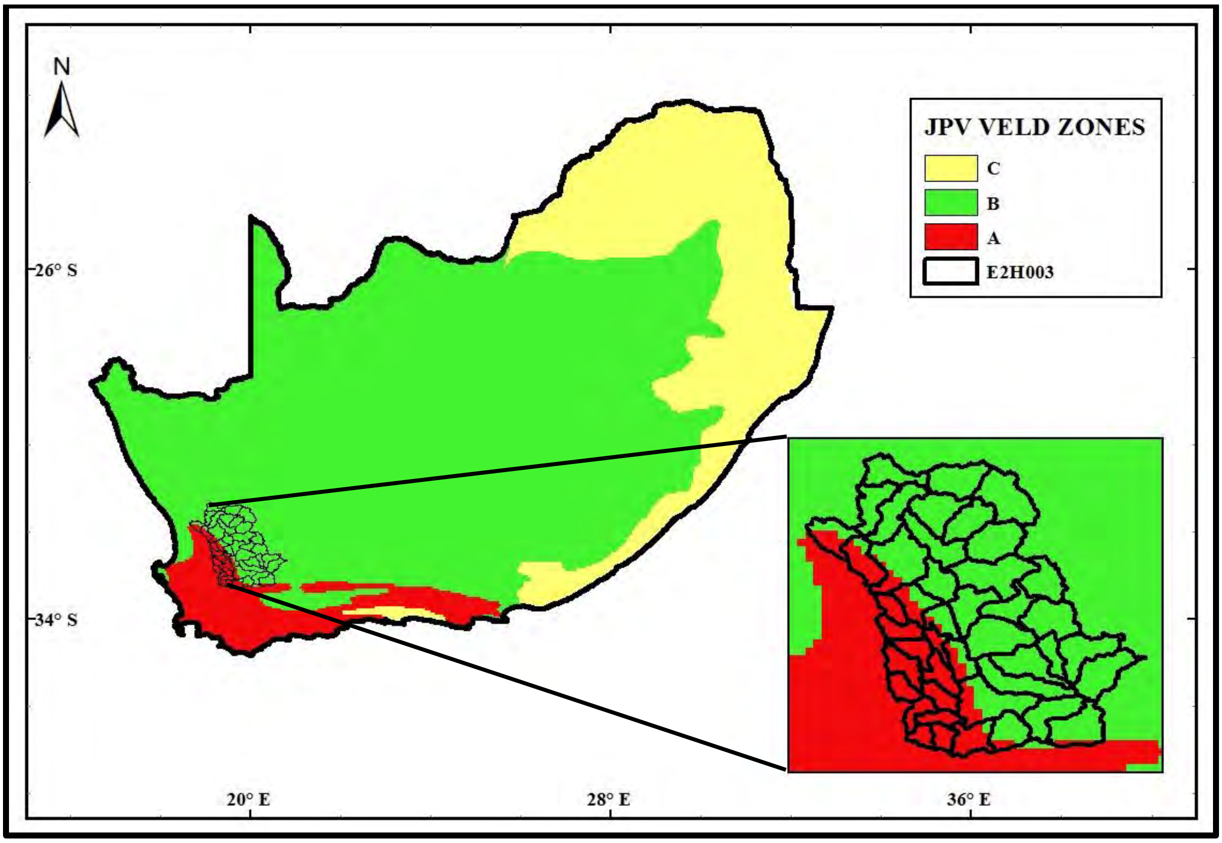

Figure C7 Dominant Veld zone upstream of station E2H003

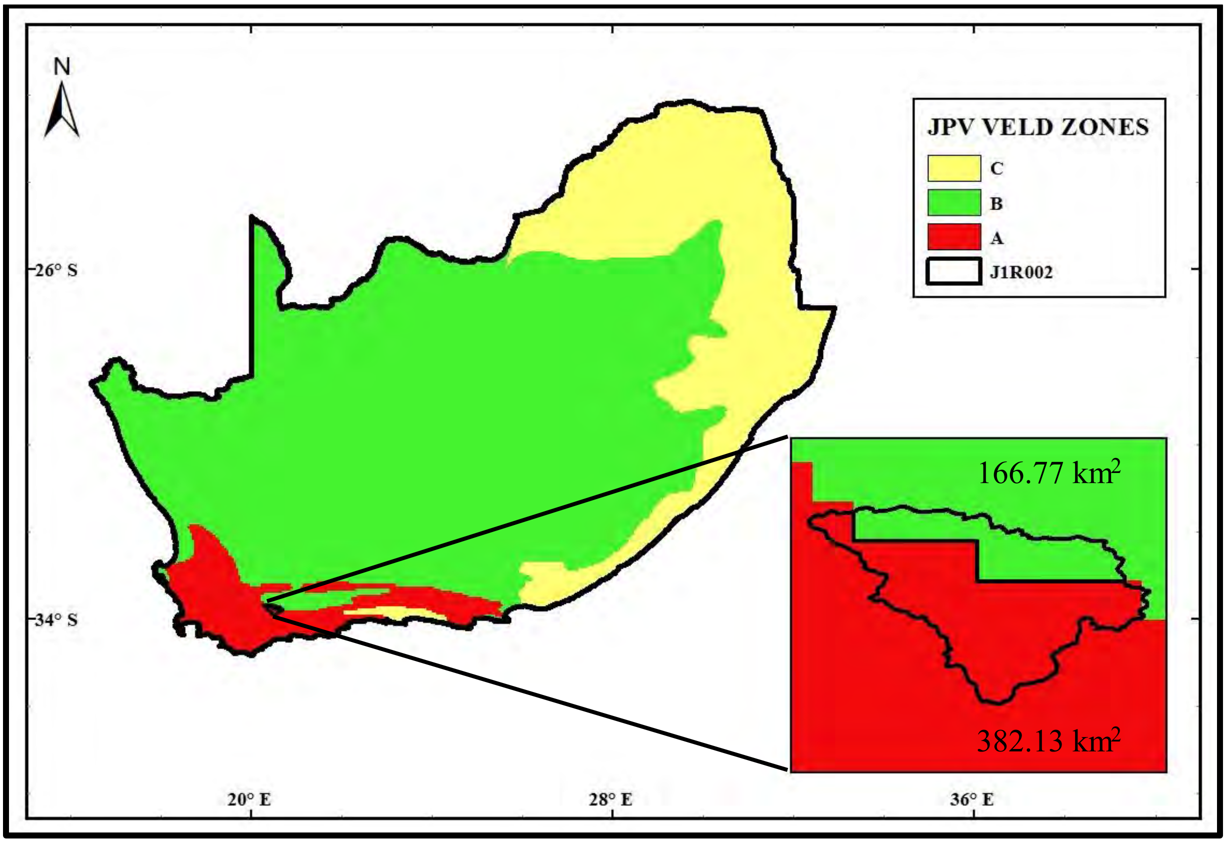

Figure C8 Dominant Veld zone upstream of station J1R002

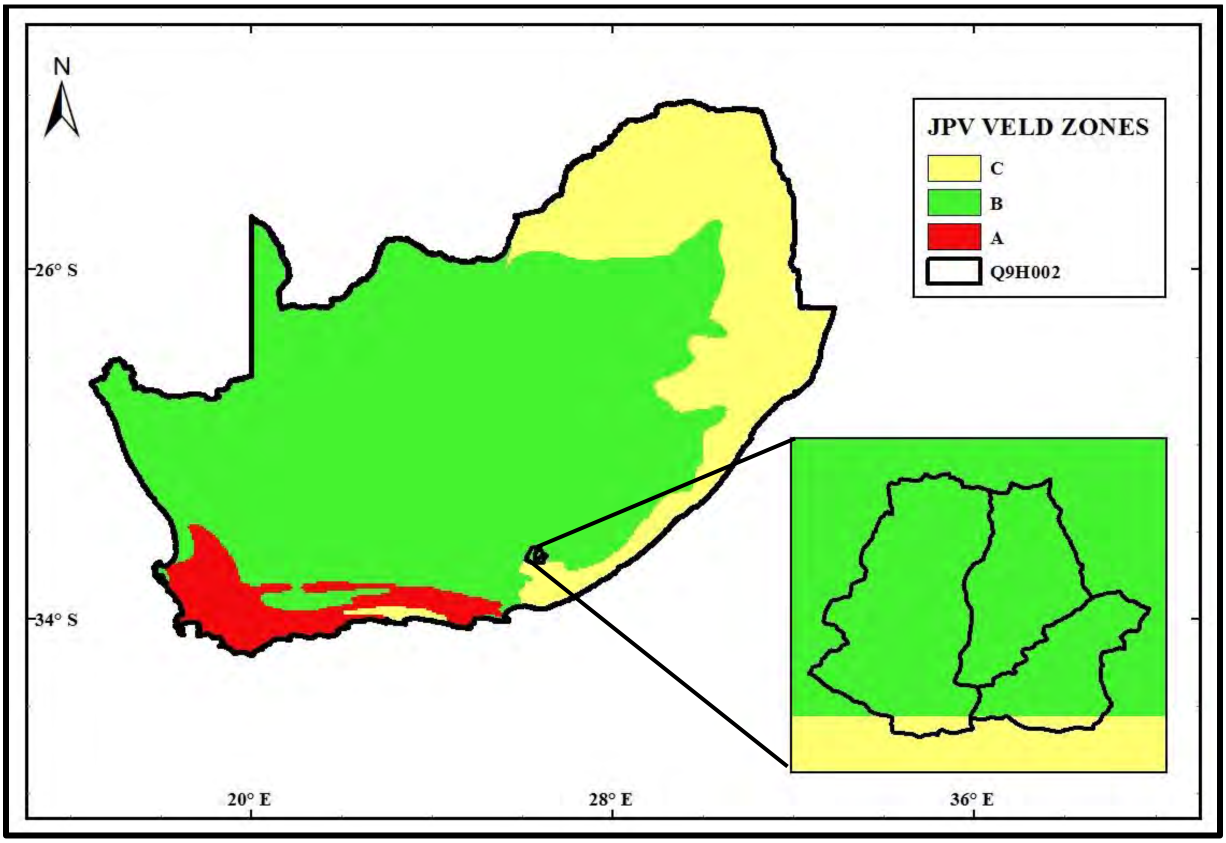

Figure C9 Dominant Veld zone upstream of station Q9H002

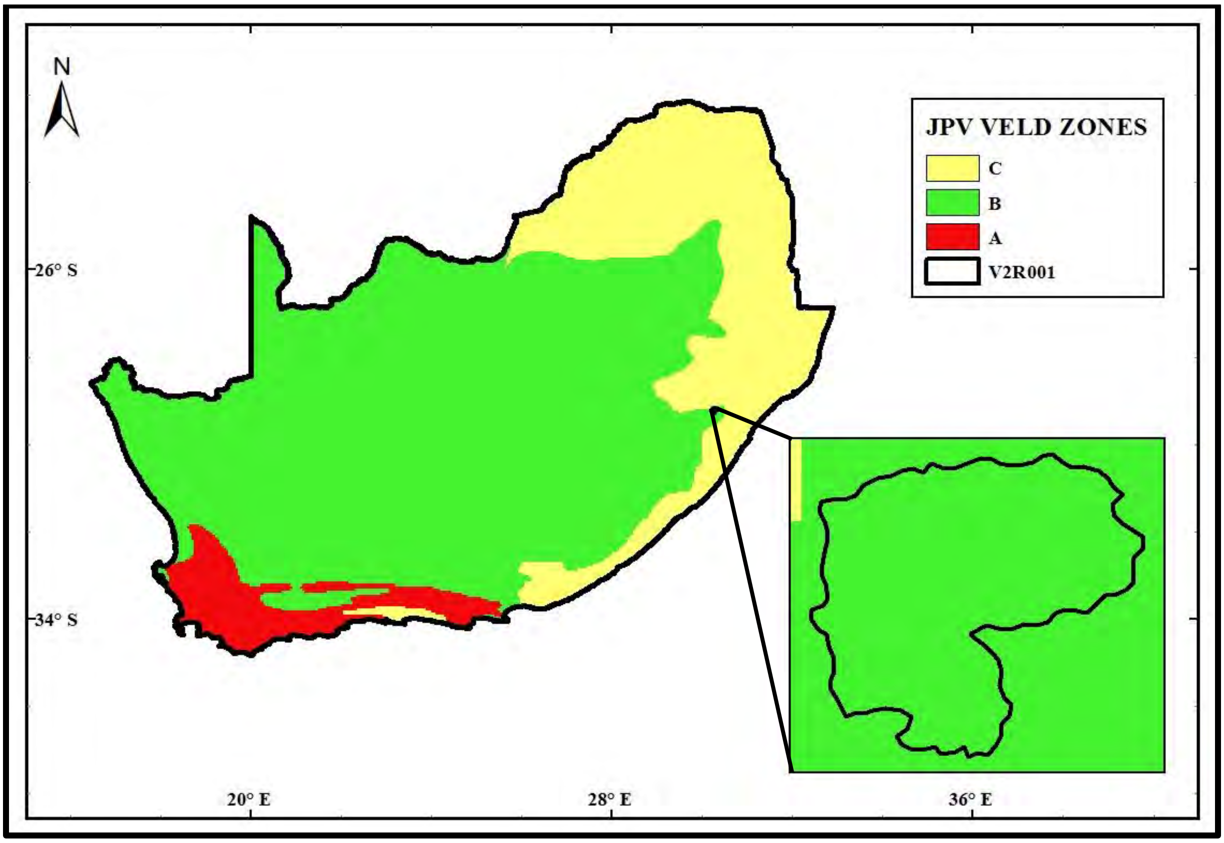

Figure C10 Dominant Veld zone upstream of station V2R001

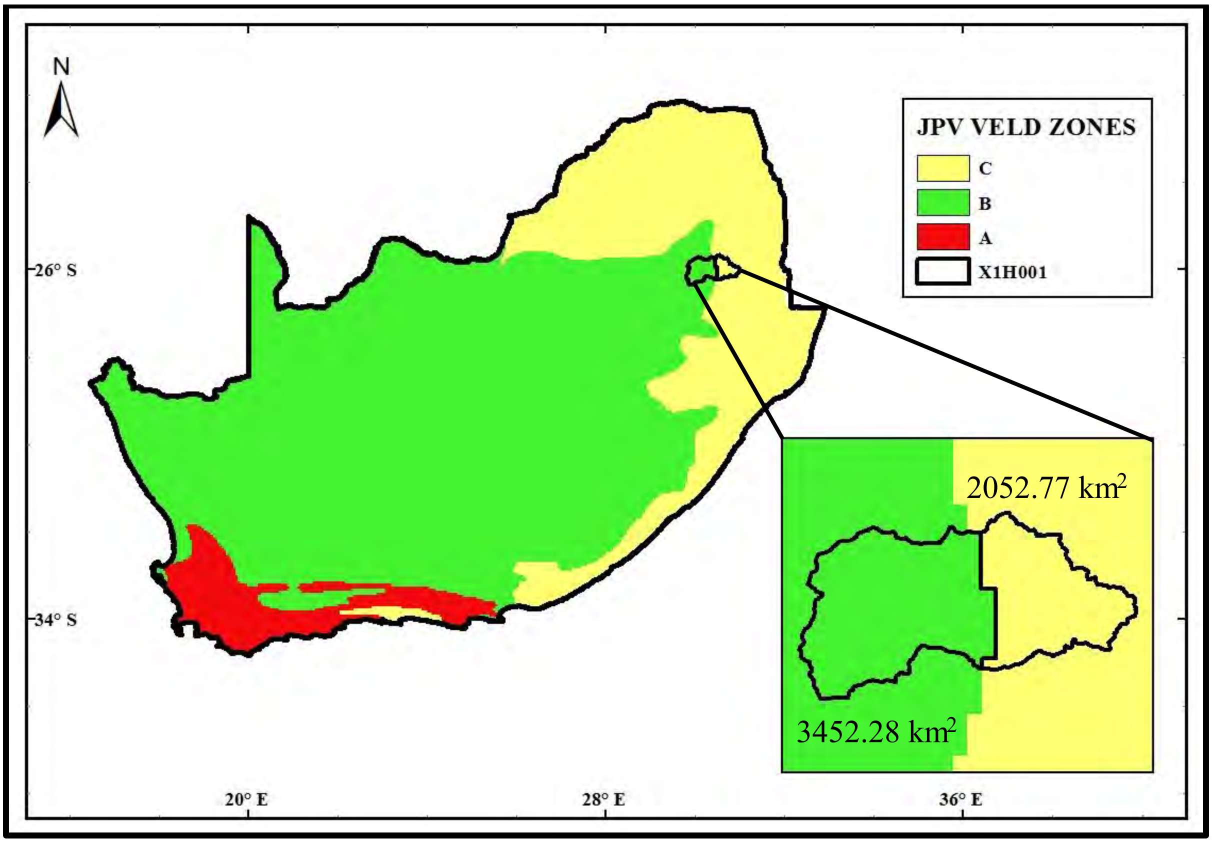

Figure C11 Dominant Veld zone upstream of station X1H001

APPENDIX D: K-REGION AT STATION LOCATION

Appendix D provides a series of maps, illustrating the K-Region at a station location. These K-Regions differ from those reported by Görgens (2007).

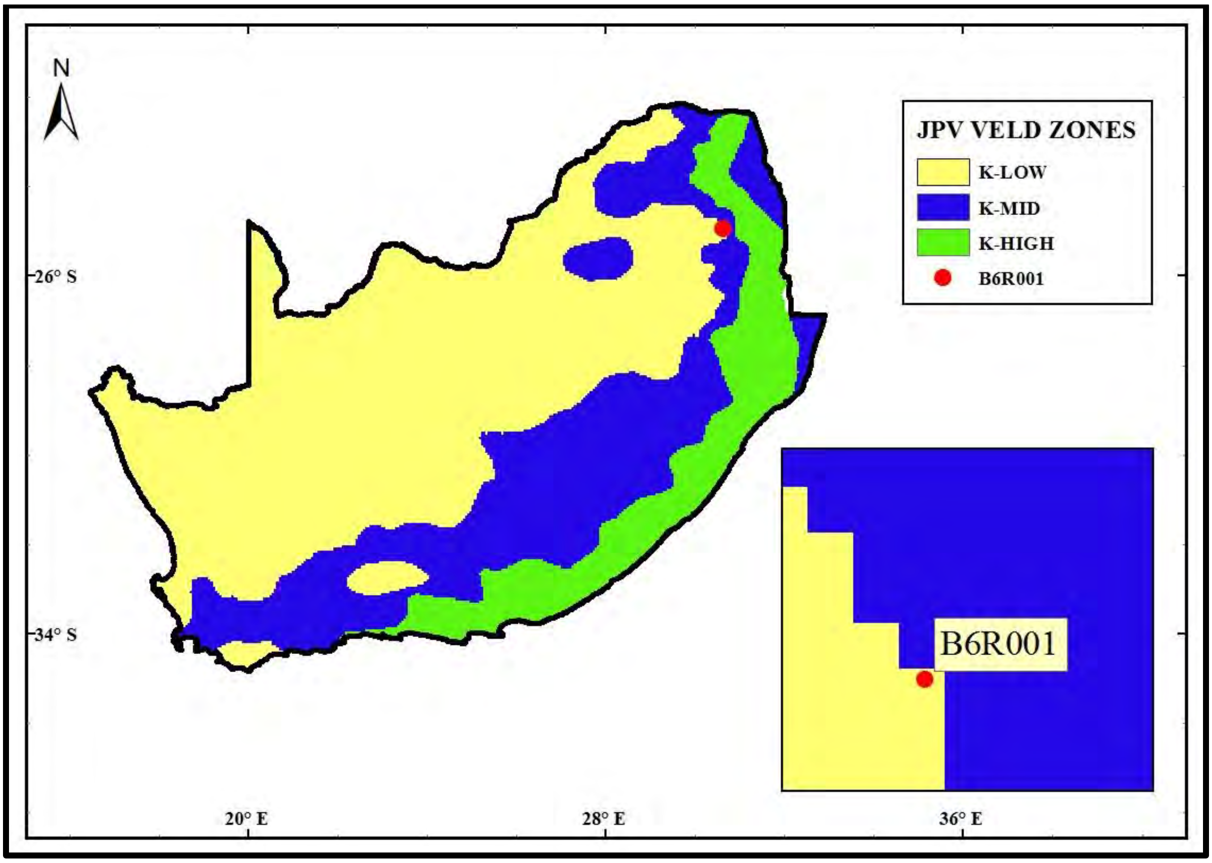

Figure D1 K-Region at station B6R001

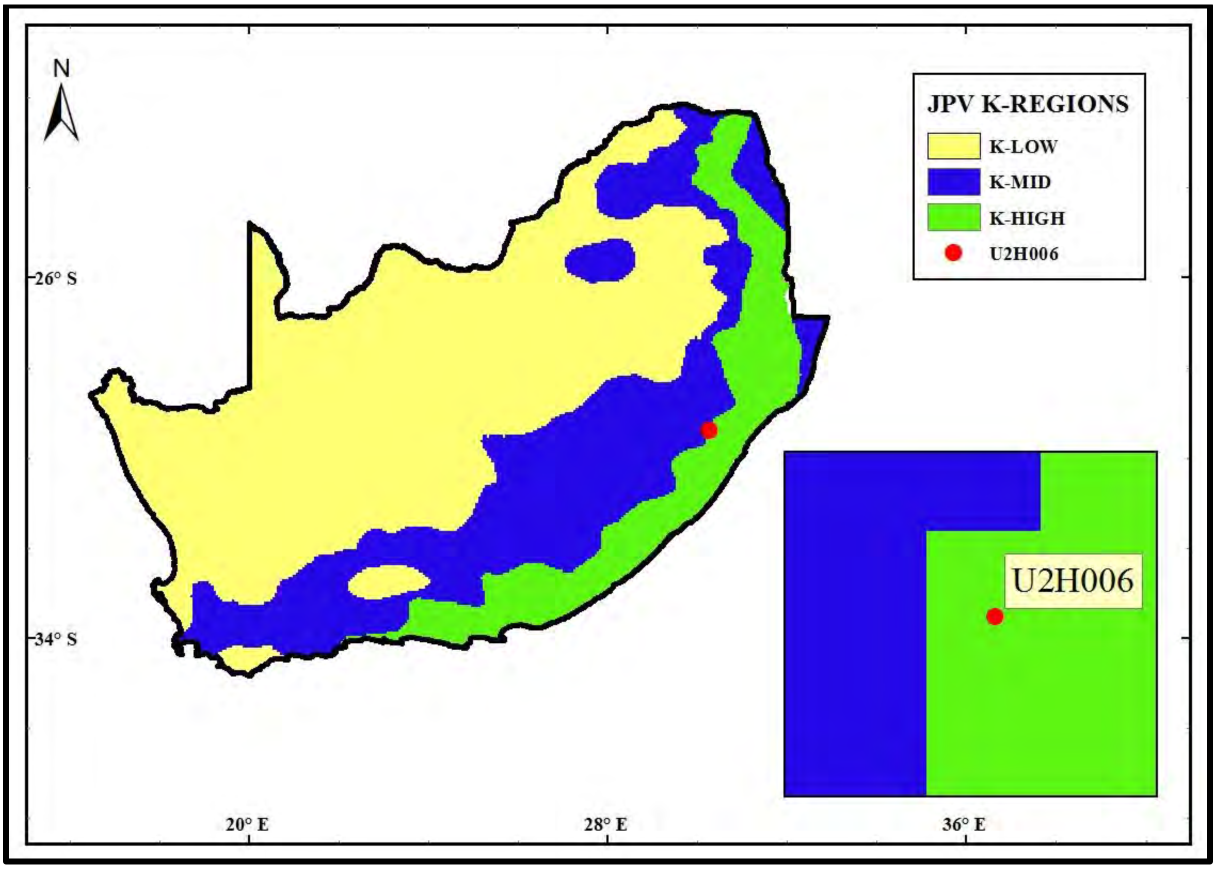

Figure D2 K-Region at station U2H006

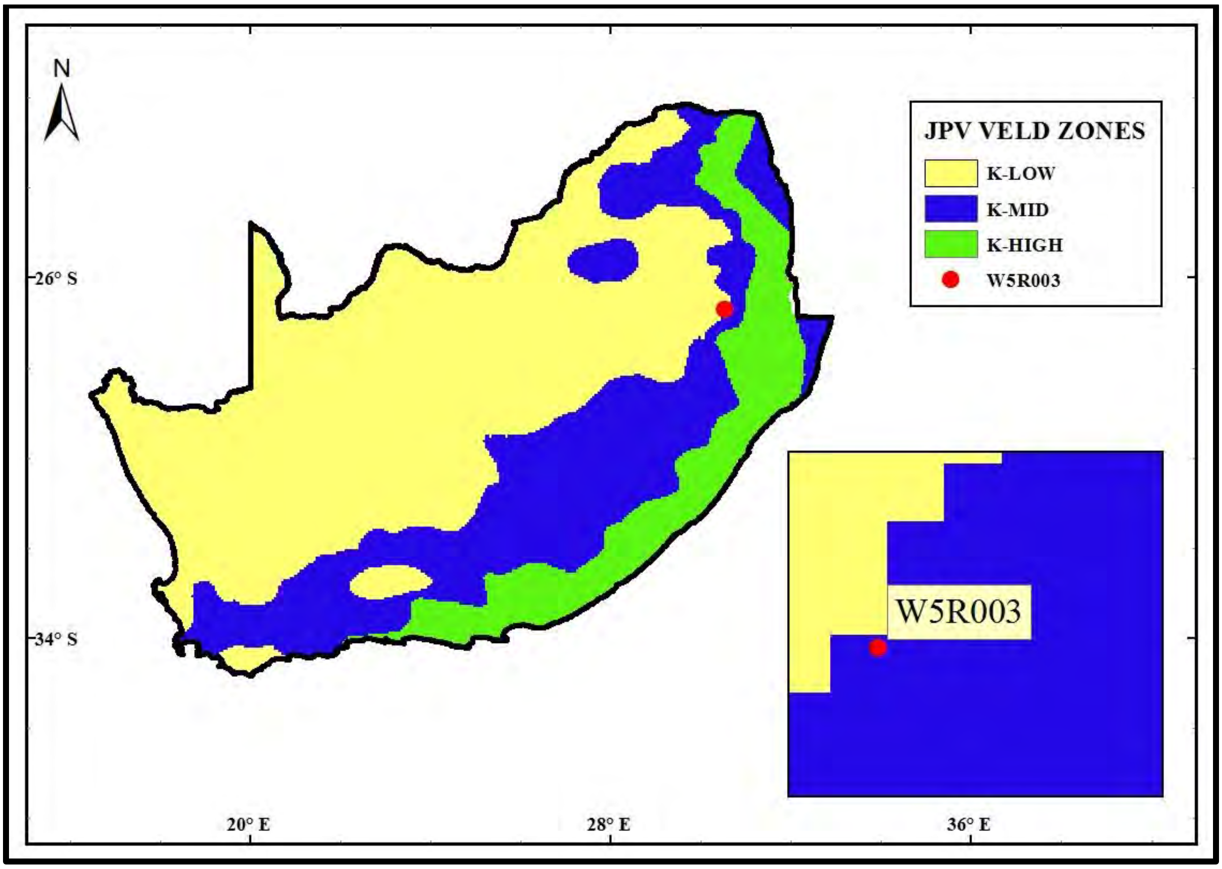

Figure D3 K-Region at station W5R003

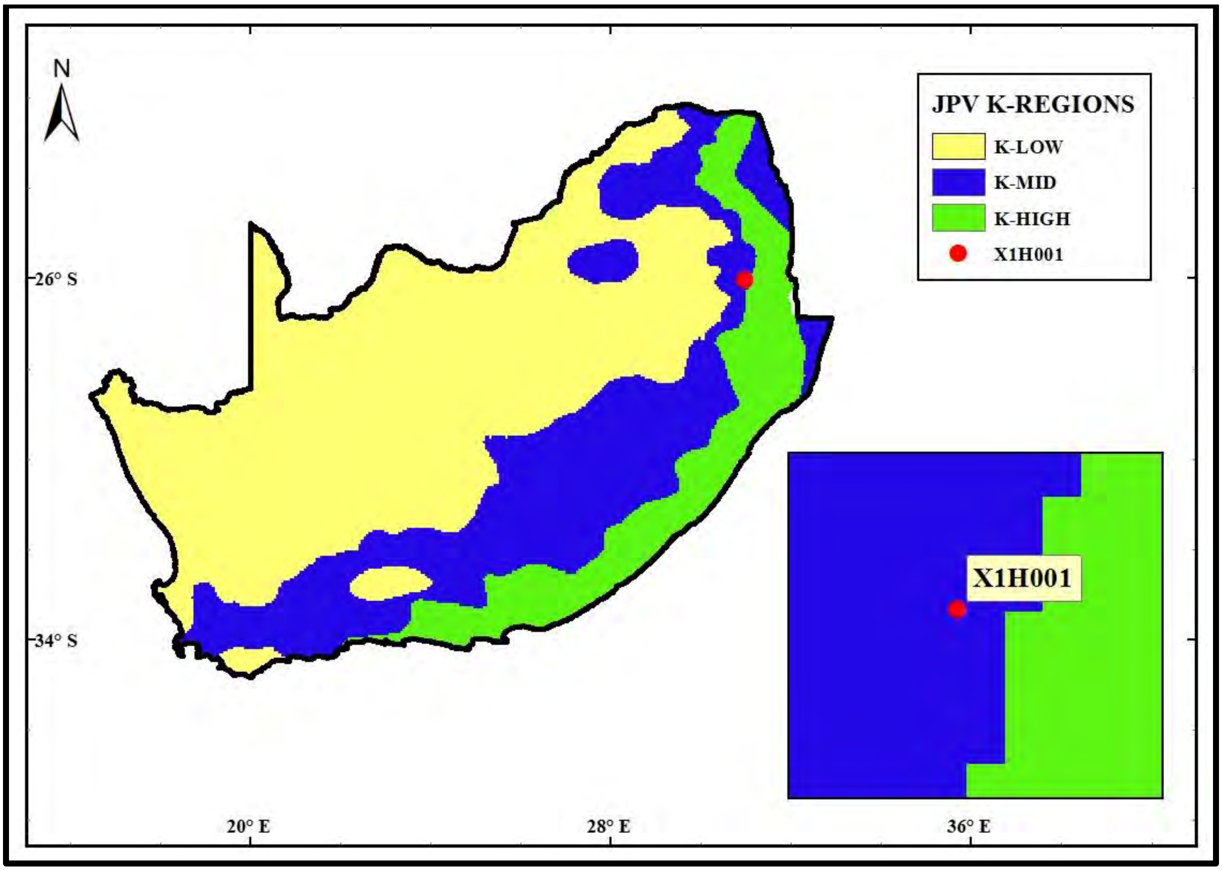

Figure D4 K-Region at station X1H001

APPENDIX E: FREQUENCY PLOTS OF ABSOLUTE RELATIVE ERRORS

Appendix E provides a series of graphs, which illustrate the frequency of absolute relative errors for the 2-, 10-, 20- and 100-year return periods.

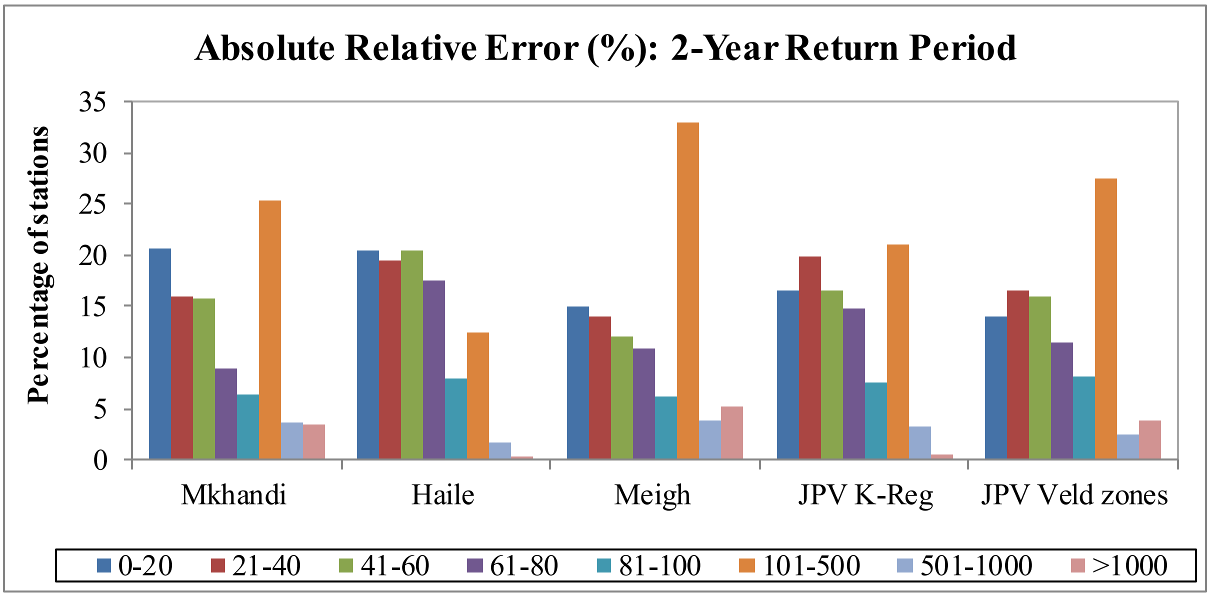

Figure E1 Performance of RFFA, based on the frequency of absolute relative errors for the 2-year Return Period

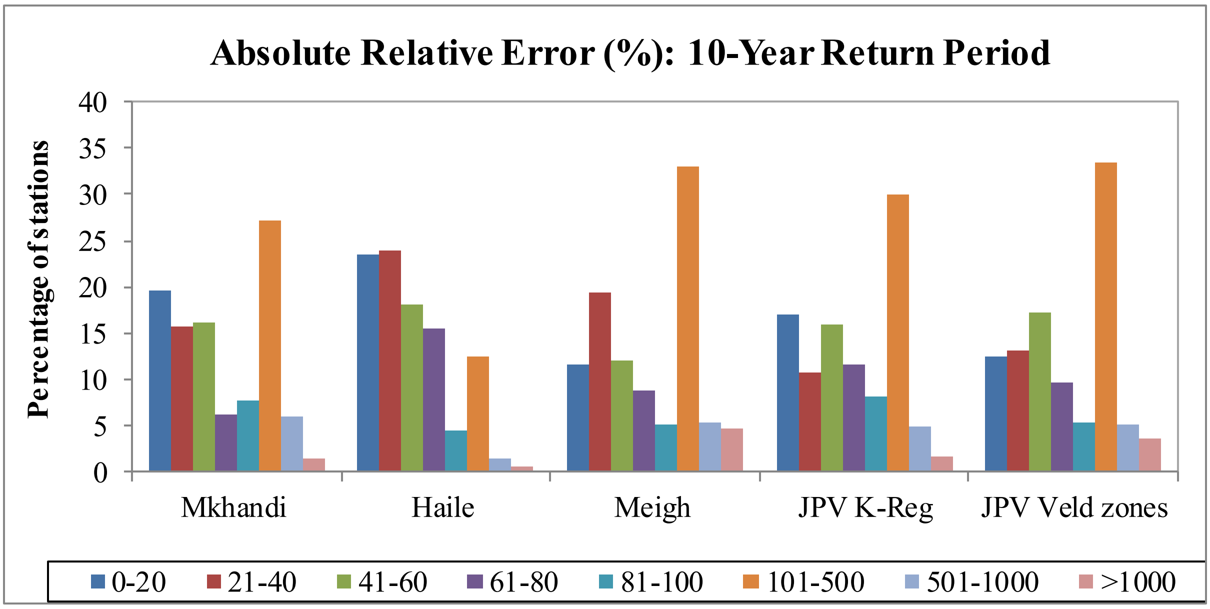

Figure E2 Performance of RFFA, based on the frequency of absolute relative errors for the 10-year Return Period

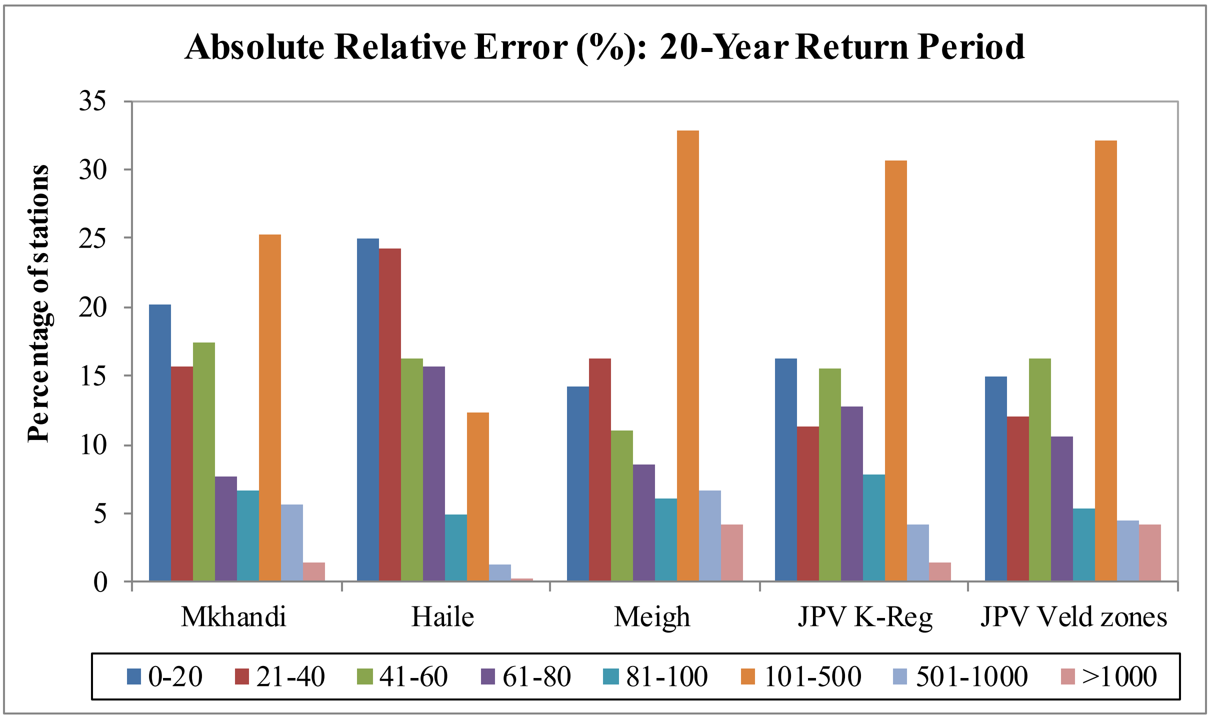

Figure E3 Performance of RFFA, based on the frequency of absolute relative errors for the 20-year Return Period

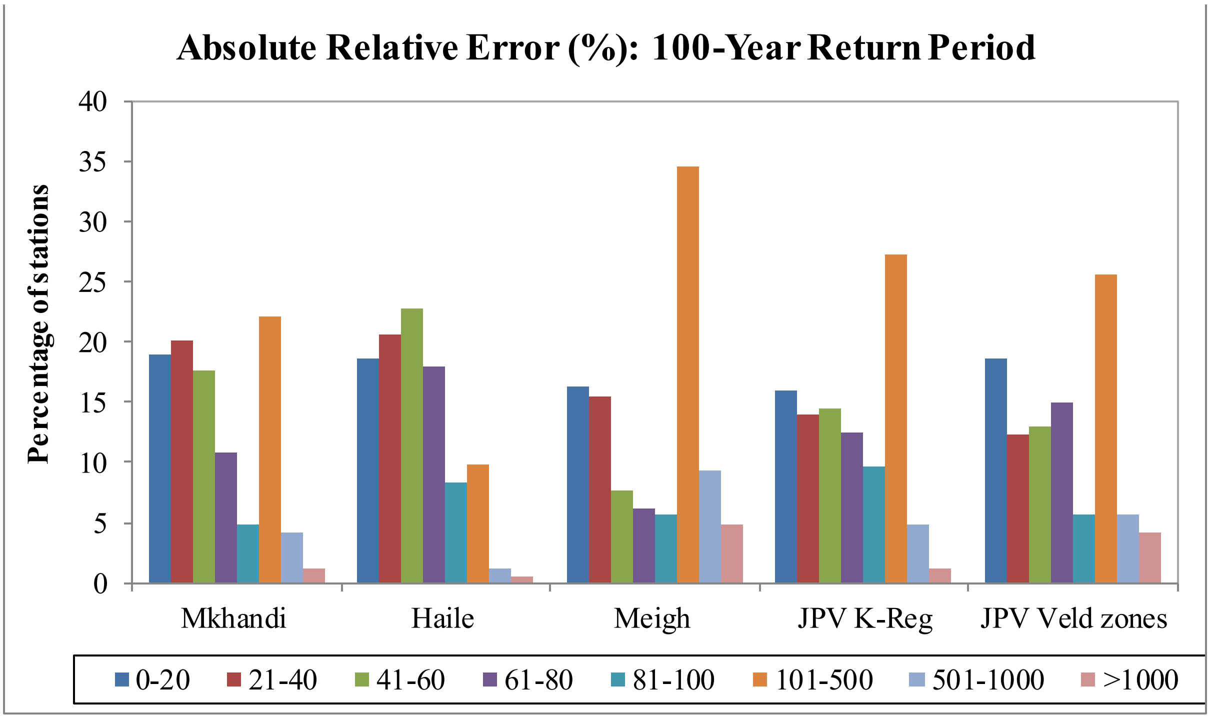

Figure E4 Performance of RFFA, based on the frequency of absolute relative errors for the 100-year Return Period

APPENDIX F: METHOD RANK

Appendix F provides a series of maps, illustrating the rank of each RFFA method.

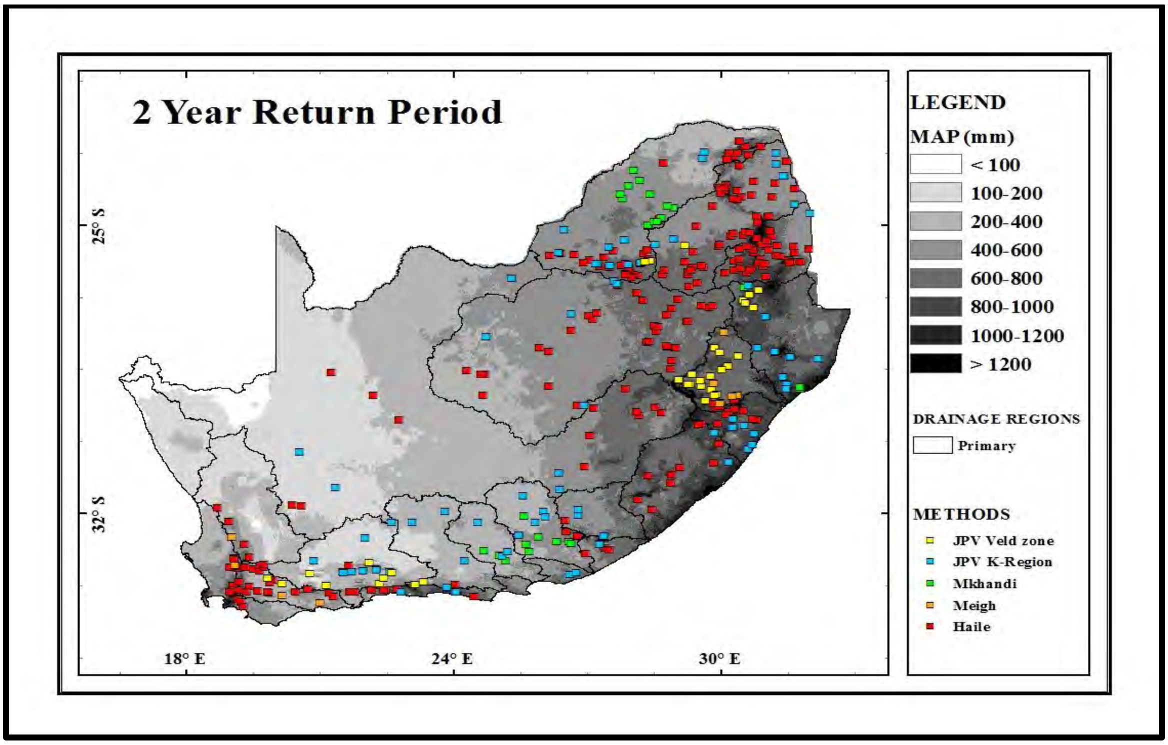

Figure F1 Method rank for the 2-year Return Period

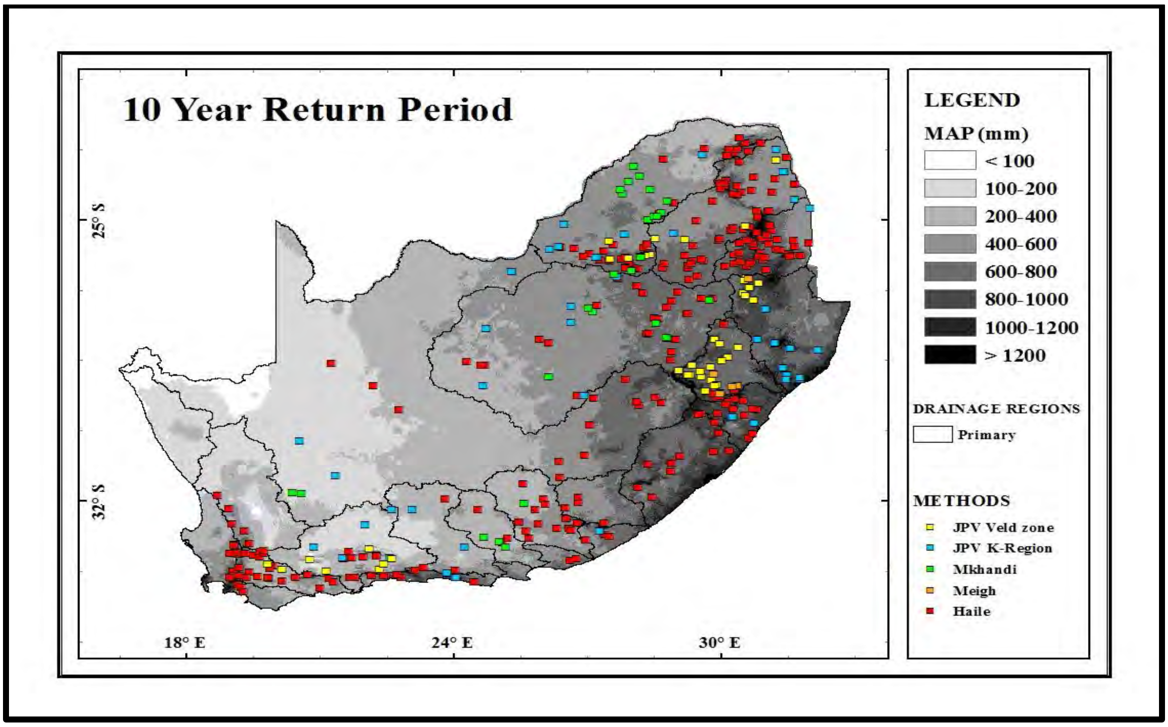

Figure F2 Method rank for the 10-year Return Period

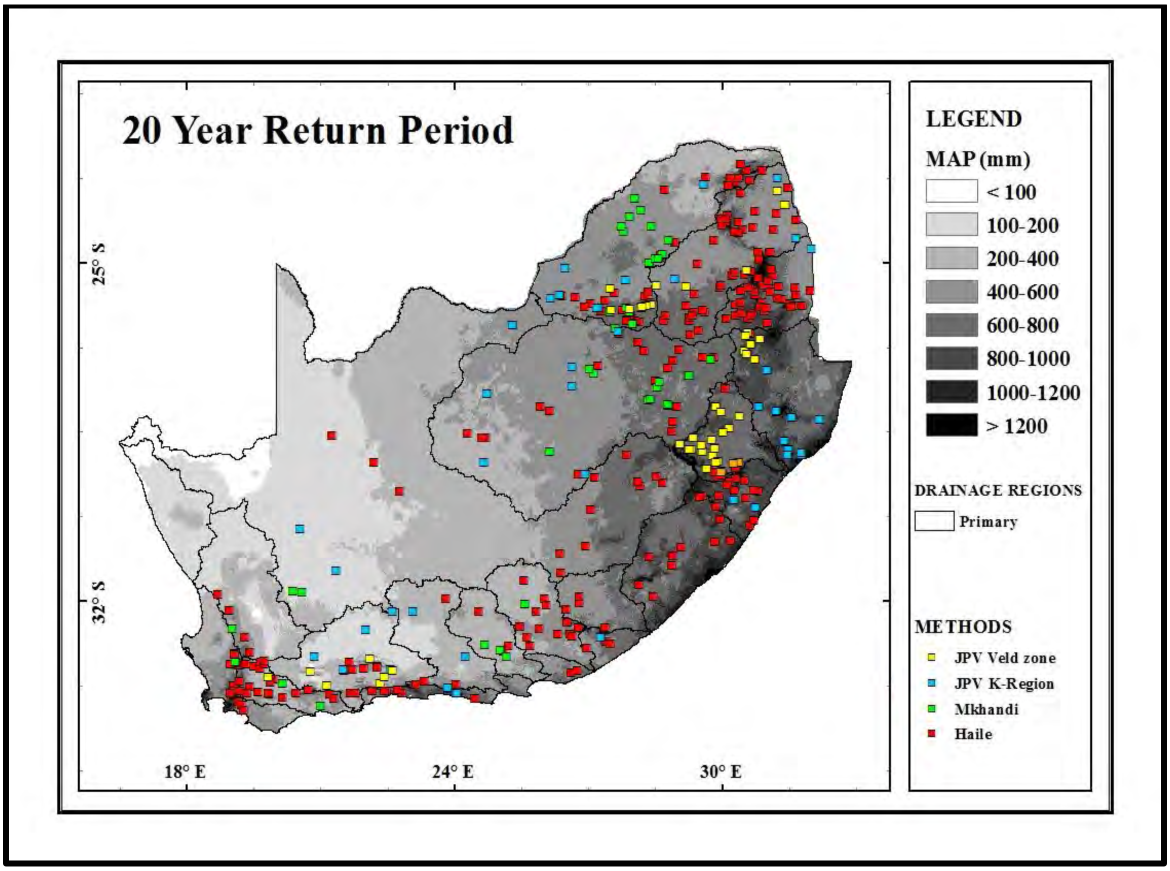

Figure F3 Method rank for the 20-year Return Period

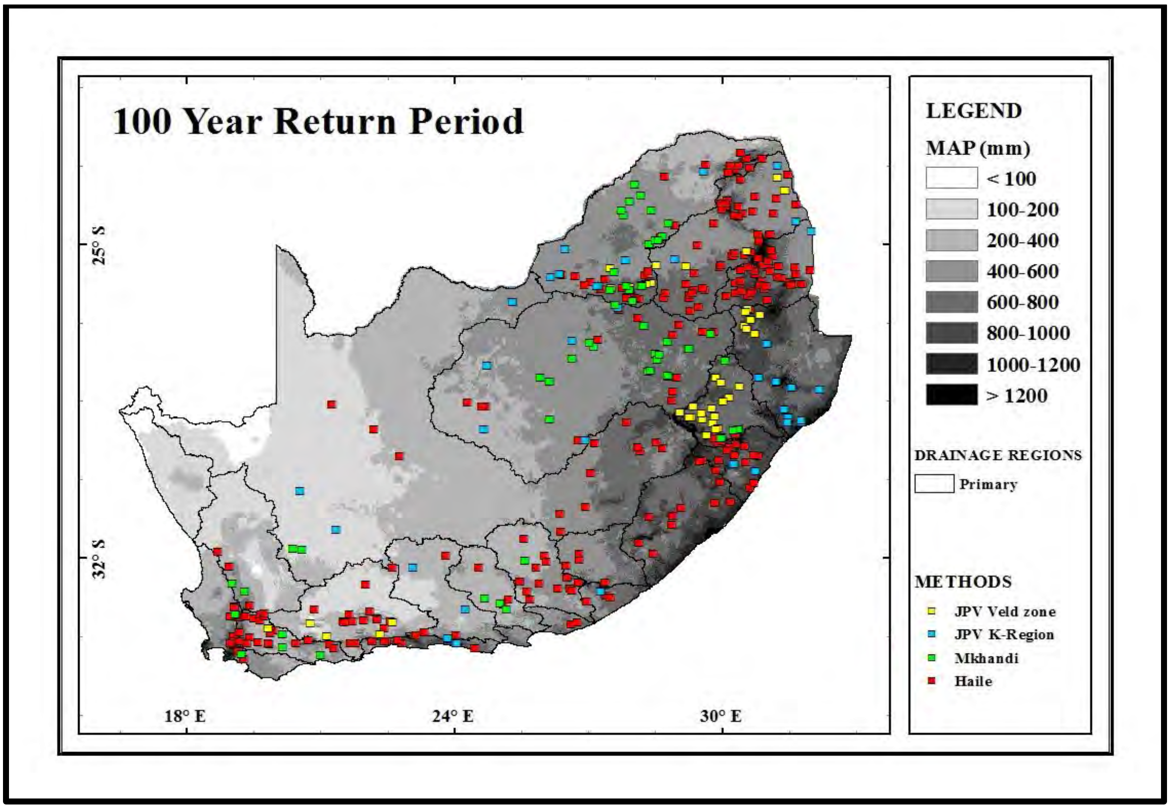

Figure F4 Method rank for the 100-year Return Period