A GUIDE TO FLOOD FREQUENCY ESTIMATION METHODS

AS APPLIED IN THE DEPARTMENT OF WATER AND SANITATION

Department Water and Sanitation, South Africa
Pretoria

April 2023

Tenth Edition

DEPARTMENT: WATER AND SANITATION

DOCUMENT DETAILS

Title:	**A Guide to Flood Frequency Estimation Methods, as applied in the DWS**		
Previous editions:	**First, Mar 2004**		
	Second, Sep 2006	-	partly revised
	Third, Jul 2008	-	revised, separate chapter and updated map re RMF
	Fourth, Apr 2010	-	partly revised
	Fifth, Jul 2012	-	partly revised
	Sixth, Jun 2014	-	revised and added new rainfall methodologies
	Seventh, Jul 2016	-	revised and added FS's approach to RMF and SEF
	Eight, Jul 2018	-	revised and HRU maps improved in GIS
	Ninth, Mar 2021	-	revised and improved figures
Report status:	**Tenth edition, Apr 2023**	-	revised

ENDORSEMENT

Authors

P Rademeyer PrSciNat
Scientific Manager
Directorate: National Hydrological Services

E Oakes PrSciNat
Scientific Manager
Directorate: Infrastructure Capacity Analysis

D van der Spuy PrEng
Consultant

Preface

The first and most important piece of information is that, inevitably, different flood frequency estimation methods will give different results and there is no single calculation method that can be presumed to be better than any of the other methods.

Consequently you, the user, will have to apply your own experience and knowledge to decide whether a method is applicable for your situation. This places a responsibility on the user to understand the strengths and limitations of each of the available methods, as well as their applicability to the site being studied.

This publication is not intended to be a handbook or manual of any sort, nor is it an attempt to prescribe any method over the other. Its purpose is merely to provide the reader with basic information on the principal methods of flood frequency analysis, currently in use by the Department of Water and Sanitation (DWS), South Africa. The reader is strongly advised to study the relevant documents and handbooks, referred to under sources and the bibliography, as well as any other publications, to understand the full extent of the complexity involved in a flood frequency analysis.

Sources

The principle sources used in the preparation of these notes were a series of lecture notes produced for courses on flood hydrology that were presented jointly by the DWS and the University of Pretoria during the period 1976 to 1985, the handbook Flood Hydrology for Southern Africa by Alexander (1990), various publications by the Hydrological Research Unit (HRU) of the University of the Witwatersrand (between 1969 and 1974), documentation on the SCS method from the University of Natal, various unpublished work/documents of the DWS Flood Studies Unit, as well as input from many years of collective practical experience.

Acknowledgements

We would like to express our sincere appreciation to all our colleagues in the DWS, present and past, who directly or indirectly contributed to the compilation of these notes. We would also like to convey our gratitude to the DWS for permission to use these notes in lectures.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS, INITIALISMS, ABBREVIATIONS AND SYMBOLS

A-	-	Total catchment area (km^2)
A_e	-	Effective catchment area (km^2)
AEP	-	Annual Exceedance Probability
C_S	-	Value dependent on catchment steepness (Rational method)
C_P	-	Value dependent on soil permeability (Rational method)
C_V	-	Value dependent on vegetal cover (Rational method)
CAPA	-	Catchment parameter method
CDF	-	Cumulative distribution function
COV	-	Coefficient of variation
DCR	-	Daily catchment-rainfall approach
DEM	-	Digital Elevation Model
DRH	-	Direct Runoff Hydrograph method (Bauer and Midgley, 1974)
DWS	-	Department of Water and Sanitation
EV	-	Extreme value
FEH	-	Flood Estimation Handbook
K	-	RMF region
K_e	-	Effective Francou-Rodier value
$K_{0.01\%}$	-	K-value, associated with estimated 0.01% AEP (10 000 year) flood peak
K_{SEF}	-	Site-specific K-value, associated with Safety Evaluation Flood
K_{Site}	-	Site-specific K-value, associated with Expected Maximum Flood
kurt	-	kurtosis
L	-	Length of longest watercourse (km)
L_c	-	Centre of gravity of catchment length (km)
LN	-	Log normal frequency distribution
LP3	-	Log Pearson III frequency distribution
GEV	-	General Extreme Value distribution
GEV_{MM}	-	General Extreme Value using Mean Moments
GEV_{PWM}	-	General Extreme Value using Probable Weighted Moments

GIS	-	Geographical Information System
GLO	-	Generalised Logistic distribution
HRU	-	Hydrological Research Unit (of the University of the Witwatersrand)
HRU1/71	-	Empirical method devised by Pitman and Midgley (1971).
MAP	-	Mean annual precipitation (mm)
MIPI	-	Empirical method devised by Pitman and Midgley (1967).
MSR_{CS}	-	Maximum station-rainfall approach using catchment statistics
MSR_{SS}	-	Maximum station-rainfall approach using station statistics
NFSP	-	National Flood Studies Programme
PILFs	-	Potentially Influential Low Flows
PP	-	Plotting Position
Q_T	-	Flood peak for return period T year (m^3/s)
Q_P	-	Flood peak for probability of exceedance of P
PDF	-	Probability density function
PMP	-	Probable maximum precipitation
RMF	-	Regional maximum flood (m^3/s) for t_C (Kovács, 1988)
SANRAL	-	South African National Road Agency
SEF	-	Safety Evaluation Flood
S_A	-	Mean catchment steepness (m/m)
S_L	-	Average slope of longest watercourse (m/m)
SCS	-	Soil Conservation Service method for small catchments
SSR	-	Smithers-Schulze regional-rainfall approach
SUH	-	Synthetic Unit Hydrograph method (Midgley, 1972)
T	-	Return period in years
t_C	-	Time of concentration (h)

1. INTRODUCTION

In 1919 the Department of Water and Sanitation (then the Department of Irrigation) issued the first paper *Maximum flood curves* in its Professional Paper series. In his foreword the then Director of Irrigation, AD Lewis, wrote:

> "*Too much importance must not be attached to the formula. No formula is likely to be discovered which will apply to all drainage areas. The maximum flood depends on too many circumstances, such as intensity of rainfall, size and shape of catchment and channel, and permeability of ground surface.*"

Now, more than 100 years later, Lewis' statement is still valid. Despite the collection of a vast amount of meteorological and hydrological data in South Africa, as well as elsewhere in the world, there is still no universally applicable method for flood frequency determination. The engineer or hydrologist who undertakes these assessments must still use judgement and experience in interpreting the results of the estimations.

The magnitude of a flood is principally a function of the fixed characteristics of the catchment (mainly its size); the properties which may vary from flood to flood (mainly those of storm rainfall); and the moisture status of the catchment prior to the storm (mainly soil moisture status and river flow).

Many methodologies and models were developed over the years to try and simulate the processes that convert rainfall into flood runoff – some of these methods are part of the bouquet of methods used by the DWS, which can be classified, generally, under three principal flood peak frequency estimation approaches:

- statistical methods, where flood peak frequencies are estimated through statistical analyses of observed annual flood peak data;
- deterministic methods, where storm rainfall is used in conceptual rainfall-runoff models. Flood peak frequencies are estimated by assuming that the statistical properties of a flood are the same as that of the storm rainfall causing it; and
- empirical methods (based, per definition, on observation or experience rather than theory or pure logic), where flood peak frequencies are estimated by using mathematical models, developed through analyses of available flow data.

All the methods make use of, or are developed from, available data. In South Africa rainfall data can be obtained from the South African Weather Services (SAWS) and other hydrological data from the Department of Water and Sanitation (DWS).

South Africa has a very wide range of climatological, and consequently hydrological, conditions. Many papers had been published about Flood Hydrology throughout the years. Probably the most comprehensive study on the subject in South Africa, up to date, was done by the Hydrological Research Unit (HRU) of the University of the Witwatersrand, at the request of the South African Institution of Civil Engineers, and their first report (HRU 4/69) was published in 1969. Some of the methods proposed in this report, as well as in some of the follow up reports (like HRU 1/72), were developed a few years before the first report was published (e.g. the MIPI method, developed in 1967).

The need arose to update most of these methods, which has become necessary for an impartial assessment of the various methods of flood frequency analyses, used in this country. More than 50 years of additional data are available that can be utilised in improving the methods. The National Flood Studies Programme (NFSP) was established, and various research studies have been completed and are currently underway to improve these methods.

In the next few chapters, the reader will be introduced to a very concise and basic description of the methods currently in use in the DWS. All three principal flood frequency estimation approaches will be addressed.

To conclude, DF Roberts, when he was asked how he would describe a Hydrologist, commented as follows:

> *A Hydrologist is a Scientist who is capable of producing an exact answer from a mass of unreliable basic data, using dubious statistical methods based on guesswork.*

2. CATCHMENT CHARACTERISTICS

2.1 Effective catchment Area, A_e

The catchment area was shown to be one of the most important catchment characteristics (Raudkivi, 1979). The volume of water to be expected at the point of interest is directly related to the area covered by rain and the rainfall thereon. In small catchments, for example, the relationship between rainfall intensity and infiltration rate of the soil is very important, whereas in large catchments the quantity of rainfall relative to the water storage capacity of the ground is more important.

The effective catchment area is that part of the total catchment area that will contribute to the peak runoff. Ineffective areas are most often surface depressions (i.e. pans) and areas separated by artificial or geological barriers.

The catchment area can be quite accurately determined by using Geographical Information Systems (GIS) with the correct projection.

Alternatively, a planimeter can be used to estimate the catchment area from 1:50 000 or 1:250 000 (larger catchments) topographical maps.

2.2 Mean catchment slope, S_A

The catchment slope is an important characteristic, which is one of the characteristics that determine the catchment response in the case of runoff (Shaw, 1988).

The catchment slope can be quite accurately determined by using GIS, provided that a hydrologically correct Digital Elevation Model (DEM) is used.

Alternatively, 1:50 000 or 1:250 000 (larger catchments) topographical maps can be used. The catchment slope is calculated by superimposing a grid of at least 50 squares over the catchment. The horizontal distance between the contour intervals is then measured for each grid point (Figure 2-1). In the grid method the catchment slope is defined as the average slope perpendicular between the nearest contour lines, through each grid-point.

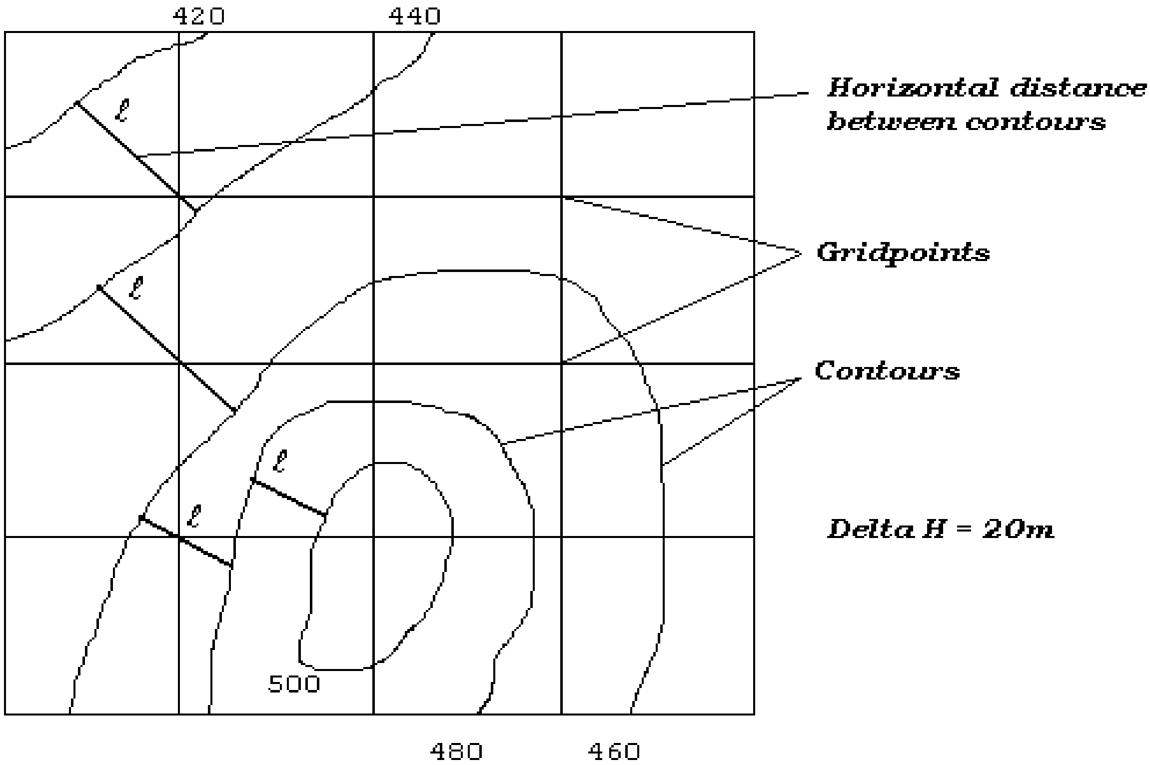


Figure 2-1: Calculating the catchment slope using the grid method

Using the grid method, the catchment slope can be determined as follows:

$$\bar{\ell} = \sum_{i=1}^{n} \frac{\ell_i}{n} \qquad \text{and} \qquad S_A = \frac{\Delta H}{\bar{\ell}}$$

where:

ℓ_i	-	horizontal distance between consecutive contours
n	-	number of grid points
ΔH	-	contour interval (m)

General guideline: The minimum number of grid points should be 20 in catchments where A < 10 km^2 and 50 where A > 10 km^2.

<u>Note:</u> this guideline is not necessarily applicable in all cases. For example, where contour intervals are far apart, in small relatively flat catchments, the analysts should apply their minds to deal with the problem.

2.3 Longest watercourse, L

The route, which will be followed by a water particle, from a point on the catchment boundary taking the longest time to reach the catchment outlet, is defined as the longest watercourse (L). This distance consists of the natural channel (L_1) and overland flow (undefined channel, L_2). L_2 is the distance between the upstream end of the natural channel and the catchment boundary.

This distance can be measured quite accurately using GIS. Noteworthy, is that the centreline of the longest stream should always be followed/delineated using high resolution images in GIS.

Alternatively, L can be measured on 1:50 000 topographical maps. Where a divider is used to determine L, the divider should be set to 0.25 km (5 mm).

2.4 Centre of gravity of catchment length, L_C

It is the distance from a point on the longest watercourse (closest to the centre of gravity of the catchment) along the watercourse to the outlet of the catchment.

The centre of gravity of a catchment and L_C can be measured accurately using GIS.

Alternatively, it can be measured on 1:50 000 topographical maps.

2.5 Mean river slope, S_L

The three most appropriate methods to estimate river slope are discussed below.

<u>10 @ 85 method:</u>

This method was developed by the US Geological Survey and is given by the formula:

$$S_L = \frac{H_{0.85L} - H_{0.10L}}{0.75L}$$

where:

L	-	length of longest watercourse (m)
$H_{0.10L}$	-	elevation height at 10% of L (m)
$H_{0.85L}$	-	elevation height at 85% of L (m)

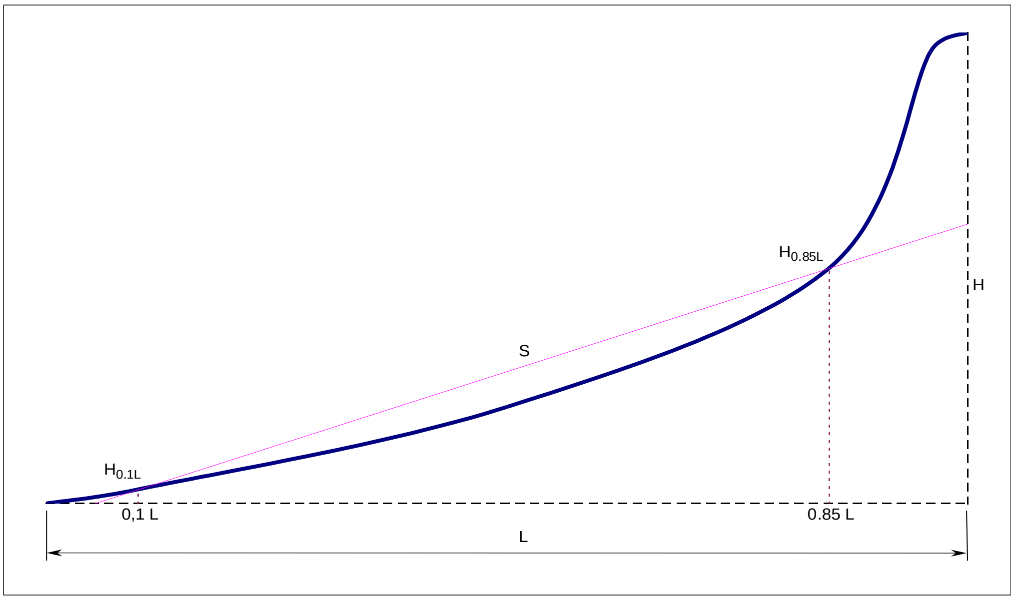


Figure 2-2: 10 @ 85 river slope calculation

Equal area method:

From Figure 2-3: The mean slope of the longest watercourse is determined as the slope of the line drawn such that A1 = A2.

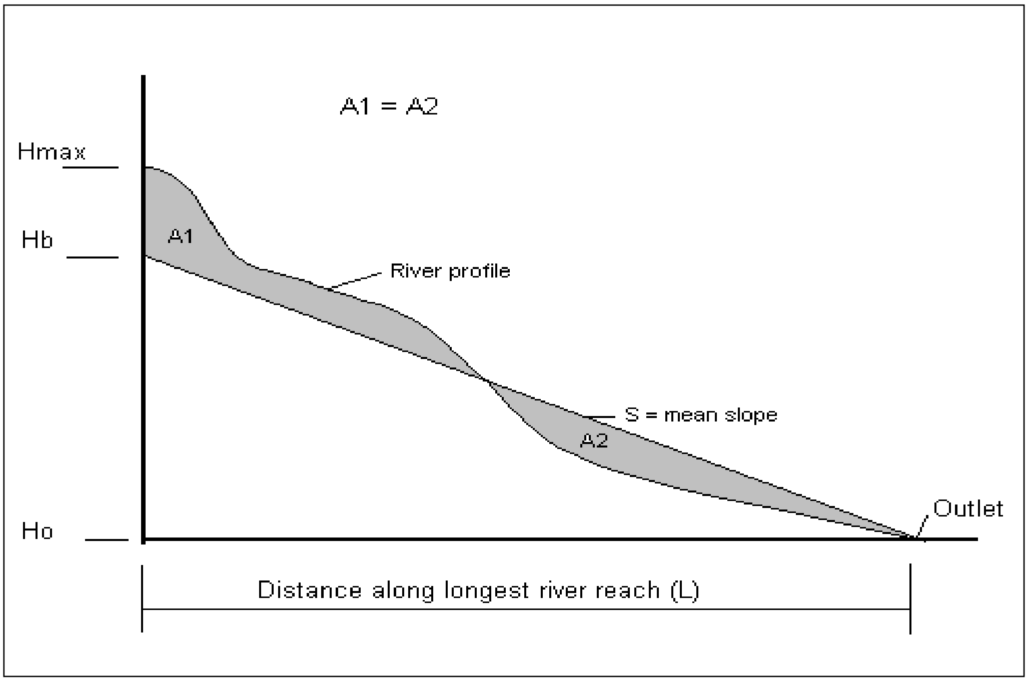


Figure 2-3: Equal area river slope

The mean river slope is given by:

$$S_L = \frac{H_b - H_0}{L}$$

where:

L	-	length of longest watercourse (m)
H_b	-	$H_0 + 2\sum A_i/L$
A_i	-	$[(h_i + h_{i+1})/2 - H_0] \times \ell_i$
h_i	-	height value of contour line (m)
ℓ_i	-	distance along river between two consecutive contours (m)

<u>Taylor-Schwarz method:</u>

The Taylor-Schwarz method (Flood Studies Report, 1975) is considered scientifically more correct than the other methods and is the method currently preferred by DWS.

The method divides the river profile into sub-reaches and uses the fact that the velocity in each reach is related, in the basic flow equations, to the square root of the slope. The index is equivalent to the slope of a uniform channel having the same length as the longest watercourse and an equal time of travel.

$$S_L = \left[\frac{L}{\sum_{i=0}^{n} \frac{\ell_i}{\sqrt{S_i}}}\right]^2$$

where:

L	-	length of longest watercourse (m)
ℓ_i	-	distance along river between two consecutive contours (m)
S_i	-	slope between two consecutive contours (m/m)

2.6 Soils and Land Surface Cover

Soil types:

For rainfall-runoff relations it is important to identify different soil types. The soil types are categorised into four-basic hydrological soil groups namely:

Group A: Low runoff potential (Very permeable)

- Infiltration rate is high
- Soils are deep (well drained)
- Texture is coarse (Gravel and coarse sand)

Dolomite areas: The knowledge of dolomite areas is of great importance for the deterministic and empirical methods. It is important to note that dolomite areas may absorb as much as 90% of the runoff for underground storage. Dolomite areas should always be classified as very permeable.

Group B: Moderately low runoff potential (Permeable)

- Infiltration rate is medium
- Medium effective soil depth (sandy, sandy loam)

Group C: Moderately high runoff potential (Semi-permeable)

- Infiltration rate is slow
- Depth is shallow
- Texture is fine (silt, loam, and clayey sand)

Group D: High runoff potential (Impermeable)

- Very low infiltration rates
- Very shallow and/or expansive soils (clay, peat, and rock)

In the most recent analyses at the DWS the soil-type classification according to the SCS approach (Schulze and Arnold, 1979) has been used. The dataset is available in GIS format. More information can also be obtained from the Agricultural Research Centre.

Land Surface Cover:

Vegetal cover causes water retention in the catchment and has an influence on the runoff coefficients used in the deterministic methods.

A good knowledge of the vegetal cover in the catchment is essential in the process of calculating the runoff coefficients. The vegetation can be categorised as follows:

- Category A: Forest, dense bush, and wood
- Category B: Thin bush and cultivated land
- Category C: Grassland
- Category D: Bare surface (no vegetation)

Forest plantation can always be classified as Category A, independently of whether the area is covered by fully grown forest or is cleared.

Dense bush and thin bush are usually marked by the same symbol on the 1:50 000 maps and should be separated on the basis of MAP (Table 2-1) or by consulting people who know the area.

Table 2-1: Separation of dense bush and thin bush on the basis of MAP

MAP (mm)	**Dense Bush % Area**	**Thin Bush % Area**
< 600	-	100
600 - 900	Photographs or consult people who know the area.	
> 900	100	-

Cultivated areas can be estimated from 1:50 000 maps or other related sources.

Grassland and bare surface have no symbols on the 1:50 000 maps and should be estimated on the basis of MAP (Table 2-2) or by consulting people who know the area.

Table 2-2: Separation of grassland and bare surface on the basis of MAP

MAP (mm)	**Grassland % Area**	**Bare Surface % Area**
< 400	-	100
400 - 600	33	67
600 -900	67	33
> 900	$100 - A_{>50\%}$	$A_{>50\%}$

$A_{>50\%}$ = % of A where the slope is 50% or steeper.

In addition, the GIS-based 2020 South African National Land-Cover Dataset (Geoterra Image (Pty) Ltd, South Africa, 2020) has been utilised to provide a more relevant estimate of the present land surface cover.

2.7 Time of concentration (t_c)

The time that a water particle requires to travel from the furthest point in the catchment to the outlet, is known as the time of concentration (t_c). In case of extreme events, it is assumed that the storm duration is similar to the t_c. It can consist of natural and overland flow components.

Natural channel:

The US Bureau of Reclamation Equation was suggested by the University of the Witwatersrand (Midgley, 1972). The equation is more applicable to rural areas and is also currently being used by DWS to calculate t_c for channel flow.

$$t_c = \tau \left[\frac{0.87 \times L_1^2}{1000 \times S_L}\right]^{0.385}$$

where:

t_c	-	time of concentration (h)
τ	-	correction factor
L_1	-	length of natural channel (km)
S_L	-	mean channel slope (m/m)

Experience has shown that the above equation may overestimate or underestimate t_c. A set of correction factors (τ) has been developed by Kovacs (unpublished) to overcome this problem.

Table 2-3: Correction factors for t_c

***A* (km²)**	τ
< 1	2
1 - 100	2 - 0.5 log *A*
100 - 5 000	1
5 000 - 100 000	2.42 - 0.385 log *A*
> 100 000	0.5

Storm duration is usually assumed as multiples of t_c. In this regard it is suggested, for ease of computation, to round off t_c as follows:

Table 2-4: Suggested rounding of t_c

t_c	Round to:
$10\ h \leq t_c$	Nearest even hour
$5\ h \leq t_c < 10\ h$	Nearest integer hour
$1h \leq t_c < 5\ h$	Nearest half hour
$t_c < 1h$	Nearest decimal hour

Overland flow:

This usually occurs in small, flat catchments or in the upper reaches of catchments where there is no clearly defined watercourse.

The Kerby equation (Drainage Manual, Pretoria 2006) is recommended for the calculation of t_c in this case. It is applicable to parts where the slope is fairly even:

$$t_c = 0.604\left[\frac{r \times L_2}{S_L^{0.5}}\right]^{0.467}$$

where:

t_c	-	time of concentration (h)
r	-	roughness coefficient (Table 2-5)
L_2	-	length of overland flow (km)
S_L	-	mean overland slope (m/m)

Table 2-5: Recommended values of r

Surface description	r
Paved areas	0.02
Clean compacted soil, no stones	0.10
Sparse grass over fairly rough surface	0.30
Medium grass cover	0.40
Thick grass cover	0.80

3. PRECIPITATION

3.1 Introduction

The depth, areal spread and duration of the rainfall, as well as the variations in intensity in space and time over the catchment primarily determines the severity of the flood.

There are four major processes that determine the type of precipitation.

i. Orographic lifting: Where a mountain range intercepts moist air flow and forces it to ascend. As it cools it sheds some or all its moisture as rain or snow. Typically, orographic rain will therefore be of low intensity but long duration, and is controlled by the local topography.

ii. Convection: Differential heating or advection can lead to the air becoming locally more buoyant. The air mass may then rise to levels were it becomes saturated, forms clouds and precipitates. Thunderstorms are examples of the convective process. The intensity of the precipitation will depend on the rate of cooling, which is a function of the vertical velocity which, in turn, is largely determined by the temperature difference between the rising and ambient air. The duration is normally very short.

iii. Low pressure: In the summer months a low-pressure system situated over South Africa may drag moist warm air from the north to give rise to heavy prolonged rainfall over a wide area. Tropical cyclones on the other hand are a low pressure system situated over the ocean. If moist warm air is available, they will be self-sustaining as they move along. They quickly subside once they move over land. They can cause prolonged high intensity rainfall.

iv. Fronts: A cold front causes cold air to move below warm air and displaces it upwards. The interface is a steep wedge having a backward slope. If the warm air is unstable with a high moisture content violent storms with a high intensity and a short duration can result. A warm front on the other hand is warm air that displaces cold air. Less intense rainfall is experienced over a longer duration.

3.2 Calculating storm rainfall

After the catchment has been identified, a plot of all the rainfall gauging stations scattered throughout the catchment has to be made. Each station needs to be investigated to determine the record length, data quality (missing, unreliable, etc.) and topographical position.

Due to a probabilistic analysis that needs to be done on each rainfall gauging station it is suggested not to use any station with a record length shorter than 30 years. Most rainfall gauging stations will give reliable results after 30 years of data.

When selecting a rainfall gauging station just outside the catchment, extreme caution should be applied, due to the topographical position of that station. The rainfall on either side of a mountain range could differ considerably even if the distance does not seem to be far.

<u>Converting daily rainfall to critical storm duration (t_c) rainfall:</u>

We will assume that the storm duration is equal to t_c, this will produce the peak flow for the catchment. Depending on the t_c, we either increase or decrease the daily storm depth to be used in the Deterministic methods.

Firstly we will look at decreasing the rain depth.

To convert daily rainfall depth to 24-h rainfall depth apply the following formula:

$$\boldsymbol{24\text{–}\,h\ depth\ (mm) = daily\ depth\ (mm) \times 1.11}$$

We use the computed ratios for D (h) duration storm depth to that of 24 h for summer rainfall regions (R1) and winter rainfall regions (R2) taken from TR102 (Adamson, 1981). Refer to Table 3-1.

Table 3-1: Ratios of storm depths for durations less than 24-h

D (h)	R1	R2
0.10	0.17	0.14
0.25	0.32	0.23
0.50	0.46	0.32
1.00	0.60	0.41
2.00	0.72	0.53
3.00	0.78	0.60
4.00	0.82	0.67
5.00	0.84	0.71
6.00	0.87	0.75
8.00	0.90	0.81
10.00	0.92	0.85
12.00	0.94	0.89
18.00	0.98	0.96
24.00	1.00	1.00

Converting daily rainfall depths to durations longer than one day, simply implies converting the 1-day rainfall depth to a 24-h depth, a 2-day depth to a 48-h depth, etc. and interpolating between 24, 48, 72, etc.-h rainfall depths.

Table 3-2: Converting daily rainfalls to hourly rainfalls

Duration		Multiply with
From	To	
1-day	24-h	1.11
2-day	48-h	1.07
3-day	72-h	1.05
4-day	96-h	1.04
5-day	120-h	1.03
7-day	168-h	1.02
> 7-day	> 168-h	1.00

Probabilistic analysis of annual storm depths (mm) for each station:

The annual storm depths (mm) for each station must now be probabilistically analysed to obtain storm depths for all relevant probabilities of exceedance.

The Log Normal (LN), Log Pearson type III (LP3), General Extreme Value using Mean Moments (GEV_{MM}) and General Extreme Value using Probable Weighted Moments (GEV_{PWM}), can be used to determine the required depths (mm).

Figure 3-1 shows a graphical illustration of the various distributions used on the annual rainfall depths.

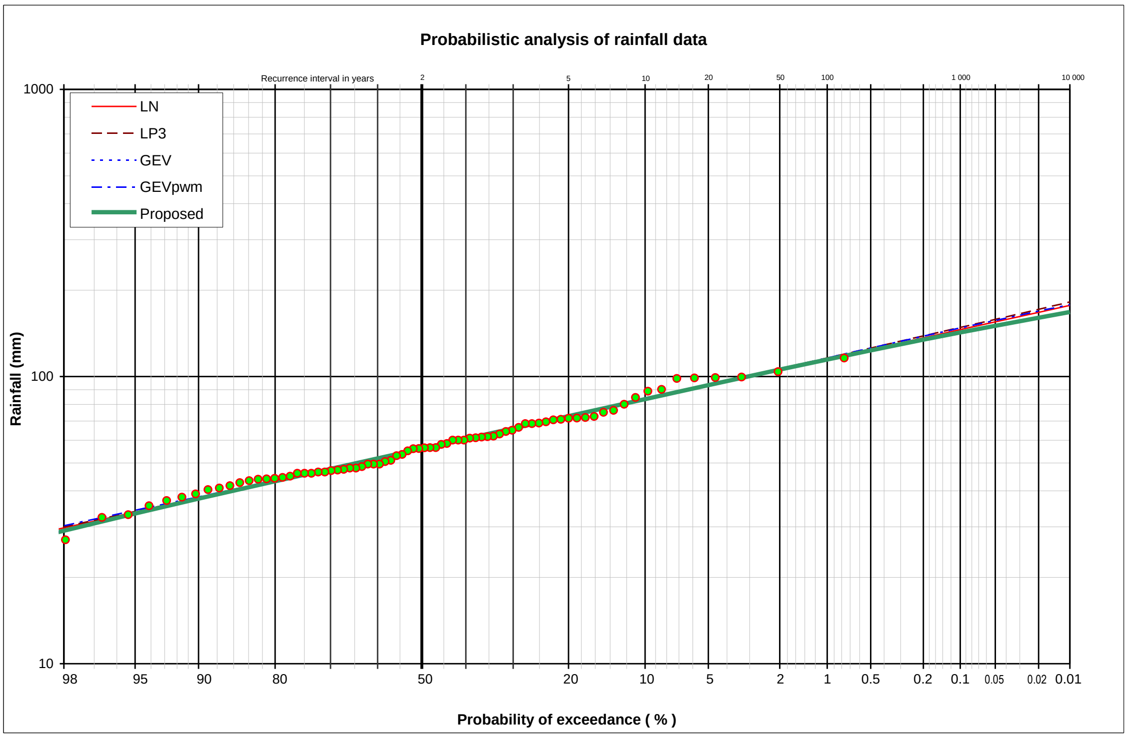


Figure 3-1: Rainfall depth frequency distributions

3.3 Depth - Area relationships

The storm rainfall may be estimated by averaging the rainfall measured at the gauge locations. For averaging of the rainfall, the Thiessen polygon method is convenient, particularly for repeated calculations (the method is available in GIS). The method is schematically illustrated in Figure 3-2:

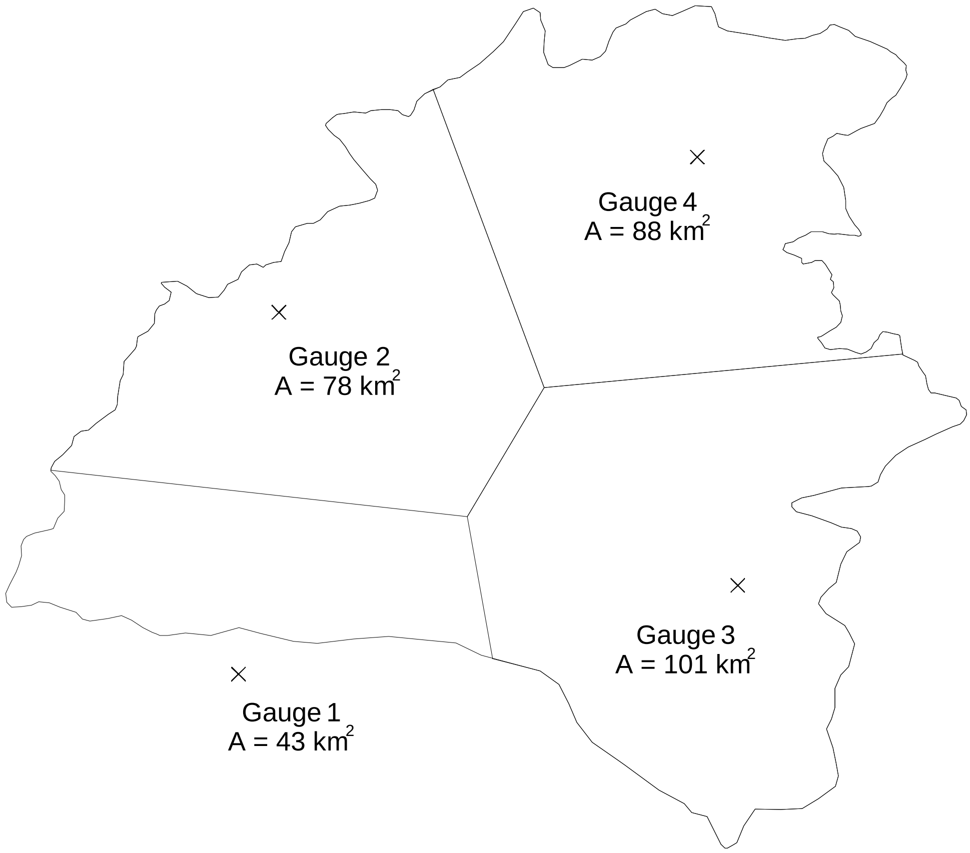


Figure 3-2: Thiessen Polygon Method

The bisectors of the lines connecting the gauges subdivide the catchment. The area allocated to each station is measured and expressed as a percentage of the total area. The percentages are then used as "weightings", e.g.:

Table 3-3: Storm rainfall, using Thiessen Polygons

Gauge	Area (km^2)	Areal weighting	Rainfall (mm)	Portion (mm)
1	43	0.14	72	9.99
2	78	0.25	73	18.37
3	101	0.33	83	27.04
4	88	0.28	85	24.13
Σ =	**310**	**1.00**	**Mean Rainfall =**	**79.53 mm**

This method assumes linear variation between stations.

The summation of the rainfall between the isohyets over the catchment is still one of the most widely used methods. An experienced analyst can make allowances for topographical effects in drawing isohyetal maps.

The most common problem in using these techniques is the availability of rain gauges in and around a catchment. For various reasons the rain gauge network in South Africa is generally declining.

There is thus a need for remote sensing of rainfall using radar and satellite imagery. These technically advanced, relatively recent innovations provide detailed information about rain-rates over large areas in sequences of images. They thus have the potential to provide detailed information (in space and time) of rain rates that can be used in many applications.

3.4 Mean Annual Precipitation (MAP)

The mean annual precipitation (MAP) for the catchment can be derived by using the same technique as used in determining the storm rainfall. The MAP for each station is multiplied by the relative contribution (areal weighting) of that station towards the total catchment area according to the Thiessen Polygon depth-area distribution.

The WR2005/WR2012 reports by the Water Research Commission (Middleton & Bailey) can also be used to determine the MAP.

3.5 Area reduction factors (ARF)

Precipitation from flood producing storms is almost never evenly spread in time and space over a catchment. For this reason, it is necessary to reduce the rainfall depth according to the catchment size (km^2) and storm duration (h).

Area reduction factors are used to estimate the average precipitation over a catchment and the DWS has found the relationship from the UK Flood Studies Report (1975) are more applicable in South Africa than those found in the HRU documentation (see Figure 3-3; also, Appendix Figure A-1).

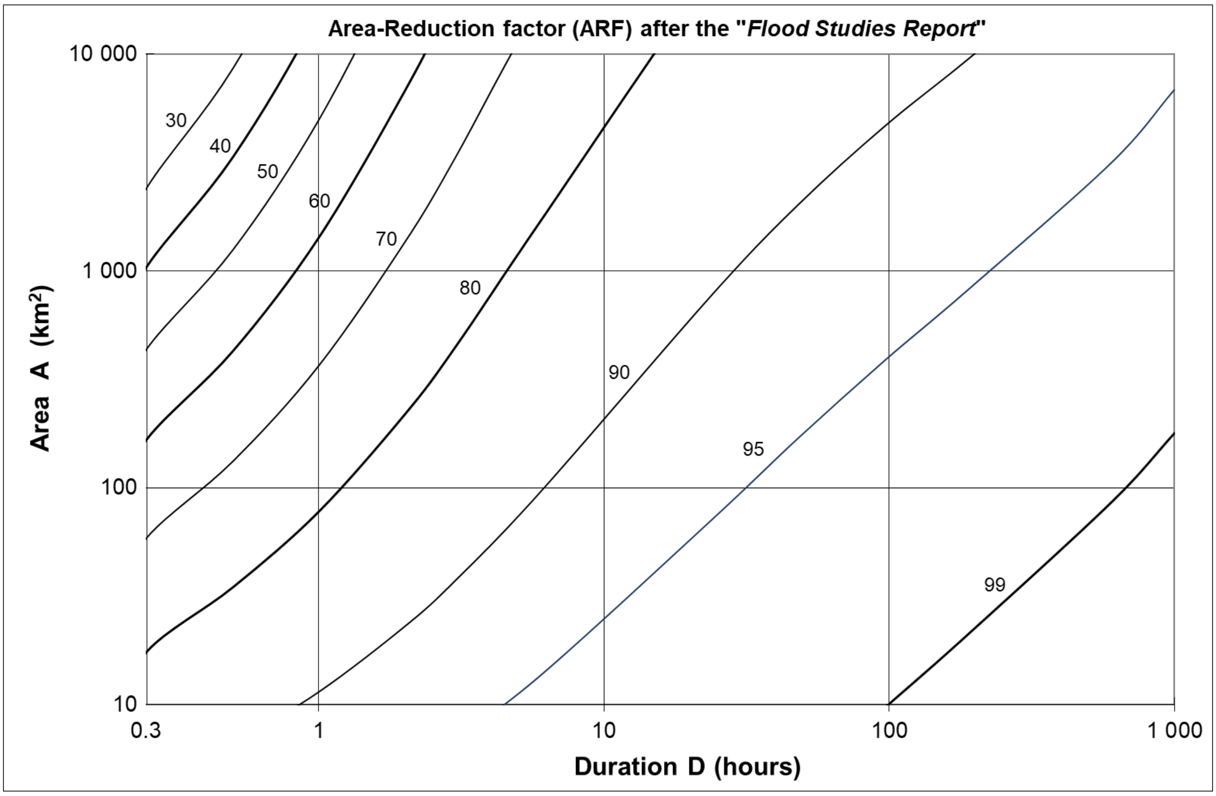


Figure 3-3: Area reduction factors

3.6 Probable Maximum Precipitation (PMP)

The term, probable maximum precipitation (PMP), is well established and is widely used to refer to the quantity of precipitation that approaches the physical upper limit of precipitation of a given duration over a particular catchment. The terms PMP and extreme rainfall have been used with roughly the same meaning. To ask, how possible or how probable such precipitation is, would be ,at best, a rhetorical question, because the definition of probable maximum is an operational one that is specified by the operations performed on the data. (Chapter 5.7, WMO, 2009).

In South Africa meteorologists have not concerned themselves with PMP estimation like most other countries. The only established PMP estimation procedure, for South Africa, is given in Report 1/72 of the University of the Witwatersrand's HRU (HRU 1/72). Envelope curves for regions experiencing similar extreme point rainfalls for South Africa were developed. These envelope curves, as well as the extreme rainfall regions, need revision. More than 30 years of additional data is now available.

An alternative approach could be to use the 0.001 % probability of exceedance rainfall depth as the PMP.

3.7 Runoff

Introduction:

When the rate of rainfall exceeds the interception requirements, and the rate of infiltration water starts to accumulate on the surface. At first the water accumulates into small depressions and hollows until the surface detention requirements are satisfied. After that water begins to move down the slopes as a thin film and tiny streams. This early stage of overland flow is greatly influenced by surface tension and friction forces. With continuing rainfall, the depth of surface detention and the rate of overland flow increase (Figure 3-4) but the paths of the small streams at the surface of the catchment are still tortuous and full of obstructions.

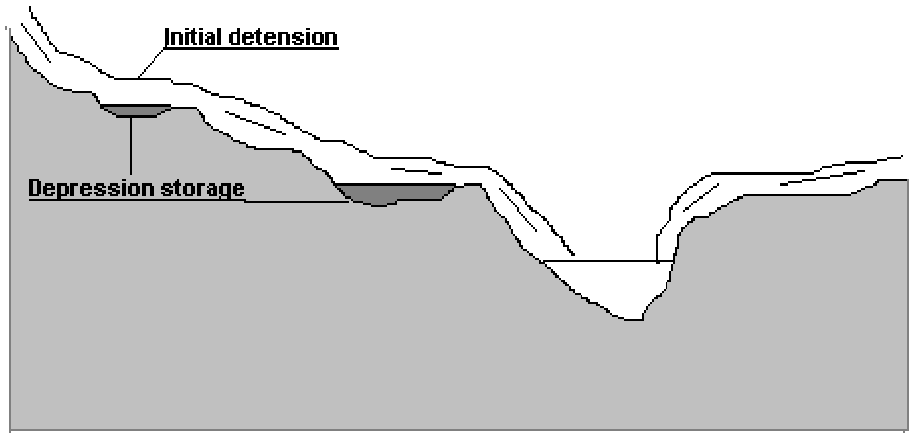


Figure 3-4: Surface Runoff

Every small obstruction causes a delay until the upstream water level has risen to overflow the obstacle or wash it away. All these small rivulets end up in the river that drains the whole catchment in question.

The runoff process contains three elements. These are (a) overland flow as a thin sheet of water, (b) small stream flow, and (c) river flow.

Physical characteristics of the catchment:

In a hydrologically large catchment the storage effects in water courses and lakes, etc. dominate and control the catchment's hydrological response. The storage effect causes the catchment to have a sluggish response to rainfall. A large catchment is not sensitive to variation of rainfall intensity and to land use. Most catchments large in size, with major rivers and lakes, fall into this category.

In a hydrologically small catchment overland flow, land use, slope, etc., have a strong influence on its hydrological response. Storage effects are small and the catchment is very sensitive to rainfall, that is, it responds quickly. Note that a small swampy catchment may behave hydrologically like a large catchment since the storage effects of the waterbodies are dominant.

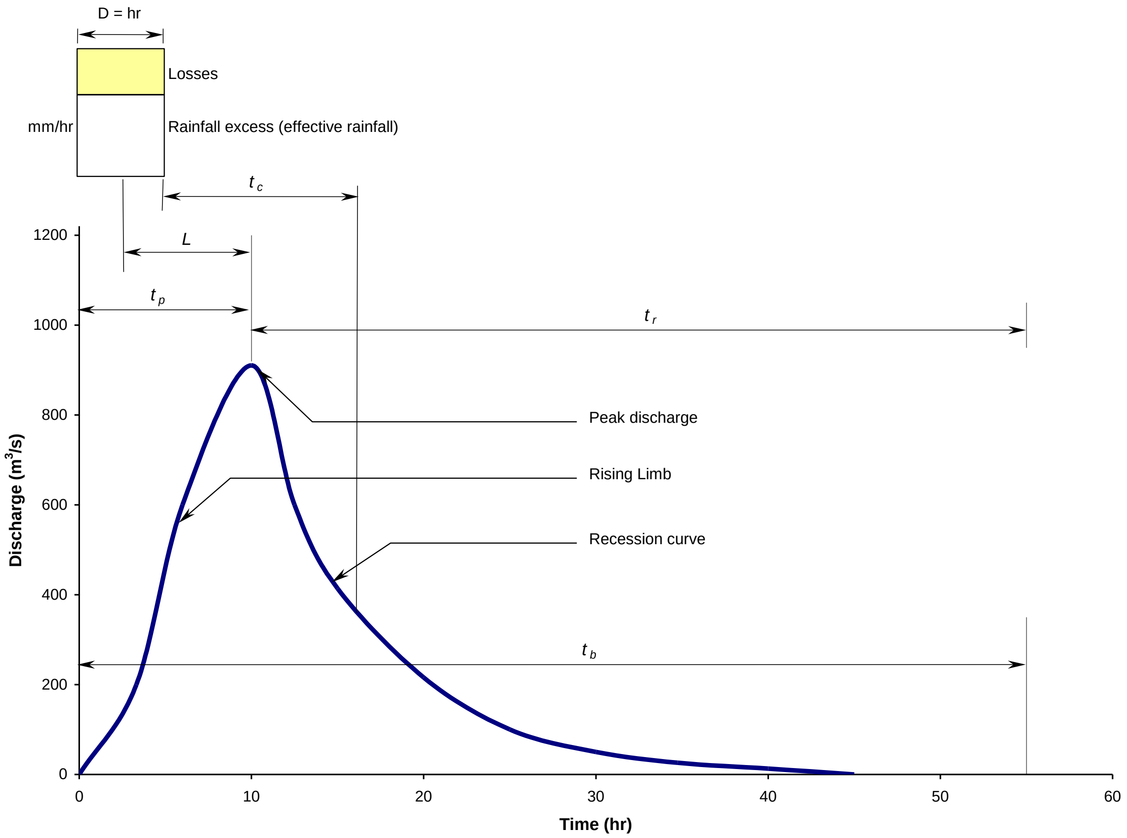
Climatic Factors:

The climatic factors which influence runoff are:

- Nature of the precipitation (rain, snow, sleet): the effect of a rainfall event is felt immediately but that of snow may be delayed for months.
- Rainfall intensity: only when the rainfall intensity exceeds the infiltration loss will any runoff occur.
- Duration of rainfall: the duration and the intensity of the rainfall are obviously the most important climatic factors. Their relationship with runoff will be discussed in more detail below.
- Areal distribution of rainfall: the areal distribution of the rainfall determines the shape of the hydrograph. High intensity rainfall near the outlet leads to a rapidly rising and falling hydrograph with a sharp peak. Rainfall that is mainly concentrated in the upper reaches of the catchment produces a lower peak that occurs later and a broader hydrograph. The distribution coefficient, which is the maximum point rainfall to the average rainfall over the catchment, is frequently used as an index.
- Spatial distribution of rainfall with time: this parameter is significant on small catchments. On large catchments the equalising effect makes the hydrograph insensitive to rainfall distribution with time. In principle, a hyetograph that starts at a high intensity and decreases gradually to zero produces a hydrograph with a convex rising limb. A rainfall excess distribution, which gradually increases from zero to a maximum and stops, leads to a hydrograph with an upwards concave rising limb. On small catchments the rising limb is followed by a flat peak (saturation segment). The recession limb of the hydrograph from a decreasing rainfall is concave upwards and convex for an increasing hyetograph.
- Direction of storm movement: The direction of storm movement has the greatest effect on elongated catchments. The amount of rain, over the same period, produces a much greater peak when the storm is moving down the valley than when the storm is moving up the valley. The rainfall from a storm moving up the valley becomes runoff long before the storm reaches the top of the catchment.

The effect of rainfall duration and the concept of the runoff hydrograph are best illustrated with the aid of Figure 3-5.

Figure 3-5: Surface Runoff Hydrograph



The following properties are identified on the graph:

- Lag Time (L). The time interval, from the centre of mass of the rainfall excess, to the peak of the resulting hydrograph.
- Time to Peak (t_p). The time interval from the start of rainfall excess to the peak of the resulting hydrograph.
- Time of Concentration (t_c). The time interval from the end of rainfall excess to the inflection point (change of slope) on the recession curve.
- Recession Time (t_r). Time from peak to the end of surface runoff.
- Time Base (t_b). Time from the beginning to the end of surface runoff.

3.8 Approaches developed and used by DWS

The DWS no longer uses the depth-duration-frequency (DDF) coaxial graph (the approach given in the HRU documents), as well as the storm rainfall information emanating from TR102 (Adamson, 1983).

The DWS adopted the use of four approaches to determine the catchment storm rainfall, when performing flood frequency analyses: These include the maximum station-rainfall approach using catchment statistics (MSRcs), the maximum station-rainfall approach using station statistics (MSRss), the daily catchment-rainfall approach (DCR) and the Smithers regional-rainfall approach (SRR). The methodologies are summarised below:

Maximum Station-Rainfall approach (MSR)

Input:

- Point rainfall record at each available station
- Weighted representative catchment area of each rainfall station (e.g. from Thiessen polygons)

MSR_{CS} (CS – Statistical analysis on Catchment Rainfall)

Note: Applicable only for patched data

Determine:

- The highest 1-day, 2-day, 3-day ... n-day point rainfall per annum, for each station
- The highest 1-day, 2-day, 3-day ... n-day weighted representative catchment rainfall per annum (Thiessen polygons)
- The highest ½t_c, t_c and 2t_c (storm durations) catchment rainfall per annum
- The estimated storm duration events catchment rainfall, for all exceedance probabilities (statistical analyses)

The ARF applicable to this approach will most probably differ to the ARF applicable to the next approach.

MSR_{SS} (SS – Statistical Analysis on Station Rainfall)

Note: Applicable for either patched or unpatched data

Determine:

- The highest 1-day, 2-day, 3-day ... n-day point rainfall per annum, for each station
- The highest ½t_c, t_c and 2t_c (storm durations) point rainfall per annum, for each station
- The estimated storm duration events point rainfall per annum, for each station, for all exceedance probabilities (statistical analyses)
- The estimated storm duration events catchment rainfall, for all exceedance probabilities (Thiessen polygons)

Apply a suitable ARF to determine the estimated effective catchment rainfall, for all considered storm events and appropriate exceedance probabilities.

Daily Catchment-Rainfall approach (DCR)

Note: Applicable only for patched data

Input:

- Point rainfall record at each available station
- Weighted representative catchment area of each rainfall station (e.g. from Thiessen polygons)

Determine:

- For every day, in each year, the weighted representative catchment rainfall (Thiessen polygons)
- The highest 1-day, 2-day, 3-day ... n-day catchment rainfall per annum
- The highest ½t_c, t_c and 2t_c (storm durations) catchment rainfall per annum
- The estimated storm duration events catchment rainfall, for all exceedance probabilities (statistical analyses)

As this method constitutes an analysis of the weighted catchment rainfall for each day, there is no need to apply an ARF.

Smithers-Schulze Regional-Rainfall approach (SSR)

The Smithers and Schulze regional-rainfall approach is a regional scale-invariance model developed for South Africa at the University of KwaZulu-Natal. The model is applied through a software package which extrapolates n-day regional rainfall from a database of rainfall stations across South Africa and provides n-day catchment rainfall frequencies specific to the limits of a defined rainfall area.

In implementing the model, the user defines the area over which the regional rainfall is required. The software generates a raster grid, from which it will then extrapolate n-day point rainfall data from within the specified area to provide point rainfall data for each point of the stipulated raster grid. The user is then able to calculate the average n-day regional catchment rainfall from the raster data obtained from the model.

The first 3 approaches make use of daily recorded rainfall data, obtainable from SAWS, and the patched rainfall database developed by Lynch (2004). The difference between them is outlined in Table 3-4.

The DWS are continuously evaluating and comparing all the above-mentioned approaches, in order to use the best possible estimates for the catchment under investigation.

Table 3-4: Approaches to determine catchment rainfall from rainfall station data

<table>
<tr><th colspan="2">Approach</th><th>Info</th><th>process →</th><th>Product</th><th>process →</th><th>Product</th><th>process →</th><th>Product</th></tr>
<tr><td rowspan="2">MSR</td><td>SS</td><td rowspan="3">Daily Rainfall event (data) per station</td><td rowspan="2">Choose (n-day) event</td><td rowspan="2">Maximum annual n-day rainfall event per station</td><td>Statistical analysis</td><td>n-day Rainfall frequencies per station</td><td>Weighting + ARF</td><td rowspan="3">n-day Catchment Rainfall frequencies</td></tr>
<tr><td>CS</td><td>Weighting</td><td rowspan="2">Maximum annual n-day Catchment Rainfall event</td><td>Statistical analysis + ARF*</td></tr>
<tr><td>DCR</td><td></td><td>Weighting</td><td>Daily catchment rainfall event</td><td>Choose (n-day) event</td><td>Statistical analysis</td></tr>
</table>

Note on the ARF: Flood Studies observed that the 'area-reduction' that should be applied to the MSR_{CS} approach (ARF*) seems to be different from the ARF applied to the (old?) MSR_{SS} approach. The ARF seems redundant in case of the DCR approach (as expected).

3.9 Post HRU and TR102 Development

There have been various attempts to try and ease the burden on the design engineers and scientists by developing a system that could determine design rainfall depths for various durations in a much simpler way.

Studies, undertaken by the School of Bioresources Engineering and Environmental Hydrology at the University of KwaZulu-Natal:

- Long duration design rainfall estimates for South Africa by JC Smithers and RE Schulze (WRC Report No. 811/1/00).
- Design Rainfall and Flood Estimation in South Africa by JC Smithers and RE Schulze (WRC Report No. 1060/1/03).
- Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa by SD Lynch (WRC Report No. 1156/1/04)

<u>What lies ahead?</u>

The DWS are continuously evaluating new studies for their applicability in its analyses.

The DWS also continuously investigates/changes/upgrades existing methodologies where necessary and applicable keeping in mind that it cannot deviate (too much) from the approach followed in developing the methods, but we can improve on the reliability of, for example, catchment rainfall.

SANCOLD, the WRC, prominent Universities and the DWS have engaged in a joint venture in an attempt to upgrade/improve the flood frequency methods in use in South Africa. This will also include interrogation of rainfall methodologies as an input to these methods.

4. DETERMINISTIC METHODS

The deterministic methods are used in cases where no flow records or very few flow records exist. Catchment characteristics and storm rainfall, of the area of concern, is used to calculate flood peaks.

4.1 Rational Method

The Rational method is a simple method that uses catchment characteristics and storm rainfall to reproduce flood peaks. Although it is generally recommended that the method only be applied to catchments smaller than 15 km^2, it has been used successfully for larger catchments, by more experienced users. The Rational equation is given by:

$$Q = C\ i\ A_e$$

where:

Q	-	Peak flow (m^3/s)
C	-	Runoff coefficient (dimensionless)
i	-	Average rainfall intensity (m/s)
A_e	-	Effective catchment area (m^2)

The equation can be transformed to a more workable format where the catchment area (A) is in km^2, and the average rainfall intensity (i) is given in mm/h:

$$Q_P = 0.278\ C_P\ i_{P(Av)}\ A_e$$

where:

P	-	Exceedance probability
Q_P	-	Peak flow (m^3/s)
0.278	-	Conversion factor
C_P	-	Runoff coefficient (dimensionless)
$i_{P(Av)}$	-	Average rainfall intensity (mm/h)
A_e	-	Effective catchment area (km^2)

Runoff coefficient, C

The runoff coefficient is an integrated value representing a number of factors, influencing the rainfall-runoff relationship. It reflects that part of the storm rainfall contributing to the peak flood runoff at the outlet of the catchment. The runoff coefficient is given by:

$$C = \alpha C_1 + \beta C_2 + \gamma C_3$$

where:

C_1 - rural runoff coefficient

C_2 - urban runoff coefficient

C_3 - lakes runoff coefficient

α, β and γ are the area weighting factors and $\alpha + \beta + \gamma = 1$.

C_1 and C_2 are determined by sub-dividing it up into components (See Table 4-1).

Table 4-1: Recommended values of runoff coefficients

RURAL: C_1			**MAP (mm)**		
Component		**Classification**	**<600**	**600 - 900**	**>900**
C_S	Catchment Slope	Very flat areas (< 3%)	0.01	0.03	0.05
		Flat areas (3% - 10%)	0.06	0.08	0.11
		Hilly areas (10% - 30%)	0.12	0.16	0.20
		Steep areas (>30%)	0.22	0.26	0.30
C_P	Soil Permeability	Very Permeable (A)	0.03	0.04	0.05
		Permeable (B)	0.06	0.08	0.10
		Semi-Permeable (C)	0.12	0.16	0.20
		Impermeable (D)	0.21	0.26	0.30
C_V	Vegetation	Dense Bush, Forest	0.03	0.04	0.05
		Thin Bush, Cultivated Land	0.07	0.11	0.15
		Grassland	0.17	0.21	0.25
		Bare Surface	0.26	0.28	0.30

Probability of exceedance	50	20	10	5	2	1	0.5
F_T	0.35	0.54	0.66	0.77	0.90	1.00	1.10

The relationship between rainfall and runoff is not linear and a catchment is often more saturated for a storm with a low exceedance probability than is the case with a storm with a high exceedance probability.

$$C_1 = F_T(C_S + C_P + C_V)$$

URBAN: C_2 ($2 \le T \le 20$)	
Use	**factor**
Lawns	
Sandy, flat (< 2%)	0.05 - 0.10
Sandy, steep (> 7%)	0.15 - 0.20
Heavy soil, flat (< 2%)	0.13 - 0.17
Heavy soil, steep (> 7%)	0.25 - 0.35
Residential areas	
Single family area (houses)	0.30 - 0.50
Apartment dwelling (flats)	0.50 - 0.70
Industrial areas	
Light industry areas	0.50 - 0.80
Heavy industry areas	0.60 - 0.90
Business areas	
City centre	0.70 - 0.95
Suburban	0.50 - 0.70
Streets	0.70 - 0.95

URBAN: C_2 ($20 < T \le 50$)	
Lawns	0.35 - 0.50
Other	0.70 - 1.00

URBAN: C_2 ($50 < T \le 200$)
1.00

Note: The values for F_T in the above table is "experience factors", introduced and refined at Flood Studies

Rural areas:

Most of the catchment analysed in the DWS' flood frequency studies fall within this category.

Steepness: steep slopes cause more runoff; thus the assumption can be made that steep slopes are less permeable than flat areas. Natural ponds also decrease with the increase in slope. Thus, the slope of the catchment has a significant contribution to the runoff.

Permeability of soil: the permeability of the soil can be classified as follows:

- Group A: (very permeable) soils are sand, loamy sand or sandy loam. It has the lowest runoff potential and high infiltration rates when thoroughly wetted.
- Group B: (permeable) soils are silt loam or loam. It has a moderate infiltration rate when thoroughly wetted.
- Group C: (semi-permeable) sandy clay loam. It has low infiltration rate when thoroughly wetted.
- Group D: (impermeable) clay loam, silt clay loam, sandy clay, silt clay or clay. It has the highest runoff potential and very low infiltration rates when thoroughly wetted.

The classification can be made from visual inspection or by using soil permeability maps.

Vegetation: vegetation can be classified as follows:

- Group A: Forest, dense bush and wood
- Group B: Thin bush and cultivated land
- Group C: Grassland
- Group D: Bare surface (no vegetation)

Runoff increases as the density of the vegetation decreases. The vegetation can be determined by visual inspection or by using the publication by Acocks (1988).

Exceedance Probability: the exceedance probability has an important effect on the runoff coefficient. The smaller the exceedance probability, the larger C will be. This is to make provision for the variation of known effects that increase with rainfall intensity but are not accounted for in the calculations.

These effects include:

- Shortened t_c.
- Higher percentage runoff.
- Greater possibility of saturated catchment prior to the storm.
- Validity of the basic assumptions and the calculation method.

Urban areas:

These should be considered only if occupying more than 5% of the catchment area.

It is normally not necessary to adjust the value of C according to the return period.

Area-reduction factor, ARF

Refer to Chapter 3 (Section 3.5 and Section 3.8).

Rainfall intensity, *i*

The intensity of a storm increases as the exceedance probability decreases (return period increases) and as the duration of the storm decreases. To obtain the largest possible flood peak, for a given exceedance probability, the storm rainfall must have a duration equal to the t_c.

The storm rainfall P, is reduced to catchment rainfall P_{Av} through multiplication by an ARF for the duration and catchment area of interest. The ARF compensates for the non-spatial uniformity of the storm rainfall.

$$P_{Av} = P \times ARF$$

The rainfall intensity i, is obtained by dividing P_{Av} by the critical storm duration D (where D can be ½t_c, t_c or 2t_c).

$$i_{Av}(mm/hr) \;=\; P_{Av}(mm) \,/\, D(h)$$

The average rainfall intensity is the value that is used in the Rational method.

4.2 SCS-SA Method

The SCS Method is widely used in the USA, Germany, France, Middle-East, Australia and other parts of Africa.
In South Africa it was introduced by Reich in 1962 but only became popular after 1979 when the first SCS user manual was introduced by Schulze and Arnold.

Why the SCS Method?

The major reasons for the SCS model's widespread usage include the following:

- the mathematical equations describing the model are simple to use;
- the main inputs required for the model are obtained readily;
- the technique is user oriented and various monographic solutions have been presented to assist in computations;
- the technique has been shown to provide realistic estimates of peak discharge and runoff volume when compared with observed data.

A verification study was undertaken by SRK (Campbell et al., 1986) to compare actual observed data with the SCS, Rational, Kinematical and Time-Area based methods. The SCS method turned out to be the best method to use for small catchments.

The DWS recommends the use of the SCS method to calculate flood peaks for small catchments.

SCS software

The "Visual SCS-SA" software package is available at the School of Bioresources Engineering and Environmental Hydrology, University of KwaZulu-Natal, Pietermaritzburg.

4.3 Unit Hydrograph Method

Background

This method was developed by the HRU (HRU 3/69 and HRU 1/72). It is recommended for catchments ranging 20 km^2 < A < 10 000 km^2.

The basic assumption in the unit hydrograph method is that a unit of effective precipitation (that part of the precipitation which results in direct runoff), uniformly distributed over the catchment in both time and space, will result in a uniquely shaped hydrograph for that catchment. Further assumptions are that the ordinates of the hydrograph are linearly proportional to the depth of effective precipitation and that the shape is independent of antecedent conditions. While a hydrograph shape must clearly be dependent on antecedent conditions, this assumption implies that the hydrograph is related to the average state of the catchment.

Unit hydrographs for 96 river measuring stations in South Africa have been compiled from historical data and, from these, synthetic unit hydrographs have been derived for 9 regions of South Africa, with similar catchment characteristics such as topography, soil type, vegetation type and rainfall characteristics (see Appendix Figure A-2).

The unit hydrograph is a typical hydrograph for a given catchment, produced by rainfall of given duration and intensity. In South Africa, the standard unit hydrograph used in the derivation of the SUH method is associated with a net rainfall of 1 mm in 1 h.

Calculation procedure

The first step is to determine the values of the dimensionalising factors, basin lag (T_L) and unitgraph peak (Q_P):

- Basin lag: T_L

$$T_L = C_t I_c^{0.36}$$

where:

$$I_c = \frac{L\ L_c}{\sqrt{S}} \qquad \text{(catchment index)}$$

and

- Unit hydrograph peak: Q_P

$$Q_P = K_u \frac{A_e}{T_L}$$

with:

L	-	length of longest watercourse
L_c	-	distance along the main watercourse to a point opposite the catchment centroid
S	-	average slope of the longest watercourse
A_e	-	Effective catchment area
C_t	-	Regional lag coefficient
K_u	-	Regional discharge coefficient

Table 4-2: Veld type zones

Veld type zone	Generalized veld type description	C_t	K_u
1	Coastal tropical forest	0.99	0.261
2	Schlerophyllous bush	0.62	0.306
3	Mountain sourveld	0.35	0.277
4	Grasslands of interior plateau	0.32	0.386
5	Highland sourveld and Dohne sourveld	0.21	0.351
5A	As for Zone 5 – but soils weakly developed	0.53	0.488
6	Karoo	0.19	0.265
7	False Karoo	0.19	0.315
8	Bushveld	0.19	0.367
9	Tall sourveld	0.13	0.321

When these have been calculated, the relationship between flow and time can be determined from the tabulated values of T/T_L vs Q/Q_P. The appropriate unit hydrograph is obtained according to the regional classification (Appendix, Table A-1).

The resulting hydrograph is the standard unit hydrograph which represents the direct runoff in cubic metres per second resulting from 1 mm of effective precipitation falling uniformly for a period of 1 h over the catchment.

- Derivation of the S-curve

The next step is the derivation of the standard S-curve which is defined as the cumulative sum of the hourly ordinates of the standard unit hydrograph and is a standardised approximation of the hydrograph that would result from a constant precipitation rate of 1 mm per hour.

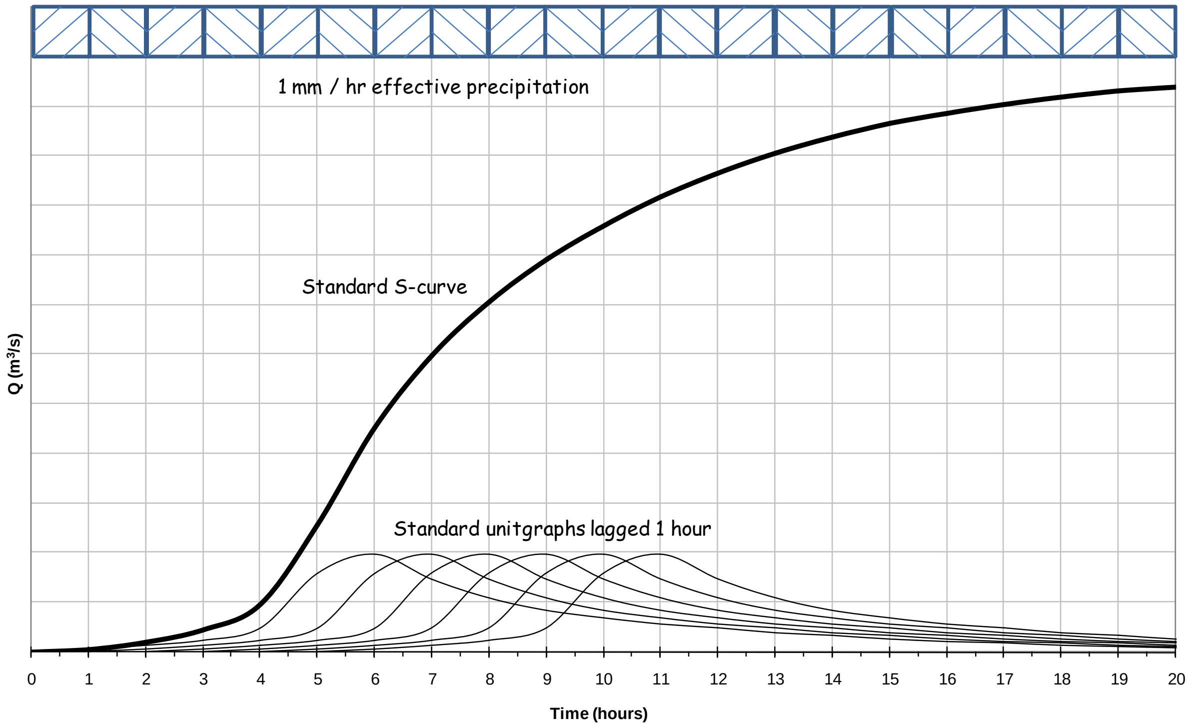


Figure 4-1: Derivation of S-curve

In practice there are often fluctuations in the horizontal part of the S-curve, particularly with small catchments. This is normally ascribed to the time intervals being too large for accurate calculation or to the duration of the unit hydrograph being too long in relation to the lag in the catchment.

Once the S-curve has been derived unit hydrographs for other durations can be determined by lagging the S-curve with respect to itself by the storm duration of interest, calculating the difference in the corresponding ordinate values and dividing this difference by the storm duration. The resulting unit hydrograph represents 1 mm of effective precipitation uniformly distributed over a period equal to the storm duration.

The derivation of the lagged S-curve is shown in Figure 4-2.

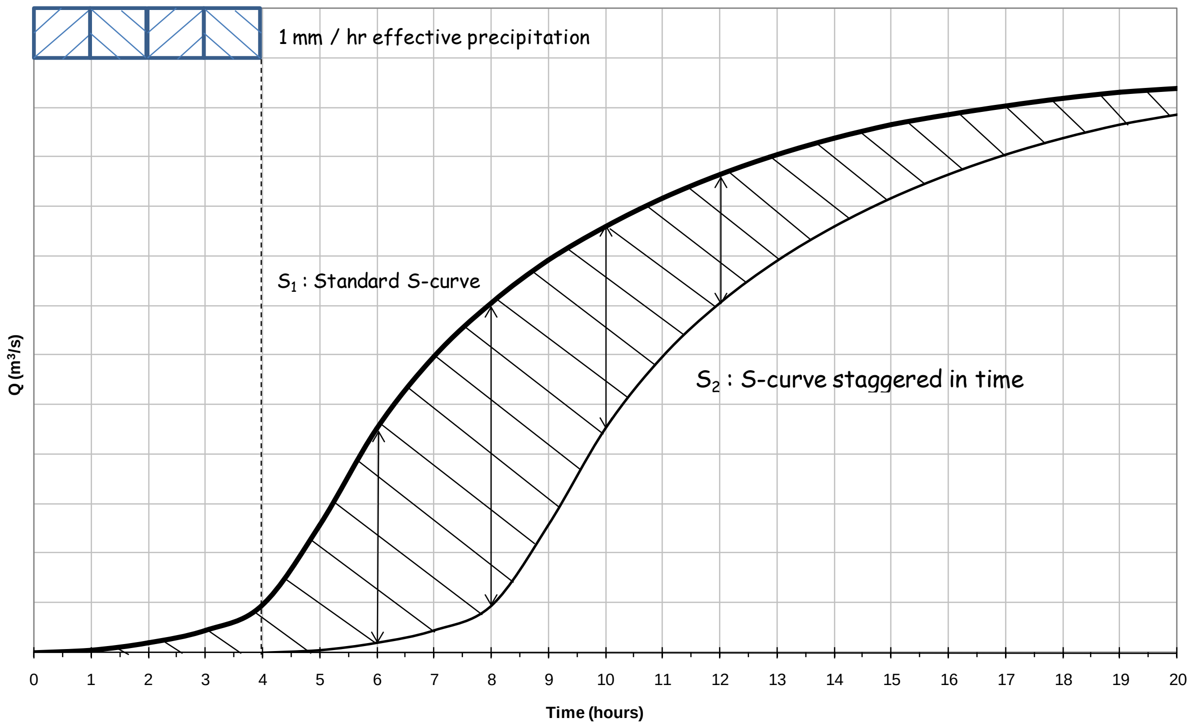


Figure 4-2: Staggered S-curve

Curve S_1 is the hydrograph which will result from continuous effective precipitation of 1 mm per hour starting at time zero, and S_2 is the identical hydrograph starting 4 h later. The difference represents the hydrograph of 1 mm per hour for 4 h. If the ordinates are divided by 4, the resulting unit hydrograph will be that of 1 mm precipitation falling in 4 h.

Rainfall

The storm rainfall P is reduced to catchment rainfall P_{Av} by using an ARF for the duration and catchment area of interest.

$$P_{Av} = P \times ARF$$

Compensation, for the fact that only a portion of the P_{Av} rainfall realises into runoff, is made by applying a storm runoff factor k. (Refer to Appendix Figure A-3 for a graph on the storm runoff factor, k).

Consequently, the effective storm rainfall (or excess rain) P_e is obtained from:

$$P_e = P_{Av} \times k$$

Example:

Effective catchment area: A_e (km^2)	310	Area reduction factor: ARF	0.878
Longest watercourse: L (km)	37	Storm runoff factor: k	0.390
Centre of gravity: L_C (km)	15	Effective storm rainfall: P_e (mm)	37.67
River slope: S (m/m)	0.004969	Catchment Index: I_C	7873
Time of concentration: t_c (h)	8	Lag Coefficient: C_t	0.320
Veld Zone	4	Basin Lag: T_L (h)	8.1
Exceedance Probability: EP (%)	1	Coefficient: K_U	0.386
Storm duration: D (h)	8	Unit hydrograph peak: Q_P	14.8
Storm rainfall: P (mm)	110	**Peak discharge: m^3/s**	**304**

Time(T) h	T/T_L	D = 1 h UH		D* = 8 h UH				HG
		Q/Qp	S-curve (S_1)	Staggered S-curve (S_2)	$S_8 = S_1 - S_2$	$S_8 \times D/D^*$	$\times Q_P$	$\times P_e$
0	0.00	0.000	0.000		0.000	0.000	0.00	0
1	0.12	0.031	0.031		0.031	0.004	0.06	2
2	0.25	0.069	0.100		0.100	0.012	0.18	7
3	0.37	0.122	0.222		0.222	0.028	0.41	15
4	0.49	0.251	0.473		0.473	0.059	0.87	33
5	0.62	0.805	1.277		1.277	0.160	2.36	89
6	0.74	0.989	2.267		2.267	0.283	4.19	158
7	0.87	0.732	2.998		2.998	0.375	5.55	209
8	0.99	0.541	3.539	0.000	3.539	0.442	6.55	247
9	1.11	0.420	3.960	0.031	3.929	0.491	7.27	274
10	1.24	0.343	4.303	0.100	4.203	0.525	7.77	293
11	1.36	0.285	4.588	0.222	4.366	0.546	8.07	**304**
12	1.48	0.240	4.828	0.473	4.355	0.544	8.06	303
13	1.61	0.199	5.027	1.277	3.749	0.469	6.94	261
14	1.73	0.164	5.191	2.267	2.924	0.366	5.41	204
15	1.85	0.135	5.326	2.998	2.327	0.291	4.30	162
16	1.98	0.111	5.437	3.539	1.898	0.237	3.51	132
17	2.10	0.089	5.525	3.960	1.566	0.196	2.90	109
18	2.23	0.071	5.596	4.303	1.293	0.162	2.39	90
19	2.35	0.056	5.652	4.588	1.065	0.133	1.97	74
20	2.47	0.045	5.697	4.828	0.869	0.109	1.61	61
21	2.60	0.035	5.733	5.027	0.706	0.088	1.31	49
22	2.72	0.028	5.760	5.191	0.570	0.071	1.05	40
23	2.84	0.021	5.782	5.326	0.456	0.057	0.84	32
24	2.97	0.017	5.798	5.437	0.362	0.045	0.67	25
25	3.09	0.011	5.809	5.525	0.284	0.036	0.53	20

4.4 Direct Runoff Hydrograph Method

This method was developed by the HRU (1974). It is based on the proven assumption that a hydrograph can be reproduced with reasonable accuracy by routing the corresponding areal rainfall which is uniformly distributed over the catchment after reducing it by storm loss. The catchment is considered to be a simple reservoir-type storage to which the Muskingum routing method is applied.

Applicability - Catchment area: 20 km^2 < A < 20 000 km^2

Rainfall

The same procedure, which is used to calculate the P_e for the SUH method, is used for the DRH method.

Storm Hyetograph

Rainfall distribution with time is the driving mechanism of this method. The shape of the hydrograph is determined by the rainfall distribution in time and the time of concentration.

After calculating the P_e its distribution with time must be determined. The percentage of excess rain that has fallen, in a certain percentage of the duration can be estimated from Figure 4-3 (see also Appendix Figure A-4).

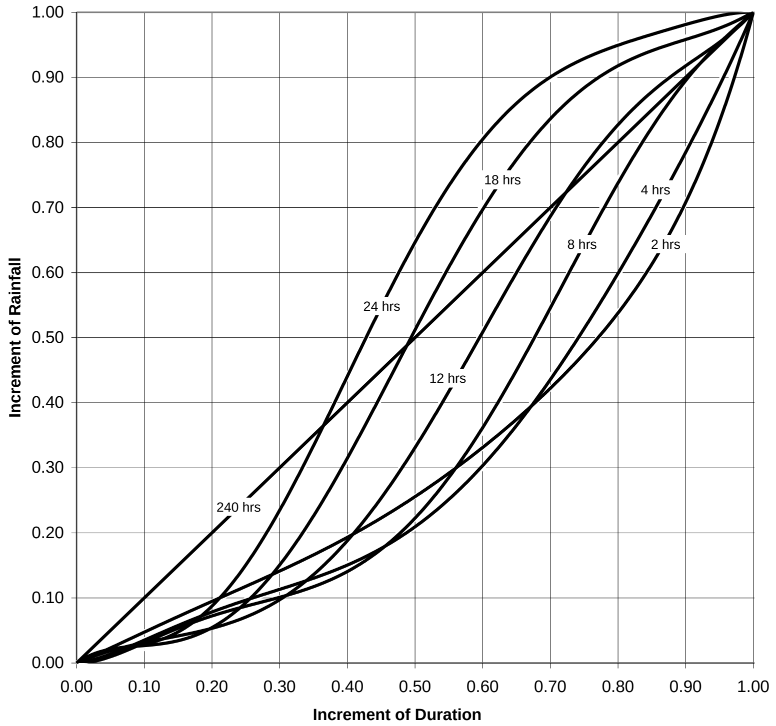


Figure 4-3: Rainfall distribution with time

In the case where the duration of interest does not correspond to the duration depicted on Figure 4-3, interpolation techniques must be applied to obtain the curve that corresponds to the duration of interest. This distribution is used to determine the hyetograph that will be used to derive the hydrograph.

The ΔP_e , in mm, is converted to ΔP_e in m^3/s (inflow):

$$\Delta P_e(m^3/s) = 0.278\ i\ A_e$$

where:

i	-	ΔExcess rain (mm) / Δt (h)
A_e	-	Effective catchment area (km^2)

Flow is considered at discrete time steps, with $0.1D \geq \Delta t \geq 0.05D$.

Muskingum routing

The hyetograph is transformed into a hydrograph, by applying the following Muskingum routing equation:

$$Q_{out(2)} = C_0 Q_{in(2)} + C_1 Q_{in(1)} + C_2 Q_{out(1)}$$

where:

Q_{out}	-	outflow (m^3/s)
Q_{in}	-	inflow (m^3/s)

C_0, C_1 and C_2 are the Muskingum coefficients with $C_0 + C_1 + C_2 = 1.$

The Muskingum coefficients are calculated as follows:

$$C_2 = e^{-\left(\frac{\Delta t}{K}\right)} \qquad C_1 = \frac{K}{\Delta t}(1 - C_2) - C_2 \qquad C_0 = 1 - \frac{K}{\Delta t}(1 - C_2)$$

with the Muskingum routing factor K estimated as follows:

$$K = a\ A_e^{0.318}$$

where a is related to the veld type zone, as shown in Table 4-3.

Table 4-3: Factor a, for calculating the routing factor K.

Veld Type	1	2	3	4	5	6	7	8	9
a	1.83	1.30	1.10	0.97	0.79	0.86	0.48	0.45	0.55

From experience and careful consideration, it has been found that, in South Africa, a way of determining the flood routing factor is to link it to the t_c by the following relation (unpublished):

$$K = 0.6\ t_c$$

The Muskingum routing equation can now be solved in a stepwise manner. Initially, inflow and outflow are known and, after the lapse of one time increment, the new inflow is known so the resulting outflow can be computed. The discharge at the end of the first time step becomes the discharge at the beginning of the next time step.

Example:

Effective catchment area: A_e (km^2)	310	Area reduction factor: ARF	0.878
Longest watercourse: L (km)	37	Storm runoff factor: k	0.390
Centre of gravity: L_C (km)	15	Effective storm rainfall: P_e (mm)	**37.67**
River slope: S (m/m)	0.004969	Delta t: Δt (h)	0.8
Time of concentration: t_c (h)	8	Muskingum routing factor: K	4.8
Veld Zone	4	C_0	0.079
Exceedance Probability: EP (%)	1	C_1	0.075
Storm duration: D (h)	8	C_2	0.846
Storm rainfall: P (mm)	110	**Peak discharge: m^3/s**	**391**

$$Q_{out(2)} = C_0 Q_{in(2)} + C_1 Q_{in(1)} + C_2 Q_{out(1)}$$

Time (h)	Increment of Duration (%)	Increment of Rainfall (%)	Excess rain (mm)	Δ Excess rain (mm)	Δ Excess rain (m^3/s)	Routing (m^3/s)
					Q_{in}	Q_{out}
0.0	0	0	0	0	0	0
0.8	10	4	1.51	1.51	162	13
1.6	20	7	2.64	1.13	122	33
2.4	30	10	3.77	1.13	122	46
3.2	40	15	5.65	1.88	203	64
4.0	50	23	8.66	3.01	325	95
4.8	60	36	13.56	4.90	527	146
5.6	70	55	20.72	7.16	771	224
6.4	80	74	27.87	7.16	771	308
7.2	90	89	33.52	5.65	609	366
8.0	100	100	37.67	4.14	446	**391**
						364
						308
						261

5. EMPIRICAL METHODS

5.1 Introduction

Unfortunately, there is no absolute test against which the numerous methods of flood frequency estimation can be compared. No commonly applicable method has yet been developed for South African conditions. The DWS Flood Studies component of (FS_DWS), has made slight improvements to methods, over the years, but the aim will be to combine the various methods into one empirical method for the entire country. This may not be such a far-fetched dream, after all, after having performed numerous flood frequency analyses it became very clear that certain empirical methods will perform better than others in certain areas whilst others will again perform better in other areas.

In this section the various methods, which, at present, are deemed suitable for application in South Africa by the DWS will be described briefly.

The criteria, normally used to evaluate methodologies, are:

- Theoretical soundness
- Simplicity of application
- General acceptability to practising engineers and hydrologists

By definition empirical methods do not meet the first requirement listed above. Nevertheless, they are relatively very easy to apply and have been in use for a long time.

Early empirical methods were of the form:

$$Q = C\,\sqrt{A} \qquad \text{or} \qquad Q = C\,A^k$$

Some of the methods that are going to be discussed are more complex but still easy to apply.

Alexander (1990) cited several empirical methods, but the UK Flood Studies Report (1975) makes no mention of similar equations.

5.2 Midgley and Pitman Method (MIPI)

Background

This method can be described as an Empirical-Probabilistic method, of the form:

$$Q_P = C\,K_P\,{A_e}^{m}$$

with

Q_P	-	flood peak (m^3/s) for the probability of exceedance P
C	-	independent catchment coefficient
K_P	-	constant derived from assumed probability distribution function of the given probability of exceedance
A_e	-	Effective catchment area (km^2)
m	-	constant

One of the earlier methods of this type, widely used in South Africa previously, was the Roberts-method. Roberts assumed a value of 0.5 for $\boldsymbol{m}$ and $\boldsymbol{K_P}$ was a coefficient derived from the Hazen frequency distribution. The major objection to the Roberts-method is that the factor $\boldsymbol{C}$ shows very wide variations from stream to stream and cannot be related to any region or measured variables. Another weakness is the assumption of the same variance and skewness, for all South African rivers, inherent in the Hazen distribution. This method gave way to other methods, of which the so-called MIPI method (Pitman and Midgley, 1967) is a good example.

Developing the method

Pitman and Midgley (1967) retained the value of 0.5 for $\boldsymbol{m}$ but used the Log-Gumbel function to derive $\boldsymbol{K_P}$ and the catchment coefficient $\boldsymbol{C}$ was regionalised. The results are presented in Figure 5-1, as a coaxial diagram.

Although the Log-Gumbel distribution is reported to have a sounder theoretical basis, American studies had shown it to be less satisfactory than the Hazen, Log-Normal and Log-Pearson-III distributions. As in the case of the Roberts method, it is also assumed in the MIPI method that annual peak distributions for all South African rivers have the same variance and skewness.

Applicability - Catchment area: $100\ km^2 < A < 200\ 000\ km^2$
- Outcome: 2 year ≤ T ≤ 200 year flood peaks

Using the method

The method is very simple to apply: Determine MIPI flood region from Figure A-5 in the Appendix and then use Figure 5-1 (Figure A-6 in the Appendix) to determine flood peaks for various recurrence intervals.

Owing to its simplicity and reasonable consistence, the MIPI method is among the standard methods adopted by the DWS.

It regularly still produces very acceptable flood peak estimations. It is thus a very useful method, to include, when comparing results of various methods.

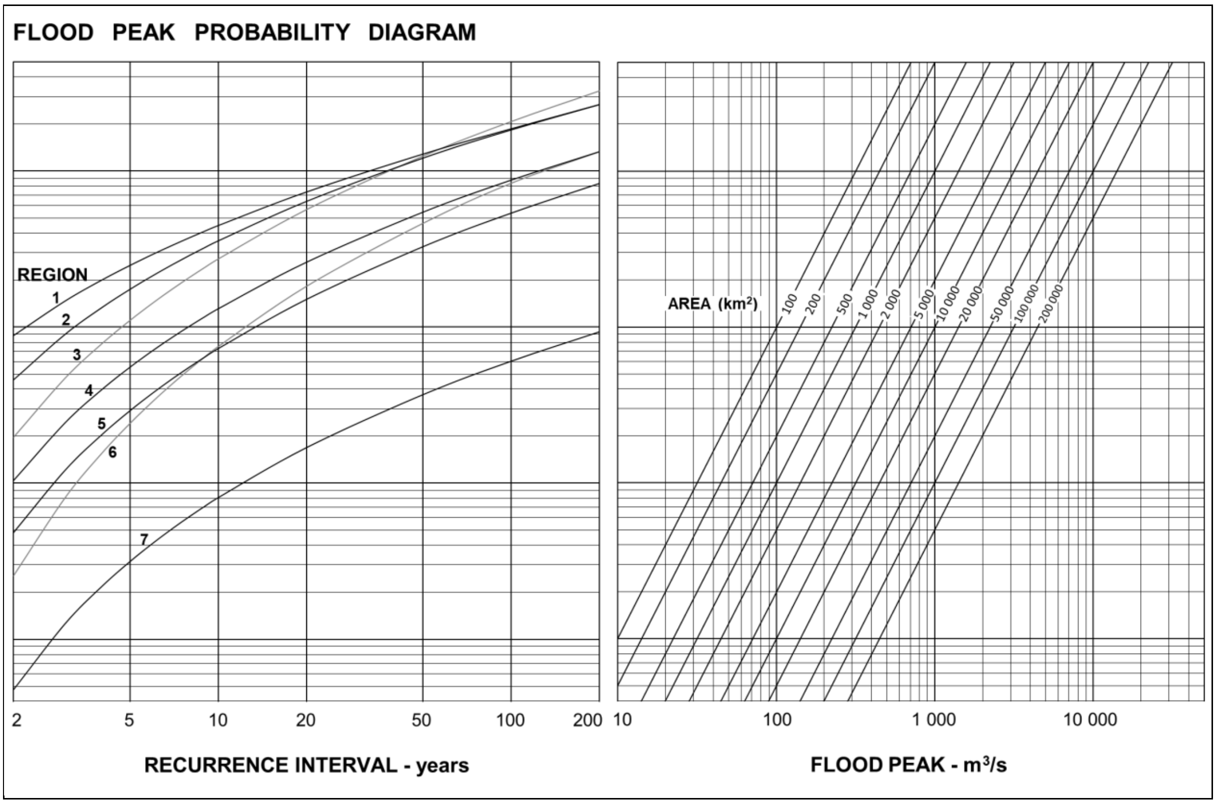


Figure 5-1: Diagram to be used with the MIPI method

5.3 HRU 1/71 (Midgley and Pitman)

Background

Shortly after HRU report 4/69 (Midgley et al, 1969) was published a need arose for a simpler and less cumbersome method than the SUH method for preliminary design or checking purposes – since only the hydrograph peak was needed. At the time, available methods to estimate flood peak probabilities had serious shortcomings in that they took account of only the catchment area and locality of catchment (e.g. the Roberts and MIPI methods). Therefore a clear motivation existed to develop an improved simple method for flood peak estimation.

Developing the method

This method was derived from the original SUH method in HRU Report No. 4/69 during the first general review of the latter (Pitman WV and Midgley DC, 1971).

Consideration of the computational steps involved in the original method led to the selection of four parameters by which to express peak discharge in m^3/s:

$$Q_T = 0.0377\ MAP\ K_T\ {A_e}^{0.6}\ C^{0.2}$$

where

A_e	-	Effective catchment area in km^2
MAP	-	Mean annual rainfall over the catchment in mm
K_T	-	Combined coefficient dependant on the meteorological region, veld type zone and return period
C	-	Catchment parameter; incorporating catchment area and catchment index to reflect the response time of the catchment in terms of area and shape; C equals the area divided by the catchment index as defined in HRU report 4/69 (Midgley et al, 1969):

$$C = \frac{A_e\sqrt{S_L}}{LL_C}$$

Applicability - Catchment area: $20\ km^2 < A < 100\ 000\ km^2$

- Outcome: 2 year ≤ T ≤ 200 year flood peaks

Using the method

The method is very simple to apply and thus provides an easy, quick check of other methods:

- determine the catchment parameters (A_e, S_L, L and L_C)
- calculate C
- determine the MAP
- establish veld type zone (from Figure A-2 in the Appendix)
- Use Table 5-1 to obtain K_T for various probabilities of exceedance.
- Compute $Q_T = 0.0377\ MAP\ K_T\ A_e^{0.6}\ C^{0.2}$

In the experience of the DWS, HRU 1/71 is a useful method of flood estimation.

Table 5-1: Values of coefficient K_T applicable to HRU 1/71

Probability of exceedance (%)			50	20	10	5	2	1	0.5
Veld type zone	1	summer / all year	0.05	0.12	0.17	0.23	0.32	0.40	0.50
	2	winter	0.14	0.30	0.42	0.53	0.68	0.80	0.93
		all year	0.28	0.61	0.83	1.05	1.36	1.60	1.86
	3	summer	0.08	0.20	0.29	0.40	0.55	0.70	0.87
	4, 5a, 9	summer	0.14	0.35	0.50	0.68	0.95	1.20	1.49
	5	summer	0.16	0.40	0.59	0.79	1.11	1.40	1.73
	6	winter	0.09	0.23	0.34	0.45	0.63	0.80	0.99
	6, 7	summer / all year	0.18	0.46	0.67	0.90	1.27	1.60	1.98
	8	summer	0.11	0.29	0.42	0.57	0.79	1.00	1.24

5.4 Catchment Parameter method (CAPA)

Background

McPherson (1983) stated that the rapid calculation of flood peaks in ungauged catchments necessitates the following steps:

- estimation of the Mean Annual Flood ($\overline{Q}$) or the flood peak which have a 50% probability of exceedance ($Q_{50\%}$)
- development of regional flood frequency growth curves by means of statistical analyses of annual maximum flood peak records
- restriction of the upper limits of frequency curves by a kind of maximum flood peak

McPherson (1983) attempted to solve the first step above. Unfortunately the catchment slope, which forms part of this method, was determined by applying a method that gave erroneous results. This was later rectified by Thobejane (2001) under the guidance of van der Spuy and Linström.

The second step was addressed to a large degree by FS_DWS where factors for determining flood peaks of 20% to 1% were determined, to be able to apply this methodology (unpublished).

The last step had been addressed by Kovacs (1988), which is discussed under Chapter 6.

This method is applied by the DWS as a separate method, which is called the Catchment Parameter Method (CAPA) and it seems to give quite good estimates of the applicable flood peaks, when compared with all the other methods in use.

Developing the method

This method was developed in the FS_DWS. The aim was to provide a quick, robust method for the estimation of the mean annual flood ($\overline{Q}$ or $Q_{50\%}$), by only using readily calculable variables. The method is based on flow, rainfall and physical geography data from more than 140 South African catchments.

McPherson identified 10 variables which were likely to have an influence on $\overline{Q}$. The preliminary analysis had shown that four variables were possibly the most significant in determining $\overline{Q}$:

- **Effective Catchment Area (A_e in km²)**: This obviously has a large influence, comprising the $\{L^2\}$ component of the flood volume $\{L^3\}$
- **Mean Annual Precipitation (MAP in mm):** For relatively small floods, such as $\overline{Q}$, the MAP was considered to be more suitable than the 2-year, 1-day storm rainfall.
- **Mean Catchment Surface Slope (S_A in m/m):** This is related to runoff retention and the velocity of runoff accumulation.
- **Shape Parameter:** defined as the **river length (L in km)** divided by the **square root of A_e):** This parameter is related to runoff accumulation and areal storm rainfall.

The variables were combined to form a single lumped parameter (M) as follows:

$$M = MAP \left(\frac{100\, S_A \sqrt{A_e}}{L} \right)^{1/2}$$

The solution was presented as a family of parallel lines representing M, drawn on a log-log plot of A_e versus $Q_{50\%}$ - from which $Q_{50\%}$ can be read, using A_e as input (see Figure 5-2; also, Appendix Figure A-7).

Applicability - Catchment area: 10 km^2 ≤ A ≤ 100 000 km^2

- Outcome: 2 year ≤ T ≤ 200 year flood peaks

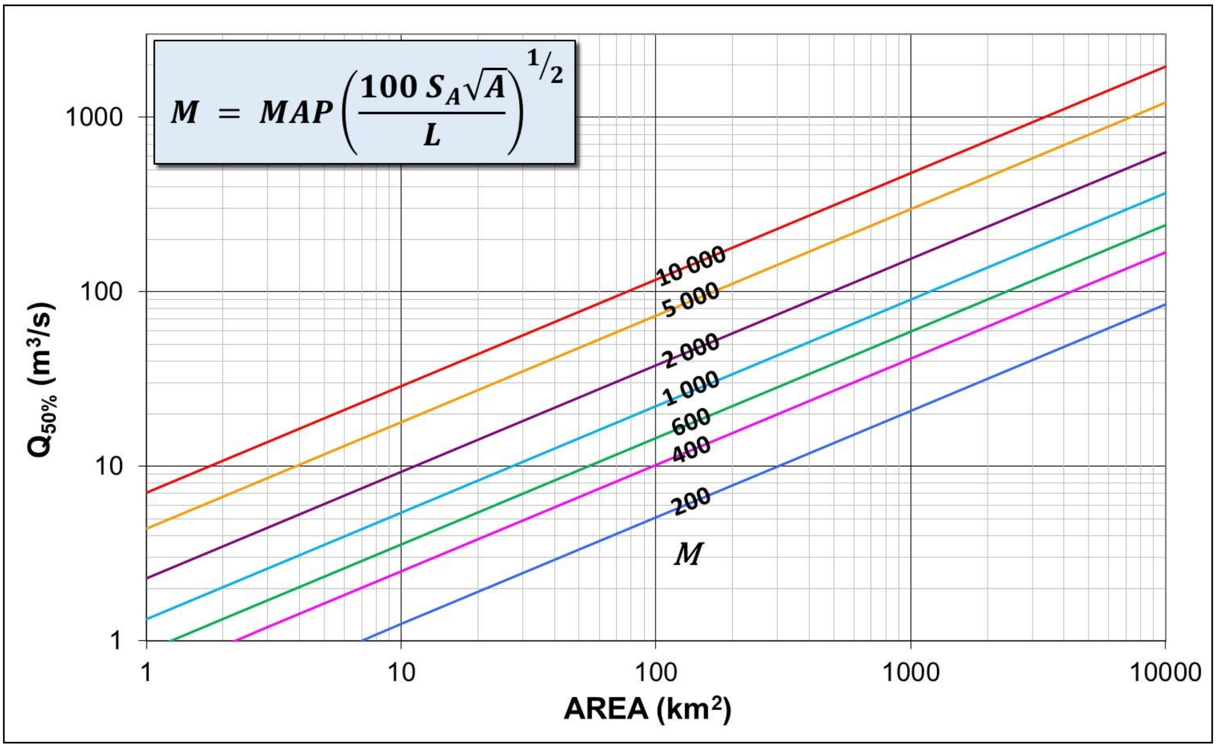


Figure 5-2: CAPA 'M' diagram to determine Mean Annual Flood

__Using the method__

The method is simple to apply and provides an easy, quick check of other methods:

- determine the four parameters (***A_e*, *MAP*, *SA*** and ***L***)
- calculate ***M***
- from Figure 5-2 determine the Mean Annual Flood (***$Q_{50\%}$***)
- Use Table 5-2 to obtain ***K_P*** for various probabilities of exceedance.
- ***$Q_P = K_P * Q_{50\%}$***

Table 5-2: Values of K_P for various probabilities of exceedance

MAP	Probability of exceedance (%)					
	20	**10**	**5**	**2**	**1**	**0.5**
100	4.49	9.49	16.97	31.41	45.36	61.51
200	3.27	5.96	9.65	16.26	22.15	28.50
400	2.47	3.97	5.89	9.13	11.81	14.52
600	2.13	3.20	4.52	6.72	8.45	10.13
800	1.93	2.76	3.79	5.46	6.75	7.96
1000	1.79	2.48	3.32	4.68	5.71	6.65
1500	1.57	2.05	2.64	3.58	4.26	4.86
2000	1.44	1.80	2.26	2.99	3.50	3.93

6. MAXIMUM EXPECTED FLOOD

6.1 Introduction

A realistic estimate of the maximum expected flood peak, at a given site, is still a matter of great debate amongst flood hydrologists and engineers. There are basically three ways in which such a flood can be estimated, namely empirically, deterministically and probabilistically.

Regional Empirical methods:

Maximum observed flood peaks in a hydrologic homogeneous region are plotted against catchment area. An upper envelope curve is then considered as the upper limit of expected flood peaks.

The biggest single advantage of the regional empirical approach is that it emanates from a vast amount of experience, and there is a good chance that a few really extreme flood events are included in the database, especially if the observations cover a relatively large area and/or period.

Fundamental shortcomings are:

- Uncertainty in determining boundaries of homogeneous regions (common drawback of all regional approaches)
- Unusual hydrological features of very small or very large catchments often not accounted for by the regional approach
- The selection of envelope curves is very subjective

Once civil engineers and hydrologists became acquainted with the more elaborate and universal deterministic and probabilistic approaches the empirical approach vanished into oblivion – to such an extent that mention of it disappeared from almost all textbooks after 1950.

Deterministic methods:

The maximum expected flood peak is denoted as the *Probable Maximum Flood* (PMF) and is calculated by unit hydrograph principles on the presumption that the *Probable Maximum Precipitation* (PMP) is falling on a saturated catchment.

The biggest advantage of this approach is that physical factors, that play an important role in the flood process, form the basis of the analysis.

However, this approach has a serious drawback (due to a lack of adequate data), in that it has to resort to not yet verified or unverifiable hypothesis and average rainfall coefficients – for example:

- storm rainfall areal reduction factor
- transposition of storms
- storm losses
- validity of unit hydrograph principles in case of extreme flood conditions

Furthermore, the PMP has basically the same shortcomings as described under the Regional Empirical methods.

The consequence is that the cumulative error may reach the magnitude of the PMF itself (Kovacs, 1988)

Probabilistic methods:

The maximum flood is normally associated with a very low probability, most often P = 0.0001 (recurrence interval of 10 000 year) or even P = 0.00001.

Principle shortcomings are:

- The representativeness of a relatively short period of flow data is unknown.
- The theoretical probability distributions vary most of the time vastly in their extrapolated estimates – and the preference for one particular method cannot be established. (These particular shortcomings will be addressed to a large degree when the probabilistic analysis is dealt with)

Regardless of the above, the results of the probabilistic approach should be more realistic in large catchments where the increase of flood peaks is limited by natural storage over extensive flood plains. Record lengths should preferably be longer than 50 years.

Concluding:

Francou and Rodier (1967) revived the empirical approach in 1967. Kovacs (1980) was convinced that, after carrying out flood frequency analyses at more than 100 dam sites, the probabilistic and deterministic approaches frequently resulted in extremely unrealistic and inconsistent figures. He investigated the Francou-Rodier empirical approach, to develop an approach that would yield more realistic and consistent maximum flood peaks.

6.2 Regional Maximum Flood (RMF)

Francou and Rodier (1967) of the Hydrological Services of *Electricité de France* compiled a catalogue of about 1200 maximum flood peak discharges, representing most regions of the world. They plotted the flood peaks against corresponding catchment areas. Francou and Rodier developed a family of enveloping curves for hydrological homogeneous regions. They found that, for catchments larger than about 100 km^2, regional envelope curves tend to be straight and converge towards a single point representing the approximate total drainage area and mean runoff of all the rivers in the world.

The Francou-Rodier diagram of flood peak classification is shown in Figure 6-1.

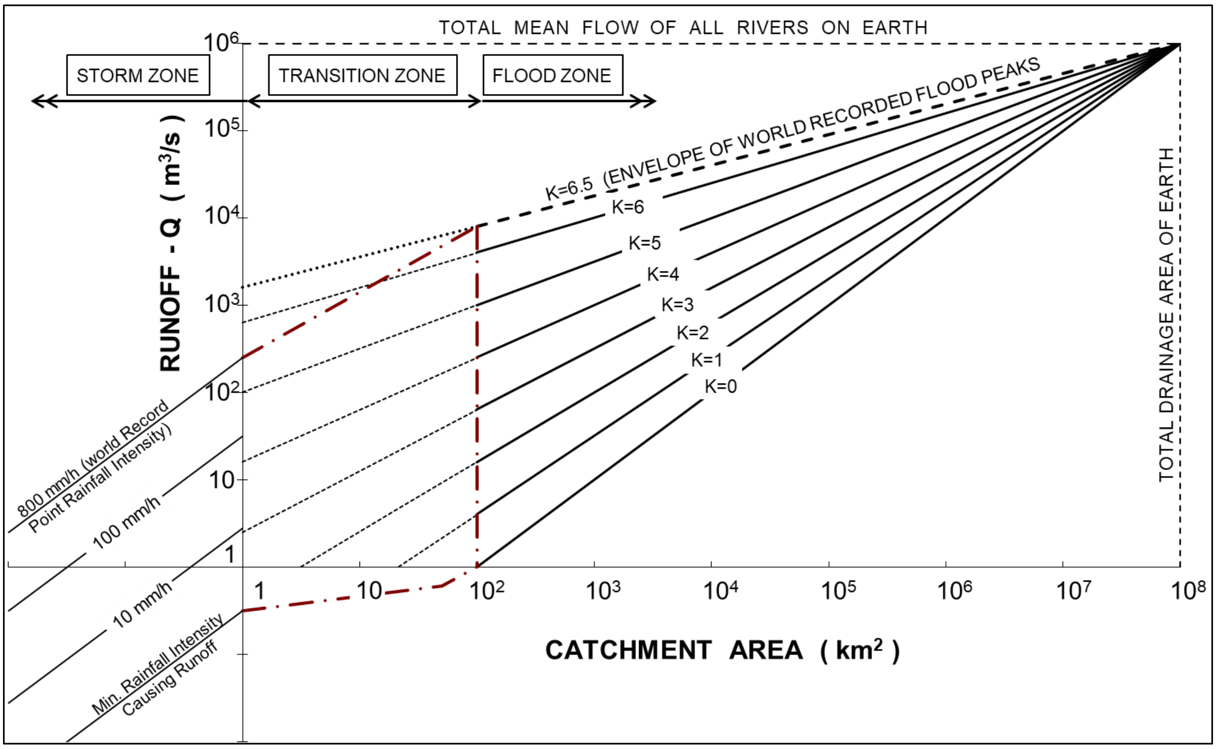


Figure 6-1: Regional Maximum Flood Peak Classification (re Francou-Rodier)

Francou (1968) made the following remarks (with reference to Figure 6-1):

- The diagram consists of three zones: the flood zone, the storm zone and the transition zone.
- In the **storm zone** (A < 1 km^2) the peak discharge depends only on rainfall intensity. For A = 1 km^2 the discharge is $Q = 0.278*\mathbf{i}$, where **i** is the maximum 15-minute rainfall intensity in mm/h (15 min. is the approximate time of concentration in a 1 km^2 catchment).

- In the **flood zone** (A > 100 km^2) the flood peak depends both upon storm rainfall (intensity, area, and duration) and catchment characteristics. This is the zone of converging envelope lines which are described by the well-known Francou-Rodier formula:

$$\frac{Q}{Q_0} = \left(\frac{A}{A_0}\right)^{1-0.1K}$$

with:

Q	-	required flood peak in m^3/s
Q_0	-	10^6m^3/s ≈ the mean annual runoff of all the rivers of the world
A	-	area of the catchment in km^2
A_0	-	10^8km^2 ≈ the total drainage area on earth
K	-	regional characteristic coefficient expressing relative flood peak magnitude

- Between the flood and storm zones lies a zone, called the **transitional zone**, where the envelope lines are supposed to provide a smooth transition between the regional 15 minute point rainfall discharge and the regional K-envelope curve in the flood zone.

Kovacs (1988) undertook a similar analysis for Southern Africa. He quoted the following typical maximum values for *K:*

2.0 to 3.0	-	Tropical Africa	
3.0 to 4.0	-	Central Europe, UK, USSR, Canada	
4.0 to 5.0	-	Argentina, Uruguay, most parts of USA	
5.0 to 5.5	-	Mediterranean Europe	
5.5 to 6.0	-	Madagascar, New Zealand, India	
6.0 to 6.5	-	Far East, Central America, Texas	
< 3	-	Southern Africa:	Kalahari
4.5 to 5.0	-		Highveld
5.0 to 5.5	-		South Eastern Coast belt

In Figure 6-2 world record flood peaks (Kovacs, 1988), as in 1960 and 1984, as well as South African record flood peaks, as in 1960 and 1988, are plotted against catchment area. Kovacs remarked that the world record peaks seem to have stabilised between K = 6.0 and K = 6.5 which is an indication that the sample is fairly complete.

<u>Note:</u> South African record flood peaks, as in 2000 (Eline) were added to Figure 6-2. The flood peaks, as in 1960 and 1988, are not yet verified and were just copied from TR137 (Kovacs, 1988).

The envelope of South African flood peaks has moved upwards during the last 28 years from K = 5.2 to K = 5.6, not the least because the sample size was much larger in 1988.

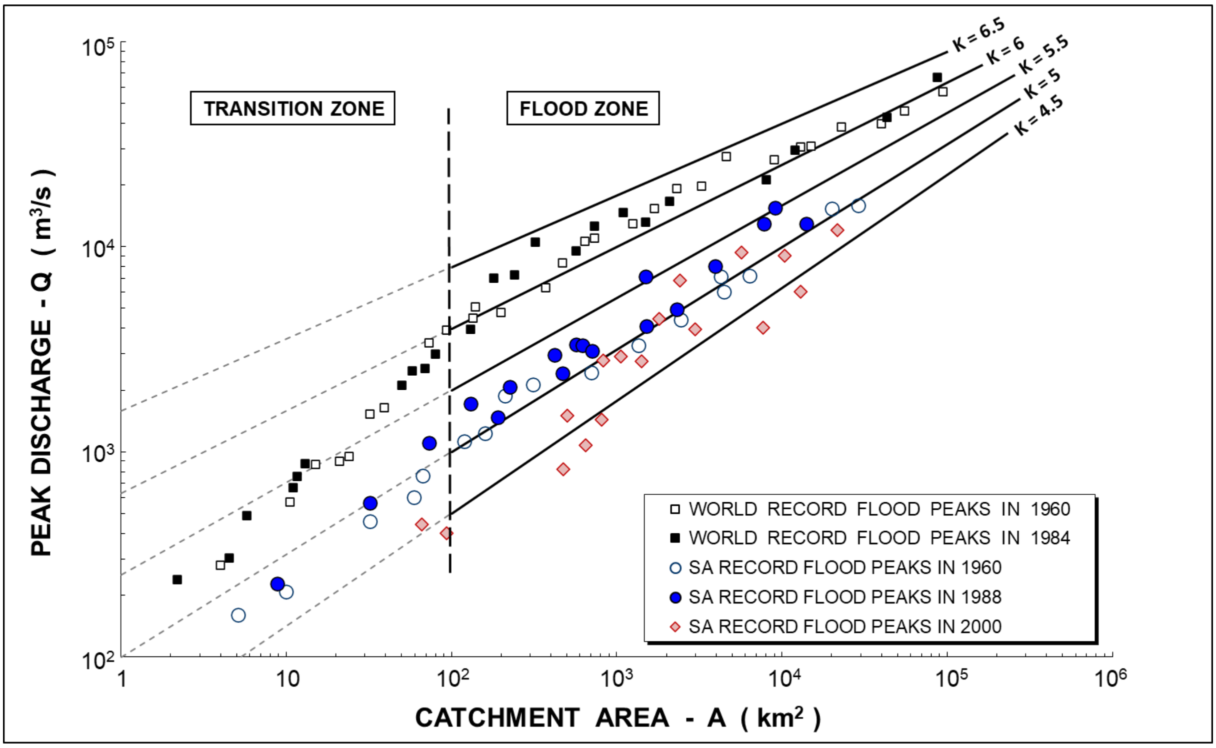


Figure 6-2: World- and South African Record Flood Peaks

The trends of the sets of data are strikingly similar; in the flood zone the points are well aligned with the direction of the corresponding K lines and the change from the transition zone to the flood zone is clearly visible between A = 100 km^2 and A = 200 km^2

(Kovacs found that for K-values of 4 and lower the transition zone increase to between 300 and 500 km^2 – see Table 6-1).

Kovacs (1988) based the K-values for Southern Africa on maximum flood peaks recorded at more than 519 sites, of which 354 correspond to South Africa and 165 were recorded in neighbouring countries (at some sites since 1856). Eight maximum flood peak regions were delimited by a joint consideration of K, maximum observed 3-day rainfall, catchment characteristics and recorded flood peaks. He recommended K-values for South Africa as shown in Figure 6-4; enlarged in the Appendices (Figure A-9: *RMF regions for South Africa*).

Figure 6-3 is an example of how the available data was plotted and how Kovacs determined the K-regions for Southern Africa.

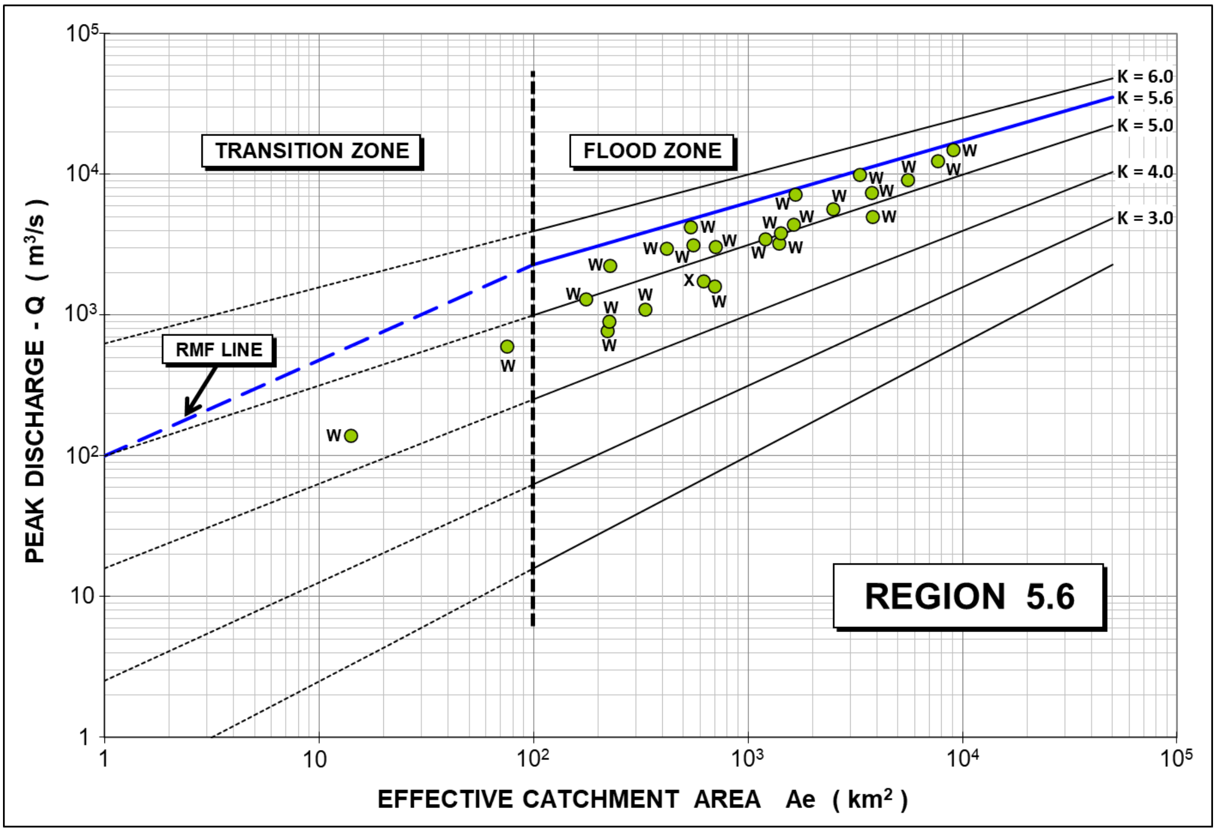


Figure 6-3: Envelope Curve for K-Region 5.6 in Southern Africa

Equations to calculate the RMF in the Transition zone, as well as simplified equations for the Flood zone, for each K-value (region) are given in Table 6-1.

Table 6-1: RMF Equations for designated regions in Southern Africa

Region	Transition zone		Flood zone	
	RMF (m³/s)	Areal range (km²)	RMF (m³/s)	Areal range (km²)
2.8	$30A_e^{0.262}$	1 - 500	$1.74\ A_e^{0.72}$	500 - 500 000
3.4*	$46.9A_e^{0.301}$	1 - 450	$5.25\ A_e^{0.66}$	450 - 500 000
4	$70A_e^{0.34}$	1 - 300	$15.8\ A_e^{0.60}$	300 - 300 000
4.6	$100\ A_e^{0.38}$	1 - 100	$47.9\ A_e^{0.54}$	100 - 100 000
5	$100\ A_e^{0.50}$	1 - 100	$100\ A_e^{0.50}$	100 - 100 000
5.2	$100\ A_e^{0.56}$	1 - 100	$145\ A_e^{0.48}$	100 - 30 000
5.4	$100A_e^{0.62}$	1 - 100	$209\ A_e^{0.46}$	100 - 20 000
5.6	$100A_e^{0.68}$	1 - 100	$302\ A_e^{0.44}$	100 - 10 000

*Note: Region 3.4 differs from Kovacs (1988) for programming purposes; results very slightly affected, though

Take Note: "The equations listed in Table 6-1 enable the instant determination of RMF if the geographic position of the site and its effective catchment area are known. The method is expected to render the best results for catchment sizes approximately between 300 km^2 and 20 000 km^2. In both the small and large catchments, one is confronted with the inherent weakness of all regionally based methods, namely that the particular characteristics of these catchments cannot be easily expressed by one common regional factor, in this case K_e." (Kovacs, 1988)

6.3 Application

The RMF should be calculated as follows (re Kovacs, 1988):

- Determine the geographical position and the effective catchment area
- Determine the Francou-Rodier K-value from Figure 6-4 keeping in mind that:
  - Generally, a site located in a given K-region is characterised by the K-value of that region, even if parts of the catchment extended into another K-region. However, Kovacs (1988) stated that if a site is located on or about a regional boundary the average K-value, or where particularly justified, the higher of the two values may be adopted.
  - In smaller to medium catchments, under special conditions, the K-value can either be lowered – for instance, very permeable conditions like dolomite areas, very flat catchments, low relative rainfall conditions, etc, - or it can be increased - for instance, very steep catchments, high relative rainfall conditions, etc.
  - Large rivers, which flow across several K-regions have distinct characteristics which may differ substantially from the K-region in which the site is situated – K-values have been adjusted along these rivers as shown on Figure 6-4 (also Appendix Figure A-8).
  - As a rule of thumb, the RMF should not be reduced because of upstream dams except under exceptional circumstances, such as where the impounding capacity exceeds the RMF volume coupled with long duration storm rainfall (typically > 3 day rainfall events).
  - Calculate the RMF from Table 6-1 or by applying the fundamental Francou-Rodier formula (where applicable)

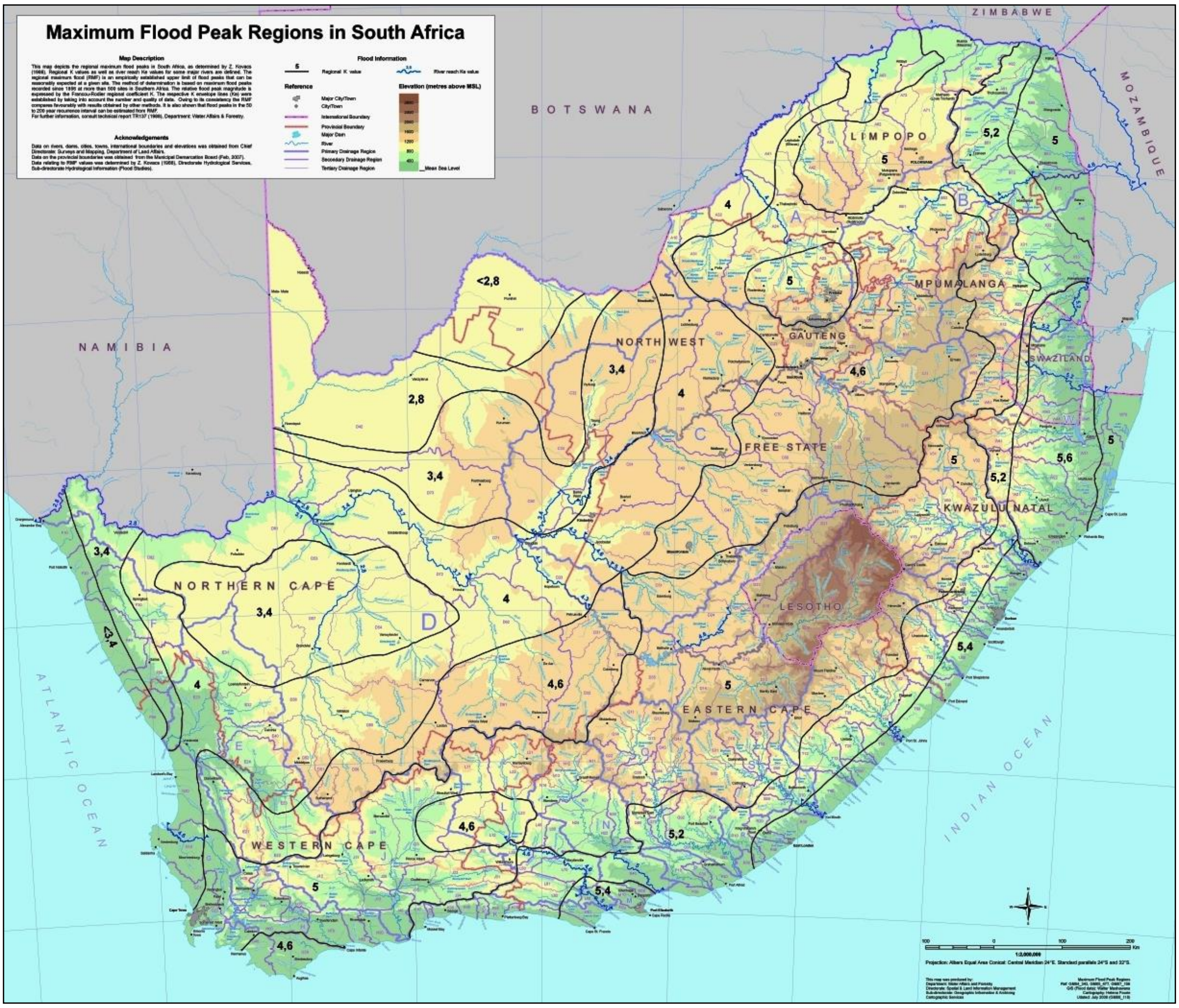


Figure 6-4: RMF regions for South Africa

6.4 Concluding remarks from TR137

The following closing remarks reiterate the concluding observations in Technical Report 137 by Kovacs (1988):

- In a general sense all scientific methods that are able to say something about nature are empirical, i.e. they are contingent and revisable. The need for revision arises if the calculation results are consistently refuted by observations.
- The empirical methods in the proper sense of the word, such as the RMF, rely on *in situ* observations to a higher degree than do other methods. The frequency of their revision will depend first of all on the representativeness of the data base.

- The data base of maximum flood peaks observed in South Africa used in Kovacs (1988) is a great improvement in comparison to the catalogue of the original 1980 report. The regional boundaries, shown in Figure 6-4, should require adjustments only if the respective K-value were consistently exceeded by more than $\Delta K = 0.1$.
- By considering the size of the database, the maximum attained K-value in Madagascar (5.78), where extreme floods are caused by frequent tropical cyclones, as well as presuming no changes in the climate, it is believed that regions 5 to 5.6 might require but modest adjustments in future whereby the increase in K-value will be limited to $\Delta K = 0.2$.

  The most likely areas for this to happen are the South Western Cape and adjacent Karoo where a K-value of 5.2 is possible and the Mpumalanga Lowveld where the data-base is small and which could be affected by tropical cyclones.
- In the dryer western and north-western areas of South Africa the data base is sparse, thus there is a greater chance for future modifications, particularly in Namaqualand where no data are available yet and the western part of Limpopo Province. However, it is not likely that the K-values recommended for certain reaches of large rivers, such as the Lower Orange, Vaal, Hartbees, Harts and Limpopo, should soon be changed.

6.5 Approach adopted by the DWS

Observations from numerous site-specific analyses, also considering the suggested screening criteria from the SANCOLD (1991) guidelines, led to the observation that the estimated 0.01% (10 000 year) flood peak is a good indicator for the maximum expected flood in a catchment (correlate very well with RMF estimations at various sites). FS-DWS, consequently, adopted the following approach in order to evaluate the effective site-specific K-values for the regional maximum flood (RMF) and safety evaluation flood (SEF), respectively labelled as K_{Site} and K_{SEF}:

- Determine the K-value for the forecasted 0.01% (10 000 year) flood peak ($K_{0.01\%}$).
- Analytically evaluate $K_{0.01\%}$, also considering the regional K-value (K_{RMF}), observed flood peaks and any other relevant information. Subsequently, determine a 'site-specific' K-value (K_{Site}) that is sound and consistent with the site-specific analysis.
- The site-specific K_{SEF} is determined as; $K_{SEF} = K_{Site} + \delta$ (where $0.1 \leq \delta \leq 0.2$)

7. BASIC STATISTICAL CONCEPTS

7.1 Introduction

Engineers and Hydrologists are often confronted with the problem of estimating the magnitude of floods, or the severity of droughts. As these are natural events which occur randomly, neither can be established with absolute certainty. It is possible, however, to quantify the measure of uncertainty by employing concepts and methods of probability. Statistical analyses provide powerful tools with which the probability of occurrence of particular events can be estimated.

The main objective of statistical analyses is to make "some sense" out of collected data. This is achieved by summarising the data, estimating certain parameters, and then choosing an appropriate theoretical distribution with which probabilities can be calculated; this technique is known as statistical inference.

Data can be summarised by using numerical- and/or graphical methods.

7.2 Numerical portrayal of data

Numerical values to describe data are more commonly used, since the calculations are fairly easy and the results can be used in formulas to calculate probabilities of occurrence, etc.

Numerical values most often used to describe data are the central values (mean, mode and median), the measures of dispersion (variance, standard deviation and coefficient of variation) and the measure of skewness (skewness coefficient).

Data

As an example, the annual maximum series (AMS) flood peak record at Grootdraai dam on the Vaal river is shown on the next page, dating from 1904 to 2020. Data at Grootdraai dam itself exist since 1979 - prior to that, data from gauging weir C1H001 (with flood section C1H014) were used to augment the record (Table 7-1).

Table 7-1: AMS flood peak record at Grootdraai Dam in the Vaal River

Vaal river at Grootdraai dam (1904/1905 - 2019/2020)											
Hydro year	Q_{max} m^3/s	Hydro year	Q_{max} m^3/s	Hydro year	Q_{max} m^3/s	Hydro year	Q_{max} m^3/s	Hydro year	Q_{max} m^3/s	Hydro year	Q_{max} m^3/s
C1H001/014		1924	121	1944	1 059	1964	243	1983	106	2003	212
1905	129	1925	565	1945	233	1965	538	1984	282	2004	255
1906	274	1926	105	1946	405	1966	43	1985	144	2005	215
1907	901	1927	213	1947	260	1967	766	1986	414	2006	1 900
1908	495	1928	164	1948	340	1968	148	1987	126	2007	855
1909	987	1929	371	1949	121	1969	131	1988	651	2008	825
1910	835	1930	592	1950	475	1970	276	1989	420	2009	1 300
1911	1 456	1931	122	1951	218	1971	127	1990	390	2010	1 335
1912	262	1932	144	1952	344	1972	900	1991	177	2011	765
1913	116	1933	57	1953	556	1973	64	1992	121	2012	160
1914	118	1934	667	1954	164	1974	357	1993	145	2013	350
1915	391	1935	251	1955	1 210	1975	1 851	1994	286	2014	1 275
1916	452	1936	581	1956	693	1976	610	1995	300	2015	310
1917	332	1937	586	1957	743	1977	893	1996	2 135	2016	315
1918	987	1938	445	1958	474	1978	184	1997	665	2017	1 265
1919	429	1939	1 011	1959	437	Grootdraai		1998	618	2018	390
1920	95	1940	515	1960	177	1979	163	1999	367	2019	169
1921	597	1941	568	1961	484	1980	252	2000	1 045	2020	885
1922	849	1942	221	1962	341	1981	245	2001	568		
1923	1 337	1943	307	1963	231	1982	143	2002	84		

Mode

The *mode* is the value that occurs the largest number of times

From example:		
	$\boldsymbol{Mode\ =\ 121.0\ m^3/s}$	(natural data)
	$\boldsymbol{Mode_{log} = 2.0828\ \rightarrow\ Mode = 121.0\ m^3/s}$	(log-transformed data)

Median

The *median* is the value of a random variable at which values above and below it are equally probable, i.e. the value of the random variable at which the cumulative frequency is 0.5.

From example:		
	$\boldsymbol{Q_{50\%}\ =\ 362.0\ m^3/s}$	(natural data)
	$\boldsymbol{logQ_{50\%} = 2.5587\ \rightarrow\ Q_{50\%} = 362.0\ m^3/s}$	(log-transformed data)

Arithmetic mean

The *sample arithmetic mean* of a set of values, $x_1, x_2, x_3, \ldots\ldots x_n$ is given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

where

n = number of observations and
$\bar{x}$ = sample arithmetic mean

From example:	$\boldsymbol{\bar{Q} = 494.6\ m^3/s}$ $\boldsymbol{\overline{logQ} = 2.5535 \rightarrow \bar{Q} = 357.7\ m^3/s}$	(natural data) (log-transformed data)

Variance, Standard deviation and Coefficient of Variation

The *standard deviation* gives an indication of the shape of the distribution. A small *standard deviation* indicates that most of the values are closely spaced around the *mean*.

The sample *standard deviation* is the positive square root of the *variance*:

$$S = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

From example:	$\boldsymbol{S = 413.5 m^3/s}$ $\boldsymbol{S = 0.3606 \rightarrow S = 2.294\ m^3/s}$	(natural data) (log-transformed data)

This measure of dispersion is sometimes expressed in terms of the *arithmetic mean*. A useful non-dimensional measure of dispersion is the *coefficient of variation*, given by:

$$COV = \frac{S}{\bar{x}}$$

From example:	$\boldsymbol{COV = 0.8361}$ $\boldsymbol{COV = 0.1412}$	(natural data) (log transformed data)

Skewness and Kurtosis

These two statistics are commonly referred to as shape parameters of a distribution. Skewness is a measure of symmetry in a distribution. Kurtosis originally was thought to measure the peakedness of a distribution. Although this is considered as its definition in many places, it is a misconception. Skewness and kurtosis actually relate to the tails of the distribution. It is discussed in more detail below.

<u>*Skewness*</u>

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. Skewness essentially measures the <u>relative size of the two tails</u>. A distribution, or data set, is symmetric (skewness of zero) if it looks the same to the left and right of the centre point.

Skewness is defined as:

$$g = \frac{1}{n}\sum_{i}^{n}\frac{(x_i - \overline{x})^3}{S^3}$$

To account for the sample size, the sample *coefficient of skewness* is given by:

$$g = \frac{n}{(n-1)(n-2)}\sum_{i}^{n}\frac{(x_i - \overline{x})^3}{S^3}$$

<u>From example</u>:	$\boldsymbol{g = 1.6232}$ $\boldsymbol{g = -0.0779}$	(natural data) (log transformed data)

Negative values for the coefficient of skewness indicate data that are skewed to the left – which implies that the left tail of the distribution is longer; the mass of the data is therefore concentrated on the right of the distribution plot – clearly, the opposite applies for skewed to the right (see Figure 7-1 for illustration).

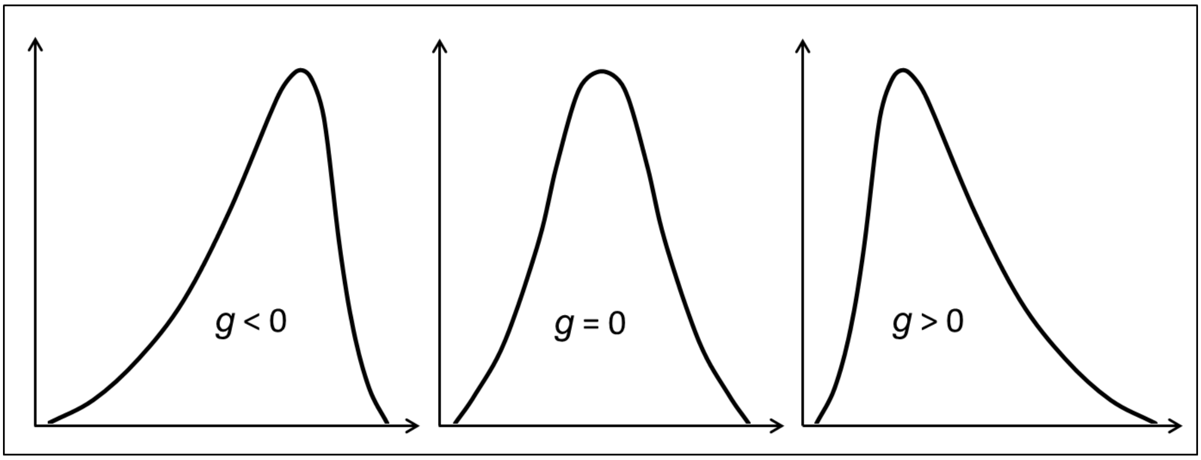


Figure 7-1: Illustration of negative and positive skewness distributions

Kurtosis

This part was added to the document, due to interest expressed and questions asked by course attendants.

As stated above, it is a misconception to refer to the kurtosis as a measure of "peakedness" of a distribution.

Westfall (2014) echoes this by stating that: *"Kurtosis tells you virtually nothing about the shape of the peak – its only unambiguous interpretation is in terms of tail extremity".*

McNeese (2016) referenced Wheeler (2011a) stating that "*Kurtosis was originally thought to be a measure the "peakedness" of a distribution. However, since the central portion of the distribution is virtually ignored by this parameter, kurtosis cannot be said to measure peakedness directly. While there is a correlation between peakedness and kurtosis, the relationship is an indirect and imperfect one at best.*"

He defines kurtosis as: *"The kurtosis parameter is a measure of the combined weight of the tails relative to the rest of the distribution."*

Kurtosis is defined as:

$$kurt = \frac{1}{n}\sum_{i}^{n}\frac{(x_i - \overline{x})^4}{S^4}$$

At this point one should be very careful: if the above equation is used, the kurtosis for a normal distribution equals 3 – however, most software packages use the equation below:

$$kurt = \left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i}^{n}\frac{(x_i - \overline{x})^4}{S^4}\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

This formula has a dual purpose: It considers the sample size plus it subtracts 3 from the kurtosis. Consequently, the kurtosis of a normal distribution becomes 0. This is really the **excess kurtosis**, but most software packages refer to it as simply kurtosis.

Therefore, excess kurtosis is a measure of how a distribution's tails compare to that of the normal distribution. So, if a dataset has a positive kurtosis, it has more weight in the tails, relative to the normal distribution. If a dataset has a negative kurtosis, it has less weight in its tails compared to the tails of the normal distribution (see also Figure 7-2).

From example:	$\boldsymbol{kurt = 2.8741}$ $\boldsymbol{kurt = -0.5714}$	(natural data) (log transformed data)

A distribution is called Mesokurtic if ***kurt* ≈ 0**, Leptokurtic if ***kurt* > 0** and Platykurtic if ***kurt* < 0**. These names originate from the previous misconception around the statistical meaning of the kurtosis; Platy meaning 'broad', Lepto meaning 'slender', Meso simply meaning 'in the middle' and kurtosis (from the Greek: kurtos) meaning 'curved'.

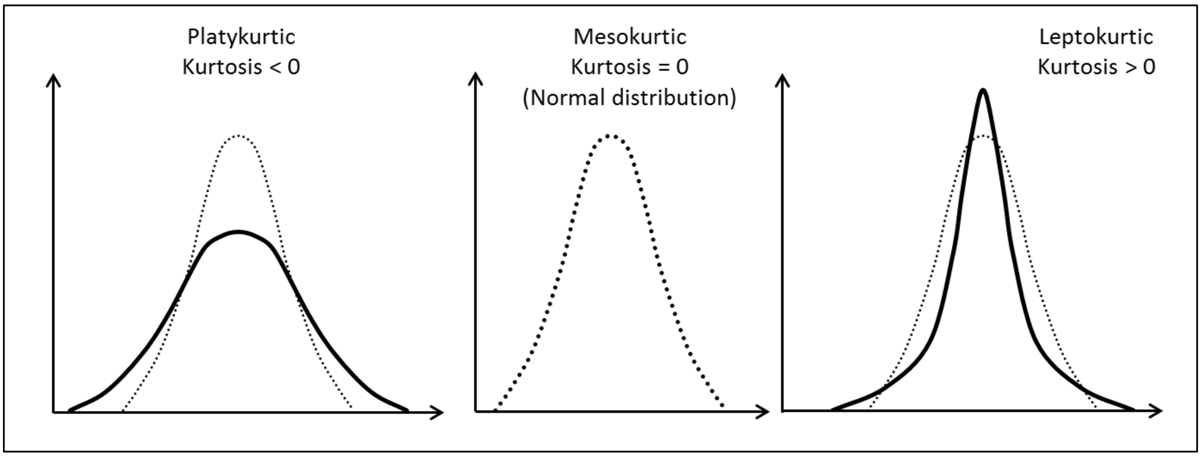


Figure 7-2: Excess Kurtosis

Some remarks by Wheeler (2011b) are worthy to take note of, though:

- To estimate the skewness with the same precision as the mean we need 6 times the data points (similarly, 24 times (!) the data points are needed to obtain the same accuracy in estimating the kurtosis). Consequently, Wheeler stated: "*This means that regardless of how many data we have, we will always have much more uncertainty in the shape statistics than we will have in the location and dispersion statistics. This limitation on what we can obtain from a collection of data is inherent in the statistics themselves, and must be respected in our analysis of the data*."
- He also, interestingly, pointed to a (forgotten?) simple visual technique that can provide valuable information at a glance: *"…the first step in any real-world analysis must always be an examination of the data for evidence of a lack of homogeneity. So, we return to the one completely general technique we have that can examine suspect data for evidence of a lack of homogeneity—the process behaviour chart.*"
- In his summary he concluded that: "*So, in consideration of the many problems with the shape statistics, I have to agree with Shewhart when he concluded that the location and dispersion statistics provide virtually all the useful information which can be obtained from numerical summaries of the data. The use of additional statistics such as skewness and kurtosis is superfluous."*

The accuracy comment is of course applicable to all sizes of data sets, but it is directed more to the accuracy of the population statistics. It might not be accurate to inform on the actual shape of the distribution, but can still be useful in sample statistics to point to the presence of possible outliers, etc.

7.3 Graphical depiction of data

The graphical depiction of data can assist the analyst to better understand the context of the data – a picture is worth a thousand words.

By summarising data as a histogram, or frequency distribution histogram, more information can be extracted. This will be best explained by means of an example:

On the next page the AMS flood peak record at Grootdraai dam on the Vaal river is repeated, for ease of reference (Figure 7-3). Natural data, as well as the log-transformed data, are presented. The explanation of the *Graphical Methods* is done for the natural data – the same applies to the log transformed data (also shown on next page). Numerical values are also repeated for comparison (see next section)

The figures below the AMS flood peak record depict the following:

First figure: In the first of the three figures the complete flood peak record is presented as a histogram, with the flood peak plotted on the ordinate and the year on the abscissa.

Second figure: The natural data were grouped into 200 m^3/s ranges and the number of flood peaks in each range was determined. The number of flood peaks in each range was divided by the total number of flood peaks to obtain the relative frequency of occurrence (which is a rough approximation of the probability of occurrence). The flood peak ranges were then plotted on the ordinate and the corresponding relative frequencies for each range, on the abscissa to produce the frequency distribution histogram.

Third figure: The cumulative frequency diagram, obtained by summing the relative frequencies of the various ranges, is the last of the three figures.

Vaal river at Grootdraai dam (1904/1905 - 2019/2020)

Hydro year	Q_{max} m³/s	$logQ_{max}$	Hydro year	Q_{max} m³/s	$logQ_{max}$	Hydro year	Q_{max} m³/s	$logQ_{max}$	Hydro year	Q_{max} m³/s	$logQ_{max}$
C1H001 and C1H014			1937	586	2.7679	1970	276	2.4409	2002	84	1.9243
1905	129	2.1106	1938	445	2.6484	1971	127	2.1038	2003	212	2.3263
1906	274	2.4378	1939	1011	3.0048	1972	900	2.9542	2004	255	2.4065
1907	901	2.9547	1940	515	2.7118	1973	64	1.8062	2005	215	2.3324
1908	495	2.6946	1941	568	2.7543	1974	357	2.5527	2006	1900	3.2788
1909	987	2.9943	1942	221	2.3444	1975	1851	3.2674	2007	855	2.9320
1910	835	2.9217	1943	307	2.4871	1976	610	2.7853	2008	825	2.9165
1911	1456	3.1632	1944	1059	3.0249	1977	893	2.9509	2009	1300	3.1139
1912	262	2.4183	1945	233	2.3674	1978	184	2.2648	2010	1335	3.1255
1913	116	2.0645	1946	405	2.6075	Grootdraai			2011	765	2.8837
1914	118	2.0719	1947	260	2.4150	1979	163	2.2122	2012	160	2.2041
1915	391	2.5922	1948	340	2.5315	1980	252	2.4014	2013	350	2.5441
1916	452	2.6551	1949	121	2.0828	1981	245	2.3892	2014	1275	3.1055
1917	332	2.5211	1950	475	2.6767	1982	143	2.1553	2015	310	2.4914
1918	987	2.9943	1951	218	2.3385	1983	106	2.0253	2016	315	2.4983
1919	429	2.6325	1952	344	2.5366	1984	282	2.4502	2017	1265	3.1021
1920	95	1.9777	1953	556	2.7451	1985	144	2.1584	2018	390	2.5911
1921	597	2.7760	1954	164	2.2148	1986	414	2.6170	2019	169	2.2279
1922	849	2.9289	1955	1210	3.0828	1987	126	2.1004	2020	885	2.9469
1923	1337	3.1261	1956	693	2.8407	1988	651	2.8136			
1924	121	2.0828	1957	743	2.8710	1989	420	2.6232			
1925	565	2.7520	1958	474	2.6758	1990	390	2.5911			
1926	105	2.0212	1959	437	2.6405	1991	177	2.2480			
1927	213	2.3284	1960	177	2.2480	1992	121	2.0828			
1928	164	2.2148	1961	484	2.6848	1993	145	2.1614			
1929	371	2.5694	1962	341	2.5328	1994	286	2.4564			
1930	592	2.7723	1963	231	2.3636	1995	300	2.4771			
1931	122	2.0864	1964	243	2.3856	1996	2135	3.3294			
1932	144	2.1584	1965	538	2.7308	1997	665	2.8228			
1933	57	1.7559	1966	43	1.6335	1998	618	2.7910			
1934	667	2.8241	1967	766	2.8842	1999	367	2.5647			
1935	251	2.3997	1968	148	2.1703	2000	1045	3.0191			
1936	581	2.7642	1969	131	2.1173	2001	568	2.7543			

Statistical properties	
Natural data (m³/s)	
Median	362.0
Mean	494.6
Standard deviation	413.5
Coeff. of variation	0.8361
Skewness	1.6232
log transformed data	
Median	2.5587
Mean	2.5535
Standard deviation	0.3606
Coeff. of variation	0.1412
Skewness	-0.0779

Natural data **log transformed data**

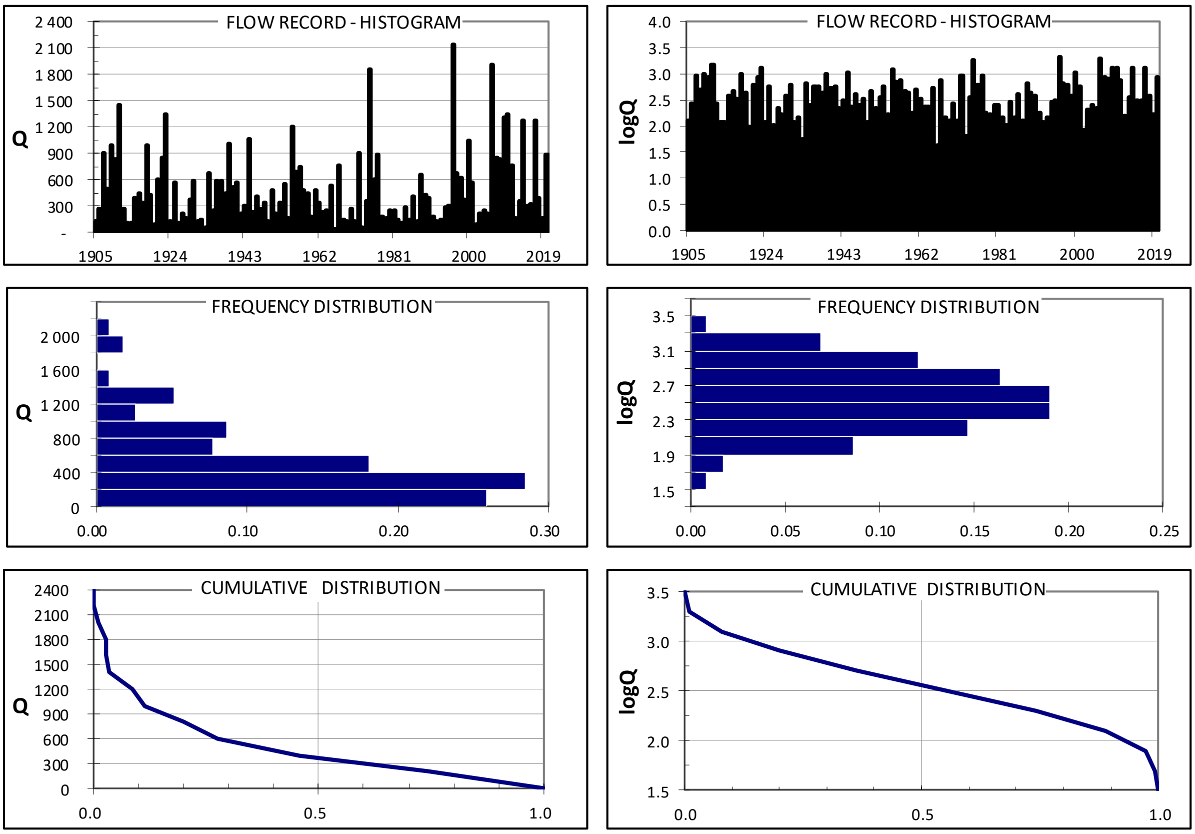


Figure 7-3: Vaal at Grootdraai

The information that can be obtained from the figures:

- The range of values (from just more than 40 m^3/s to just more than 2200 m^3/s)
- The range of values that occur most frequently (200 m^3/s to 400 m^3/s)
- The high variability of values about the mean, with the natural data. In case of the log transformed data, however, the variability seems much lower
- The distribution of flood peaks is asymmetrical (skew) about the mean (frequency distribution), however, in the case of the log transformed data the frequency distribution seems much more symmetrical
- 50% of the peaks are larger and 50% smaller than a flood peak of about 360 m^3/s (cumulative frequency distribution)

8. PROBABILITY DISTRIBUTIONS

8.1 Introduction

Public Agencies are very keen on amassing statistics - they collect them, add them, raise them to the n^{th} power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of those figures comes in the first instance from the village watchman, who just puts down what he damn well pleases.

Sir Josiah Stamp as quoted in Wonnacott and Wonnacott (1972)

Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting and analyzing data, as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

Spiegel (1961)

The old saying is that "figures will not lie," but a new saying is "liars will figure." It is our duty, as practical statisticians, to prevent the liar from figuring; in other words, to prevent him from perverting the truth, in the interest of some theory he wishes to establish.

Carroll D. Wright (1889 – statistician)

The basic issue in hydrology is the estimation of the probable magnitude of future events based on historical observations. As it is extremely unlikely that the historical records will be repeated in future, the statistical properties of the past records have to be examined, and applied to make estimates of the likelihood of events of given severity.

The most powerful tool available to hydrologists for this purpose is that of probabilistic analysis. The reason for the earlier mistrust in statistics was the failure on the part of the analysts to appreciate the weakness, as well as the power, of the basic statistical methodology. Another reason for continuing mistrust is the differences in nomenclature and approach. This is particularly true in the field of statistics where no two authors appear to use the same nomenclature, and even the form of the equations used for the same function may be different.

As with all the other methods – *you have to apply your mind to the problem, if you do not want to end up with misleading results!*

These notes will hopefully serve as a guide to carry out routine hydrological analyses, sensibly.

8.2 Data acquisition

The quotation from Josiah Stamp, in the beginning of this chapter, should not be ignored. The statistical analyses and the conclusions that are drawn from them can only be as accurate as the data on which they are based. Data published by the public agencies will have gone through a checking procedure, but may still contain errors. Use the data accordingly and in case of doubt contact the agency which produced the data.

The fundamental assumptions that have to be made, when performing analyses with collected data, is:

- Each observation is independent of previous and subsequent observations – this is reasonable to assume for annual flood maxima.
- The data are free of measurement errors. This should be verified in all cases, especially where major structures are involved.
- The data are identically distributed. This means that it can safely be assumed that the data came from a single parent population, which in turn implies a single type of rainfall producing meteorological phenomenon. This is clearly not the case in areas subject to infrequent tropical cyclones for example, which will require special treatment.

  Analysts should be aware of the strong possibility that many of the apparent anomalies in statistical analyses arise from the mixture of different meteorological phenomena and different states of antecedent conditions that determine the magnitude of runoff events.

8.3 Probability distributions

In the previous chapter probability distributions were introduced, using histograms and frequency diagrams.

If the top midpoints of the frequency histograms were joined the result would be a frequency polygon. A mathematical equation which best fits this polygon is called a probability density function (PDF) while the equation which fits the cumulative frequency polygon is called a cumulative distribution function (CDF) or simply the distribution function. An important property of the PDF is that the area beneath the curve must be unity, and similarly the CDF must have exceedance probability values in the range between zero and one.

Some of the more common probability distributions are discussed briefly, followed by a more in-depth review of the most applicable distributions for South African conditions.

Normal distribution

The normal distribution was first developed by de Moivre (1753). The distribution is widely used in hydrology as well as in other civil engineering applications, such as measurement errors (survey).

This distribution is symmetrical about the mean and is therefore only suitable for data where the skewness coefficient ***g*** is equal to, or close to zero. The distribution has certain deficiencies when used for examining the minima of a data set or when generating synthetic data (some negative flows may be generated). Notwithstanding the deficiencies, *the normal distribution of log transformed data (discussed in next paragraph) is still the most widely used distribution in hydrological analyses.*

Log Normal distribution (LN)

The log normal distribution is a normal distribution using the logarithms of the observed values. Hazen (1914) is credited with having observed that, while hydrological data are usually strongly skewed, the logarithms of the data have a near symmetrical distribution.

Exponential distribution

This is the simplest of the one-tailed distributions and is based on the equation $y = e^{-x}$ which is equal to 1.0 when $x = 0$ and decays rapidly to 0.00674 at $x = 5$.

It is seldom used directly in hydrological analyses but like the gamma distribution (below) it is incorporated in the more complex equations derived from it. It is the most common model for a partial duration series.

Gamma distribution

This is a strongly skewed distribution with a lower bound at zero, and makes use of the factorial series ***1/n!*** Where ***n*** need not be an integer. The gamma distribution is the distribution of the sum of a number of independent exponentially distributed random variables.

Pearson Type III distribution

This is essentially a Gamma distribution but with the mean displayed by a constant x_0 from the origin. It includes the normal distribution as a special case when the skewness equals 0.

Log Pearson Type III distribution (LP3)

This is the form in which the Pearson Type III distribution is most commonly used in hydrological analyses and is the distribution of the logarithms of the observed values. *It will fit most sets of hydrological data.*

Extreme value (EV) distributions

If the distribution of the events within the year is such that the tail of the distribution decays exponentially, then the family of extreme value (EV) distributions can be applied to the annual maxima. The most commonly used extreme value distributions are:

- EV1 (Gumbel) distribution: This distribution has a constant positive skewness coefficient of 1.13957 and should only be used when the data set has a value of ***g*** close to this figure.
- EV2 (Fretchet) distribution: This is a positively skewed distribution with ***g > 1.1396***. If the raw data are EV2 distributed then their logarithms will be EV1 distributed.
- EV3 (Weibull) distribution: The Weibull distribution is negatively skewed.

General extreme value distribution (GEV)

This is the generalised form of the above three extreme value distributions and is described in detail in the *Flood Studies Report (1975)*.

It is a family of three sub-types of distributions which are classified according to the value of the skewness coefficient **g**. This is a very flexible distribution and was the distribution recommended for use in the United Kingdom (*Flood Studies Report*, 1975) until it was replaced, in the 5-volume Flood Estimation Handbook (FEH), by the Generalised Logistic distribution (GLO) as the preferred distribution (Institute of Hydrology, 1999).

Wakeby distribution

The Wakeby distribution is one of the more recently introduced distributions, which was developed by Thomas (1978) for flood frequency analyses. The distribution has five parameters which makes it very flexible.

8.4 Probability analysis

A visual comparison between the data and the estimated probability distributions is indispensable – a picture is always worth a thousand words. The graphical representation of the data, on log-probability paper, will be explained, which will enable the analyst to compare the data with the results from the probability analysis.

Graphical representation

The objective in graphical estimation is to reduce the cumulative distribution function to a linear relationship by adjusting the horizontal scale of the graph. The horizontal scale can be linearised by expressing the exceedance probabilities (and also the corresponding return period) in units of standard deviations.

The most useful graphical representations are those with linear horizontal scales in units of standard deviations, but calibrated in exceedance (or non-exceedance) probabilities, as well as return periods. The two distributions commonly used for determining these horizontal scales are the Normal and EV1 distributions. The vertical scale, for the probability distribution methods that will be discussed in this chapter, is a logarithmic scale.

Plotting position (PP) of data

The PP technique is summarised as follows:

- Rank the data, starting with the largest value and give each value (flood peak) a rank number, starting with one (i = 1) at the largest flood peak. (*This ranking order is preferred, since it relates directly to an AEP, which directly relates to risk. If the flood peak data are sorted in ascending order (noted in some references) the probability value, assigned to a flood peak data point, indicates probability of non-exceedance.*)
- Calculate the PP (relating to probability of exceedance) corresponding to each flood peak, using the following equation (which is applicable to most existing PP techniques):

$$P_i = \frac{i + a}{n + 1 + 2a}$$

with

i	-	i^{th} order statistic (rank number)
a	-	‘unbiased’ plotting parameter
n	-	sample size
P_i	-	plotting probability of i^{th} order statistic

Some common PPs, recommended for use in hydrological analyses:

Plotting proponent	Parameter *a*	Suggested Distributions
Weibull (1939):	0	All
Adamowski (1981):	-0.25	All
Blom (1958):	-0.375	Normal
Cunane (1977):	-0.40	GEV, PIII, General
Gringorten (1963):	-0.44	Exponential, GEV and EV1
Hazen (1913), Foster (1936)	-0.50	Extreme Value

The Cunane and the Blom PPs are effectively very close to the average of all the PPs of this form. The Cunane is the suggested PP if two or more distributions are plotted on the same graph and is also currently the preferred PP by the DWS.

Having calculated the PP for each event (say annual flood peak) it is a simple matter to plot the flood peak versus PP (**Figure 8-1**).

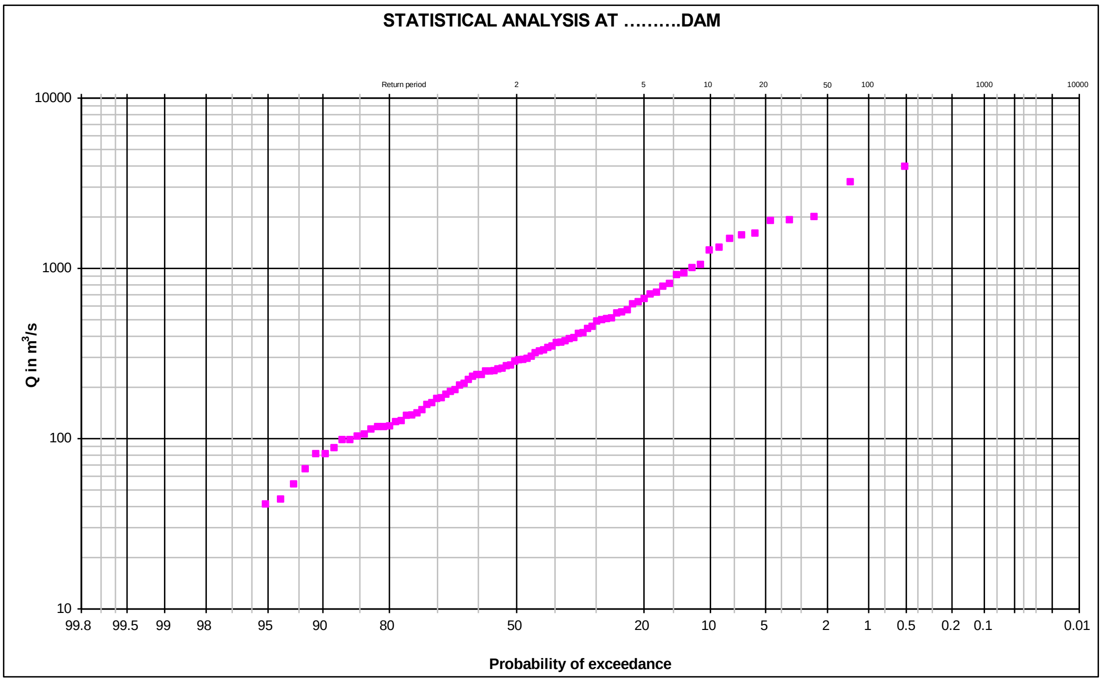


Figure 8-1: An example of a probability plot of flood peak data

Suitable Probability Distributions

Only the most suitable probability distributions, for flood frequency analyses, namely the ***Log Normal (LN), Log Pearson Type III (LP3)*** and the ***General Extreme Value (GEV)*** distributions will be discussed.

In the United States the LP3 distribution is accepted as being the most general and most objective of their best three distributions and is hence recommended for general use. The *GEV distribution* was preferred in the UK prior to 1999 and is one of the recommended distributions in Australia (Vogel, *et a*l 1993). Numerous flood frequency studies by the DWS confirmed that both these distributions are applicable for South African conditions. Similar to all other methods, probability distributions have their limitations and should never be applied without applying one's mind to the problem.

LP3 and LN Distributions

LP3 is a 3-parameter distribution. The LN distribution, a special case of the LP3 distribution with a skewness value of zero, is consequently only a two parameter distribution. However, studies world-wide have demonstrated the LN distribution to fit flood peak data more closely than any other two parameter distribution.

The general form of the prediction equation, for estimating Q_P, is:

$$\log Q_P = \overline{\log Q} + S_{log}.W_P$$

Where:

$\log Q_P$	-	The log of the required flood peak for exceedance probability P
$\overline{\log Q}$	-	The sample mean of the log transformed data
S_{log}	-	The sample standard deviation of the log transformed data
W_P	-	The standardised variate from **Table 8-1**

<u>Note:</u> for the LN distribution – determine W_P at $g = 0$ (a skewness of zero)

Table 8-1: LN and LP3 distributions: Standardised variate W_P

Skewness (g)	Probability of exceedance (%)											
	50	20	10	5	2	1	0.5	0.2	0.1	0.05	0.02	0.01
	Standardised variate W_P											
-1.4	0.225	0.832	1.041	1.168	1.270	1.319	1.352	1.380	1.394	1.404	1.412	1.416
-1.2	0.195	0.844	1.086	1.243	1.380	1.450	1.502	1.551	1.578	1.598	1.618	1.628
-1.0	0.164	0.852	1.128	1.317	1.492	1.589	1.664	1.741	1.787	1.824	1.862	1.885
-0.8	0.132	0.856	1.166	1.389	1.606	1.733	1.837	1.948	2.018	2.077	2.143	2.186
-0.7	0.116	0.857	1.183	1.423	1.663	1.806	1.926	2.057	2.140	2.213	2.296	2.351
-0.6	0.099	0.857	1.200	1.458	1.720	1.880	2.016	2.168	2.267	2.355	2.457	2.525
-0.5	0.083	0.857	1.216	1.491	1.777	1.954	2.108	2.282	2.398	2.502	2.625	2.708
-0.4	0.067	0.855	1.231	1.524	1.833	2.029	2.200	2.399	2.532	2.653	2.798	2.899
-0.3	0.050	0.853	1.245	1.555	1.889	2.104	2.294	2.517	2.668	2.808	2.977	3.096
-0.2	0.033	0.850	1.258	1.586	1.945	2.178	2.388	2.636	2.807	2.966	3.161	3.299
-0.1	0.017	0.846	1.270	1.616	2.000	2.252	2.482	2.757	2.948	3.127	3.349	3.507
0.0	0.000	0.842	1.282	1.645	2.054	2.326	2.576	2.878	3.090	3.291	3.540	3.719
0.1	-0.017	0.836	1.292	1.673	2.107	2.400	2.670	3.000	3.234	3.456	3.734	3.935
0.2	-0.033	0.830	1.301	1.700	2.159	2.472	2.763	3.122	3.378	3.622	3.930	4.154
0.3	-0.050	0.824	1.309	1.726	2.211	2.544	2.857	3.244	3.522	3.789	4.128	4.375
0.4	-0.067	0.816	1.317	1.750	2.261	2.616	2.949	3.366	3.667	3.957	4.327	4.598
0.5	-0.083	0.808	1.323	1.774	2.311	2.686	3.041	3.488	3.812	4.125	4.527	4.822
0.6	-0.099	0.799	1.328	1.797	2.359	2.755	3.132	3.609	3.956	4.294	4.728	5.048
0.7	-0.116	0.790	1.333	1.819	2.407	2.824	3.223	3.730	4.100	4.462	4.929	5.274
0.8	-0.132	0.780	1.336	1.839	2.453	2.891	3.312	3.850	4.244	4.631	5.130	5.501
1.0	-0.164	0.758	1.340	1.877	2.542	3.022	3.488	4.087	4.530	4.966	5.533	5.955
1.2	-0.195	0.733	1.341	1.910	2.626	3.149	3.660	4.322	4.814	5.300	5.935	6.410
1.4	-0.225	0.705	1.337	1.938	2.706	3.271	3.828	4.553	5.095	5.632	6.336	6.864

General Extreme Value (GEV) distribution

This is a 3-parameter distribution, which is the generalised form of the three extreme value distributions EV1, EV2and EV3.

The general form of the prediction equation can be written as follows:

$$Q_P = \overline{Q} + f_g . S \left[\frac{k . W_P + E(y) - 1}{[var(y)]^{0.5}} \right]$$

where:

Q_P	-	The required flood peak value for exceedance probability P
$\overline{Q}$	-	The mean of the flood peak sample data
f_g	-	f_g = 1 for g < 1.139566, otherwise f_g = -1
S	-	The standard deviation of the sample data
W_P	-	The standardised variate (Table 8-2)
k	-	Shape parameter, which is a function of g (Table 8-2)
$E(y), var(y)$	-	Moments of y, as a function of **k** (Table 8-2).

Note: $E(y)$ for g < 1.139566 will be shown as negative in most other publications (EV3).
This is because $E(y_2) = \Gamma(1+k)$ and $E(y_3) = -\Gamma(1+k)$
In order to use only one equation for both EV2 and EV3 (the one above), $E(y) = \Gamma(1+k)$.

<u>At g = 1.139566:</u> $k = 0$, $E(y) = 1$ and $var(y) = 0$
which will cause the equation, above, to be invalid. However, both GEV2 and GEV3 converge towards the following simple relationship for GEV1:

$$Q_P = \overline{Q} + S(0.78071 W_P - 0.45064)$$

It is suggested for hand calculations, because of interpolation difficulty, that the GEV1 distribution be used for values of **g** between 1.136 and 1.144.

Table 8-2: GEV distribution parameters

Distribution	Skew (g)	Probability of exceedance (%)												k	E(y)	var(y)
		50	20	10	5	2	1	0.5	0.2	0.1	0.05	0.02	0.01			
		Standardised variate W_P														
GEV3	-2.0	0.3062	0.7713	0.886	0.939	0.969	0.978	0.983	0.986	0.987	0.987	0.987	0.988	1.0000	1.0000	1.0000
	-1.8	0.3102	0.8057	0.938	1.002	1.040	1.053	1.060	1.065	1.066	1.067	1.067	1.068	0.9319	0.9731	0.8235
	-1.6	0.3142	0.8412	0.993	1.070	1.119	1.137	1.147	1.153	1.156	1.157	1.158	1.158	0.8617	0.9492	0.6729
	-1.4	0.3183	0.8791	1.052	1.145	1.209	1.233	1.248	1.258	1.261	1.264	1.265	1.266	0.7892	0.9286	0.5446
	-1.2	0.3224	0.9201	1.119	1.232	1.313	1.347	1.367	1.382	1.389	1.393	1.396	1.397	0.7149	0.9115	0.4360
	-1.0	0.3267	0.9646	1.193	1.330	1.435	1.481	1.511	1.534	1.545	1.552	1.557	1.560	0.6394	0.8986	0.3447
	-0.8	0.3311	1.012	1.275	1.442	1.577	1.641	1.685	1.721	1.738	1.750	1.760	1.764	0.5635	0.8899	0.2686
	-0.6	0.3356	1.063	1.365	1.568	1.743	1.831	1.893	1.949	1.977	1.998	2.016	2.025	0.4883	0.8859	0.2055
	-0.4	0.3400	1.117	1.463	1.707	1.933	2.053	2.143	2.227	2.273	2.307	2.340	2.358	0.4149	0.8866	0.1536
	-0.2	0.3443	1.172	1.566	1.860	2.147	2.309	2.436	2.563	2.635	2.693	2.750	2.783	0.3443	0.8917	0.1113
	0.0	0.3485	1.227	1.674	2.023	2.383	2.598	2.774	2.960	3.073	3.166	3.264	3.323	0.2777	0.9011	0.0772
	0.2	0.3524	1.281	1.783	2.193	2.637	2.917	3.156	3.422	3.590	3.735	3.896	3.999	0.2158	0.9141	0.0504
	0.4	0.3560	1.334	1.891	2.366	2.905	3.260	3.576	3.943	4.187	4.405	4.658	4.827	0.1594	0.9300	0.0299
	0.6	0.3593	1.384	1.996	2.538	3.179	3.619	4.025	4.515	4.855	5.170	5.551	5.816	0.1088	0.9479	0.0152
	0.8	0.3622	1.430	2.096	2.705	3.453	3.985	4.492	5.127	5.583	6.019	6.566	6.959	0.0640	0.9669	0.00579
	1.00	0.3649	1.473	2.189	2.864	3.721	4.350	4.966	5.763	6.353	6.934	7.686	8.243	0.0246	0.9864	0.000936
	1.10	0.3661	1.492	2.233	2.941	3.851	4.530	5.203	6.086	6.749	7.410	8.278	8.931	0.0067	0.9962	0.0000729
	1.136	0.3665	1.499	2.249	2.968	3.897	4.594	5.287	6.202	6.893	7.583	8.496	9.185	0.0006	0.99965	0.00000059
GEV1	1.1396	0.3665	1.500	2.250	2.970	3.902	4.600	5.296	6.214	6.907	7.601	8.517	9.210	0.0000	1.00000	0.00000000
GEV2	1.144	0.3666	1.501	2.252	2.973	3.908	4.608	5.306	6.228	6.925	7.622	8.544	9.242	-0.0007	1.00043	0.00000091
	1.15	0.3666	1.502	2.255	2.978	3.915	4.619	5.320	6.247	6.949	7.651	8.581	9.285	-0.0017	1.0010	0.00000501
	1.18	0.3670	1.507	2.267	3.000	3.953	4.672	5.391	6.345	7.069	7.797	8.765	9.500	-0.0067	1.0039	0.0000747
	1.28	0.3680	1.525	2.308	3.071	4.078	4.846	5.623	6.668	7.472	8.288	9.386	10.23	-0.0224	1.0135	0.000878
	1.4	0.3692	1.546	2.355	3.154	4.222	5.050	5.898	7.053	7.954	8.881	10.15	11.13	-0.0399	1.0247	0.00293
	1.6	0.3710	1.577	2.426	3.281	4.451	5.375	6.340	7.682	8.752	9.872	11.43	12.68	-0.0660	1.0427	0.00866
	1.8	0.3725	1.604	2.491	3.398	4.663	5.681	6.760	8.290	9.532	10.85	12.73	14.25	-0.0887	1.0597	0.0169
	2.0	0.3739	1.629	2.549	3.505	4.858	5.966	7.157	8.872	10.29	11.81	14.01	15.82	-0.1086	1.0756	0.0270
	2.5	0.3766	1.679	2.670	3.730	5.281	6.591	8.039	10.19	12.02	14.05	17.08	19.65	-0.1480	1.1106	0.0584
	3.0	0.3787	1.718	2.764	3.907	5.620	7.103	8.773	11.32	13.53	16.04	19.85	23.18	-0.1769	1.1392	0.0941
	3.5	0.3802	1.747	2.838	4.048	5.894	7.520	9.381	12.26	14.82	17.75	22.30	26.34	-0.1986	1.1627	0.1311
	4.0	0.3814	1.771	2.896	4.161	6.118	7.864	9.887	13.06	15.92	19.23	24.44	29.13	-0.2155	1.1821	0.1675
	4.5	0.3823	1.789	2.943	4.253	6.302	8.150	10.31	13.74	16.85	20.50	26.31	31.58	-0.2288	1.1983	0.2022
	5.0	0.3831	1.805	2.982	4.329	6.454	8.389	10.67	14.31	17.66	21.60	27.92	33.72	-0.2395	1.2117	0.2345
	5.5	0.3837	1.817	3.014	4.392	6.583	8.591	10.97	14.81	18.35	22.55	29.34	35.60	-0.2483	1.2232	0.2647
	6.0	0.3842	1.828	3.042	4.447	6.694	8.767	11.23	15.24	18.95	23.39	30.59	37.28	-0.2564	1.2340	0.2956
	6.5	0.3847	1.838	3.066	4.495	6.795	8.926	11.47	15.63	19.51	24.16	31.76	38.85	-0.2656	1.2468	0.3349

8.5 Some practical considerations

One of the great debates regarding statistical (probabilistic) analyses and distributions, is the confidence that can be attached to the predictions of the lower exceedance flood peaks.

The intention of this chapter is not to participate in this debate but rather to present the reader with some suggestions and pointers which might be of help when evaluating the results from such an analysis.

The application of any method, without applying one's mind to the problem, can be considered as the so-called "black box" methodology.

In other words, if the input and output of a method is not carefully considered and evaluated that method can be considered a 'black box'.

It is thus not the method that is at fault but the user

A big computer, a complex algorithm and a long time does not equal science

Robert Gentleman

Most common questions asked:

Q: How long is a long enough data record?

A: This question forms the basis of most doubts when a distribution is fitted to an observed flood peak record, especially to estimate lower exceedance probability flood peaks.

Many analysts claim that existing periods of observed flows are too short, to do proper probabilistic analyses and that they have much more confidence in the '*proven'* empirical or deterministic methods.

Proven against what? – do they really grasp what they are implying? These empirical and deterministic methods only attempt to predict runoff from rainfall, whereas flow records provide us with actual runoff. Furthermore, these methods were developed using statistical evaluations of sets of flow records – almost 50 years ago! Is it really possible to believe that these methods will be more reliable than up to date statistical analyses, today, with nearly 50 years of additional data?!

The main advantage of any flow record is that it provides the analyst with additional data. If a flow record is available, no matter how short, the runoff-from-rainfall is no longer completely unknown. Consequently, even with a relative short record, the performance of the empirical and deterministic model(s) can be measured against actual runoff values.

Also: a general perception exists that flood peaks will be estimated too low, at lower probabilities of exceedance, with a short record – the opposite is actually more probable!

Q: What is an outlier?

A: Sometimes a data set will have one or more observations with unusually large or unusually small values. These extreme values are called outliers.

Ball, *et al.* (2019) consider outliers as observations that are inconsistent with the general trend of the rest of the data. Frost (s.a.) stated that, while there is no strict statistical rule or mathematical definition to identify outliers, guidelines exist through which possible outliers can be identified. He emphasised that a sound knowledge of the subject-area and understanding of the data collection process is crucial to accurately identify outliers. He described five methods, including: (1) sorting data, (2) graphing data, (3) using Z-scores, (4) using the interquartile range and (5) hypothesis tests, to identify outliers in datasets, noting the advantages and disadvantages of each. Frost (s.a.) indicated that the biggest disadvantage of the Z-score approach is that a high outlier in the dataset inflates the mean and standard deviation.

However, the Z-score provides a measure to compare relative probabilities, associated with relative magnitudes of the data:

- Standardised values (Z-scores) can be used statistically to identify outliers.

$$Z_i = \frac{x_i - \overline{x}}{S}$$

- Since c. 100% of the data will be within 3 standard deviations of the mean, data with a Z-score higher than 3, or lower than -3, can be considered an outlier.

The problem that faces the analyst is what to do with that particular flood peak (outlier) when performing the analysis.

There is no simple answer to the problem – the solution probably lies in a combination of possibilities, like ignoring the outliers (or not), visually comparing the data with the estimated probability distributions, performing a partial series flood frequency analysis (as well) if sufficient additional data are available and various other aids, like historical data and palaeo-flood data (again, if available).

Conducting data analysis is like drinking a fine wine. It is important to swirl and sniff the wine, to unpack the complex bouquet and to appreciate the experience. Gulping the wine doesn't work.

Daniel B Wright (2003)

In essence the message is: "*Do not just accept a distribution fit, blindly –* ***apply your mind to the problem***".

<u>It is thus also important to appreciate your information</u>

It is important to be aware of the variety of information your flood peak record contains: e.g. how many "severe" flood events are represented in a particular record – an experienced flood hydrology engineer or hydrologist will immediately be able to tell you that (e.g.) 2012, 2011, 2008, 2007, 2006, 2000, 1996, 1988, 1987, 1984, 1981, 1974, 1959, 1945, 1937, 1925, etc. were years of high flood peak occurrences in the record, or not.

This kind of information (or knowledge) is vital when some sensible deduction has to be made from a series of results, especially when the validity of an outlier has to be assessed.

Do not put your faith in what statistics say, until you have carefully considered what they do not say. William W. Watt

<u>Q: What about historical flood peak and palaeo-flood data?</u>

<u>A:</u> Historical flood peak data are considered to be data collected prior to the continuous flow monitoring period which, if added to the flood peak record from continuous monitoring, can increase the total period of observation.

Palaeo-flood data, on the other hand, should rather not be considered in the estimation of any probabilistic distribution. It can, however, provide an 'upper boundary' to assist in determining the maximum expected flood peak.

Historical flood peak data, as a rule, are very scarce and difficult to get hold of. Although quite useful to increase the period of observation, utmost care should be taken when these observations are included in the sample, or flood peak record.

The following line of approach is suggested, if historical observations are to be included in a statistical analysis:

- A lower limit for these observations must be established (not necessarily the lowest observed historical flood peak value), above which it will be reasonably certain that all flood peaks, higher than this value, are accounted for in the historical observations.
- All flood peaks, within the continuous record, higher than this lower limit, should also be identified.
- These *high flood peaks*, together with their historical counterparts, will be temporarily excluded from the combined flood peak record and the mean, standard deviation and skewness parameters of the remaining flood peaks are calculated.
- These estimated statistical parameters are assumed to be valid for the period of observation of no flow (that is excluding the number of years for which the *high flood peaks* were temporarily excluded).
- Ultimately the *high flood peaks* are added to the record and the adjusted mean, standard deviation and skewness parameters are estimated.

The impact of historical data on statistical parameters

- The reader is encouraged to consult appropriate literature if the need ever arises to add historical flood peak data to a continuous flow record, for statistical analysis purposes.
- For those who are interested, the impact of the suggested approach, on the adjustment of the mean, for a record with historical flood peak observations, is illustrated:

For observed flood peak record, lower than the identified lower limit for historical flood peak observations:

$$\overline{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i$$

After adding the high flood peaks from historical observations:

$$\overline{Q_h} = \frac{1}{N}\left(n'\overline{Q} + \sum_{i=1}^{n''} Q_{h_i}\right)$$

with:

N – Total observation period (years)

n – Continuous observation period (years)

n'' – Number of 'historical flood peaks' (high flood peaks)

n' – $n - n''$

$\overline{Q}$ – The mean of the flood peak sample data

$\overline{Q_h}$ – The mean of the flood peak sample data + historical data

Q_{h_i} – historical flood peak

Q: Can there be a 100 year flood event in the record, or not?

Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital

Aaron Levenstein (2010)

A: It is important to realise that it is indeed much more probable to experience an extreme flood event, than most people would think – as indicated in the next table.

For instance, the probability of observing a 1% probability of exceedance (100 year) flood peak in a 50 year observation period is almost 40%! – the probability that the peak could have been the 0.5% probability of exceedance flood peak is also just more than 22%, etc.

HHP	P	Observation (design) period (in years)							
		1	2	5	10	20	25	50	100
2	50%	50.0%	75.0%	96.9%	99.9%	100.0%	100.0%	100.0%	100.0%
5	20%	20.0%	36.0%	67.2%	89.3%	98.8%	99.6%	100.0%	100.0%
10	10%	10.0%	19.0%	41.0%	65.1%	87.8%	92.8%	99.5%	100.0%
20	5%	5.0%	9.8%	22.6%	40.1%	64.2%	72.3%	92.3%	99.4%
50	2%	2.0%	4.0%	9.6%	18.3%	33.2%	39.7%	63.6%	86.7%
100	1%	1.0%	2.0%	4.9%	9.6%	18.2%	22.2%	39.5%	63.4%
200	0.5%	0.5%	1.0%	2.5%	4.9%	9.5%	11.8%	22.2%	39.4%
500	0.2%	0.2%	0.4%	1.0%	2.0%	3.9%	4.9%	9.5%	18.1%

Subconsciously the perception will always exist that "*one cannot experience an event with a recurrence interval higher in value than one's own age*".

The problem stems (again) from relative short records combined with the PPs that we apply to our observed data – the so-called "outlier events" cannot be accommodated satisfactorily and generally causes some distributions to go astray.

Q: Which statistical method is the best for South Africa?

All models are wrong, but some are useful George E. P. Box

A: In the United States the LP3 distribution is accepted as being the most general and most objective of their best three distributions and is hence recommended for general use.

The Flood Studies Report of the U.K. gives preference to the *GEV distribution* and it is discussed in detail in the report. The GEV is also one of the recommended distributions in Australia (Vogel, *et a*l 1993).

Numerous flood frequency studies by the DWS confirmed that both these distributions are applicable for South African conditions. In addition, the FS_DWS observed that the LP3 distribution appears to be very sensitive to low flows/outliers. This is most probably the reason for the suggested censoring of Potentially Influential Low Flows (PILFs), in the US (England, *et al.*, 2019) – supported by Ball, *et al.* (2019) in Australia. In statistics, censoring implies that the value of an observation is only partially known. In the UK Reed (1999) proposes four possible treatments for outliers, of which an erroneous observation is the only reason considered for rejecting the outlier, since he considers it as "*bad practice*" to ignore outliers in FFA. This view is not supported in the DWS.

The GEV distribution, probably because it is an extreme value distribution, seems to be almost oblivious to low outliers. Both the LP3 and GEV are affected by high outliers but, in general, the GEV seems to be less affected by it, than the LP3.

Similar to all other methods, probability distributions have their limitations and should never be applied without applying one's mind to the problem.

A quote to conclude with (remember, many a true word hath been spoken in jest):

Statistics - A subject which most statisticians find difficult but in which nearly all physicians are experts. *Stephen Senn*

BIBLIOGRAPHY

Acocks JPH (1988)
Veld types of South Africa. Memoirs of the Botanical Survey of South Africa No.57, Botanical Research Institute, Department of Agriculture and Water Supply, Pretoria.

Adamson PT (1983)
Technical Report TR102, Southern African storm rainfall, Department of Environment Affairs, Pretoria

Alexander WJR (1982)
Course in Hydrological Analyses, University of Pretoria 1982

Alexander WJR (1990)
Flood Hydrology for Southern Africa, Department of Civil Engineering, University of Pretoria, prepared for SANCOLD

Ball J, Babister M, Nathan R, Weeks W, Weinmann E, Retallick M and Testoni I (ed.) (2019)
Australian Rainfall and Runoff: A Guide to Flood Estimation. © Commonwealth of Australia (Geoscience Australia).

Bauer SW and Midgley DC (1974)
A simple procedure for synthesising Direct Runoff Hydrographs, HRU Report 1/74, Department of Civil Engineering, University of the Witwatersrand, Johannesburg.

Campbell GV, Ward AD and Middleton BJ (1986)
An evaluation of hydrological techniques for estimating floods from small ungauged catchments, SRK (Civil) Inc., Johannesburg, SRK Report No. CI 3392/10.

England JF, Cohn TA, Faber BA, Stedinger JR, Thomas WO, Veilleux AG, Kiang JE and Mason RR (2019)
Guidelines for determining flood flow frequency - Bulletin 17C (ver. 1.1, May 2019). U.S. Geological Survey Techniques and Methods, book 4, chap. B5 148 pp.

Flood Studies Report (1975)
National Environment Research Council, London, United Kingdom. 1975

Francou J and Rodier JA (1967)
Essai de classification des crues maximales. Proceedings of the Leningrad Symposium on floods and their computation. UNESCO, August 1967

Francou J (1968)
La prévision des crues. X^{e}Journées de l' Hydraulique, Tome I: Discussion de la Question I, Paris, 1968

Frost J (s.a.)
5 Ways to Find Outliers in Your Data. Statistics by Jim
[Online] Available: https://statisticsbyjim.com/basics/outliers/

Geoterra Image (Pty) Ltd, South Africa (2020)
South African National Land-Cover 2020 Dataset
Department of Environmental Affairs, South Africa

Institute of Hydrology (1999)
Flood Estimation Handbook (five volumes). Centre for Ecology & Hydrology, Wallingford, Oxfordshire, UK.

Kovacs Z (1980)
Maximum Flood Peak discharges in Southern Africa: An empirical approach. TR105. Department of Water Affairs, Pretoria

Kovacs Z (1988)
Regional Maximum Flood Peaks in Southern Africa. TR137. Department of Water Affairs, Pretoria

Kovacs Z and Zürich HMG (compilers)
Flood Frequency Manual, Part I and II. Department of Water Affairs, Pretoria (unpublished)

Lynch SD (2004)
Development of a raster database of annual, monthly and daily rainfall for Southern Africa. WRC Report No.1156/1/04, Water Research Commission

MacVicar CN, de Villiers JM, Loxton RF, Verster E, Lamprechts JJN, Merryweather FR, le Roux J, van Rooyen TH and Harmse HJ von M (1977)
Binomial Soil Classification system for South Africa. Department Agricultural Tech. Serv., Pretoria.

McNeese B (2016)
Are the Skewness and Kurtosis Usefull Statistics?. SPC EXCEL, BPI Consulting LLC, https://www.spcforexcel.com/home

McPherson DR (1983)
Comparison of mean annual flood peaks calculated by various methods. Technical Report TR119. South African National Hydrology Symposium Proceedings. Department of Environment Affairs and WRC, Pretoria

Midgley DC, Pullen RA and Pitman WV (1969)
HRU report no. 4/69 –Design Flood Determination in South Africa, Department of Civil Engineering, University of the Witwatersrand, Johannesburg.

Midgley DC (1972)
HRU report no. 1/72 – Design Flood Determination in South Africa, Department of Civil Engineering, University of the Witwatersrand, Johannesburg.

Pitman WV and Midgley DC (1967)
Flood studies in South Africa: Frequency analysis of peak discharges. Transactions South African Institution of Civil Engineers. August 1967.

Pitman WV and Midgley DC (1971)
HRU Report No. 1/71 - Amendments to Design Flood Manual, HRU 4/69. Department of Civil Engineering, University of the Witwatersrand, Johannesburg.

Raudkivi AJ (1979)
Hydrology, University of Auckland, New Zealand. 1979

Reed, D.W. 1999.
Overview. Volume 1 of the Flood Estimation Handbook. Centre for Ecology & Hydrology, Wallingford, Oxfordshire, UK.

Reich BM (1962)
Soil Conservation service design hydrographs, Trans South African Inst. Civil Eng.

SANCOLD (1991)
Guidelines on Safety in Relation to Floods. Safety Evaluation of Dams, Report No.4. South African National Committee on Large Dams.

SANRAL (2013)
Drainage Manual 6th Edition. The South African National Roads Agency Ltd.

Schulze RE and Arnold H (1979)
Estimation of volume and rate of runoff from small catchments in South Africa, based on the SCS technique. University of Natal, Pietermaritzburg, Department Agricultural Engineering. CRU Report 8.

Schulze RE, Schmidt EJ and Smithers JC (1992)
SCS-SA User Manual. University of Natal, Pietermaritzburg, Department Agricultural Engineering. CRU Report 40.

SCWG (1991)
Taxonomic Soil Classification system for South Africa. Department Agricultural Development, Soil Classification Workgroup, SIRI, Pretoria.

Shaw EM (1988)
Hydrology In Practice.

Smithers JC and Schulze RE (2001)
Long Duration Design Rainfall Estimates for South Africa. School of Bioresources Engineering and Environmental Hydrology, University of Natal, Pietermaritzburg.

Sokolov AA Rantz SE and Roche M (1976)
Flood flow computation. Methods compiled from world experience. UNESCO Press. 1976

Thobejane A (2001)
An Evaluation of the Catchment Parameter Method (CAPA). A Project Report Submitted for Special Hydrology, BSc.Honours study, University of Pretoria.

Vogel RM, McMahon TA and Chiew FHS (1993)
Floodflow frequency model selection in Australia. *Journal of Hydrology 146:421-449.*

Weddepohl JP (1988)

Design Rainfall distributions for South Africa. A Project Report Submitted for MSc. Dissertation, University of Natal, Pietermaritzburg.

Westfall PH (2014)

Kurtosis as Peakedness, 1905 – 2014. R.I.P. The American Statistician, 68(3), 191–195. http://doi.org/10.1080/00031305.2014.917055

Wheeler DJ (2011a)

Problems with Skewness and Kurtosis, Part One; What do the shape parameters do? Qualitydigest. https://www.qualitydigest.com/inside/quality-insider-article/problems-skewness-and-kurtosis-part-one.html

Wheeler DJ (2011b)

Problems with Skewness and Kurtosis, Part Two; What do the shape statistics do? Qualitydigest. https://www.qualitydigest.com/inside/quality-insider-article/problems-skewness-and-kurtosis-part-two.html

WMO (2009)

Guide to Hydrological Practices. World Meteorological Organization.

Appendix

List of graphs, maps, and tables

Deterministic methods

Empirical methods

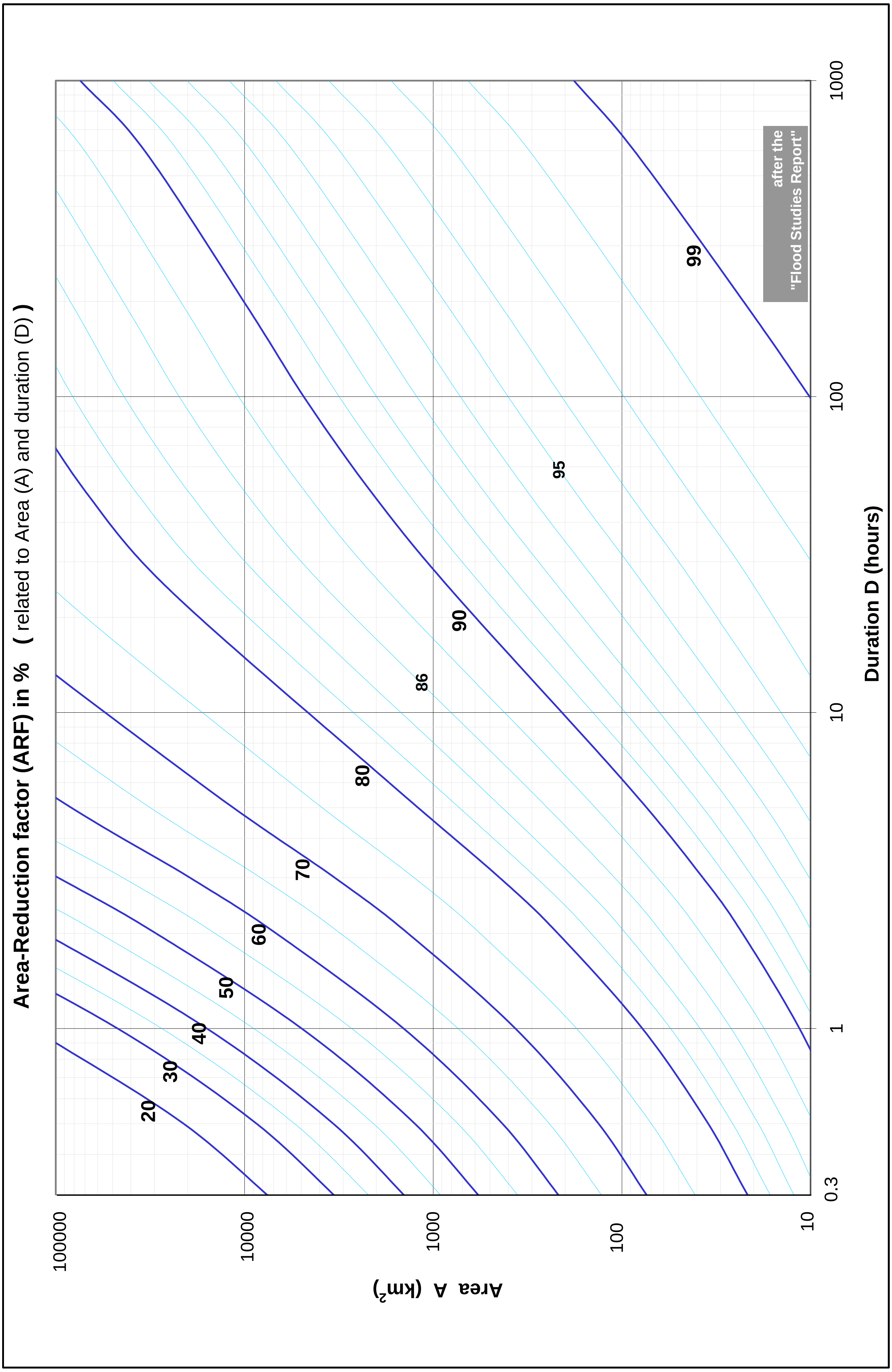


Figure A-1: ARF (Area Reduction Factor)

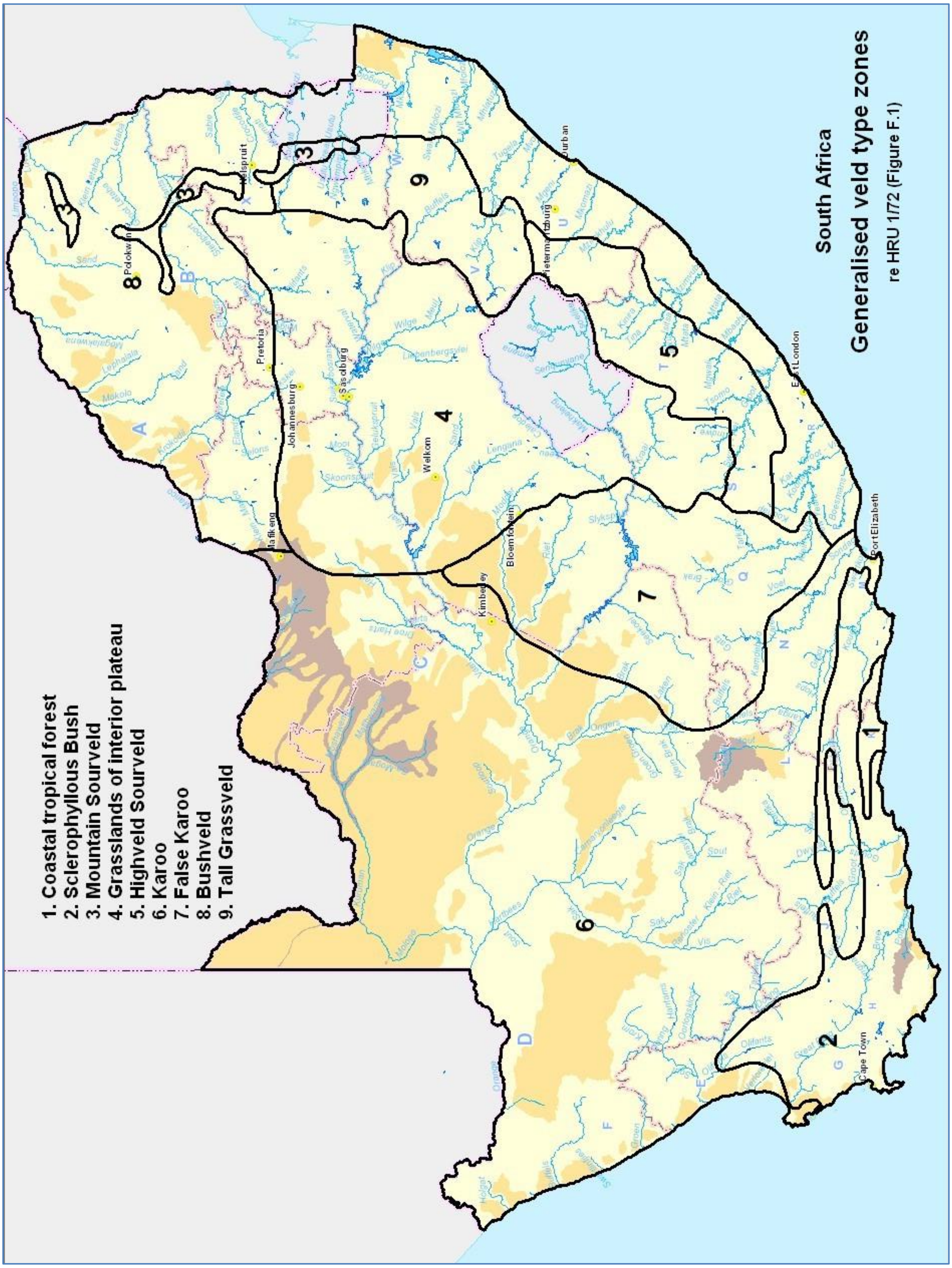


Figure A-2: Generalized veld type zones

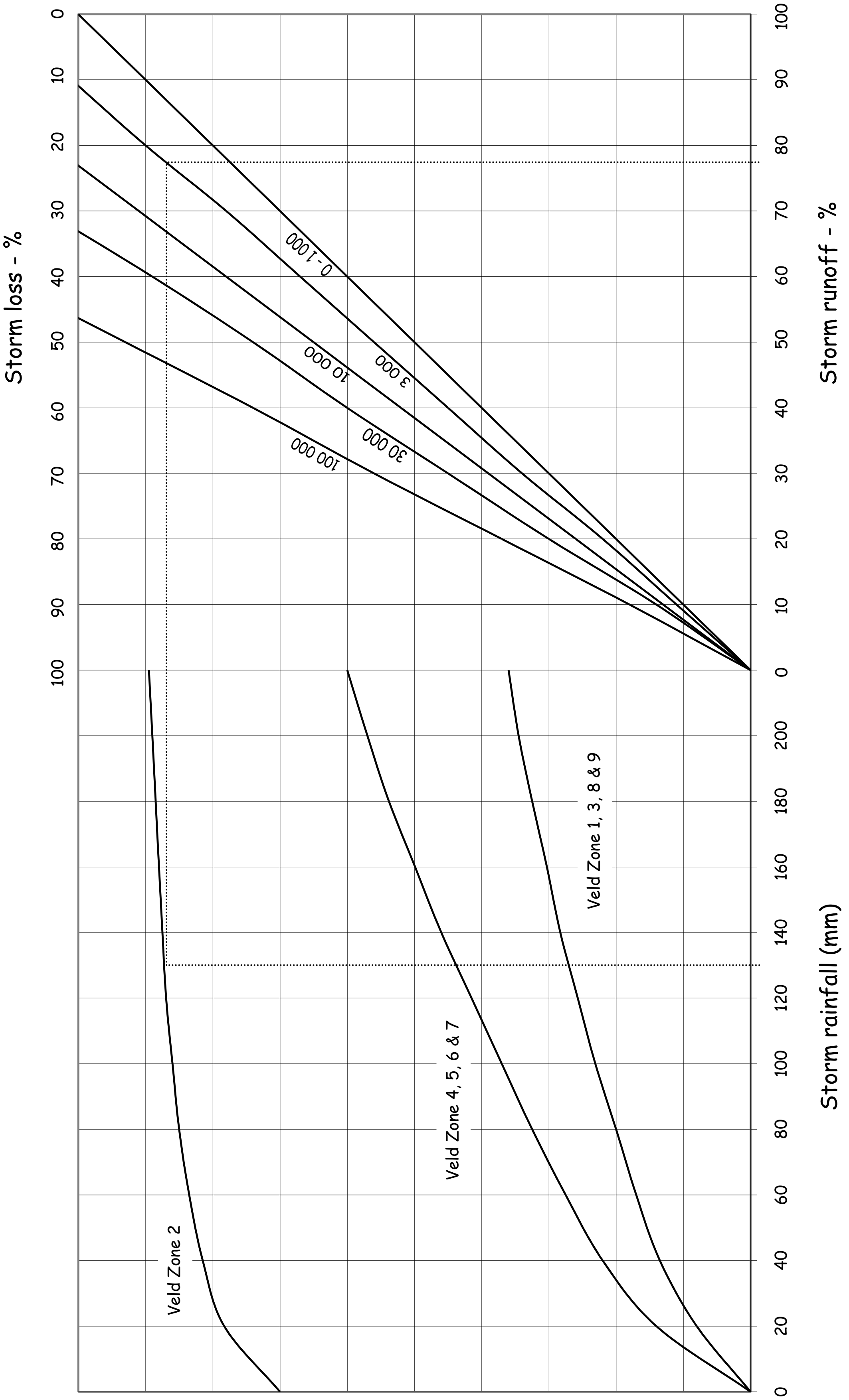


Figure A-3: Storm runoff factor, k

Table A-1: Regionally generalized dimensionless 1-hour unit hydrographs

Ratio	Discharge expressed as ratio Q/Q_P for veld zones									
T/T_L	1	2	3	4	5	6	7	8	9	5A
0.00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.035	0.012	0.010	0.011	0.018	0.024	0.006	0.006	0.011	0.004
0.10	0.070	0.024	0.023	0.024	0.038	0.052	0.014	0.014	0.027	0.011
0.15	0.112	0.036	0.039	0.038	0.063	0.087	0.024	0.025	0.043	0.019
0.20	0.163	0.052	0.057	0.051	0.095	0.140	0.032	0.035	0.065	0.027
0.25	0.228	0.072	0.074	0.070	0.142	0.260	0.044	0.050	0.093	0.037
0.30	0.306	0.091	0.106	0.089	0.220	0.700	0.058	0.069	0.142	0.050
0.35	0.414	0.121	0.139	0.111	0.315	0.983	0.074	0.100	0.225	0.064
0.40	0.524	0.152	0.184	0.138	0.500	1.000	0.095	0.150	0.350	0.083
0.45	0.709	0.198	0.261	0.175	0.685	0.970	0.121	0.245	0.570	0.107
0.50	0.921	0.258	0.376	0.260	0.810	0.915	0.160	0.655	0.772	0.140
0.55	0.983	0.342	0.518	0.350	0.936	0.848	0.275	0.905	0.930	0.210
0.60	0.996	0.472	0.670	0.700	0.985	0.795	0.480	0.980	0.982	0.425
0.65	0.998	0.676	0.809	0.980	1.000	0.754	0.700	0.994	1.000	0.885
0.70	0.964	0.940	0.970	1.000	0.960	0.714	0.950	0.991	0.985	0.958
0.75	0.893	0.991	1.000	0.987	0.800	0.678	0.975	0.966	0.945	0.993
0.80	0.826	0.995	0.990	0.885	0.675	0.641	0.993	0.860	0.900	0.991
0.85	0.758	0.973	0.935	0.760	0.588	0.605	1.000	0.755	0.814	0.955
0.90	0.700	0.888	0.840	0.670	0.524	0.572	0.995	0.655	0.750	0.740
0.95	0.652	0.807	0.755	0.580	0.473	0.540	0.980	0.565	0.670	0.535
1.00	0.605	0.741	0.675	0.530	0.432	0.514	0.900	0.500	0.600	0.440
1.05	0.563	0.678	0.612	0.470	0.397	0.488	0.805	0.440	0.530	0.385
1.10	0.525	0.622	0.546	0.430	0.365	0.465	0.730	0.392	0.472	0.340
1.15	0.491	0.567	0.500	0.393	0.340	0.443	0.655	0.355	0.413	0.300
1.20	0.463	0.513	0.460	0.364	0.315	0.422	0.590	0.322	0.364	0.265
1.25	0.437	0.467	0.424	0.336	0.295	0.402	0.530	0.294	0.316	0.235
1.30	0.411	0.425	0.395	0.310	0.276	0.382	0.477	0.270	0.280	0.209
1.35	0.387	0.394	0.368	0.288	0.260	0.365	0.432	0.250	0.260	0.187
1.40	0.362	0.364	0.347	0.271	0.242	0.347	0.388	0.231	0.241	0.169
1.45	0.341	0.338	0.325	0.252	0.228	0.330	0.350	0.215	0.225	0.152
1.50	0.321	0.313	0.305	0.235	0.214	0.315	0.308	0.200	0.210	0.140
1.55	0.302	0.291	0.290	0.218	0.200	0.300	0.280	0.186	0.198	0.128
1.60	0.283	0.272	0.276	0.201	0.187	0.287	0.255	0.174	0.188	0.116
1.65	0.265	0.253	0.264	0.187	0.174	0.274	0.232	0.164	0.176	0.105
1.70	0.252	0.236	0.252	0.172	0.163	0.260	0.211	0.155	0.168	0.097
1.75	0.238	0.220	0.238	0.159	0.152	0.249	0.194	0.146	0.158	0.088
1.80	0.226	0.206	0.228	0.147	0.143	0.237	0.177	0.137	0.151	0.081
1.85	0.215	0.192	0.216	0.136	0.134	0.225	0.164	0.130	0.144	0.074
1.90	0.204	0.181	0.208	0.125	0.126	0.214	0.152	0.122	0.137	0.067
1.95	0.194	0.171	0.200	0.115	0.120	0.203	0.140	0.115	0.131	0.061
2.00	0.183	0.160	0.194	0.108	0.112	0.193	0.130	0.110	0.124	0.055
2.05	0.174	0.152	0.186	0.098	0.106	0.183	0.120	0.103	0.119	0.050
2.10	0.165	0.143	0.178	0.089	0.101	0.173	0.111	0.098	0.113	0.046
2.15	0.157	0.136	0.171	0.081	0.094	0.164	0.102	0.091	0.108	0.041
2.20	0.149	0.130	0.165	0.074	0.088	0.155	0.094	0.086	0.103	0.038
2.25	0.142	0.123	0.158	0.068	0.084	0.147	0.087	0.081	0.097	0.034
2.30	0.135	0.118	0.152	0.062	0.079	0.138	0.081	0.075	0.093	0.031
2.35	0.128	0.114	0.147	0.056	0.074	0.130	0.075	0.070	0.087	0.028
2.40	0.121	0.108	0.142	0.052	0.070	0.122	0.069	0.066	0.085	0.025
2.45	0.116	0.104	0.139	0.047	0.066	0.115	0.063	0.062	0.079	0.023

Ratio T/T_L	Discharge expressed as ratio Q/Q_P for veld zones									
	1	2	3	4	5	6	7	8	9	5A
2.50	0.110	0.100	0.132	0.043	0.062	0.109	0.058	0.058	0.075	0.021
2.55	0.105	0.096	0.128	0.039	0.058	0.102	0.053	0.054	0.071	0.019
2.60	0.100	0.093	0.124	0.035	0.055	0.097	0.049	0.050	0.070	0.017
2.65	0.096	0.089	0.120	0.032	0.051	0.090	0.045	0.047	0.063	0.015
2.70	0.091	0.085	0.114	0.029	0.048	0.085	0.041	0.044	0.061	0.013
2.75	0.087	0.081	0.111	0.026	0.045	0.080	0.039	0.041	0.055	0.012
2.80	0.082	0.078	0.107	0.023	0.042	0.075	0.036	0.038	0.053	0.011
2.85	0.078	0.074	0.103	0.021	0.039	0.069	0.033	0.035	0.049	0.010
2.90	0.074	0.070	0.099	0.019	0.036	0.064	0.030	0.032	0.045	0.009
2.95	0.070	0.066	0.095	0.017	0.033	0.059	0.029	0.029	0.041	0.008
3.00	0.066	0.063	0.091	0.016	0.030	0.054	0.026	0.026	0.038	0.006
3.05	0.062	0.060	0.087	0.012	0.027	0.049	0.023	0.024	0.035	0.004
3.10	0.057	0.056	0.084	0.011	0.025	0.044	0.021	0.022	0.030	0.003
3.15	0.054	0.053	0.081	0.009	0.022	0.040	0.019	0.020	0.027	0.002
3.20	0.050	0.050	0.078	0.008	0.020	0.036	0.017	0.019	0.022	0.001
3.25	0.047	0.047	0.075	0.006	0.018	0.031	0.015	0.017	0.018	0.000
3.30	0.043	0.044	0.071	0.004	0.016	0.027	0.013	0.015	0.014	0.000
3.35	0.039	0.040	0.068	0.003	0.013	0.022	0.011	0.013	0.010	0.000
3.40	0.036	0.037	0.064	0.002	0.011	0.018	0.010	0.011	0.007	0.000
3.45	0.032	0.034	0.062	0.001	0.010	0.013	0.008	0.009	0.004	0.000
3.50	0.029	0.031	0.059	0.000	0.008	0.010	0.006	0.007	0.002	0.000
3.55	0.025	0.027	0.056	0.000	0.006	0.005	0.005	0.005	0.000	0.000
3.60	0.022	0.024	0.051	0.000	0.004	0.000	0.004	0.004	0.000	0.000
3.65	0.019	0.021	0.048	0.000	0.002	0.000	0.002	0.002	0.000	0.000
3.70	0.016	0.018	0.046	0.000	0.001	0.000	0.001	0.001	0.000	0.000
3.75	0.012	0.015	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.80	0.009	0.011	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.85	0.005	0.008	0.037	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.90	0.003	0.005	0.035	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.95	0.000	0.002	0.032	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.00	0.000	0.000	0.029	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.05	0.000	0.000	0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.10	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.15	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.20	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.25	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.30	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.35	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.40	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.45	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.50	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.55	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

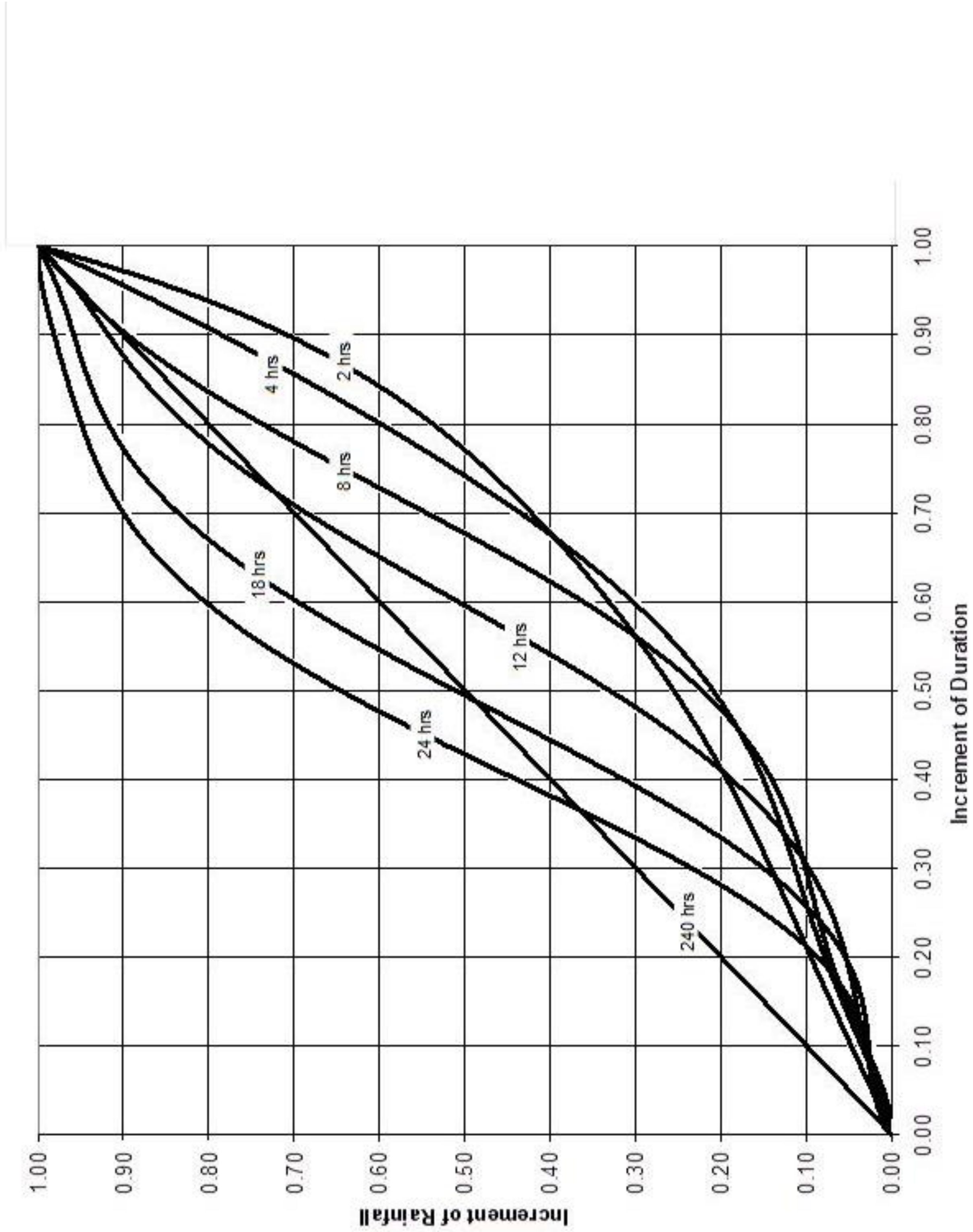


Figure A-4: Rainfall distribution with time

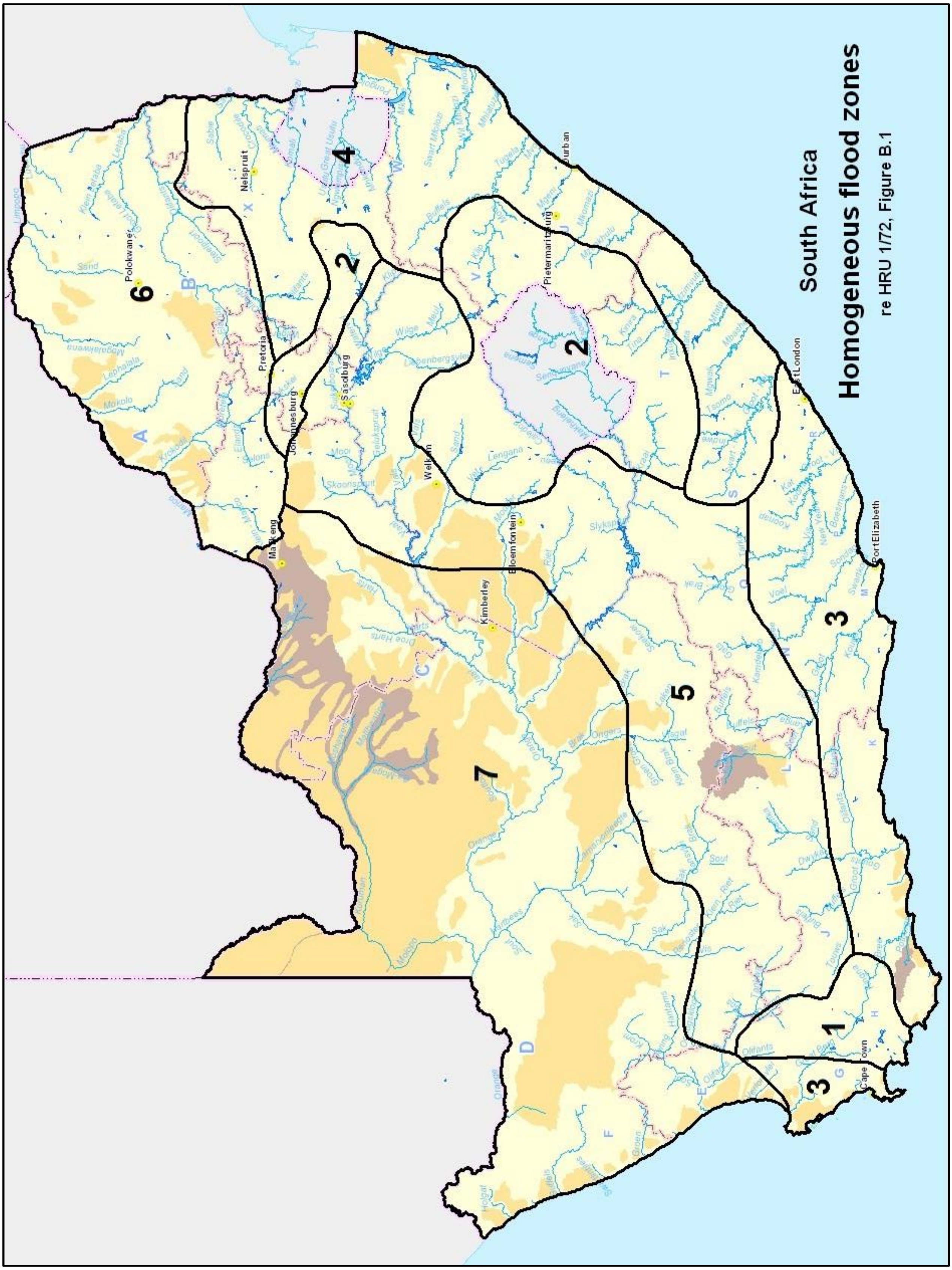


Figure A-5: MIPI flood regions

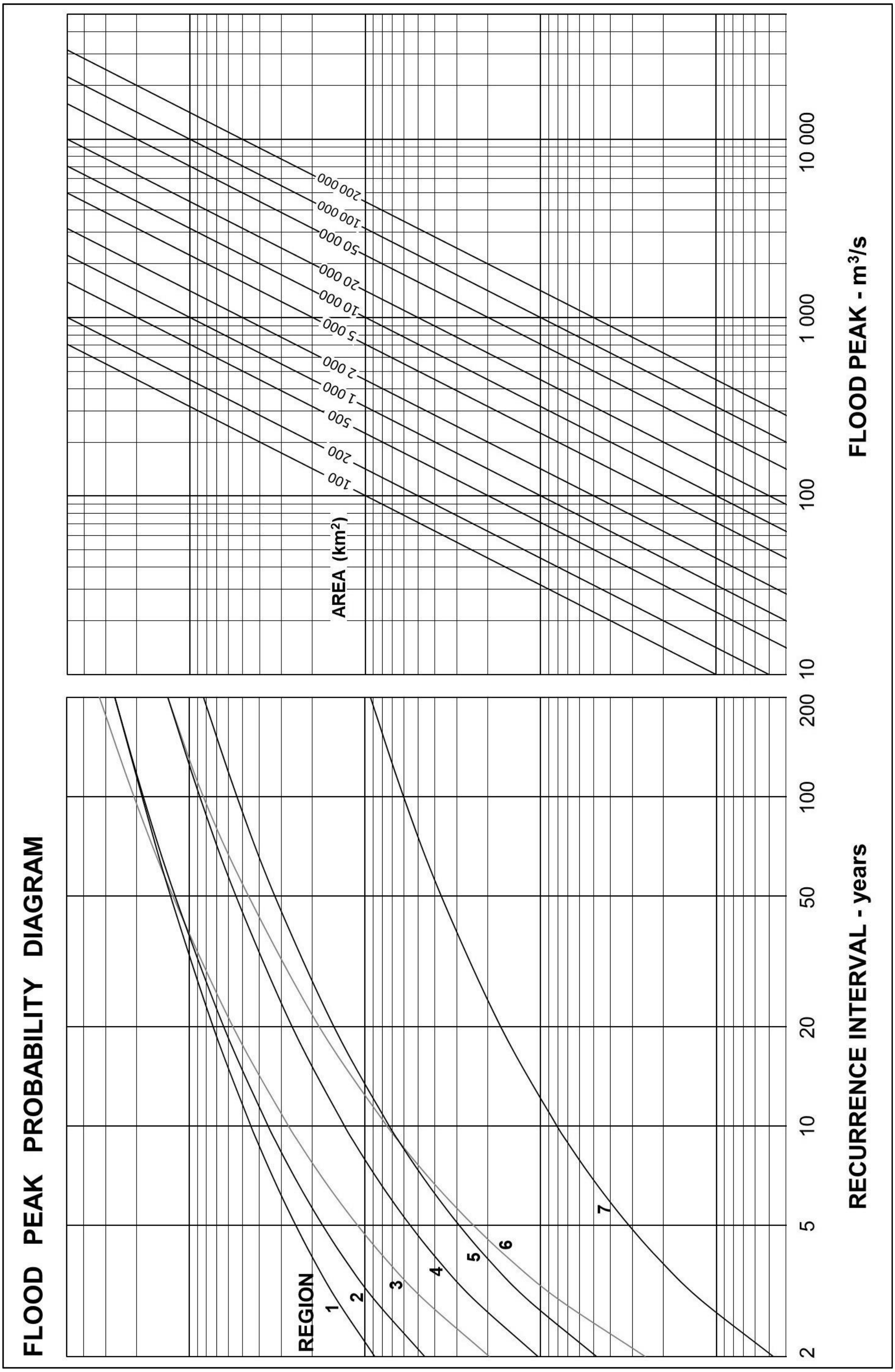


Figure A-6: MIPI-method

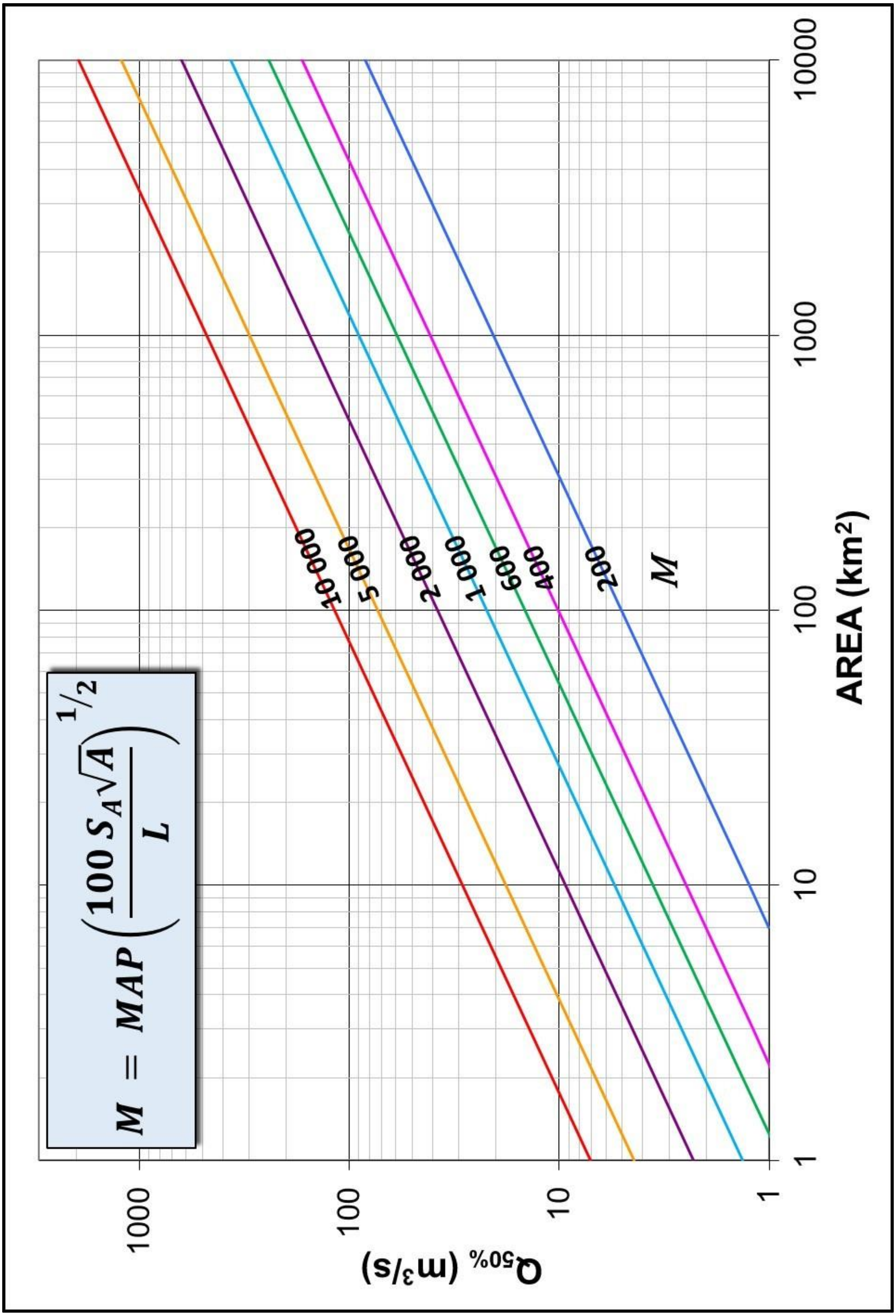


Figure A-7: CAPA-method

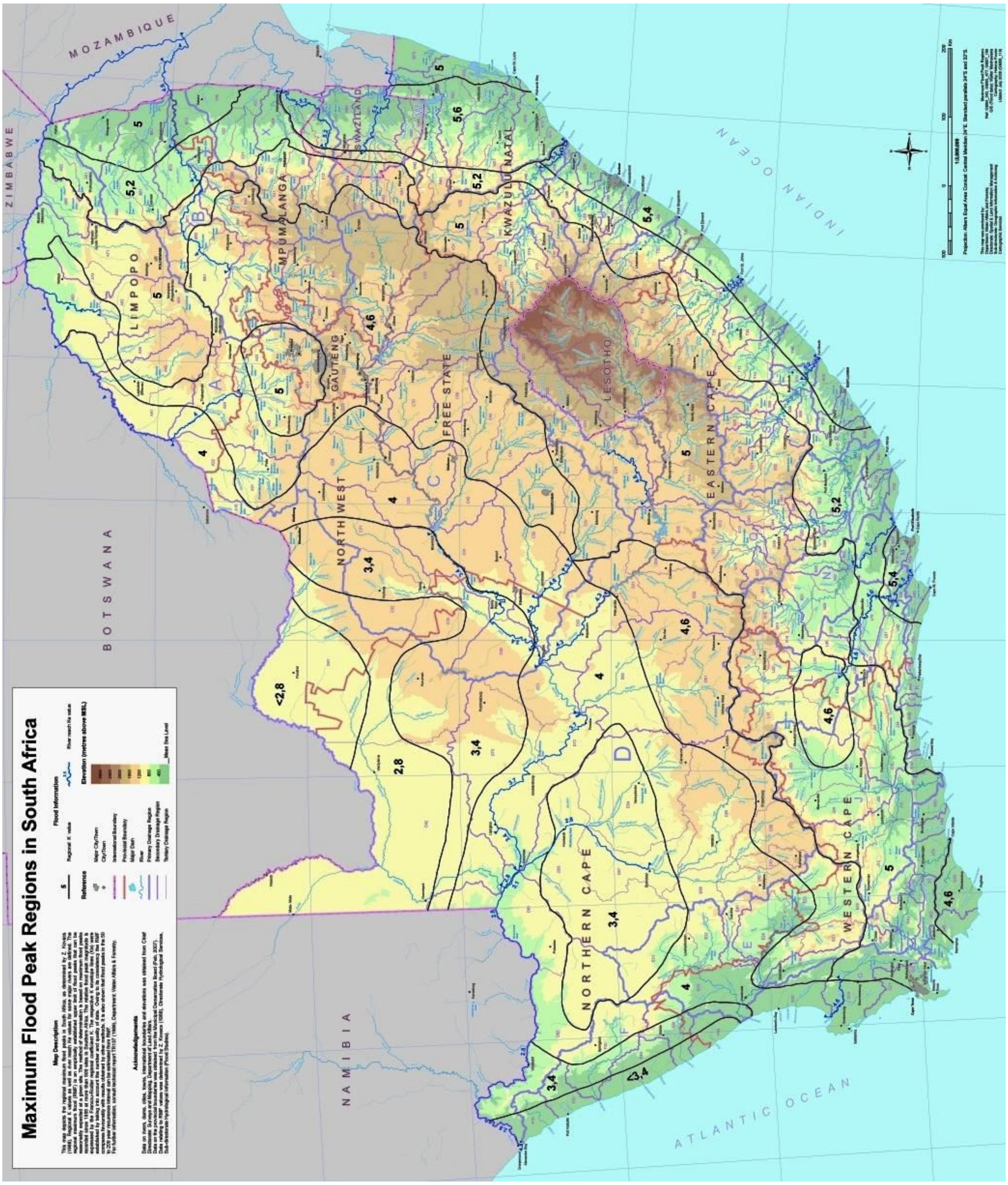


Figure A-8: RMF regions for South Africa