

# Estimation of Flow Characteristics of Ungauged Catchments

Case Study in Zimbabwe

Dominic MAZVIMAVI

Promotor: Prof. dr. Alfred Stein  
Professor of Spatial Statistics  
Wageningen University, Wageningen, and  
International Institute for Geo-information Science  
and Earth Observation, Enschede.

Co-Promotors: Prof. dr. A.M.J. Meijerink  
Professor of Water Resources Surveys and  
Watershed Management  
International Institute for Geo-information Science  
and Earth Observation, Enschede.

Prof. dr. ir. H.H.G. Savenije  
Professor of Water Resources Management  
UNESCO-IHE, Delft.

Members of the Examination Committee

Prof. dr. C. C. Mutambirwa, University of  
Zimbabwe  
Prof.dr. P. Troch, Wageningen University  
Dr. C.J.F. Ter Braak, Biometris  
Dr. E. Seyhan, Free University Amsterdam

# Estimation of Flow Characteristics of Ungauged Catchments

Case Study in Zimbabwe

Dominic MAZVIMAVI

Thesis  
to fulfil the requirements for the degree of doctor  
on the authority of the rector magnificus  
of Wageningen University  
Prof.dr.ir. L. Speelman  
to be publicly defended on 18 December 2003  
at 11.00 hours in the auditorium of ITC, Enschede

ISBN 90-5808-950-9  
ITC Dissertation number 107

Dominic Mazvimavi, 2003  
Estimation of Flow Characteristics of Ungauged Catchments: Case Study  
in Zimbabwe/ Mazvimavi, D.  
Ph.D. Thesis, Wageningen University – with references and summaries  
in English and Dutch.

## ACKNOWLEDGEMENTS

I have received tremendous assistance and encouragement from several people during my PhD studies in both the Netherlands and Zimbabwe. Firstly, I would like to express my gratitude to Prof. dr. A.M.J. Meijerink who accepted the challenge of supervising a student who was going to spend most of the time away from ITC, knowing fully well problems associated with such an arrangement. Several discussions we had with Prof Meijerink and his comments on several drafts of this thesis have been very illuminating. At all stages of preparing this thesis Prof Meijerink went out of his way to provide me with the necessary support.

I am most grateful to Prof. dr. A. Stein for accepting to be my promotor. Prof Stein meticulously went over several drafts of this thesis, and each time I handed some material for his consideration, I always looked forward to receive his comments. Prof H.H.G. Savenije, my co-promotor, encouraged me to embark on a PhD on several occasions we met in Africa, and I am most grateful for the assistance he has provided. He always created time for discussions during his short but busy visits to the University of Zimbabwe.

I am thankful to Prof D. Hughes of the Institute of Water Research, Rhodes University, South Africa, who kindly assisted me on rainfall-runoff modelling. The Institute of Water Research provided me with an office, and this is sincerely acknowledged.

My PhD studies would not have been possible without the financial support provided by the Netherlands Government funded ITC/Dept of Geography & Environmental Science (University of Zimbabwe) Project on Strengthening the Masters in Environmental Policy (MEPP). Dr Herman Huizing (ITC MEPP Project Director) and Dr Charles Kunaka (Dept of Geography, University of Zimbabwe, MEPP Project Director) were instrumental in encouraging me to start on a PhD. Dr Huizing went out of his way organizing my several visits to ITC. His comments which he termed “layman comments” on my thesis were very encouraging. To both Herman and Charles, thank you.

The support provided by my colleagues in the Department of Geography and Environmental Science, University of Zimbabwe, is greatly appreciated. They took over my teaching obligations despite the department not having a full staff complement. Prof C.C. Mutambirwa’s encouragement, and comments are greatly appreciated. Dr Pieter van der Zaag at the University of Zimbabwe has been very supportive over the years. He translated the English Summary of this thesis into Dutch, and I am very grateful for all the assistance that Pieter has provided. Sam Kusangaya and Tinos Madondo worked tirelessly even during

off-working hours capturing data, and I am grateful for their assistance. The University of Zimbabwe is thanked for granting me leave to undertake my PhD studies.

During periods when I was at ITC, Zimbabwean students; Onisimo Mutanga, Amon Murwira, Liliosa Magombedza, Angela, Robert Maruziva, Gilbert Mhlanga and Itai made me feel at home. Onisimo, Amon and Karin, fellow PhD students have been such wonderful colleagues and I thank them. Other Zimbabweans in the Netherlands, Sam and Patience Matsangaise, Joyce Nyanyira welcomed me in their homes whenever I needed to take a break and be out of a student environment. Their support is gratefully acknowledged.

Staff in the Department of Water Resources, and the Cluster Manager at ITC were helpful during my stay at ITC. I thank them all. The support provided by fellow PhD students; Mobin-ud-Din Ahmad, Masoud Kheirkhah, and Obokile Obakeng whom I shared an office with is greatly appreciated.

Data used in this study were provided by the Zimbabwe National Water Authority, and the Department of Meteorological Services. The assistance provided by these organizations is gratefully acknowledged.

The encouragement from several friends in Zimbabwe in particular Stan and Grace Mangoma, Charles Kunaka, Lazarus Zanamwe, Davison Gumbo, Chris and Esther Mandizvidza, Felicitus, Netsai, Gladys, Tendai and Bathsheba Biti is sincerely appreciated.

I am most grateful for the support and encouragement throughout my education from my father and late mother. I am thankful for the continuous encouragement from my brothers and sisters. My PhD studies have required that every year I was in the Netherlands for about three months, during which my family had to stand on their own. I wish to express my gratitude for their sacrifice, support and encouragement. To my wife Gertrude, and children Tafadzwa, Vimbiso and Tapiwa, thank you.

## TABLE OF CONTENTS

<i>ACKNOWLEDGEMENTS</i> .....	<i>i</i>
<i>TABLE OF CONTENTS</i> .....	<i>iii</i>
<i>LIST OF FREQUENTLY USED SYMBOLS</i> .....	<i>vi</i>
<b>1 INTRODUCTION</b> .....	<b>1</b>
1.1 Background.....	1
1.2 Research objectives.....	3
1.3 Outline of the thesis.....	4
<b>2 THE STUDY AREA</b> .....	<b>5</b>
2.1 Introduction.....	5
2.2 Selection of the catchments.....	8
2.3 Derivation of flow characteristics.....	11
2.3.1 Base flow and recession constant.....	12
2.4 Selection and derivation of catchment characteristics.....	14
2.4.1 Mean annual precipitation and number of rainy days.....	15
2.4.2 Maximum, average and minimum elevations, and relief.....	17
2.4.3 Slope.....	18
2.4.4 Drainage density.....	20
2.4.5 Geology.....	22
2.4.6 Land cover.....	26
2.4.7 Normalized Difference Vegetation Index.....	28
2.4.8 Evaporation.....	30
2.5 Flow characteristics.....	33
<b>3 PREDICTION OF FLOW CHARACTERISTICS</b> .....	<b>37</b>
3.1 Introduction.....	37
3.2 Neural networks.....	37
3.3 Results.....	40
3.3.1 Correlation between flow and catchment characteristics.....	40
3.3.2 Mean annual runoff.....	42
3.3.3 Base flow index.....	46
3.3.4 Flow duration curves.....	50
3.3.5 Average number of days with zero flows ( $\bar{N}_{DZ}$ ).....	53
3.3.6 Mean monthly runoff distribution.....	53
3.4 Summary.....	56
<b>4 ORDINATION</b> .....	<b>59</b>
4.1 Introduction.....	59
4.2 Methodology.....	59
4.3 Results.....	62
4.3.1 Relationships between ordination axes.....	62
4.3.2 Relationships between catchment characteristics and their ordination axis.....	64
4.3.3 Relationship between flow characteristics and their ordination axes.....	65

4.3.4	Relationship between flow and catchment characteristics .....	66
4.4	Discussion and Conclusion .....	69
5	<i>IDENTIFICATION OF CATCHMENTS WITH SIMILAR HYDROLOGICAL RESPONSES</i> .....	71
5.1	Introduction .....	71
5.2	Methodology .....	71
5.2.1	Selection of catchment descriptors.....	71
5.2.2	Determination of the number of clusters and validation .....	73
5.3	Results and Discussion.....	75
5.3.1	Classification using catchment characteristics .....	75
5.3.2	Classification using flow characteristics .....	75
5.3.3	Number of clusters .....	78
5.3.4	Catchment characteristics of clusters .....	81
5.3.5	Flow characteristics.....	82
5.3.6	Comparison of weighted and unweighted clustering .....	85
5.3.7	Prediction of flow characteristics of clusters .....	87
5.4	Conclusion .....	90
6	<i>REGIONALISATION OF SELECTED RAINFALL-RUNOFF MODELS</i> .....	93
6.1	Introduction.....	93
6.2	Methodology .....	94
6.2.1	Selection of rainfall-runoff models .....	95
6.2.2	Model calibration .....	95
6.2.3	Model validation .....	97
6.2.4	Structure of models selected .....	97
6.3	Results and discussion .....	107
6.3.1	Comparison of simulated and observed monthly flows .....	108
6.3.2	Prediction of model parameters .....	117
6.4	Conclusion .....	135
7	<i>CONCLUSIONS AND RECOMMENDATIONS</i> .....	137
7.1	Introduction.....	137
7.2	Prediction of flow characteristics using univariate methods.....	137
7.3	Identification of clusters of catchments with similar hydrological responses .....	138
7.4	Prediction of flow characteristics based on hydrological homogenous regions.....	139
7.5	Prediction of model parameters of conceptual models .....	140
7.6	Comparison of performances of neural networks and multiple regression.....	141
7.7	Recommendations .....	141
	<i>REFERENCES</i> .....	143
	<i>ENGLISH SUMMARY</i> .....	157
	<i>SAMENVATTING</i> .....	161
	<i>CURRICULUM VITAE</i> .....	165

*APPENDIX 1: LIST OF SELECTED CATCHMENTS..... 167*  
*APPENDIX 2: CODES USED TO REFER TO SELECTED CATCHMENTS 169*  
*APPENDIX 3: ITC DISSERTATION LIST..... 171*

## LIST OF FREQUENTLY USED SYMBOLS

Symbol	Representation	Units
$A$	Catchment area	$\text{km}^2$
$AI$	Proportion of the catchment which is impervious	
$\overline{N}_{DZ}$	Annual average number of days with no flow	days year <sup>-1</sup>
$BFI$	Base flow index	
$CV$	Coefficient of variation	%
$Dd$	Drainage density	$\text{km km}^{-2}$
$exp$	Exponential function	
$E_{pot,t}$	Monthly potential evaporation	mm month <sup>-1</sup>
$E_t$	Monthly actual evaporation	mm month <sup>-1</sup>
$\overline{E}_{pot,yr}$	Annual average potential evaporation	mm year <sup>-1</sup>
$FT$	Maximum rate of drainage when soil is saturated	mm month <sup>-1</sup>
$GL_{GG}$	Proportion of a catchment with gneiss and granite	%
$GL_{DO}$	Proportion of a catchment with dolerite dykes and sills	%
$GL_{AL}$	Proportion of a catchment with alluvial deposits	%
$GL_{BA}$	Proportion of a catchment with upper Karoo basalt	%
$GL_{GD}$	Proportion of a catchment on which the Great Dyke occurs	%
$GL_{GR}$	Proportion of a catchment with greenstones	%
$GL_{KL}$	Proportion of a catchment with Kalahari sands	%
$GL_{LM}$	Proportion of a catchment with the Umkondo assemblage	%
$GL_{SA}$	Proportion of a catchment with upper Karoo sandstones	%
$GL$	Time of lag of subsurface flow from lower zone in the Pitman model	month
$G_w$	Maximum monthly groundwater flow rate in the Pitman model	mm month <sup>-1</sup>
$I_{cap}$	Interception capacity	mm
$I_{m,t}$	Monthly interception rate	mm month <sup>-1</sup>
$K_b$	Monthly recession constant	
$LC_{PL}$	Proportion of a catchment under forest plantation	%
$LC_{FO}$	Proportion of a catchment under moist evergreen and deciduous species	%
$LC_{WD}$	Proportion of a catchment under woodlands	%
$LC_{WG}$	Proportion of a catchment under wooded grasslands	%
$LC_{GR}$	Proportion of a catchment under grasslands	%

$LC_{BU}$	Proportion of a catchment under bushland	%
$LC_{CU}$	Proportion of a catchment which is cultivated	%
$NDVI$	Normalized Difference Vegetation Index	
$n_r$	Number of rainy days in a month	days
$n_m$	Number of days in a month	days
$\overline{N}_{yr}$	Average number of rainy days per year	days year <sup>-1</sup>
$P_t$	Monthly precipitation	mm month <sup>-1</sup>
$\overline{P}_{yr}$	Annual average precipitation	mm year <sup>-1</sup>
$POW$	Power of the curve relating rate of soil moisture drainage to stream from the soil moisture storage	
$q_t$	Average daily flow	m <sup>3</sup> s <sup>-1</sup>
$q_{g,t}$	Base flow at a daily interval	m <sup>3</sup> s <sup>-1</sup>
$q_{s,t}$	Surface (direct) runoff at a daily interval	m <sup>3</sup> s <sup>-1</sup>
$q_{5...90}$	Dimensionless daily flow with a 5...90% exceedance probability, and derived as a proportion of the average daily flow	
$Q_t$	Monthly runoff	mm month <sup>-1</sup>
$Q_{g,t}$	Base flow rate at monthly interval	mm month <sup>-1</sup>
$Q_{s,t}$	Surface (direct) runoff at monthly interval	mm month <sup>-1</sup>
$\overline{Q}_{yr}$	Annual average runoff	mm year <sup>-1</sup>
$r$	Correlation coefficient	
$r^2$	Coefficient of determination	
$R$	Parameter relating rate of decrease in evaporation rate to the decrease in soil moisture storage	
$R_g$	Index measuring level of agreement between results of 2 different clustering procedures	
$R_{chg,t}$	Rate of groundwater recharge per month	mm month <sup>-1</sup>
$S_t$	Soil moisture storage at the end of month	mm
$S_{cap}$	Maximum soil moisture storage capacity	mm
$S_{g,t}$	Groundwater storage at the end of month	mm
$TL$	Time lag of subsurface runoff from the upper zone	month
$yr$	year	
$Z_{min}$	Minimum absorption rate of rainfall by a catchment	mm
$Z_{max}$	Maximum absorption rate of rainfall by a catchment	mm



# 1 INTRODUCTION

## 1.1 Background

Sustainable water resources planning and management requires data to enable quantification of water quality and quantity (Oyebande, 2001). Information is required on the rates of transfers and storage of water within a catchment. Lack of adequate hydrological data introduces uncertainty in both the design and management of water resources systems. Much of sub-Saharan Africa has as a result of low conversion of rainfall to runoff a precarious balance between available water resources and water demand. The rapid population growth characteristic of this region, which is increasing water demand for domestic, agricultural, and industrial purposes, is causing water scarcity (Oyebande, 2001). The magnitude of this scarcity and its variation in both space and time are largely unknown because of lack of hydrological data. Catchment degradation in its various forms continues without effective control measures due partly to uncertainty regarding the adverse effects on water resources. This uncertainty again arises from lack of adequate hydrological data that should enable quantification of effects of specific land use practices on quality and quantity of water resources. In addition, floods and droughts occur with frequencies and magnitudes that are poorly defined in sub-Saharan Africa because of lack of relevant hydrological data. These cause annually major social, economic and environmental tragedies.

Lack of information about the quantity and quality of water resources arises from poorly developed hydrological networks. Most sub-Saharan countries lack financial, human and technical resources for developing and maintaining networks that can provide data for sustainable water resources planning, design, and management. While the needs for hydrological information for these countries are increasing, their technical and human capacities are declining as noted by the reduction in the number of meteorological stations in Africa during the last 30 years (Bonifacio and Grimes, 1998; Oyebande, 2001). If resources were to be made available for the extension of hydrometric networks, it will take 10 to 30 years before adequate data are collected. Therefore, problems arising from lack of data will persist in the foreseeable future. It is also not possible to set up an ideal network as some sites are inaccessible. Furthermore, some existing monitoring sites have already been affected by anthropogenic influences such as upstream abstractions and impoundments on rivers that render the data collected unsuitable for long-term planning. Consequently, there is a need to develop methods for predicting flow characteristics at ungauged sites. The International Association of Hydrological Sciences (IAHS) recognized this need in 2002, and adopted the Prediction of Ungauged Basins

(PUBS) as a research agenda for the coming decade <[www.cig.enscm.fr/~iahs/PUBs/PUB-proposal](http://www.cig.enscm.fr/~iahs/PUBs/PUB-proposal)>.

Estimation of flow characteristics of ungauged catchments is usually based on transferring or extrapolating information from gauged to ungauged sites, a process called regionalisation (Nathan and McMahon, 1990a; Bullock and Andrews, 1997; Hall and Minns, 1999). Several regionalisation approaches have been used, and the most common method involves derivation of empirical relationships between flow and catchment characteristics (Gan *et al.*, 1990; Riggs, 1990). These relationships are in most cases region specific. Therefore, regions within which they are applicable have to be delimited, for example using hydrometric zones (Mimikou, 1984). Flow characteristics at ungauged sites are estimated by applying a predictive equation developed for a particular hydrometric zone (NERC, 1975; IH, 1980). Catchments that belong to the same hydrometric zone, however, do not necessarily have similar hydrological responses since geographical proximity is not a sufficient condition for hydrological homogeneity (Acreman and Sinclair, 1986). Meijerink (1985) found that morpho-lithological characteristics could be used to identify catchment groups with similar hydrological responses. The delimitation of regions with similar hydrological responses or hydrologically homogenous regions has alternatively been done using multivariate techniques such as multiple regression, cluster and discriminant analysis (Tasker, 1982; Nathan and McMahon, 1990a; Burn and Boorman, 1993; Zrinji and Burn, 1994). Catchment characteristics that influence flow characteristics should ideally be used for cluster analysis. This enables determination of membership of an ungauged catchment on the basis of its catchment characteristics, to a region with a known relationship between flow and catchment characteristics. The selection of these catchment characteristics is problematic since different sets of predictive variables will identify different clusters. Nathan and McMahon (1990a) demonstrated that a combination of multiple regression, cluster analysis and multi-dimensional plotting improved the delimitation of these hydrologically homogenous regions within which predictive equations for flow characteristics can be developed.

The use of ordination techniques like redundancy analysis to select catchment characteristics for classifying catchments into homogenous regions has not been explored. Ordination techniques have been found to be suitable for identifying explanatory variables for multidimensional responses (Kent and Coker, 1992; Ter Braak and Smilauer, 1998; McGarrial *et al.*, 2000). Neural networks also offer an alternative approach for classifying catchments using catchment characteristics (Hall and Minns, 1999).

Another approach that has been used for estimating flow characteristics of ungauged catchments is the use of rainfall-runoff models whose parameters have been regionalised. Previous studies such as in Ivory Coast (Servat and Dezetter,

1993), Australia (Post and Jakeman, 1996, 1999; Post *et al.*, 1998), and the United Kingdom (Manley, 1978) have demonstrated the feasibility of regionalising lumped conceptual models. Regionalisation of the Pitman rainfall-runoff model in South Africa has been the basis of assessing water resources for all drainage basins in that country. Hughes (1997) concluded that there appears to be a potential for regionalising the Pitman model in Zimbabwe. There have been few studies in southern Africa that have investigated the feasibility of deriving relationships between parameters of lumped models and catchment characteristics. If possible, this will offer an opportunity for estimating time series of flows at ungauged sites.

A case study approach is used in this study to investigate the problem of estimating flow characteristics of ungauged catchments. The case study is based on selected catchments in Zimbabwe, a country with a tropical climate and rainfall occurring in one distinct season, mid-November to March. The spatial variation of rainfall is greatly dependent on altitude with lowlying areas receiving  $350 \text{ mm yr}^{-1}$ , and  $2000 \text{ mm yr}^{-1}$  for highland regions. Most rivers dry up during the dry season, resulting in water scarcity. Water scarcity also results from high inter-annual variability of rainfall. Dam construction to create over-the-year storage is necessary to alleviate water shortages. Most dams are sited on ungauged catchments, and yet their design and subsequent management of reservoirs require hydrological data.

## **1.2 Research objectives**

The main objective is to identify and assess the suitability of statistical methods and conceptual rainfall-runoff models to estimate flow characteristics of ungauged catchments in Zimbabwe.

The specific objectives are:

1. To identify catchment characteristics that can be used for predicting flow characteristics of ungauged catchments.
2. To examine the feasibility of using catchment characteristics for identifying catchments with similar hydrological responses or delimiting hydrologically homogenous regions.
3. To assess the potential of using hydrologically homogenous regions as the basis for estimating flow characteristics of ungauged catchments.
4. To determine the possibility for regionalising parameters of selected lumped rainfall-runoff models on the basis of catchment characteristics, and using these to estimate flow characteristics of ungauged catchments.

5. To assess if neural networks have a better capability than multiple regression methods to predict flow characteristics and parameters of conceptual rainfall-runoff models from catchment characteristics.

### **1.3 Outline of the thesis**

Chapter 2 describes the study area. The types of catchment and flow characteristics selected for use in this study, and justification for this selection are presented. A description of the variation of these characteristics among the selected catchments is also given.

Chapter 3 develops relationships between catchment and flow characteristics using multiple regression, and neural networks. These relationships are used for estimating mean annual runoff, base flow index, average number of days per year with no flow, flow duration curves, and distribution of mean annual runoff into monthly flows.

In Chapter 4 a direct gradient analysis method, redundancy analysis, is used to investigate the effects of catchment characteristics on multidimensional hydrological responses. The relative importance of each of the catchment characteristics in explaining the variance of all the flow characteristics is investigated.

Chapter 5 investigates the possibility of clustering catchments using catchment characteristics into clusters with similar hydrological responses. An assessment of whether clustering of catchments improves the prediction of flow characteristics done in Chapter 3 is made.

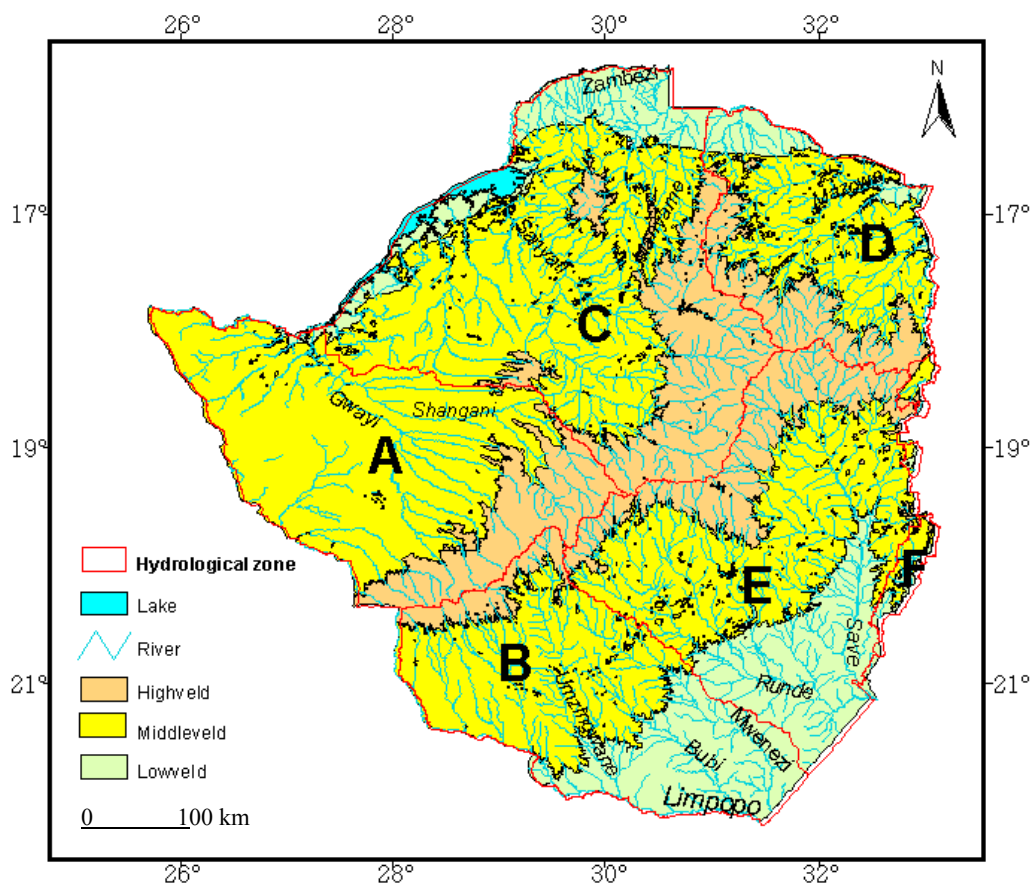
The feasibility of regionalising the *abcd* model and the Pitman model, both of which are lumped conceptual models, is investigated in Chapter 6. The possibility of predicting model parameters from catchment characteristics using multiple regression and neural networks is examined.

Chapter 7 presents the conclusions of this study.

## 2 THE STUDY AREA

### 2.1 Introduction

Zimbabwe has an area of 390 757 km<sup>2</sup> of which over 60% is underlain by Archaen granite and greenstone that have considerable influence on relief. Altitude varies from 162 to 2592 m above sea level. The country can be divided into four physiographic regions on the basis of relief (Table 2.1; Figure 2.1).



**Figure 2.1:** Relief map of Zimbabwe. A = rivers draining into the Gwayi River and then into the Zambezi River. B = rivers draining into the Limpopo River. C = catchment of Manyame and Sanyati Rivers which drain into the Zambezi River. D = basin of Mazowe River which drains into the Zambezi River. E = area drained by Save and Runde Rivers. F = rivers draining from the Eastern Highlands towards the east into Mozambique

The topography is generally flat to undulating on the lowveld, undulating and rolling on the middleveld, and the highveld comprises an undulating plateau. The Eastern Highlands region is mountainous with the highest peak on Mt Nyangani, 2592 m. There are four seasons which can be identified (Table 2.2).

**Table 2.1:** Physiographic Regions of Zimbabwe

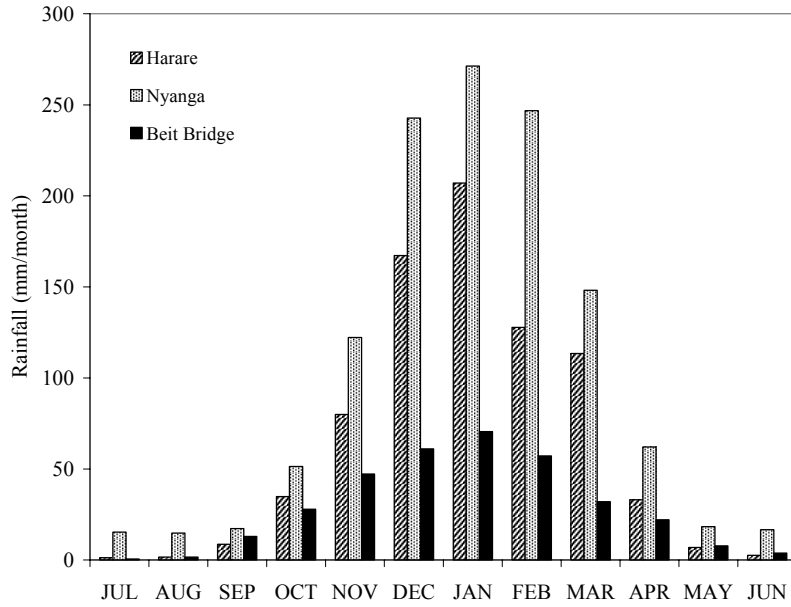
Relief Region	Altitude (m)	Min and Max Temperatures (°C)	Mean Annual Rainfall (mm yr <sup>-1</sup> )
Lowveld	162 – 600	9.4 – 33.7	344 – 600
Middleveld	600 – 1200	5.5 – 30.7	600 – 700
Highveld	1200 – 1800	5.0 – 27.5	700 – 1200
Eastern Highlands	1800 - 2592	5.0 – 22.0	1200 - 2000

**Table 2.2:** Seasons of Zimbabwe

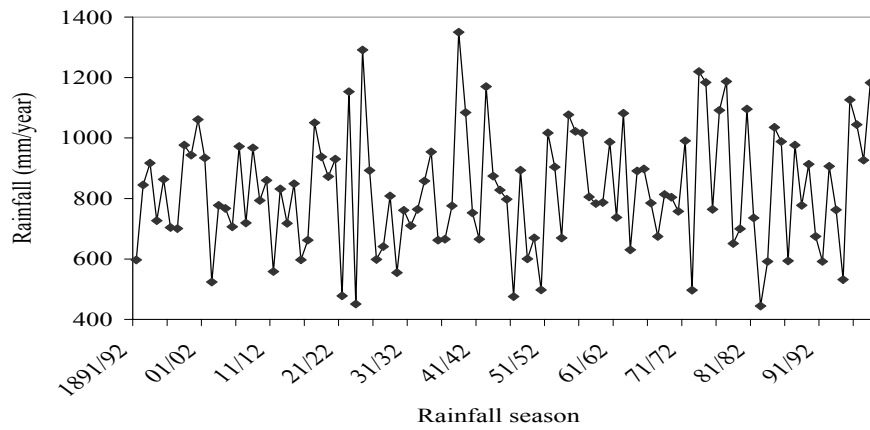
Season	Duration
Rainy season	mid-November to mid-March
Post-rainy season	mid-March to mid-May
Cool season	mid-May to August
Hot season	September to mid-November

Figure 2.2 shows the seasonal variation of rainfall at Beit Bridge in the lowveld with mean annual rainfall of 345 mm yr<sup>-1</sup>, Harare 850 mm yr<sup>-1</sup> on the highveld, and Nyanga 1227 mm yr<sup>-1</sup> on the Eastern Highlands.

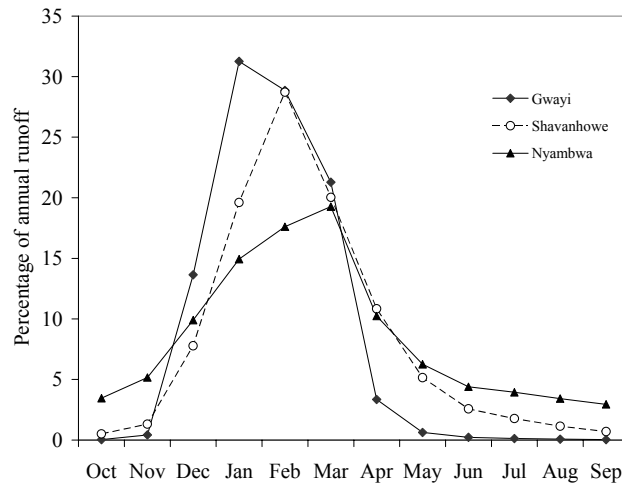
Potential evaporation rates exceed rainfall except during the rainy season, and the aridity index is 6.0 for Beit Bridge, 2.3 for Harare, and 1.1 for Nyanga in the Eastern Highlands. The aridity index is the ratio of mean annual evaporation to the mean annual precipitation (Oyebande, 2001). Rainfall has a high inter-annual variability as shown by the example of Harare in Figure 2.3. The northern part of the country is drained by the Gwayi, Sanyati, Manyame, and Mazowe Rivers which flow into the Zambezi River. The Save River drains the south-eastern part of the country, while the rest of the southern area flows into the Limpopo River (Figure 2.1). Runoff is highly variable in space and the mean annual runoff is approximately 20 mm yr<sup>-1</sup> for most catchments on the north-western, and extreme southern parts of the country. Mean annual runoff for catchments on the central part of the country ranges from 60 to 80 mm yr<sup>-1</sup>, and from 200 to 600 mm yr<sup>-1</sup> in the Eastern Highlands.



**Figure 2.2:** Mean monthly rainfall for Beit Bridge on the lowveld, Harare on the highveld, and Nyanga on the Eastern Highlands



**Figure 2.3:** Illustration of high inter-annual variability of rainfall in Zimbabwe using Harare annual rainfall



**Figure 2.4:** Seasonality of river flows. Gwayi River with a mean annual runoff of 17 mm/yr, Shavanhowe 195 mm/yr, and Nyambwa 290 mm/yr

Most rivers flow only during the rainy season except for those on the Eastern Highlands. This is illustrated in Figure 2.4 with Gwayi River representing the dry northern and southern parts; Shavanhowe River, northern part; and Nyambwa River, Eastern Highlands. Water scarcity due to both high inter-annual variability and seasonality of flows is a major problem for water resources management. Severe competition for these resources occurs between a) the commercial and peasant farming sectors, and b) urban and rural areas. Conflicts over water are numerous within the commercial farming sector. For some catchments, 50 to 90% of renewable annual water resources have already been allocated to existing water users (Mazvimavi, 1998). Water allocation to prospective new water users becomes highly problematic in the absence of adequate flow data as is common for several catchments. Most flow measuring stations are located on the developed central part of the country (Figure 2.5). The need to allocate some water for environmental purposes has been acknowledged recently, but very limited data is available for determining flow regimes to be maintained along rivers for this purpose.

## 2.2 Selection of the catchments

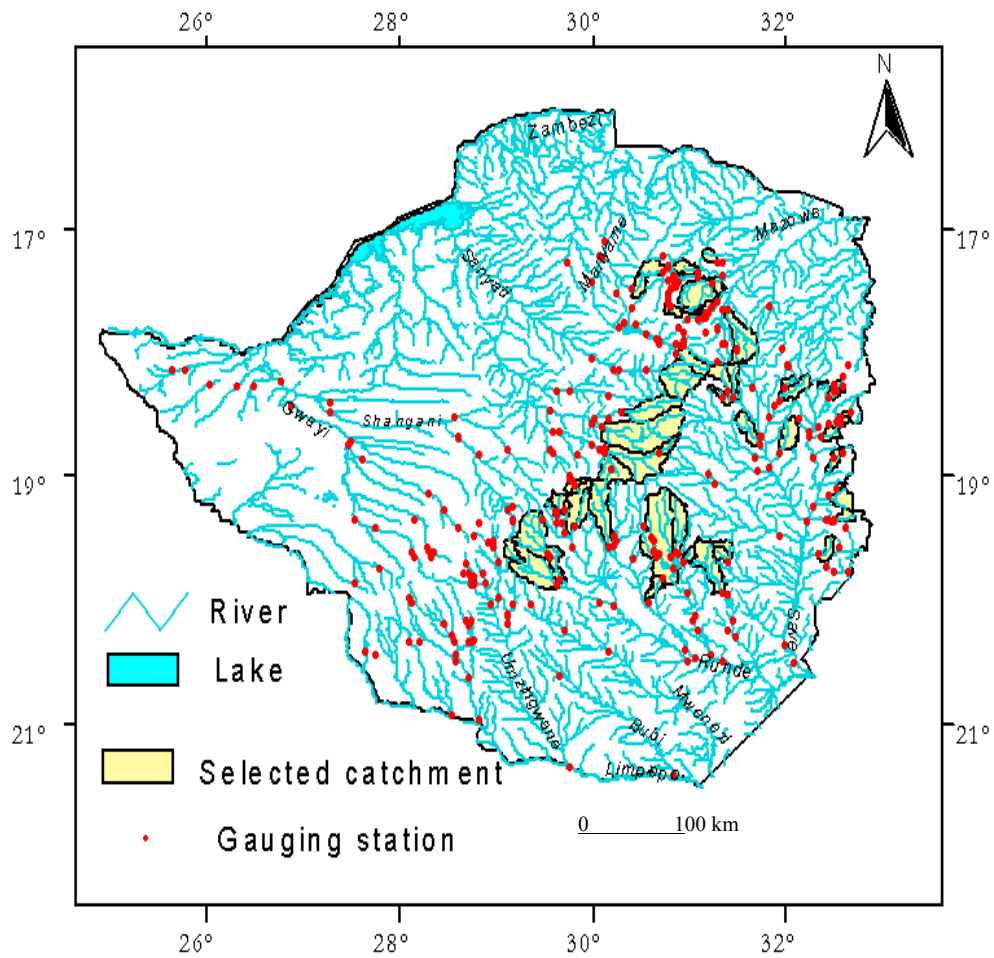
The main consideration in selecting catchments for inclusion in this study is the availability of flow data on each catchment to enable accurate estimation of flow statistics like means of daily and monthly flows, flow duration curves, and separation of base flows from total flow. With regards to base flows, flow data should not have gaps in most of the seasons so that base flow separation can be undertaken. Previous studies in Zimbabwe showed that a minimum of 10 years

of flow data gives a reasonable estimation of most flow statistics (Bullock, 1988). This criterion is used to select catchments to be included in this study.

Almost all river flow measurements in Zimbabwe are done using flumes and weirs. Discussions were held with staff of the Department of Water Development (DWD), and the Zimbabwe National Water Authority (ZINWA) to identify flow measuring stations with accurate flow data. These organizations have for each station a file documenting maintenance of the station and the accuracy of the rating curve of the flow measuring structure. These files were reviewed to identify stations with acceptable flow data. DWD undertook detailed assessments of rating curves of stations within the Odzi, Manyame, and Mazowe catchments in connection with rainfall-runoff modelling exercises during the 1990's. The reports of these assessments were considered in selecting flow measuring stations.

Non-parametric tests were applied to monthly and annual flows of stations that had a potential for being included in this study (Kite, 1988). These tests aimed at eliminating those stations with data showing statistically significant changes in their time series that can be due to measurement errors, upstream abstractions or impoundments. Stations with abstractions or impoundments that amounted to more than 10% of their estimated mean annual runoff were excluded.

The number of selected catchments should be large enough to enable application of both univariate and multivariate analysis methods. In the case of cluster analysis the number of catchments should ideally be larger than 30. This study selected 52 catchments with areas varying from 3.5 to 2630 km<sup>2</sup> (Table 2.3 and Appendix 1). The locations of these catchments are shown in Figure 2.5, while Appendix 1 provides the name of river and flow measuring station, and catchment area for each of these catchments. The station codes (C6, C13, etc) are used throughout this thesis to refer to these catchments, and Appendix 2 shows locations of these catchments and their codes.



**Figure 2.5:** Locations of flow measuring stations in Zimbabwe, and selected catchments comprising the study area.

**Table 2.3:** Distribution of selected catchments according to catchment area

Size of Catchment km <sup>2</sup>	Number of Catchments
< 100	12
101 – 500	23
501 – 1000	6
1001 – 2000	10
>2001	1

The range of areas of selected catchments is representative of catchment sizes used for water resources planning and design in Zimbabwe. The hydrological year is used in this study, and it starts in October and ends in September of the following year.

### 2.3 Derivation of flow characteristics

This study develops techniques for use at the drainage basin level to provide information for the following purposes;

- allocation or licensing of abstraction and storage of water,
- determination of instream or environmental flow requirements,
- yield analysis of proposed reservoirs,
- water quality management in terms of estimating effluent dilution requirements.

These tasks are undertaken at the drainage basin or meso-scale level. Therefore, selected flow and catchment characteristics should reflect meso-scale hydrological characteristics. Relevant flow characteristics are

- mean annual flow,
- mean monthly flows,
- coefficient of variation of flows,
- flow duration curves,
- base flow statistics,
- average number of days without flow in a year.

These flow characteristics are generally referred to as low flow measures (IH, 1980; Gustard, *et al.*, 1989; Smakhatin and Toulouse, 1998). According to a survey undertaken by Smakhatin *et al.* (1995) in South Africa, these flow characteristics were used for environmental impact assessment by 65% of the water resources practitioners, for water resources research by 55%, and for water supply and water quality management purposes by 50% of the same practitioners. The same study showed that information about flow duration and low flow frequency was required by 70% of water resources specialists, and 40% of the same specialists or practitioners indicated that they also required information on base flows. Although no similar survey has been undertaken in Zimbabwe, it is likely that the demands for such information are of similar importance. Consequently these flow characteristics are selected for use in this study.

The flow duration curve is estimated from daily flows using the method described by IH (1980) and Gustard *et al.* (1989). Daily flows,  $q_t$ , for each catchment are divided by the average daily flow ( $\bar{q}$ ) to give dimensionless flows so as to exclude the effects of catchment size as was the case in the study

by Gan *et al.* (1990) in Australia. The flow duration curve is used to derive percentile flows  $q_p$  where  $p$  is the exceedance probability.

### 2.3.1 Base flow and recession constant

Base flow consists of the contribution of groundwater flow and delayed interflow to total flow, and this makes up most of the dry weather flow (Kirkby, 1978; Linsley *et al.*, 1982). Most regionalisation studies attempt to determine how magnitudes of base flows vary from one catchment to the other. A common quantitative measure of base flows is the base flow index (*BFI*) which is the proportion of the volume of base flows to the volume of total flows within a specified period (IH, 1980; Bullock, 1988; Gustard *et al.*, 1989). Automated base flow separation techniques have been developed to estimate *BFI* from flow time series. The most commonly used are (a) the smoothed minima technique (IH, 1980), and (b) the recursive digital filter (Lyne and Hollick, 1974).

The following procedure is used when separating base flows using the smoothed minima technique.

- Daily flows,  $q_t$ , are divided into  $m$  non-overlapping five day blocks starting at the beginning of the daily flow time series.
- For each block the minimum daily flow is identified, and these form the  $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \dots, \tilde{q}_m$  series of minima.
- Turning points among the  $\tilde{q}_i$  are identified such that values multiplied by 0.9 are smaller than both neighbours, i.e.  $\tilde{q}_i$  is a turning point if  $0.9\tilde{q}_i < \tilde{q}_{i-1}$  and  $0.9\tilde{q}_i < \tilde{q}_{i+1}$ .
- The turning points become base flow ordinates. Values between these ordinates are interpolated under the condition that the base flow cannot exceed the total daily flow, since base flow is part of the daily flow.

A recursive digital filter separates the high frequency signals that are produced by surface runoff, from low frequency signals due to base flow (Arnold, *et al.*, 1995):

$$q_{s,t} = \beta q_{s,t-1} + \frac{(1+\beta)}{2}(q_t - q_{t-1}) \quad (2.1)$$

where  $q_{s,t}$  is the filtered surface runoff at time  $t$ ,  $\beta$  is the filter parameter which has to be estimated, and  $q_t$  is the total daily flow. The base flow,  $q_{g,t}$ , is defined as

$$q_{g,t} = q_t - q_{s,t} \quad (2.2)$$

and the base flow index is given by

$$BFI = \frac{\sum_{t=1}^{n_d} q_{g,t}}{\sum_{t=1}^{n_d} q_t} \quad (2.3)$$

where  $n_d$  = total number of days in the flow record. Nathan and McMahon (1990b) found a correlation coefficient of 0.94 between *BFI*'s estimated using smoothed minima and recursive digital filter techniques. Arnold *et al.* (1995) found that the *BFI* estimated using the smoothed minima and recursive digital filter technique was comparable to that obtained using manual methods. No significant differences between these two techniques have been reported in the literature. Recent studies in southern Africa have used the smoothed minima technique (Bullock, 1988; Bullock *et al.*, 1997). This justifies selection of the smoothed minima technique in this study so that the results are comparable with those of other studies. For each year with flow data on a particular catchment, a *BFI* is estimated, and then the average of all the annual *BFI*'s gives the catchment *BFI*.

The one day recession constant,  $\alpha_d$ , was derived from the flow recession equation (Arnold *et al.*, 1995; Chapman, 1999):

$$\alpha_d = \frac{1}{\Delta t} \log \left( \frac{q_0}{q_t} \right) \quad (2.4)$$

where  $q_0$  = daily flow at  $t = 0$

$\alpha_d$  = recession constant derived from daily flows

$\Delta_t$  = difference in time (days) between  $q_t$  and  $q_0$ .

The recession constant,  $\alpha_d$ , varies from 0 to 1 and gives an indication of the rate of depletion of flows. This constant depends mainly on topography, drainage pattern and soil types. Rivers draining areas with substantial

subsurface storage, and therefore gradual depletion of dry season flows have  $\alpha_d$  close to 1, e.g. limestone region with karst features. Those rivers draining areas with minor base flow contribution such as in a granitic terrain have  $\alpha_d$  close to 0.

## 2.4 Selection and derivation of catchment characteristics

Problems occur in selecting catchment characteristics or descriptors for use in developing methods to estimate flow characteristics of ungauged catchments. Ideally those catchment characteristics that have the strongest influence on flow characteristics of interest should be selected, but these are not known *a priori*. The literature provides some guidelines (Table 2.4).

**Table 2.4:** Catchment characteristics used in similar studies

Catchment Characteristic	Study
Area	Tasker (1982), NERC (1975), Gustard <i>et al.</i> (1989), Gan <i>et al.</i> (1990), Nathan and McMahon (1990a), Riggs (1990), Burn and Boorman (1993), Sefton and Howarth (1998), Bullock (1988), Bullock <i>et al.</i> (1990).
Elevation	Nathan and McMahon (1990a), Gustard <i>et al.</i> (1989), Tasker (1982)
Main stream length	Nathan and McMahon (1990a), Gustard <i>et al.</i> (1989), Burn and Boorman (1993)
Slope	Nathan and McMahon (1990a), Gustard <i>et al.</i> (1989), Sefton and Howarth (1998), Burn and Boorman (1993), Lacey and Grayson (1998), Berger and Entekhabi (2001)
Stream frequency	Sefton and Howarth (1998), Nathan and McMahon (1990a), Gustard <i>et al.</i> (1989), Burn and Boorman (1993).
Drainage density	Nathan and McMahon (1990a), Lacey and Grayson (1998), Berger and Entekhabi (2001)
Proportion of catchment under various soil types	Sefton and Howarth (1998), Burn and Boorman (1993), Gustard <i>et al.</i> (1989) Tasker (1982)
Land cover	Sefton and Howarth (1998), Lacey and Grayson (1998), Nathan and McMahon (1990a)
Proportion of catchment under various types of geological formations	Sefton and Howarth (1998), Nathan and McMahon (1990a), Gustard <i>et al.</i> (1989), Yokoo <i>et al.</i> (2001)
Location - latitude and longitude	Sefton and Howarth (1998), Nathan and McMahon (1990a).

This study took into account that selected catchment characteristics should be derived from sources that are readily available to practising hydrologists, i.e., maps, satellite imagery, and national databases. Selected catchment characteristics are given Table 2.5.

**Table 2.5: Catchment characteristics selected for use in the study**

Catchment Characteristic	Description and Data Source
1. Means of monthly and annual rainfall, and the average number of rainy days per year	Estimated from rain gauge data.
2. Maximum, average, and minimum catchment elevation	Derived from a digital elevation model (DEM) with a geographic projection of 30 Arc seconds
3. Drainage density	1:50 000 topographical maps and blue lines on these maps are assumed to represent streams.
4. Slope	Estimated from a DEM
5. Proportions of the catchment with different lithologies	Derived from the 1:500 000 Hydrogeological Map of Zimbabwe (Interconsult A/S, 1985)
6. Proportions of the catchment with different land cover types	Determined from the 1:250 000 Vegetation Map of Zimbabwe produced by the Forestry Commission from the 1992 LANDSAT TM images (Kweshu, 2000)
7. Means of monthly and annual potential evaporation	USA Class A pan evaporation measurements
8. Normalized difference vegetation index ( <i>NDVI</i> )	Obtained from SADC/RRSU archive (SADC/RRSU, 2000)

#### 2.4.1 Mean annual precipitation and number of rainy days

Rainfall stations that are within or close to each of the 52 catchments were identified. This study uses all stations with over 10 years of continuous data since some catchments do not have a minimum of the recommended 30 years of rainfall data (Figure 2.6) (Dent *et al.*, 1990). Mean annual precipitation ( $\bar{P}_{yr}$ ) for each of the catchments was estimated using the arithmetic mean method, and this varies from 604 to 1770 mm yr<sup>-1</sup> (Figure 2.6). Catchments that occur along the central part of the country have mean annual precipitation in the 600 to 800 mm yr<sup>-1</sup> range. Mean annual precipitation of catchments in the northern part is in the 800 to 1000 mm yr<sup>-1</sup> range. The highest rainfall is on catchments located on the Eastern Highlands (1300 to 1770 mm yr<sup>-1</sup>) which indicates that precipitation increases with altitude.

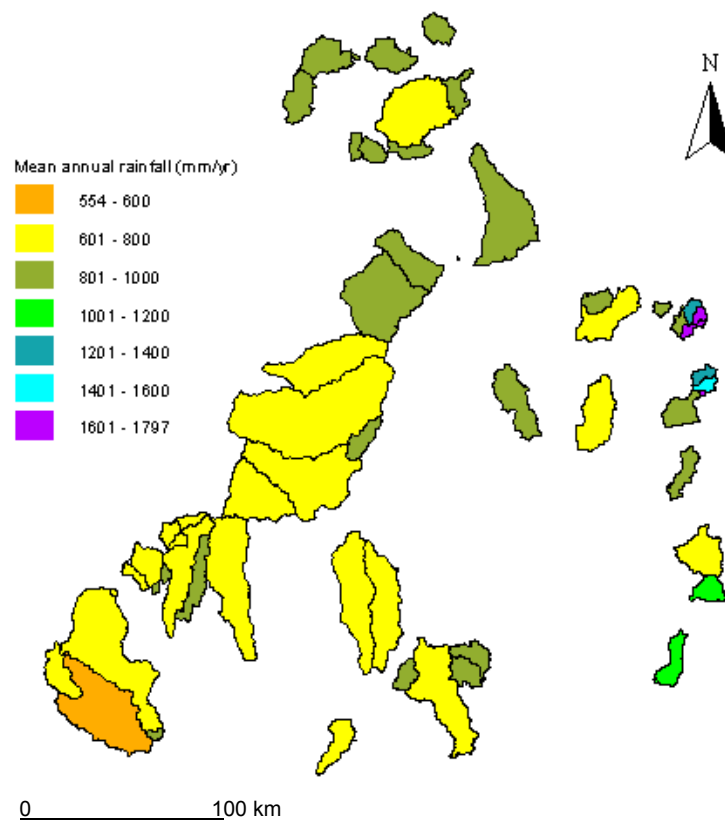
De Groen (2002) demonstrated that the amount of rainfall intercepted per month is related to the number of rainy days. Therefore, the average number of rainy days per year is likely to affect the volume of runoff, and is included as a

catchment characteristic. The average number of rainy days per year,  $\bar{N}_{yr}$ , varies within the study area from 42 days  $yr^{-1}$  in the extreme south-western part, 52-58 days  $yr^{-1}$  on the central part, 62-71 days  $yr^{-1}$  on northern catchments, and 105-125 days  $yr^{-1}$  on the Eastern

Highlands. A strong linear relationship occurs between the average annual rainfall and the average annual number of rainy days. A high correlation,  $r = 0.83$ , exists between  $\bar{P}_{yr}$  and  $\bar{N}_{yr}$ .  $\bar{N}_{yr}$  at a rainfall station can be predicted from the following equation

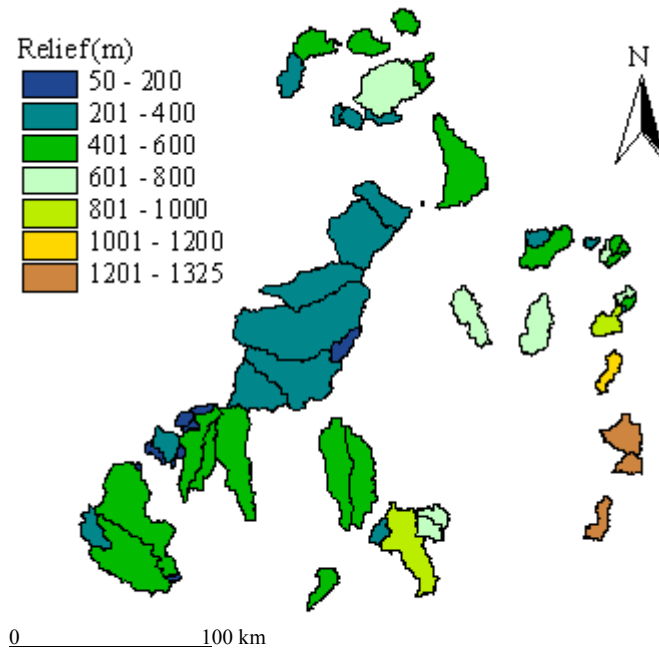
$$\bar{N}_{yr} = 22.905 + 0.053\bar{P}_{yr} \quad \text{days } yr^{-1} \quad r^2 = 0.69 \quad (2.5)$$

This result is in agreement with De Groen (2002) who established a relationship between monthly rainfall and the number of rainy days in a month.



**Figure 2.6:** Mean annual rainfall of the study area.

## 2.4.2 Maximum, average and minimum elevations, and relief



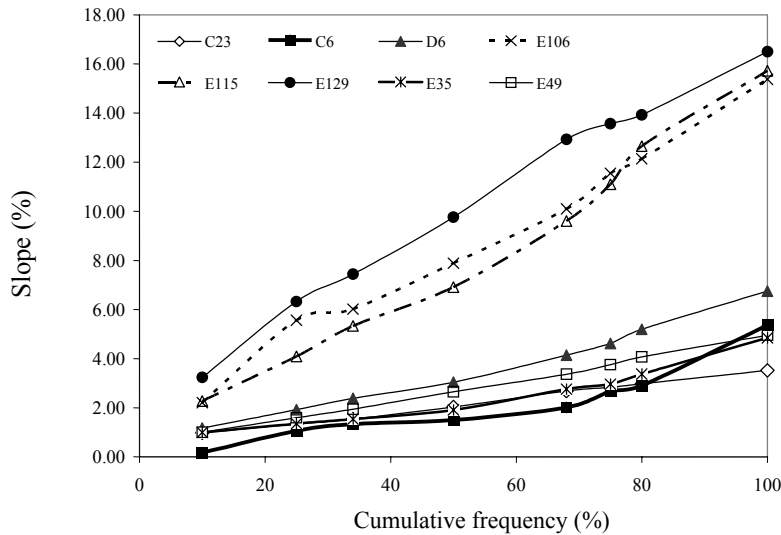
**Figure 2.7:** Catchment relief derived from the difference between the highest and lowest pixel within a DEM.

The estimation of drainage basin characteristics such as elevation has traditionally been hindered by the time consuming nature of this exercise when topographic maps are used (Meijerink, *et al.*, 1994). A digital elevation model (DEM) with a geographic projection of 30 Arc seconds which is approximately one km was used to derive maximum, average, and minimum elevations of catchments. The maximum elevation varies from 1100 to 2250 m (Figure 2.7). Catchments that are along the Eastern Highlands have the highest elevation. C23, C47 and C70 have the lowest relief (50 to 305 m). The highest relief occurs on Eastern Highlands catchments where it varies from 1000 to 1300 m.

### 2.4.3 Slope

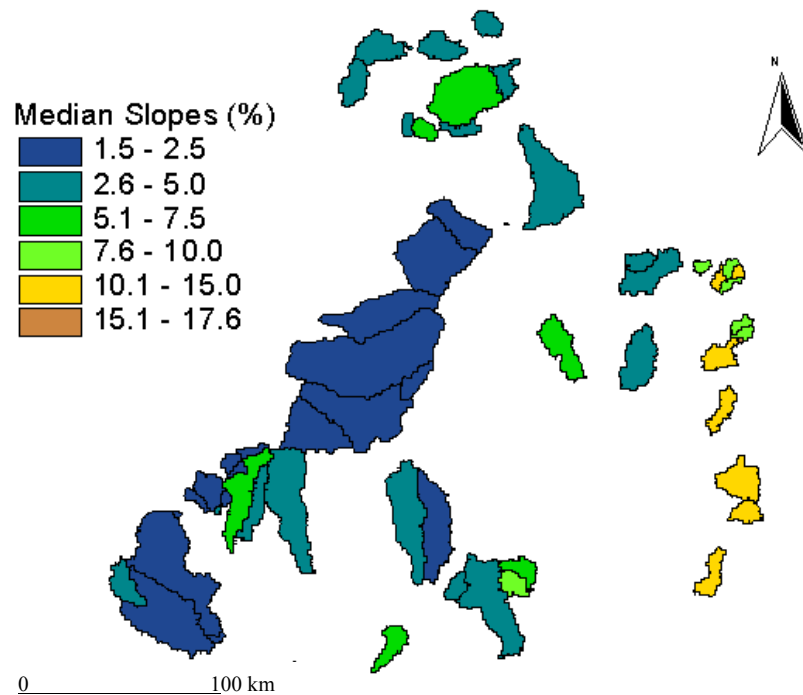
Slope is an important characteristic of a catchment as it gives an indication of the kinetic energy available for water to move towards the basin outlet, and it has been found to be related to total runoff and base flows (Bullock, 1988; Vogel and Kroll, 1992). Slope is highly variable within a basin, and hence no single measure of slope is commonly agreed upon. Before the advent of DEMs, it was almost practically impossible to derive slopes for all the landscapes within a basin. Consequently various slope measures assumed to be representative for the effects of slope on runoff processes have been used (Drayton *et al.*, 1980; IH, 1980; Seyhan and Keet, 1981; Bullock, 1985; Gustard *et al.*, 1989; Nathan and McMahon 1990a). A single slope index for the whole basin may not be representative for all the landscapes that affect runoff processes. This study uses a DEM to estimate slopes for all pixels within a catchment, and then constructs a cumulative frequency distribution of slopes from which slope indices  $S_{\psi}$  are derived.  $S_{\psi}$  denotes a slope value for which  $\psi$  % of the pixels in a basin are equal to or less than this value. Berger and Entekhabi (2001) recommended the use of  $S_{50}$ , the median slope, instead of the average slope which they considered to be unrepresentative. This study includes several slope indices for values of  $\psi$  from 5 to 95% to identify  $S_{\psi}$  that best explains each of the flow characteristics.

Figure 2.8 shows cumulative frequency diagrams of slopes for 8 selected catchments. The shapes of these curves indicate whether relatively flat or steep slopes dominate in particular catchments. Curves for E106 and E129 in the Eastern Highlands show major differences between the largest and smallest slopes. These catchments will have fast flowing rivers. There are no major differences between the largest and smallest slopes for catchments C6, C23 and E49 on the central part of the country, and therefore river flows will have relatively low velocities. The median slope,  $S_{50}$ , is shown in Figure 2.9.



**Figure 2.8:** Cumulative frequency distribution of slopes on 8 selected catchments.

Catchments located on the central part of the country have the lowest median slopes, varying from 1.5 to 3.04%. The effects of the Shurugwi Hills are evident on E42 which has a median slope of 5.5% while neighbouring catchments have median slopes in the 1.5 to 3% range. The hilly topography around the Great Zimbabwe National Monument has also resulted in E107 having a median slope of 6% while neighbouring catchments have lower slopes. E114 and E115 drain the Bikita Highlands that have numerous dwalas with steep slopes, and therefore these catchments have median slopes of 10% and 7% respectively. The highest median slopes range from 12 to 17% and occur on the Eastern Highlands catchments.

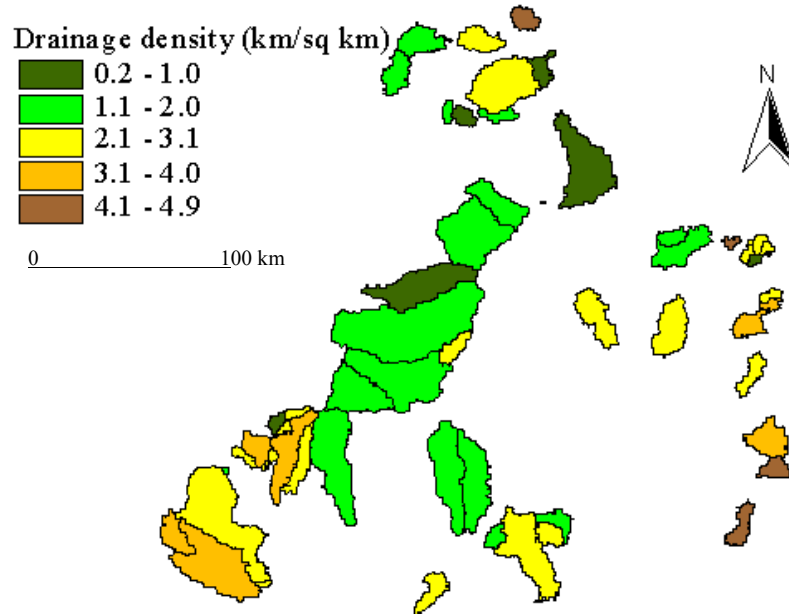


**Figure 2.9:** Median slopes of the study area.

#### 2.4.4 Drainage density

Drainage density ( $Dd$ ) is derived by dividing the total stream length within a catchment by the catchment area, and is regarded as an important landscape characteristic (Gregory and Walling, 1973; Seyhan, 1977). It is a measure of how dissected a basin is, and it is expected that  $Dd$  affects the transformation of rainfall into runoff (Seyhan and Keet, 1981; Pitlick, 1994; Tucci, 1995; Berger and Entekhabi, 2001). The main deterrent to the use of  $Dd$  is that it is laborious and time consuming to estimate from aerial photographs or topographical maps. In addition the definition of a stream is not consistent among mapping agencies (Gregory and Walling, 1973; Seyhan and Keet, 1981). This study assumed that blue lines on 1:50 000 topographical maps produced by the Surveyor General of Zimbabwe are representative of stream networks. While drainage densities estimated from these maps may not be accurate in an absolute sense, they allow a comparison of the effects of differing intensities of dissection on runoff among catchments. Streams on all selected catchments are digitized, and the total stream length for each catchment is then estimated using standard routines available in most GIS packages. Values of catchment areas used here are those

which are officially used by the Department of Water Development in Zimbabwe.



**Figure 2.10:** Drainage density derived from blue lines on 1:50 000 topographical maps.

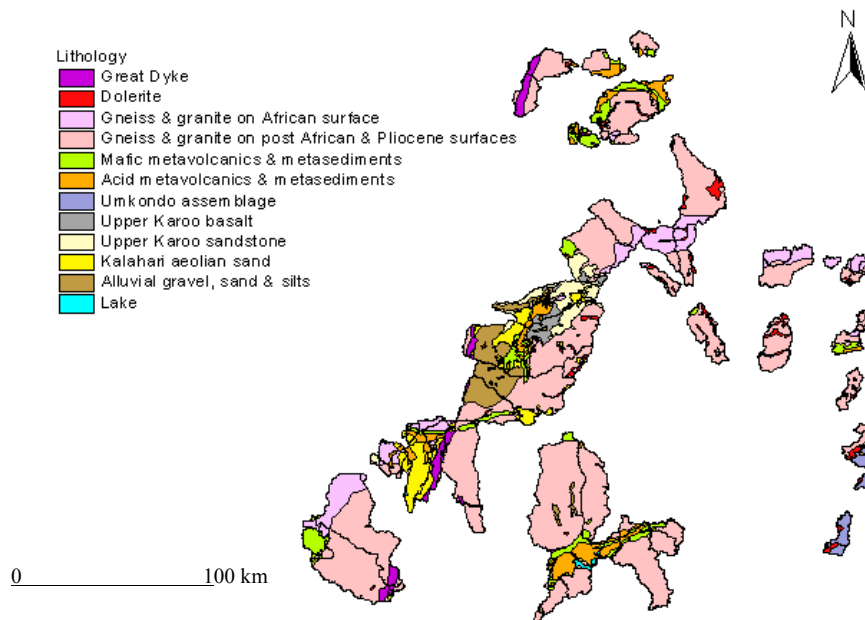
Drainage density ( $Dd$ ) varies from 0.22 to 6.30 km km<sup>-2</sup> (Figure 2.10). In general the central part of the country has the lowest  $Dd$  values, while the Eastern Highlands region has the highest values. The region within which C23, C70, C6, C47, and C41 are located has numerous dambos. Dambos are seasonally waterlogged valleys with grasslands. Similar features occur on the E49. These basins have  $Dd$  of approximately 1.5 km km<sup>-2</sup>.  $Dd$  is positively correlated with the median slope,  $r = 0.65$ , and relief,  $r = 0.42$ . Thus areas with steep slopes and high  $Dd$  are expected to have fast channel flow. Low  $Dd$  in Zimbabwe is associated with pedimentation (sheet wash plains) and therefore low groundwater contribution to streamflows. A somewhat weak relationship exists between  $Dd$  and  $\bar{P}_{yr}$  ( $r = 0.35$ ). All these correlation coefficients are significant at the 5% level. The relationship between  $Dd$  and  $\bar{P}_{yr}$  can be explained by both being positively correlated with elevation. The lack of a strong relationship between  $Dd$  and  $\bar{P}_{yr}$ , shows that current drainage networks reflect the long geomorphological history of the landscape and its associated

previous climates. Geological processes such as faulting, fracturing, uplifting, and pediplanation influenced the creation of current drainage networks. In addition previous changes in climate like the drying during the Miocene which resulted in the deposition of large quantities of aeolian sands caused changes in drainage patterns in Zimbabwe (Lister, 1987). Current precipitation patterns have not had major effects on the development of the main drainage lines. There is no relationship between  $Dd$  and the proportion of the catchment that is covered by different lithologies. The type of lithology underlying a particular area may not be as important in affecting drainage density as geomorphological processes that have shaped the landscape.

#### **2.4.5 Geology**

The derivation of quantitative geological indices that express the geological effects on runoff processes at the basin level is a major challenge in hydrology. Hydrogeological characteristics like permeability and depth to the water table that have been used in some studies are highly variable in space. In developing countries such data are rarely available, since most countries only developed and maintained geological databases relevant for planning and managing mining operations. Therefore, most regionalisation studies used the proportions of catchments with different lithologies (Gustard *et al.*, 1989; Nathan and McMahon, 1990a; Sefton and Howarth, 1998; Yokoo *et al.* 2001). This study uses the same approach since detailed geological mapping has not covered the whole of Zimbabwe. The generalised 1:500 000 Hydrogeological Map of Zimbabwe is used (Interconsult A/S, 1985). This map has 17 lithological classes developed for assessing the potential for groundwater occurrence, and development of rural water supply systems abstracting groundwater.

The study area is underlain mostly by rocks belonging to the crystalline basement of Africa comprising granite-gneiss-greenstone belts of the Archaean craton (Key, 1997). These rocks lack primary porosity, and aquifers only occur within the weathered regolith and fractured bedrock (Key, 1997; Wright, 1987; Wright, 1997). Aquifers have localized groundwater flow systems, with recharge on the interfluvial and discharge in valley bottomlands (Wright, 1997).



**Figure 2.11:** Lithology of the catchments derived from the 1: 500 000 Hydrogeological map (Interconsult A/S, 1985).

Figure 2.11 shows the different types of lithologies occurring within the study area and Table 2.6 below shows the coverage of the dominant lithologies. Average water table depths and borehole yields presented in Table 2.6 were obtained from Interconsult A/S (1985).

**Table 2.6:** Lithological types of the study area, their coverage, average water tables and borehole yields

Lithology	No. of Catchments	% Area	Water Table Depth (m)	Borehole Yield ( $\text{m}^3 \text{ day}^{-1}$ )
Gneiss and young intrusive granite on the African surface	30	0.3 – 100.0	< 10	50 – 100
Gneiss and young intrusive granite on post African and Pliocene surface	53	0.6 – 99.1	< 10	10 - 50
Mafic metavolcanics (Greenstone)	31	0.2 – 70.0	10 – 20	10 – 250
Acid metavolcanics (Greenstone)	27	0.2 – 63.5	< 10	10 - 25
Dolerite dykes and sills	31	0.1 – 30.4	< 10	25 - 100
Kalahari aeolian sands	13	1.3 – 80.9	> 20	100 - 1000
Alluvial deposits	7	2.5 – 40.8	variable	100 - 5000
Umkondo assemblage	3	51.6 – 92.6	5 – 20	10 - 100
Upper Karoo basalt	3	5 – 10	5 – 15	20 – 100
Upper Karoo sandstone	3	17.0 – 43.0	> 20	50 – 300
Great Dyke	13	0.1 – 50.0	unknown	unknown

The following notation is used to denote the proportion in percentages of a catchment with the above lithologies;

$GL_{GG}$ granite and gneiss	$GL_{GR}$ greenstones
$GL_{DO}$ dolerite dykes and sills	$GL_{KL}$ Kalahari sands
$GL_{AL}$ alluvial deposits	$GL_{LM}$ Umkondo group
$GL_{BA}$ upper Karoo basalt	$GL_{SA}$ upper Karoo sandstone
$GL_{GD}$ Great Dyke	

The dominant lithologies are gneiss and young intrusive granite (Table 2.6 and Figure 2.11). The potential for groundwater occurrence in granites is variable depending on the depth and areal extent of both fracturing and secondary weathering. Weathering depths tend to be considerable on the African surface, and rather limited on the Post-African surface with prominent rocks outcrops (Interconsult A/S, 1985). The potential for groundwater occurrence in granite is favourable in areas where it is closely jointed resulting in castle koppies and balancing rocks as part of the landscape. Areas with bornhardts that are rounded or whale-back shaped have limited potential due to shallow depth of weathering. Well yields in granites in general are variable. Water yielding

properties of granite also depend on its texture (Jordan, 1968; Meijerink, 1974). Therefore, the contribution of groundwater to runoff may vary between catchments although all of them could be underlain by granite. Granite often gives rise to sandy soils that are moderately deep, and with high infiltration rates. Due to low water holding capacity of these sandy soils and shallow depth of impermeable underlying granite, shallow water tables generally occur during the rainy season (Thompson and Purves, 1978).

Mafic and acid metavolcanics commonly referred to as the greenstone formations are frequent on catchments in the northern part of the study area, and on parts of C6, C18, E45, E49, and E1 (Figure 2.11). Mafic metavolcanics tend to have considerable depth of weathering, and high potential of groundwater occurrence. Groundwater contribution to runoff is expected to be relatively high in comparison to granites. The high potential for groundwater occurrence in these formations has led in some cases to over extraction, and this has adversely affected groundwater contribution to some rivers (Jordan, 1968). Mafic rocks give rise to red clays with considerable water holding properties (Thompson and Purves, 1978).

Acid metavolcanics have rather limited potential for groundwater occurrence due to variable depth of weathering.

Dolerite dykes and sills are numerous within the study area, and occur as intrusions into the granite and gneiss (Figure 2.11). They generally occur on small portions of the catchments. The degree of weathering of the dykes and sills is variable, ranging from fresh to decomposed. Groundwater occurrence is favourable where the dolerite has been weathered, or along the lower contact zone between the sill and granite or gneiss. Where aquifers occur, they tend to have water tables with depths less than 10 m.

Kalahari sands cover over 50% of E156, E40, and E42. They also cover 1-30% of C13, C18, C41, and C47 (Figure 2.11). Kalahari sands are generally fine to medium grained unconsolidated sands (Interconsult A/S, 1985). They have primary porosity and permeability, and the potential for groundwater occurrence is very high. Aquifers in these formations are unconfined with water table depths greater than 20 m. Rivers draining Kalahari sands may not benefit from groundwater flow because water tables are often below river beds.

Alluvial deposits comprising gravel sand and silts cover 26% of C41, and 41% of C47 (Figure 2.11). They also cover small portions of C6, and C18. Alluvial deposits have primary porosity and permeability, and a high potential for groundwater occurrence. Water table depths are variable.

The Umkondo assemblage consists of quartzite, shale, limestone and dolerite intrusions, and occur on three catchments in the Eastern Highlands, E37, E121, and E125. They lack primary porosity. Groundwater tends to occur along the shale/dolerite, and quartzite/dolerite contacts, and fractures within quartzites. Karst features have not been reported on these catchments (Interconsult A/S, 1985).

Upper Karoo basalt and sandstone occur on C6, C18 and C70. The Great Dyke which is 3-10 km wide and 550 km long running from NNE to SW in Zimbabwe occurs on 13 catchments within the study area, for example C25, D28, D48, E29 and E40. It is an elongated intrusion comprising mafic and ultramafic rocks such as pyroxenites and gabbro. Very little hydrogeological information exists about the dyke.

#### **2.4.6 Land cover**

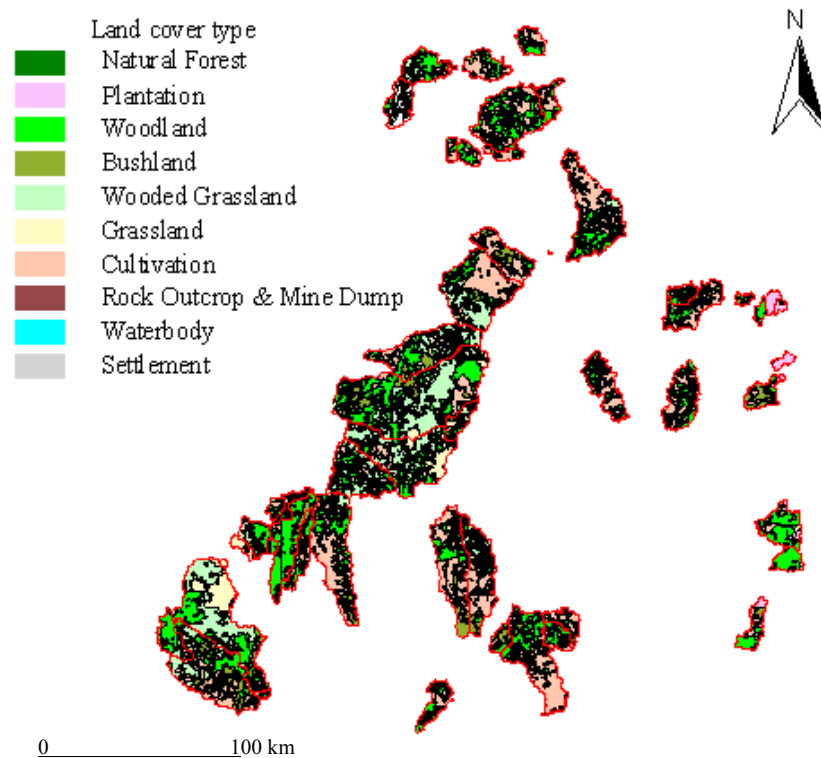
Land cover has been shown in several studies to affect flow characteristics (Edwards and Blackie, 1981; Bosch, 1979; Mumeka, 1986; Bosch and Hewlett, 1982; Andrews and Bullock, 1994). Land cover is selected for use in this study. The Vegetation Map of Zimbabwe has land cover classes given in Table 2.7, and these are used in this study (Kweshu, 2000).

**Table 2.7:** Land cover classes used derived from the 1:250 000 Vegetation Map of Zimbabwe

Land Cover Type (Notation)	Description
Forest plantation ( $LC_{PL}$ )	80 – 100% canopy cover and with height > 15 m. Exotic species.
Natural forest ( $LC_{FO}$ )	Canopy cover > 80%, tree height > 15 m. Moist evergreen and deciduous species.
Woodland ( $LC_{WD}$ )	Open to dense stand of indigenous trees. Canopy cover of 20 – 80%. Tree height 5 – 15m. Trees are widely spaced. An incomplete under storey of small trees and large bushes.
Bushland ( $LC_{BU}$ )	Indigenous woody cover with 20 – 80% canopy cover, and heights of 1-5m. Usually multi-stemmed.
Wooded grassland ( $LC_{WG}$ )	Clumped or scattered trees, or bushes with height 1 – 15 m. Canopy cover 2 – 20%. Bushes or tree clumps often occur on termite mounds.
Grassland ( $LC_{GR}$ )	Absent or scattered trees. Canopy cover of bushes or trees less than 2%. Often occur in areas that are seasonally waterlogged, e.g. dambos.
Cultivation ( $LC_{CU}$ )	Agricultural crop production including tea, coffee, banana, sugar plantations, and orchards.

The sum of the proportions of areas with wooded grassland and grassland is denoted by  $LC_{CG}$ .

Woodlands are the most dominant land cover within the whole study area, and cover 38% of the area, followed by cultivation which covers 30%. Grasslands and wooded grasslands cover 16%, and forest plantation and bushland each cover approximately 7% of the area (Figure 2.12).



**Figure 2.12:** Land cover types obtained from the Vegetation Map of Zimbabwe which was derived from 1992 Landsat TM images by the Forestry Commission of Zimbabwe.

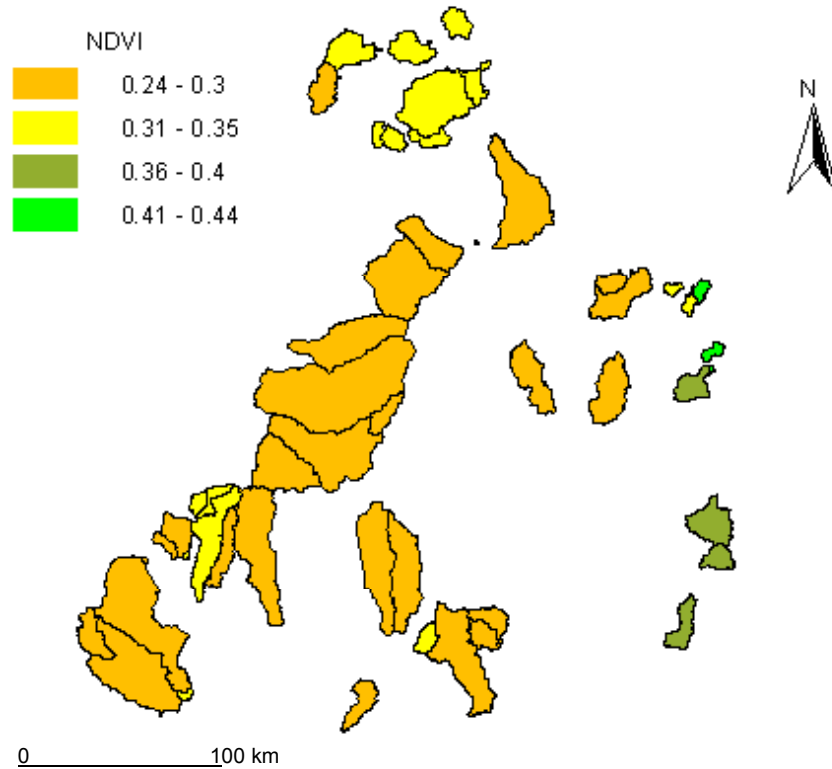
Forest plantations occur mostly on the Eastern Highlands and cover almost all of E72, E106, E127, E128, and E129. There are also small portions of some catchments not on the Eastern Highlands with plantation forests, e.g. 1.7% on C41 due to Mtao Forest. Woodlands, and cultivation occur on almost all catchments. Wooded grasslands occur mostly on catchments on the central part, where they cover 30-49% of the areas of C6, C18, C41, and C47. Grasslands cover almost the whole of E23 (90%), and E33 (97%). They also cover about 28% of C47. Catchments with both wooded grasslands and grasslands usually have dambos.

#### 2.4.7 Normalized Difference Vegetation Index

The normalized difference vegetation index (*NDVI*) gives an indication of the photosynthetic activity of vegetation, and is related to vegetation density

(Bastiaanssen, 1998; Meijerink, *et al.*, 1994). Several studies have found a relationship between *NDVI* and rates of evaporation, which suggests that *NDVI* is likely to affect flow characteristics (Hendricksen and Durkin, 1986; Cihlar *et al.*, 1991; Yang *et al.*, 1994). *NDVI* is therefore selected as one of the catchment characteristics. Dekadal *NDVI* data starting from October 1981 to 1999 are used to estimate monthly and annual *NDVI* for each catchment (SADC/RRSU, 2000).

The average annual *NDVI* varies from 0.24 to 0.44 (Figure 2.13). Catchments which occur on the central part of the country show some interesting differences in *NDVI* values. C23 and C70 have lower *NDVI* values than their neighbouring C6, C18, and C47. The differences are due to communal areas that are densely populated on the former catchments, while the latter are under large scale commercial farming. Most communal lands are dominated by cultivated lands on which vegetation has been cleared, resulting in low *NDVI* values. The Eastern Highlands region which is well vegetated has the highest *NDVI* values. The increase in *NDVI* from the central part of the country to the Eastern Highlands reflects the close relationship between *NDVI* and the leaf area index (*LAI*). An approximately linear relationship exists between *NDVI* and *LAI*, up to *LAI* = 3 to 4, after which the *NDVI* does not change significantly. A positive correlation exists between *NDVI* and  $\overline{P}_{yr}$  ( $r = 0.65$ ). This is a reflection of active vegetation growth in areas with high rainfall, which is the case in the Eastern Highlands, and poor vegetation cover in areas with low rainfall.



**Figure 2.13:** Average annual *NDVI* estimated from dekadal *NDVI* values from October 1981 to October 1999

### 2.4.8 Evaporation

Evaporation is an important component of the water budget in Zimbabwe where about 90% of the annual rainfall returns back to the atmosphere through this process. Evaporation has therefore considerable effects on runoff. The Penman type of equations for estimating potential evaporation cannot be used for most parts of the study area as they do not have all of the required meteorological data (Penman, 1948 & 1956; Schulze and Kunz, 1995). Methods that use temperature for estimating potential evaporation have a capability of being used in the study area (Hargreaves and Samani, 1985; Hargreaves and Hargreaves, 1985; ASCE, 1996). The Hargreaves and Samani (1985) method performed better than other temperature based methods in South Africa (Schulze and Kunz, 1995), and therefore has a capability for estimating potential evaporation rates of selected catchments. The following equation derived by Hargreaves and

Samani (1985) with an additional factor (1.25) to convert reference crop potential evaporation rate to a pan evaporation equivalent (Schulze and Kunz, 1995) is used:

$$E_{pan,t} = (1.25)0.0023R_a (T_{max} - T_{min})^{0.5}(T_a + 17.8) \quad (2.6)$$

where

$E_{pan,t}$  = monthly pan evaporation equivalent (mm/day),

$R_a$  = average monthly extra-terrestrial solar radiation (mm equivalent per day),

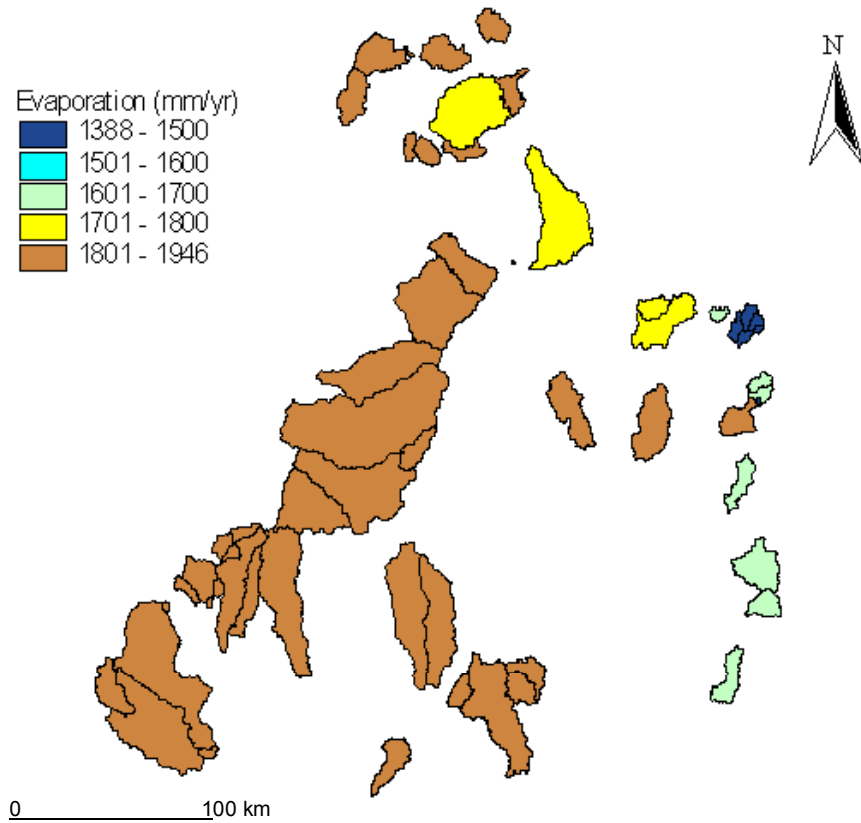
$T_{max}$  = maximum monthly temperature (°C),

$T_{min}$  = minimum monthly temperature (°C),

$T_a$  = mean monthly air temperature (°C).

Pan evaporation measurements are undertaken at 58 stations in Zimbabwe using the screened USA Class. These measurements are used to estimate potential evaporation rates for the selected catchments. Evaporation rates for catchments without pan measurements are estimated by extrapolating from nearby sites. A comparison of potential evaporation rates estimated by the Hargreaves and Samani (1985) method with pan evaporation measurements is conducted to identify the most suitable method for estimating monthly potential evaporation rates for selected catchments.

Figure 2.14 shows mean annual evaporation rates based on USA Class A pan measurements. Catchments on the southern and central part of the study area have mean annual evaporation rates of 1800 – 2000 mm yr<sup>-1</sup>. D6, D24, E136, E37, E121, and E125 have pan evaporation rates of 1600 – 1800 mm yr<sup>-1</sup>. Catchments with very high altitudes such as E127, E128, E129, and E132 have low pan evaporation rates, 1300 – 1400 mm yr<sup>-1</sup>. Evaporation rates decline with altitude since temperature also decreases with altitude. Figure 2.15 shows that low evaporation rates occur during the cool season, from May to July. Evaporation rates then increase and reach a maximum in October. The increase in cloud cover and humidity during the rainy season reduces evaporation rates.



**Figure 2.14:** Mean annual potential evaporation estimated from pan evaporation records

Evaporation rates estimated from the Hargreaves and Samani (1985) method are 1-10% less than pan evaporation rates during the August to October period, while these are 5-40% greater than pan evaporation during the rainy season, November to May (Figure 2.16) In general the differences between these two methods are less than 10% during the May to November period. It seems that the increase in relative humidity which reduces evaporation rates during the rainy season is not accurately reflected by the Hargreaves and Samani (1985) method. Consequently, this study uses evaporation rates based on USA Class A pan.

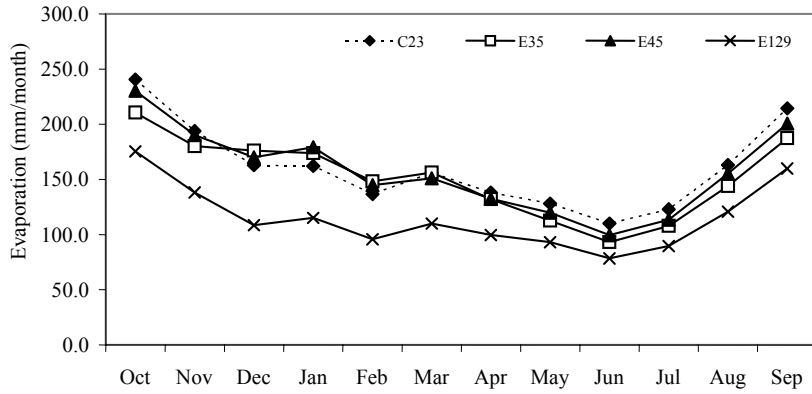


Figure 2.15: Mean monthly evaporation

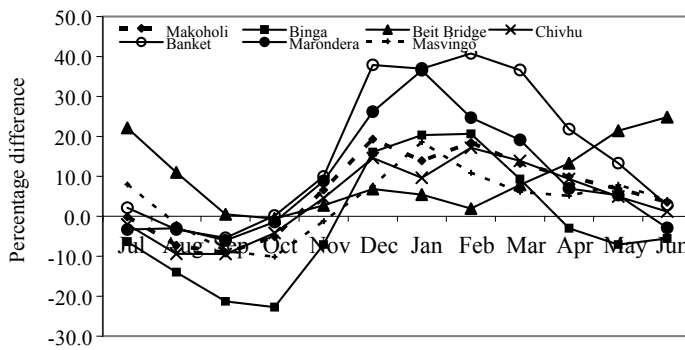
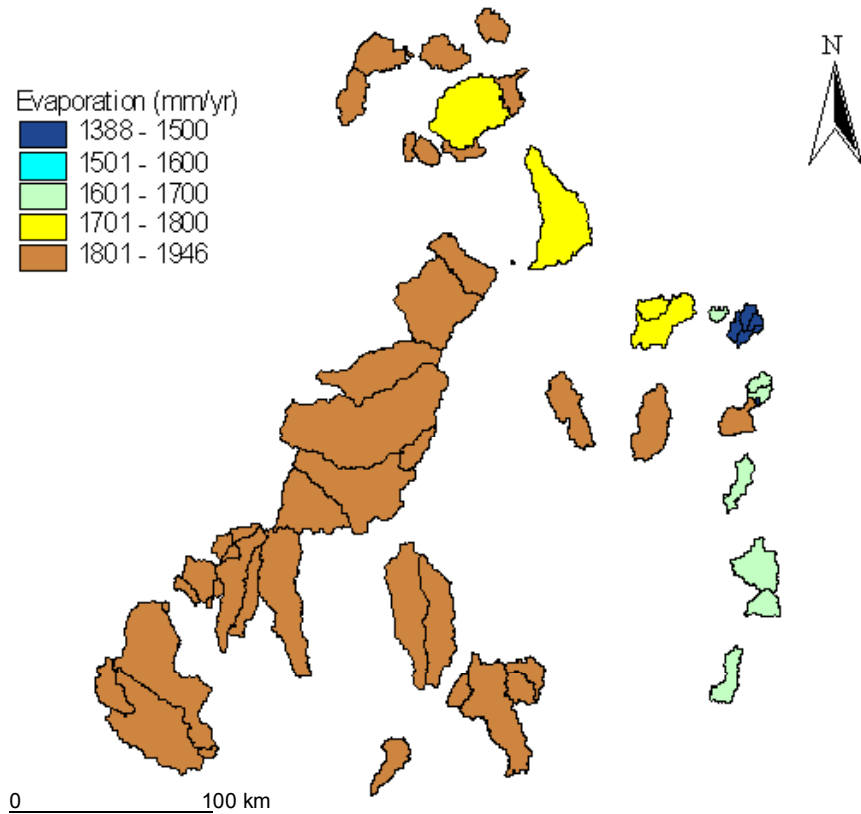


Figure 2.16: Differences between Hargreaves and Samani (1985) and pan evaporation rates

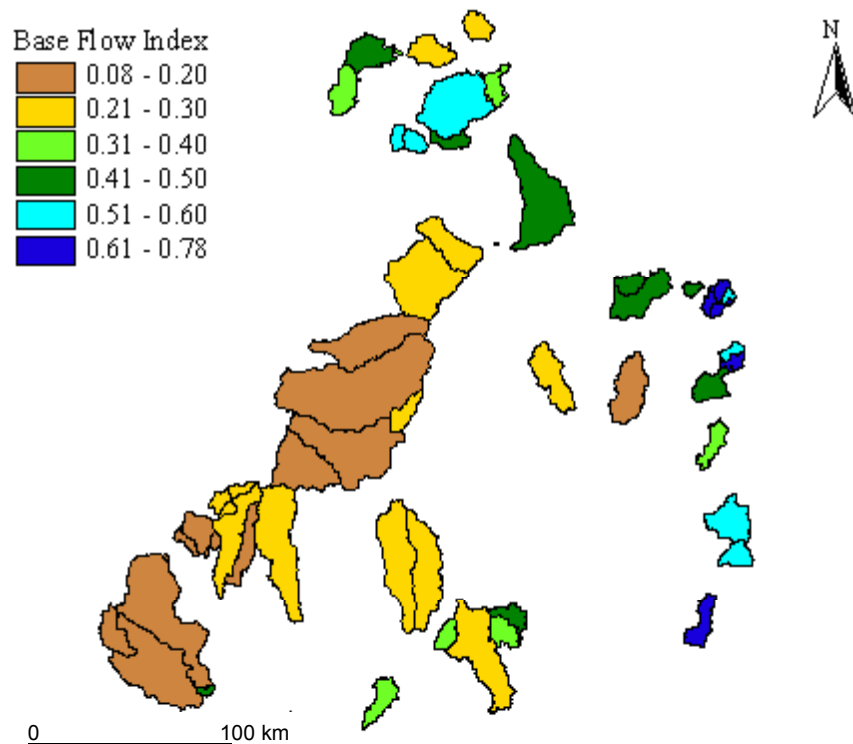
## 2.5 Flow characteristics

The mean annual runoff,  $\bar{Q}_{yr}$ , ranges from 38 to 45 mm yr<sup>-1</sup> for the south-western catchments, and from 45 to 85 mm yr<sup>-1</sup> for the central catchments (Figure 2.17). Most of the northern catchments have  $\bar{Q}_{yr}$  in the 100 to 200 mm yr<sup>-1</sup> range. The Eastern Highlands catchments have the highest  $\bar{Q}_{yr}$ , 200 to 460 mm yr<sup>-1</sup>. There is only one catchment, E72, with  $\bar{Q}_{yr}$  of 778 mm yr<sup>-1</sup>. The coefficient of variation ( $CV$ ) of annual flows is inversely related to the  $\bar{Q}_{yr}$ . Thus Eastern Highlands catchments have the lowest  $CV$ , 55 to 75%, while the south-western and central catchments have  $CV$ s between 120 and 160%.



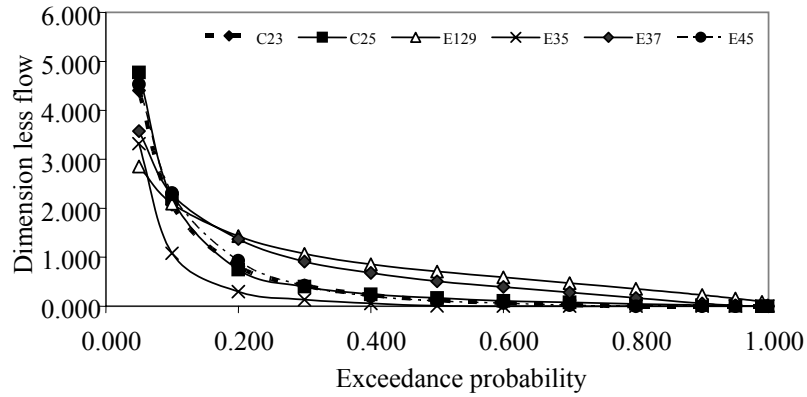
**Figure 2.17:** Mean annual runoff estimated from flow records

*BFI* ranges from 0.08 to 0.76 and the average *BFI* for all the catchments is 0.36 (Figure 2.18). Bullock (1988), and Bullock, *et al.* (1992) found a similar value for different sets of catchments in Zimbabwe. Central catchments have the lowest *BFI* values, while most of the northern and all the eastern catchments have high *BFI* from 0.44 to 0.76. Catchments with low *BFI* have the greatest variability in their annual *BFI*s. The *CV* of *BFI* was found to vary from 70 to 105% for the central and south-western catchments, while the eastern catchments had values ranging from 5 to 35%. *BFI* has a linear relationship with the recession constant,  $\alpha_d$ , derived using daily flows ( $r = 0.80$ ). The recession constant varied from 0.83 to 0.96 among the selected catchments. Catchments with small *BFI* have small recession constants indicating rapid depletion of subsurface storage.



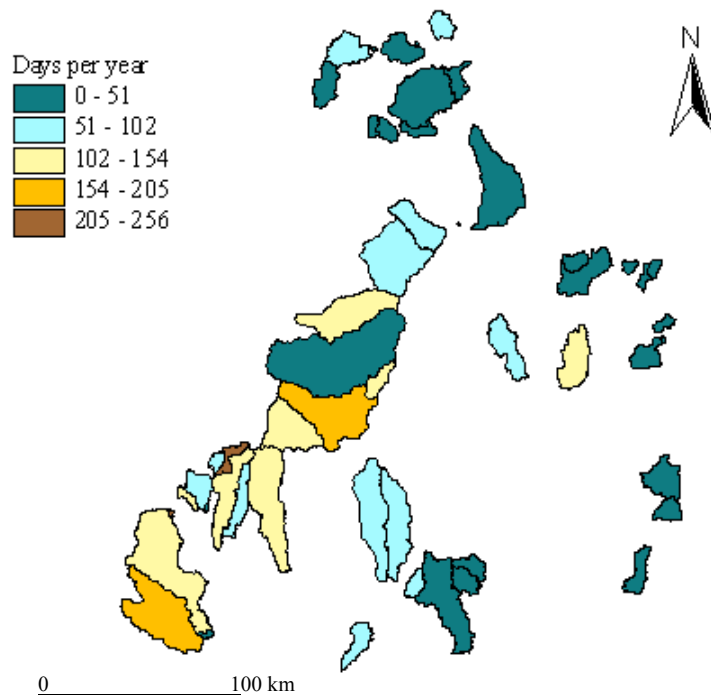
**Figure 2.18:** Base flow index estimated using the smoothed minima method.

Flow duration curves have a gradation from relatively dry catchments with curves at the bottom of Figure 2.19, to wet catchments with top most curves. The very bottom curves are representative of relatively dry catchments in the south-western and central parts of the study area. For these catchments the flow with an exceedance probability greater than 0.40 is the zero flow. These catchments have small *BFI* values and experience quick recession. For the relatively wet catchments, the flow with an exceedance probability of 0.30 is the annual average flow. The shapes of these curves reflect flow regimes of these catchments. Catchments with steep flow duration curves dry quickly, while flat curves indicate gradual depletion of flows.



**Figure 2.19:** Flow duration curves for some catchments

Most catchments in the eastern and northern parts have the number of days with no flow per year,  $\bar{N}_{DZ}$ , varying from 0 to 10 days, while for the extreme southwestern catchments this ranges from 130 to 244 days in a year (Figure 2.20).  $\bar{N}_{DZ}$  was found to be negatively correlated to both  $\bar{Q}_{yr}$  ( $r = -0.42$ ) and  $BFI$  ( $r = -0.70$ ), and positively correlated to  $CV$  ( $r = 0.50$ ).



**Figure 2.20:** Average number of days without flow per year

## 3 PREDICTION OF FLOW CHARACTERISTICS

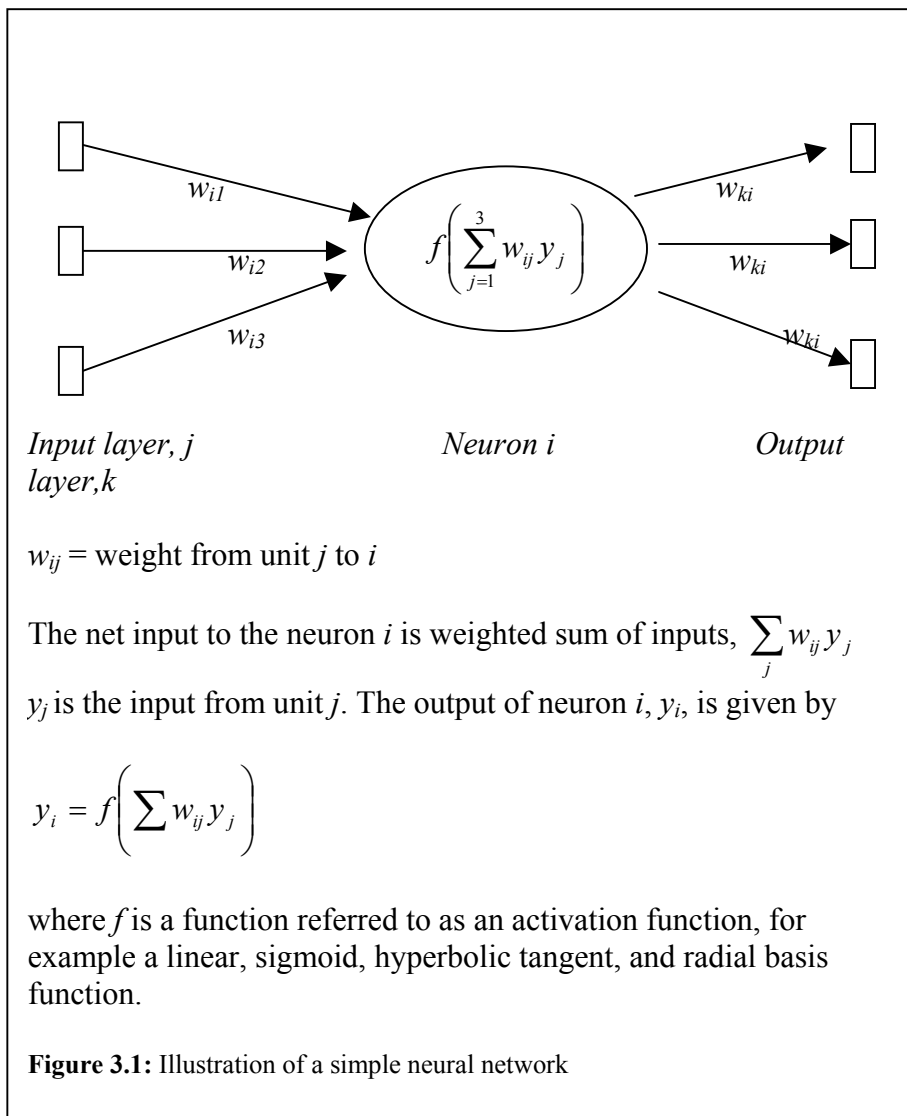
### 3.1 Introduction

Variations of flow characteristics in space and time are influenced by catchment characteristics such as climate, catchment morphometry, lithology, and land cover. Climate sets the broad limits to the transfer of water between the atmospheric system and the drainage basin. Other factors affect the transformation of net rainfall into various transfer processes and storages that occur within a drainage basin hydrological system. Flow characteristics are indicators of the status of some of these processes and storages. Catchment characteristics affect differently each of the flow characteristics. Consequently, prediction of each of the flow characteristics requires identification of influential factors. For some of the flow characteristics, multiple regression methods are capable of identifying the influential catchment characteristics. Prediction of such flow characteristics can be done using these methods. Multiple regression methods may be inappropriate for some flow characteristics, as these methods assume a) a linear relationship between flow and catchment characteristics, and b) that variables have distributions that approximate a normal distribution. Both assumptions are not always valid, and therefore other methods have to be used for identifying influential catchment characteristics, and the subsequent development of predictive techniques. One such method is the use of neural networks that are capable of modelling non-linear relationships, and do not assume a specific underlying distribution for the data (Ripley, 1994; Ardo, *et al.*, 1997; Hall and Minns, 1999; Orr, *et al.*, 1999). Neural networks have mostly been used in hydrology for flow forecasting (Chibanga *et al.*, 2003; Cigizzoghi, *et al.*, 2003; Dolling and Varas, 2003).

The aim of this chapter is to identify catchment characteristics that influence selected flow characteristics using univariate statistical methods and neural networks. The possibility of predicting these flow characteristics from catchment characteristics is explored using the same methods.

### 3.2 Neural networks

A neural network consists of computational units or nodes that are linked (Figure 3.1).



The output of a neuron can be an input to the next neuron. The most commonly used neural network architecture is one with three layers, comprising (i) an input layer, (ii) a hidden layer, and (iii) an output layer. A neural network is denoted for example by MLP5-7-6 meaning a multilayer perceptron with 5 units in the input layer, 7 units in the hidden layer, and 6 units in the output layer. A neural network without a hidden layer, and describing a linear input-output relationship is denoted by for example L5-2 meaning 5 units in the input layer, and 2 units in the output layer. The number of units in the input layer, and

those in the output layer depend on the problem being analysed. The number of units or neurons within the hidden layer are selected by trial and error so as to best describe the input-output relationship. A general recommendation is that the number of hidden units should be half the sum of the number of units in the input and output layers. It is possible to have more than one hidden layer. In most neural networks, each neuron in a preceding layer is connected to all the neurons in the next layer. Such a neural network is described as being fully connected. Neural networks in which information passes from one preceding layer to the next, and not backwards, are called feedforward networks, and being the most commonly used.

Neural networks can be used for modelling input – output relationships, supervised and unsupervised classification (Ripley, 1994; Orr, *et al.*, 1999; Bals, 2002). The process of adjusting weights of the network to minimise differences between the outputs predicted by the network, and the actual outputs is called network training. A back propagation algorithm is commonly used for training a neural network (Ripley, 1994; Ardo, *et al.*, 1997; Skidmore, *et al.*, 1997; Orr, *et al.*, 1999). This algorithm is used in this study. Two approaches can be used to select explanatory variables. First, all relevant variables are included, and through trial and error some variables are removed and the effect of this on the prediction error is assessed. This is done until only those variables that are necessary for describing relationships between variables are retained in a network. Second, a small number of variables are initially used, and other variables are added until the inclusion of additional variables has no effect on the prediction accuracy. The variables used in this study for neural network analysis are given in Table 3.1 below. Different configurations of neural networks for predicting each of the flow characteristics are assessed.

**Table 3.1:** Variables for neural network analysis

Catchment (Inputs)	Descriptors	Flow Characteristics (Outputs)
$\bar{P}_{yr}$ , $S_{10}$ , $S_{25}$ , $S_{34}$ , $S_{50}$ , $S_{68}$ , $S_{75}$ , $S_{80}$ , $S_{90}$ , $Dd$ , $\bar{E}_{pot, yr}$ , $GL_{GG}$ , $GL_{GR}$ , $GL_{KL}$ , $GL_{LM}$ , $LC_{BU}$ , $LC_{WD}$ , $LC_{WG}$ , $LC_{GR}$ , and $LC_{CU}$ .		$\bar{Q}_{yr}$ , $BFI$ , $q_{90}$ , $q_{80}$ , $q_{70}$ , $q_{60}$ , $q_{50}$ , $q_{40}$ , $q_{30}$ , $q_{20}$ , $q_{10}$ .

When training a neural network there is a need to prevent overtraining of the network, and this is done by training a network on a separate data set (training sub-sample). Training continues as long as the root mean square error (RMSE) decreases. But when this error no longer decreases on a second sub-sample

(selection sub-sample), this indicates that the network is being overtraining. A third sub-sample is used to validate the network (testing sub-sample). There are 52 catchments in this study, and these are split by random sampling into 26 catchments for training, 13 for selection, and 13 for testing (StatSoft Inc., 2001).

The software used for neural network analysis in this study is STATISTICA Version 6 by StatSoft Inc (2001). This allows for testing of as many different configurations of networks (e. g., 500 networks) as possible, and the best network in terms of minimising the RMSE is selected.

### **3.3 Results**

#### **3.3.1 Correlation between flow and catchment characteristics**

Table 3.2 shows the correlation between flow characteristics and catchment descriptors. Only correlation coefficients significant at the 5% significance level are presented.

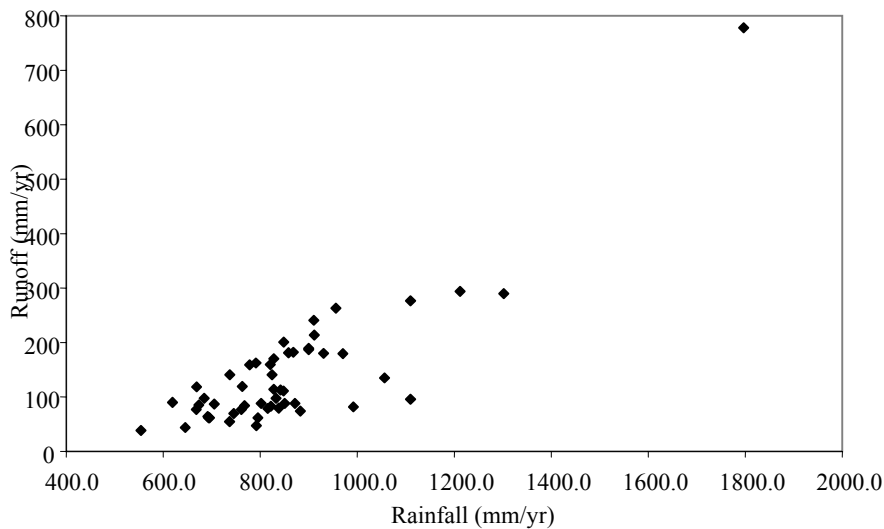
**Table 3.2:** Correlation Between of Flow Characteristics and Catchment Descriptors

	$\bar{Q}_{yr}$	$CV$	$BFI$	$q_{90}$	$q_{70}$	$q_{50}$	$q_{20}$	$\bar{N}_{DZ}$
$\bar{P}_{yr}$	0.85		0.71	0.87	0.83	0.75	0.56	-0.52
$NDVI$	0.64	-0.41	0.71	0.63	0.74	0.72	0.61	-0.44
$Dd$	0.29			0.31	0.39	0.37		
$GL_{GG}$	0.33							
$GL_{LM}$			0.30			0.28		
$GL_{BA}$							-0.28	
$GL_{SA}$			-0.30				-0.33	
$GL_{KL}$			-0.30					
$GL_{AL}$			-0.32				-0.31	
$S_{10}$	0.49	-0.43	0.75	0.50	0.70	0.78	0.70	-0.48
$S_{25}$	0.68	-0.44	0.80	0.72	0.84	0.84	0.68	-0.56
$S_{34}$	0.72	-0.43	0.77	0.71	0.80	0.81	0.65	-0.55
$S_{50}$	0.65	-0.45	0.76	0.60	0.73	0.77	0.69	-0.53
$S_{68}$	0.59	-0.40	0.76	0.60	0.74	0.76	0.65	-0.56
$S_{75}$	0.58	-0.42	0.73	0.54	0.67	0.71	0.66	-0.52
$S_{80}$	0.44	-0.36	0.72	0.49	0.67	0.71	0.64	-0.56
$S_{90}$	0.46	-0.37	0.73	0.51	0.69	0.72	0.64	-0.56
$DB$			-0.39				-0.43	0.38
$LC_{WD}$	-0.27			-0.33				
$LC_{WG}$			-0.45		-0.36	-0.42	-0.51	0.29
$LC_{GR}$			-0.37			-0.29	-0.42	0.49
$LC_{CU}$				-0.43	-0.38	-0.29		
$LC_{CG}$	-0.31		-0.53		-0.38	-0.45	-0.60	0.53

Notes:  $DB$  is the proportion of the catchment covered by dambos. Dambo figures are available for only 37 catchments (Bullock,1988).

### 3.3.2 Mean annual runoff

Table 3.2 above shows a strong correlation between  $\bar{Q}_{yr}$  and  $\bar{P}_{yr}$ , ( $r = 0.85$ ) caused by the correspondence of rainfall with runoff increases (Figure 3.2). The mean annual runoff has also a moderately strong relationship with  $NDVI$ . Both  $NDVI$  and  $\bar{Q}_{yr}$  are positively related with  $\bar{P}_{yr}$ , and therefore the correlation between  $\bar{Q}_{yr}$  and  $NDVI$  is just showing that both variables are dependent on one variable.



**Figure 3.2:** Mean annual runoff against mean annual rainfall

A weak positive relationship exists between  $\bar{Q}_{yr}$  and  $GL_{GG}$  ( $r = 0.33$ ). Areas under gneiss and granite ( $GL_{GG}$ ) tend to have rock outcrops, such as dwalas which promote the formation of runoff as they are impervious.  $\bar{Q}_{yr}$  also increases with slope, for example  $S_{50}$  ( $r = 0.65$ ). Steep slopes promote the formation of runoff. Areas with steep slopes generally have high altitudes and high rainfall, e.g. Eastern Highlands. The positive correlation between  $\bar{Q}_{yr}$  and slope also reflects this phenomenon. A weak negative relationship seems to exist between  $\bar{Q}_{yr}$  and  $LC_{CG}$  ( $r = -0.31$ ). Areas with wooded grasslands and grasslands tend to have dambos, and the correlation between  $LC_{CG}$  and  $DB$  is  $r^2 = 0.72$ . The negative correlation between  $\bar{Q}_{yr}$  and  $LC_{CG}$  is likely to reflect increased evaporation losses within dambos (McCartney, 1998; Wolski, 1999).

The following linear equations for predicted  $\bar{Q}_{yr}$  using catchment descriptors were developed using a step-wise multiple regression technique (Box 3.1).

**Box 3.1 Predictive equations for  $\bar{Q}_{yr}$  (mm yr-1) from catchment characteristics using multiple regression methods**

$$\bar{Q}_{yr} = 0.393\bar{P}_{yr} - 197.4 \quad r^2=0.75 \quad (3.1)$$

$$\bar{Q}_{yr} = 0.376\bar{P}_{yr} + 0.755GL_{GG} - 232.5 \quad r^2=0.78 \quad (3.2)$$

$$\bar{Q}_{yr} = 0.293\bar{P}_{yr} + 0.803GL_{GG} + 9.574S_{34} - 200.0 \quad r^2=0.81 \quad (3.3)$$

$$\bar{Q}_{yr} = 0.247\bar{P}_{yr} + 0.644GL_{GG} + 30.160S_{34} - 10.3000S_{80} - 143.9 \quad r^2=0.86 \quad (3.4)$$

$$\bar{Q}_{yr} = 0.279\bar{P}_{yr} + 0.600GL_{GG} + 1.411S_{34} + 30.637S_{75} - 27.4 \quad r^2=0.89 \quad (3.5)$$

$$\bar{Q}_{yr} = 0.282\bar{P}_{yr} + 0.6001GL_{GG} + 31.693S_{75} - 27.870S_{80} - 27.9 \quad r^2=0.89 \quad (3.6)$$

Bullock (1988) derived the following equation using 108 catchments located on the central part of Zimbabwe

$$\bar{Q}_{yr} = 0.5\bar{P}_{yr} - 258 \quad (3.7)$$

with  $r^2 = 0.42$ . Predictive equations presented in Box 3.1 above have higher levels of explained variance of  $\bar{Q}_{yr}$  than Eqn (3.7). Eqn (3.6) is recommended for estimating  $\bar{Q}_{yr}$  from catchment descriptors for ungauged catchments within the study area. If slope data are not available, then Eqn (3.1) can be used. The intercept in Eqn (3.1) reflects the effects of interception on runoff production (Savenije, 1997). Equation (3.1) can be generalized to

$$\bar{Q}_{yr} = \gamma(\bar{P}_{yr} - \bar{I}_{yr}) \quad (3.8)$$

where  $\bar{I}_{yr}$  = average annual interception (mm yr<sup>-1</sup>)

$\gamma$  = effective runoff coefficient (Savenije, 1997) .

$(\bar{P}_{yr} - \bar{I}_{yr})$  is the effective rainfall that reaches the surface and then partitioned into surface runoff, transpiration, and subsurface storage. Savenije (1997) derived the effective runoff coefficient from analysis of monthly flows and this gives an estimate of an annual effective runoff coefficient. When the annual average number of rainy days per year,  $\bar{N}_{yr}$ , is included as an explanatory variable for  $\bar{Q}_{yr}$ , the step-wise multiple regression method indicates that this variable does not improve the prediction of  $\bar{Q}_{yr}$ .  $\bar{N}_{yr}$  is highly correlated with  $\bar{P}_{yr}$  ( $r = 0.83$ ), and therefore the variability of  $\bar{Q}_{yr}$  that  $\bar{N}_{yr}$  may explain is already accounted for by  $\bar{P}_{yr}$  which is included in Eqn (3.1).

An assessment of whether neural networks are better than Eqn (3.1) to (3.6) which assume that  $\bar{Q}_{yr}$  is linearly related to catchment characteristics was undertaken. Several neural networks were calibrated with the restriction that the possible number of hidden layers does not exceed one. Table 3.3 gives neural networks which were considered best at predicting  $\bar{Q}_{yr}$  on the basis of the coefficient of determination. The coefficient of determination was estimated by comparing predicted  $\bar{Q}_{yr}$  and that estimated from observed flows.

The explanatory variables are given in Table 3.3 for each network in their order of importance. A linear neural network performs equally well as the multi-layer perceptron with 2 units in the hidden layer. This analysis shows again that  $\bar{P}_{yr}$ , is the most important explanatory variable for  $\bar{Q}_{yr}$ . The slope of a catchment also influences  $\bar{Q}_{yr}$ .  $GL_{LM}$ ,  $\bar{E}_{pot, yr}$ , and  $LC_{CU}$  have also been identified as influencing  $\bar{Q}_{yr}$  but they were not identified during multiple regression analysis.

**Table 3.3:** Neural networks for predicting  $\bar{Q}_{yr}$  from catchment characteristics

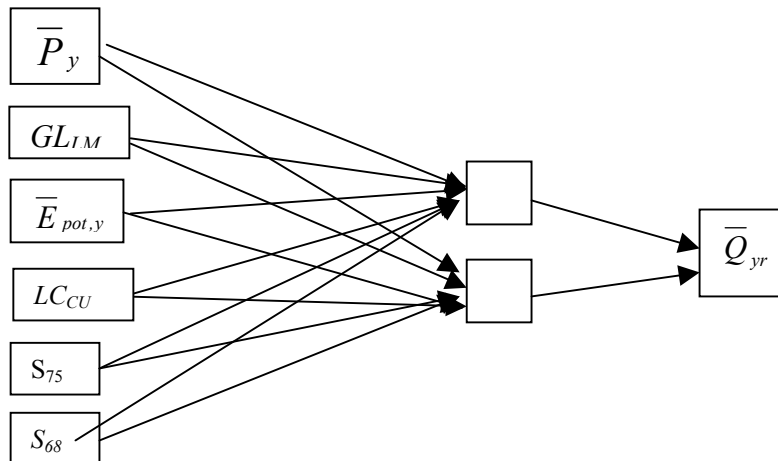
Type of Network	No. of Units in the Hidden Layer	Explanatory Variables	$r^2$
L 3-1	None	$\bar{P}_{yr}, \bar{E}_{pot,yr}, GL_{LM}$	0.76
MLP 5-3-1	3	$\bar{P}_{yr}, GL_{LM}, \bar{E}_{pot,yr}, S_{75}$	0.72
MLP 6-2-1	2	$\bar{P}_{yr}, GL_{LM}, \bar{E}_{pot,yr}, LC_{CU}, S_{75}, S_{68}$	0.76

*Notes:*

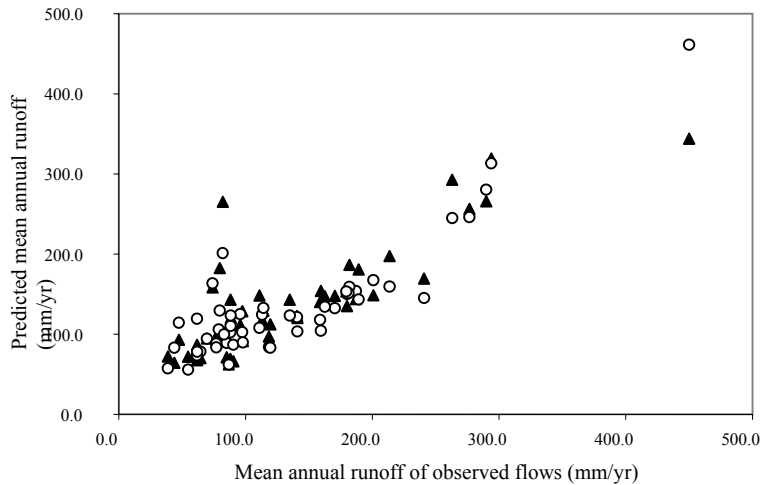
- L3-1: linear neural network with 3 input unit, no hidden layer, and 1 output unit.
- MLP5-3-1: multi-layer perceptron with 5 input units, 3 units in the first and only hidden layer, and 1 output unit.
- MLP6-2-1: multi-layer perceptron with 6 input units, 2 units in the first and only hidden layer, and 1 output unit.

An illustration of the structure of one of the above neural networks is provided in Figure 3.3.

Neural networks can model non-linear relationships and the inclusion of  $GL_{LM}$ ,  $\bar{E}_{pot,yr}$ , and  $LC_{CU}$  indicates that these are non-linearly related to  $\bar{Q}_{yr}$ . The coefficients of determination given in Table 3.3 are smaller than those of Eqn (3.4) to (3.6) presented in Box 3.1. A comparison of predictions of  $\bar{Q}_{yr}$  produced by Eqn (3.6) and those of an MLP6-2-1 does not show major differences (Figure 3.4).



**Figure 3.3:** The MLP 6-2-1 neural network, with 6 units in the input layer, 2 units in the hidden layer, and one unit in the output layer.

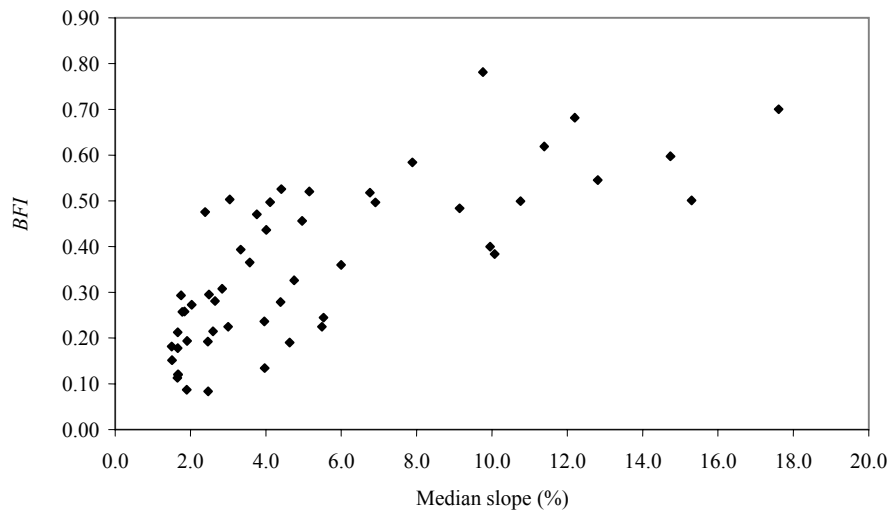


**Figure 3.4:** Comparison of the performance of multiple regression and an MLP6-2-1 neural network in predicting mean annual runoff.  
 ▲ = MLP6-2-1, ○ = multiple regression equation.

An MLP6-2-1 underestimates  $\bar{Q}_{yr}$  for one catchment with very high  $\bar{Q}_{yr}$  value. Prediction of  $\bar{Q}_{yr}$  for ungauged catchment using multiple regression is recommended as this method is simpler than the use of neural networks.

### 3.3.3 Base flow index

*BFI* has a strong relationship with  $\bar{P}_{yr}$ , *NDVI*, and slope. If all other factors are constant, areas with high rainfall will have larger amounts of water stored in subsurface storage than areas with low rainfall. A large subsurface storage will increase delayed interflow and groundwater flows, and therefore has high *BFI*. Both *BFI* and *NDVI* depend upon  $\bar{P}_{yr}$ , and hence the relationship between *BFI* and *NDVI* reflects this dependence on a common variable. Slopes within a catchment have a considerable effect on base flows (Figure 3.5).



**Figure 3.5:** Influence of slope on base flow index

Subsurface water can only contribute to runoff if a hydraulic gradient exists. The slope of the water table usually conforms to the slope of the land above (Freeze and Cherry, 1979). According to Darcy's Law catchments with steep slopes will have high hydraulic gradients, resulting in high base flow rates. Besides the effect of slope on the hydraulic gradient, steep slopes occur in mountainous and hilly parts of catchments. Some of these hills and mountains have substantial fractures and fissures that store water during the rainy season, and then release it as base flow during the dry season. This was observed on some catchments in South Africa (Hughes, personal communication). Thus, steep slopes may also indirectly reflect areas with substantial subsurface storage due to these fractures and fissures.

*BFI* has a weak and negative relationship with  $GL_{KL}$ ,  $GL_{AL}$ , and  $GL_{SA}$ . Water tables within Kalahari sands are usually at depths greater than 20 m, and therefore river beds do not intersect aquifers in these formations (Interconsult A/S, 1985). This reduces the potential for groundwater to contribute to stream flows. A negative relationship exists between *BFI* and  $LC_{CG}$ . Most catchments with grasslands have dambos. It is likely that the negative relationship between  $LC_{CG}$  and *BFI* reflects that catchments with dambos have high evaporation losses resulting in reduced base flows. Studies by Faulkner and Lambert (1991), McCartney (1998), and Wolski (1999) have shown that evapotranspiration rates are higher within dambos than on the surrounding interfluvies. Dambos have in comparison to interfluvies denser vegetation, shallower water tables, and heavier

textured soils which have high capillary rises, and all these contribute towards higher evapotranspiration rates than on interfluves.

Box 3.2 gives the equations for predicting *BFI* derived using the step-wise multiple regression technique.

**Box 3.2 Predictive equations for *BFI* from catchment characteristics using multiple regression methods**

$$\hat{BFI} = 0.162 + 0.064S_{25} \quad r^2 = 0.63 \quad (3.9)$$

$$\hat{BFI} = 0.213 + 0.056S_{25} - 0.002LC_{CG} \quad r^2 = 0.66 \quad (3.10)$$

$$\hat{BFI} = 0.237 + 0.056S_{25} - 0.003GL_{KL} - 0.002LC_{CG} \quad r^2 = 0.69 \quad (3.11)$$

$$\hat{BFI} = 0.072S_{10} + 0.0003\bar{P}_{yr} - 0.003GL_{KL} - 0.001LC_{CG} \quad r^2 = 0.75 \quad (3.12)$$

In the study by Bullock (1988) the following Eqn (3.13) for predicting *BFI* was derived for catchments on the central part of Zimbabwe

$$\hat{BFI} = 0.0407BFISOIL^{0.706} \bar{P}_{yr}^{0.800} (1 + FALAKE)^{19.9} A^{0.170} S1085^{0.346} \quad (3.13)$$

with  $r^2 = 0.48$ . *BFISOIL* was defined as a function of the soil types and is given by

$$\hat{BFISOIL} = -0.04Dambo + 0.6SOIL1 + 0.6SOIL2 + 0.45SOIL3 + 0.3SOIL4 + 0.6SOIL5 \quad (3.14)$$

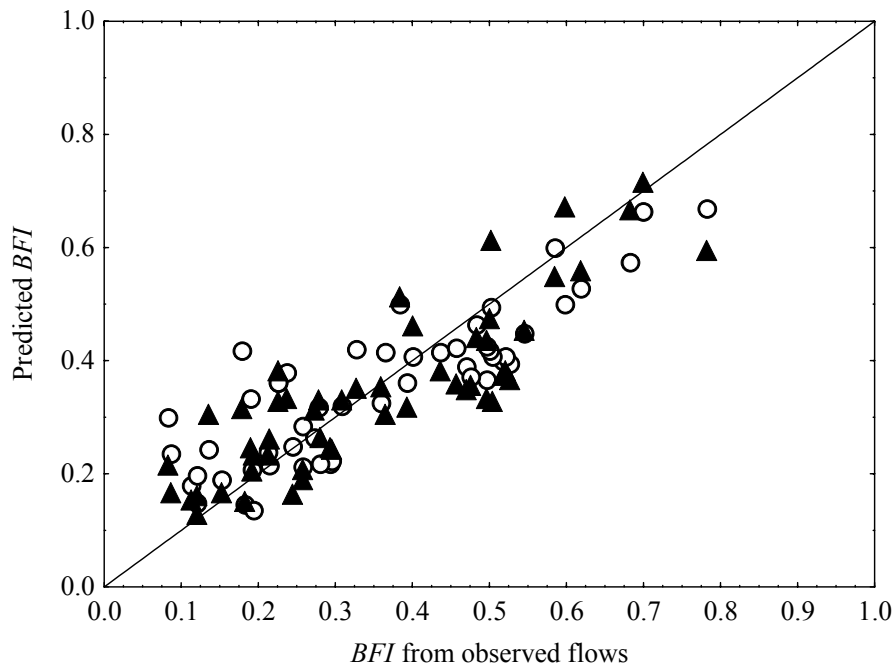
*SOIL1...SOIL5* are the proportions of the catchment with soil types 1 to 5 distinguished in that area. *FALAKE* is the proportion of the catchment covered by small dams. The predictive equation for *BFI* derived by Bullock (1988) is non-linear and has a lower coefficient of determination than Eqn (3.9) to (3.12). *BFI* for catchments in western Europe was found to be related to soil types, but for most countries the coefficient of determination of the predictive equations developed was low ( $< 0.50$ ) (Gustard *et al.* 1989).

The potential for predicting *BFI* using neural networks was investigated, and Table 3.4 gives neural networks for predicting *BFI*.

**Table 3.4:** Neural networks for predicting *BFI* from catchment characteristics.

Type of Network	No. of Units in the Hidden Layer	Explanatory Variables	$r^2$
L 5-1	None	$S_{50}, Dd, \bar{P}_{yr}, \bar{E}_{pot, yr}, GL_{KL}$ .	0.71
MLP 4-3-1	3	$\bar{E}_{pot, yr}, LC_{CG}, GL_{KL}, \bar{P}_{yr}$ .	0.72
MLP 2-2-1	2	$\bar{P}_{yr}, S_{10}$	0.71
MLP 4-5-1	9	$S_{50}, \bar{P}_{yr}, \bar{E}_{pot, yr}, GL_{KL}$	0.75

Neural networks identified catchment characteristics that are almost the same as those for multiple regression. The only additional catchment characteristics are  $\bar{E}_{pot, yr}$  and  $GL_{GG}$ . An MLP 4-5-1 is marginally best at predicting *BFI*. When the coefficients of determination of predictive equations for *BFI*, Eqns (3.9) to (3.12), derived using multiple regression are compared with those of neural networks, these show that neural networks are marginally better at predicting *BFI* than multiple regression equations.



**Figure 3.6:** Prediction of *BFI* by an MLP 4-5-1 neural network and multiple regression. Circle = neural network, shaded triangle = multiple regression.

However, Figure 3.6 shows no significant differences between predictions made by Eqn (3.12) and an MLP 4-5-1. Both methods are recommended for predicting *BFI* for ungauged as there is no difference in the root mean square error.

### 3.3.4 Flow duration curves

An examination of shapes of each of the flow durations curves of catchments considered in this study, suggested that the following functions are likely to be appropriate for modelling the relationship between dimensionless daily flows,  $q_p$ , and their exceedance probabilities,  $p$ :

$$q_p = b_0 + \frac{b_1}{p} \quad (3.15)$$

$$q_p = b_0 - b_1 + b_2 p^2 \quad (3.16)$$

$$q_p = b_0 - b_1 + b_2 p^2 - b_3 p^3 \quad (3.17)$$

$$q_p = b_0 \exp(-b_1 p) \quad (3.18)$$

where  $b_0, \dots, b_3$  are coefficients, and  $\exp$  is the exponential function

For each of the catchments, dimensionless daily flows were regressed against their exceedance probabilities. Eqn (3.15) and (3.18) best describe the relationship between  $q_p$  and  $p$ . Mimikou and Kaemake (1985) also found that the exponential model described flow duration curves of rivers in Greece.  $b_0$  in Eqn (3.15) can be related to *BFI*, and  $b_1$  is not related to any catchment characteristic for catchments considered in this study. Both coefficients in Eqn (3.18) can be related to *BFI*, which enables derivation of flow duration curves for ungauged catchments, if *BFI* is known. For all catchments, the exponential model is capable of explaining over 99% of the variation of  $q_p$  for given values of  $p$ .

The following equations were derived

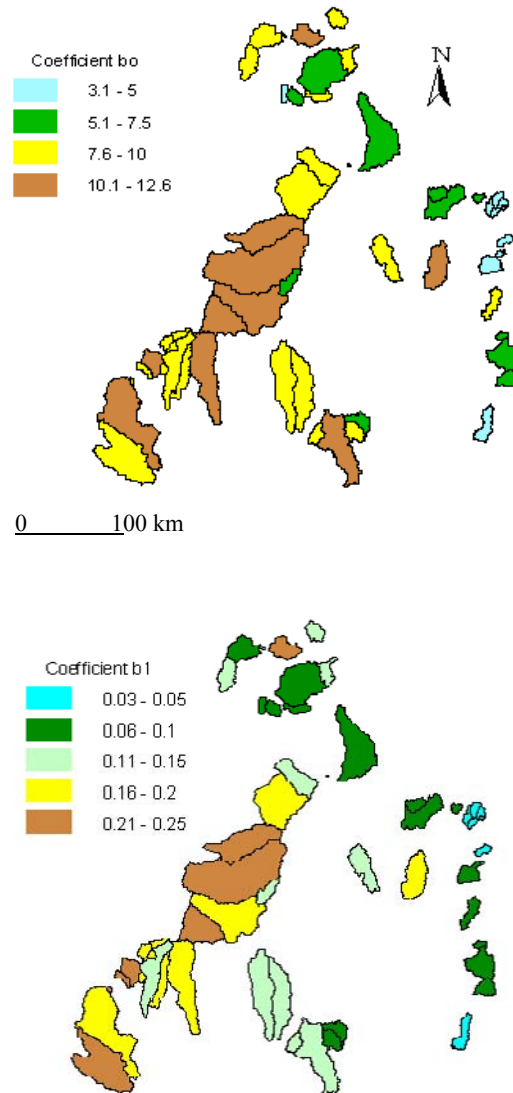
$$\hat{b}_0 = 11.852 - 11.519 BFI \quad r^2 = 0.74 \quad (3.19)$$

$$\hat{b}_1 = 0.311 \exp(-2.878 BFI) \quad r^2 = 0.92 \quad (3.20)$$

The exponential model for the flow duration curve then becomes

$$\hat{q}_p = (11.852 - 11.519BFI) \exp\{-[0.311 \exp(-2.878BFI)]p\} \quad (3.21)$$

Figure 3.7 shows the spatial variation of coefficients  $b_0$  and  $b_1$ .



**Figure 3.7:** Spatial patterns of coefficients of the exponential model of flow duration curves

Coefficient  $b_0$  tends to be large (7.0 -12.0) for catchments on the central part, and small (3.0 - 6.0) for Eastern Highlands catchments. Flow duration curves were constructed using dimensionless flows which were derived by dividing daily flows by mean daily flows. Coefficient  $b_0$  gives the magnitude of flow when the exceedance probability ( $p$ ) approaches zero. Catchments in dry areas will have this flow being many times the mean daily flow, while for the Eastern Highlands this flow will not be many times the mean daily flow. Hence  $b_0$  is relatively large for catchments on the central part, while this is small for Eastern Highlands catchments. Catchments on the central part of the country have rapid flow depletion and therefore steep flow duration curves. Consequently  $b_1$  which determines the slope of a flow duration curve is relatively large (0.120 - 0.200) on these catchments. The Eastern Highlands catchments have a gradual flow depletion shown by the flattish flow duration curves, and hence  $b_1$  is relatively small (0.032 - 0.095) for these catchments.

The possibility of predicting flow duration curves from catchment characteristics using neural networks was explored. For this purpose the inputs for the neural network comprise catchment characteristics shown in Table 3.1. As the outputs comprise nine percentile flows,  $q_{90}, q_{80}, q_{70}, q_{60}, q_{50}, q_{40}, q_{30}, q_{20}, q_{10}$ , neural networks also have 9 output units. The best prediction of flow duration curves was made by a multi-layer perceptron with 5 input units and 17 units in the single hidden layer. (MLP 5-17-9). Catchment characteristics for predicting flow duration curves in their order of importance are  $S_{25}, S_{75}, \bar{E}_{pot, yr}, \bar{P}_{yr}$ , and  $Dd$ . The coefficients of determination for the various percentile flows ranged from 0.64 to 0.92 as shown below.

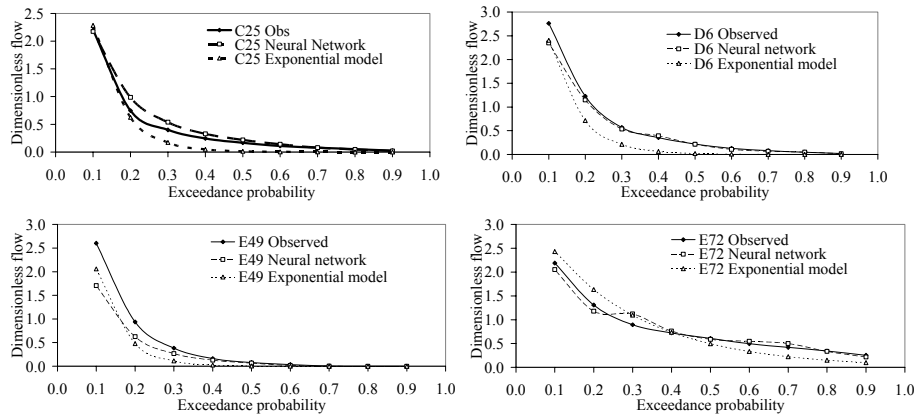
**Table 3.5:** Coefficients of determination for percentile flows predicted using a neural network.

Percentile Flow	$r^2$	Percentile Flow	$r^2$
$q_{90}$	0.92	$q_{40}$	0.83
$q_{80}$	0.88	$q_{30}$	0.77
$q_{70}$	0.85	$q_{20}$	0.72
$q_{60}$	0.86	$q_{10}$	0.64
$q_{50}$	0.83		

The coefficients of determination in Table 3.5 show that a neural network has a higher ability to predict low flows,  $q_{90}, q_{80}, q_{70}, q_{60}$ , than flood flows,  $q_{10}$ . Flood flows mainly depend upon characteristics of specific rainfall events rather than catchment characteristics used in this study.

Figure 3.8 compares flow duration curves derived using observed flows, those predicted using a) the MLP 5-17-9 neural network, and b) the exponential model with  $BFI$  estimated from catchment characteristics using Eqn (3.12). For

the catchments presented in Figure 3.8, the neural network has a smaller root mean square error than the exponential model on C25, D6, and E72. This shows that a neural network gives better predictions of flow duration curves than the exponential model. Figure 3.8 also shows that both the neural network and exponential model have problems predicting flows with exceedance probabilities less than 0.20 which are flood flows. Flood flows are mainly influenced by specific rainfall events.



**Figure 3.8:** Comparison of flow duration curves based on observed daily flows and those predicted using an exponential model, and a neural network for C25, D6, E49 and E72 respectively.

### 3.3.5 Average number of days with zero flows ( $\bar{N}_{DZ}$ ).

$\bar{N}_{DZ}$  has a negative exponential relationship with  $\bar{Q}_{yr}$  because rivers with high runoff tend to be perennial.  $\bar{N}_{DZ}$  also decreases with increasing groundwater flow contribution to river flows. This is appears from the negative correlation between  $BFI$  and  $\bar{N}_{DZ}$  ( $r = -0.72$ ).

$$\bar{N}_{DZ} = 391.7 \exp(-0.0171\bar{Q}_{yr}) \quad r^2 = 0.60 \quad (3.22)$$

The best prediction of  $\bar{N}_{DZ}$  was made using Eqn (3.22).

### 3.3.6 Mean monthly runoff distribution

The prediction of 12 mean monthly runoff values from catchment characteristics is rather problematic, as each of the monthly values requires its

own model. If simple linear regression models for predicting each of the monthly values are used, then 24 model parameters have to be estimated. With such a large number of model parameters the effects of estimation errors will be considerable. The possibility of predicting the mean monthly runoff distribution from flow and catchment characteristics using neural networks is therefore investigated. For each catchment the proportion of mean annual runoff occurring in each of the months,  $h_j$ , is given by

$$h_j = \frac{\bar{Q}_j}{\bar{Q}_{yr}} \quad (3.23)$$

where  $j = 1, 2, \dots, 12$ , and  $\bar{Q}_j$  = mean monthly runoff for month  $j$ . Thus

$$1.0 = \sum_{j=1}^{12} h_j \quad \text{and} \quad \bar{Q}_{yr} = \sum_{j=1}^{12} \bar{Q}_j$$

Neural network analysis is firstly undertaken with catchment characteristics only as inputs and the outputs are the 12 values of the hydrograph,  $h_j$ . For the second case, inputs to the neural network consists of flow characteristics and catchment characteristics.

For the first case an MLP 8-13-12 gave the best predictions, and the correlation coefficients between the predicted  $h_j$  and those derived from measured flows are given in Table 3.6 below.

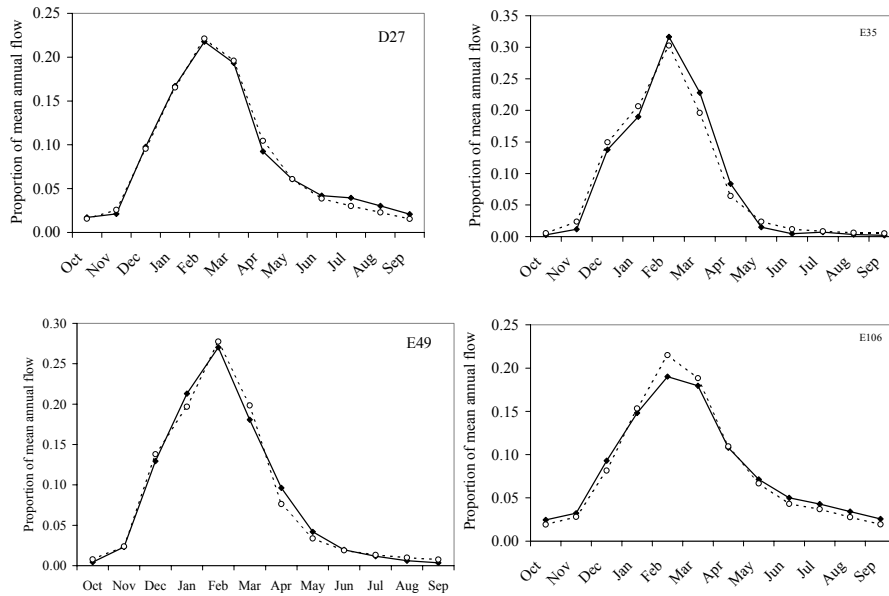
**Table 3.6:** Correlation coefficients between predicted and observed annual hydrograph values

Month	Catchment Characteristics MLP8-13-12	Flow and Catchment Characteristics MLP4-12-12
October	0.91	0.96
November	0.54	0.61
December	0.70	0.87
January	0.74	0.81
February	0.81	0.85
March	0.42	0.37
April	0.65	0.82
May	0.85	0.95
June	0.85	0.90
July	0.91	0.97
August	0.91	0.96
September	0.91	0.96

Predictive catchment characteristics for an MLP8-13-12 are  $GL_{GR}$ ,  $\bar{E}_{pot, yr}$ ,  $S_{10}$ ,  $S_{50}$ ,  $GL_{GG}$ ,  $S_{34}$ , and  $GL_{KL}$  in order of importance. No problems occur when predicting the recession limb of the hydrograph from May to October. Flows start to rise in November and the neural network cannot predict accurately the hydrograph value for this month. The onset of the rainy season is highly variable in both space and time and hence leads to poor prediction. The inflection point on the annual hydrograph occurs around March, and again is difficult to predict from catchment characteristics. Lithology, slope and potential evaporation rates affect runoff distribution within a year. These variables were earlier identified as affecting  $BFI$ . Catchments with relatively high  $BFI$  values will have a gradual depletion of subsurface storage, and hence a significant proportion of the annual flow will occur during the recession period (relatively large  $h_j$  values from May to October). Slope affects the rate of drainage of both surface and subsurface water, and therefore affects  $h_j$ . Catchments with relatively high values of  $\bar{E}_{pot, yr}$  are expected to have rapid depletion of subsurface storage, and hence low  $h_j$  values during the dry season May to October.

Figure 3.9 compares hydrographs predicted by this neural network, and those derived from observed flows on some of the catchments that were used for validation during neural network training. A close agreement occurs between these hydrographs. Therefore this neural network can be used to distribute  $\bar{Q}_{yr}$  into monthly flows, and hence predicting mean monthly flows.

For the second case an MLP 4-12-12 with  $BFI$ ,  $LC_{CG}$ ,  $S_{75}$  and  $S_{10}$  in that order of importance, gave the best predictions when both flow and catchment characteristics were inputs to the neural network. The correlation coefficients are higher than when catchment characteristics alone are used.  $BFI$  is the most important variable for predicting the annual hydrograph. All lithological effects on the distribution of runoff are reflected in the  $BFI$  and hence these are not included as inputs to the network.  $S_{75}$  and  $S_{10}$  are identified as explanatory variables. The inclusion of  $LC_{CG}$  is likely to reflect that catchments with this land cover type have dambos that enhance evaporation rates, and therefore reducing dry season,  $h_j$ .

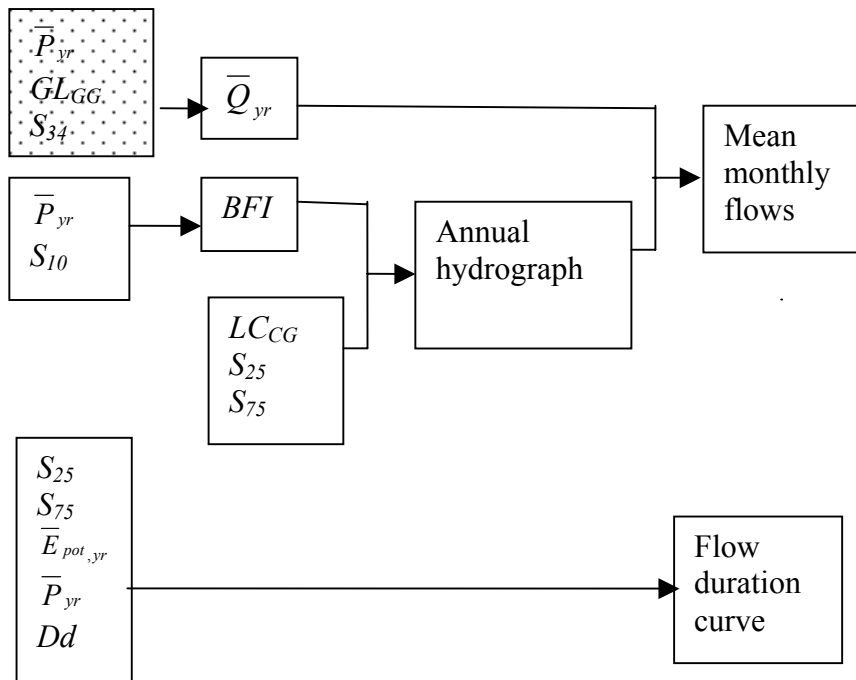


**Figure 3.9:** Comparison of hydrographs of the distribution of mean annual flow into monthly flows for D27, E35, E49 and E106. Solid line = estimated from observed flows, dashed line = predicted by a neural network.

### 3.4 Summary

This chapter has shown that  $\overline{P}_{yr}$ , the input to the catchment hydrological system, and slope are the main explanatory variables for flow characteristics.  $\overline{Q}_{yr}$  is best predicted using a linear equation as shown in Figure 3.10.

*BFI* can be predicted using a linear equation, but a neural network yields a more parsimonious model. The mean annual hydrograph can be predicted from *BFI*, slope, and *LC<sub>CG</sub>* using a neural network. However, ordinates at the onset of the rise of the hydrograph, and its inflection point after the crest are not predicted accurately. The predicted  $\overline{Q}_{yr}$  and mean annual hydrograph allow estimation of mean monthly flows.,



**Figure 3.10:** Prediction of flow characteristics. A linear equation is used for the shaded box, while neural networks are used for other boxes.

The flow duration curve can be predicted using a neural network with slope,  $\bar{E}_{pot, yr}$ ,  $\bar{P}_{yr}$  and  $Dd$  as inputs. This chapter has demonstrated that it is possible to predict some of the flow statistics that are required for water resources planning and management.



## 4 ORDINATION

### 4.1 Introduction

The landscape on any catchment is made up of several combinations of physiographic attributes. These combinations are usually variable among catchments, giving rise to different hydrological responses. The effect on flow characteristics of a change in a single catchment characteristic such as land use is likely to be identifiable (Bosch, 1979; Edwards and Blackie, 1981; Bosch and Hewlett, 1982; Mumeka, 1986). But effects of changes of several catchment characteristics are not easily identifiable. Different combinations of catchment characteristics give rise to different responses. Changes in some catchment characteristics may counteract the effects of other catchment characteristics. In addition effects of some of the catchment characteristics are only identifiable within certain ranges of catchment areas, and outside these ranges, their effects are masked by other catchment characteristics (Pitlick, 1994). Several flow characteristics such as mean annual runoff,  $\bar{Q}_{yr}$ , daily flows with 0.90, 0.75, and 0.50 exceedance probabilities ( $q_{90}$ ,  $q_{75}$ ,  $q_{50}$ ), base flow index,  $BFI$ , and average number of days per year with no flow,  $\bar{N}_{DZ}$ , have to be examined in order to identify the effects of different combinations of catchment characteristics.

Multivariate analysis methods such as ordination are best suited at determining effects of different combinations of catchment characteristics on several flow characteristics. These methods enable explanation of the variability of a set of flow characteristics by a set of catchment characteristics. The relative importance of explanatory catchment characteristics can also be determined by ordination techniques. The aim of this chapter is to identify among the set of catchment characteristics those which explain the variability of flow characteristics. If catchment characteristics that influence flow characteristics can be identified, then these can hopefully be used to cluster catchments into clusters with similar hydrological responses.

### 4.2 Methodology

Ordination techniques fall into two categories that are indirect and direct gradient analysis (Kent and Coker, 1992; Legendre and Legendre, 1998; Ter Braak and Smilauer, 1998) (Table 4.1).

**Table 4.1:** Methods for indirect and direct gradient analysis

Indirect Gradient Analysis	Direct Gradient Analysis
Principal component analysis (PCA)	Redundancy Analysis (RDA)
Correspondence analysis (CA)	Canonical correlation (CanCor)
Detrended correspondence analysis (DCA)	Canonical Correspondence analysis (CCA)
	Detrended canonical correspondence analysis (DCCA)

Indirect gradient analysis aims to explain the variability of variables such as flow characteristics by a relatively small number of components. Let  $y_{ik}$  be a flow characteristic  $k$ ,  $k = 1, 2, \dots, n_q$ , and  $i = 1, 2, \dots, n_c$  denote the number of catchments. The variation of flow characteristics  $y_{ik}$  is explained by an unknown explanatory variable  $x_i$

$$y_{ik} = a_k + b_k x_i \quad (4.1)$$

where  $a_k$  and  $b_k$  are unknown regression coefficients. Since  $x_i$  is unknown, the regression coefficients are derived to explain the variation of  $y_{ik}$ . This is similar to PCA. The derivation of the components does not take into account catchment characteristics that explain the variation of flow characteristics. This is a major weakness of indirect gradient analysis methods. PCA has been applied in some studies to identify catchment characteristics for use in regionalisation (Seyhan and Keet, 1981).

A direct gradient analysis aims to identify the underlying structure in a data set by considering the relationships between response variables (flow characteristics) and explanatory variables (catchment characteristics). In the case of RDA,  $x_i$  is a linear combination of explanatory variables,  $z_{ij}$ , where  $j = 1, 2, \dots, n_p$  is the number of explanatory variables (Ter Braak and Prentice, 1988; Ter Braak and Smilauer, 1998). For example, if  $p = 2$ ,  $x_i$  is given by

$$x_i = c_1 z_{i1} + c_2 z_{i2} \quad (4.2)$$

where  $c_1$  and  $c_2$  are the weights applied to the measured explanatory variables  $z_{ij}$  to derive  $x_i$  the theoretical explanatory variable. Substituting for  $x_i$  in Eqn (4.1) gives the RDA model

$$y_{ik} = a_k + b_k c_1 z_{i1} + b_k c_2 z_{i2} \quad (4.3)$$

The aim of RDA is to estimate  $a_k$  and  $b_k$  being parameters of the response variables, and  $c_1$  and  $c_2$  weights from response and explanatory variables. These

equations can be extended to more than two response and explanatory variables. RDA is also referred to as constrained ordination since the weights on environmental variables are constrained to fit the response variables.

CCA and RDA are both direct gradient analysis methods that combine correlation and multiple regression to identify explanatory variables for response variables. The selection of which technique to use depends on the assumed underlying relationship between flow and catchment characteristics. CCA assumes a unimodal relationship, in which increases in magnitudes of catchment characteristics are associated with increases in flow characteristics up to a certain level, beyond which flow characteristics will decrease as catchment characteristics increase. RDA assumes that flow characteristics increase linearly as the catchment characteristics increase. This model is attractive for hydrological applications since flows increase with increasing precipitation, slope and decreasing evaporation rates. RDA is selected for use in this study for the following reasons:

- The main aim of the study is to determine the effect of catchment characteristics on flow characteristics, and therefore it is a problem of direct gradient analysis.
- Univariate analysis shows that most variables in this study have some form of linear relationship between flow and catchment characteristics,
- RDA has no restriction on the number of variables in both data sets in relationship to the number of catchments or samples.

Monte Carlo permutation tests identify catchment characteristics that are significant in explaining variability in flow characteristics. Table 4.2 shows flow and catchment characteristics that are selected for RDA. Catchment characteristics given in Table 4.2 are selected on the basis that they are likely to explain the various hydrological responses reflected by flow characteristics given in the same table. RDA is undertaken with standardized variable since different measurement units are used for most variables.

**Table 4.2:** Flow and catchment characteristics selected for redundancy analysis

Flow Characteristics	Catchment Characteristics
i. Mean annual runoff ( $\bar{Q}_{yr}$ )	i. Mean annual precipitation ( $\bar{P}_{yr}$ )
ii. Coefficient of variation of annual runoff ( $CV$ )	ii. Mean annual potential evaporation ( $\bar{E}_{pot,yr}$ )
iii. Base flow index ( $BFI$ )	iii. Catchment area ( $A$ )
iv. $\bar{N}_{DZ}$	iv. Median slope ( $S_{50}$ )
v. $q_{90}$	v. Drainage density ( $Dd$ )
vi. $q_{70}$	vi. Proportions of each catchment under the various lithologies
vii. $q_{50}$	vii. Proportions of each catchment with various land cover types

Ter Braak and Smilauer (1998) have developed the CANOCO software package for ordination using techniques given in Table 4.1. Both RDA and CCA derive two sets of ordination axes. One set is a linear combination of flow characteristics, whereas the second set is a linear combination of catchment characteristics. The first ordination axis of catchment characteristics explains as much as possible the variability of flow characteristics within the first ordination axis of flow characteristics. The second ordination axis is again a linear combination of catchment characteristics, orthogonal to the first axis. It explains as much as possible the variance of the flow characteristics that is not contained in the first ordination axis. Similarly the third ordination axis is again a linear combination of catchments characteristics, orthogonal to the first and second ordination axes, and explaining as much as possible the unaccounted for variance of flow characteristics.

### 4.3 Results

#### 4.3.1 Relationships between ordination axes

Table 4.3 presents a summary of the results of RDA in terms of the proportion of the total variance of the flow characteristics explained by the ordination axes of the catchment characteristics.

**Table 4.3:** Proportion of variance of flow characteristics explained by the ordination axes of catchment characteristics.

	AXIS 1	AXIS 2	AXIS 3	AXIS 4
Eigenvalues	0.637	0.059	0.034	0.014
Flows-catchment correlations	0.916	0.763	0.776	0.552
Cumulative percentage of variance of flow characteristics	63.7	69.6	73.0	74.4
Cumulative percentage of variance of flow-catchment relation	84.8	92.7	97.1	99.1

The eigenvalue for each axis shows the amount of variance of the flow characteristics that is explained by the particular axis. The first axis explains 63.1% of the variance of the flow characteristics, while the second axis explains an additional 5.9% of the variance. The third axis accounts for 3.4% of the variance of flow characteristics, while the fourth axis explains 1.4%.

The flow-catchment correlation coefficients in Table 4.3 indicate the strength of the linear association between the derived ordination axes. Thus, the first flow characteristics axis has a correlation coefficient of 0.916 with the first catchment characteristics axis. A high correlation indicates that the derived axis have strong linear association with the catchment characteristics. This does not necessarily indicate that the catchment characteristics explain the variability of flow characteristics. It is possible that the derived axes are closely correlated but without significantly explaining the variability of the flow characteristics. The amount of variance of flow characteristics that is explained by the catchment axes is shown as “Cumulative percentage of variance of flow characteristics”. In RDA the amount of variance explained is equal to the eigenvalue for each axis, and therefore the cumulative percentage of variance is obtained as the sum of the relevant eigenvalues as is shown in Table 4.3. The four ordination axes explain 74.7% of the variance of flow characteristics.

Table 4.3 also shows the contribution of each catchment axis in accounting for the total variance of flow characteristics that is explained by all the catchment axes. The “Cumulative percentage variance of flow-environment relation” shows that 84.8% of this explained variance is accounted for by the first axis, and the first and second axes account for 92.5%.

### 4.3.2 Relationships between catchment characteristics and their ordination axis

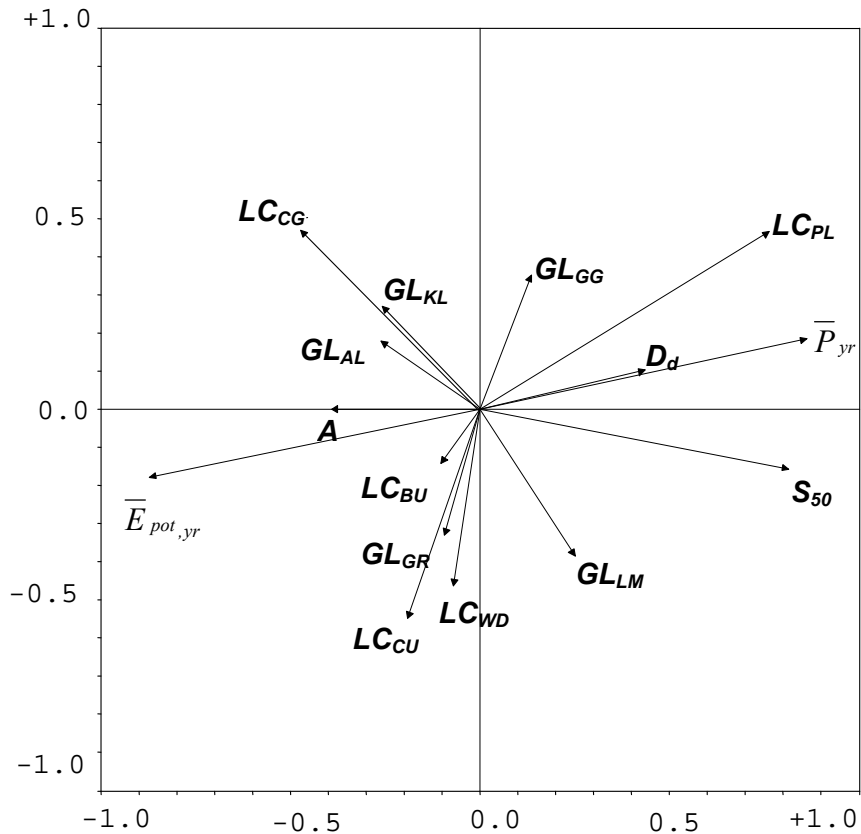
Table 4.4 presents the correlation coefficients between catchment characteristics and their ordination axes or intra-set correlations. The first ordination axis (Axis 1) has a strong positive correlation with  $\bar{P}_{yr}$ ,  $S_{50}$ , and  $LC_{PL}$ .  $\bar{E}_{pot,yr}$  has a negative relationship with this axis. These variables are the most important in defining this ordination axis as is also shown in Figure 4.1, in which the correlation coefficient between a catchment characteristic and the first ordination axis (horizontal axis), is plotted against correlation coefficient between the same catchment characteristics with the second ordination axis (vertical axis).

**Table 4.4:** Correlation between ordination axes of catchment characteristics with the catchment characteristics used to derive these axes

	Axis 1	Axis 2	Axis 3	Axis 4
$\bar{P}_{yr}$	0.861			
$\bar{E}_{pot,yr}$	-0.869			
$A$	-0.391			
$Dd$	0.436			
$GL_{GG}$		0.361		
$GL_{GR}$		-0.331		-0.460
$GL_{LM}$		-0.391	0.428	
$GL_{KA}$				
$GL_{AL}$	-0.280			
$S_{50}$	0.814			
$LC_{PL}$	0.763	0.464		
$LC_{WD}$		-0.468		
$LC_{BU}$				
$LC_{CU}$		-0.530	-0.451	
$LL_{CG}$	-0.450	0.490		

$LC_{PL}$  has a positive correlation with the first ordination axis ( $r = 0.76$ ), which is likely due to plantations occurring mostly on catchments within the Eastern Highlands, a region associated with high  $\bar{P}_{yr}$ , steep slopes, and relatively low  $\bar{E}_{pot,yr}$ . A weak positive correlation ( $r = 0.44$ ) exists between  $Dd$  and the first ordination axis. The weak negative correlation with catchment area,  $A$ , is likely due to areas of gauged catchments generally decreasing with altitude, while  $\bar{P}_{yr}$  and  $S_{50}$  are increasing. This is not a causal relationship. The first ordination axis is defined by climatological and morphometric properties of a catchment.

Land cover types ( $LL_{CG}$ ,  $LC_{CU}$ ,  $LL_{BU}$ ) and lithology ( $GL_{GG}$ ,  $GL_{GR}$ ,  $GL_{LM}$ ) are important in defining the second ordination axis. Similarly the third and fourth ordination axes depend on lithology and land cover types.

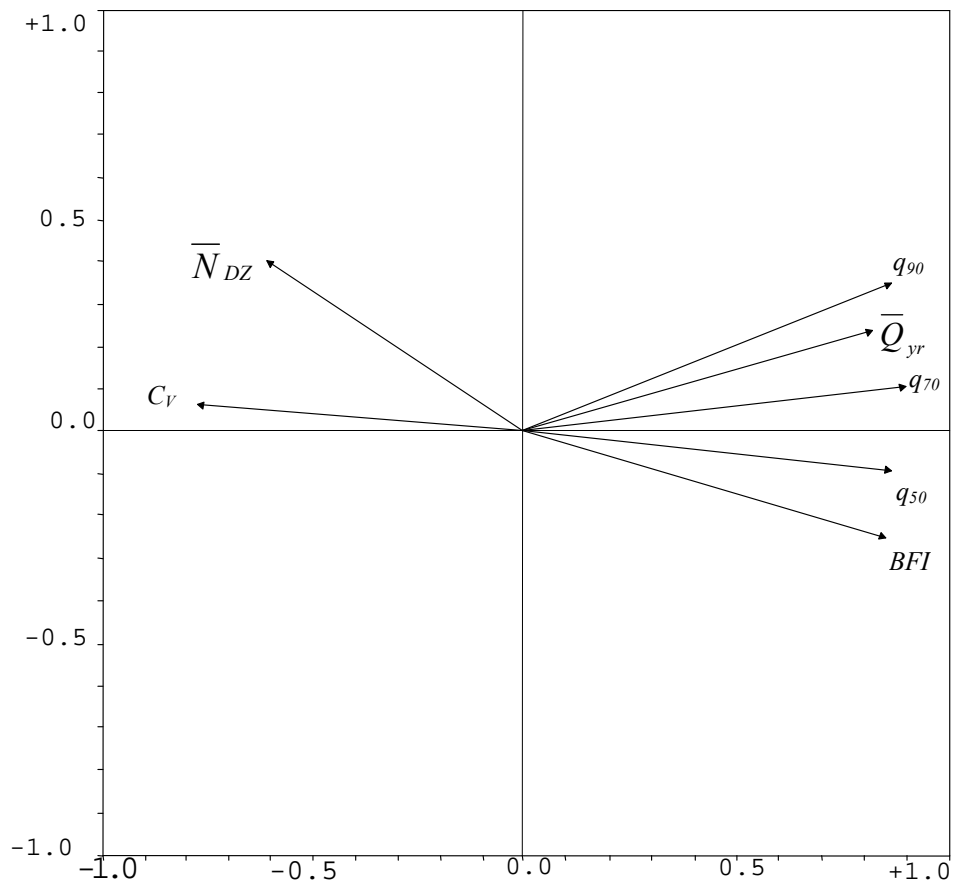


**Figure 4.1:** Relationship between catchment characteristics and their ordination axes.

### 4.3.3 Relationship between flow characteristics and their ordination axes

The correlation coefficient between each flow characteristic and the first ordination axis is plotted against the correlation coefficient between the same flow characteristic and the second ordination axis in Figure 4.2.  $\bar{Q}_{yr}$  has a positive and strong relationship with the first axis. Other flow characteristics such as  $q_{90}$ ,  $q_{70}$ ,  $q_{50}$ , and  $BFI$  have also positive relationships with the first ordination axis. This shows that catchments with large  $\bar{Q}_{yr}$  will also have large

values of these flow characteristics.  $\bar{N}_{DZ}$  and  $CV$  have negative relationships with the first axis (Figure 4.2).  $BFI$  and  $\bar{N}_{DZ}$  have also moderate correlation with the second axis.

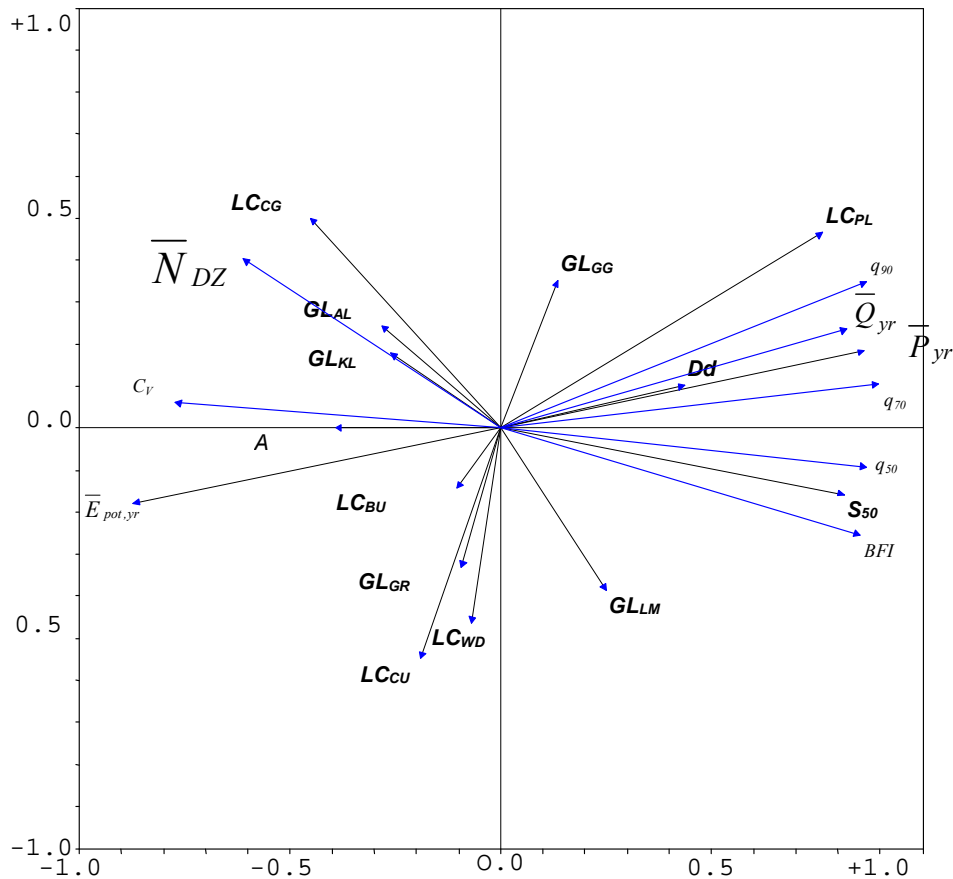


**Figure 4.2:** Relationship between flow characteristics with the first and second ordination axes

#### 4.3.4 Relationship between flow and catchment characteristics

The relationships between catchment characteristics and flow characteristics are illustrated in Figure 4.3. In addition, Table 4.5 shows correlation coefficients between catchment characteristics, and ordination axes of flow characteristics, or inter-set correlations.  $\bar{P}_{yr}$ ,  $\bar{E}_{pot,yr}$  and  $S_{50}$  have the highest correlation with the first ordination axis and are therefore the most important in explaining the variation of flow characteristics contained in this axis. The positive correlation

between flow characteristics, and areas under plantation is as result of plantations being mostly confined to the Eastern Highlands that have high rainfall and consequently high runoff. This is not a causal relationship.



**Figure 4.3:** Relationship between flow characteristics and catchment characteristics

$LC_{CG}$  has a moderate negative correlation with the first flow axis ( $r = -0.42$ ). The number of days with zero flows per year,  $\overline{N}_{DZ}$ , are close to  $LC_{CG}$  which mostly occurs in areas with relatively low rainfall, and therefore the association of  $LC_{CG}$  and  $\overline{N}_{DZ}$ . In addition,  $LC_{CG}$ , in Figure 4.3 has a negative association with  $BFI$  (Figure 4.3). Catchments with large proportions under grasslands usually have damboes, and therefore have higher evaporation rates than the surrounding interflues. (Bullock, 1988; McCartney, 1998; Wolski, 1997). Drainage density ( $D_d$ ) is also correlated with the first ordination axis of flow characteristics suggesting that  $D_d$  contributes in explaining flow characteristics.

**Table 4.5:** Correlation coefficients between catchment characteristics and ordination axes of flow characteristics

	Axis 1	Axis 2	Axis 3
$\bar{P}_{yr}$	0.808		
$\bar{E}_{pot, yr}$	-0.816		
$A$	-0.367		
$Dd$	0.409		
$GL_{GG}$		0.282	
$GL_{LM}$		-0.306	0.344
$S_{50}$	0.764		
$LC_{PL}$	0.716	0.363	
$LC_{WD}$		-0.366	
$LC_{CU}$		-0.415	-0.363
$LL_{CG}$	-0.422	0.383	

The proportions of a catchment under a) cultivation, ( $LC_{CU}$ ), b) woodlands ( $LC_{WD}$ ), underlain by c) Umkondo assemblage, ( $GL_{LM}$ ), and d) granites and gneiss ( $GL_{GG}$ ) are correlated to the second ordination axis. These catchment characteristics explain the variance of flow characteristics that is not accounted for by the first ordination axis. There is no significant correlation between catchment characteristics and the fourth ordination axis of flow characteristics.

The Monte Carlo permutation test was used to determine those catchment characteristics that are significant at the 5% level in explaining the variance of the flow characteristics. Table 4.6 below shows the catchment characteristics that were found to be significant, and the proportion of the variance of flow characteristics that each variable explains.

$\bar{P}_{yr}$ ,  $S_{50}$ ,  $LC_{PL}$ , and  $LC_{CU}$  are the only variables that significantly explain the variance of flow characteristics.

**Table 4.6:** Proportion of variance flow characteristics explained by catchment characteristics

Catchment Characteristic	Percentage Explained	Cumulative Percentage
$\bar{P}_{yr}$	50	50
$S_{50}$	7	58
$LC_{PL}$	5	62
$LC_{CU}$	3	65

The following variables improved the fit between ordination axes, but were not significant at the 5% level;  $GL_{GG}$ ,  $GL_{KL}$ ,  $GL_{AL}$ ,  $LC_{WD}$ ,  $Dd$ ,  $A$ ,  $LL_{CG}$  and  $LL_{BU}$ . The inclusion of these variables increased the cumulative percentage of explained variance to 75 %.

#### 4.4 Discussion and Conclusion

Direct gradient analysis has shown the variation of flow characteristics is explained by  $\bar{P}_{yr}$ ,  $S_{50}$ ,  $LC_{PL}$ , and  $LC_{CU}$ . The study area is located in a region where annual potential evaporation exceeds annual rainfall. Thus runoff is greatly dependent on the availability of excess rainfall which is determined by  $\bar{P}_{yr}$  and  $\bar{E}_{pot,yr}$ . The dominance of climatological characteristics over topographical characteristics in explaining runoff characteristics has also been observed in other studies (Meijerink, 1985; Pitlick, 1994).  $S_{50}$  greatly determines the kinetic energy available for water droplets to move towards stream channels, and along these channels. Slope can also be a surrogate for soil thickness, since thin soils occur on areas with steep slopes (Pitlick, 1994). Fast overland flow occurs on thin soils.

The proportion of the catchment under cultivation ( $LC_{CU}$ ) has been found to be significant in explaining part of the variability of flow characteristics.  $LC_{CU}$  tends to explain the variation of flow characteristics not explained by the first ordination axis. Effects of cultivation greatly depend on tilling methods and crop types. Tilling across the slopes and construction of contour bunds for the purpose of reducing soil erosion from cultivated lands encourage retention of surface runoff within cultivated lands. These are standard land use practices recommended within the study area. It is likely that these land uses practices result in some reduction in flows as is suggested by the  $LC_{CU}$  arrow in Figure 4.3 pointing in a slightly opposite direction to  $q_{90}$  and  $\bar{Q}_{yr}$ .

Kalahari sands and alluvial deposits are plotted pointing in a direction opposite that of  $BFI$  in Figure 4.3 revealing that these formations have minor negative effects on groundwater contribution to streams. Kalahari sands are most

extensive in the north-western part of Zimbabwe, and rivers in that part of the country have insignificant dry weather flows due to limited groundwater contribution. Groundwater often occurs in Kalahari sands at some depth below river beds. The same is true for alluvial deposits.

Catchment characteristics cannot account for 25% of the variation in flow characteristics. The unaccounted variation in flow characteristics can be due to random behaviour in the hydrological response. There is a possibility that effects of some of catchment characteristics used in this study cannot be detected at the catchment scale. The influence of these factors is masked by other factors, and cannot explain this unaccounted variation of flow characteristics. Alternatively this suggests that further investigations have to be made to identify additional catchment factors that account for this unexplained variation. There is also the possibility of using alternative quantitative measures of catchment characteristics to those that have been used in this study. For example, instead of using area covered by different lithologies, hydrogeological parameters like transmissivity and well yields, could be considered. Data for these parameters are only available for sites at which wells have been drilled. Groundwater exploration for well development aims at selecting sites with high transmissivity and well yield. Therefore, the available transmissivity and well yield data are not representative of hydrogeological conditions within the whole catchment. This inhibits application of these parameters in regionalisation studies.

The analysis in this chapter has demonstrated the value of direct gradient techniques like RDA in regionalisation studies. Nathan and McMahon (1990a) highlighted one of the major problems encountered in regionalisation studies, which is the selection of those catchment characteristics that affect the multi-dimensional hydrological response. Catchment characteristics identified in this chapter provide a basis for delimiting hydrologically homogenous regions. Univariate correlation analysis will only suggest explanatory variables for each of the response variables but without taking into account the multi-dimensional nature of the hydrological response. Under these circumstances several different sets of catchment characteristics would have to be used during cluster analysis aimed at delimiting hydrologically homogenous regions. In contrast RDA does determine those catchment characteristics that significantly affect the several different flow characteristics.  $\bar{P}_{yr}$ ,  $S_{50}$ ,  $LC_{PL}$ , and  $LC_{CU}$  explain the variability of the set of selected flow characteristics. Although other catchment characteristics vary between catchments, they are not significant in explaining the variability of flow characteristics. Cluster analysis can be undertaken using these catchment characteristics that have been identified as being significant. This is done in the next chapter.

## 5 IDENTIFICATION OF CATCHMENTS WITH SIMILAR HYDROLOGICAL RESPONSES

### 5.1 Introduction

Classification of catchments into groups with similar hydrological responses has been proposed as a feasible approach for regionalising flow characteristics (Gustard et al., 1989; Nathan and McMahon, 1990a). Flow characteristics of an ungauged catchment can then be estimated from those of the cluster to which it belongs. A major problem with cluster analysis for purposes of regionalisation is the selection of catchment descriptors, which ensure that clusters derived have similar hydrological responses. Redundancy analysis done in Chapter 4 enables identification of explanatory catchment descriptors for several flow characteristics, and these catchment descriptors can be used for classifying catchments. The aim of this chapter is to use these catchment descriptors to classify catchments into clusters with similar hydrological responses, and then determine if these clusters improve the prediction of flow characteristics from catchment characteristics.

### 5.2 Methodology

Cluster analysis aims to partition a set of objects into similar sub-sets. For these clusters the within-group dissimilarity should be minimised while maximising dissimilarity between clusters (Klastorin, 1983; Punj and Stewart, 1983; Everitt, 1993; Legendre and Legendre, 1998; Gordon, 1999; McGarigal *et al.*, 2000). Catchments are partitioned using catchments characteristics so that an ungauged catchment can hopefully be allocated using the same catchment characteristics to one of the clusters. Flow characteristics are used to verify if clusters have similar hydrological responses. This study uses hierarchical cluster analysis because the number of clusters is not known *a priori*. Ward's minimum variance linkage method together with the Euclidean distance similarity measure is used.

#### 5.2.1 Selection of catchment descriptors

Legendre and Legendre (1998) pointed out that the inclusion of insignificant descriptors of entities introduces noise in clustering, and therefore the results may not reveal clusters that represent the variation of hydrological responses. Punj and Stewart (1983) emphasized that proper selection of variables used for classification is critical in order to ensure that clusters identified satisfy classification objectives. Redundancy analysis revealed that the following catchment characteristics explain the variation of flow characteristics

- mean annual precipitation ( $\bar{P}_{yr}$ ),

- median slope ( $S_{50}$ ),
- proportion of the catchment with plantations ( $LC_{PL}$ ),
- proportion of the catchment cultivated ( $LC_{CU}$ ),
- proportion of the catchment underlain by granite and gneiss ( $GL_{GG}$ )
- proportion of the catchment underlain by Kalahari sands ( $GL_{KL}$ )
- proportion of the catchment with alluvial deposits ( $GL_{AL}$ )
- proportion of the catchment with woodlands ( $LC_{WD}$ ),
- drainage density ( $D_d$ )
- proportion of the catchment underlain greenstones ( $GL_{GR}$ )
- size of the catchment area ( $A$ ),
- proportion of the catchment with bushlands ( $LC_{BU}$ ).

These catchment characteristics are standardized (Eqn 5.1) to ensure that the analysis is independent of measurement units used for these variables.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (5.1)$$

where

$i = 1, \dots, n_c$  catchments,

$j = 1, \dots, n_p$  explanatory variables,

$z_{ij}$  = standardized variable  $j$  at catchment  $i$ ,

$x_{ij}$  = value of variable  $j$  at catchment  $i$ ,

$\bar{x}_j$  = mean of variable  $j$  for the  $n_c$  catchments,

$s_j$  = standard deviation of variable  $j$  over all the  $n_c$  catchments.

Redundancy analysis showed that different catchment characteristics were not equally important in explaining the variance of flow characteristics (Table 4.6). There is justification to take into account these differences during cluster analysis. Everitt (1993) argued against weighting variables used for classification since weights cannot be established before undertaking cluster analysis. But redundancy analysis provides a basis for weighting variables. This study explores the utility of weighting catchment characteristics on the basis of the amount of variance of flow characteristics that each variable accounts for. The standardized variables  $z_{ij}$  are weighted by  $w_j$ . All the selected catchment characteristics together explain 75.% of the variance of flow characteristics (Chapter 4).  $\bar{P}_{yr}$  explains 50%, and therefore a weight,  $w_j = 0.67 (=50/75)$  is used. Weights of other catchment characteristics are similarly derived. A weighted explanatory variable,  $y_{ij}$  is hence given by

$$y_{ij} = w_j z_{ij} \quad (5.2)$$

## 5.2.2 Determination of the number of clusters and validation

There is no completely satisfactory method for testing the significant number of clusters (Everitt, 1993; Punj and Stewart, 1983; Sarle, 1996). The null hypothesis to be tested is that catchments have been randomly classified. Sarle (1996) reviewed several tests and shows that most of these have major weaknesses. Everitt (1993) and McGarigal *et al.* (2000) recommend a visual inspection of the fusion levels displayed on a dendrogram. Fusion of clusters at large distances suggests dissimilarity between clusters. Therefore, an examination of distances at which clusters fuse is used in this study to assist in determining the number of clusters.

Validation of the results of cluster analysis aims at ascertaining whether the results are hydrologically sensible. This can be done in a non-statistical sense by means of a visual inspection of cluster membership. For example, a clustering of catchments in the high rainfall and high runoff Eastern Highlands region together with catchments in the dry south-western part of Zimbabwe is not hydrologically sensible. This validation method is not effective for those catchments with no easily discernible differences, especially those occurring along the central watershed of Zimbabwe.

Nathan and McMahon (1990a) suggested the use of curves developed by Andrews (1972) to assess the homogeneity of cluster membership and allocating catchments with doubtful cluster membership to the most feasible cluster. These curves are produced as

$$f(v) = \frac{z_{i1}}{\sqrt{2}} + z_{i2} \sin(v) + z_{i3} \cos(v) + z_{i4} \sin(v) + z_{i5} \cos(v) + \dots \quad (5.3)$$

where  $z_{i1}, z_{i2}, \dots$  are standardized catchment characteristics, and  $v$  ranges from  $-\pi$  to  $+\pi$ . This function preserves the distance measure and can therefore be used to visually assess the homogeneity of clusters. The shapes of curves are affected by the order which catchment characteristics are entered into  $f(v)$ . Nathan and McMahon (1990a) suggested that the most important variable in terms of explaining flow characteristics should be designated as  $z_{i1}$ , and the second most important variable as  $z_{i2}$ , and so on. The reason being that variables entered at the beginning have low frequency cycles that are readily discerned, while later variables have high frequency cycles that are not easily discerned. The following order which reflects the importance of the variables based on the results of redundancy analysis is used in constructing the Andrews' curves

$$\begin{array}{lll} z_{i1} = \bar{P}_{yr}, & z_{i2} = S_{50}, & z_{i3} = LC_{PL}, \\ z_{i4} = LC_{CU}, & z_{i5} = GL_{GG}, & z_{i6} = GL_{KL}, \\ z_{i7} = GL_{AL}, & z_{i8} = LC_{WD}, & z_{i9} = Dd \end{array}$$

$$z_{i10} = GG_{GR} \quad z_{i11} = A \quad z_{i12} = LC_{BU}$$

If a catchment has a curve with a shape that differs from those of other members of the same cluster, then the membership of such a catchment to that cluster is doubtful (Nathan and McMahon, 1990a).

Within the context of regionalisation studies, it is important that grouping of catchments explains the variability of flow characteristics. Therefore an important validation technique is to compare clusters based on a) catchment characteristics, and b) with those derived from flow characteristics. Flow characteristics that are of interest to this study are

- mean annual runoff ( $\bar{Q}_{yr}$ ),
- coefficient of variation of annual runoff ( $CV$ ),
- base flow index ( $BFI$ ), and
- $q_{90}$ ,  $q_{70}$ , and  $q_{50}$ .

The above flow characteristics are used to cluster catchments. If there is agreement between clusters based on catchment characteristics and those based on flow characteristics, this will indicate the validity of delineated hydrologically homogenous regions. This procedure is referred to as external validation of cluster analysis (Punj and Stewart, 1983; Everitt, 1993; Legendre and Legendre, 1998). Everitt (1993) recommends the use of an  $R_g$  index to establish the level of agreement between clusters based on catchment characteristics and those derived from flow characteristics.  $R_g$  is the ratio of the total number of pairs of catchments that are grouped in the two clustering procedures (i.e. one based on catchment characteristics, and the other on flow characteristics), and those which occur in different groups, to the total number of possible pairs.

$$R_g = \left[ T_g - \frac{U_g}{2} - \frac{V_g}{2} + \binom{n}{2} \right] / \binom{n}{2} \quad (5.4)$$

$$T_g = \sum_{i=1}^g \sum_{j=1}^g m_{ij}^2 - n \quad (5.5)$$

$$U_g = \sum_{i=1}^g m_{j.}^2 - n \quad (5.6)$$

$$V_g = \sum_{i=1}^g m_{.j}^2 - n \quad (5.7)$$

where  $g$  is the number of cluster,  $m_{ij}$  the number of catchments in common between the  $i$ th cluster based on catchment characteristics, and  $j$ th cluster based on flow characteristics. These form a matrix  $M$ .  $m_{.j}$  equals the marginal column total of  $M$ , and  $m_{j.}$  equals the marginal row total of  $M$ .  $R_g$  has values ranging from 0 to 1. Values close to 1 show agreement between two clustering methods, whereas values close to 0 show disagreement.

Another approach for validating clustering results is to undertake a canonical variate analysis (CVA) which is the same as Fisher's linear discriminant analysis (Ter Braak and Smilauer, 1998). Flow characteristics are assumed to be the discriminating variables for clusters derived using catchment characteristics. If flow characteristics can discriminate between these clusters, this indicates that clusters derived from catchment characteristics are homogenous with respect to flow characteristics.

### 5.3 Results and Discussion

#### 5.3.1 Classification using catchment characteristics

Table 5.1 shows the cluster membership for 2 to 10 clusters identified using Ward's clustering technique. The above results show that the increase in the number of clusters from 5 to 7 does not cause significant changes to the compositions of clusters. Catchments for which cluster membership is affected by the increase in the number of clusters are E29, E127, and E144.

When the number of clusters increased from 5 to 8, this resulted in the subdivision of Cluster 1. There are no major changes in cluster membership for the other clusters. Further increases in the number of clusters resulted in minor changes in cluster membership.

#### 5.3.2 Classification using flow characteristics

Table 5.2 shows cluster membership obtained using flow characteristics; mean annual runoff ( $\bar{Q}_{yr}$ ), coefficient of variation of annual runoff ( $CV$ ), base flow index ( $BFI$ ), and percentile flows of the flow duration curve;  $q_{90}$ ,  $q_{70}$ , and  $q_{50}$  for clustering.

No major changes occur in cluster membership when the number of clusters increases from 6 to 10. Catchments which belong to Clusters 1, 2, and 3 when the total number of clusters is set to 6, have very stable cluster membership as they remain in the same clusters when the number of clusters increases.

**Table 5.1:** Cluster membership for 2 to 10 clusters based on cluster analysis of catchment characteristics.

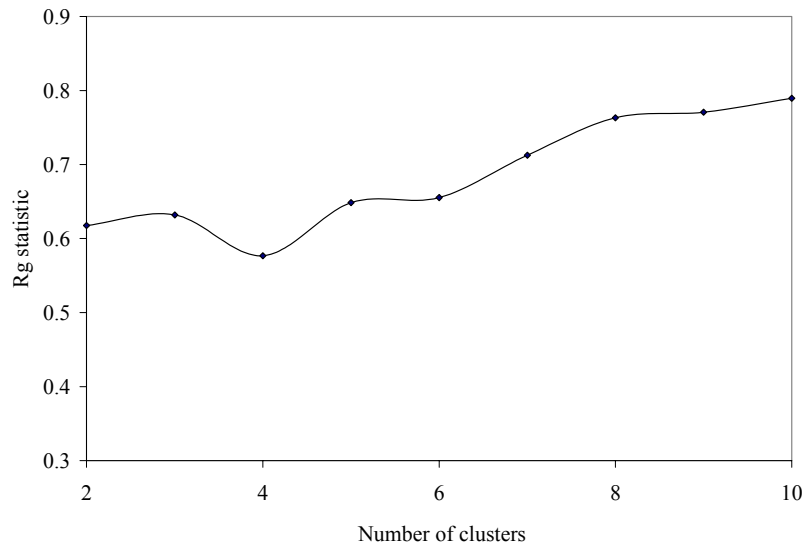
Catchments	Number of Clusters									
	10	9	8	7	6	5	4	3	2	
C13	1	1	1	1	1	1	1	1	1	
C6	1	1	1	1	1	1	1	1	1	
D24	1	1	1	1	1	1	1	1	1	
E107	1	1	1	1	1	1	1	1	1	
E108	1	1	1	1	1	1	1	1	1	
E125	1	1	1	1	1	1	1	1	1	
E42	1	1	1	1	1	1	1	1	1	
C14	2	2	2	1	1	1	1	1	1	
C23	2	2	2	1	1	1	1	1	1	
C33	2	2	2	1	1	1	1	1	1	
C70	2	2	2	1	1	1	1	1	1	
D42	2	2	2	1	1	1	1	1	1	
D44	2	2	2	1	1	1	1	1	1	
D70	2	2	2	1	1	1	1	1	1	
E1	2	2	2	1	1	1	1	1	1	
E114	2	2	2	1	1	1	1	1	1	
E136	2	2	2	1	1	1	1	1	1	
E30	2	2	2	1	1	1	1	1	1	
E40	2	2	2	1	1	1	1	1	1	
E44	2	2	2	1	1	1	1	1	1	
C18	3	3	3	2	2	2	2	1	1	
C41	3	3	3	2	2	2	2	1	1	
C47	3	3	3	2	2	2	2	1	1	
E112	3	3	3	2	2	2	2	1	1	
E23	3	3	3	2	2	2	2	1	1	
E24	3	3	3	2	2	2	2	1	1	
E28	3	3	3	2	2	2	2	1	1	
E33	3	3	3	2	2	2	2	1	1	
E45	3	3	3	2	2	2	2	1	1	
E49	3	3	3	2	2	2	2	1	1	
C25	4	4	4	3	3	3	1	1	1	
D45	4	4	4	3	3	3	1	1	1	
D48	4	4	4	3	3	3	1	1	1	
D50	4	4	4	3	3	3	1	1	1	
E115	4	4	4	3	3	3	1	1	1	
C43	5	4	4	3	3	3	1	1	1	
D27	5	4	4	3	3	3	1	1	1	
D28	5	4	4	3	3	3	1	1	1	
D6	5	4	4	3	3	3	1	1	1	
E123	5	4	4	3	3	3	1	1	1	
E132	5	4	4	3	3	3	1	1	1	
E152	5	4	4	3	3	3	1	1	1	
E106	6	5	5	4	4	4	3	2	2	
E129	6	5	5	4	4	4	3	2	2	
E121	7	6	6	5	5	4	3	2	2	
E16	7	6	6	5	5	4	3	2	2	
E37	7	6	6	5	5	4	3	2	2	
E127	8	7	7	6	3	3	1	1	1	
E144	8	7	7	6	3	3	1	1	1	
E29	8	7	7	6	3	3	1	1	1	
E35	9	8	3	2	2	2	2	1	1	
E72	10	9	8	7	6	5	4	3	2	

**Table 5.2:** Cluster membership based on cluster analysis of flow characteristics

Catchments	Number of Clusters									
	10	9	8	7	6	5	4	3	2	
C13	1	1	1	1	1	1	1	1	1	
C18	1	1	1	1	1	1	1	1	1	
C23	1	1	1	1	1	1	1	1	1	
C33	1	1	1	1	1	1	1	1	1	
C70	1	1	1	1	1	1	1	1	1	
D42	1	1	1	1	1	1	1	1	1	
D70	1	1	1	1	1	1	1	1	1	
E108	1	1	1	1	1	1	1	1	1	
E112	1	1	1	1	1	1	1	1	1	
E123	1	1	1	1	1	1	1	1	1	
E40	1	1	1	1	1	1	1	1	1	
E42	1	1	1	1	1	1	1	1	1	
E44	1	1	1	1	1	1	1	1	1	
E45	1	1	1	1	1	1	1	1	1	
E49	1	1	1	1	1	1	1	1	1	
C14	2	2	2	2	2	2	2	1	1	
C43	2	2	2	2	2	2	2	1	1	
E33	2	2	2	2	2	2	2	1	1	
C25	3	3	3	3	1	1	1	1	1	
D44	3	3	3	3	1	1	1	1	1	
D45	3	3	3	3	1	1	1	1	1	
D48	3	3	3	3	1	1	1	1	1	
D50	3	3	3	3	1	1	1	1	1	
E144	3	3	3	3	1	1	1	1	1	
C41	4	4	2	2	2	2	2	1	1	
C47	4	4	2	2	2	2	2	1	1	
C6	4	4	2	2	2	2	2	1	1	
E23	4	4	2	2	2	2	2	1	1	
E24	4	4	2	2	2	2	2	1	1	
E28	4	4	2	2	2	2	2	1	1	
E30	4	4	2	2	2	2	2	1	1	
E35	4	4	2	2	2	2	2	1	1	
D24	5	5	4	4	3	3	3	2	2	
D27	5	5	4	4	3	3	3	2	2	
D28	5	5	4	4	3	3	3	2	2	
D6	5	5	4	4	3	3	3	2	2	
E114	5	5	4	4	3	3	3	2	2	
E115	5	5	4	4	3	3	3	2	2	
E132	5	5	4	4	3	3	3	2	2	
E136	5	5	4	4	3	3	3	2	2	
E152	5	5	4	4	3	3	3	2	2	
E16	5	5	4	4	3	3	3	2	2	
E29	5	5	4	4	3	3	3	2	2	
E1	6	6	5	4	3	3	3	2	2	
E121	6	6	5	4	3	3	3	2	2	
E125	6	6	5	4	3	3	3	2	2	
E37	6	6	5	4	3	3	3	2	2	
E106	7	7	6	5	4	4	4	3	2	
E127	7	7	6	5	4	4	4	3	2	
E107	8	8	7	6	5	5	2	1	1	
E129	9	7	6	5	4	4	4	3	2	
E72	10	9	8	7	6	4	4	3	2	

### 5.3.3 Number of clusters

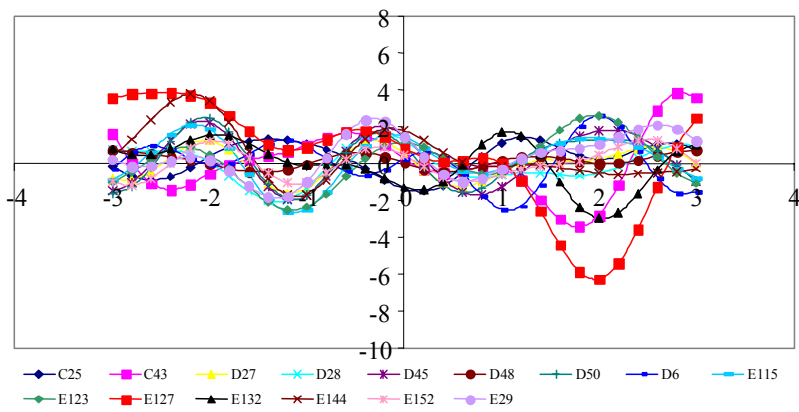
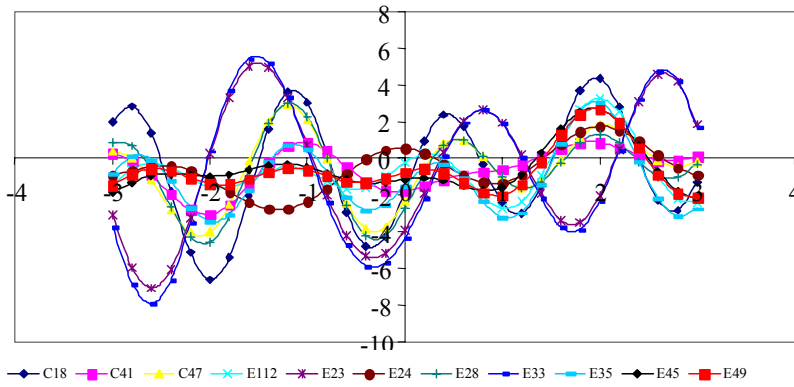
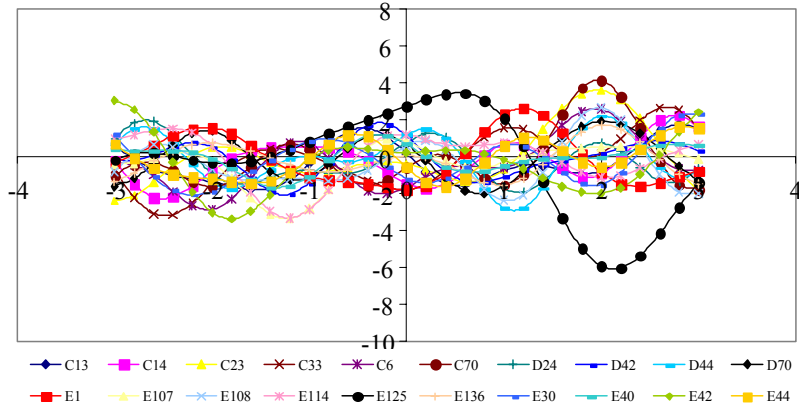
Figure 5.1 presents the  $R_g$  statistic which measures the level of agreement in cluster membership between clusters formed using a) catchment descriptors, and b) flow characteristics.

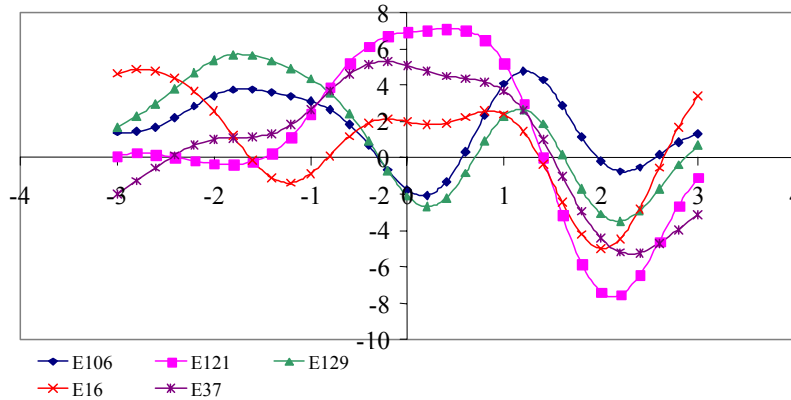


**Figure 5.1:** Variation of  $R_g$  statistic with number of clusters

Increasing the number of clusters from 2 to 5 causes  $R_g$  to increase from 0.62 to 0.65. No significant changes of  $R_g$  occur when the number of clusters increases from 8 to 10. If  $R_g$  is used as the only criterion for determining the number of clusters, then there should be 8 clusters.

A canonical variate analysis was undertaken to determine whether flow characteristics are discriminatory variables for clusters derived from catchment descriptors. Since the  $R_g$  statistic suggests 5 or 8 as the desirable number of clusters, canonical variate analysis was undertaken for 5 and 8 clusters. Flow characteristics explain 29.1% of the variance of clusters when the number of clusters based on catchment descriptors is 8, and with 5 clusters 43.2%. An increase in the explained variance of clusters is likely to result in an improvement in the identification of catchments with similar hydrological responses. Therefore, 5 clusters are used in subsequent analysis. Andrew's curves for these clusters are plotted in Figure 5.2.





**Cluster 4**

**Figure 5.2:** Andrew's curves constructed from catchment characteristics which explain flow characteristics

E107 and E125 have Andrew's curves different from other members of Cluster 1 (Figure 5.2). E125 is an Eastern Highland catchment underlain by the Umkondo ( $GL_{LM}$ ) assemblage and was clustered with catchments that represent physiographic conditions occurring on the central part of the country. The median slope on this catchment is 13%, whereas it is less than 5% for most of the members of this cluster.  $\bar{E}_{pot, yr}$  of E125 is  $1611 \text{ mm yr}^{-1}$ , while the other catchments have  $\bar{E}_{pot, yr}$  greater than  $1800 \text{ mm yr}^{-1}$ . E107 lies in a drier region in comparison to the other members of this cluster. These catchments were removed from this cluster. Misclassification could be a result of the Euclidean distance measure not discriminating adequately between some catchments.

Within Cluster 2, E24 has a curve that has some differences with the rest of the group (Figure 5.2). Although this catchment has a similar  $\bar{P}_{yr}$  to the other members of this cluster, it has a high median slope (4%) in comparison to the rest of the group ( $< 2.5\%$ ). This catchment has the largest proportion of the area with woodlands, 45%, while this is less than 33% for the other catchments. Consequently this catchment was removed from Cluster 2.

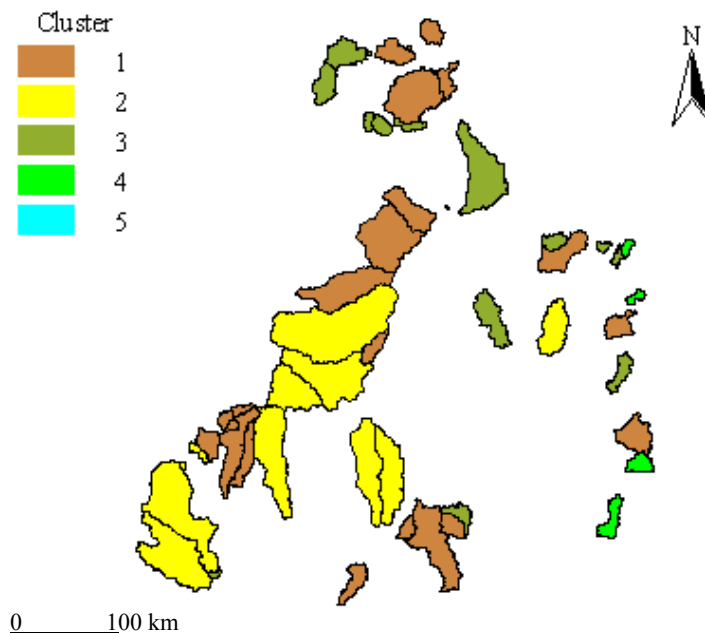
C43, E127 and E132 have curves that differ from other members of Cluster 3 (Figure 5.2). C43 has a median slope of 1.7% while other members of this cluster have median slopes greater than 4%. This is the smallest catchment of the study area ( $3.5 \text{ km}^2$ ). E127 and E132 are Eastern Highlands catchments that have been clustered with catchments occurring on the highveld. Both

catchments have  $\overline{E}_{pot,yr}$  in the 1400 – 1700 mm yr<sup>-1</sup> range, while other members have  $\overline{E}_{pot,yr}$  greater than 1800 mm yr<sup>-1</sup>. The median slopes of both catchments are in the 9-12% range, while this is 4 – 5% for the other catchments. These three catchments were therefore removed from Cluster 3.

The Andrews' curves for Cluster 4 show that E121, and E37 have curves differing from the rest of the group (Figure 5.2). The underlying lithology on both catchments is the Umkondo assemblage which has hydrogeological properties that differ from granite and gneiss occurring on the other catchments. These two catchments were therefore removed from Cluster 4. Andrews' curves were not constructed for Cluster 5 that has only a single catchment.

### 5.3.4 Catchment characteristics of clusters

Catchments belonging to Clusters 1 and 3 do not occur in specific geographical regions (Figure 5.3). Cluster 2 catchments tend to occur on the central and southern parts of the study area. E16, E106 and E129 belong to Cluster 4 and are located on the Eastern Highlands. Cluster 5 has only one catchment, and this is also located on the Eastern Highlands.



**Figure 5.3:** Cluster membership of catchments derived using weighted catchment characteristics

Table 5.3 presents for each cluster the average values of catchment characteristics used for classification. The most distinguishing feature of the clusters is  $\bar{P}_{yr}$  (Table 5.3 and Figure 5.4).  $\bar{P}_{yr}$  increases from Cluster 1, 3 to 5 while Cluster 2 has the lowest value. There are almost no overlaps in the range of values of  $\bar{P}_{yr}$  between clusters (Figure 5.4). Clusters 1 and 2 have similar  $\bar{E}_{pot,yr}$  values (Table 5.3 and Figure 5.4), and this decreases from Cluster 3 to 5.

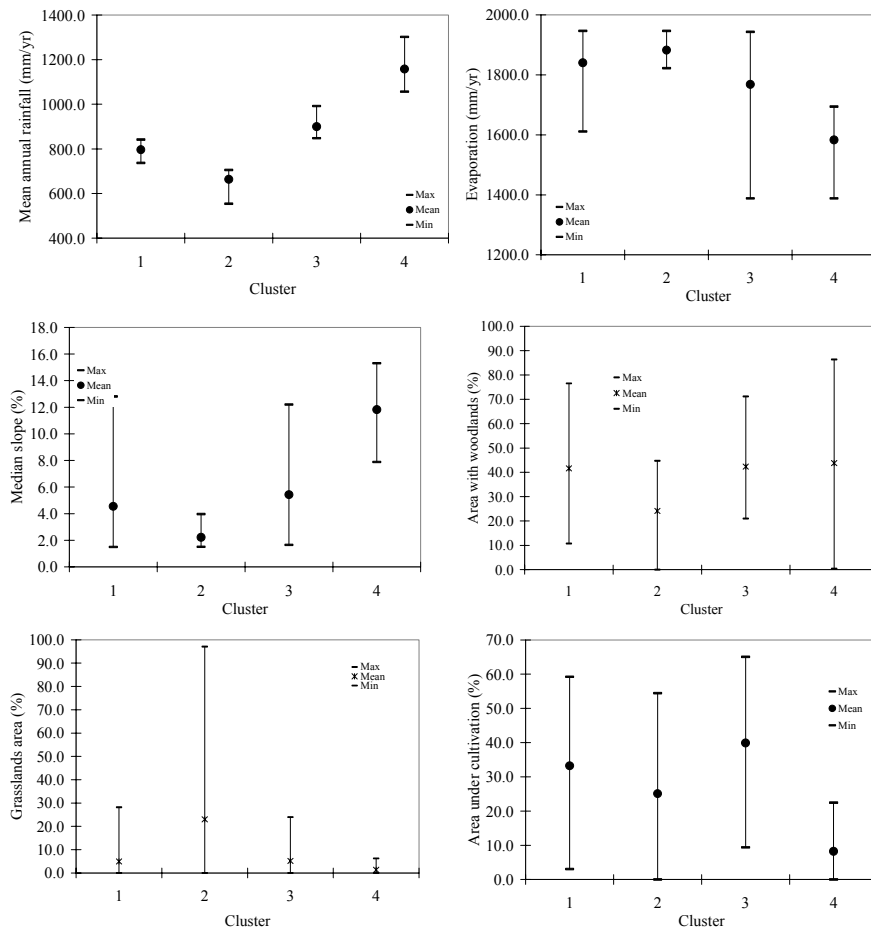
**Table 5.3: Average values of catchment characteristics for clusters**

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
$\bar{P}_{yr}$ (mm yr <sup>-1</sup> )	800.7	660.0	899.9	1207.9	1796.8
$\bar{E}_{pot,yr}$ (mm yr <sup>-1</sup> )	1848.2	876.3	1768.	1564.3	1388.0
$S_{50}$ (%)	4.02	2.05	5.43	10.98	17.61
$LC_{WD}$ (%)	39.83	22.07	42.35	28.12	0.29
$LC_{GR}$ (%)	4.98	25.3	5.17	0.13	0.00
$LC_{CU}$ (%)	33.65	22.45	39.92	4.88	0.00

Catchments belonging to Clusters 1 and 2 have gentle slopes, but slopes increase from Cluster 3 to 5 (Figure 5.4). Cluster 5 has the steepest slopes with a median slope of 17.6%. No major differences occur between Cluster 1 and 3 in the proportion of the area under woodlands (Figure 5.4). Cluster 5 has no woodlands and is all under plantations. With regards to the area with grasslands, Cluster 2 is the only cluster with substantial areas with grasslands (Figure 5.4). Clusters 1 to 3 do not differ in the proportion of the area cultivated (Figure 5.4). The area cultivated is not very significant in Cluster 4, while this is absent in Cluster 5. In general land cover type is not a distinguishing characteristic for Clusters 1 to 3.

### 5.3.5 Flow characteristics

An analysis of flow characteristics of clusters based on catchment descriptors is presented below. This analysis will enable an assessment of whether each of the identified clusters can be considered to have homogenous hydrological responses. Table 5.4 summarises the flow characteristics of these clusters.



**Figure 5.4:** Catchment characteristics of clusters.

**Table 5.4:** Average values of flow characteristics for clusters derived using catchment characteristics

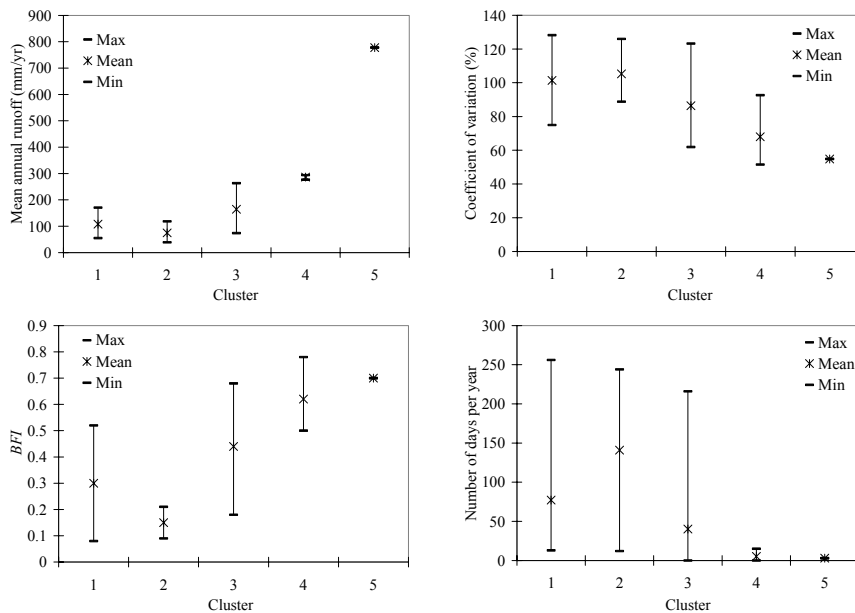
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
$\bar{Q}_{yr}$ (mm yr <sup>-1</sup> )	108.3	75.0	164.1	286.8	778.0
$BFI$	0.30	0.15	0.44	0.62	0.70
$\bar{N}_{DZ}$	4.02	2.05	5.43	10.98	17.61
$q_{70}$	0.042	0.001	0.105	0.305	0.419

$\bar{Q}_{yr}$  increases from Cluster 1 to Clusters 3, 4 and 5 (Figure 5.5), and Cluster 2 has the lowest  $\bar{Q}_{yr}$ . The  $CV$  has a similar pattern across these clusters. The are

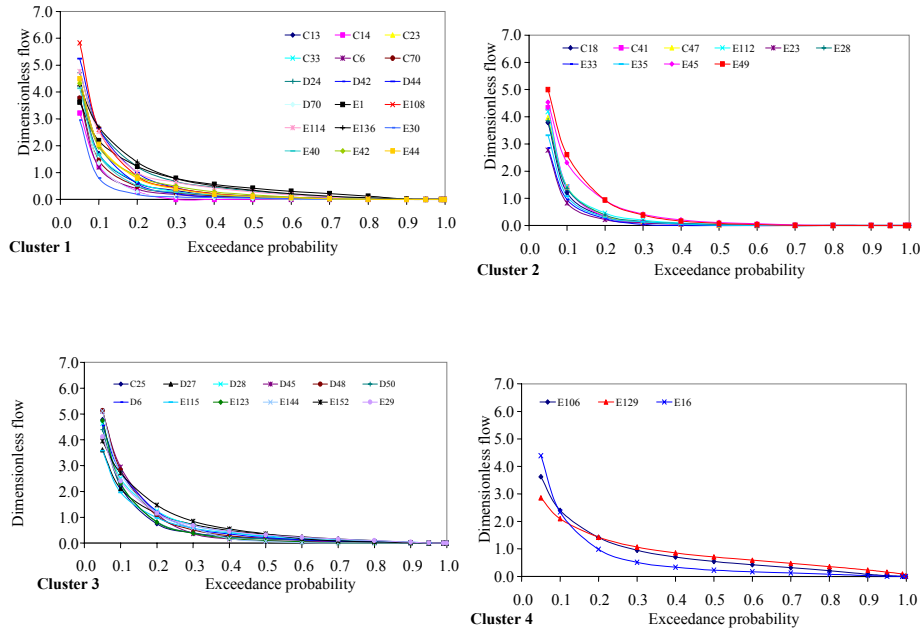
no major differences in the  $CV$  between Cluster 1 and 2, and this decreases from Cluster 3 to 5.

Figure 5.5 shows that Cluster 2 has the lowest  $BFI$  indicating that rivers in this cluster will dry up for prolonged periods. This is confirmed by the average number of days per year without flow for Cluster 2, 140 days/yr.  $BFI$  increases from Cluster 1, 3 to 5. Similarly  $\bar{N}_{DZ}$  decreases from Cluster 1, 3 to 5. The high  $BFI$  and small  $\bar{N}_{DZ}$  for Clusters 4 and 5 show that rivers in these clusters are perennial.

Figure 5.6 shows flow duration curves for the four clusters, and for most of the rivers in Cluster 1 the flow with a 0.6 – 0.70 exceedance probability is the zero flow. Cluster 2 has very steep flow duration curves indicating that rivers in this cluster run dry soon after the wet season. The zero flow has a 0.5 exceedance probability for this cluster. E45 and E49 have curves that differ in their slopes from all the other catchments. Flow duration curves for Cluster 3 are rather flat for flows with over 0.1 exceedance probability. This indicates that flow depletion is gradual for all other flows except peak flows. Rivers in Cluster 4 have perennial flow as is evident from their flow duration curves. Even for the 0.9 exceedance probability, the flows are still greater than zero



**Figure 5.5:** Flow characteristics of clusters derived using weighted catchment characteristics.



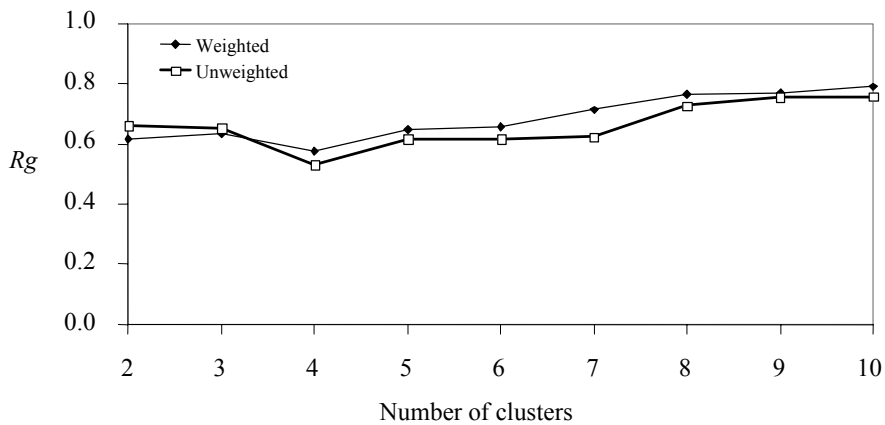
**Figure 5.6:** Flow duration curves of clusters derived from weighted catchment characteristics

The Kruskal-Wallis test was used to test the hypothesis that there were no significant differences of flow characteristics (i.e.  $\overline{Q}_{yr}$ ,  $CV$ ,  $BFI$ ,  $\overline{N}_{DZ}$ ,  $q_{70}$  and  $q_{50}$ ) between catchments belonging to different clusters. This hypothesis was rejected at the 5% significance level for each of these flow characteristics. Another hypothesis tested using the same test was that there were no significant differences in percentile flows of the flow duration curves between catchments belonging to different clusters. Percentile flows tested are  $q_{90}$ ,  $q_{80}$ ,  $q_{70}$ ,  $q_{60}$ ,  $q_{50}$ ,  $q_{40}$ ,  $q_{30}$ ,  $q_{20}$ ,  $q_{10}$  and  $q_5$ . This hypothesis was rejected for all percentile flows except  $q_5$  which are flood flows. The results of these two tests indicate that each of the clusters has unique flow characteristics. Each cluster can be regarded as being hydrologically homogenous with respect to these flow characteristics.

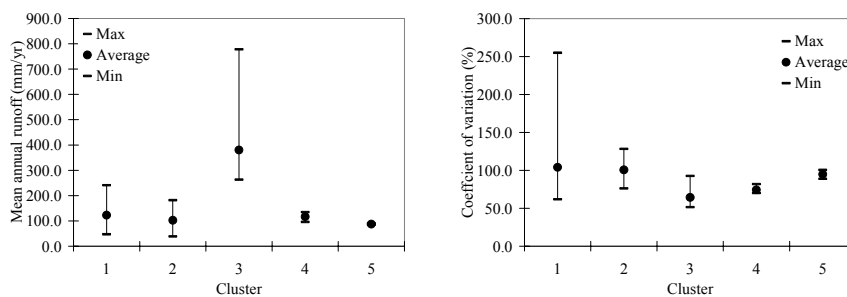
### 5.3.6 Comparison of weighted and unweighted clustering

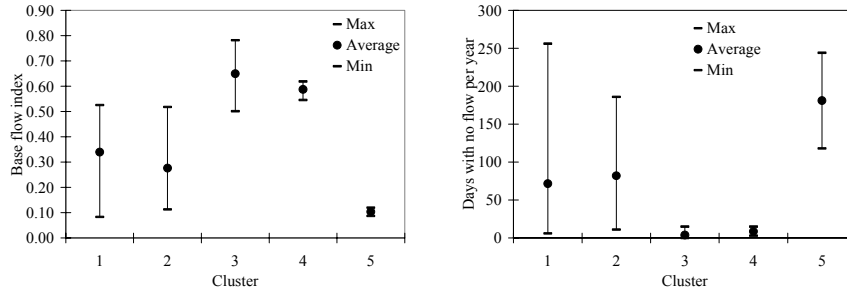
Flow characteristics discussed above are for clusters derived using weighted catchment characteristics. Clustering was also undertaken without weighting these catchment characteristics, so as to determine whether weighting of catchment characteristics improves the identification of clusters with homogenous hydrological responses. The  $R_g$  statistic was estimated by comparing the resemblance between clusters derived using catchment

characteristics and those based on flow characteristics. Figure 5.7 shows that weighted catchment characteristics give a better resemblance to clusters derived using flow characteristics, than clusters based on unweighted catchment characteristics. A canonical variate analysis was also undertaken to determine if flow characteristics can explain the variance of clusters derived using unweighted catchment characteristics. This was done for 5 clusters. Flow characteristics explain 31.0% of the variance of these clusters. But when weighted catchment characteristics were used to derive clusters, the explained variance was 43.2% as discussed in 5.3.3. Thus both the  $R_g$  statistic and canonical variate analysis indicate that weighted catchment characteristics are superior in terms of discriminating catchments with different hydrological responses.



**Figure 5.7:** Comparison of  $R_g$  statistics for cluster analysis using weighted and unweighted catchment characteristics.





**Figure 5.8:** Flow characteristics of clusters derived from unweighted catchment characteristics

Figure 5.8 shows  $\bar{Q}_{yr}$  for clusters derived from unweighted catchment characteristics. Clusters 2, 4 and 5 have all  $\bar{Q}_{yr}$  within the range of  $\bar{Q}_{yr}$  for Cluster 1. But with clusters based on weighted catchment characteristics, the upper limit of one cluster overlapped with the lower limit of the next cluster (Figure 5.5), and not for the whole range to be contained in the range of another cluster. Figure 5.8 shows again that the range of values of the  $CV$  of all other clusters derived from unweighted catchment characteristics were contained within the range of Cluster 1. But this was not the case when weighted catchment characteristics were used (Figure 5.5). Figure 5.8 shows that Cluster 1, 2 and 5, do not differ in their  $BFI$ , and the same is true for Cluster 3 and 4. Clusters derived from unweighted catchment characteristics do not differ greatly in their  $\bar{N}_{DZ}$ .

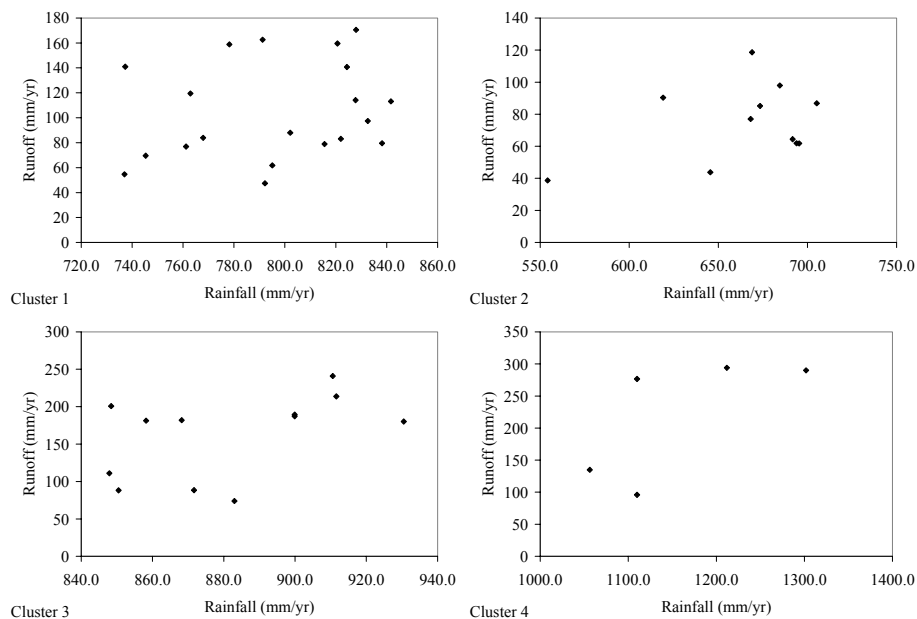
Thus weighting of catchment characteristics before clustering improves the identification of clusters with similar hydrological responses. Flow characteristics of catchments clustered in this manner can be easily discerned.

### 5.3.7 Prediction of flow characteristics of clusters

Chapter 3 identified catchment characteristics that can be used for estimating flow characteristics at ungauged sites. The relevant predictive equations were developed using data for all the 52 catchments. One of the hypotheses of this study is that the prediction of flow characteristics at ungauged sites can be improved by grouping catchments into clusters that have each similar hydrological responses. The validity of this hypothesis is examined below.

### Prediction of $\bar{Q}_{yr}$ , $CV$ and $BFI$

It was established in Chapter 3 that  $\bar{Q}_{yr}$  can be predicted using  $\bar{P}_{yr}$ , granite, and slope indices. Figure 5.9 shows that there is no significant relationship between  $\bar{Q}_{yr}$  and  $\bar{P}_{yr}$  for catchments in Cluster 1 to 4. Cluster 5 has one catchment and it was therefore not possible to determine any relationships between catchment descriptors and flow characteristics. Both  $\bar{Q}_{yr}$  and  $CV$  were also not related to any other catchment descriptors within Clusters 1, 2, and 3. Thus clustering of catchments does not seem to improve the prediction of  $\bar{Q}_{yr}$  and  $CV$  for ungauged sites. Clustering has subdivided catchments into subsets with narrow ranges of physiographic conditions. The relationships between  $\bar{Q}_{yr}$  and  $CV$ , and other flow characteristics are no longer discernible within these narrow ranges of physiographic conditions. This observation suggests that classification of catchments into hydrologically homogenous regions, and the subsequent use of these regions as the basis for developing methods for estimating flow characteristics of ungauged catchments, is only applicable to regions with wide ranges of physiographic conditions.



**Figure 5.9:** Relationship between mean annual runoff and mean annual rainfall for clusters derived using weighted catchment characteristics

$BFI$  was found to be related to  $\bar{P}_{yr}$  in Chapter 3, but this relationship does not exist for each of the clusters. Slope indices from  $S_{25}$  to  $S_{90}$  were found to be correlated to the  $BFI$  in Clusters 1 to 3. The correlation coefficients obtained when  $BFI$  was correlated with  $S_{50}$  are

Cluster 1	$r = 0.63$
Cluster 2	$r = 0.82$
Cluster 3	$r = 0.63$

When  $S_{50}$  was correlated with  $BFI$  for all the catchments without clustering in Chapter 3, the correlation coefficient was 0.76. Therefore, clustering of catchments has not improved the relationship between  $BFI$  and slope for all clusters except Cluster 2.

### Flow duration curves

Percentile flows  $q_{90}$ ,  $q_{70}$ ,  $q_{50}$  and  $q_{20}$  were found to be correlated to  $\bar{P}_{yr}$  in Chapter 3. But these relationships were absent in each cluster. Clustering produced sub-sets of catchment with narrow ranges in the values of physiographic characteristics, and therefore the lack of relationship between percentile flows and  $\bar{P}_{yr}$ .

Chapter 3 established that the following exponential model is suitable for describing the relationship between flows,  $q_p$ , and their exceedance probabilities,  $p$

$$q_p = b_0 \exp(-b_1 p) \quad (5.8)$$

It was established that  $b_0$  and  $b_1$  can be estimated using  $BFI$ , and the coefficient of determination of the predictive equations was 0.74 and 0.92 respectively. Predictive equations for estimating these parameters using  $BFI$  for Cluster 1, 2 and 3 tend to have lower coefficient of determination, and in some cases it was not possible to derive such equations as shown below

#### Cluster 1

$$b_0 = 11.9806 - 11.4488BFI \quad r^2 = 0.49 \quad (5.9)$$

$$b_1 = 0.3227 \exp(-2.9581BFI) \quad r^2 = 0.85 \quad (5.10)$$

## Cluster 2

$b_0$  no significant relationship with  $BFI$

$$b_1 = 0.3241 \exp(-3.1932) \quad r^2 = 0.79 \quad (5.11)$$

## Cluster 3

$$b_0 = 12.6595 - 12.9389 BFI \quad r^2 = 0.61 \quad (5.12)$$

$$b_1 = 0.2816 \exp(-2.6188 BFI) \quad r^2 = 0.75 \quad (5.13)$$

Clustering of catchments has therefore not improved the prediction of parameters of the exponential model for flow duration curves.

The Kruskal-Wallis test was used to test the hypothesis that the values of each of the parameters of the exponential model do not differ between clusters. This hypothesis was rejected for each of the parameters. Each cluster has unique hydrological responses, but it is problematic to relate these responses to catchment characteristics.

## 5.4 Conclusion

This chapter has demonstrated that redundancy analysis provides a basis for selecting catchment characteristics for use during clustering. Although Everitt (1993) argued against weighting variables used for clustering since the importance of these variables were not known a priori, this study has shown that the results of a redundancy analysis provide a basis for deriving weights to be applied to the selected variables. The use of both the Rg statistic and canonical variate analysis enables the determination of the desirable number of clusters. Andrews' curves enable identification of catchments that are outliers within clusters. RDA results are useful in determining the order with which variables are used to construct Andrews' curves.

The most distinguishing features of clusters identified are  $\bar{P}_{yr}$  and S50. Land cover types do not differ greatly between clusters. Clusters have flow characteristics that are significantly different between them. Therefore, catchments forming each cluster can be regarded as having similar hydrological responses. However, when clustering is done using unweighted catchment characteristics, the clusters do not have distinct flow characteristics. This proved that weighting of catchment characteristics using the amount of variance of the flow characteristics that each catchment characteristic explains, leads to the delimitation of clusters with distinct hydrological responses. Clusters will

however overlap in their characteristics at their margins since clustering subdivides continuous variables into distinct subsets.

Each of the clusters identified using weighted catchment characteristics has narrow ranges in the variation of their physiographic attributes, and consequently there are no discernible relationships between catchment descriptors and flow characteristics. Therefore clustering did not provide a basis for developing equations for predicting flow characteristics at ungauged sites. It is however possible to estimate flow characteristics of an ungauged site using the range or mean values for each cluster. The cluster membership of an ungauged site can be determined by comparing the Andrews' curve of such a site with those of clusters identified for gauged sites.



## 6 REGIONALISATION OF SELECTED RAINFALL-RUNOFF MODELS

### 6.1 Introduction

So far the study has considered estimation of flow statistics such as mean annual flow, flow duration curves, and mean monthly flows of ungauged catchments. These statistics are relevant for most water resources planning and management problems. Other water resources planning problems however require time series of flows, e.g. estimation of yields of reservoirs (McMahon and Mein, 1978). A common approach for estimating time series of flows at ungauged sites is the extrapolation of flow records from gauged sites. Catchments for which flow time series are to be estimated may not have comparable gauged catchments hence prohibiting extrapolation. For such cases the use of a rainfall-runoff model with regionalised parameters may be a feasible option.

Rainfall-runoff models fall into 2 main classes; a) lumped conceptual models, and b) distributed physically based models (Refsgaard and Knudsen, 1996; Beven and O'Connell, 1982). Lumped conceptual models describe mathematically processes within a hydrological system such as interception, surface runoff, and groundwater. Spatial variations of these processes are not accounted for, but rather spatially averaged values are used (Blackie and Eeles, 1985; Wood, 1995). Physically based models use continuum equations to describe temporal and spatial variations of hydrological processes (Beven and O'Connell, 1982).

Estimation of parameter values of physically based models for ungauged catchments should theoretically be feasible because these parameters are supposedly measurable. There has been limited success in the use of these models on ungauged catchments, because the data required for estimating model parameter values are not available at both spatial and temporal scales to enable a truly physically based modelling. A physically based model, SHE, was applied on six catchments in India, but the physiographic data required for estimation of model parameters were only available at a coarse spatial resolution. Therefore, values of model parameters were derived as spatially averaged values, and their physical interpretation was questionable (Refsgaard, *et al.* 1992; Jain, *et al.* 1992; Lohani, *et al.* 1993). Pilgrim (1983) noted that hydrological processes are highly irregular in both space and time, and thus derivation of measurable model parameters is problematic. Refsgaard and Knudsen (1996) compared the ability of a) NAM, a lumped conceptual model, b) WATBAL, a semi-distributed model, and c) MIKE SHE a physically based model, to model flows at ungauged catchments in Zimbabwe. The NAM model

performed as well as the other two models. These studies concluded that when the objective is to derive time series of runoff only, a lumped conceptual model performs better than a physically based model. This study is therefore directed towards regionalisation of lumped conceptual models.

Regionalisation of lumped conceptual models has been done by either extrapolating parameters from gauged sites, or by relating model parameters to catchment characteristics. Pitman (1973) used the knowledge of physiographic conditions in various parts of South Africa to produce maps for regionalising his model, but the validity of suggested parameter values was never ascertained. Wolski (1999) noted that calibrated model parameter values reflected a specific combination of physiographic conditions of a particular catchment, and caution must be exercised when extrapolating these parameters.

An increasing interest exists in predicting parameter values of conceptual models from catchment characteristics. A limitation of this approach in some of the studies has been that while some of the parameters can be predicted, a few require calibration, which is not possible on ungauged catchments (Manley, 1978; Ibrahim and Cordery, 1995). There is increasing evidence that prediction of parameter values for some models is feasible (Post and Jakeman, 1996, 1999; Post *et al.*, 1998; Soften and Howarth, 1998; Fernandez *et al.*, 2000; Berger and Entekhabi, 2001, Yokoo *et al.*, 2001). Most of the previous studies have used multiple regression to predict model parameter values. However, some of the parameters are non-linearly related to and/or have complex relationships with catchment characteristics which cannot be described by multiple regression. For such non-linear and complex relationships, neural networks may be suitable for predicting model parameter values from catchment characteristics. The aim of this chapter is to investigate the possibility of predicting parameter values of conceptual models from catchment characteristics using multiple regression and neural networks.

## **6.2 Methodology**

This study derived catchment descriptors like morphometric properties, proportions of catchments with different lithologies and land cover, and climatic properties. The feasibility of establishing relationships between these catchment descriptors and parameters of selected lumped conceptual models is investigated. Models with a monthly time interval are used which is the time interval appropriate for most water resources planning and management in southern Africa. In addition, the rain gauge network within the study area is sparse and therefore catchment rainfall cannot be estimated accurately at time intervals less than a month for example weekly or daily intervals.

Catchments in the study area have already been classified into clusters using catchment descriptors. An assessment of whether model parameters can be regionalised on the basis of these clusters is made. Regionalisation is possible if each cluster has model parameter values that significantly differ from those of other clusters. A non-parametric test is used to ascertain the validity of the null hypothesis that there are no significant differences in the values of model parameters between clusters, and any differences are due to chance.

### **6.2.1 Selection of rainfall-runoff models**

Several lumped models exist and the selection of models for use in this study is guided by the following points:

- The model addresses the problem, i.e. simulation of monthly flows. (Klemes, 1986; Simmers, 1984; Hendriks, 1990).
- Applicability of the model to the hydroclimatic region of the study area.
- Using the simplest model if possible. Perrin *et al.* (2001) noted that simple conceptual rainfall-runoff models had fewer problems arising from parameter uncertainty than complex models. Uncertainty in the values of model parameters will limit the potential for regionalising model parameters.

Lumped conceptual models selected for use in this study are a) *abcd* model (Fernandez *et al.*, 2000; Alley, 1984), and b) Pitman (1973) model. Fernandez *et al.* (2000) related parameters of the *abcd* model to catchment descriptors, while Vandewiele *et al.* (1992) demonstrated that models similar in structure to the *abcd* model can be regionalised. This model is very simple and has four parameters. The Pitman model has been widely used in southern Africa (Pitman and Middleton, 1994; Hughes, 1995; Hughes and Metzler, 1998). This model has four to six parameters that require calibration. Hughes (1985) found that some parameters of this model can be estimated from catchment descriptors. The potential for regionalising this model in Zimbabwe was highlighted by Hughes (1997).

### **6.2.2 Model calibration**

Calibration of model parameters is done in this study by combining manual trial-and-error and automatic optimisation methods (Green and Stephenson, 1986; Refsgaard and Storm, 1996). The use of these two approaches is meant to minimise their limitations. Manual calibration of model parameters has the advantage that parameter values can be selected so that there are hydrologically meaningful. But this method does not always result in optimal parameter values. Automatic optimisation will be used to fine tune parameter values. If automatic

optimisation is used only without manual calibration, values of parameters may not have any physical relevance. In addition, optimisation routines do not always identify the global optimum, and parameters values may be based on local optima. Ndiritu and Daniel (1999) did caution that parameters at the global optimum do not necessarily give superior simulations. Problems with the model structure, and data errors can be more important than locating the global optimum. Equifinality of model parameters which means that different sets of model parameter values produce similar simulations on the same catchment is also a major constraint when attempting to identify optimal parameters values (Gupta and Sorooshian, 1983; Liden and Harlin, 2000). This affects the feasibility of relating model parameters to catchment characteristics. Those sets of model parameter values that result in some of the state variables assuming values that are hydrologically not meaningful will be eliminated.

This study considers a simulation to be acceptable if simulated monthly flows satisfy the following conditions (Lorup *et al.*, 1998; Schulze and Smithers, 1995):

- i. The difference between the mean of observed and that of simulated monthly flows is within the +/-10% range.
- ii. The difference between the standard deviation of observed flows and that of simulated flows is within the +/-15%.
- iii. Coefficient of efficiency > 0.70
- iv. An acceptable agreement between the flow duration curves of observed and simulated flows based on visual inspection.

The coefficient of efficiency,  $CI$ , was defined by Nash and Sutcliffe (1970) as:

$$CI = \frac{F_o^2 - F^2}{F_o^2} \quad (6.1)$$

with

$$F^2 = \sum_{t=1}^{n_s} (Q_{obs,t} - Q_{sim,t})^2 \quad (6.2)$$

$$F_o^2 = \sum_{t=1}^{n_s} (Q_{obs,t} - \bar{Q}_t)^2 \quad (6.3)$$

$Q_{obs,t}$  = observed or measured monthly flow

$Q_{sim,t}$  = simulated monthly flow

$\bar{Q}_t$  = average of observed monthly flows

$t$  = month interval

$n_s$  = total number of months simulated.

Values of  $CI$  close to 1 indicate close agreement between observed and simulated flows.

The  $abcd$  model is calibrated using optimisation routines that are part of the Microsoft Excel Solver since this model is written on a spreadsheet. Microsoft Excel Solver has quasi-Newton and conjugate gradient optimisation routines (Fylstra, *et al.*, 1998). The version of the Pitman model used in this study is one produced by Hughes and Forsyth (2002), and incorporates a genetic optimisation routine (Ndiritu; 1998). The objective function used for automatic optimisation is the maximisation of  $CI$ . Optimised model parameters are also adjusted to satisfy goodness-of-fit criteria relating to the mean, and standard deviation.

### 6.2.3 Model validation

Calibrated model parameters can result in simulations that satisfy goodness-of-fit criteria, but parameter values may not have any hydrological meaning. Values of model parameters will be a result of curve fitting. This is also reflected in having different sets of parameters values producing simulations which satisfy these criteria. It is necessary to test if parameter values reflect the underlying hydrological processes, and are not a result of curve fitting. This is called model validation (Klemes, 1986, Refsgaard and Storm, 1996). Use is made of the split-sample test and proxy-basin test in this study. The split-sample test involves splitting the available time series into two parts. One part is used to calibrate the model, and the second part is used for testing if calibrated parameters can produce simulations which satisfy goodness-of-fit tests. The split sample test is suitable for catchments with long time series, and it is applied in this study to catchments with over 20 years of data. For such catchments the available record is split into two equal parts. With regards to the proxy-basin test, calibration is done on one catchment, and the parameters are tested on a similar catchment.

### 6.2.4 Structure of models selected

#### 6.2.4.1 The $abcd$ Model

The  $abcd$  model is a nonlinear watershed model introduced by Thomas (1981). In this model the available water,  $W_t$ , is defined as

$$W_t = \Delta t P_t + S_{t-1} \quad (6.4)$$

where

$P_t$  = monthly precipitation during month  $t$  (mm month<sup>-1</sup>),

$S_{t-1}$  = soil moisture at the beginning of month  $t$  (mm),  
 $\Delta t$  = time unit equal to one month included to ensure consistency of measurement units.

Thomas (1981) defined  $Y_t$  referred to as the evaporation opportunity as

$$Y_t = \Delta t E_t + S_t \quad (6.5)$$

where

$E_t$  = actual evaporation during month  $t$  (mm month<sup>-1</sup>),

$S_t$  = soil moisture at the end of month  $t$  (mm).

Since there are no measurements of actual evaporation rates,  $E_t$ , the evaporation opportunity,  $Y_t$  is modelled as a nonlinear function of the available water.

$$Y_t = \frac{W_t + b}{2a} - \sqrt{\left(\frac{W_t + b}{2a}\right)^2 - \frac{W_t b}{a}} \quad (6.6)$$

Parameters  $a$  and  $b$  in Eqn (6.6) have to be calibrated. The function expressed in Eqn (6.6) ensures that  $Y_t \leq W_t$  because  $0 \leq a \leq 1$ , and that the upper limit of  $Y_t = b$ . It is assumed that the rate of actual evaporation is proportional to the soil moisture storage, and the change in soil moisture storage due to evaporation can be expressed as

$$\frac{dS}{dt} = -E_{pot,t} \frac{S_t}{b} \quad (6.7)$$

where  $E_{pot,t}$  = monthly potential evaporation.

If it is assumed that soil moisture storage at the beginning of the month is equal to  $Y_t$ , then

$$S_t = Y_t \exp\left(-\frac{E_{pot,t}}{b}\right). \quad (6.8)$$

The difference between  $W_t$  and  $Y_t$ , gives the total volume of water that forms direct runoff,  $Q_{s,t}$ , and recharges groundwater,  $R_{chg,t}$

$$R_{chg,t} = c(W_t - Y_t) / \Delta t \quad (6.9)$$

$$Q_{s,t} = (1 - c)(W_t - Y_t) / \Delta t \quad (6.10)$$

where  $c$  is a model parameter. Discharge of groundwater to streams,  $Q_{g,t}$ , is assumed to be a function of groundwater storage,  $S_{g,t-1}$ , and modelled by a parameter  $d$ . Thus

$$Q_{g,t} = dS_{g,t-1}. \quad (6.11)$$

The changes in groundwater storage are described by the continuity equation

$$S_{g,t} = S_{g,t-1} + c(W_t - Y_t) - \Delta t Q_{g,t} \quad (6.12)$$

Streamflow,  $Q_t$ , is given by

$$Q_t = (1 - c)(W_t - Y_t) / \Delta t + Q_{g,t} \quad (6.13)$$

The model has four parameters. Parameter  $a$  describes the tendency for runoff to occur before saturation of the soil. Most studies have found values of  $0.95 \leq a \leq 0.98$  (Alley, 1984; Vandewiele *et al.*, 1992; Fernandez *et al.*, 2000) related this parameter to permeability of the soil. Model parameter  $c$  is the proportion of groundwater contribution to streamflow, and therefore should approximate the baseflow index. If streamflow comprises groundwater flow only,

$$Q_t = Q_{g,t} \quad (6.14)$$

and  $K_b$  is defined as the monthly recession constant, then

$$Q_t = K_b Q_{t-1}. \quad (6.15)$$

When streamflow comprises groundwater flows only, then (Vogel and Kroll, 1996; Savenije, 2001)

$$Q_t = -\frac{-\ln(K_b)S_{g,t}}{\Delta t}. \quad (6.16)$$

Dry season flows are described by the recession equation written as

$$Q_t = Q_0 \exp(-\alpha_m t). \quad (6.17)$$

Linsley *et al.* (1982) pointed out that the volume of groundwater discharged during time  $dt$  is equal to the change in storage  $dS_{g,t}$ . Therefore

$$dS_{g,t} = Q_t dt = Q_0 \exp(-\alpha_m t) dt \quad (6.18)$$

The integration of Eqn (6.18) gives

$$S_{g,t} = -\frac{Q_t}{\alpha_m} . \quad (6.19)$$

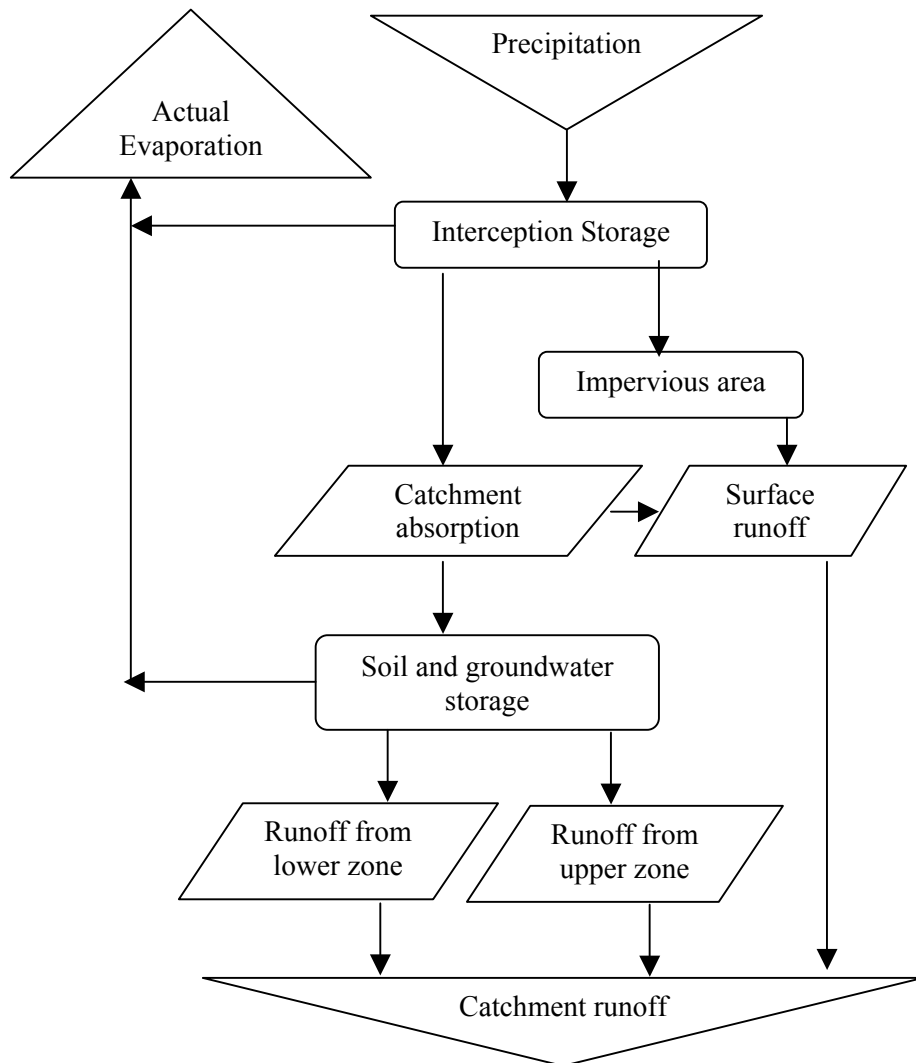
Equations (6.11), (6.16) and (6.19) show that parameter  $d$  of the *abcd* model is related to the recession constant in the following manner

$$d = \frac{\ln(K_b)}{\Delta t} = \frac{1}{\alpha_m} \quad (6.20)$$

In Eqn (6.20)  $\Delta t$  has been included in order to have consistency in the units, since  $K_b$  is dimensionless while  $\alpha_m$  has the dimension of month<sup>-1</sup>. The reciprocal of parameter  $d$  as defined in Eqn (6.20) is the residence time of groundwater (Alley, 1984; Tallaksen, 1995).

#### 6.2.4.2 Pitman model

The Pitman model treats a catchment as having two storages, which are the interception storage, and subsurface storage (Figure 6.1). Pitman originally referred to the subsurface storage as soil moisture storage which combined soil moisture and groundwater. The term subsurface storage is regarded as the most appropriate (Hughes, 2002). The model simulates four processes which are; interception, surface runoff, evaporation from the subsurface storage, and runoff from the subsurface storage.



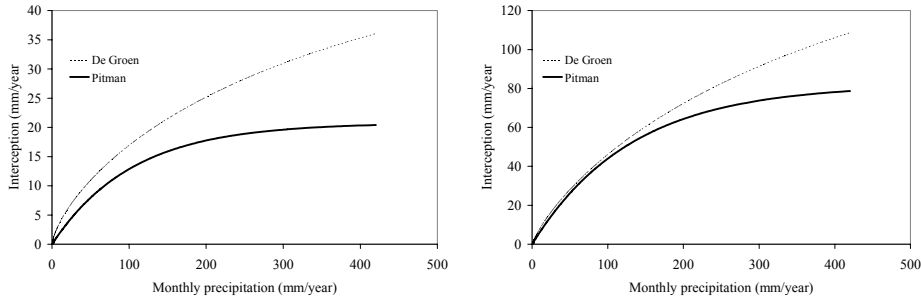
**Figure 6.1:** Structure of the Pitman monthly rainfall-runoff model

## Interception

The model assumes that all the rainfall that is intercepted on a particular day evaporates on the same day. The interception loss is a function of interception capacity,  $I_{cap}$ , and the amount of daily rainfall. The following equation for estimating interception loss,  $I_t$ , was derived empirically in South Africa and has been assumed to be valid in southern Africa

$$I_t = 13.08 I_{cap}^{1.144} \left[ 1 - \exp \left\{ P_t \left( 0.00099 I_{cap}^{0.75} - 0.11 \right) \right\} \right] \text{ mm month}^{-1} \quad (6.21)$$

Measurements made in South Africa showed that  $I_{cap}$  varies from 0 – 8 mm/day (Schulze, 1995). An  $I_{cap} = 1.5$  mm/day for most land cover types and  $I_{cap} = 4$  mm/day for forests were recommended by Pitman (1973). The original model allowed a single value of  $I_{cap}$  to be used for the whole catchment, but the version used in this study allows for subdivision into two sub-catchments with different  $I_{cap}$  values. The proportion of the catchment to which each of the  $I_{cap}$  values is applicable has to be specified.



**Figure 6.2:** Comparison of interception rates estimated by the De Groen and Pitman interception models.  $I_{cap} = 1.5$  mm for the left graph, and  $I_{cap} = 5.0$  mm for the right graph.

The Pitman interception model assumes that the rate of increase of interception gradually decreases as the rainfall increases, while De Groen (2002) assumed that interception continues to increase with increasing rainfall (Figure 6.2). According to De Groen (2002) monthly interception is a function of the amount of rainfall in a particular month,  $P_t$ , number of rainy days,  $n_r$ , and the mean interception capacity of the area,  $I_{cap}$ , and is given by

$$I_t = P_t \left( 1 - \exp \left( \frac{-I_{cap} n_r}{P_t} \right) \right) \text{ mm month}^{-1}. \quad (6.22)$$

De Groen (2002) developed predictive equations for  $n_r$ , the number of rainy days in any month.

This study will investigate whether there are any significant differences in values of Pitman model parameters, and the simulation of runoff when either Eqn (6.21) or (6.22) is used.

### Surface runoff

The model assumes that surface runoff can be formed in two ways. Firstly, all the rainfall on impervious portions of the catchment will form surface runoff. These impervious areas can only contribute to stream flows if they discharge directly into streams, otherwise runoff generated on these parts of the catchment will be absorbed by the surrounding soils. The model requires that the proportion of impervious areas,  $AI$ , be specified if these areas are to contribute to stream flows.

Secondly, surface runoff can be formed as a result of rainfall which does not infiltrate on pervious parts of the catchment. The absorption rate of rainfall into the subsurface storage is considered to vary spatially because of variations in vegetation, soils and geology. It is assumed that as the rainfall increases, it will increasingly exceed absorption rates of increasing proportions of the catchment and therefore contribute to surface runoff. Absorption rates within a catchment are assumed to vary from a minimum of  $Z_{min}$  mm month<sup>-1</sup> to a maximum of  $Z_{max}$  mm month<sup>-1</sup>. The proportions of the catchment area that have absorption rates varying from  $Z_{min}$  to  $Z_{max}$  are assumed to follow a triangular distribution. Thus if the rainfall is less than  $Z_{min}$ , no surface runoff will be formed, and all the rainfall is absorbed. As the rainfall increases above  $Z_{min}$ , the proportion of the catchment that will form runoff will increase. The model assumes that the average absorption rate,  $Z_{ave}$ , is given by

$$Z_{ave} = \frac{Z_{min} + Z_{max}}{2}. \quad (6.23)$$

Using the properties of a triangular frequency distribution of the absorption rates the following equation holds for determining the proportion of the catchment,  $a_{abs}$ , with a specified absorption rate,  $Z$  mm month<sup>-1</sup>

$$a_{abs} = \frac{2(Z - Z_{min})^2}{(Z_{max} - Z_{min})^2} \quad \text{for } Z \leq Z_{ave} \quad (6.24)$$

$$a_{abs} = 1 - \frac{2(Z_{max} - Z)^2}{(Z_{max} - Z_{min})^2} \quad \text{for } Z \geq Z_{ave}. \quad (6.25)$$

When  $Z = Z_{min}$ ,  $a_{abs} = 0$ , and with  $Z = Z_{max}$ ,  $a_{abs} = 1$ . These equations are used to estimate monthly surface runoff,  $Q_{s,t}$  (mm month<sup>-1</sup>) for any given rate of effective rainfall,  $Peff_t = P_t - I_t$ :

$$Q_{s,t} = \frac{2(Peff_t - Z_{min})^3}{3(Z_{max} - Z_{min})^2} \quad \text{for } Z_{min} \leq Peff_t \leq Z_{ave} \quad (6.26)$$

$$Q_{s,t} = \frac{(Z_{max} - Z_{min})}{12} \quad \text{for } Peff_t = Z_{ave} \quad (6.27)$$

$$Q_{s,t} = Peff_t - Z_{ave} + \frac{2(Z_{max} - Peff_t)^3}{3(Z_{max} - Z_{min})^2} \quad \text{for } Z_{ave} \leq Peff_t \leq Z_{max} \quad (6.28)$$

$$Q_{s,t} = \frac{(Z_{max} - Z_{min})}{2} \quad \text{for } Peff_t = Z_{max} \quad (6.29)$$

$$Q_{s,t} = Peff_t - Z_{ave} \quad \text{for } Peff_t \geq Z_{max}. \quad (6.30)$$

As the amount of rainfall increases from  $Z_{min}$  to  $Z_{max}$ , an increasing proportion of rainfall will form surface runoff. Any rainfall in excess of  $Z_{max}$  will not be absorbed and will form surface runoff.

### Evaporation from the subsurface store

When the subsurface storage has the maximum amount of water ( $S_{cap}$ ), the actual evaporation rate ( $E_t$ ) equals the potential evaporation rate ( $E_{pot,t}$ ). But when the subsurface storage declines below this maximum amount, the actual evaporation rate declines from the potential rate. The model assumes two possible relationships exist between actual evaporation rates and the amount of subsurface storage. The first is the linear relationship

$$E_t = \frac{S_t}{S_{cap}} E_{pot,t}. \quad (6.31)$$

Eqn (6.31) assumes that evaporation from the subsurface storage will continue until  $S_t = 0$ . Such conditions are likely to be approximated in areas with deep rooting plants. Parameter  $S_{cap}$  in Eqn (6.31) is similar to parameter  $b$  in Eqn (6.7) of the *abcd* model, and these two parameters are expected to be correlated.

Alternatively the rate of decline of actual evaporation rate from the potential rate ( $E_{pot,t} - E_t$ ) is linearly related to the moisture deficit in the subsurface store ( $S_{cap} - S_t$ ). It is assumed that evaporation will cease when  $S_t = S_0$ . If  $0 < S_0 < S_{cap}$ , then

$$E_t = E_{pot,t} \frac{(S_t - S_0)}{(S_{cap} - S_0)}. \quad (6.32)$$

In this particular case when the subsurface storage,  $S_t$ , decreases below a specified  $S_0$ , then evaporation ceases. Pitman introduced parameter  $R$  which varies from 0.0 – 1.0 and controls the relationship between ratio of actual evaporation to potential evaporation and subsurface storage. If  $R = 0.0$  evaporation ceases when  $S_t = 0.0$ , and therefore Eqn (6.31) is applicable. Increasing the value of  $R$  means that evaporation will cease at high moisture content ( $S_0$ ), which is applicable in areas with shallow rooting plants.

### Subsurface runoff

The following power curve was assumed to describe the rate at which water drains from subsurface storage to streams. Subsurface runoff,  $Q_{g,t}$ , is given by

$$Q_{g,t} = \Lambda (S_t - S_l)^{POW} \quad (6.33)$$

and

$$\Lambda = \frac{FT}{(S_{cap} - S_l)^{POW}} \quad (6.34)$$

where

$S_l$  = subsurface moisture content below which subsurface runoff ceases

$FT$  = rate of drainage when subsurface storage is at its maximum,  $S_{cap}$ .

$POW$  = power of the curve.

It is assumed that when  $Q_{g,t} \leq G_w$ , then all the water is coming from the lower zone of the subsurface storage. Water coming from the lower zone of the subsurface storage ( $\leq G_w$ ) will take longer time to reach the catchment outlet than ( $Q_{g,t} - G_w$ ) originating from the upper zone. Two parameters,  $TL$  and  $GL$ , are introduced in order to incorporate the different time lags that are applicable

to  $(Q_{g,t} - G_w)$  and  $(\leq G_w)$  respectively. The Muskingum routing procedure is used to effect this lagging.

The Pitman model has 12 model parameters and these are presented in Table 6.1 below. The most important parameters are  $S_{cap}$ ,  $FT$ ,  $Z_{min}$ ,  $Z_{max}$ , and  $POW$ .  $S_{cap}$  and  $FT$  in particular have a major effect on the water balance.  $S_{cap}$  varies from 100 to 500 mm in South Africa, and geology has a strong influence on its values. An increase in  $S_{cap}$  decreases  $\bar{Q}_{yr}$ .  $FT$  was found to have values in the 0 to 30 range. A value of  $S_l = 0.0$  was found to be applicable for most catchments in South Africa.  $Z_{min}$  varied from 0 – 110 mm month<sup>-1</sup> in South Africa and reduces  $\bar{Q}_{yr}$ , and increases the standard deviation of flows.  $Z_{max}$  varied from 280 – 1100 mm, and was not significant in humid catchments in South Africa. This parameter reduces  $\bar{Q}_{yr}$ .

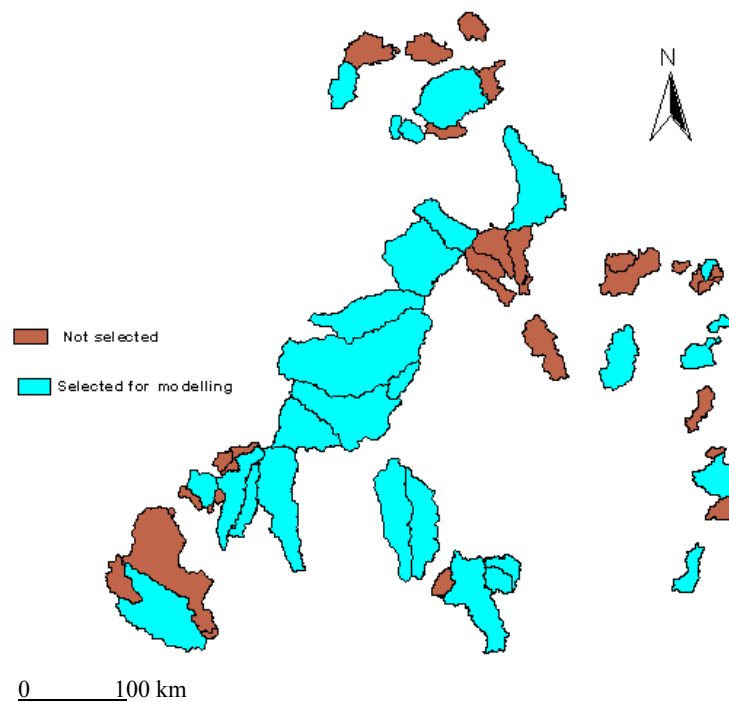
**Table 6.1:** Parameters of the Pitman model. *Calibrate* means a parameter requiring calibration, while *Estimate* means the value for this parameter is estimated from available data of a catchment.

PARAMETER	Value	Description
$POW$	<i>Calibrate</i>	Power of the soil moisture-runoff relationship
$S_l$	0.0	Subsurface moisture content at which subsurface runoff ceases (mm)
$S_{cap}$	<i>Calibrate</i>	Maximum water content of the subsurface store (mm)
$FT$	<i>Calibrate</i>	Rate of subsurface runoff drained when the subsurface storage is at its maximum capacity (mm month <sup>-1</sup> )
$G_w$	<i>Estimate</i>	Maximum rate of groundwater flow (mm month <sup>-1</sup> )
$AI$	<i>Estimate</i>	Proportion of the catchment which is impervious
$Z_{min}$	<i>Calibrate</i>	Minimum catchment absorption rate (mm month <sup>-1</sup> )
$Z_{max}$	<i>Calibrate</i>	Maximum catchment absorption rate (mm month <sup>-1</sup> )
$I_{cap}$	<i>Estimate</i>	Interception storage capacity (mm/day)
$TL$	0.25	Time lag of $Q_{g,t} - G_w$ (months)
$GL$	<i>Estimate</i>	Time lag of subsurface runoff from the lower zone ( $\leq GW$ ) (months)
$R$	0.5	Defines the relationship between evaporation and soil moisture content

$POW$  affects both the seasonal distribution of flows and  $\bar{Q}_{yr}$ . Catchments with perennial rivers have low values of  $POW \approx 2.0$ , while high  $POW \approx 3.0$  on those catchments which tend to dry up during the dry season. Parameters  $G_w$ ,  $GL$ , and  $TL$  only affect the seasonal distribution of monthly flows.  $G_w$  and  $GL$  can initially be set to zero, and  $TL = 0.25$ . Parameter  $R$  has values in the  $0 \leq R \leq 1.0$  range, and determines the rate at which actual evaporation decreases as the soil moisture storage declines below  $S_{cap}$ . An  $R = 0.5$  was found to be appropriate for the summer region of South Africa, and this value is assumed to be valid for catchments considered in this study.

### 6.3 Results and discussion

There are 30 catchments that fulfilled the criteria for selection for the purposes of rainfall-runoff modelling (Figure 6.3).

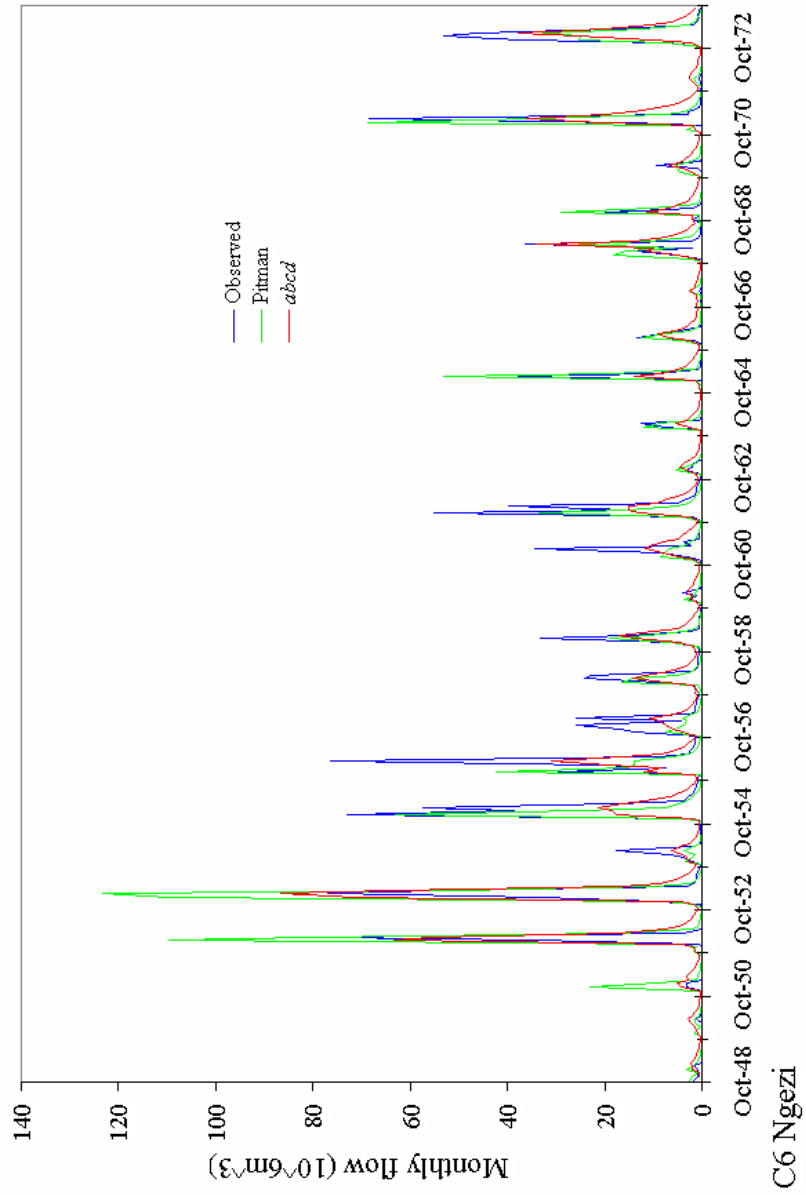


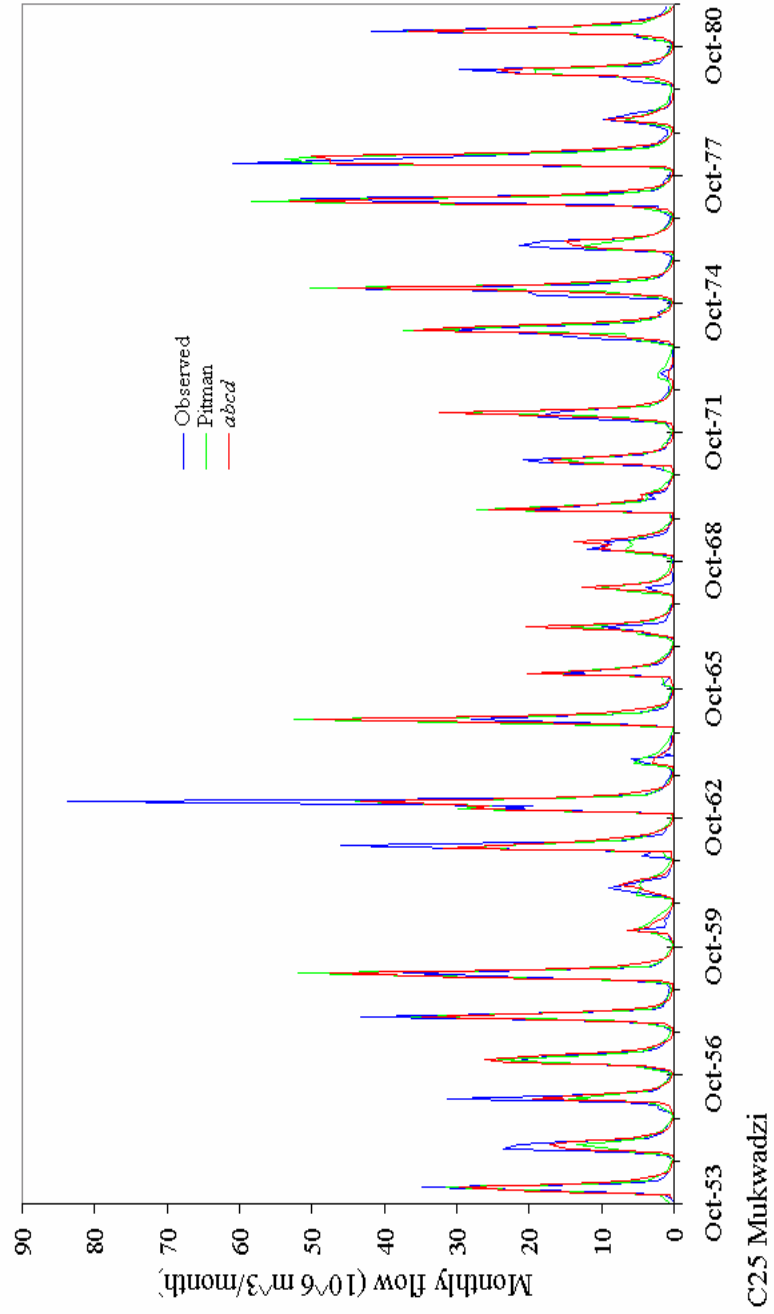
**Figure 6.3:** Catchment selected for rainfall-runoff modelling

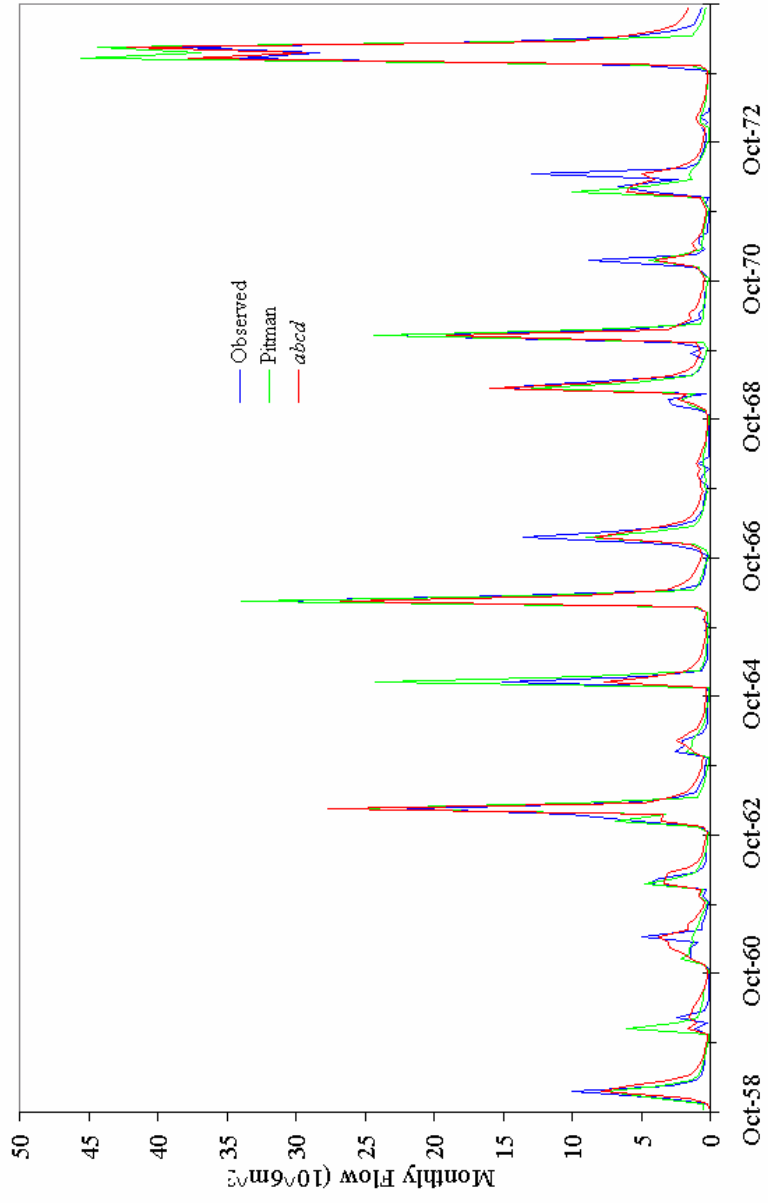
### 6.3.1 Comparison of simulated and observed monthly flows

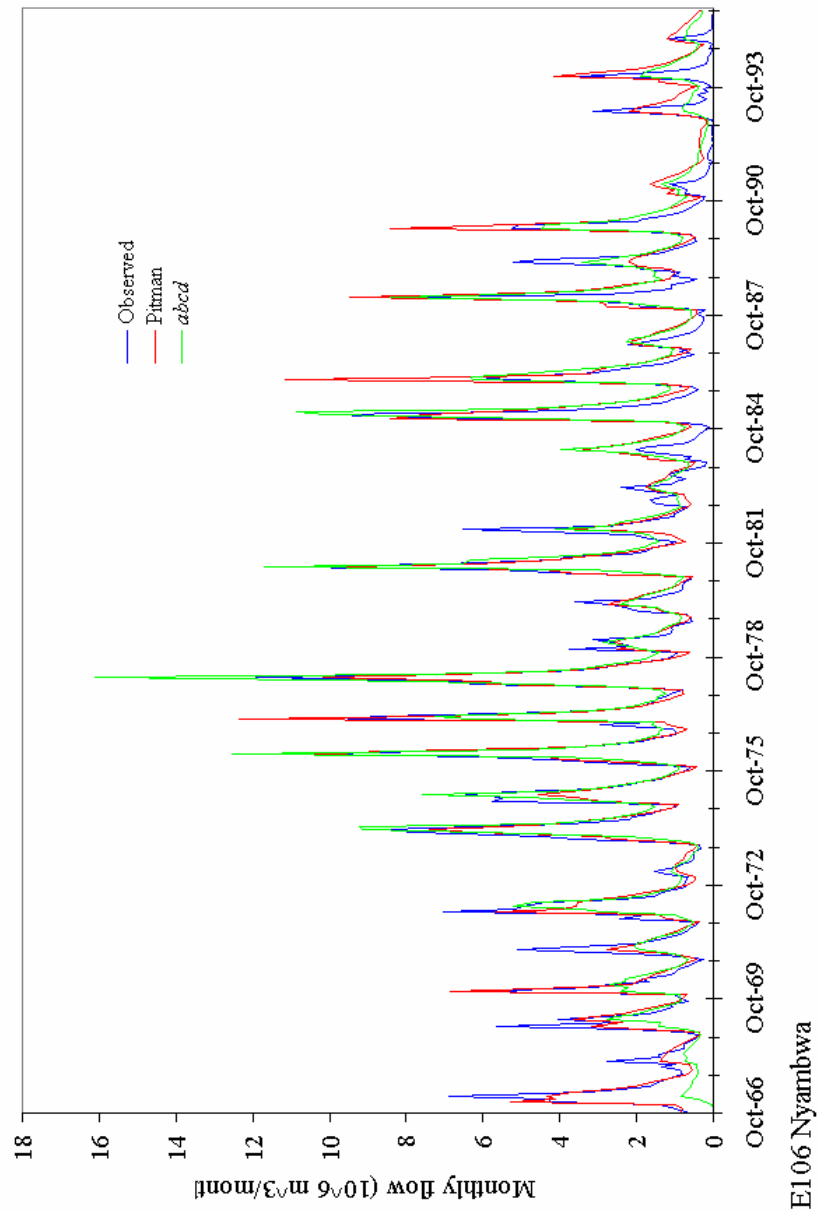
The *abcd* and Pitman models preserved the seasonal and inter-annual variations of flows (Figure 6.4). There are few peak flows that are not correctly simulated for some years by both models. Most of the rainfall in Zimbabwe occurs in the form of thunderstorms, and as a result the total monthly rainfall can be due to a few storms within a month that are not captured by sparse rain gauge networks. Hence, peak flows are not always modelled correctly due to inaccurate estimation of catchment rainfall. Both models simulated accurately relatively high dry season flows during the generally wet 1973-81, but low dry season flows for the 1983-91 period were overestimated on some of the catchments (Figure 6.4). It was observed during calibration of both models on some catchments, that if calibrated parameters simulated accurately dry season flows during the 1983-91 period, then the 1973-81 dry season flows are underestimated. Both models seem not to adequately represent the rise and decline in groundwater storage over several years with its subsequent effects on dry season flows.

Figure 6.5 shows close agreement between the means and standard deviations of simulated flows, and with those of observed flows. The goodness-of-fit criteria on the mean and standard deviation were satisfied on 76.7% and 86.7% of the catchments by the *abcd* and Pitman models respectively (Table 6.2). When the De Groen (2002) interception model (Eqn 6.22) is used in the Pitman model, instead of of Eqn (6.21), both the mean and standard deviation were preserved on 63.3% of the catchments. Thus, there were no differences between using Eqn (6.21) or Eqn (6.22) in preservation of flow statistics by simulated monthly flows.

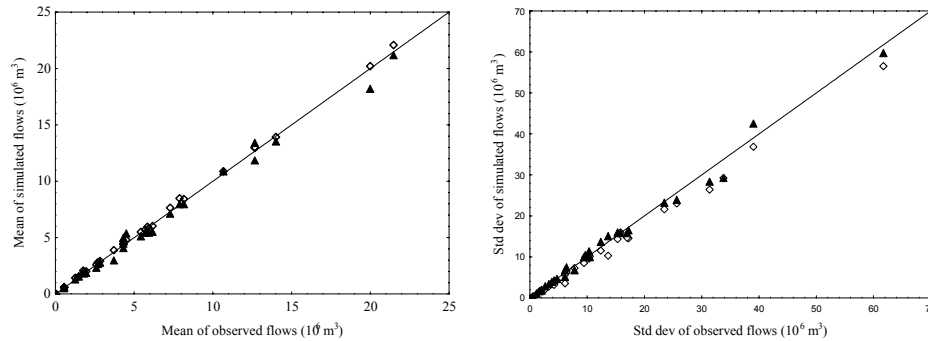








**Figure 6.4:** Comparison of observed monthly flows and those simulated using the Pitman and *abcd* models



**Figure 6.5:** Comparison of means and standard deviations of simulated monthly flows with those of observed monthly flows (Std dev = standard deviation,  $\diamond$  = *abcd* model,  $\blacktriangle$  = Pitman model)

**Table 6.2:** Preservation of mean and standard deviation of the monthly flows by the *abcd* and Pitman models – number (percentage) of catchments

<b><i>abcd</i> model</b>					
Differences in the Standard Deviation					
Differences in the	$\leq 5\%$	5 – 10%	10 – 15%	$>15\%$	Total
Mean					
<5%	4 (13.3%)	7 (23.3%)	7 (23.3%)	6 (20.0%)	24 (80.0%)
5 – 10%	1 (3.3%)	3 (10.0%)	1 (3.3%)		5 (16.7%)
>10%				1 (3.3%)	1 (3.3%)
<b>Total</b>	<b>5 (16.7%)</b>	<b>10 (33.3%)</b>	<b>8 (26.7%)</b>	<b>7 (23.3%)</b>	<b>30 (100.0%)</b>
<b>Pitman model</b>					
<5%	9 (30.0%)	13 (43.3%)	1 (3.3%)	3 (10.0%)	26 (86.7%)
5 – 10%	2 (6.7%)		1 (3.3%)		3 (10.0%)
>10%	1 (3.3%)				1 (3.3%)
<b>Total</b>	<b>12 (40.0%)</b>	<b>13 (43.3%)</b>	<b>2 (6.7%)</b>		<b>30 (100.0%)</b>

Each of the models satisfied the three goodness-of-fit criteria including the criterion that  $CI \geq 0.70$  on 70% of the catchments (Table 6.3). If the criteria relating to the mean and standard deviation are relaxed by  $\pm 2\%$ , then these are satisfied by 80% and 90% of the catchments by the *abcd* model and Pitman model respectively. Transformed flows in Table 6.3 means the given statistics were estimated from logarithms of simulated and observed monthly flows, while untransformed means the statistics were estimated from monthly flows that have not been transformed in this manner.

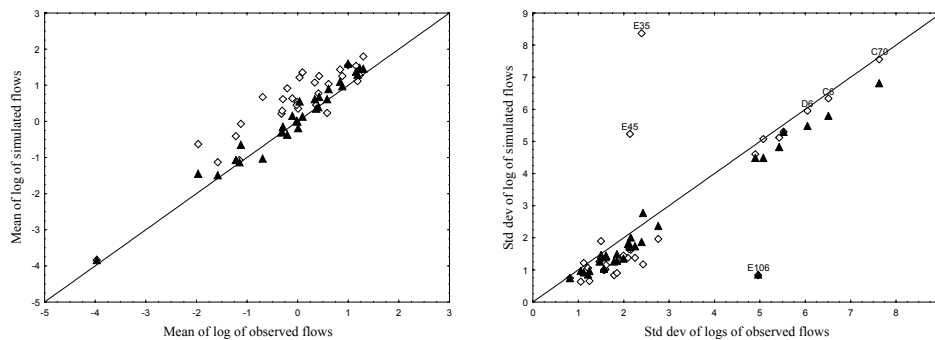
**Table 6.3:** Coefficient of efficiency for untransformed and transformed monthly flows

Catchment	Period	Untransformed Flows		Transformed Flows	
		<i>abcd</i>	Pitman	<i>abcd</i>	Pitman
C6	1948-73	0.62	0.40	0.92	0.60
C18	1962-84	0.84	0.81	0.60	0.80
C23	1956-88	0.80	0.72	0.56	0.68
C25	1954-82	0.82	0.81	0.65	0.86
C33	1956-65	0.82	0.70	0.66	0.76
C41	1956-88	0.79	0.65	0.91	0.69
C43	1959-85	0.60	0.60	0.60	0.60
C47	1956-78	0.55	0.52	0.87	0.70
C70	1969-83	0.88	0.85	0.98	0.73
D6	1949-70	0.77	0.75	0.98	0.77
D24	1958-70	0.77	0.75	0.69	0.72
D27	1962-95	0.75	0.78	0.57	0.50
D28	1962-95	0.74	0.69	0.24	0.56
E1	1925-45	0.52	0.60	0.40	0.57
E24	1955-95	0.86	0.85	0.23	0.32
E30	1957-97	0.73	0.69	0.83	0.50
E35	1958-75	0.81	0.72	-10.56	0.52
E37	1958-87	0.75	0.61	0.44	0.72
E40	1958-76	0.93	0.88	0.68	0.76
E42	1961-93	0.83	0.69	0.40	0.52
E45	1960-91	0.81	0.70	-4.75	0.61
E49	1960-95	0.89	0.75	0.48	0.54
E72	1961-87	0.75	0.74	0.99	0.75
E106	1966-95	0.77	0.82	0.07	0.07
E108	1966-90	0.86	0.80	0.53	0.43
E112	1966-86	0.76	0.67	0.62	0.66
E114	1967-86	0.73	0.61	0.61	0.53
E115	1967-85	0.84	0.76	0.72	0.68
E125	1970-80	0.83	0.74	0.81	0.52
E129	1970-98	0.69	0.70	0.65	0.67

This transformation makes dry season flows much more prominent, which enables an assessment of how well these have been simulated.

A comparison of the means of simulated and observed flows that have been transformed shows that the *abcd* model tended to overestimate low flows, while there is reasonable agreement for the Pitman model (Figure 6.6). This lack of fit for low flows by the *abcd* model seems to reflect that the structure of the model

does not adequately describe processes relevant to dry season flows. The variance of low flows was poorly preserved on very few catchments (Figure 6.6). The coefficient of efficiency for the transformed flows was greater than 0.70 on 30.0% and 33.% catchment for the *abcd* and Pitman models respectively. This again shows that both models have some weaknesses in describing dry season flows. The three outliers on the graph comparing standard deviations of logarithms of simulated monthly flows are E35, E45, and E106. The coefficient of efficiency for transformed flows in Table 6.3 was less than 0.07 for all these three catchments indicating inaccurate simulation of dry season flows. E35 has a sharp decline to zero flows while the model has a rather gradual decline to zero flows. Dry season flows for E45 were over-estimated during the 1964-74 and 1982-95 periods, while this was the case for the 1984-96 period on E106. In general the differences between transformed observed and simulated flows should not be over-emphasized since the logarithmic transformation exaggerates small differences.



**Figure 6.6:** Comparison of means of logarithms of simulated monthly flows with those of logarithms of observed monthly flows. Diamond = *abcd* model, shaded triangle = Pitman model

When the untransformed simulated monthly flows are considered, the *abcd* model satisfied the three goodness-of-fit criteria (mean, standard deviation, *CI*) on 70% of the catchments while this is 57% for the Pitman model.

### Model validation

The ability of calibrated model parameters to preserve flow statistics of a data set different from that used during calibration for the same catchment is presented in Table 6.4 and 6.5 for the *abcd* and Pitman models respectively.

**Table 6.4:** Split-sample test for the *abcd* model – comparison of observed and simulated monthly flows using differences between the mean and standard deviation, and coefficient of efficiency.

Catchment	Calibration			Validation		
	Percentage Difference			Percentage Difference		
	Mean	Standard Deviation	<i>CI</i>	Mean	Standard Deviation	<i>CI</i>
C25	-3.8	-9.7	0.79	1.4	6.8	0.85
C18	-0.2	-2.1	0.80	8.0	10.3	0.81
C23	-5.1	3.8	0.73	-58.3	-24.4	0.74
C41	-9.2	2.9	0.92	-17.3	4.4	0.66
D6	-13.0	-1.4	0.85	7.1	20.7	0.70
D27	-2.7	27.0	0.75	-54.6	-9.8	0.74
D28	5.0	11.0	0.64	-100.5	-56.6	-1.40
E24	-4.6	5.7	0.87	-27.6	6.0	0.84
E30	-4.9	24.1	0.78	-7.0	14.8	0.51
E42	-5.1	5.6	0.88	0.5	70.2	0.66
E45	-5.0	3.3	0.94	35.2	62.9	0.67
E49	-4.9	11.1	0.90	-0.3	33.8	0.84
E72	0.3	11.1	0.77	27.7	45.7	0.61
E106	-3.6	-1.5	0.90	-26.6	-2.6	0.70
E112	-5.1	9.9	0.88	32.3	47.5	0.53
E129	-5.0	3.2	0.67	-24.9	17.5	0.72

For both models calibrated model parameters failed to satisfy goodness-of-fit criteria on the validation data sets on almost all catchments. Similar results were obtained when a proxy-basin test was done. Model parameters which were used to simulate monthly flows of catchments different from those used during calibration, were only able to satisfy goodness-of-fit criteria on 3 out of 15 catchments for the *abcd* model, and 5 out of 15 catchments for the Pitman model. Catchments whose monthly flows could be simulated with model parameters calibrated at another catchment are shown below:

*abcd* model

C47 with C41 parameters  
D6 with D24 parameters  
E129 with E106 parameters

Pitman model

C23 with C70 parameters, E49 with E45 parameters  
E106 with E129 parameters, E112 with E49 parameters  
E129 with E106 parameters

**Table 6.5:** Spilt sample test for the Pitman model with model parameters calibrated on the first half of the data, and validation on the second half

Catchment	Calibration			Validation		
	Percentage Difference			Percentage Difference		
	Mean	Standard Deviation	<i>CI</i>	Mean	Standard Deviation	<i>CI</i>
C18	10.3	5.7	0.74	16.8	12.4	0.84
C23	-4.1	-24.4	0.51	-45.2	-48.4	0.50
C25	3.8	10.3	0.77	6.3	-4.0	0.84
C41	15.4	8.9	0.80	23.4	11.5	0.59
D27	4.9	0.2	0.78	-19.6	-9.9	0.73
D28	17.0	-12.2	0.70	2.3	9.2	0.63
D6	5.2	-10.9	0.70	18.7	19.4	0.55
E106	7.6	5.1	0.89	-26.6	-0.9	0.72
E112	-0.8	12.6	0.84	22.4	27.5	0.49
E129	3.8	-1.0	0.82	-7.5	5.3	0.62
E24	5.1	-14.8	0.82	15.5	14.0	0.89
E30	-8.9	5.8	0.68	-9.7	-0.4	0.69
E42	9.0	-7.9	0.77	25.9	49.0	0.51
E45	2.0	10.1	0.83	29.6	31.3	0.62
E49	15.0	9.0	0.84	12.8	9.5	0.76
E72	-16.6	-0.6	0.67	15.1	19.0	0.80

It is of interest to note that while model parameters of say catchment A can simulate monthly flows of catchment B, but parameters of catchment B are generally not valid for catchment A. This is only possible for E106 and E129 with the Pitman model parameter. This reinforces a conclusion made by Wolski (1999) that calibrated model parameters reflect a specific combination of physiographic conditions on a catchment. Thus extrapolating model parameters for purposes of simulating flows of ungauged catchments appears not possible for both models. This casts doubts on the feasibility of predicting values of model parameters from catchment characteristics.

### 6.3.2 Prediction of model parameters

#### 6.3.2.1 *abcd* model

Table 6.6 gives the values of calibrated parameters of the *abcd* model. The derivation of relationships between model parameters and catchment descriptors can be affected by equifinality of model parameters. Equifinality of model parameters means that different sets of model parameter values give similar results. Table 6.7 illustrates on E49 Popotekwe that four different sets of values of model parameters can result in similar outputs.

**Table 6.6:** Estimated values of parameters of the *abcd* model

Catchment	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
C18	0.9869	403.2	0.1208	0.8904
C23	0.9883	599.9	0.2727	0.7024
C25	0.9941	402.2	0.3545	0.5998
C33	0.9798	456.8	0.2934	0.7840
C41	0.9891	534.0	0.1100	0.0011
C43	0.9915	728.3	0.1778	0.0900
C47	0.9908	530.7	0.1516	0.0900
C6	0.9869	810.9	0.1816	0.0001
C70	0.9911	475.4	0.2126	0.5445
D24	0.9798	487.3	0.5178	0.5544
D27	0.9945	560.3	0.5255	0.1492
D28	0.9941	777.8	0.5203	0.0944
D6	0.9668	599.9	0.5029	0.9932
E1	0.9518	589.6	0.2747	0.2000
E106	0.9929	1708.9	0.2700	0.2000
E108	0.9445	394.7	0.2788	0.6671
E112	0.9720	358.2	0.3593	0.9755
E114	0.9638	504.7	0.3996	0.1638
E115	0.8814	297.7	0.4969	0.4327
E125	0.9738	450.3	0.5453	0.3032
E129	0.9983	682.6	0.7816	0.0910
E24	0.9887	510.8	0.1345	0.2439
E30	0.9914	564.2	0.0834	0.0161
E35	0.9867	366.6	0.1939	0.6277
E37	0.9962	1198.1	0.5148	0.0401
E40	0.9864	561.7	0.1605	0.0900
E42	0.9902	697.5	0.2446	0.0336
E45	0.9674	508.7	0.2950	0.9999
E49	0.9799	504.5	0.2807	0.8163
E72	0.9810	2371.9	0.0800	0.1376
Mean	0.9793	654.6	0.3112	0.3844
CV (%)	2.3	64.6	55.5	89.2

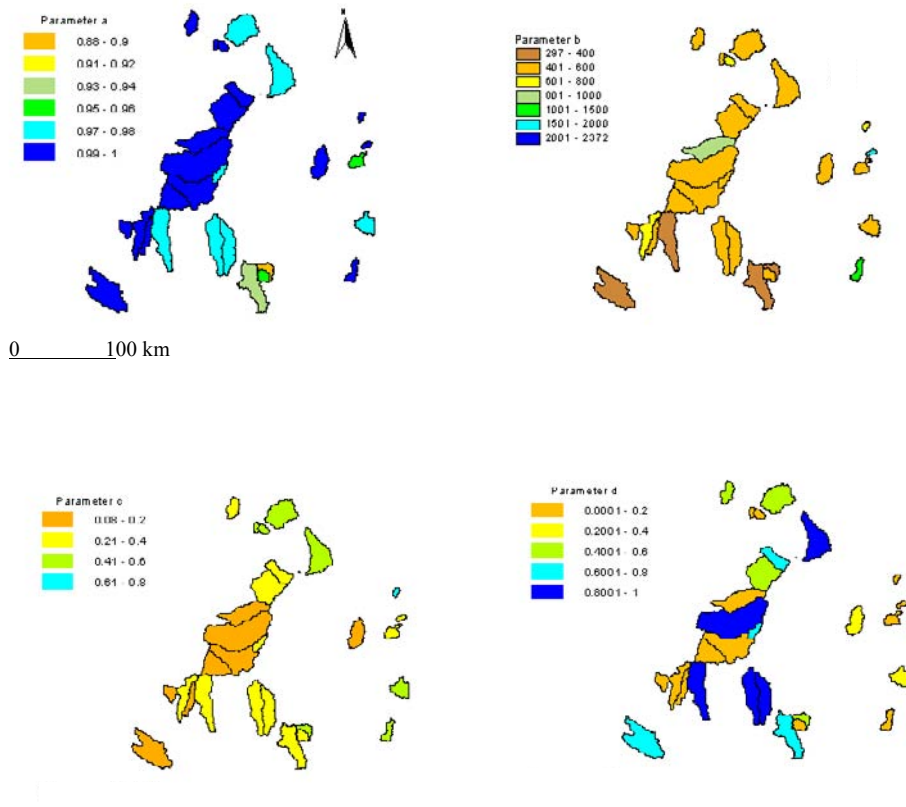
**Table 6.7:** Illustration of equifinality of model parameters on E49 with four sets of parameters resulting in simulated monthly flows with similar statistics.

Parameter	Set 1	Set 2	Set 3	Set 4
<i>a</i>	0.9799	0.9852	0.9780	0.9766
<i>b</i>	504.5	571.4	558.3	618.1
<i>c</i>	0.2807	0.2807	0.1566	0.0000
<i>d</i>	0.8163	0.0000	1.0000	1.0000
Mean ( $10^6\text{m}^3$ )	7.55	7.55	7.55	7.55
Std Dev ( $10^6\text{m}^3$ )	14.20	12.66	13.82	13.32
Coef Eff	0.89	0.87	0.89	0.89

In both Sets 1 and 2 parameter *c* was fixed equal to *BFI*, while in Set 3 this parameter was also calibrated. Alley (1984) noted that parameters *c* and *d* could have fixed values of  $c = 0.000$  and  $d = 1.000$ , while the other parameters are calibrated. This was done on E49 and the optimized values of parameters *a* and *b* are given under Set 4 (Table 6.7). A value of  $d = 0.0000$  in Set 2 results in no groundwater discharge to streams or base flows, but since groundwater recharge takes place, groundwater storage continuously increases. This is not feasible and therefore Set 2 is not appropriate. With  $d = 1.0000$  in Set 3, groundwater storage will always be none existent since all the water recharging groundwater will be discharged within the same month. This is rarely possible, and therefore Set 3 is inappropriate. A value of  $c = 0.000$  in Set 4 implies that there is no groundwater recharge, and therefore no base flow. If  $c = 0.000$ , then parameter *d* is no longer necessary as they will be no groundwater storage. Catchments without groundwater recharge and discharge are rare, and therefore Set 4 is inappropriate. Set 1 allows for recharge and storage of groundwater and depletion of part of the storage within a particular month, and therefore is appropriate. The results show that equifinality of model parameters exists in conceptual models (Liden and Harlin, 2000). Elimination of parameter values that result in state variables assuming unrealistic values is likely to lead to parameter values that can be related to catchment descriptors.

### Multiple regression

The range of values of parameter *a* is similar to that obtained in other studies (Alley, 1984; Vandewiele *et al.*, 1992; Fernandez *et al.*, 2000). This parameter is not highly variable among the 30 catchments and has a low *CV* of 2.3%. Catchments on the central and eastern parts of the country have  $a = 0.9800 - 0.9983$  (Figure 6.7).

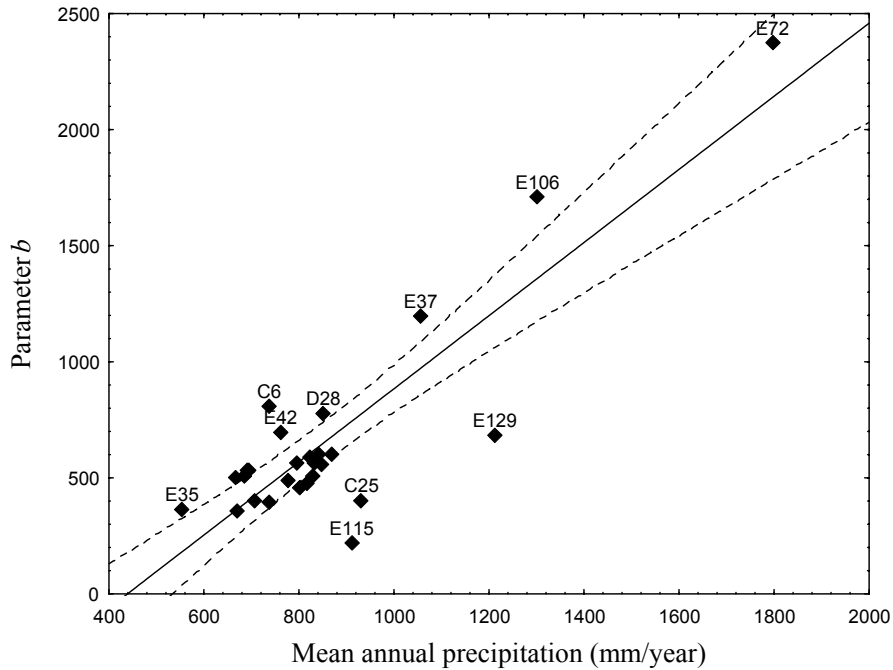


**Figure 6.7:** Spatial variation of the values of parameters of the *abcd* model

The only catchment with  $a < 0.9400$  is E115 with  $a = 0.8814$ . This parameter has a significant effect on  $\bar{Q}_{yr}$ , and a 5% increase in this parameter resulted in 50-70% decrease in  $\bar{Q}_{yr}$ , and a 10% decrease in the standard deviation. Parameter  $a$  is not related to flow statistics, and catchment characteristics. Fernandez *et al.* (2000) related parameter  $a$  to permeability which is influenced by lithology. However, in this study parameter  $a$  is not related to the proportions of a catchment with various lithologies. It is likely that this parameter is affected by several factors with none being dominant, since climatological effects (evaporation) and those effects due to soil-water relationships on runoff formation are reflected in this parameter.

Parameter  $b$  has no relationship with flow statistics, and no distinguishable spatial distribution (Figure 6.7). There is a tendency for catchments on the

Eastern Highlands to have high values of parameter  $b$ . This parameter has a linear relationship with  $\bar{P}_{yr}$  (Figure 6.8).



**Figure 6.8:** Relationship between parameter  $b$  of the  $abcd$  model and mean annual precipitation

The following predictive equation for  $b$  was derived for all the catchments where the coefficient of efficiency  $\geq 0.70$

$$b = 1.744\bar{P}_{yr} - 826.244 \quad (6.35)$$

with  $r^2 = 0.86$ . The positive relationship between  $b$  and  $\bar{P}_{yr}$  reflects that catchments with high rainfall have also high actual evaporation ( $\bar{E}_{yr}$ ). Parameter  $b$  was also found to have a positive linear relationship with  $\bar{E}_{yr}$ , and can be predicted using the following equation

$$b = 3914.16 - 11.483\bar{E}_{yr} + 0.0094\bar{E}_{yr}^2 \quad r^2 = 0.92 \quad (6.36)$$

Parameter  $c$  was set equal to  $BFI$ , and this has a range of values of 0.0800 – 0.7816. It has no effect on the mean of simulated monthly flows, but affects the seasonal distribution flows. It has already been established in Chapter 3 that  $BFI$

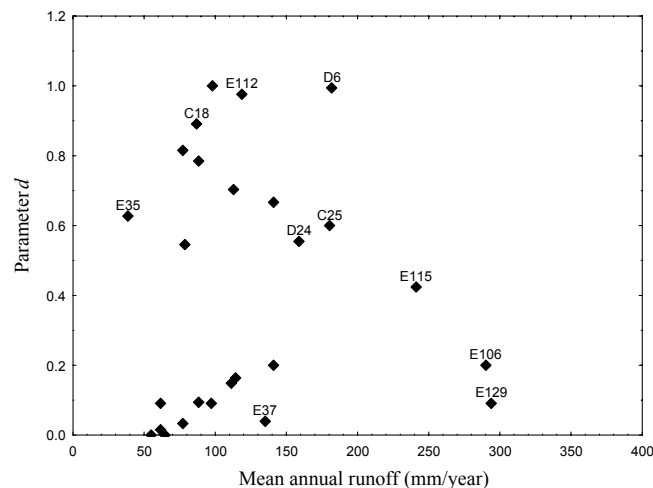
is related to  $\bar{Q}_{yr}$ , recession constant, and flows with specified exceedance probabilities and can be estimated by the following equation

$$c = 0.0003\bar{P}_{yr} + 0.072S_{10} - 0.003GL_{KL} - 0.001LC_{CG} \quad (6.37)$$

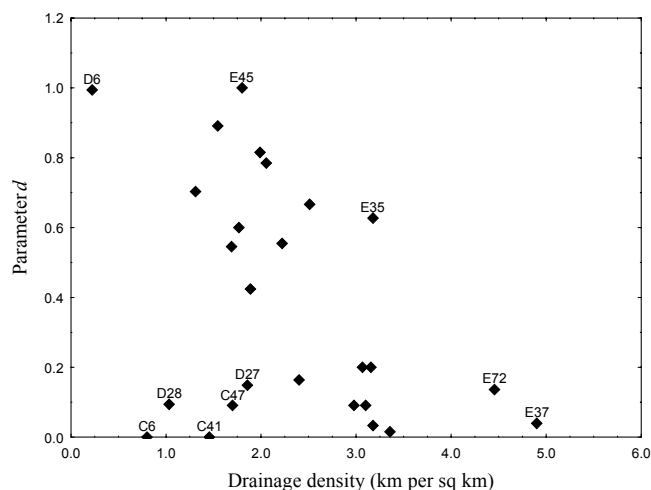
with  $r^2 = 0.92$ . Parameter  $d$  has no relationship with flow statistics. There is some grouping of catchments when this parameter is plotted against  $\bar{Q}_{yr}$  (Figure 6.9).

These patterns suggest that catchments with high  $\bar{Q}_{yr}$  have small values of parameter  $d$ . Small values of parameter  $d$  indicate that groundwater storage on catchments with high  $\bar{Q}_{yr}$  is depleted gradually. Catchments with low  $\bar{Q}_{yr}$  dry up soon after the rainy season, and therefore have high  $d$  values showing the rapid depletion of groundwater storage. This parameter has no clear spatial distribution, although there is some tendency for catchments on the Eastern Highlands to have low values (Figure 6.7).

Parameter  $d$  was found to have a weak negative relationship with drainage density,  $Dd$ , (Figure 6.10). Thus catchments with high  $Dd$  and therefore low permeability have low values of parameter  $d$  indicating slow rates of groundwater discharge.



**Figure 6.9:** Relationship between parameter  $d$  of the  $abcd$  model with mean annual runoff



**Figure 6.10:** Relationship between parameter  $d$  of the  $abcd$  model with drainage density

The following equation relating parameter  $d$  to  $Dd$  was derived

$$d = 1.0493 - 0.2379Dd \quad r^2 = 0.53 \quad (6.38)$$

Although some of Eqns 6.35 to 6.38 show weak relationships, they demonstrate that there is some potential for regionalising three out of the four model parameters.

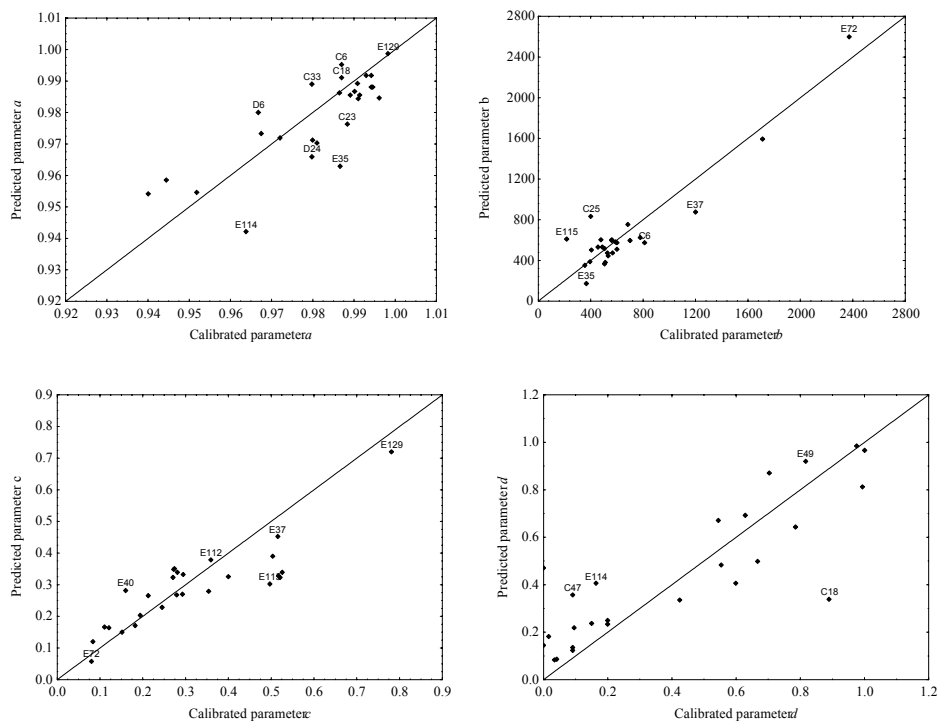
### Neural networks

The potential of predicting parameters of the  $abcd$  model using neural networks was investigated. Neural networks were configured with the four parameters as outputs. The best prediction was produced by an MLP6-7-4 neural network that has  $S_{34}$ ,  $\bar{P}_{yr}$ ,  $S_{10}$ ,  $\bar{E}_{yr}$ ,  $GLGG$ , and  $LC_{CU}$  as inputs. The coefficient of determination between the calibrated parameters and those predicted by the neural network are given in Table 6.8 below.

**Table 6.8:** Coefficient of determination for the prediction of  $abcd$  model parameters using multiple regression and an MLP6-7-4 neural network

Parameter	Linear Regression	MLP 6-7-4
a	0.00	0.62
b	0.92	0.86
c	0.92	0.74
d	0.53	0.66

When this neural network is used to predict parameters of the *abcd* model, all the other parameters except parameter *a* are estimated with acceptable accuracy (Figure 6.11). Several calibrations (training) of this neural network were done with random selection of catchments constituting test or validation sub-samples. When training of the neural network was completed, the predicted values of model parameters for catchments comprising the validation set, were noted. This was repeated with different sets of catchments comprising the validation set. Monthly flows were simulated using values of model parameters predicted by the neural network on 14 catchments that constituted validation sub-samples during these calibrations. The goodness-of-fit criteria were satisfied by only one catchment (E108) out of the 14 catchments. The differences between means and standard deviations of simulated flow and those of observed flows were in the +/- 20 to 50%. The coefficient of efficiency was greater than 0.70 on two catchments only. The *abcd* model is very sensitive to values of its parameters, and hence there is very limited potential to successfully simulate monthly flows using parameter values predicted from catchment characteristics. Alternatively, a large sample is required for training neural networks.



**Figure 6.11:** Comparison of values of *abcd* model parameters predicted from catchment characteristics by an MLP 6-7-4 neural network, with those values estimated from model calibration.

## Cluster analysis

An assessment of whether parameter values of the *abcd* model differed between clusters based on catchment descriptors identified in Chapter 5 was undertaken. Table 6.9 below shows the average values of the parameters within each of the clusters.

**Table 6.9:** Range of values of *abcd* model parameters in clusters

Cluster	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
1	0.978	558.4	0.265	0.784
2	0.982	458.0	0.216	1.000
3	0.978	511.7	0.480	0.993
4	0.996	1195.8	0.526	0.200
5	0.981	2371.9	0.080	0.138

Clusters 4 and 5 had less than five catchments with calibrated model parameters. The test of the null hypothesis that there are no differences between clusters in the values of model parameters was therefore restricted to Cluster 1 to 3. This hypothesis was not rejected at the 5% significance level by the Kruskal-Wallis test for parameters *a*, *b* and *d* except for parameter *c* which is the *BFI*. It was established in Chapter 5 that flow characteristics differed significantly between clusters, and therefore this hypothesis was rejected for parameter *c*. The clusters do not have unique values for parameters of the *abcd* model, and cannot be used to regionalise this model.

### 6.3.2.2 Pitman model

Calibrated Pitman model parameters are presented in Table 6.10 below. The most important parameters for this model as has already been discussed are *POW*, *S<sub>cap</sub>*, *FT*, *Z<sub>min</sub>*, and *Z<sub>max</sub>*. *I<sub>cap</sub>* was assumed to be equal to 1.5 mm for all those parts of the catchments that are under woodland, bushland, wooded grasslands, grasslands, and cultivation. An *I<sub>cap</sub>*=4.0 mm was assumed for parts of the catchments that are under forest plantations. *G<sub>w</sub>* and *GL* were all set to zero.

The relationships between model parameters, and a) flow statistics, and b) catchment descriptors are likely to be affected by equifinality. For example, *POW* and *FT* are closely related in their effects on simulated flows. If other parameters are held constant an increase in *POW* will cause an increased rate of subsurface runoff during the wet season. This will rapidly deplete this storage resulting in reduced dry season flows. The same effect can be achieved by decreasing *FT*. Table 6.11 shows that two different sets of values of model parameters can produce the same outputs on C25.

**Table 6.10:** Calibrated Parameters for the Pitman model

NAME	PERIOD	<i>POW</i>	$S_{cap}$	<i>FT</i>	<i>ZMIN</i>	<i>ZMAX</i>
C18	1963 - 83	3.6	217.6	9.2	60.9	1025.5
C23	1956 - 88	3.3	465.0	8.9	31.3	800.0
C25	1954 - 79	2.1	346.6	21.4	22.9	1151.0
C33	1954 - 65	3.3	261.5	20.4	16.1	1225.2
C41	1956 - 93	3.3	217.0	8.0	41.3	913.2
C43	1959 - 84	2.0	567.2	5.6	29.1	970.9
C47	1956 - 93	3.5	290.0	9.0	46.6	933.9
C6	1948 - 73	2.8	500.0	6.3	10.7	1099.7
C70	1969 - 82	3.2	380.0	6.4	79.3	1156.3
D24	1958 - 69	3.4	498.0	47.2	21.6	1024.2
D27	1962 - 95	3.0	689.9	23.3	43.7	990.8
D28	1962 - 95	2.8	770.0	7.4	46.6	1300.9
D6	1949 - 69	3.1	355.0	74.0	24.0	1128.3
E1	1925 - 45	1.5	240.0	24.1	20.0	900.0
E106	1966 - 95	2.4	1121.8	77.1	41.2	1174.0
E108	1966 - 89	2.6	268.4	44.9	21.3	786.2
E112	1966 - 85	3.0	210.0	17.7	40.2	766.4
E114	1967 - 85	1.5	260.0	20.0	31.3	918.1
E115	1968 - 84	1.4	146.4	87.1	24.9	1044.5
E125	1970 - 81	1.4	268.7	37.4	57.6	921.0
E129	1970 - 97	2.7	1061.2	96.3	32.1	1107.1
E24	1955 - 95	3.1	350.0	3.8	39.7	908.9
E30	1956 - 95	3.1	394.9	5.2	26.1	1114.0
E35	1958 - 74	3.0	220.0	8.0	62.2	877.6
E37	1958 - 86	3.5	850.0	42.0	17.8	1804.0
E40	1958 - 75	1.9	384.3	6.7	68.1	981.2
E42	1961 - 93	2.8	559.5	10.6	45.9	1280.3
E45	1960 - 87	3.3	250.0	64.0	69.5	928.0
E49	1960 - 93	3.3	220.0	38.8	88.0	949.0
E72	1961 - 85	1.0	697.0	98.0	37.3	1163.7

**Table 6.11:** Illustration of equipfinality of Pitman model parameters on C25 with two sets of parameter values producing monthly flows with similar flow statistics

	Set 1	Set 2
<i>POW</i>	2.1	2.7
<i>S<sub>cap</sub></i>	346.6	343.3
<i>FT</i>	21.4	35.0
<i>ZMIN</i>	22.9	25.9
<i>ZMAX</i>	1151.0	1205.6
UNTRANSFORMED MONTHLY FLOWS		
% Difference in the mean	-3.3%	2.5%
% Difference in the standard deviation	-5.6%	-3.6%
Coefficient of efficiency	0.81	0.81
TRANSFORMED MONTHLY FLOWS		
% Difference in the mean	+13.7%	+4.3%
% Difference in the standard deviation	-4.5%	+7.5%
Coefficient of efficiency	0.86	0.82

The “% Difference” given above are those between the statistics of observed and those of simulated monthly flows. In Set 2 *POW* was increased which necessitated an increase in *FT*, and the two sets of parameters produce almost identical flows. The value of *S<sub>cap</sub>* can also be affected by the estimated monthly potential evaporation. If potential evaporation has been underestimated, the value of *S<sub>cap</sub>* will increase and thus reducing  $\bar{Q}_{yr}$ . These interactions of model parameters will affect relationships between these parameters, and flow statistics and catchment descriptors.

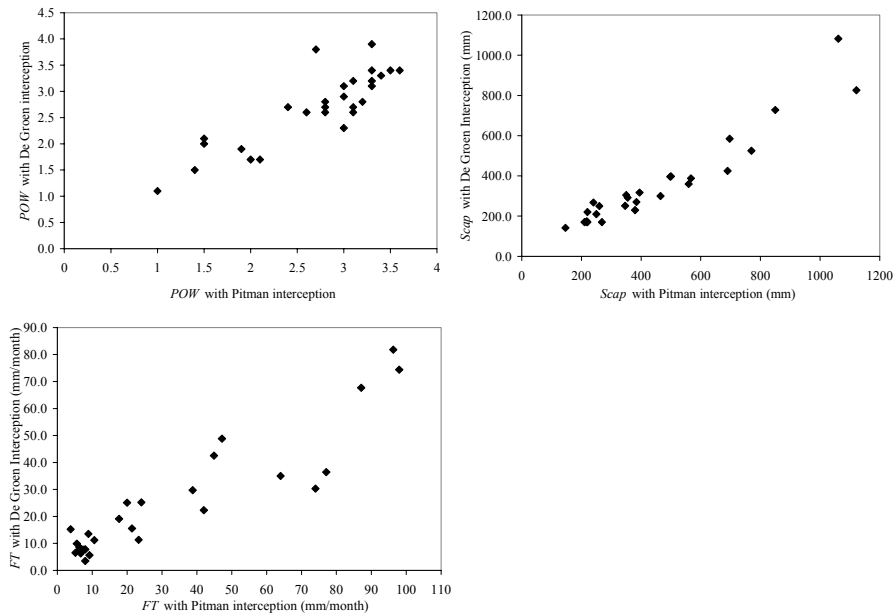
### Multiple regression

Table 6.12 presents correlation coefficients that are significant at the 5% level between model parameters, flow statistics, and catchment characteristics.

**Table 6.12:** Correlation coefficients between Pitman model parameters and, flow statistics and catchment characteristics

FLOW STATISTIC	<i>POW</i>	<i>S<sub>cap</sub></i>	<i>FT</i>
$\bar{Q}_{yr}$	-0.52	0.40	0.74
<i>BFI</i>	-0.40	0.60	0.77
One day recession coefficient		0.40	0.76
Runoff coefficient	-0.56		0.80
$\bar{N}_{DZ}$			-0.58
<i>q<sub>90</sub></i>	-0.38	0.61	0.73
<i>q<sub>20</sub></i>		0.54	0.71
$\bar{P}_{yr}$	-0.48	0.66	0.64
$\bar{E}_{yr}$		0.86	
$\bar{E}_{pot,yr}$		-0.58	-0.40
<i>S<sub>50</sub></i>	-0.66	0.39	0.58
<i>LC<sub>CG</sub></i>			-0.40
<i>LC<sub>CU</sub></i>		-0.41	

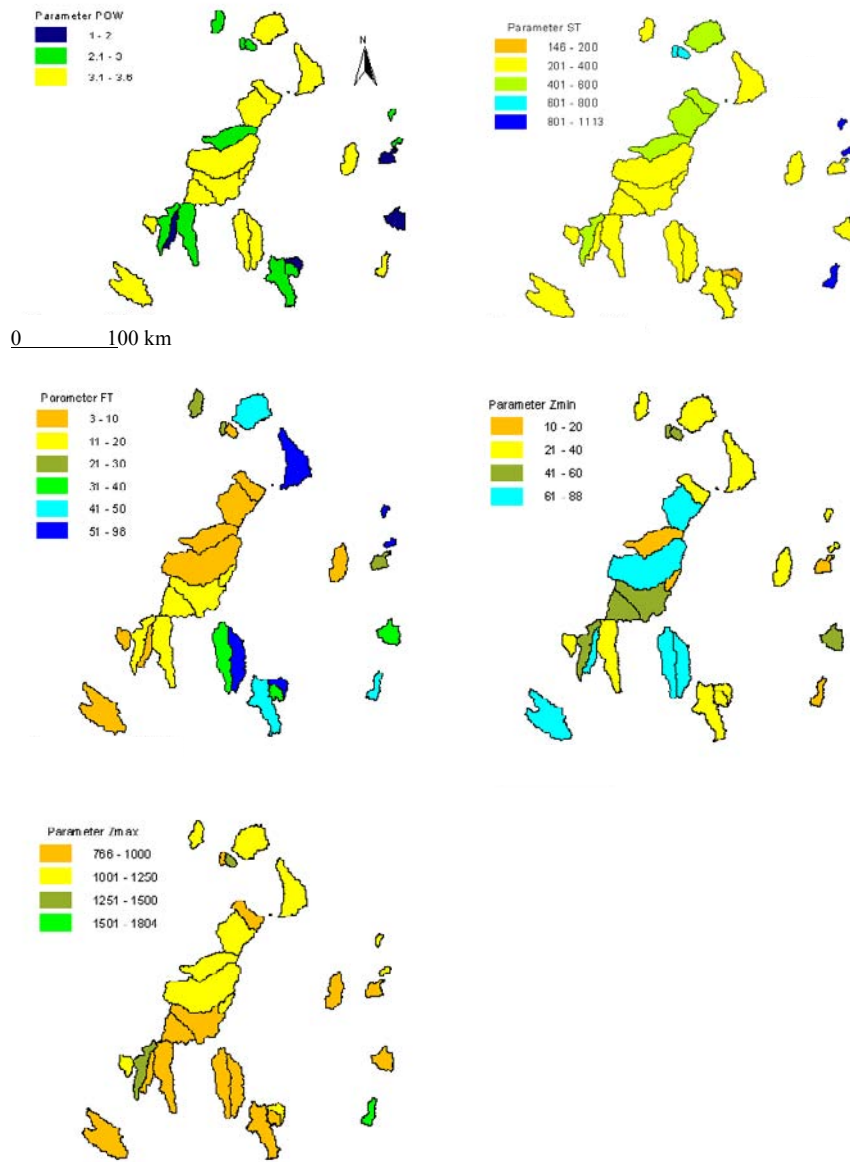
Most catchments have *POW* values close to 3.0 which is similar to the results of Hughes (1997) and Santa Clara (1980). Use of the De Groen (2002) interception model instead of the Pitman interception model did not cause systematic changes to this parameter (Figure 6.12). *POW* has a negative relationship with  $\bar{Q}_{yr}$  and therefore catchments on the Eastern Highlands with large  $\bar{Q}_{yr}$ , have low values of *POW*, 1.0 – 2.0 (Figure 6.13). It was established in Chapter 2 that *BFI* and the runoff coefficient are both positively related to  $\bar{Q}_{yr}$ , and hence both have a negative relationship with *POW*.



**Figure 6.12:** Comparison of values of Pitman model parameters calibrated when the De Groen (2002) interception equation is used and those obtained when the Pitman interception equation is used

High  $POW$  values cause the distribution of monthly flows to be peaky, while low values result in well distributed flows throughout the year. Thus catchments in areas with low rainfall have high values of  $POW$  since flows dry out during the dry season. Catchments with high rainfall that is well distributed throughout the year, and with perennial flow have low  $POW$  values.  $POW$  has a negative linear relationship with  $S_{50}$  (Table 6.12). Within the study area, catchments with perennial flows tend to be located in highland regions with relatively high  $S_{50}$ . The following predictive equation for  $POW$  was developed using the stepwise multiple regression technique.

$$POW = 3.419 - 0.139 S_{50} \quad r^2 = 0.58 \quad (6.39)$$



**Figure 6.13:** Spatial variation of calibrated Pitman model parameters

Most catchments particularly on the central part of the country have  $S_{cap}$  values in the 200 - 300 mm range (Figure 6.13).  $S_{cap}$  was found to be the most important parameter in determining the catchment water balance. Use of the De



the amount of rainfall received on a particular catchment. If  $\overline{E}_{yr}$  can be predicted from other variables then  $S_{cap}$  is given by

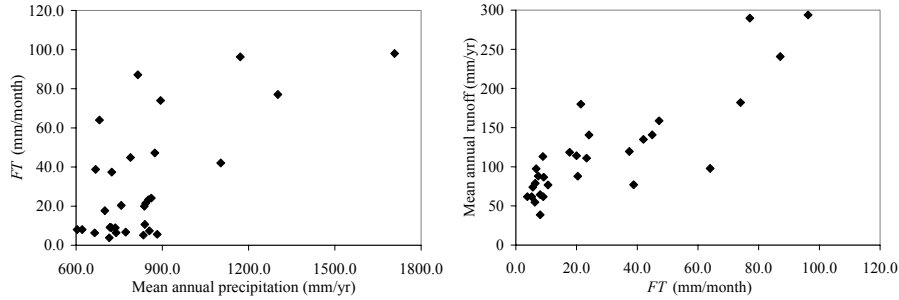
$$S_{cap} = 2.069 \overline{E}_{yr} - 993.11 \quad r^2 = 0.85 \quad (6.40)$$

It has to be emphasized that  $\overline{E}_{yr}$  is the dependent variable, and the above equation is useful for regionalisation purposes if  $\overline{E}_{yr}$  is predicted from other variables, or by remote sensing (Menenti, 2000).  $\overline{E}_{yr}$  used to develop Eqn (6.40) was derived from  $\overline{P}_{yr} - \overline{Q}_{yr}$ , and on ungauged catchment  $\overline{Q}_{yr}$  is unknown.

The maximum rate at which water is drained from saturated soils,  $FT$ , has a strong relationship with  $\overline{Q}_{yr}$ ,  $BFI$ , runoff coefficient,  $q_{90}$  and  $q_{20}$ . Very high  $FT$  values, 65.0 – 98.0 mm month<sup>-1</sup>, occur on catchments on the Eastern Highlands that are perennial and with high  $\overline{Q}_{yr}$  (Figure 6.13). These catchments have generally relatively steep slopes which result in high hydraulic gradients, and therefore high rates of drainage of subsurface water. The central part of the country with gentle slopes has low  $FT$  values, 4.0 – 10.0 mm month<sup>-1</sup>, e.g. C23, C70, C18. Gentle slopes result in rather low hydraulic gradients and therefore low rates of maximum drainage of subsurface water, i.e. low  $FT$  values. Catchments in the southern part of the study area such as E49, E45 and E108 have moderately high  $FT$  values, 30.0 – 65.0 mm month<sup>-1</sup>. For those catchments with high  $FT$  values, application of the De Groen (2002) interception models tends to decrease values of this parameter (Figure 6.12). There is a weak relationship between  $FT$  and catchment descriptors (Table 6.10). Figure 6.15 suggests that rainfall does to some degree affect the maximum rate of drainage from saturated soils.  $FT$  increases with annual rainfall, but this tends to a constant of about 80 – 100 mm month<sup>-1</sup> where mean annual rainfall is greater than 1000 mm yr<sup>-1</sup>.

$Z_{min}$  has values in the 10.0 – 88.0 mm month<sup>-1</sup> range and has no relationship with flow statistics. Monthly flow simulations were not very sensitive to this parameter. Catchments located on the central part with  $S_{50} \approx 1.70\%$  have  $Z_{min}$  values of 40 – 88 mm month<sup>-1</sup> (Figure 6.13). Catchments with  $S_{50} = 4 – 10\%$  have low  $Z_{min}$  values, 20 – 40 mm month<sup>-1</sup>, for example D6, D24, E108, E114, E115, E106 and E129. With such slopes and shallow soils, very low amounts of rainfall are required to initiate surface runoff, and therefore low  $Z_{min}$  values.  $Z_{min}$  is not related to any catchment descriptor within the study area. Pitman (1973)

suggested that  $Z_{min}$  increased with  $\overline{P}_{yr}$ , but there was no discernible relationship between these two variables on catchments used in this study.



**Figure 6.15:** Relationship between Pitman model parameter  $FT$  with a) mean annual runoff, and b) mean annual precipitation

$Z_{max}$  has values in the 766.0 – 1804.0 mm month<sup>-1</sup> range and has no relationship with flow statistics. Most of the catchments in the southern part of the study area have  $Z_{max}$  of 766 – 1000.0, while those in the northern part have values of 1000.0 – 1250.0 mm month<sup>-1</sup> (Figure 6.13).

### Neural networks

Neural networks were trained to predict all the five parameters,  $POW$ ,  $S_{cap}$ ,  $FT$ ,  $Z_{min}$ , and  $Z_{max}$ . The best prediction was made by an MLP 5-7-5 neural network that had  $S_{34}$ ,  $\overline{P}_{yr}$ ,  $S_{10}$ ,  $\overline{E}_{pot}$ ,  $GL_{GG}$  and  $LC_{CU}$  as inputs. The coefficients of determination for predicted parameters are given in Table 6.13.

The most important explanatory variables for these parameters are slope, precipitation and evaporation. Lithology and land cover type are the least important explanatory variables. This neural network has potential to estimate  $POW$  and  $S_{cap}$  with fair accuracy, but this is rather limited for  $FT$  and  $Z_{max}$  (Figure 6.16). However, neural networks offer a better potential for estimating Pitman model parameters from catchment descriptors than linear regression.

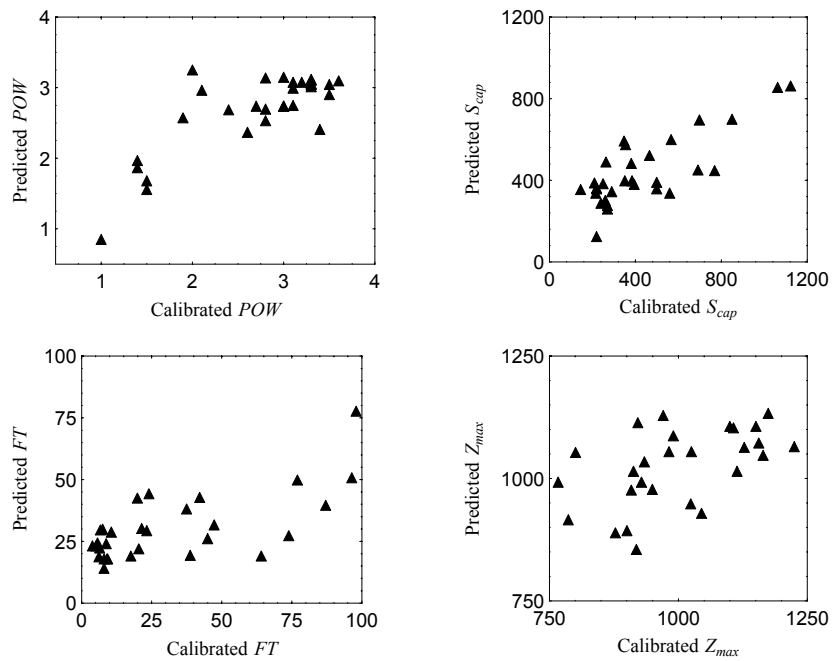
### Cluster analysis

The Kruskal-Wallis test was used to determine if the values of  $POW$ ,  $S_{cap}$ ,  $FT$ ,  $Z_{min}$  and  $Z_{max}$  differed between clusters derived using catchment descriptors in Chapter 5. Values of  $POW$ ,  $S_{cap}$  and  $FT$  differ significantly between clusters, but not those of  $Z_{min}$  and  $Z_{max}$ . These results indicate that clusters derived in Chapter 5 have values of  $POW$ ,  $S_{cap}$  and  $FT$  that are unique to each cluster. Cluster 2

which is relatively dry has high  $POW$  values.  $POW$  tends to decrease as the  $\bar{Q}_{yr}$  increases from Cluster 1, 3 to 5.  $S_{cap}$  increases with high  $\bar{Q}_{yr}$ , from Cluster 3 to 5.  $FT$  generally decreases from Cluster 1 to 5.

**Table 6.13:** Coefficient of determination for the prediction of Pitman model parameters using multiple regression and an MLP 6-7-4 neural network

Parameter	Linear Regression	MLP 6-7-4
$POW$	0.58	0.62
$S_{cap}$	0.85	0.64
$FT$	none	0.46
$Z_{min}$	none	0.00
$Z_{max}$	none	0.55



**Figure 6.16:** Comparison of parameters of the Pitman model predicted from catchment characteristics by an MLP 5-7-5 neural network, and those obtained through calibration

## 6.4 Conclusion

Monthly flows simulated with the *abcd* model and Pitman model satisfied all the three goodness-of-fit criteria on 70% and 57% of the catchments respectively. Lack of fit on some catchments was due to some peak flows that were not correctly simulated. Such flows are likely to have been caused by thunderstorms occurring on locations not covered by rain gauges, and therefore not reflected in the estimated catchment monthly rainfall. Another reason was that both models could not simultaneously simulate accurately relatively high dry season flows, arising from an increase in groundwater storage during a prolonged period of wet years, such as the 1973-1980 period, and low dry season flows due to reduced groundwater storage during a subsequent dry period such as the 1982 – 1989. This suggests a need for improving the representation of groundwater storage in both models. Equifinality of model parameters exists in both models, and this can affect the possibility of relating model parameters to catchment characteristics.

The proxy-basin test showed that extrapolation of parameters of both model was only possible on 3 out of 15 catchments for the *abcd* model, and 5 out of 15 catchments for the Pitman model. Thus use of parameters of any of these two models calibrated on one catchment to simulate flows of another ungauged catchment is generally not possible. Calibrated model parameters reflect a unique combination of physiographic characteristics on a particular catchment which are not likely to exist in a similar manner on another catchment.

Parameters *b*, *c*, and *d* of the *abcd* model have linear relationships with catchment characteristics, and can therefore be predicted by multiple regression. The coefficient of determination for derived multiple regression equations are 0.9 for both parameter *b* and *c*, while this is 0.5 for parameter *d*. An MLP 6-7-4 neural network can predict all the four parameters of the *abcd* model from catchment descriptors, and the coefficients of determination are in the 0.6 – 0.9 range. The *abcd* model is sensitive to parameters values, hence monthly flows simulated from parameters predicted by this neural network do not satisfy the goodness-of-fit criteria.

Parameter *POW* of the Pitman model can be predicted from median slope using a linear regression equation, and the coefficient of determination is 0.6. The subsurface storage capacity,  $S_{cap}$ , is related to mean annual actual evaporation, and the coefficient of determination for the predictive equation is 0.9. However, the relationship between  $S_{cap}$  and mean annual actual evaporation cannot be used to estimate  $S_{cap}$  for an ungauged catchment, since mean annual actual evaporation is unknown. Other important model parameters, *FT*,  $Z_{MIN}$ , and  $Z_{MAX}$  did not have linear relationships with catchment characteristics. An MLP 6-7-4

can predict  $POW$ ,  $S_{cap}$ ,  $FT$ , and  $Z_{MAX}$  with coefficients of determination of 0.5 – 0.6, but  $Z_{MIN}$  cannot be predicted.

All the parameters of the  $abcd$  model except parameter  $c$  do not have values that differed significantly between clusters derived using catchment characteristics. Thus clusters identified do not offer a basis for regionalising parameters of this model. With regard to the Pitman model,  $POW$ ,  $S_{cap}$ , and  $FT$  have values that differ significantly between clusters. However, it is not possible to develop predictive equations for these parameters for each of the clusters.

## 7 CONCLUSIONS AND RECOMMENDATIONS

### 7.1 Introduction

The goal of this study was to identify and assess the suitability of statistical methods and conceptual rainfall-runoff models to estimate flow characteristics of ungauged catchments. A case study approach was adopted using 52 catchments in Zimbabwe. Flow characteristics that were of interest are  $\bar{Q}_{yr}$ ,  $CV$ ,  $BFI$ , flow duration curve, mean monthly flows, and monthly flow time series. A central issue in this study was that methods developed in this study should be based on data that are readily available to practising hydrologists. Catchment characteristics that were considered for purposes of predicting the selected flow characteristics for ungauged catchments are  $\bar{P}_{yr}$ ,  $\bar{E}_{pot,yr}$ , elevation, slope, drainage density, lithology, land cover, and  $NDVI$ .

### 7.2 Prediction of flow characteristics using univariate methods

Mean annual flows are mainly dependent and positively related to mean annual precipitation and catchment slope. The proportion of a catchment underlain by granite and gneiss has a weak positive relationship with mean annual flows. Areas with granites tend to have impervious rock outcrops and thin soils that promote runoff formation, and hence the positive relationship. Kalahari sands and alluvial deposits have negative effects on mean annual flows due to deep percolation within these formations. Grasslands also have negative effects on mean annual flows. Most areas with grasslands have dambos which have higher evaporation rates than interfluves, resulting in a decrease in mean annual flows. Mean annual flows within the study area can be predicted from mean annual rainfall, slope, and the proportion of the catchment underlain by granite and gneiss using a linear regression equation. These variables particularly mean annual rainfall and slope are likely to be applicable to other parts of Zimbabwe. Use of neural networks does not improve prediction of mean annual flows.

Mean annual precipitation and slope are positively related to  $BFI$ , while evaporation has a negative effect. Kalahari sands, alluvial deposits, and grasslands have also negative effects on  $BFI$ . Areas with Kalahari sands and alluvial deposits tend to have deep water tables below river beds. Thus groundwater contribution to streams is reduced, and hence the negative relationship with  $BFI$ .  $BFI$  can be predicted using a linear regression equation from slope, and proportion of a catchment with wooded grasslands and grasslands. A neural network with mean annual precipitation and slope as inputs predicts  $BFI$  with comparable accuracy. Thus prediction of  $BFI$  using neural networks is recommended.

An exponential model describes flow duration curves of the study area. Parameters of this model can be predicted from *BFI*. Neural networks give the best prediction for flow duration curves. Inputs to this neural network are slope, mean annual potential evaporation, slope, mean annual rainfall, and drainage density.

A neural network with *BFI*, proportion of the catchment with wooded grasslands and grasslands, and slope as inputs can predict the hydrograph describing the proportion of mean annual flow occurring in each of the twelve months of the year. However, the onset of the rise and inflection point of this hydrograph are not accurately predicted. Mean monthly flows of an ungauged catchment can be predicted by distributing the predicted mean annual flow into mean monthly flows using this hydrograph.

One of the interesting findings of this study is that if the *BFI* of a catchment is known, then several flow characteristics of a catchment can be estimated, for example, recession constant, flow duration curve, and the distribution of mean annual flow into monthly flows. The study area is located in a region where rainfall occurs during one distinct season of approximately four months. During the rest of the year, river flows are sustained by contribution from subsurface storage. Flow characteristics during these eight months are therefore dependent on the hydrogeological properties of subsurface storages. *BFI* is a representative measure of these hydrogeological properties. At the beginning of the rainy season, the rate of recharge to subsurface storage partly affects the amount rainfall which will form runoff and causing the rise in the hydrograph. Hydrogeological properties of a catchment of which *BFI* seems to reflect affect the recharge rate. Further the rate of depletion of subsurface storage and its contribution to stream flows depends again on the properties of subsurface which *BFI* reflects. This study has therefore shown that the key to understanding flow regimes of regions with a distinct rainy season followed by a dry season is to understand factors influencing base flows.

This study has managed to identify catchment characteristics that explain variations of individual flow characteristics. The study has also developed methods for predicting these flow characteristics using catchment characteristics. Therefore, the first objective of the study has been fulfilled.

### **7.3 Identification of clusters of catchments with similar hydrological responses**

Redundancy analysis identified a set of catchment characteristics that explain variations of flow characteristics between catchments. These factors are mean annual precipitation, proportion of a catchment with the Umkondo rocks

(quartzite, shale, limestone and dolerite intrusion), proportion of a catchment with grasslands, and median slope. Redundancy analysis estimated the relative importance of these factors in explaining all the flow characteristics, and mean annual precipitation explains 59% of the variance of flow characteristics. The study area is located in a semi-arid to sub-humid area in which the occurrence of river flows greatly depends on the availability of precipitation. Hence mean annual precipitation alone explains such a high proportion of the variance of all flow characteristics. Redundancy analysis provided a basis for selecting those catchment characteristics, which explain flow characteristics, for use in classifying catchments into clusters with similar hydrological responses. In addition, estimation of the relative importance of these catchment characteristics enabled their weighting before cluster analysis.

Use of the  $R_g$  index and canonical variate analysis jointly provided an objective procedure for selecting the number of clusters. This procedure is recommended. Cluster analysis using weighted catchment characteristics and application of Andrews' curves to identify catchments that are outliers within clusters, resulted in delimitation of clusters that had distinct flow characteristics. In contrast, use of unweighted catchment characteristics resulted in clusters without distinct hydrological responses.

This study has shown that it is feasible to classify catchments using catchment characteristics into clusters with each cluster having similar hydrological responses. Catchments constituting a cluster do not necessarily form a contiguous region. A hydrologically homogenous region is therefore not necessarily continuous in space, but is made up of catchments with similar hydrological responses, and these catchments can be in different locations. Therefore, the second objective of the study has been fulfilled.

#### **7.4 Prediction of flow characteristics based on hydrological homogenous regions**

There are no identifiable relationships between catchment characteristics and flow characteristics for each of the derived clusters. Clustering partitioned catchments into sub-sets with narrow ranges in both their catchment and flow characteristics. Within these narrow ranges, it is possible that variations of flow characteristics cannot be explained at the catchment scale. A spatial resolution smaller than a catchment may be appropriate. Alternatively, catchment characteristics not included in this study may be appropriate. There is also the possibility that these unexplained variations are due to random behaviour of flow characteristics.

The third objective of this study has therefore been fulfilled as it was demonstrated that hydrologically homogenous clusters do not in all cases enable prediction of flow characteristics from catchment characteristics. However, if catchments clustered have wide ranges of both flow and catchment characteristics, there seems to be a possibility of predicting flow characteristics from catchment characteristics for clusters delimited.

### **7.5 Prediction of model parameters of conceptual models**

The *abcd* model and Pitman model were able to simulate accurately monthly flows of 77% and 87% of the catchments respectively. Accuracy of simulated flows was not affected by the size of the catchment for both models. Dry season flows for some years were not simulated accurately by both models. This indicates that components of both models describing the interactions between subsurface water and surface runoff require improvement. Parameters of both models that are related to subsurface moisture content can be predicted from catchment characteristics. The best prediction of parameters of the *abcd* model from catchment characteristics is made by a neural network with slope, mean annual precipitation, mean annual evaporation, proportion of the catchment with granites and gneiss, and that under cultivation as inputs. The predicted parameter values failed to simulate accurately monthly flows due to the high sensitivity of this model to parameter values. There is limited potential to predict all parameters of the Pitman model from catchment characteristics, except for *POW* and  $S_{cap}$ . The lack of relationships between some parameters of both models and catchment characteristics likely reflects that these parameters do not describe hydrological processes. This conclusion is valid for parameter *a* of the *abcd* model. Model parameters may describe hydrological processes, but errors in data used to calibrate these parameters can inhibit estimation of parameter values with physical relevance. Simulated monthly flows will agree with observed flows, but values of model parameters will have been estimated so as to fit errors in data. There is also a possibility that while model parameters are related to catchment characteristics used in this study, but values of these catchment characteristics were derived at scales at which their effects cannot be identified.

Values of the parameters of the *abcd* model did not differ significantly between clusters derived using catchment characteristics. Therefore, identification of hydrologically homogenous regions does not provide a basis for estimating parameters of this model for ungauged catchments. The most important three parameters of the Pitman model, *POW*,  $S_{cap}$  and *FT*, differed significantly between clusters, which indicates that there is a potential to estimate these parameters using such clusters. However, values of these parameters for each of the clusters could not be related to catchment characteristics. The conclusion to be made with regards to the fourth objective is that while some parameters of

lumped rainfall-runoff models used in this study can be estimated from catchment characteristics, it is not possible to simulate flows of ungauged catchments because other parameters still require to be estimated through calibration. Thus the fourth objective of the study has been fulfilled.

## **7.6 Comparison of performances of neural networks and multiple regression**

Neural networks could predict *BFI*, flow duration curves, and the distribution of mean annual flow into mean monthly flows. Multiple regression could only predict mean annual runoff and *BFI*. The majority of flow characteristics are non-linearly related to catchment characteristics, and therefore neural networks perform better than multiple regression in predicting flow characteristics.

Three out four parameters of the *abcd* model could be predicted from catchment characteristics using multiple regression, while a neural network was able to predict all the four parameters. With regards to the Pitman model, two out of five parameters could be predicted by multiple regression. A neural network managed to predict three parameters of this model. These results show that some model parameters are non-linearly related to catchment characteristics, and therefore neural networks perform better than multiple regression. The fifth objective of this study was therefore addressed, however further research on the use of neural networks to predict parameters of rainfall-runoff models from catchment characteristics is recommended.

## **7.7 Recommendations**

The use of methods developed in this study for estimating  $\bar{Q}_{yr}$ , *BFI*, and flow duration curves of ungauged catchments is recommended. Mean monthly flows of ungauged catchment can be estimated after predicting the hydrograph describing the distribution of mean annual flow into monthly flows. Use of this method is recommended.

Further studies aimed at improving some of the methods developed are recommended. In particular, the extension of this study to other parts of the country in order to identify if there are any additional factors that may improve the prediction of *BFI* as this is important in predicting other flow characteristics.

This study has demonstrated the feasibility of predicting a hydrograph describing the distribution of the mean annual flow into monthly flows at ungauged sites. However, the inflection point of this hydrograph was not accurately estimated. Further studies aimed at improving predicting this hydrograph from catchment characteristics are recommended.

River flows during the dry season which is about eight months greatly depend on hydrogeological properties of the subsurface storage. Prediction of these flows will improve if quantitative measures representing these hydrogeological properties at the catchment scale are available. Thus studies aimed at developing these measures are recommended.

Clusters with similar hydrological responses identified in this study had narrow ranges in the values of their catchment characteristics. Relationships between catchment and flow characteristics could not be determined. Further studies clustering catchments with wide ranges in the values of both catchment and flow characteristics are recommended in order to determine whether this approach will improve the prediction of flows of ungauged catchments.

Rainfall-runoff modelling was undertaken in this study using catchment rainfall estimated from point rainfall measurements. The study noted the possibility of errors in estimating catchment rainfall, which affect values of calibrated model parameters, and hence lack of relationship between these parameters and catchment characteristics. Further studies using rainfall and evaporation estimated from satellite imagery are recommended in order to explore if these data improve the relationship between model parameters and catchment characteristics.

## REFERENCES

- Acreman, M.C. and Sinclair, C.D. 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *Journal of Hydrology*, Vol 84, 365-380.
- Alley, W.M. 1986. On the treatment of evaporation, soil moisture accounting, and aquifer recharge in monthly water balance models, *Water Resources Research*, Vol. 20, No. 8, 1137-1149.
- Andrews, D.F. 1972. Plots of high dimensional data. *Biometrics*, Vol. 28, 25-136.
- Andrews, A.J. and Bullock, A. 1994. *Hydrological impact of afforestation in eastern Zimbabwe*. Overseas Development Report No. 94/5, Institute of Hydrology, Wallingford, UK.
- Ardo, J., Pilesjo, P. and Skidmore, A. 1997. Neural networks, multitemporal Landsat thematic mapper data and topographic data to classify forest damages in the Czech Republic. *Canadian Journal of Remote Sensing*, Vol. 23, No. 3, 217-228.
- Arnold, J.G., Allen, P.M., Muttiah, R., and Bernhardt, G. 1995. Automated base flow separation and recession analysis techniques. *Ground Water* Vol. 33, No. 6, 1010-1018.
- ASCE. 1996. *Hydrology handbook*. American Society of Civil Engineers. New York.
- Bals, J. 2002. *Classification of European estuaries*. Master's Thesis, TU Delft, The Netherlands.
- Bastiaanssen, W.G. 1998. *Remote sensing in water resources management: The state of the art*. International Water Management Institute, Colombo, Sri Lanka.
- Berger, K.P. and Entekhabi, D. 2001. Basin hydrologic response relations to distributed physiographic descriptors and climate. *Journal of Hydrology*, Vol. 247, 169-182.
- Beven, K. and O'Connell, P.E. 1982. *On the role of physically-based distributed modelling in hydrology*. Report No. 81, Institute of Hydrology, Wallingford, UK.

- Bonifacio, R. and Grimes, D.I.F. 1998. *Drought and flood warning in southern Africa*. IDNDR Flagship Programme - Forecasts and Warnings, UK National Coordination Committee for the IDNDR, Thomas Telford, London.
- Blackie, J.R. and Eeles, C.W.O. 1985. Lumped catchment models. In: Anderson, M.G. and Burt, T.P. (Ed) *Hydrological forecasting*. John Wiley and Sons, 311-345.
- Bosch, J.M. 1979. Treatment effects on annual and dry period streamflow at Cathedral Peak. *South African Forestry Journal*, Vol. 108, 29-38.
- Bosch, J.M. and Hewlett, J.D. 1982. A review of catchment experiments to determine the effect of vegetation changes on water yield and evaporation. *Journal of Hydrology*, Vol. 55, 3-23.
- Bullock, A., 1988. *Dambos and discharge in central Zimbabwe*. Unpublished PhD thesis, University of Southampton.
- Bullock, A., Chirwa, A.B., Matondo, J.I. and Mazvimavi, D. 1990. *Analysis of flow regimes in Malawi, Tanzania and Zimbabwe: A feasibility study for Africa FRIEND*. Overseas Development Report, Institute of Hydrology, Wallingford, UK.
- Bullock, A. and Andrews, A. 1997. Introduction to Southern Africa FRIEND. In: *Southern African FRIEND*. Technical Documents in Hydrology No 15, UNESCO, Paris.
- Bullock, A., Andrews, A. and Mngodo, R. 1997. Regional surface water resources and drought assessment. In: *Southern African FRIEND*. Technical Documents in Hydrology No 15, UNESCO, Paris.
- Burn, D.H. 1990a. An appraisal of the "region of influence" approach to flood frequency analysis. *Hydrological Sciences Journal*, Vol. 35, 149-165
- Burn, D.H. 1990b. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, Vol. 26, No.10, 2257-2265.
- Burn, D.H. and Boorman, D.B. 1993. Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology*, Vol. 143, 429-454
- Chapman, T. 1999. A comparison of algorithms for stream flow recession and base flow separation. *Hydrological Processes*, Vol. 13, 701-714.

- Chibanga, R., Berlamont, J., and Vandewalle, J. 2003. Modelling and forecasting of hydrological variables using artificial neural networks: the Kafue River sub-basin. *Hydrological Sciences Journal*, Vol. 48, 363-379.
- Chiew, F.H.S. and McMahon, T.A. 1990. Estimating groundwater recharge using a surface watershed modelling approach. *Journal of Hydrology*, Vol. 114, 285-304.
- Cigizoglu, H.K., 2003. Estimation, forecasting and extrapolation of river flows by artificial neural networks. *Hydrological Sciences Journal*, Vol. 48, 349-361.
- Cihlar, J., Laurent, L.St. and Dyer, J.A. 1991. Relation between the normalized difference vegetation index and ecological variables. *Remote Sensing Environment* Vol. 35, 279-298.
- De Groen, M. 2002. *Modelling interception and transpiration at monthly time steps: Introducing daily variability through Markov chains*. PhD Thesis, IHE Delft, The Netherlands.
- Dent, M.C., Lynch, S.D. and Tarboton, H. 1989. Detailed delimitation of rainfall regions in southern Africa. *Water SA*, Vol. 16, 1-4.
- Department of Meteorological Services. 1981. *Climate handbook of Zimbabwe*. Department of Meteorological Services Zimbabwe, Harare.
- Dolling, O.R. and Varas, E.A. 2003. Artificial neural networks for streamflow prediction. *Journal of Hydraulic Research*, Vol. 40, No. 5, 547-554.
- Drayton, R.S., Kidd, C.H.R., Mandeville, A.N. and Miller, J.B. 1980. *A regional analysis of river flows and low flows in Malawi*. Report No. 72, Institute of Hydrology, Wallingford, U.K.
- Edwards, K.A. and Blackie, J.R. 1981. Results of the East African catchment experiments 1958-1974. In: Lall, R. and Russel, E.W. (Eds) *Tropical agriculture hydrology*. Wiley, 163-188.
- Everitt, B.S. 1993. *Cluster analysis*. Edward Arnold, London.
- Faulkner, R.D. and Lambert, R.A. 1991. The effect of irrigation on dambo hydrology: A case study. *Journal of Hydrology*, Vol. 123, 147-161.
- Fernandez, W. Vogel, R.M. and Sankarasubramanian, A. 2000. Regional calibration of a watershed model. *Hydrological Sciences Journal*, Vol. 45, No. 5, 689-707.

- Freeze, R.A. and Cherry, J. A. 1979. *Groundwater hydrology*. Prentice-Hall, 604p.
- Fylstra, D., Lasdon, L., Watson, J. and Waren, A. 1998. Design and use of the Microsoft Excel Solver. *Interface Journal*, Vol. 28, No. 5, 28-55. Also available on <http://www.solver.com/academic.htm>.
- Gan, K.C., McMahon, T.A. and O'Neill, I.C. 1990. Errors in estimated streamflow parameters and storage for ungauged catchments. *Water Resources Bulletin*, Vol 26. No. 3, 443-450.
- Gordon, A.R. 1999. *Classification methods for the exploratory analysis of multivariate data*. Chapman and Hall, London.
- Green, I.R.A. and Stephenson, D. 1986. Criteria for comparison of single event models. *Hydrological Sciences Journal*, Vol. 31(3), 395-410.
- Gregory, K.J. and Walling, D.E. 1973. *Drainage basin form and process: A geomorphological approach*. Edward Arnold, London.
- Gupta, V.K. and Sorooshian, S. 1983. Uniqueness and observability of conceptual rainfall-runoff model parameters: The percolation process examined. *Water Resources Research*, Vol. 19, No. 1, 269-276.
- Gustard, A. 1983. Regional variability of soil characteristics for flood and low flow estimation. *Journal of Agricultural Water Management*, Vol. 6, 255-268.
- Gustard, A., Roald, L.A., Demuth, S., Lumadjeng, H.S. and Gross, R. 1989. *Flow Regimes from Experimental and Network Data (FRIEND), Volume 1 Hydrological Studies*. Institute of Hydrology, Wallingford, UK.
- Hall, M.J. and Minns, A.W. 1999. The classification of hydrologically homogenous regions. *Hydrological Sciences Journal*, Vol. 44, No. 8, 693-704.
- Hargreaves, G.L. and Hargreaves, G.H. 1985. Irrigation water requirements for Senegal River Basin. *Journal of Irrigation and Drainage Engineering*, Vol. 111, 265-275.
- Hargreaves, G.H. and Samani, Z.A. 1985. Reference crop evaporation from temperature. *Transactions of the American of Agricultural Engineers*, Vol. 1, 96-99.

- Hendriks, M.R. 1990. *Regionalisation of hydrological data: Effects of lithology and land use on storm runoff in east Luxembourg*. Netherlands Geographical Studies 114, Amsterdam/Utrecht.
- Hendricksen, B.L. and Durkin, J.W. 1986. Growing period and drought early warning in Africa using satellite data. *International Journal of Remote Sensing*, Vol 7. No. 11, 1583-1608.
- Hughes, D.A. 1985. Conceptual catchment model parameter transfer investigations in the Southern Cape. *WaterSA*, Vol. 11, No. 3, 149-156.
- Hughes, D. 1997. Rainfall-runoff modelling. In: *Southern African FRIEND*. Technical Documents in Hydrology, No. 15, UNESCO, Paris.
- Hughes, D.A. 1995. Monthly rainfall-runoff models applied to arid and semiarid catchments for water resource estimation process. *Hydrological Sciences Journal*, Vol. 40, No.6, 751-769.
- Hughes, D.A. 1999. Towards the incorporation of magnitude-frequency concepts into the building block methodology used for quantifying ecological flow requirements of South African rivers. *Water SA*, Vol. 25, No. 3, 279-284.
- Hughes, D.A. and Metzler, W. 1998. Assessment of three monthly rainfall-runoff models for estimating water resource yield of semiarid catchments in Namibia. *Hydrological Sciences Journal*, Vol. 43, No. 2, 283-297.
- Hughes, D. 2002. *The Pitman model*. Unpublished notes for calibrating the Pitman Model. Institute for Water Research, Rhodes University, Grahamstown, South Africa.
- Hughes, D. and Forsyth, D. 2002. *SPATSIM – Spatial and Time Series Information Modelling Software User's Guide*. Institute for Water Research, Rhodes University, Grahamstown, South Africa.
- Ibrahim, A.B. and Cordery, I. 1995. Estimation of recharge and runoff volumes from ungauged catchments in eastern Australia. *Hydrological Sciences Journal*, Vol. 40, No. 4, 499-515.
- IH, 1980. *Low flow studies*. Research Report 1, Institute of Hydrology, Wallingford, UK.
- Interconsult A/S. 1985. *National master plan for rural water supply and sanitation, Volume 2.2 Hydrogeology*. Ministry of Energy and Water Resources and Development, Republic of Zimbabwe.

- Jain, S.K., Storm, B., Bathurst, J.C., Refsgaard, J.C. and Singh, R.D. 1992. Application of the SHE to catchments in India Part 2. Field experiments and simulation studies with the SHE of the Kolar subcatchments of the Narmada River. *Journal of Hydrology*, Vol. 140, 25-47.
- Jordan, J.N. 1968. Ground-water in the Rhodesian basement complex. Geological Society of South Africa, Vol. (LXXXI), 103-111.
- Kent, M. and Coker, P. 1992. *Vegetation description and analysis: A practical approach*. Belhaven Press. London.
- Key, R.M., 1997. An introduction to the crystalline basement of Africa. In: Wright, E.P. and Burgess, W.G. (Ed.) *The hydrogeology of crystalline basement aquifers in Africa*. Geological Society Special Publication No. 66, London, 29-57.
- Kirkby, M.J. (Ed) 1978. *Hillslope hydrology*. John Wiley and Sons.
- Kirkby, M.J. 1985. Hillslope hydrology. In: Anderson, M.G. and Burt, T.P. (Ed) *Hydrological forecasting*. John Wiley and Sons, 37-74.
- Kite, G.W. 1988. *Frequency and risk analyses in hydrology*. Water Resources Publications, Fort Collins, Colorado.
- Klastorin, T.D. 1983. Assessing cluster analysis results. *Journal of Marketing*, Vol 20, 92-98.
- Klemes, V. 1986. Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, Vol. 31, No. 3, 13-24.
- Kweshu, D. 2000. *Gathering key information about indigenous forests of Zimbabwe*. Forestry Commission, Harare, Zimbabwe (unpublished).
- Lacey, G.C. and Grayson, R.B. 1998. Relating baseflow characteristics to catchment properties in south-eastern Australia. *Journal of Hydrology*, Vol 204, 231-250.
- Legendre, P. and Legendre, L. 1998. *Numerical ecology*. Developments in Environmental Modelling, 20. Elsevier, Amsterdam.
- Liden, R. and Harlin, J. 2000. Analysis of conceptual rainfall-runoff modelling performance in different climates. *Journal of Hydrology*, Vol. 238, 231-247.

- Linsley, R.K., Kohler, M.A. and Paulhus, J.L.H. 1982. *Hydrology for engineers*. McGraw-Hill, New York.
- Lister, L.A. 1987. *The erosion surfaces of Zimbabwe*. Zimbabwe Geological Survey Bulletin No. 90, Geological Survey Department, Harare.
- Lohani, V.K., Refsgaard, J.C., Clausen, T., Erlich, M. and Storm, B. 1993. Application of SHE for irrigation-command area studies in India. *Journal of Irrigation and Drainage Engineering*, Vol. 119, No. 1, 34-49.
- Lorup, J.K. 1995. *Iringa soil and water conservation project, Iringa District, Tanzania*. HIMA (Iringa)/DANIDA. Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark, Denmark.
- Lorup, J.K., Refsgaard, J.C. and Mazvimavi, D. 1998. Assessing the effect of land use change on catchment runoff by combined statistical tests and hydrological modelling: Case studies from Zimbabwe. *Journal of Hydrology*, Vol. 205, 147-163.
- Lyne, V. and Hollick, M. 1979. Stochastic time-variable rainfall-runoff modelling. *Institute of Engineers Australian National Conference Publication 79/10*. Institute of Australian Engineers, Canberra, 89-98.
- Manley, R.E. 1978. Simulation of flows in ungauged basins. *Hydrological Sciences Journal*, Vol. 23, No. 1, 85-101.
- Mather, J.R. 1981. Using computed stream flow in watershed analysis. *Water Resources Bulletin*, Vol. 17, No. 3, 474-482.
- Mazvimavi, D. 2002. Watershed degradation and management. IN; Hriji, R., Johnson, P., Maro, P. and Matiza-Chiuta, T. (Ed) *Defining and mainstreaming environmental sustainability in water resources management in southern Africa*. SADC, IUCN, SaRDC, World Bank, Maseru/Harare/Washington DC, 179-203.
- Mazvimavi, D., 1998. Water availability and utilization in Zimbabwe, *Geographical Journal of Zimbabwe*, Vol. 29, 23-36.
- McCartney, M. 1998. *The hydrology of a headwater catchment containing a dambo*. Unpublished PhD Thesis, University of Reading.
- McGarigal, K., Cushman, S. and Stafford, S. 2000. *Multivariate statistics for wildlife and ecology research*. Springer, New York.

- McMahon, T.A. and Mein, R.G. 1978. *Reservoir yield analysis*. Developments in Water Science 9, Elsevier, Amsterdam.
- Meijerink, A.M.J. 1974. *Photo hydrological reconnaissance surveys*. PhD Thesis Free University Amsterdam, and International Institute for Aerial Survey and Earth Sciences Publication, Enschede, The Netherlands.
- Meijerink, A.M.J. 1985. Estimates of peak runoff from hilly terrain with varied lithology. *Journal of Hydrology*, Vol. 77, 227-236.
- Meijerink, A.M.J., de Brouwer, H.A.M., Mannaerts, C.M. and Valenuela, C.R. 1994. *Introduction to the use of geographical information systems for practical hydrology*. Publication No. 23, International Institute for Aerospace Survey and Earth Sciences, Enschede, Netherlands.
- Menenti, T.M. 2000. Evaporation. IN: Schultz, G.A. and Engman, E.T. (Ed) *Remote sensing in hydrology and water management*. Springer Verlag, Berlin, Heidelberg, New York, 157-188.
- Mimikou, M. 1984. Regional relationships between basin size and runoff characteristics. *Hydrological Sciences Journal*, Vol. 29, No. 1, 63-73.
- Mimikou, M. and Kaemaki, S. 1985. Regionalisation of flow duration characteristics. *Journal of Hydrology*, Vol. 82, 77-91.
- Mumeka, A. 1986. Effect of deforestation and subsistence agriculture on runoff on the Kafue headwaters, Zambia. *Hydrological Sciences Journal*, Vol. 31, No. 4, 543-554.
- Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models Part I – a discussion of principles. *Journal of Hydrology*, Vol. 10, 282-290.
- Nathan, R.J. and McMahon, T.A. 1990a. Identification of homogenous regions for purposes of regionalisation. *Journal of Hydrology*, Vol 121, 217-238.
- Nathan, R.J. and McMahon, T.A. 1990b. Evaluation of automated techniques for base flow and recession analyses. *Water Resources Research*, Vol 26, No. 7, 1465-1473.
- Ndiritu, J.G. 1998. *An improved genetic algorithm for rainfall-runoff model calibration*. PhD Thesis, University of Adelaide, Australia.

- Ndiritu, J.G. and Daniel, T.M. 1999. Assessing model calibration adequacy via global optimisation. *Water SA*, Vol. 25, No. 3, 317-326.
- NERC. 1975. *Flood studies reports*. National Environmental Research Council, UK.
- Orr, G., Schraudolph, N. and F. Cummins, 1999. CS\_449 *Neural networks lecture notes*. Williamette University, USA,  
<http://www.williamette.edu/~gorr/classes/cs449/>
- Oyebande, L. 2001. Water problems in Africa-how can sciences help? *Hydrological Sciences Journal*, Vol. 46, No. 6, 947-961.
- Penman, H.L. 1948. Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London, Series A, Mathematics and Physical Sciences*, 120 – 145.
- Penman, H.L. 1956. Estimating evaporation. *Transactions, American Geophysical Union*, Vol. 37, No. 1, 43-50.
- Perrin C., Michel, C. and Andreassin, V. 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, Vol. 242, 275-301.
- Pilgrim, D.H. 1983. Some problems in transferring hydrological relationships between small and large drainage basins and between regions. *Journal of Hydrology*, Vol. 65, 49-72.
- Pitlick, J. 1994. Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. *Journal of Hydrology*, Vol. 158, 219-240.
- Pitman, W.V. 1973. *A mathematical model for generating monthly rivers flows from meteorological data in South Africa*. Report No. 2/73, Hydrological Research Unit, Department of Civil Engineering, University of the Witwatersrand, S. Africa.
- Pitman, W.V. and Middleton, B.J. 1994. *Surface water resources of South Africa, Vol I to VI*. Water Research Commission Report. Pretoria, South Africa.
- Post, D.A. and Jakeman, A.J. 1996. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, Vol. 10, 877-892.

Post, D.A. and Jakeman, A.J. 1999. Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, Vol. 123, 91-104.

Post, D.A., Jones, J.A. and Grant, G.E. 1998. An improved methodology for predicting daily hydrologic response of ungauged catchments. *Environmental Modelling and Software*, Vol. 13, 395-403.

Punj, G. and Stewart, D.W. 1983. Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, Vol. 20, pp 134-148.

Refsgaard, J.C., Seth, S.M., Bathurst, J.C., Erlich, M., Storm, B., Jorgensen, G.H. and Chandra, S. 1992. Application of the SHE to catchments in India Part 1: General Results. *Journal of Hydrology*, Vol. 140, 1-23.

Refsgaard, J.C. and Knudsen, J. 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, Vol. 32, No. 7, 2189-2202.

Refsgaard, J.C. and Storm, B. 1996. Models. In: Abbot, M.B. and Refsgaard, J.C. (Ed) *Distributed hydrological modelling*. Kluwer Academic Publishers, The Netherlands, 41-54.

Riggs, H.C. 1990. Estimating flow characteristics at ungauged sites. In: *Regionalisation in Hydrology, Proceedings of the Ljubljana Symposium, April 1990*, IAHS Publication No. 191.

Ripley, B.D. 1994. Statistical aspects of neural networks. In: Barndorff-Nielsen, O.E., Jensen, J.L. and Kendall, W.S. (Ed) *Networks and Chaos-Statistical and Probabilistic Aspects*. Chapman and Hall, London.

SADC/RRSU. 2000. *Training paper series No. 6 Background information and exercises*. SADC Regional Remote Sensing Project, Harare, Zimbabwe.

Santa Clara, J.M.A. 1980. *A comparison between automatic and trial-and-error calibration of a hydrologic simulation model*. Unpublished MSc dissertation, Department of Civil Engineering, University of Birmingham, UK.

Sarle, W.S. 1996. *The number of clusters*. <http://www.pitt.edu/~wpilib/clusfaq.html>, or [saswss@unx.sas.com](mailto:saswss@unx.sas.com).

Savenije, H.H.G. 1997. Determination of evaporation from a catchment water balance at a monthly time scale. *Hydrology of Earth Systems Sciences*, Vol. 1, No. 1, 93-100.

Savenije, H.H.G. 2001. *Water resources management: Concept and tool*. Lecture Notes, IHE Delft, The Netherlands.

Schulze, R.E. (Ed) 1995. *Hydrology and agrohydrology – A text to accompany the ACRU 3.0 agrohydrological modelling system*. Report No. TT69/95, Water Research Commission, Pretoria, South Africa.

Schulze, R.E. and Kunz, R.P. 1995. Reference potential evaporation. In: Schulze, R.E. (Ed) *Hydrology and agrohydrology: A text to accompany the ACRU 3.00 Agrohydrological modelling systems*. Water Research Commission, Pretoria, Report TT69/95. pp AT4-1 to AT4-38.

Schulze, R.E. and Smithers, J.C. 1995. Procedures to improve and verify streamflow simulations. In: Smithers, J.C. and Schulze, R.E. (Ed) *ACRU Agrohydrological modelling system: User Manual Version 3.0*. Water Research Commission, Pretoria, Report TT70/95, pp AM9-1 to AM9-14.

Sefton, C.E.M. and Howarth, S.M. 1998. Relationship between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology*, Vol. 211, 1-16.

Servat, E. and Dezetter, A. 1993. Rainfall-runoff modelling and water resources assessment in northwestern Ivory Coast. Tentative extension to ungauged catchments. *Journal of Hydrology*, Vol. 148, pp. 231-248.

Seyhan, E. 1977. *A morphometrical analysis in the Ardeche River basin, southern France*. Utrechtse Geografische Studies 7, Geografisch Instituut Rijksuniversiteit Utrecht, The Netherlands.

Seyhan, E. and Keet, B. 1981. *Multivariate statistical analysis (Part I): Application to hydromorphometrical data. (Case Study: Ahyr River Basin, Bolzano, Italy)*. Communications of the Institute of Earth Sciences, Free Reformed University – Amsterdam, Series A. No. 8.

Simmers, I. 1984. A systematic problem-oriented approach to hydrological data regionalisation. *Journal of Hydrology*, Vol. 73, 71-87.

Skidmore, A.K., Turner, B.J., Brinkhof, W. and Knowles, E. 1997. Performance of a neural network: Mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, Vol. 63, No. 5, 501-514.

- Smakhtin, V.Y., Watkins, D.A. and Hughes, D.A. 1995. Preliminary analysis of low-flow characteristics of South African Rivers. *Water SA*, Vol. 21, No. 3, 201-210.
- Smakhtin, V.Y. and Toulouse, M. 1998. Relationships between low-flow characteristics of South African streams. *Water SA*, Vol 24, No. 2, 107-112.
- SPSS Inc. 1999. *SPSS 10.0 Syntax reference guide*. SPSS Inc., Chicago.
- StatSoft Inc. 2001. *STATISTICA (data analysis software system), version 6*. [www.statsoft.com](http://www.statsoft.com).
- Tallaksen, L.M. 1995. A review of baseflow recession analysis. *Journal of Hydrology*, Vol. 165, 349-370.
- Tasker, G.D. 1982. Comparing methods of hydrologic regionalisation. *Water Resources Bulletin*, Vol 18, No 6, pp 965-970.
- Ter Braak, C.J.F. and Prentice, I.C. 1988. A theory of gradient analysis. *Advances in Ecological Research*, Vol. 18, pp 271 – 317.
- Ter Braak, C.J.F. and Smilauer, P. 1998. *CANOCO Reference manual and user's guide to CANOCO for Windows: Software for Canonical Community Ordination (version 4)*. Microcomputer Power, Ithaca, New York.
- Thomas, H.A. 1981. *Improved methods for national water assessment report*. U.S. Water Resources Council, Washington, D.C.
- Thompson, J.G. and Purves, W.D. 1978. *A guide to the soils of Zimbabwe*. Zimbabwe Agricultural Journal Technical Handbook No.3.
- Tucci, C., Silveira, A. and Sanchez, J. 1995. Flow regionalization in the upper Paraguay basin, Brazil. *Hydrological Sciences Journal*, Vol. 40, No. 4, 485-497.
- Vandewiele, G.L., Chong-Yu Xu and Ni-Lar-Win. 1992. Methodology and comparative study of monthly water balance models in Belgium, China and Burma. *Journal of Hydrology*, Vol. 1992, 315-347.
- Vogel, R.M. and Kroll, C.N. 1996. Estimation of baseflow recession constants. *Water Resources Management*, Vol. 10, 303-320.

Wolski, P. 1999. *Application of reservoir modelling to hydrotopes identified by remote sensing*. PhD Thesis, International Institute for Aerospace Survey and Earth Sciences, Publication No. 69, Enschede, The Netherlands

Wood, E.F. 1995. Heterogeneity and scaling land-atmospheric water and energy fluxes in climate systems. In: Feddes, R.A (Ed), *Space and time scale variability and interdependencies in hydrological processes*. Cambridge University Press, 3-19.

Wright, E.P. 1987. Groundwater occurrence and groundwater flow systems in basement aquifers. In: *Groundwater exploration and development in crystalline basement aquifers. Proceedings of the 15-24 June 1987, Harare (Zimbabwe) Workshop Volume 11*. Commonwealth Science Council Technical Paper 273, 251-256.

Wright, E.P. 1997. The hydrogeology of crystalline basement aquifers in African. In: Wright, E.P. and Burgess, W.G. (Ed), *The hydrogeology of crystalline basement aquifers in Africa*. Geological Society Special Publication No. 66, London, 1-28.

Yang, W., Yang, L. and Merchant, J.W. 1994. *AVHRR-derived NDVI and ecoclimatological parameters: Relationships, spatial and temporal variabilities*. <http://www.odyssey.maine.edu/gisweb/spatdb/acsm/ac94087.html>.

Yokoo, Y., Kazama, S., Sawamoto and Nishimura, H. 2001. Regionalisation of lumped water balance model parameters based on multiple regression. *Journal of Hydrology*, Vol. 246, 209-222.

Zrinji, Z. And Burn, D.H. 1994. Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, Vol. 153, 1-21.



## ENGLISH SUMMARY

Sustainable water resources planning and management require hydrological data. These data are needed for example when assessing available water resources, planning, design and management of reservoirs, allocation of water for abstraction and impoundment, and estimating instream environmental water requirements. Most parts of sub-Saharan Africa have inadequate hydrological data, which introduces uncertainty in water resources planning and management. Poorly developed hydrological networks, inaccessibility of some sites, and the decline in technical and financial capacities for hydrological monitoring are the main reasons for inadequate hydrological data. The development of methods for estimating hydrological characteristics of areas with inadequate data is therefore necessary, since planning and management of water resources has to be done for these areas. Thus the main objective of this study is to identify and assess the suitability of statistical methods and conceptual rainfall-runoff models to estimate flow characteristics of ungauged catchments.

A case study approach based on 52 selected catchments in Zimbabwe is used. Zimbabwe experiences a tropical climate characterised by the occurrence of rainfall in one distinct season, mid-November to mid-March, while the rest of the year is dry. This study investigates the feasibility of predicting the following flow characteristics; mean annual runoff, coefficient of variation of annual runoff, base flow index, average number of days per year with no flow, a dimensionless flow duration curve, mean monthly flows, and monthly flow time series. These flow statistics are required for planning, design, and management of various types of water resources projects. Mean annual runoff of selected catchments varies from 38 to 778 mm yr<sup>-1</sup>, and the coefficient of variation of annual runoff is from 55 to 160%. Peak flows occur during the December to March period, and most rivers run dry from August to November.

Catchment characteristics that are investigated for predicting the selected flow statistics are mean annual precipitation, monthly precipitation, average number of rainy days per year, mean annual potential evaporation, elevation, catchment area, drainage density, slope, proportions of a catchment covered by different lithologies, and proportions of a catchment with various land cover types. Mean annual precipitation of the selected catchments varies from 604 to 1770 mm yr<sup>-1</sup>, and mean annual potential evaporation from 1300 to 2000 mm yr<sup>-1</sup>. Crystalline basement complex rocks comprising granite, gneiss, and greenstones are the dominant lithological types. Aquifers only occur where these formations have been weathered, in fractures and fissures. Woodlands are the dominant land cover type, followed by cultivation, and then grasslands.

Chapter 3 investigates prediction of flow statistics using multiple regression methods and neural networks. Mean annual flows are best predicted using multiple regression methods from mean annual precipitation, slope, and proportion of a catchment underlain by granite and gneiss. A neural network with slope, the combined proportion of a catchment with wooded grasslands and grasslands, and proportion of the catchment with Kalahari sands as inputs gives the best prediction for the base flow index. An exponential model describes flow duration curves of the study area. Coefficients of this model can be predicted from the base flow index. However, best predictions of flow duration curves are done with a neural network with slope, mean annual potential evaporation, mean annual rainfall, and drainage density as inputs. A hydrograph of mean monthly flow can be predicted by a neural network with base flow index, proportion of a catchment with both wooded grasslands and grasslands, and slope.

A direct gradient analysis method, redundancy analysis, is used in Chapter 4 to identify catchment characteristics that explain the variation of all flow characteristics. Variations of all flow characteristics are explained by the first and second ordination axes of catchment characteristics. Mean annual precipitation is identified as the most important catchment characteristic, and this accounts for 59% of the explained variance of flow characteristics. Other catchment characteristics identified are proportion of a catchment underlain by rocks belonging to the Umkondo assemblage (quartzite, shale, limestone, and dolerite intrusions), proportion of a catchment with grasslands, and slope.

The selected catchments are classified into clusters with homogenous hydrological responses in Chapter 5. Catchment characteristics identified in Chapter 4 as explaining variations of flow characteristics are used for this classification. A hierarchical clustering method is used. The most appropriate number of clusters was found to be five. Mean annual precipitation is the most distinguishing characteristic of these clusters. Flow characteristics differ significantly between clusters. Therefore, each of the clusters derived is regarded as having homogenous hydrological responses. No significant relationships exist between flow characteristics and catchment characteristics for each of the clusters. Clustering partitioned catchments into sub-sets with narrow ranges in the values of both flow and catchment characteristics. Within these narrow ranges, relationships between flow and catchment characteristics are not identifiable. Therefore clustering does not improve the prediction of flow characteristics from catchment characteristics of the study area. Flow duration curves for each of the clusters can however be used to estimate a flow duration curve of an ungauged catchment provided the cluster to which such a catchment belongs to is established.

Chapter 6 examines the possibility of predicting parameter values of conceptual rainfall-runoff models from catchment characteristics. The models investigated are a) the *abcd* model with four parameters, and b) Pitman model with five parameter requiring calibration. Both models are used on monthly interval, and values of model parameters are calibrated by using manual and automatic optimisation methods. Equifinality affects values of model parameters. Monthly flows simulated by the *abcd* model and Pitman model satisfied goodness-of-fit criteria on 77% and 87% of the catchments respectively. Two of the parameters of the *abcd* model are linearly related to catchment characteristics. However, a neural network has the potential to predict all the four parameters from catchment characteristics. But this model is highly sensitive to minor changes in parameter values and simulations done with parameters predicted by a neural network do not satisfy the goodness-of-fit criteria. Four out of the five parameters of the Pitman model requiring calibration can be predicted from catchment characteristics using a neural network, but the coefficient of determination for the predictions are rather modest (0.5 - 0.6).

The following conclusions are made from this study.

- Most of the flow statistics required for water resources planning can be predicted from catchment characteristics. Neural networks give better predictions than multiple regression methods for most of these statistics.
- The base flow index is an important characteristic of a catchment from which several other flow statistics can be predicted from.
- Mean annual precipitation is the most important catchment characteristics in terms of explaining variations of flow characteristics within the study area.
- Cluster analysis using catchment characteristics enables the identification of catchments with similar hydrological responses, provided this classification uses and weights catchment characteristics that explain the relevant flow characteristics.
- Classification of catchments into clusters with similar hydrological responses does not improve relationships between flow characteristics and catchment characteristics, if clusters derived have narrow ranges in the values of both the flow and catchment characteristics.
- Prediction of parameter values of some conceptual rainfall-runoff models from catchment characteristics may be possible, but the usefulness of these predictions can be affected by the sensitivity of a model to changes in parameter values. Each of the models has some parameters which cannot be related to catchment characteristics, and have to be calibrated. Hence use of these models on an ungauged catchment is not possible because not all parameters can be predicted from catchment characteristics.



## SAMENVATTING

Hydrologische gegevens zijn noodzakelijk voor duurzaam waterbeheer en -planning. Deze gegevens zijn nodig voor het bepalen van de hoeveelheid water dat beschikbaar is, voor het plannen, ontwerpen en beheren van stuwwerken, het vastleggen van water voor direct gebruik en opslag, en voor het schatten van de waterbehoefte van de natuurlijke omgeving. Duurzaam waterbeheer wordt in grote delen van Afrika ten zuiden van de Sahara bemoeilijkt door een gebrek aan hydrologische gegevens. Dit komt door gebrekkige hydrologische meetnetwerken, de ontoegankelijkheid van sommige gebieden, en de verminderende technische en financiële capaciteit om hydrologische processen te monitoren. Omdat ook hier planning en waterbeheer noodzakelijk zijn, zijn er methoden nodig om in gebieden met onvoldoende gegevens hydrologische karakteristieken te kunnen schatten. Het doel van deze studie is om statistische en conceptuele modellen voor de afvoer van regenval te identificeren en te evalueren of deze modellen geschikt zijn om afvoercharacteristieken van "ongemeten" stroomgebieden te schatten.

De studie is gebaseerd op 52 geselecteerde stroomgebieden in Zimbabwe. Zimbabwe kent een tropisch klimaat: het regenseizoen begint half november en eindigt half maart; gedurende de rest van het jaar is het droog. Deze studie onderzoekt de haalbaarheid om de volgende afvoer kenmerken te voorspellen: gemiddelde jaarlijkse afvoer, variatie coëfficiënt, basis afvoer index, gemiddeld aantal dagen zonder afvoer, een dimensieloze duurlijn voor afvoer, gemiddelde maandelijkse afvoer, en reeksen van maandelijkse afvoer. Deze afvoerstatistieken zijn nodig voor planning, ontwerp en beheer van verschillende soorten waterprojecten. De gemiddelde afvoer van de geselecteerde stroomgebieden varieert tussen 38 en 778 mm per jaar, met een variatie coëfficiënt van de jaarlijkse afvoer tussen 55 en 160%. Van december tot maart treedt afvoer als gevolg van storm op, terwijl de meeste rivieren vanaf augustus tot november droogvallen.

Om de geïdentificeerde afvoerstatistieken te voorspellen zijn de volgende kenmerken van stroomgebieden onderzocht: gemiddelde jaarlijkse regenval, maandelijkse regenval, gemiddeld aantal regendagen per jaar, gemiddelde jaarlijkse potentiële verdamping, hoogte, oppervlakte van het stroomgebied, dichtheid van drainage, helling, de verhouding van verschillende lithologieën in een stroomgebied en de verhouding van verschillende typen landgebruik in een stroomgebied. De gemiddelde jaarlijkse regenval varieert tussen 604 en 1.770 mm jaar<sup>-1</sup> in de geselecteerde stroomgebieden, terwijl de gemiddelde jaarlijkse potentiële verdamping varieert tussen 1.300 en 2.000 mm jaar<sup>-1</sup>. Gesteenten samengesteld uit een kristallijne fundering bestaande uit graniet, gneis en groensteen zijn dominante lithologische typen. Aquifers komen alleen voor

waar deze formaties verweerd zijn, dus in breuken en scheuren. Bosgebied is het dominante type landbedekking, gevolgd door akkerbouw en grasland.

Hoofdstuk 3 onderzoekt het voorspellen van afvoerstatistieken met behulp van meervoudige regressie methoden en neurale netwerken. Meervoudige regressie modellen op basis van de gemiddelde jaarlijkse regenval, helling, en de verhouding van graniet en gneis gesteente in een stroomgebied voorspellen de gemiddelde jaarlijkse afvoer het best. Een neuraal netwerk met als afhankelijke variabelen helling, de gecombineerde fractie van een stroomgebied dat bestaat uit bebost grasland en grasland, en de fractie van een stroomgebied dat bestaat uit Kalahari zand, geeft de beste beschrijving van de basis afvoer index. Een exponentieel model kan de durlijnen voor afvoer van de bestudeerde stroomgebieden beschrijven. Model coëfficiënten kunnen worden afgeleid van de basis afvoer index. De beste beschrijving van durlijnen voor afvoer wordt daarentegen behaald met een neuraal netwerk dat de volgende invoer gegevens gebruikt: helling, gemiddelde jaarlijkse potentiële verdamping, gemiddelde jaarlijkse regenval, en de drainage dichtheid. Een verlooplijn voor de afvoer van de gemiddelde maandelijkse neerslag kan worden afgeleid van een neuraal netwerk model, op basis van de basis afvoer index, de verhouding van zowel bebost grasland en grasland in een stroomgebied, en de helling.

Hoofdstuk 4 gebruikt een directe methode om het overschot binnen de stroomgebiedkenmerken te identificeren die de variatie van alle afvoer karakteristieken kunnen beschrijven. Variatie van alle afvoer karakteristieken wordt beschreven door de eerste en tweede coördinaat assen van kenmerken binnen een stroomgebied. De gemiddelde jaarlijkse regenval blijkt het belangrijkste kenmerk van een stroomgebied te zijn, omdat het 59% van de variatie in afvoer karakteristieken kan verklaren. Andere belangrijke stroomgebiedkenmerken zijn de fractie van een stroomgebied met gesteente behorende tot de Umkondo assemblage (kwarts, schalie, kalksteen, en doleriet intrusie gesteente), het deel van een stroomgebied met grasland, en de helling.

Hoofdstuk 5 classificeert de bestudeerde stroomgebieden in clusters met homogene hydrologische response volgens de hiërarchische clustering methode. Hierbij is gebruik gemaakt van de stroomgebiedkenmerken die de variatie in afvoer karakteristieken kunnen beschrijven, zoals geïdentificeerd in hoofdstuk 4. Het beste aantal clusters bleek vijf te zijn. De gemiddelde jaarlijkse regenval is het meest onderscheidende kenmerk van deze clusters. Er is een significant verschil in afvoer karakteristieken tussen de clusters. Daarom mag verondersteld worden dat ieder cluster een homogene hydrologische response heeft. Er bestaat geen significante relatie tussen afvoer karakteristieken en de kenmerken van een stroomgebied voor welk cluster dan ook. Clustering verdeelde de stroomgebieden in groepen waarbij de waarden van zowel de

afvoercharacteristieken als van de kenmerken van het stroomgebied weinig verschilden. Binnen deze nauwe marges zijn er geen relaties identificeerbaar tussen kenmerken van de afvoer en die van het stroomgebied. Clustering geeft daarom geen betere voorspelling van de afvoercharacteristieken op basis van stroomgebiedkenmerken in de onderzochte stroomgebieden. Wel kunnen de duurlijn van de afvoer voor ieder van de clusters gebruikt worden om de duurlijn van de afvoer van een ongemeten stroomgebied te schatten, maar dan moet eerst de cluster waartoe dit stroomgebied behoort vastgesteld zijn.

Hoofdstuk 6 bestudeert de mogelijkheid om parameters van conceptuele modellen voor de afvoer van regenval af te leiden uit de kenmerken van een stroomgebied. De bestudeerde modellen zijn: a) het *abcd* model met vier parameters, en b) het Pitman model met vijf parameters. Beide modellen zijn gebaseerd op een maandelijkse tijdstap. De waarden van de model parameters worden gekalibreerd waarbij gebruik gemaakt wordt van zowel handmatige als automatische optimalisatie methoden. Equifinaliteit kan de waarden van deze parameters beïnvloeden. Maandelijkse afvoer gesimuleerd door het *abcd* en het Pitman model beantwoordde aan geschiktheids criteria voor respectievelijk 77% en 87% van de stroomgebieden. Twee parameters van het *abcd* model zijn lineair gerelateerd aan kenmerken van het stroomgebied. Een neurale netwerk, daarentegen, heeft in principe de mogelijkheid om alle parameters te schatten op basis van de kenmerken van het stroomgebied. Dit model is echter erg gevoelig voor kleine veranderingen in de waarden van parameters. Simulaties met parameters die door een neurale netwerk werden gegenereerd voldeden niet aan de criteria voor geschiktheid. Vier van de vijf parameters van het Pitman model die gekalibreerd moeten worden kunnen geschat worden op basis van de kenmerken van een stroomgebied met een neurale netwerk, maar de determinatie coëfficiënten voor de schattingen waren nogal laag (0.5-0.6).

De volgende conclusies kunnen uit deze studie worden getrokken:

- De meeste afvoercharacteristieken die nodig zijn voor waterbeheer en planning kunnen voorspeld worden op basis van kenmerken van het stroomgebied. De meeste van deze karakteristieken worden beter door neurale netwerken voorspeld dan door meervoudige regressie methoden.
- De basis afvoer index is een belangrijk kenmerk van een stroomgebied, met behulp waarvan verscheidene andere afvoerstatistieken kunnen worden afgeleid.
- De gemiddelde jaarlijkse regenval is het belangrijkste kenmerk van een stroomgebied, in die zin dat het de variaties van afvoercharacteristieken in de bestudeerde stroomgebieden goed kan verklaren.
- Met behulp van een cluster analyse gebaseerd op kenmerken van het stroomgebied kunnen stroomgebieden met een vergelijkbare hydrologisch gedrag geïdentificeerd worden. Dan moet deze classificatie wel gebruik

maken van, en gewichten toekennen aan, die stroomgebiedkenmerken die de relevante afvoercharacteristieken kunnen beschrijven.

- De classificatie van stroomgebieden in clusters met een vergelijkbaar hydrologisch gedrag verbetert de relatie tussen afvoercharacteristieken en stroomgebiedkenmerken niet als binnen deze clusters de waarden van zowel de afvoercharacteristieken als de kenmerken van het stroomgebied weinig verschillen.
- De voorspelling van parameters van sommige conceptuele regenvalafvoer modellen op basis van kenmerken van een stroomgebied is mogelijk, maar de bruikbaarheid van deze voorspellingen kan beïnvloed worden door de gevoeligheid van een model voor veranderingen in parameterwaarden. Elk van de modellen heeft parameters die niet gerelateerd kunnen worden aan stroomgebiedkenmerken, en die wel moeten worden gekalibreerd. Deze modellen kunnen niet in ongemeten stroomgebieden worden gebruikt omdat niet alle parameters afgeleid kunnen worden van de kenmerken van dat stroomgebied.

## **CURRICULUM VITAE**

Dominic Mazvimavi was born on 17 May 1957 at Mvuma in Zimbabwe. He attended Kutama Secondary School (1972-75) and Gokomere High School (1976-77). He was admitted in 1978 to read for the BSc General Degree at the then University of Rhodesia, and obtained a BSc Honours in Geography in 1980. In 1981 he joined the Department of Water Resources Development, Zimbabwe as a Hydrologist. He was awarded a Belgian Fellowship to read for the Post-graduate Diploma in Hydrology from 1983 to 1984 at the Free University of Brussels, and proceeded to do an MSc in Hydrology, (1984-85), at the same university. Upon his return to the Department of Water Resources Development in Zimbabwe in 1985, he was responsible at the national level for undertaking various hydrological analysis for water resources planning, and estimation of flow characteristics of ungauged catchments was a recurring problem.

Dominic Mazvimavi was appointed in 1988 as a Lecturer in the Department of Geography and Environmental Science at the University of Zimbabwe, teaching hydrology, environmental impact assessment, and water resources management. In addition to teaching at this university, he has undertaken several water resources planning and management projects within east and southern Africa, and again estimation of flow characteristics of ungauged catchments is a major problem within this region. He started his PhD studies at the International Institute for Geo-Information and Earth Sciences (ITC), in the Netherlands in 2000, while registered at Wageningen University, the Netherlands. He is currently employed as a Senior Lecturer at the University of Zimbabwe.

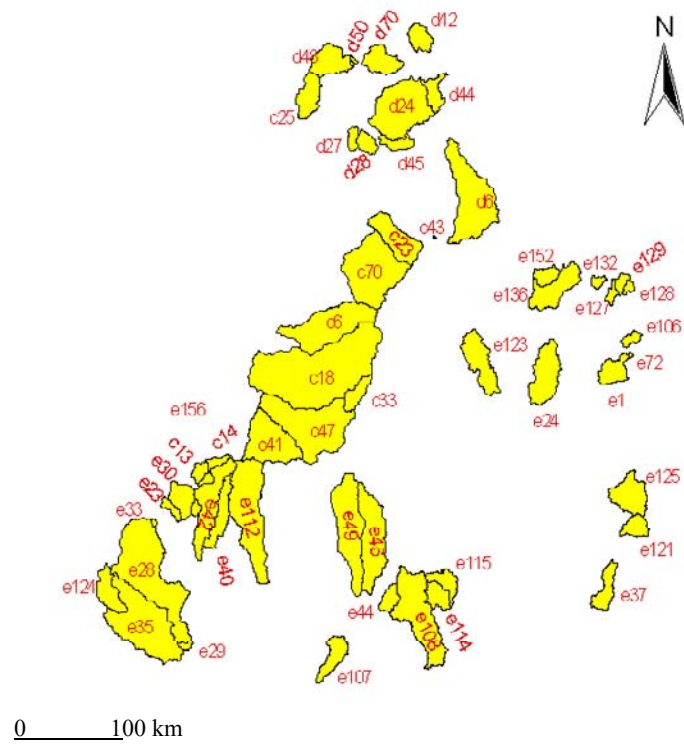


## APPENDIX 1: LIST OF SELECTED CATCHMENTS

Code	Station Name	River	Latitude (°)	Longitude (°E)	Area (km <sup>2</sup> )	Opened
C13	Whitewaters Dam U/S	Kwekwe	-19.367	30.017	150.0	10/26/1950
C14	Whitewaters Dam U/S	Kanuka	-19.400	30.050	101.0	10/27/1950
C18	Dyke G/W	Munyati	-18.817	30.317	2630.0	9/28/1952
C23	Edinburgh	Nyatsime	-18.067	31.067	50.0	9/10/1953
C25	Ayre's Poort	Mukwadzi	-17.433	30.617	282.0	11/3/1953
C33	Chivhu	Sebakwe	-19.083	30.90	194.0	9/9/1954
C41	Sebakwe Dam U/S	Umvumi	-19.067	31.350	855.0	8/15/1955
C43	Grasslands	Manyame Trib	-18.167	31.483	3.50	9/14/1955
C47	Sebakwe Dam U/S	Sebakwe	-19.067	30.350	1550.0	9/26/1956
C6	Ngezi Dam U/S	Ngezi	-18.667	30.517	1040.0	1/1/1948
C70	Beatrice	Mupfure	-18.250	30.767	1210.0	2/20/1969
D24	Arcadia U/S	Pote	-17.367	31.433	1060.0	2/15/1958
D27	Mazowe Dam U/S	Dassura	-17.567	31.00	70.0	11/9/1961
D28	Mazowe Dam U/S	Mazowe	-17.583	31.017	223.0	10/1/1962
D42	Mufurudzi Dam U/S	Mufurudzi	-17.167	31.50	163.0	1/1/1968
D44	Myross	Pote	-17.267	31.550	1220.0	9/30/1968
D6	Mutoko Road Bridge	Shawanhowe	-17.633	31.60	1170.0	10/3/1949
D45	Bally Vaughan	Munenga	-17.67	31.38	137.0	31/10/1968
D48	Mwenje Dam U/S	Wengi	-17.25	30.57	399.0	22/8/1969
D50	Mwenje Dam U/S	Nyamasanga	-17.22	30.98	13.0	31/7/1970
D70	Foothills	Sambi	-17.250	31.283	316.0	12/14/1974
E106	Drennan	Nyambwa	-18.750	32.717	77.70	1/1/1966
E107	Gadziguru Flumes	Musokwesi	-20.57	30.78	249.0	24/9/1966
E108	Manjirenj	Chiredzi	-20.483	31.533	1041.0	9/26/1966
E114	Roswa Tur	Roswa	-20.167	31.60	197.0	1/10/1967
E115	U/S Roswa	Turgwe	-20.167	31.60	223.0	2/15/1967
E120	Nyanyadzi Dam U/S	Piriviri	-19.767	32.667	150.0	9/25/1968
E121	Nyanyadzi Dam U/S	Nyanyadzi	-19.750	32.683	186.0	11/15/1938

E123	Condo Dam U/S	Mare	-19.00	31.917	492.0	7/18/1969
E125	Old Cashel Road Bridge	Umvumvumu	-19.517	32.617	433.0	3/25/1970
E127	Selbourne	Nyamazi	-18.533	32.633	67.30	1/1/1970
E128	Minnehaha	Nyakupinga	-18.467	32.70	46.60	1/1/1970
E129	Minnehaha	Odzi	-18.467	32.683	75.10	1/1/1970
E132	Lisnakea	Umvumira	-18.417	32.517	34.0	1/1/1970
E136	Rusape Dam U/S	Rusape	-18.550	32.117	635.0	12/1/1971
E152	Glenfarg	Chimbi	-18.433	32.183	146.0	2/8/1974
E16	Mharapara	Wengezi	-19.433	32.70	47.0	7/7/1952
E23	Gwenoro D	Nyamadziw	-19.683	29.850	85.50	8/12/1955
E24	Condo U/S	Tsungwesi	-19.050	32.117	557.0	8/31/1955
E28	Mberengwa	Ngezi	-20.367	29.90	1680.0	1/1/1956
E29	Zimunya	Mupudzi	-19.133	32.667	75.10	6/29/1956
E30	Gwenoro D	Runde	-19.683	29.867	254.0	8/26/1956
E33	Standhope Dam	Gwetshetshe	-19.73	29.43	18.0	7/10/1957
E35	Mberengwa Road	Muchingwe	-20.42	29.87	1630.0	27/11/1957
E37	Buffels Drift	Tanganda	-20.10	32.517	246.0	1/29/1958
E40	Mt. Bouga	Littl.Mut	-19.817	30.050	285.0	11/20/1958
E42	Rietfonte	Mutevekwe	-19.883	29.950	648.0	2/11/1959
E44	Mutirikwi Dam U/S	Bevumi	-20.17	31.13	114.0	21/7/1959
E45	Mutirikwi	Mutirikwi	-20.083	31.067	847.0	9/15/1959
E49	Mutirikwi	Popotekwe	-20.117	31.017	1010.0	12/11/1959
E72	Odzani Dam U/S	Nyakawunga	-18.783	32.750	8.40	11/20/1961

## APPENDIX 2: CODES USED TO REFER TO SELECTED CATCHMENTS





### APPENDIX 3: ITC DISSERTATION LIST

1. **Akinyede**, 1990, Highway cost modelling and route selection using a geotechnical information system
2. **Pan He Ping**, 1990, 90-9003757-8, Spatial structure theory in machine vision and applications to structural and textural analysis of remotely sensed images
3. **Bocco Verdinelli, G.**, 1990, Gully erosion analysis using remote sensing and geographic information systems: a case study in Central Mexico
4. **Sharif, M.**, 1991, Composite sampling optimization for DTM in the context of GIS
5. **Drummond, J.**, 1991, Determining and processing quality parameters in geographic information systems
6. **Groten, S.**, 1991, Satellite monitoring of agro-ecosystems in the Sahel
7. **Sharifi, A.**, 1991, 90-6164-074-1, Development of an appropriate resource information system to support agricultural management at farm enterprise level
8. **Zee, D. van der**, 1991, 90-6164-075-X, Recreation studied from above: Air photo interpretation as input into land evaluation for recreation
9. **Mannaerts, C.**, 1991, 90-6164-085-7, Assessment of the transferability of laboratory rainfall-runoff and rainfall - soil loss relationships to field and catchment scales: a study in the Cape Verde Islands
10. **Ze Shen Wang**, 1991: 90-393-0333-9, An expert system for cartographic symbol design
11. **Zhou Yunxian**, 1991, 90-6164-081-4, Application of Radon transforms to the processing of airborne geophysical data
12. **Zuviria, M. de**, 1992, 90-6164-077-6, Mapping agro-topoclimates by integrating topographic, meteorological and land ecological data in a geographic information system: a case study of the Lom Sak area, North Central Thailand
13. **Westen, C. van**, 1993, 90-6164-078-4, Application of Geographic Information Systems to landslide hazard zonation
14. **Shi Wenzhong**, 1994, 90-6164-099-7, Modelling positional and thematic uncertainties in integration of remote sensing and geographic information systems
15. **Javelosa, R.**, 1994, 90-6164-086-5, Active Quaternary environments in the Philippine mobile belt
16. **Lo King-Chang**, 1994, 90-9006526-1, High Quality Automatic DEM, Digital Elevation Model Generation from Multiple Imagery
17. **Wokabi, S.**, 1994, 90-6164-102-0, Quantified land evaluation for maize yield gap analysis at three sites on the eastern slope of Mt. Kenya
18. **Rodriguez, O.**, 1995, Land Use conflicts and planning strategies in urban fringes: a case study of Western Caracas, Venezuela

19. **Meer, F. van der**, 1995, 90-5485-385-9, Imaging spectrometry & the Ronda peridotites
20. **Kufoniya, O.**, 1995, 90-6164-105-5, Spatial coincidence: automated database updating and data consistency in vector GIS
21. **Zambezi, P.**, 1995, Geochemistry of the Nkombwa Hill carbonatite complex of Isoka District, north-east Zambia, with special emphasis on economic minerals
22. **Woldai, T.**, 1995, The application of remote sensing to the study of the geology and structure of the Carboniferous in the Calañas area, pyrite belt, SW Spain
23. **Verweij, P.**, 1995, 90-6164-109-8, Spatial and temporal modelling of vegetation patterns: burning and grazing in the Paramo of Los Nevados National Park, Colombia
24. **Pohl, C.**, 1996, 90-6164-121-7, Geometric Aspects of Multisensor Image Fusion for Topographic Map Updating in the Humid Tropics
25. **Jiang Bin**, 1996, 90-6266-128-9, Fuzzy overlay analysis and visualization in GIS
26. **Metternicht, G.**, 1996, 90-6164-118-7, Detecting and monitoring land degradation features and processes in the Cochabamba Valleys, Bolivia. A synergistic approach
27. **Hoanh Chu Thai**, 1996, 90-6164-120-9, Development of a Computerized Aid to Integrated Land Use Planning (CAILUP) at regional level in irrigated areas: a case study for the Quan Lo Phung Hiep region in the Mekong Delta, Vietnam
28. **Roshannejad, A.**, 1996, 90-9009284-6, The management of spatio-temporal data in a national geographic information system
29. **Terlien, M.**, 1996, 90-6164-115-2, Modelling Spatial and Temporal Variations in Rainfall-Triggered Landslides: the integration of hydrologic models, slope stability models and GIS for the hazard zonation of rainfall-triggered landslides with examples from Manizales, Colombia
30. **Mahavir, J.**, 1996, 90-6164-117-9, Modelling settlement patterns for metropolitan regions: inputs from remote sensing
31. **Al-Amir, S.**, 1996, 90-6164-116-0, Modern spatial planning practice as supported by the multi-applicable tools of remote sensing and GIS: the Syrian case
32. **Pilouk, M.**, 1996, 90-6164-122-5, Integrated modelling for 3D GIS
33. **Duan Zengshan**, 1996, 90-6164-123-3, Optimization modelling of a river-aquifer system with technical interventions: a case study for the Huangshui river and the coastal aquifer, Shandong, China
34. **Man, W.H. de**, 1996, 90-9009-775-9, Surveys: informatie als norm: een verkenning van de institutionalisering van dorp - surveys in Thailand en op de Filippijnen
35. **Vekerdy, Z.**, 1996, 90-6164-119-5, GIS-based hydrological modelling of alluvial regions: using the example of the Kisaföld, Hungary

36. **Pereira, Luisa**, 1996, 90-407-1385-5, A Robust and Adaptive Matching Procedure for Automatic Modelling of Terrain Relief
37. **Fandino Lozano, M.**, 1996, 90-6164-129-2, A Framework of Ecological Evaluation oriented at the Establishment and Management of Protected Areas: a case study of the Santuario de Iguaque, Colombia
38. **Toxopeus, B.**, 1996, 90-6164-126-8, ISM: an Interactive Spatial and temporal Modelling system as a tool in ecosystem management: with two case studies: Cibodas biosphere reserve, West Java Indonesia: Amboseli biosphere reserve, Kajiado district, Central Southern Kenya
39. **Wang Yiman**, 1997, 90-6164-131-4, Satellite SAR imagery for topographic mapping of tidal flat areas in the Dutch Wadden Sea
40. **Saldana-Lopez, Asunción**, 1997, 90-6164-133-0, Complexity of soils and Soilscape patterns on the southern slopes of the Ayllon Range, central Spain: a GIS assisted modelling approach
41. **Ceccarelli, T.**, 1997, 90-6164-135-7, Towards a planning support system for communal areas in the Zambezi valley, Zimbabwe; a multi-criteria evaluation linking farm household analysis, land evaluation and geographic information systems
42. **Peng Wannin**, 1997, 90-6164-134-9, Automated generalization in GIS
43. **Lawas, C.**, 1997, 90-6164-137-3, The Resource Users' Knowledge, the neglected input in Land resource management: the case of the Kankanaey farmers in Benguet, Philippines
44. **Bijker, W.**, 1997, 90-6164-139-X, Radar for rain forest: A monitoring system for land cover Change in the Colombian Amazon
45. **Farshad, A.**, 1997, 90-6164-142-X, Analysis of integrated land and water management practices within different agricultural systems under semi-arid conditions of Iran and evaluation of their sustainability
46. **Orlic, B.**, 1997, 90-6164-140-3, Predicting subsurface conditions for geotechnical modelling
47. **Bishr, Y.**, 1997, 90-6164-141-1, Semantic Aspects of Interoperable GIS
48. **Zhang Xiangmin**, 1998, 90-6164-144-6, Coal fires in Northwest China: detection, monitoring and prediction using remote sensing data
49. **Gens, R.**, 1998, 90-6164-155-1, Quality assessment of SAR interferometric data
50. **Turkstra, J.**, 1998, 90-6164-147-0, Urban development and geographical information: spatial and temporal patterns of urban development and land values using integrated geo-data, Villaviciencia, Colombia
51. **Cassells, C.**, 1998, Thermal modelling of underground coal fires in northern China
52. **Naseri, M.**, 1998, 90-6164-195-0, Characterization of Salt-affected Soils for Modelling Sustainable Land Management in Semi-arid Environment: a case study in the Gorgan Region, Northeast, Iran
53. **Gorte B.G.H.**, 1998, 90-6164-157-8, Probabilistic Segmentation of Remotely Sensed Images

54. **Tenalem Ayenew**, 1998, 90-6164-158-6, The hydrological system of the lake district basin, central main Ethiopian rift
55. **Wang Donggen**, 1998, 90-6864-551-7, Conjoint approaches to developing activity-based models
56. **Bastidas de Calderon, M.**, 1998, 90-6164-193-4, Environmental fragility and vulnerability of Amazonian landscapes and ecosystems in the middle Orinoco river basin, Venezuela
57. **Moameni, A.**, 1999, Soil quality changes under long-term wheat cultivation in the Marvdasht plain, South-Central Iran
58. **Groenigen, J.W. van**, 1999, 90-6164-156-X, Constrained optimisation of spatial sampling: a geostatistical approach
59. **Cheng Tao**, 1999, 90-6164-164-0, A process-oriented data model for fuzzy spatial objects
60. **Wolski, Piotr**, 1999, 90-6164-165-9, Application of reservoir modelling to hydrotopes identified by remote sensing
61. **Acharya, B.**, 1999, 90-6164-168-3, Forest biodiversity assessment: A spatial analysis of tree species diversity in Nepal
62. **Akbar Abkar, Ali**, 1999, 90-6164-169-1, Likelihood-based segmentation and classification of remotely sensed images
63. **Yanuariadi, T.**, 1999, 90-5808-082-X, Sustainable Land Allocation: GIS-based decision support for industrial forest plantation development in Indonesia
64. **Abu Bakr**, Mohamed, 1999, 90-6164-170-5, An Integrated Agro-Economic and Agro-Ecological Framework for Land Use Planning and Policy Analysis
65. **Eleveld, M.**, 1999, 90-6461-166-7, Exploring coastal morphodynamics of Ameland (The Netherlands) with remote sensing monitoring techniques and dynamic modelling in GIS
66. **Yang Hong**, 1999, 90-6164-172-1, Imaging Spectrometry for Hydrocarbon Microseepage
67. **Mainam, Félix**, 1999, 90-6164-179-9, Modelling soil erodibility in the semiarid zone of Cameroon
68. **Bakr, Mahmoud**, 2000, 90-6164-176-4, A Stochastic Inverse-Management Approach to Groundwater Quality
69. **Zlatanova, Z.**, 2000, 90-6164-178-0, 3D GIS for Urban Development
70. **Ottichilo, Wilber K.**, 2000, 90-5808-197-4, Wildlife Dynamics: An Analysis of Change in the Masai Mara Ecosystem
71. **Kaymakci, Nuri**, 2000, 90-6164-181-0, Tectono-stratigraphical Evolution of the Cankori Basin (Central Anatolia, Turkey)
72. **Gonzalez, Rhodora**, 2000, 90-5808-246-6, Platforms and Terraces: Bridging participation and GIS in joint-learning for watershed management with the Ifugaos of the Philippines

73. **Schetselaar, Ernst**, 2000, 90-6164-180-2, Integrated analyses of granite-gneiss terrain from field and multisource remotely sensed data. A case study from the Canadian Shield
74. **Mesgari, Saadi**, 2000, 90-3651-511-4, Topological Cell-Tuple Structure for Three-Dimensional Spatial Data
75. **Bie, Cees A.J.M. de**, 2000, 90-5808-253-9, Comparative Performance Analysis of Agro-Ecosystems
76. **Khaemba, Wilson M.**, 2000, 90-5808-280-6, Spatial Statistics for Natural Resource Management
77. **Shrestha, Dhruba**, 2000, 90-6164-189-6, Aspects of erosion and sedimentation in the Nepalese Himalaya: highland-lowland relations
78. **Asadi Haroni, Hooshang**, 2000, 90-6164-185-3, The Zarshuran Gold Deposit Model Applied in a Mineral Exploration GIS in Iran
79. **Raza, Ale**, 2001, 90-3651-540-8, Object-oriented Temporal GIS for Urban Applications
80. **Farah, Hussein**, 2001, 90-5808-331-4, Estimation of regional evaporation under different weather conditions from satellite and meteorological data. A case study in the Naivasha Basin, Kenya
81. **Zheng, Ding**, 2001, 90-6164-190-X, A Neural - Fuzzy Approach to Linguistic Knowledge Acquisition and Assessment in Spatial Decision Making
82. **Sahu, B.K.**, 2001, Aeromagnetics of continental areas flanking the Indian Ocean; with implications for geological correlation and Gondwana reassembly
83. **Alfestawi, Y.**, 2001, 90-6164-198-5, The structural, paleogeographical and hydrocarbon systems analysis of the Ghadamis and Murzuq Basins, West Libya, with emphasis on their relation to the intervening Al Qarqaf Arch
84. **Liu, Xuehua**, 2001, 90-5808-496-5, Mapping and Modelling the Habitat of Giant Pandas in Foping Nature Reserve, China
85. **Oindo, Boniface Oluoch**, 2001, 90-5808-495-7, Spatial Patterns of Species Diversity in Kenya
86. **Carranza, Emmanuel John**, 2002, 90-6164-203-5, Geologically-constrained Mineral Potential Mapping
87. **Rugege, Denis**, 2002, 90-5808-584-8, Regional Analysis of Maize-Based Land Use Systems for Early Warning Applications
88. **Liu, Yaolin**, 2002, 90-5808-648-8, Categorical Database Generalization in GIS
89. **Ogao, Patrick**, 2002, 90-6164-206-X, Exploratory Visualization of Temporal Geospatial Data using Animation
90. **Abadi, Abdulbaset M.**, 2002, 90-6164-205-1, Tectonics of the Sirt Basin – Inferences from tectonic subsidence analysis, stress inversion and gravity modelling
91. **Geneletti, Davide**, 2002, 90-5383-831-7, Ecological Evaluation for Environmental Impact Assessment

92. **Sedogo, Laurent G.**, 2002, 90-5808-751-4, Integration of Participatory Local and Regional Planning for Resources Management using Remote Sensing and GIS
93. **Montoya, Lorena**, 2002, 90-6164-208-6, Urban Disaster Management: a case study of earthquake risk assessment in Carthago, Costa Rica
94. **Ahmad, Mobin-ud-Din**, 2002, 90-5808-761-1, Estimation of Net Groundwater Use in Irrigated River Basins using Geo-information Techniques: A case study in Rechna Doab, Pakistan
95. **Said, Mohammed Yahya**, 2003, 90-5808-794-8, Multiscale perspectives of species richness in East Africa
96. **Schmidt, Karin**, 2003, 90-5808-830-8, Hyperspectral Remote Sensing of Vegetation Species Distribution in a Saltmarsh
97. **Lopez Binnquist, Citlalli**, 2003, 90-3651-900-4, The Endurance of Mexican Amate Paper: Exploring Additional Dimensions to the Sustainable Development Concept
98. **Huang, Zhengdong**, 2003, 90-6164-211-6, Data Integration for Urban Transport Planning
99. **Cheng, Jianquan**, 2003, 90-6164-212-4, Modelling Spatial and Temporal Urban Growth
100. **Campos dos Santos, Jose Laurindo**, 2003, 90-6164-214-0, A Biodiversity Information System in an Open Data/Metadatabase Architecture
101. **Hengl, Tomislav**, 2003, 90-5808-896-0, PEDOMETRIC MAPPING, Bridging the gaps between conventional and pedometric approaches
102. **Barrera Bassols, Narciso**, 2003, 90-6164-217-5, Symbolism, Knowledge and management of Soil and Land Resources in Indigenous Communities: Ethnopedology at Global, Regional and Local Scales
103. **Zhan, Qingming**, 2003, 90-5808-917-7, A Hierarchical Object-Based Approach for Urban Land-Use Classification from Remote Sensing Data
104. **Daag, Arturo S.**, 2003, 90-6164-218-3, Modelling the Erosion of Pyroclastic Flow Deposits and the Occurrences of Lahars at Mt. Pinatubo, Philippines
105. **Bacic, Ivan**, 2003, 90-5808-902-9, Demand-driven Land Evaluation with case studies in Santa Catarina, Brazil
106. **Murwira, Amon**, 2003, 90-5808-951-7, Scale matters! A new approach to quantify spatial heterogeneity for predicting the distribution of wildlife